Topology of statistical systems : a cohomological approach to information theory

Juan Pablo Vigneaux

université PARIS DIDEROT PARIS 7

Université de Paris

Thèse de doctorat

de l'Université de Paris
Préparée à l'Université Paris Diderot
École doctorale de Sciences Mathématiques de Paris Centre
Institut de Mathématiques de Jussieu–Paris Rive Gauche

Topology of statistical systems
A cohomological approach to information theory

Presentée par
Juan Pablo Vigneaux

Thèse de doctorat de mathématiques
Dirigée par **Daniel Bennequin**

Soutenue publiquement à Paris le 14 juin 2019 devant le jury composé de:

Pr. Samson Abramsky, University of Oxford, Rapporteur.
Pr. Daniel Bennequin, Université Paris Diderot, Directeur de thèse.
Pr. Stéphane Boucheron, Université Paris Diderot, Examinateur.
Pr. Antoine Chambert-Loir, Université Paris Diderot, Président du jury.
Pr. Philippe Elbaz-Vincent, Université Grenoble Alpes, Rapporteur.
Pr. Mikhail Gromov, Institut des Hautes Études Scientifiques, Examinateur.
Pr. Kathryn Hess, École Polytechnique Fédérale de Lausanne, Examinatrice.
Pr. Olivier Rioul, Télécom ParisTech, Examinateur.

On peut donc dire que la notion de topos, dérivé naturel du *point de vue faisceautique* en Topologie, constitue à son tour un élargissement substantiel de la notion d'espace topologique, englobant un grand nombre de situations qui autrefois n'étaient pas considérées comme relevant de l'intuition topologique. Le trait caractéristique de telles situations est qu'on y dispose d'une notion de « localisation », notion qui est formalisée précisément par la notion de site et, en dernière analyse, par celle de topos (via le topos associé au site). Comme le terme de « topos » lui-même est censé précisément le suggérer, il semble raisonnable et légitime aux auteurs du présent Séminaire de considérer que l'objet de la Topologie est l'étude des *topos* (et non des seuls espaces topologiques).

—A. Grothendieck and J.-L. Verdier, SGA 4

Topology of Statistical Systems: A Cohomological Approach to Information Theory

Abstract

This thesis extends in several directions the cohomological study of information theory pioneered by Baudot and Bennequin. We introduce a topos-theoretical notion of statistical space and then study several cohomological invariants. Information functions and related objects appear as distinguished cohomology classes; the corresponding cocycle equations encode recursive properties of these functions. Information has thus topological meaning and topology serves as a unifying framework.

Part I discusses the geometrical foundations of the theory. Information structures are introduced as categories that encode the relations of refinement between different statistical observables. We study products and coproducts of information structures, as well as their representation by measurable functions or hermitian operators. Every information structure gives rise to a ringed site; we discuss in detail the definition of *information cohomology* using the homological tools developed by Artin, Grothendieck, Verdier and their collaborators.

Part II studies the cohomology of discrete random variables. Information functions—Shannon entropy, Tsallis α-entropy, Kullback-Leibler divergence—appear as 1-cocycles for appropriate modules of probabilistic coefficients (functions of probability laws). In the combinatorial case (functions of histograms), the only 0-cocycle is the exponential function, and the 1-cocycles are generalized multinomial coefficients (Fontené-Ward). There is an asymptotic relation between the combinatorial and probabilistic cocycles.

Part III studies in detail the q-multinomial coefficients, showing that their growth rate is connected to Tsallis 2-entropy (quadratic entropy). When q is a prime power, these q-multinomial coefficients count flags of finite vector spaces with prescribed length and dimensions. We obtain a combinatorial explanation for the nonadditivity of the quadratic entropy and a frequentist justification for the maximum entropy principle with Tsallis statistics. We introduce a discrete-time stochastic process associated to the q-binomial probability distribution that generates finite vector spaces (flags of length 2). The concentration of measure on certain *typical subspaces* allows us to extend Shannon's theory to this setting.

Part IV discusses the generalization of information cohomology to continuous random variables. We study the functoriality properties of conditioning (seen as disintegration) and its compatibility with marginalization. The cohomological computations are restricted to the real valued, gaussian case. When coordinates are fixed, the 1-cocycles are the differential entropy as well as generalized moments. When computations are done in a coordinate-free manner, with the so-called *grassmannian categories*, we recover as the only degree-one cohomology classes the entropy and the dimension. This constitutes a novel algebraic characterization of differential entropy.

Keywords: information cohomology, topos theory, information theory, entropy, multinomial coefficients, type theory, sheaves, nonextensive statistics

Topologie des systèmes statistiques : une approche cohomologique à la théorie de l'information

Résumé

Cette thèse étend dans plusieurs directions l'étude cohomologique de la théorie de l'information initiée par Baudot et Bennequin. On introduit une notion d'espace statistique basée sur les topos, puis on étudie plusieurs invariants cohomologiques. Les fonctions d'information et quelques objets associés apparaissent comme des classes de cohomologie distinguées ; les équations de cocycle correspondantes codent les propriétés récursives de ces fonctions. L'information a donc une signification topologique et la topologie sert de cadre unificateur.

La première partie traite des fondements géométriques de la théorie. Les *structures d'information* sont présentées sous forme de catégories qui codent les relations de raffinement entre différents observables statistiques. On étudie les produits et coproduits des structures d'information, ainsi que leur représentation par des fonctions mesurables ou des opérateurs hermitiens. Chaque structure d'information donne lieu à un site annelé ; la *cohomologie de l'information* est introduite avec les outils homologiques développés par Artin, Grothendieck, Verdier et leurs collaborateurs.

La deuxième partie étudie la cohomologie des variables aléatoires discrètes. Les fonctions d'information — l'entropie de Shannon, l'α-entropie de Tsallis, et la divergence de Kullback-Leibler — apparaissent sous la forme de 1-cocycles pour certains modules de coefficients probabilistes (fonctions de lois de probabilité). Dans le cas combinatoire (fonctions des histogrammes), le seul 0-cocycle est la fonction exponentielle, et les 1-cocycles sont des coefficients multinomiaux généralisés (Fontené-Ward). Il existe une relation asymptotique entre les cocycles combinatoires et probabilistes.

La troisième partie étudie en détail les coefficients q-multinomiaux, en montrant que leur taux de croissance est lié à la 2-entropie de Tsallis (entropie quadratique). Lorsque q est une puissance première, ces coefficients q-multinomiaux comptent les drapeaux d'espaces vectoriels finis de longueur et de dimensions prescrites. On obtient une explication combinatoire de la non-additivité de l'entropie quadratique et une justification fréquentiste du principe de maximisation d'entropie quadratique. On introduit un processus stochastique à temps discret associé à la distribution de probabilité q-binomial qui génère des espaces vectoriels finis (drapeaux de longueur 2). La concentration de la mesure sur certains *sous-espaces typiques* permet d'étendre la théorie de Shannon à ce cadre.

La quatrième partie traite de la généralisation de la cohomologie de l'information aux variables aléatoires continues. On étudie les propriétés de fonctorialité du conditionnement (vu comme désintégration) et sa compatibilité avec la marginalisation. Les calculs cohomologiques sont limités aux variables réelles gaussiennes. Lorsque les coordonnées sont fixées, les 1-cocycles sont l'entropie différentielle ainsi que les moments généralisés. Les *catégories grassmanniennes* permettent de traiter les calculs canoniquement et retrouver comme seuls classes de cohomologie de degré 1 l'entropie et la dimension. Ceci constitue une nouvelle caractérisation algébrique de l'entropie différentielle.

Mots-clés: cohomologie de l'information, théorie des topos, théorie de l'information, entropie, coefficients multinomiaux, théorie des types, faisceaux, statistique nonextensive

Remerciements

La rédaction d'une thèse en mathématiques serait une tâche impossible sans l'aide d'un directeur; je veux d'abord remercier le mien, Daniel Bennequin. J'ai eu le privilège d'être son élève pendant ces cinq dernières années et d'apprendre la Géométrie comme un outil pour comprendre le monde : les particules élémentaires, le cerveau, les catastrophes, l'information... toujours à la frontière des idées qui ne sont pas encore complètement mathématisées. Les choses les plus importantes qu'il m'a transmises dépassent les simples connaissances: l'enthousiasme pour tout sujet scientifique concevable, un optimisme sans faille, sa générosité, sa bienveillance.

Je veux exprimer ma gratitude aux rapporteurs, Samson Abramsky et Philippe Elbaz-Vincent, pour avoir accepté cette responsabilité et pour toutes les remarques qui m'ont aidé à améliorer la thèse, ainsi qu'à Stéphane Boucheron, Antoine Chambert-Loir, Mikhail Gromov, Kathryn Hess et Olivier Rioul qui ont bien voulu intégrer le jury en tant qu'examinateurs. Je remercie aussi l'invitation de Samson à présenter mes travaux à l'Université d'Oxford, ce qui m'a permis d'échanger avec son équipe de recherche.

Je tiens également à remercier mes "frères de thèse": Grégoire Sergeant-Perthuis, Olivier Peltre et Alexandre Afgoustidis. Je ne saurais compter combien d'heures j'ai passées avec Olivier et Grégoire: aux cours de Daniel, au séminaire de Géométrie et Physique Mathématique, et surtout au sein de notre groupe de travail sur BP (qui parfois était un peu tendu, car chaque séance durait quatre ou cinq heures); je pense que ma recherche aurait été une expérience beaucoup plus solitaire sans cette collaboration. À son tour, Alexandre a toujours trouvé le temps de passer par mon bureau et me donner plusieurs conseils sur l'avenir.

Je dois exprimer aussi ma reconnaissance à d'autres personnes qui ont contribué directement à cette thèse: à Matilde Marcolli, qui m'a parlé de l'interprétation combinatoire des coefficients multinomiaux lors d'une conversation au CIRM (cela est à l'origine de la partie III de cette thèse); à Jean-Michel Fischer, pour les innombrables discussions mathématiques et ses commentaires à l'introduction de la thèse; à Daniel Juteau, qui m'a aidé aussi à corriger le manuscrit.

Merci à ceux qui se sont intéressés par mes travaux pendant la thèse: Pierre Baudot, qui est aussi à l'origine de cette topologie de l'information; Robert Niven, pour ses conseils et ses lettres de recommandation; Muriel Livernet et François Métayer, qui m'ont invité à parler au séminaire "Catégories supérieures, polygraphes et homotopie"; Paul-André Melliès, qui m'a mis en contact avec Samson; Olgica Milenkovic et Marcelo Firer, qui ont fait plusieurs commentaires à ma théorie de Shannon généralisée; Alexander Barg, qui m'a encouragé à envoyer mon article à IEEE Transactions on Information Theory; Hussein Mourtada, qui a toujours suivi l'état de ma thèse; Max Leyton, qui m'a proposé de faire une exposé à l'Université de Talca; Jürgen Jost et Nihat Ay, qui m'ont donné l'opportunité de parler à l'Institut Max Planck de Leipzig. Ma recherche a aussi bénéficié d'une correspondance relativement régulière avec Felipe Pérez, Darrick Lee et Fernando Rosas.

Je remercie aussi Frédéric Hélein, Sergei Barannikov, Christian Brouder et tous les participants du séminaire de Géométrie et Physique Mathématique, ainsi que Mattia Cavicchi, Réda Chaneb, Julien Page et (encore) Jean-Michel, membres du groupe de travail sur la cohomologie de Galois. Ces collectifs ont rendu beaucoup plus intéressantes et agréables mes années de doctorat.

La thèse n'est que le résultat d'une très longue formation mathématique et c'est sûrement impossible de remercier à tous ceux qui y ont joué un rôle capital. Je dois au moins mentionner quelques personnes. Rolando Rebolledo a été mon mentor pendant plusieurs années et m'a toujours encouragé à prendre le chemin de la recherche: il m'a appris non seulement la modélisation de systèmes ouvertes, mais aussi qu'on peut agir politiquement en tant qu'ingénieur mathématique; il est une personne d'une générosité extraordinaire et une curiosité remarquable vers l'ensemble des connaissances humaines. Pablo Marquet m'a enseigné un peu d'écologie et m'a accueilli dans son laboratoire quand j'ai dû rester au Chili plus que prévu; René Schott m'a appris aussi plusieurs choses dans le cadre de notre collaboration sur les marches aléatoires en écologie. J'ai eu aussi la possibilité de faire un stage de recherche à l'INRIA: je remercie tous les membres du groupe RAP (Philippe Robert, Christine Fricker, Henning Sulzbach...) et surtout Nicolas Broutin, qui m'a encadré et m'a aidé à développer mon intuition probabiliste. Je tiens aussi à remercier David Cozmar, Matías López et Carolina Urzúa, qui ont été mes compagnons d'études (et mes amis) pendant plusieurs années.

Je remercie tous mes collègues doctorant·e·s de Paris 7, ainsi que tou·te·s mes ami·e·s à Paris : sans eux, j'aurais eu du mal à survivre ici, à 11 mil kilomètres du Chili. J'ai peur de mentionner quelques un·e·s, parce que forcément plusieurs personnes importantes ne vont pas apparaître dans la liste. (Vous savez que je vous aime.)

Finalmente, quiero agradecer a mi familia, así como a todos los amigos que dejé en Chile, por su amor, comprensión y apoyo. Si no fuera por ustedes, nada de esto hubiese sido posible.

Contents

List of Figures

List of Tables

Introduction

0.1 Axiomatic characterizations of entropy

In his seminal paper on the mathematical foundations of communication [78], Claude Shannon proposed the following axiomatic characterization for a 'measure of choice':

> Suppose we have a set of possible events whose probabilities of occurrence are $p_1, p_2, \cdots, p_n$. These probabilities are known but that is all we know concerning which event will occur. Can we find a measure of how much "choice" is involved in the selection of the event or of how uncertain we are of the outcome?
>
> If there is such measure, say $H(p_1, p_2, \cdots, p_n)$, it is reasonable to require of it the following properties:
>
> 1. H should be continuous in the p_i.
> 2. If all the p_i are equal, $p_i = \frac{1}{n}$, then H should be a monotonic increasing function of n. Whith equally likely events there is more choice, or uncertainty, when there are more possible events.
> 3. If a choice be broken down into two successive choices, the original H should be the weighted sum of the individual values of H. The meaning of this is illustrated [in Figure 1]. At the left we have three possibilities each with probabilities $\frac{2}{3}, \frac{1}{3}$. The final results have the same probabilities as before. We require, in this special case, that
>
> $$H(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) = H(\frac{1}{2}, \frac{1}{2}) + \frac{1}{2} H(\frac{2}{3}, \frac{1}{3})$$
>
> The coefficient $\frac{1}{2}$ is because this second choice only occurs half the time.

Then he proved:

Theorem 0.1 (Shannon)**.** *The only H satisfying the assumptions above is of the form*

$$H(p_1, ..., p_n) = -K \sum_{i=1}^{n} p_i \log p_i, \tag{0.1}$$

where K is a positive constant.

The function H is called entropy, sometimes preceded by the names of Shannon or Gibbs.

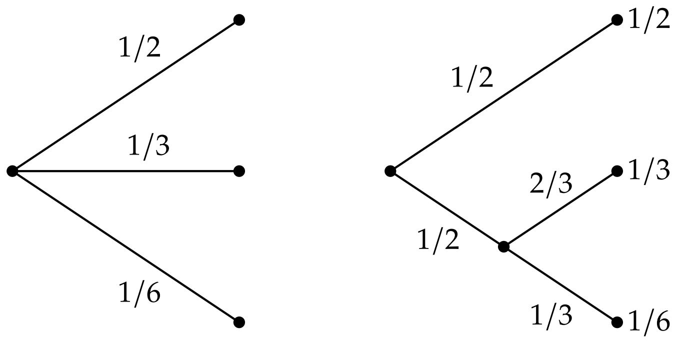

Figure 1: Decomposition of a choice from three possibilities. Figure 6 in [78].

These and analogous axioms have been taken traditionally as intuitive, natural or expected properties of information itself or other related concepts. However, several questions could (and should) be raised: What does "natural" mean in this context? Which "intuitions" are we talking about? What is our mental picture of information or even probabilities? These questions are rather philosophical, but they can be clarified answering first a mathematical one: What is the specific role that these axioms play in information theory? Maybe the situation is similar to axiomatizations of geometry: Euclid's fifth postulate makes sense and seems quite natural, but meaningful theories can be built without it.

Information theory and statistical mechanics make use of several generalizations of entropy that do not satisfy the third axiom. For example, in 1967 Jan Havrda and František Charvát [38] introduced the structural α-entropy: for $\alpha > 0$, $\alpha \neq 1$, it is given by the formula

$$S_\alpha(p_1, \dots, p_n) = K_\alpha \left(\sum_{i=1}^{n} p_n^\alpha - 1 \right), \tag{0.2}$$

where K_α is some constant chosen in such a way that $S_\alpha \to S_1 := H$ when $\alpha \to 1$ [22]. It was characterized as the only function satisfying certain axioms, including one analogous to axiom 3 above where the probabilities in front of each H were raised to the power α, cf. (0.11). These α-entropies were popularized in physics by Constantino Tsallis [87,88], that uses them as a foundation for nonextensive statistical mechanics. As a consequence, the most common name for S_α is Tsallis α-entropy.

It is important to elucidate the relation between these axiomatic characterizations of information functions and their applications in information theory and statistical mechanics, that usually involve combinatorial or probabilistic reasoning. Shannon says:

> [Theorem 0.1], and the assumptions required for its proof, are in no way necessary for the present theory. It is given chiefly to lend a certain plausibility to some of our later definitions. The real justification of this definitions, however, will reside in their implications.

We could be tempted to dismiss completely the axiomatic approach, which is not so far from Kolmogorov's position [52, p. 42]:

> The deduction of limiting theorems of the type indicated above has been carried out in many remarkable papers [...] We feel that much must still

> be done in this direction. [...] Since by its original nature "information" is not a scalar magnitude, we feel that the axiomatic study allowing to characterize [the mutual information] $I(\xi, \eta)$ uniquely (or uniquely characterize the entropy $H(\xi)$) by means of simple formal properties have in this respect a lesser importance. Here we believe that the situation is similar to the one related to the fact that of all the methods for obtaining a foundation of the normal distribution law for errors proposed by Gauss, today we favour the methods based on limiting theorems for sum of large numbers of small summands. Other methods (for example, the method based on the arithmetical mean) only explains why no other distribution law of errors can be as pleasant and convenient as the normal one, but does not answer the question why the normal distribution law so often appears in real problems. Exactly in the same way, the beautiful properties of the expressions $H(\xi)$ and $I(\xi, \eta)$ cannot explain why they suffice in many problems for finding a complete solution (at least from the asymptotical point of view).

In any case, the algebraic properties of entropy and the normal law are remarkable enough to be mentioned, and there is more to them than beauty. For example, Linnik [59] gave a (complicated) information-theoretic proof of the Central Limit Theorem (later improved and clarified, see [9]), so the limiting properties of the normal distribution are not independent from the fact that it maximizes the entropy when the mean and variance are fixed. In turn, this information-maximization property is a bridge to establish connections between the remarkable algebraic properties of entropy on the one hand and of normal distributions on the other.

It seems to us that the *connections* between the different points of view on entropy—algebraic, probabilistic, combinatorial, dynamical—are still poorly understood. This constitutes a first motivation to introduce the categorical framework of Section 0.5 as a formalism general enough to integrate different theories.

0.2 Functional equations

Let us come back to Shannon's axiomatization: it refers to a set of possible events, but those events do not appear in the notation. It would be more precise to introduce a random variable (or "random object" or "experiment") X, that can take values in a finite set E_X. A probability law is a function $P : E_X \to [0,1]$ such that $\sum_{x \in E_X} P(x) = 1$.[1] Then the Shannon entropy associated to the random variable X and the law P is[2]

$$S_1[X](P) = -K \sum_{x \in E_X} P(x) \log P(x). \tag{0.3}$$

[1]Strictly speaking, the probability law is a measure ρ on the algebra of subsets of E_X such that $\int_{E_X} \mathrm{d}\rho = 1$ and P is its density with respect to the counting measure, but here we are identifying both things, as is customary. The distinction becomes important in Section 0.8.

[2]We reserve the character H for (co)homology, so we follow other texts in statistical mechanics (e.g. Tsallis' book [88]) denoting the entropy by S.

Entropy already appears in this form in Shannon's paper, just after the axiomatization. Similarly, the α-entropy is

$$S_\alpha[X](P) = K_\alpha \left(\sum_{x \in E_X} P(x)^\alpha - 1 \right). \tag{0.4}$$

Consider now two random variables X and Y, valued in sets E_X and E_Y, respectively. The joint measurement, represented as a vector (X, Y), is also a random variable (joint variable), whose possible values belong to certain subset E_{XY} of $E_X \times E_Y$. Suppose that a law P on E_{XY} is given; following Shannon, it can be represented by a tree as in Fig. 2-(a). The probability of observing $X = x$ is computed as the sum of all the outputs of (X, Y) that contain x in the first component: $X_*P(x) := P(X = x) = \sum_{(x,y) \in E_{XY}} P(x, y)$. This defines a probability X_*P on E_X, usually called "marginal distribution". Instead of measuring directly (X, Y) one could measure first X, which constitutes a first random "choice", and then update the probabilities on E_{XY} taking into account this first result: probability represents uncertain knowledge about the "true value" of (X, Y), that is updated each time a measurement is performed. Only the pairs in $\{(x_0, y)\}_{y \in Y}$ are compatible with $X = x_0$; therefore, the conditional probability law $P|_{X=x_0} : E_X \to [0, 1]$, that represents the uncertainty after obtaining the result $X = x_0$, is defined by

$$P|_{X=x_0}(x, y) = \begin{cases} \frac{P(x,y)}{X_*P(x_0)} & \text{if } x = x_0 \\ 0 & \text{otherwise} \end{cases}. \tag{0.5}$$

This iterated choice/measurement can in turn be represented as a tree, e.g. Fig. 2-(b). According to Shannon's third axiom, the "chain rule"

$$S_1[(X, Y)](P) = S_1[X](X_*P) + \sum_{x \in E_X} X_*P(x) S_1[Y](Y_*P|_{X=x}) \tag{0.6}$$

must hold. Similarly, if the measurement of Y is performed first, we obtain another tree, Fig. 2-(c), that entails the equation

$$S_1[(X, Y)](P) = S_1[Y](Y_*P) + \sum_{y \in E_Y} Y_*P(y) S_1[X](X_*P|_{Y=y}) \tag{0.7}$$

It turns out that (0.6) and (0.7), taken as a system of functional equations with measurable unknowns $S_1[X]$, $S_1[Y]$, and $S_1[(X, Y)]$, imply that each of these functions must be the corresponding Shannon entropy defined by (0.3). This is true even for the situation pictured in Fig. 2, that is evidently the simplest possible choice that can be broken down in two different ways (see Proposition 3.10).

Tverberg [89] was the first to deduce a simple functional equation from the third axiom, nowadays called "fundamental equation of information theory":

$$f(x) + (1 - x) f\left(\frac{y}{1 - x}\right) = f(y) + (1 - y) f\left(\frac{x}{1 - y}\right), \tag{0.8}$$

where $f : [0, 1] \to \mathbb{R}$ is an unknown function, and $x, y \in [0, 1]$ are such that $x + y \in [0, 1]$. The only symmetric, measurable solutions of this equation are the real

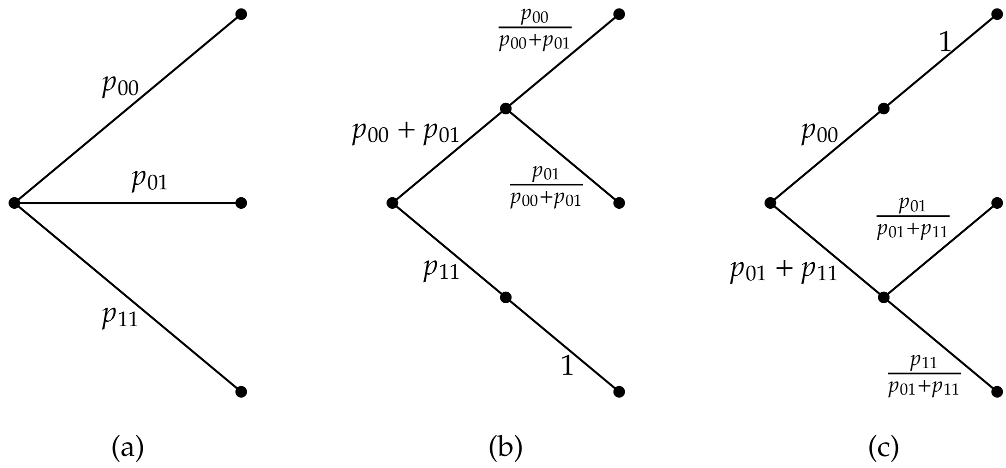

Figure 2: Different goupings when $E_X = E_Y = \{0,1\}$ and $E_{XY} = \{(0,0),(0,1),(1,1)\}$. We denote by p_{ij} the probability of the point $(i,j) \in E_{XY}$. In (b) and (c), the probabilities to the left are the marginals X_*P and Y_*P, respectively, and those to the right are the conditional laws on the appropriate subset of E_{XY}.

multiples of $s_1(x) := -x\ln(x) + (1-x)\ln(x)$ [57]. Daroczy [25] proposed a similar equation solved by the α-entropy $s_\alpha(x) := x^\alpha + (1-x)^\alpha - 1$. [3]

The situation is already quite striking, because Shannon's characterization asks for an infinite number of conditions—certain equations for *any* set of events and *any* possible grouping of them—along with strong regularity of the functions H (an infinite family indexed by n), and actually just one set, two different groupings, and measurability of the unknowns are enough to reach the same conclusion. Maybe this would only be a nice mathematical curiosity, if these chain-rule-like functional equations did not accept a much deeper interpretation. Let us define,[4] for any probabilistic functional $P \mapsto f(P)$, a new functional $X.f$ given by

$$(X.f)(P) := \sum_{x\in E_X} X_*P(X) f(Y_*P|_{X=x}). \tag{0.9}$$

in order to rewrite (0.6) as

$$0 = X.S_1[Y] - S_1[(X,Y)] + S_1[X]. \tag{0.10}$$

The notation is meant to suggest an action of random variables on probabilistic functionals, and in fact the equality $Z.(X.f) = (Z,X).f$ holds. There is an strong resemblance between (0.10) and a cocycle equation in group cohomology. Baudot and Bennequin [10] formalized this analogy: it is possible to use the general constructions of homological algebra to recover the equations (0.10) as cocycle conditions in an adapted cohomology theory that they called *information cohomology*. Since the entropy is the their only solution, the argument constitutes an alternative characterization of entropy. This description is not *axiomatic*, but *algebro-geometrical*: it has a

[3]For a detailed historical introduction and a comprehensive treatment of the subject, up to 1975, see the book by Aczél and Daróczy [3].

[4]More precise definitions are given in Section 0.7.

meaning in the context of *topos theory*, developed by Artin, Grothendieck, Verdier and their collaborators [4,5] as a tool for algebraic geometry.[5]

The α-entropy satisfies a deformed chain rule,

$$S_\alpha[(X,Y)](P) = S_\alpha[X](X_*P) + \sum_{x\in E_X} (X_*P(x))^\alpha S_\alpha[Y](Y_*P|_{X=x}). \tag{0.11}$$

As an extension of Baudot and Bennequin's results, we prove that $S_\alpha[\cdot]$ is the only family of measurable real-valued functions that satisfy these deformed functional equations for generic pairs of random variables and probabilities, up to a multiplicative constant K (Proposition 3.13). This is in turn connected to Daroczy's fundamental equation (0.8) cf. Chapter 5.

If the random variables X, Y represent the possibles states of two systems (e.g. physical systems or random sources of messages) that are supposed to be independent in the usual probabilistic sense, $P(x,y) = X_*P(x)Y_*P(y)$, then

$$S_1[(X,Y)](P) = S_1[X](X_*P) + S_1[Y](Y_*P). \tag{0.12}$$

This property of Shannon entropy is called additivity. Under the same assumptions, Tsallis entropy is nonadditive;[6] when $K = 1$,

$$S_\alpha[(X,Y)](P) = S_\alpha[X](X_*P) + S_\alpha[Y](Y_*P) - (\alpha-1)S_\alpha[X](X_*P)S_\alpha[Y](Y_*P). \tag{0.13}$$

As we already said, this property is problematic from the point of view of heuristic justifications for information functions, that assume as "intuitive" that the amount of information given by two independent events should be computed as the sum of the amounts of information given by each one separately.

0.3 Entropy in combinatorics

Before introducing the cohomological formalism, it is important to provide evidence of the mathematical and practical relevance of these generalized entropies. As Kolmogorov and Shannon said, entropy is justified mainly by the its implications and its relation to certain limiting theorems.

One of the most fundamental results in this direction relates Shannon entropy S_1 to the growth of multinomial coefficients. More precisely: given a probability law $(p_1,\dots,p_s)$,[7]

$$\lim_n \frac{1}{n}\ln\binom{n}{p_1n,\dots,p_sn} = -\sum_{i=1}^s p_i\ln p_i =: S_1(p_1,\dots,p_s). \tag{0.14}$$

---

[5]Cathelineau [16] was the first to find a cohomological interpretation for the fundamental equation (0.8): an analogue of it is involved in the computation of the homology of SL_2 over a field of characteristic zero, with coefficients in the adjoint action; however, this result was not explicitly connected to Shannon entropy or information theory. The first published work in this direction is a note by Kontsevich (reproduced as an appendix in [27]), that introduces $H_p(x) = \sum_{k=1}^{p-1}\frac{x^k}{k}$ as "a residue modulo p" of entropy, being the only continuous map $f:\mathbb{Z}/p\mathbb{Z}\to\mathbb{Z}/p\mathbb{Z}$ that verifies $f(x) = f(1-x)$ and an equation equivalent to (0.8). He proves that a related function defines a cohomology class in $H^2(F,F)$, for $F=\mathbb{R}$ or $\mathbb{Z}/p\mathbb{Z}$. Several works connected to motives or polylogarithms have emphasized the role of the fundamental equation, for instance [12,17,27,28].

[6]Originally, this was called *nonextensivity*, which explains the name 'nonextensive statistical mechanics'.

[7]The reader can either assume that each p_i is rational and the limit is taken over the n that verify $p_in\in\mathbb{Z}$, or that the multinomial coefficients are defined for complex arguments using the Γ-function.

This is a first indication of the relevance of entropy in communication theory: it approximates the counting of words of length n, made of s different symbols, each appearing with probability p_i. In statistical mechanics, it counts the number of "configurations" of n particles, when s energy levels are available and p_i is the proportion of particles with energy E_i, for $i \in \{1, ..., s\}$.

The multinomial coefficients have an interesting q-analog. Given an indeterminate q, define the q-integers $\{[n]_q\}_{n\in\mathbb{N}}$ by $[n]_q := (q^n - 1)/(q-1)$ and the q-factorials by $[n]_q! := [n]_q[n-1]_q \cdots [1]_q$. The q-multinomial coefficients are

$$\begin{bmatrix} n \\ k_1, ..., k_s \end{bmatrix}_q := \frac{[n]_q!}{[k_1]_q! \cdots [k_s]_q!}, \tag{0.15}$$

where $k_1, ..., k_s, n \in \mathbb{N}$ satisfy $\sum_{i=1}^s k_i = n$. When q is a prime power, the coefficients in (0.15) counts the number of flags of vector spaces $V_1 \subset V_2 \subset ... \subset V_n = \mathbb{F}_q^n$ such that $\dim V_i = \sum_{j=1}^i k_j$ (here $\mathbb{F}_q$ denotes the finite field of order q); we refer to the sequence $(k_1, ..., k_s)$ as the *type* of the flag. In particular, the q-binomial coefficient $\begin{bmatrix} n \\ k \end{bmatrix}_q := \begin{bmatrix} n \\ k, n-k \end{bmatrix}_q$ counts vector subspaces of dimension k in $\mathbb{F}_q^n$.

In Section 6.2 we study in detail the asymptotic behavior of the q-multinomial coefficients. In particular, we obtain the following limit.

Proposition 0.2. *Given a probability law $(p_1, ..., p_s)$,*

$$\lim_n \frac{2}{n^2} \log_q \begin{bmatrix} n \\ p_1 n, ..., p_s n \end{bmatrix}_q = 1 - \sum_{i=1}^s p_i^2 =: S_2(p_1, ..., p_s). \tag{0.16}$$

The function S_2 is Tsallis 2-entropy, also known as quadratic entropy*.*

There is a connection between these combinatorial results and the algebraic characterizations of entropy. To see this, remark first that the multinomial coefficients satisfy some multiplicative identities, that can be interpreted as a recursive enumeration. For instance,

$$\binom{n}{k_1, k_2, k_3} = \binom{n}{k_1 + k_2}\binom{k_1 + k_2}{k_1} \tag{0.17}$$

means that the number of words of length n composed of three different symbols, say $\{a_1, a_2, a_3\}$, and such that a_i appears k_i times *equals* the number of words of length n composed of two different symbols, say $\{a_{12}, a_3\}$, such that a_{12} appears $k_1 + k_2$ times *multiplied by* the number of ways of replacing the symbols a_{12} with k_1 symbols a_1 and k_2 symbols a_2 (which reduces to count certain sub-words of length $k_1 + k_2$). In the same spirit, if $P = (p_0, p_1)$, $Q = (q_0, q_1)$ are two probability laws on $\{0, 1\}$, then

$$\binom{n}{p_0q_0n, p_0q_1n, p_1q_0n, p_1q_1n} = \binom{n}{p_0n}\binom{p_0n}{p_0q_0n}\binom{p_1n}{p_1q_0n}. \tag{0.18}$$

Applying $\frac{1}{n}\ln(-)$ to both sides and taking the limit $n \to \infty$, we recover the additive relation (0.12):

$$S_1(p_0q_0, p_0q_1, p_1q_0, p_1q_1) = S_1(p_0, p_1) + S_1(q_0, q_1). \tag{0.19}$$

Equation (0.18) remains valid for the q-multinomial coefficients, but in this case one should apply $\lim_n \frac{2}{n^2}\log_q(-)$ to both sides to obtain

$$\begin{aligned}\frac{2}{n^2}\log_q \begin{bmatrix} n \\ p_0q_0n, p_0q_1n, p_1q_0n, p_1q_1n\end{bmatrix}_q &= \frac{2}{n^2}\log_q \begin{bmatrix} n \\ p_0n\end{bmatrix}_q \\ &+ p_0^2\frac{2}{(p_0n)^2}\log_q \begin{bmatrix} p_0n \\ p_0q_0n\end{bmatrix}_q + p_1^2\frac{2}{(p_1n)^2}\log_q \begin{bmatrix} p_1n \\ p_1q_0n\end{bmatrix}_q, \end{aligned} \tag{0.20}$$

which in the limit gives

$$\begin{aligned} S_2(p_0q_0, p_0q_1, p_1q_0, p_1q_1) &= S_2(p_0, p_1) + p_0^2 S_2(q_0, q_1) + (1-p_0)^2 S_2(q_0, q_1) \\ &= S_2(p_0, p_1) + S_2(q_0, q_1) - S_2(p_0, p_1)S_2(q_0, q_1). \end{aligned} \tag{0.21}$$

Thus, asymptotically, the number of flags $V_{00} \subset V_{01} \subset V_{10} \subset V_{11} = \mathbb{F}_q^n$ of type $(p_0q_0n, p_0q_1n, p_1q_0n, p_1q_1n)$ can be computed in terms of the number of flags $W_0 \subset W_1 = \mathbb{F}_q^n$ of type (p_0n, p_1n) and those flags $W_0' \subset W_1' = \mathbb{F}_q^m$ of type (q_0m, q_1m)—where m can take the values p_0n or p_1n—through this nonadditive formula. This example is discussed in Chapter 6, followed by combinatorial justification of a maximum 2-entropy principle.

More generally, for any sequence $D = \{D_i\}_{i\geq 1}$ such that $D_1 = 1$, define $[n]_D!$ as $D_nD_{n-1}\cdots D_1$ and the corresponding Fontené-Ward multinomial coefficients by

$$\begin{Bmatrix} n \\ k_1, \dots, k_s\end{Bmatrix}_D := \frac{[n]_D!}{[k_1]_D!\cdots[k_s]_D!}. \tag{0.22}$$

In Section 4.5, we prove that, for every $\alpha > 0$, there is a generalized multinomial coefficient asymptotically related to the corresponding α-entropy.

Proposition 0.3. *If $D_n = q^{n^{\beta-1}-1}$, for any $q > 0$, then*

$$\begin{Bmatrix} n \\ p_1n, \dots, p_sn\end{Bmatrix}_D = \exp\left\{ n^\beta \frac{\ln q}{\beta} S_\beta(p_1, \dots, p_s) + o(n^\beta)\right\}. \tag{0.23}$$

The Fontené-Ward multinomial coefficients clearly satisfy the same multiplicative relations as the usual multinomial coefficients, so their logarithms (properly normalized) are connected in the limit $n \to \infty$ to the deformed chain rule (0.11), as we already showed for the particular cases $D_n = n$ (linked to $\alpha = 1$, Shannon entropy) and $D_n = \frac{q^n-1}{q-1}$ (linked to $\alpha = 2$, quadratic entropy). Even more, these generalized coefficients are also the solution of certain functional equations, that corresponds to a cocycle condition in a *combinatorial* version of information cohomology.

Our motivation to study the q-multinomial coefficients was to understand better the generalized information functions of degree α. Tsallis used them as the foundation of nonextensive statistical mechanics, a generalization of Boltzmann-Gibbs statistical mechanics that was expected to describe well some systems with long-range correlations. It is not completely clear which kind of systems follow these "generalized statistics".[8] There is extensive empirical evidence about the pertinence of

[8]Tsallis says: "...the entropy to be used for thermostatistical purposes would be not universal but would depend on the system or, more precisely, on the nonadditive universality class to which the system belongs" [88, p. xii].

the predictions made by nonextensive statistical mechanics [88]. However, very few papers address the microscopical foundations of the theory (for instance, [37,44,73]). The considerations above prove that S_2 can be understood as counting microstates given by flags, still following Boltzmann intuitions [79]. We are not aware of any physical situation where flags are a natural representation for the microstates. However, we were able to come up with a version of information theory where messages correspond to vector spaces (flags of length 2) and to extend Shannon's asymptotic insights to this setting. This makes plausible an eventual application of these results in statistical mechanics.[9]

0.4 A q-deformation of Shannon's theory

We explain here how the combinatorial ideas introduced in the previous section, combined with a new probabilistic construction, can be used to build a generalization of Shannon's theory where messages are vector spaces. Formula (0.16) already suggests that the quadratic entropy plays an essential role in it. Most results summarized here and in the following section were published as [92].

The asymptotic expansion of the multinomial coefficients (0.14) is of great importance in Shannon's theory. Consider a random source that emits at time $n \in \mathbb{N}$ a symbol Z_n in $S_Z = \{z_1, ..., z_s\}$, each Z_n being an independent realization of a S_Z-valued random variable Z with law P. A *message* (at time n) corresponds to a random sequence $(Z_1, ..., Z_n)$ taking values in S_Z^n with law $P^{\otimes n}(z_1, ..., z_n) := P(z_1) \cdots P(z_n)$. The *type* of a sequence $\mathbf{z} \in S_Z^n$ is the probability distribution on S_Z given by the relative frequency of appearance of each symbol in it; for example, when $S_Z = \{0, 1\}$, the type of a sequence with k ones is $(1 - \frac{k}{n})\delta_0 + \frac{k}{n}\delta_1$. By the law of large numbers, a "typical sequence" is expected to have type P, and therefore its probability $P^{\otimes n}(\mathbf{z})$ is approximately $\prod_{z \in S_Z} P(z)^{nP(z)} = \exp\{-nS_1[Z](P)\}$. The cardinality of the set of sequences of type P is $\binom{n}{P(z_1)n, ..., P(z_s)n} \approx \exp\{nS_1[Z](P)\}$. This implies, according to Shannon, that "it is possible for most purposes to treat the long sequences as though there were just 2^{Hn} of them, each with a probability 2^{-Hn}" [78, p. 397]. This result is known nowadays as the *asymptotic equipartition property* (AEP), and can be stated more precisely as follows:

Theorem 0.4 (AEP, [20, Th. 3.1.2]). *Given $\varepsilon > 0$ and $\delta > 0$, it is possible to find $n_0 \in \mathbb{N}$ and sets $\{A_n\}_{n \geq n_0}$, $A_n \subset S_Z^n$, such that, for every $n \geq n_0$,*

1. *$P^{\otimes}(A_n^c) < \varepsilon$, and*
2. *for every $\mathbf{z} \in A_n$,*

$$\left| \frac{1}{n} \ln(P^{\otimes n}(\mathbf{z})) - S_1[Z](P) \right| < \delta. \tag{0.24}$$

The size of A_n is optimal: if $s(n, \varepsilon)$ denotes the minimal cardinality of a set $B_n \subset S_Z^n$ that accumulates probability $1 - \varepsilon$,

$$s(n, \varepsilon) = \min\{\, |B_n| \mid B_n \subset S_Z^n \text{ and } \mathbb{P}((Z_1, ..., Z_n) \in B_n) \geq 1 - \varepsilon \,\},$$

then

$$\lim_n \frac{1}{n} \ln |A_n| = \lim_n \frac{1}{n} \ln s(n, \varepsilon) = S_1[Z](P). \tag{0.25}$$

[9] Cf. Jaynes' emblematic article on the connection between information theory and statistical mechanics [42].

The set A_n can be defined to contain all the sequences whose type Q is close to P, in the sense that $\sum_{z\in S_Z} |Q(z) - P(z)|$ is upper-bounded by a small quantity; this is known as *strong typicality* (see [23, Def. 2.8]).

Similar conclusions can be drawn for a system of n independent physical particles, the state of each one being represented by a random variable Z_i; in this case, the vector $(Z_1, \dots, Z_n)$ is called a *configuration*. The set A_n can be thought as an approximation to the effective phase space ("reasonable probable" configurations) and the entropy as a measure of its size, see [43, Sec. V]. In both cases—messages and configurations—the underlying probabilistic model is a process $(Z_1, \dots, Z_n)$ linked to the multinomial distribution, and the AEP is a result on measure concentration around the expected type.

We propose a new type of statistical model, such that a message at time n (or a configuration of n particles) is represented by a flag of vector spaces $V_1 \subset V_2 \subset \dots \subset V_s = \mathbb{F}_q^n$. In the simplest case ($s = 2$) a message is just a vector space V in $\mathbb{F}_q^n$. While the type of a sequence is determined by the number of appearances of each symbol, the type of a flag is determined by its dimensions or equivalently by the numbers $(k_1, \dots, k_s)$ associated to it; by abuse of language, we refer to $(k_1, \dots, k_s)$ as the type. The cardinality of the set of flags $V_1 \subset \dots \subset V_s \subset \mathbb{F}_q^n$ that have type $(k_1, \dots, k_s)$ is $\begin{bmatrix} n \\ k_1,\dots,k_s \end{bmatrix}_q \sim C(q) q^{n^2 S_2(k_1/n,\dots,k_s/n)/2}$, where $C(q)$ is an appropriate constant.

To build a correlative of Shannon's theory of communication, it is fundamental to have *a probabilistic model for the source.* In our case, this means a random process $\{F_i\}_{i\in\mathbb{N}}$ that produces at time n a flag F_n that would correspond to a generalized message. We can define such process if we restrict our attention to the binomial case ($s = 2$). This is the content of Chapter 7, summarized in the next paragraph.

Let θ be a positive real number, and let $\{X_i\}_{i\geq 1}$ be a collection of independent random variables that satisfy $X_i \sim \mathrm{Ber}(\theta q^{i-1}/(1+\theta q^{i-1}))$, for each i. We fix a sequence of linear embeddings $\mathbb{F}_q^1 \hookrightarrow \mathbb{F}_q^2 \hookrightarrow \dots$, and identify $\mathbb{F}_q^{n-1}$ with its image in $\mathbb{F}_q^n$. The n-dilations of a subspace w of $\mathbb{F}_q^{n-1}$ are defined as

$$\mathrm{Dil}_n(w) := \{\, v \subset \mathbb{F}_q^n \mid \dim v - \dim w = 1, w \subset v \text{ and } v \not\subset \mathbb{F}_q^{n-1} \,\}. \tag{0.26}$$

We then define a stochastic process $\{V_i\}_{i\geq 0}$ such that each V_i is a vector subspace of $\mathbb{F}_q^i$, as follows: $V_0 = 0$ and, at step n, the dimension of V_{n-1} increases by 1 if and only if $X_n = 1$; in this case, V_n is picked at random (uniformly) between all the n-dilations of V_{n-1}. When $X_n = 0$, one sets $V_n = V_{n-1}$. We prove that, for any subspace $v \subset \mathbb{F}_q^n$ of dimension k,[10]

$$\mathbb{P}\,(V_n = v) = \frac{\theta^k q^{k(k-1)/2}}{(-\theta; q)_n}. \tag{0.27}$$

This implies that

$$\mathbb{P}\,(\dim V_n = k) = \begin{bmatrix} n \\ k \end{bmatrix}_q \frac{\theta^k q^{k(k-1)/2}}{(-\theta; q)_n}, \tag{0.28}$$

which appears in the literature as q-binomial distribution [48].

For the multinomial process, the probability $P^{\otimes n}$ concentrates on types close to P i.e. appearances close to the expected value $nP(z)$, for each $z \in S_Z$. In the case of V_n, the probability also concentrates on a restricted number of dimensions (types).

[10] We use the q-Pochhammer symbols $(a; q)_n := \prod_{i=0}^{n-1}(1 - aq^i)$, with $(a; q)_0 = 1$.

Table 1: Correspondence between Shannon's information theory in the case of memoryless Bernoulli sources and our q-deformed version for vector spaces. The number q is supposed to be a primer power; $\xi \in [0,1]$ and $\theta > 0$ are parameters.

Concept	**Shannon case**	**q-case**
Message at time n (n-message)	Word $w \in \{0,1\}^n$	Vector subspace $v \subset \mathbb{F}_q^n$
Type	Number of ones	Dimension
Number of n-messages of type k	$\binom{n}{k}$	$\begin{bmatrix} n \\ k \end{bmatrix}_q$
Probability of an n-message of type k	$\xi^k(1-\xi)^{n-k}$	$\dfrac{\theta^k q^{k(k-1)/2}}{(-\theta;q)_n}$

In fact, it is possible to prove an analog of the asymptotic equipartition partition property in this setting. Its statement requires the introduction of a left-continuous function $\Delta : [0,1) \to \mathbb{N}$ that depends on the asymptotic behavior of the Grassmannian process (see Section 7.4), whose discontinuity points have Lebesgue measure zero.

Theorem 0.5 (Generalized AEP). *Let $\{V_n\}_{n\in\mathbb{N}}$ be a Grassmannian process, $\delta \in (0,1)$ an arbitrary number, and $\varepsilon > 0$ such that $p_\varepsilon := 1 - \varepsilon$ is a continuity point of Δ. Define $A_n = \bigcup_{k=0}^{a_n} \mathrm{Gr}(n-k,n)$ as the smallest set of the form $\bigcup_{k=0}^{m} \mathrm{Gr}(n-k,n)$ such that $\mathbb{P}(V_n \in A_n^c) \leq \varepsilon$. Then, there exists $n_0 \in \mathbb{N}$ such that, for every $n \geq n_0$,*

1. *$A_n = \bigcup_{k=0}^{\Delta(p_\varepsilon)} \mathrm{Gr}(n-k,n)$;*
2. *for any $v \in A_n$ such that $\dim v = k$,*

$$\left| \frac{\log_q(\mathbb{P}(V_n = v)^{-1})}{n} - \frac{n}{2} S_2(k/n) \right| \leq \delta. \tag{0.29}$$

The size of A_n is optimal, up to the first order in the exponential: let $s(n,\varepsilon)$ denote $\min\{\, |B_n| \mid B_n \subset \mathrm{Gr}(n)$ and $\mathbb{P}(V_n \in B_n) \geq 1 - \varepsilon \,\}$; then

$$\lim_n \frac{1}{n} \log_q |A_n| = \lim_n \frac{1}{n} \log_q s(n,\varepsilon) = \lim_n \frac{n}{2} S_2(\Delta(p_\varepsilon)/n) = \Delta(p_\varepsilon). \tag{0.30}$$

The set A_n correspond to the "typical subspaces", in analogy with the typical sequences introduced above. We then deduce an optimal compression rate for Grassmannian sources (Section 8.3); the definition of an optimal coding scheme remains an open problem, whose solution is probably connected to Schubert calculus.[11]

The arguments used in this section can be taken as a reproducible scheme towards the generalization of conventional ("additive") information theory and statistical mechanics. They do not only require a combinatorial interpretation of a generalized

[11]Other interesting questions concern (1) the extension of the Grassmannian process in order to generate flags of arbitrary length; (2) the possible relations between this q-deformation of Shannon theory and subspace codes, where messages are coded as vector spaces, and (3) the transmission of subspaces over noisy channels.

multinomial coefficient, but also the introduction of a stochastic process that generates the objects that are counted by them. For the moment we do not know how to define a process that generates arbitrary flags, nor are aware of other combinatorial interpretations apart of those already mentioned.

0.5 Information structures

We turn now to the promised categorical formalization leading to information cohomology. The departing point is the introduction of information structures, which encode the relations between different random variables and their possible results. The content of Part I, summarized here and in the following section, can be found in [91].

Traditionally, random variables are defined as measurable functions $X : (\Omega, \mathfrak{F}) \to (\mathbb{R}^n, \mathfrak{B}(\mathbb{R}))$, where $(\Omega, \mathfrak{F})$ corresponds to certain sample space. This space has some technical importance, but its arbitrary nature must not be forgotten. For example, suppose that we want to study the tossing of a die. As we know in advance that there are six possible outcomes, we could choose $\Omega = \{1, 2, 3, 4, 5, 6\}$, but we could equally well define $\Omega = [1, 7)$ with the algebra of sets generated by $\{[j, j+1) : j = 1, ..., 6\}$ and associate each interval $[j, j+1)$ with the result "observe j dots". This example is discussed by Doob [26], who adds

> [The natural objection that the model is] needlessly complicated, is easily refuted. In fact, 'needlessly complicated' models cannot be excluded, even in practical studies. How a model is setup depends on what parameters are taken as basic. [...] If, however, the die is thought as a cube to be tossed into the air and let fall, the position of the cube in the space is determined by six parameters, and the most natural Ω might well be considered to be the twelve dimensional space of initial positions and velocities.

Random variables, also called "observables", are introduced to model measurements subject to unavoidable variability. In general, we know the possible outcomes of our experiments, already constrained by the limitations of our measuring devices (including our own perceptors). Like Gromov in [34], we want to approach measurements from a categorical point of view, describing directly the relations between them.[12] A sample space $(\Omega, \mathfrak{F})$ just serves as a model that allow us to treat observables as concrete measurable functions.[13] To suppose that such space exists reflects in fact a

[12]Apart from the articles by Baudot and Bennequin [10], and Gromov [34], we have been at least indirectly influenced by other works that apply categorical ideas to probability, statistics or information theory. One of the most important is Čensov's book [90], that introduces the categorical language to study the equivariance of optimal statistical decision rules (for inference) from a geometrical point of view; he calls this "geometrical statistics". An extension of these results is the subject of a recent monograph by Ay et al. [7]. Categories also play an important role in Holevo's book on probabilistic and statistical aspects of quantum theory [40], where they are used to formalize the notion of "measurement." Baez and Fritz derived in [8] a new characterization of relative entropy studying an appropriate category of finite probability spaces. See also their related work with Leinster [58]. With respect to those works, the main novelty of this thesis resides in its topos-geometric approach and the computation of cohomological invariants related to information theory. A related cohomology theory was introduced by Abramsky and his collaborators [1] in order to detect contextuality and paradoxes; this problem is detailed below.

[13]Such classical model does not always exists. For example, when observables do not commute,

belief in *reality*, a unified underlying structure that accounts for all our observations.[14] The probabilistic properties of the observables should not depend on the particular model that has been chosen. As Terence Tao [84] says, "sample [probability] spaces (and their attendant structures) will be used to *model* probabilistic concepts, rather than to actually *be* the concepts themselves."

In Doob's example, there is a variable X taking six possible values

$$E_X := \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\},$$

and this variable can be implemented as a category of partitions for multiple sets Ω. We can represent the situation by a diagram

$$E_X \to \{*\}. \tag{0.31}$$

The set $\{*\}$ represents the output of constant random variable, corresponding to "certainty." Given two variables X and Y, a measurable map $\pi_{YX} : (E_X, \mathfrak{E}_X) \to (E_Y, \mathfrak{E}_Y)$ between their possible outcomes says that the event of observing Y in specified measurable set B_Y is compatible only with the event $\{X \in \pi^{-1}(B_Y)\}$. As a consequence, the determination of $X = x$ implies that $Y = \pi_{YX}(y)$; more generally, any probability on $(E_X, \mathfrak{B}_X)$ induces a unique probability on $(E_Y, \mathfrak{B}_Y)$.[15] For example, we could add a phase space $(S, \mathfrak{S})$ of initial positions and velocities, and relate through a map π every output of the die with the initial conditions leading to it according to Newtonian mechanics, which gives the diagram

$$(S, \mathfrak{S}) \xrightarrow{\ \pi\ } (E_X, 2^{E_X}) \longrightarrow \{*\} \tag{0.32}$$

where 2^{E_X} denotes the atomic σ-algebra. Normally, such map π is surjective, but the system could be constrained in such a way that some outputs are not attainable from any feasible initial conditions.

In practice, the position P and the velocity V are measured by different devices, valued in certain sets $(E_P, \mathfrak{P})$ and $(E_V, \mathfrak{V})$, and the phase space $(S, \mathfrak{S})$ corresponds to certain subset of the product $(E_P \times E_V, \mathfrak{P} \otimes \mathfrak{V})$ given by the possible values of the joint measurement, denoted here $P \wedge V$. We should also take into account the joint measurement of X and P, denoted $X \wedge V$, as well as $X \wedge V$; the measurement $X \wedge P \wedge V$ is equivalent to $P \wedge V$ under the hypothesis that initial conditions determine

the underlying space E is defined as a Hilbert space and observable values appear as the spectrum of hermitian operators on E.

[14] This is not an exaggeration: we shall see that some information structures are noncontextual (Section 1.4) and therefore violate generalized Bell inequalities [2, Prop. III.1] (our noncontextual structures are the possibilistically noncontextual structures there). In this sense, some information structures are incompatible with hidden-variable models.

[15] Mathematically, we study probabilities according to Kolmogorov's axiomatization [53]. Epistemologically, we regard them as a quantification of uncertain knowledge adapted to the rules of plausible reasoning, a viewpoint elaborated by Jaynes in [41]; his Appendix A discusses the compatibility between these two perspectives. By no means we limit ourselves to frequentist scenarios.

the output of X. At the level of variables, we have the dependencies

$$\begin{tikzcd}
 & X \wedge P \arrow[r] \arrow[dr] & P \arrow[dr] & \\
P \wedge V \arrow[ur] \arrow[rr] \arrow[dr] & & X \arrow[r] & \mathbf{1} \\
 & X \wedge V \arrow[r] \arrow[ur] & V \arrow[ur] &
\end{tikzcd} \tag{0.33}$$

where $\mathbf{1}$ is again certainty, and an arrow $A \to B$ means that the value taken by B can be deduced from the value taken by A; in this sense, A is more refined. We have analogous relations at the level of values,

$$\begin{tikzcd}
 & (E_X \times E_P, 2^{E_X} \otimes \mathfrak{P}) \arrow[r] \arrow[dr] & (E_P, \mathfrak{P}) \arrow[dr] & \\
(S, \mathfrak{S}) \arrow[ur] \arrow[rr] \arrow[dr] & & (E_X, 2^{E_X}) \arrow[r] & \{*\} \\
 & (E_X \times E_V, 2^{E_X} \otimes \mathfrak{V}) \arrow[r] \arrow[ur] & (E_V, \mathfrak{V}) \arrow[ur] &
\end{tikzcd} \tag{0.34}$$

where $\otimes$ is the product of σ-algebras, and the arrows represent measurable maps. Let $\mathbf{S}$ be the free category generated by the diagram (0.33); the diagram (0.34) can be seen as the image of $\mathbf{S}$ under an appropriate functor $\mathscr{E}$.

We introduce now a general definition in this sense.

Definition 0.6. An *information structure* is a pair $(\mathbf{S}, \mathscr{E})$, where $\mathbf{S}$ ('the variables') is a small category such that

1. $\mathbf{S}$ has a terminal object, denoted $\mathbf{1}$;
2. $\mathbf{S}$ is a skeletal partially ordered set (poset).[16]
3. for objects $X, Y, Z \in \operatorname{Ob} S$, if $Z \to X$ and $Z \to Y$, then the categorical product $X \wedge Y$ exists;[17]

and $\mathscr{E}$ is a conservative[18,19] covariant functor ('the values') from $\mathbf{S}$ into the category **Meas** of measurable spaces, $X \mapsto \mathscr{E}(X) = (E(X), \mathfrak{B}(X))$, that satisfies

4. $E(\mathbf{1}) \cong \{*\}$, with the trivial σ-algebra $\mathfrak{B}(\mathbf{1}) = \{\emptyset, E(\mathbf{1})\}$;
5. for every $X \in \operatorname{Ob} \mathbf{S}$ and any $x \in E(X)$, the σ-algebra $\mathfrak{B}(X)$ contains the singleton $\{x\}$;
6. for every diagram $X \xleftarrow{\pi} X \wedge Y \xrightarrow{\sigma} Y$ the measurable map $E(X \wedge Y) \hookrightarrow E(X) \times E(Y), z \mapsto (x(z), y(z)) := (\pi_*(z), \sigma_*(z))$ is an injection.

These information structures generalize those introduced by Baudot and Bennequin in [10]: the objects of the latter were partitions of some set Ω. The new

[16]Being a poset means that, for any objects A and B, $\operatorname{Hom}(A, B)$ has at most one element. The poset is skeletal if it has no isomorphisms different from the identities: if $A \neq B$ and $A \to B$, then $B \not\to A$.

[17]This could be called "*conditional* meet semi-lattice".

[18]Given a functor $\mathscr{F} : \mathbf{S} \to \mathbf{Sets}$, we denote its value at $X \in \operatorname{Ob} \mathbf{S}$ by $\mathscr{F}(X)$ or $\mathscr{F}_X$.

[19]Conservative means that, if $\mathscr{E}(f)$ is an isomorphism, then f is an isomorphism. Since $\mathbf{S}$ is skeletal, this condition implies that, for every arrow $\pi : X \to Y$ such that $X \neq Y$, the measurable map $\pi_* := \mathscr{E}(\pi) : E(X) \to E(Y)$ is not a bijection.

ones are sufficiently general to provide a common ground for classical and quantum information; the general algebraic constructions of Chapter 2 are valid for all the known versions of information cohomology. Besides, the morphisms between them are defined more naturally.

Definition 0.7. Given two structures $(\mathbf{S}, \mathscr{E}), (\mathbf{S}', \mathscr{E}')$, a morphism $\phi = (\phi_0, \phi^\#) : (\mathbf{S}, \mathscr{E}) \to (\mathbf{S}', \mathscr{E}')$ between them is a functor $\phi_0 : \mathbf{S} \to \mathbf{S}'$ and a natural transformation $\phi^\# : \mathscr{E} \Rightarrow \mathscr{E}' \circ \phi_0$, such that

1. $\phi_0(\mathbf{1}) = \mathbf{1}$;
2. if $X \wedge Y$ exists, then $\phi_0(X \wedge Y) = \phi_0(X) \wedge \phi_0(Y)$;
3. for each $X \in \operatorname{Ob} S$, the component $\phi^\#_X : \mathscr{E}(X) \to \mathscr{E}'(\phi_0(X))$ is measurable.

Given $\phi : (\mathbf{S}, \mathscr{E}) \to (\mathbf{S}', \mathscr{E}')$ and $\psi : (\mathbf{S}', \mathscr{E}') \to (\mathbf{S}'', \mathscr{E}'')$, their composition $\psi \circ \phi$ is defined as $(\psi_0 \circ \phi_0, \psi^\# \circ \phi^\# : \mathscr{E} \Rightarrow \mathscr{E}'' \circ \psi_0 \circ \phi_0)$.

Proposition 1.8 establishes that this category has finite products and coproducts.

Given an information structure $(\mathbf{S}, \mathscr{E})$, one can define a presheaf (contravariant functor) of monoids that maps $X \in \operatorname{Ob} \mathbf{S}$ to the the set $\mathscr{S}_X := \{ Y \in \operatorname{Ob} \mathbf{S} \mid X \to Y \}$ equipped with the product $(Y, Z) \mapsto YZ := Y \wedge Z$, where $\wedge$ denotes the meet (categorical product) in $\mathbf{S}$; an arrow $X \to Y$ is mapped to the inclusion $\mathscr{S}_Y \hookrightarrow \mathscr{S}_X$. The associated presheaf of induced algebras is $X \mapsto \mathscr{A}_X := \mathbb{R}[\mathscr{S}_X]$. Probabilities are also a functor $\Pi : \mathbf{S} \to \mathbf{Sets}$, that associates to each $X \in \operatorname{Ob} S$ the set $\Pi(X)$ of measures on $\mathscr{E}(X)$ such that $\int_{E_X} \mathrm{d}\rho = 1$. Each arrow $f : X \to Y$ induces a measurable map $\mathscr{E}(f) : \mathscr{E}(X) \to \mathscr{E}(Y)$, and $\Pi(f) : \Pi(X) \to \Pi(Y)$ is defined to be the push-forward of measures: for every $B \in \mathfrak{B}_Y$,

$$(\Pi(f)(\rho))(B) = \rho(\mathscr{E}(f)^{-1}(B)). \tag{0.35}$$

This operation is called *marginalization*. We write f_* or Y_* instead of $\Pi(f)$ (normally the source of the arrow is clear from context); of course, this is compatible with the notations of Section 0.2.

If I is any set, let $\mathbf{\Delta(I)}$ be the category of its finite subsets, with arrows $A \to B$ indicating that $B \subset A$. Fpr us, a *simplicial subcomplex* of $\mathbf{\Delta(I)}$ is a full subcategory $\mathbf{K}$ such that, for any given object of $\mathbf{K}$ ('a cell'), all its subsets are also objects of $\mathbf{K}$ ('faces'). Associate to each vertex $\{i\}$ of $\mathbf{K}$ a measurable set $(E_i, \mathfrak{B}_i)$, and to every other $A \in \operatorname{Ob} \mathbf{K}$ the measurable set $\mathscr{E}(A) := (E_A, \mathfrak{B}(E_A))$, where $E_A = \prod_{i \in A} E_i$ and $\mathfrak{B}(E_A)$ is the corresponding Borel σ-algebra, that is also the product σ-algebra $\bigotimes_{i \in A} \mathfrak{B}_i$ (see Section 9.1). This defines a functor $\mathscr{E}$ if we associate to every arrow in $\mathbf{K}$ the corresponding canonical projector. The pair $(\mathbf{K}, \mathscr{E})$ is a *simplicial information structure*.[20]

The simplicial structures are flexible enough to cover *graphical models* [67], [63, Ch. 9]: the Ising model and its generalizations, Markov fields, Bayesian networks... In these examples there is a set of random variables $\{X_i\}_{i \in I}$ and a distinguished collection $\mathfrak{C}$ of subsets that represent some sort of local interactions; we define $\mathbf{K}$ as the smallest simplicial subcomplex of $\mathbf{\Delta(I)}$ that contains $\mathfrak{C}$. The local information can come as a subfunctor $\mathscr{Q}$ of Π that associates to each variable $(X_{i_1}, \dots, X_{i_m})$ represented by $\{i_1, \dots, i_m\} \in \operatorname{Ob} \mathbf{K}$ a set of possible probability laws (joint distributions). In fact, the functor $\mathscr{E}$ itself can be seen as a representation of admissible local configurations.

[20] It is worth noting that *abelian* (co)presheaves on simplicial structures are cellular (co)sheaves in the sense of [24].

Remark that in the simplicial case $\mathscr{E}$ and Π can be extended easily to the whole category $\mathbf{\Delta}(\mathbf{I})$, that should be thought as a larger geometrical space that contains $\mathbf{K}$. The so-called *marginal problem* asks when a section s of Π on $\mathbf{K}$, i.e. an element $s \in \Gamma(\mathbf{K}, \Pi)$,[21] can be extended to a section $\tilde{s}$ of Π on $\mathbf{\Delta}(\mathbf{I})$ that coincides with s over each $A \in \operatorname{Ob} \mathbf{K}$; this extension $\tilde{s}$ is meant to represent a joint state of all the X_i compatible with the known local interactions. (This problem is analog to that of extensions of holomorphic functions in complex analysis.) It is well known that such extension does not always exist: this correspond to *frustration* in physics [62, 67], to *contextuality* in quantum mechanics, and paradoxes in logic [1, 2, 30].

Section 1.4 studies the conditions under which the variables of an information structure can be represented as measurable functions on a unique sample space $(\Omega, \mathfrak{F})$, supposing that the nerve of $\mathbf{S}$ has finite dimension and each space E_X in the image of $\mathscr{E}$ is finite; such representation is a *classical model*. A necessary condition is the existence of a global section $s(x)$ of E compatible with any given value $x \in E_X$ assigned to any variable X. We determine in which case the collections of compatible measurements, elements of $\lim_{\mathbf{S}} E$, constitute themselves a classical model.

0.6 Topoi and cohomology

In modern treatments of algebraic geometry, category theory and related subjects, the appropriate notion of "space" is a *topos*. The initial motivation comes from topology: from the category $\mathbf{Sh}(T)$ of sheaves on a topological space T, one can recover its lattice of open sets, provided that each point is determined by its open neighborhoods [65, Sec. I.2]. Sheaves are particular **Sets**-valued functors on the category of open sets of T, such that a global section is uniquely determined by compatible local data prescribed on any covering of T. Grothendieck and his collaborators realized that this still makes sense if the category of open sets is replaced by any other category; actually, he introduced a notion of topology (site) on an arbitrary category, which is defined in terms of its arrows. The category of sheaves thus obtained is called a (Grothendieck) topos. The motivation behind is captured in the following quotation from the SGA 4 [4, IV 0.4], the book that introduced the theory:

> So we can say that the notion of topos, a natural derivative of the *sheafy point of view* in Topology, constitutes a substantial enlargement of the notion of topological space, encompassing a large number of situations that in the past were not considered to depend on the topological intuition. The characteristic feature of such situations is that there is a notion of "localization", which is formalized precisely by the notion of site and, in the final analysis, by that of topos (via the topos associated with the site). As the term "topos" itself is meant to suggest, it seems reasonable and legitimate to the authors of this Seminar to consider that the object of Topology is the study of *topos* (and not only topological spaces).

Lawvere and his collaborators introduced a more general notion of topos: these are

[21] A section $s \in \Gamma(\mathbf{K}, \Pi)$ is an element of the set $\operatorname{Hom}_{[\mathbf{K},\mathbf{Sets}]}(*, \Pi)$, where $[\mathbf{K}, \mathbf{Sets}]$ is the category of **Sets**-valued functors on $\mathbf{K}$ and $*$ is the functor that associates to each $X \in \operatorname{Ob} K$ a singleton. Then s is a collection of probabilities, that are mutually compatible under marginalizations. This also appears in the literature as pseudo-marginals [94] and many other names.

categories that behave like the category of sets with respect to certain operations; they also formalize generalized logical theories. We refer the reader to [15,55,61,65].

The category of presheaves on a given category can be seen as a topos for the trivial topology (*topologie grossière*). For us the main examples are the presheaf topoi on $\mathbf{S}$ or $\mathbf{S}^{op}$ for a given information structure $(\mathbf{S}, \mathscr{E})$; we shall see that in these topoi the relevant notion of localization is always linked to *marginalizations* (also called coarse-graining in the physics literature). The advantage of the topological viewpoint for the study of information structures is that new geometrical intuitions become available, as well as very general algebro-geometrical tools. We did not explore the connections to logic, but this undoubtedly constitutes a very interesting problem.

It should also be noted that in modern geometry spaces are studied though their (co)homology, that to some extent encodes their "shape". Bennequin and Baudot [10] introduced a cohomological formalism connected to information functions, called *information cohomology*. The definition utilizes the general algebraic machinery of topos theory, as we now explain.

Recall from the previous section that to any structure $(\mathbf{S}, \mathscr{E})$ we can associate a presheaf of algebras $\mathscr{A}$. An $\mathscr{A}$-module is a functor $\mathscr{F}$ from $\mathbf{S}^{op}$ to the category of abelian groups such that, for each $X \in \operatorname{Ob}\mathbf{S}$, the group $\mathscr{F}(X)$ has the structure of an $\mathscr{A}_X$-module; the action by $\mathscr{A}$ is supposed to respect functoriality. These presheaves and the natural transformations between them form an *abelian category* denoted $\mathbf{Mod}(\mathscr{A})$,[22] which entails the possibility of defining cohomological functors. The right derived functor of $\operatorname{Hom}(-, \mathscr{W})$ (where $\mathscr{W}$ is a fixed sheaf of $\mathscr{A}$-modules) is $\operatorname{Ext}^\bullet(-, \mathscr{W})$; it is a cohomological functor (exact δ-functor, see Section 2.1.3): for an exact sequence $0 \to \mathscr{V}' \to \mathscr{V} \to \mathscr{V}'' \to 0$ of sheaves, it induces a long exact sequence

$$\begin{aligned} 0 \to \operatorname{Hom}(\mathscr{V}', \mathscr{W}) \to \operatorname{Hom}(\mathscr{V}, \mathscr{W}) \to \\ \operatorname{Hom}(\mathscr{V}'', \mathscr{W}) \to \operatorname{Ext}^1(\mathscr{V}', \mathscr{W}) \to \operatorname{Ext}^1(\mathscr{V}, \mathscr{W}) \to \dots \end{aligned} \tag{0.36}$$

as singular or cellular cohomology do in basic topology.

The information cohomology of $\mathbf{S}$ with coefficients in the $\mathscr{A}$-module $\mathscr{F}$ is

$$H^\bullet(\mathbf{S}, \mathscr{F}) := \operatorname{Ext}^\bullet(\mathbb{R}_\mathbf{S}, \mathscr{F}), \tag{0.37}$$

where $\mathbb{R}_\mathbf{S}$ is the sheaf that associates to every $X \in \operatorname{Ob} S$ the set $\mathbb{R}$ with trivial $\mathscr{A}_X$ action (i.e. for all $Y \in \mathscr{S}_X$ and all $r \in \mathbb{R}$, $Y.r = r$).

Recurring to another algebraic construction, called the (relative) bar resolution [60, Ch. IX], we can give a more computable description of this cohomology, that serves as an alternative definition explicitly connected to information measures as explained below.

The bar construction gives a sequence of projective $\mathscr{A}$-modules

$$0 \longleftarrow \mathbb{R}_\mathbf{S} \xleftarrow{\ \varepsilon\ } \mathscr{B}_0 \xleftarrow{\ \partial_1\ } \mathscr{B}_1 \xleftarrow{\ \partial_2\ } \mathscr{B}_2 \xleftarrow{\ \partial_3\ } \dots , \tag{0.38}$$

that is a resolution, meaning that $\ker \varepsilon = \operatorname{im} \partial_1$ and, for every $n \geq 1$, $\ker \partial_n = \operatorname{im} \partial_{n+1}$. Moreover, for each $X \in \operatorname{Ob}\mathbf{S}$, $n \in \mathbb{N}$, the module $\mathscr{B}_n(X)$ is freely generated over $\mathscr{A}_X$

[22] In [35], Grothendieck defined abelian categories and derived functors as a general ground for homological algebra. Essentially, an abelian category has an appropriate notion of kernel and cokernel, and it allows the introduction of homological algebra mimicking the case of modules.

by the symbols $[X_1|...|X_n]$; in the case of $\mathscr{B}_0(X)$, simply by a symbol $[\,]$. Proposition 2.13 proves that, due to the conditional existence of products in the definition of $\mathbf{S}$, these $\{\mathscr{B}_i\}_i$ are projective objects in the category $\mathbf{Mod}(\mathscr{A})$.[23]

Given any $\mathscr{A}$-module $\mathscr{F}$, we obtain a sequence

$$0 \longrightarrow \mathrm{Hom}_{\mathscr{A}}(\mathbb{R}_{\mathbf{S}}, \mathscr{F}) \longrightarrow \mathrm{Hom}_{\mathscr{A}}(\mathscr{B}_0, \mathscr{F}) \xrightarrow{\delta^0} \mathrm{Hom}_{\mathscr{A}}(\mathscr{B}_1, \mathscr{F}) \xrightarrow{\delta^1} \mathrm{Hom}_{\mathscr{A}}(\mathscr{B}_2, \mathscr{F}) \xrightarrow{\delta_2} \dots , \tag{0.39}$$

that is no longer a resolution, but $\delta^2 = 0$ holds. Define $C^n(\mathscr{F}) := \mathrm{Hom}_{\mathscr{A}}(\mathscr{B}_n, \mathscr{F})$; an element $\phi \in C^n(\mathscr{F})$ is uniquely determined by its image on the generators $[X_1|...|X_n]$, that we denote $\phi[X_1|...|X_n]$. The map $\delta_n : C^n(\mathscr{F}) \to C^{n+1}(\mathscr{F})$ is given by the formula

$$\begin{aligned}\delta_n\phi[X_1|...|X_{n+1}] = X_1.\phi[X_2|...|X_{n+1}] + \sum_{k=1}^{n}(-1)^k\phi[X_1|...|X_kX_{k+1}|...|X_n] \\ + (-1)^{n+1}\phi[X_1|...|X_n]\end{aligned} \tag{0.40}$$

The cohomology of the differential complex (C^n, δ) measures the difference between $\ker \delta_n$ and $\mathrm{im}\,\delta_{n-1}$; it is defined as

$$H^0(C^\bullet(\mathscr{F}), \delta) := \ker \delta^0 \quad \text{and} \quad H^n(C^\bullet(\mathscr{F}), \delta) := \ker \delta^n / \mathrm{im}\,\delta^{n-1} \tag{0.41}$$

when $n \geq 1$. It can be proved that there is a unique ∂-functorial identification of $H^\bullet(\mathbf{S}, \mathscr{F})$ with $H^\bullet(C^\bullet(\mathscr{F}), \delta)$ and in this sense both constructions are regarded as equivalent, cf. Section 2.1.3.

The elements of $\ker \delta^0$ are called 0-cocycles; they satisfy, for every $X \in \mathrm{Ob}\,\mathbf{S}$ and $Y \in \mathscr{S}_X$,

$$0 = Y.\phi_X[\,] - \phi_X[\,]. \tag{0.42}$$

The elements of $\mathrm{im}\,\delta^0$ are 1-coboundaries and those of $\ker \delta^1$ are 1-cocycles. The latter are characterized by

$$0 = Y.\phi_X[Z] - \phi_X[YZ] + \phi_X[Z], \tag{0.43}$$

for any $X \in \mathrm{Ob}\,\mathbf{S}$ and $Y, Z \in \mathscr{S}_X$. The reader should compare these cocycle conditions with the recurrence formulae in Section 0.2 and 0.3; this is clarified below.

0.7 Cohomology of discrete variables

This section summarizes the results in Part II of this thesis. It is supposed everywhere that the nerve of $\mathbf{S}$ has finite dimension and that each set E_X, for $X \in \mathrm{Ob}\,\mathbf{S}$, is finite. We obtain cohomological results for two different modules of coefficients, the first corresponding to probabilistic functionals and the second to combinatorial ones. We determine completely H^0 and H^1. The computation of higher cohomology groups remains an open problem because the functional equations involved are very complicated.

The probabilistic case appeared initially in [91].

[23] The general construction only gives a sequence of *relatively* projective objects, as explained in Section 2.3.

0.7.1 Probabilistic cohomology

Let $\mathscr{Q}$ denote any subfunctor of Π (the functor of probability measures) *stable under conditioning*. We represent every probability law by its density P with respect to the counting measure. Let $\mathscr{F}(X)$ be the additive abelian group of measurable[24] real-valued functions on $\mathscr{Q}(X)$ and, for any arrow $\pi : X \to Y$, let $\mathscr{F}(\pi) : \mathscr{F}(Y) \to \mathscr{F}(X)$ be precomposition with marginalization: $\mathscr{F}(\pi)(\phi) = \phi \circ \pi_*$. We obtain in this way a contravariant functor $\mathscr{F}$ on $\mathbf{S}$.

For each $Y \in \mathscr{S}_X$, $\phi \in \mathscr{F}(X)$ and $P \in \mathscr{Q}(X)$, define

$$(Y.\phi)(P) = \sum_{y \in E_Y} (Y_* P(y))^\alpha \phi(P_X|_{Y=y}). \tag{0.44}$$

By convention, a summand is simply 0 if $Y_*P(y) = P(Y = y) = 0$. This turns each $\mathscr{F}(X)$ into an $\mathscr{A}_X$-module and this action is functorial, in such a way that $\mathscr{F}$ becomes an $\mathscr{A}$-module $\mathscr{F}_\alpha$ (there is a family, one for each $\alpha > 0$).

We call *probabilistic* the information cohomology with coefficients in some $\mathscr{F}_\alpha$. Using the bar-resolution description introduced in the last section, we can determine the following facts.

A 0-cochain is a collection $\{\phi_X[]\}_{X \in \operatorname{Ob} S}$ that is local: for any X, $\phi_X[](P) = \phi_1[](1)$, so $\phi_X[]$ equals a constant $K \in \mathbb{R}$. The boundary of ϕ is $(\delta\phi)[Y] = Y.\phi[] - \phi[]$ which evaluated on a probability $P \in \mathscr{Q}(X)$ reads

$$(\delta\phi)[Y](P) = \sum_{y \in E_Y} (Y_* P(y))^\alpha K - K = \begin{cases} 0 & \text{if } \alpha = 1 \\ K S_\alpha[Y](Y_* P) & \text{otherwise} \end{cases}.$$

In other words: every cochain is a 0-cocycle if $\alpha = 1$. There are no 0-cocycles if $\alpha \neq 1$, but Tsallis entropy appears as 1-coboundary, multiplied by a global constant K.

The 1-cochains are characterized by collections of functionals $\{\phi[X] : \mathscr{Q}(X) \to \mathbb{R}\}_{X \in \operatorname{Ob} \mathbf{S}}$, which takes into account joint locality: $\phi_Y[X] = \phi_X[X] =: \phi[X]$. The 1-cocycles additionally satisfy

$$0 = X.\phi[Y] - \phi[XY] + \phi[X] \tag{0.45}$$

as functions on $\mathscr{Q}(XY)$, where marginalizations are implicit. As explained in Section 0.2, this equation and its analogue with X and Y permuted imply that $\phi[\cdot] = K_{XY} S_\alpha[\cdot]$, for a constant $K_{XY} \in \mathbb{R}$. This holds as long as $\mathscr{Q}_{XY}$ contains enough probabilities; a precise sufficient condition is stated in the definition of nondegeneracy for the product of two variables (Definition 3.12).

The number of free constants is determined in Theorem 3.14, the main result of Chapter 3. For a simplicial structure $(\mathbf{K}, \mathscr{E})$, it says that

$$H^1(\mathbf{S}, \mathscr{F}_1(\Pi)) \cong \mathbb{R}^{\beta_0(\mathbf{K})} \quad \text{and} \quad H^1(\mathbf{S}, \mathscr{F}_\alpha(\Pi)) \cong \mathbb{R}^{\beta_0(\mathbf{K})-1} \text{ when } \alpha \neq 1, \tag{0.46}$$

where $\beta_0(\mathbf{K})$ is the number of connected components of (the geometric realization of) $\mathbf{K}$. On each component the entropy S_α appears as the unique generator of Z^1. We conjecture that higher cohomology groups are linked to the higher Betti numbers.

[24] In all the cases we consider, the measures have the structure of a topological space, that becomes a measurable space with the corresponding Borel σ-algebra. For a discrete variable X, $\Pi(X) \cong \Delta^{|E_X|-1} \subset \mathbb{R}^{|E_X|}$.

0.7.2 Combinatorial cohomology

Let $\mathscr{C} : \mathbf{S} \to \mathbf{Sets}$ be the functor that associates to each $X \in \operatorname{Ob}\mathbf{S}$ the set $\mathscr{C}(X)$ of *counting functions*: $\nu : E_X \to \mathbb{N}$ such that $\|\nu\| := \sum_{x\in E_X} \nu(x) > 0$. Given $\pi : X \to Y$, the arrow $\pi_* := \mathscr{C}(\pi) : \mathscr{C}(X) \to \mathscr{C}(Y)$ is given by the corresponding marginalization $\pi_*\nu(y) = \sum_{x\in\pi^{-1}(y)} \nu(x)$.

Let $G(X)$ be the multiplicative abelian group of measurable $(0,\infty)$-valued functions on $\mathscr{C}(X)$. Every arrow $\pi : X \to Y$ induces a map $G(\pi) : G(Y) \to G(X)$ given by precomposition with marginalizations, $G(\pi)(\phi) = \phi \circ \pi_*$.

For each $Y \in \mathscr{S}_X$, $\phi \in G(X)$ and $\nu \in \mathscr{C}(X)$, define

$$(Y.\phi)(P_X) = \prod_{\substack{y\in E_Y\\ \nu(Y=y)\neq 0}} \phi(\nu|_{Y=y_i}). \tag{0.47}$$

As in the probabilistic case, this turns G into an $\mathscr{A}$-module. The computation of $H^\bullet(\mathbf{S}, G)$ gives the following results. The 0-cochains correspond again to collections $\left\{\phi_X[]\right\}_{X\in\operatorname{Ob} S}$ and locality implies that, for any X, $\phi_X[](\nu) = \phi_1[](\pi_{1X*}\nu) =: \varphi(\|\nu\|)$. The 0-cocycle condition is $1 = (\delta\phi)[Y] = (Y.\phi[])(\phi[])^{-1}$; evaluated on a counting function $\nu \in \mathscr{C}(Y)$ it reads

$$\varphi(\|\nu\|) = \varphi(\nu_1)\varphi(\nu_2)\cdots\varphi(\nu_s), \tag{0.48}$$

and evidently the only solutions are $\varphi_k(x) = \exp(kx)$, for $k \in \mathbb{R}$. The 1-cochains are characterized by collections of functionals $\left\{\phi[X] : \mathscr{C}(X) \to \mathbb{R}\right\}_{X\in\operatorname{Ob} S}$, and they define a 1-cocycle if, for every admissible product XY,

$$\phi[XY] = (X.\phi[Y])\phi[X] \tag{0.49}$$

as functions on $\mathscr{C}(XY)$, marginalizations being implicit. In virtue of Proposition 4.8, the solutions to these equations are of the form

$$\phi[Z](\nu) = \frac{[\|\nu\|]_D!}{\prod_{z\in E_Z}[\nu(z)]_D!} \tag{0.50}$$

where $[0]_D! = 1$ and $[n]_D! = D_n D_{n-1}\cdots D_1 D_0$, for any sequence $\{D_i\}_{i\geq 1}$ such that $D_1 = 1$. These are the Fontené-Ward multinomial coefficients introduced in Section 0.3. Remark that (0.49) serves as a general statement of the recursive formulae satisfied by the multinomial coefficients and their generalizations; particular cases are (0.17) and (0.18).

The asymptotic relation between the usual multinomials and S_1, on the one hand, and the q-multinomials and S_2, on the other, become particular cases of a general correspondence principle (Proposition 4.11).

Proposition 0.8. *Let g be a combinatorial n-cocycle. Suppose that, for every $X_1, \dots, X_n \in$* $\operatorname{Ob}\mathbf{S}$ *such that $X_i \cdots X_n \in$* $\operatorname{Ob}\mathbf{S}$, *there exists a measurable function*

$$f[X_1|\dots|X_n] : \Pi(X_1\cdots, X_n) \to \mathbb{R}$$

with the following property: for every sequence of counting functions $\{\nu_n\}_{n\geq 1} \subset \mathscr{C}_{X_1\cdots X_n}$ such that

1. $\|\nu_n\| \to \infty$, *and*
2. *for every* $z \in E_{X_1 \cdots X_n}$, $\nu_n(z)/\|\nu_n\| \to p(z)$ *as* $n \to \infty$

the asymptotic formula

$$g[X_1|...|X_n](\nu_n) = \exp(\|\nu_n\|^\alpha f[X_1|...|X_n](p) + o(\|\nu_n\|^\alpha))$$

holds. Then f *is a* n*-cocycle of type* α*, i.e.* $f \in Z^n(\mathbf{S}, F_\alpha(\Pi))$.

0.8 Differential entropy and relative entropy

According to Shannon [78], the analogue of (0.3) for $\mathbb{R}^n$-valued random variables with densities is

$$h[X](\lambda_n, f) = -\int_{\mathbb{R}^n} f(x) \ln f(x) \,\mathrm{d}x. \tag{0.51}$$

where we have highlighted the dependence on a variable X, the Lebesgue measure λ_n on $E_X = \mathbb{R}^n$, and the density f with respect to this measure. This function is called *differential entropy*. Shannon points to some apparent differences with the discrete case:

> In the discrete case the entropy measures in an *absolute way* the randomness of the chance variable. In the continuous case the measurement is *relative to the coordinate system.* If we change the coordinates the entropy will in general change [...] the new entropy is the old entropy less the expected logarithm of the Jacobian. In the continuous case the entropy can be considered a measure of randomness *relative to an assumed standard*, namely the coordinate system chosen with each small volume element $dx_1 \cdots dx_n$ given equal weight. [...]
>
> The entropy of a continuous distribution can be negative. The scale of measurements sets an arbitrary zero corresponding to a uniform distribution over a unit volume. A distribution which is more confined than this has less entropy and will be negative.

Kolmogorov [52, p. 16] even says that "It is well known that [the differential entropy] does not have a straightforward meaningful interpretation and is even noninvariant with respect to transformations of coordinates in the space $x_1, .., x_n$," but in fact such meaningful interpretation exists and its related to asymptotic concentration of measure.

We prove in Chapter 12 the following version of the Asymptotic Equipartition Property. We do not claim originality here (it is proved as [20, Thm. 8.2.2]). Let $(E_X, \mathfrak{B}, \mu)$ be a σ-finite measure space and $\{X_i\}_{i\in\mathbb{N}}$ a collection of iid $(E_X, \mathfrak{B})$-valued random variables, each following a law ρ with density $f := \mathrm{d}\rho/\mathrm{d}\lambda$ with respect to μ. The (relative) entropy is defined by[25]

$$S_\mu(\rho) := \mathbb{E}_\rho\left(-\ln\frac{\mathrm{d}\rho}{\mathrm{d}\mu}\right) = -\int_{\operatorname{supp}\mu} f(x) \ln f(x) \,\mathrm{d}\mu(x). \tag{0.52}$$

[25]This generalizes the entropy already treated in an article by Csiszár [21], that discusses its approximation by means of discretizations of the variable X.

Theorem 0.9 (Asymptotic Equipartition Property). *Suppose that $S_\mu(\rho)$ is finite. For every $\delta > 0$ and $n \in \mathbb{N}$, set*

$$A_\delta^{(n)} := \left\{ (x_1, \dots, x_n) \in E_X^n \,\middle|\, |-\frac{1}{n} \log f_{X_1,\dots,X_n}(X_1, \dots, X_n) - S_\mu(\rho)| \leq \delta \right\}. \tag{0.53}$$

Then,

1. *for every $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that, for all $n \geq n_0$,*
$$\mathbb{P}\left(A_\delta^{(n)}\right) > 1 - \varepsilon;$$
2. *for every $n \in \mathbb{N}$,*
$$\mu^{\otimes n}(A_\delta^{(n)}) \leq \exp\{n(S_\mu(\rho) + \delta)\};$$
3. *for every $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that, for all $n \geq n_0$,*
$$\mu^{\otimes n}(A_\delta^{(n)}) \geq (1 - \varepsilon) \exp\{n(S_\mu(\rho) - \delta)\}.$$

When E_X is a countable set (possibly infinite), and μ the counting measure, a probability law ρ on E_X is always absolutely continuous with respect to μ, and its density is a function $p : E_X \to [0,1]$ such that $\sum_{x \in E_X} p(x) = 1$, that is usually taken as the definition of a probability law in the discrete case.[26] Then $S_\mu(\rho)$ is the familiar expression $-\sum_{x \in E_X} p(x) \log p(x)$ and the previous theorem corresponds to Proposition 0.4. Note that it is also possible to consider any multiple of the counting measure, $\nu = \alpha\mu$, for $\alpha > 0$. In this case, the probability density $\mathrm{d}\rho/\mathrm{d}\nu$ sums α^{-1} and $S_\nu(\rho) = S_\mu(\rho) + \ln(\alpha)$. Hence, the "absolute character" of the discrete entropy is illusory, it also depends on the reference measure.

If $E_X = \mathbb{R}^n$, μ is some Lebesgue measure, and ρ a probability law such that $\rho \ll \mu$, then the derivative $\mathrm{d}\rho/\mathrm{d}\mu \in L^1(\mathbb{R}^n)$ corresponds to the elementary notion of density, and the quantity $S_\mu(\rho)$ is the differential entropy introduced by Shannon. As Shannon already explained, a coordinate transformation changes the reference measure. The expected value of the Jacobian gives the necessary factor to correct the volume estimates. For example, if T is an invertible linear transformation,

$$\mu^{\otimes n}(TA_\delta^{(n)}) \approx \exp(nS_\mu(\rho))(\det T)^n \approx (\det T)^n \mu^{\otimes n}(A_\delta^{(n)}). \tag{0.54}$$

For any E_X, if μ is a probability law, the expression $S_\mu(\rho)$ equals $-D_{KL}(\rho||\mu)$, where D_{KL} is the Kullback-Leibler divergence.

In the same vein, Section 12.2 explains how the divergence of entropy to $-\infty$ when ρ approaches a singular measure is necessary for the consistence of this theorem.

0.9 Cohomology of continuous variables

Let $(\mathbf{K}, \mathscr{E})$ be a *simplicial information structure* as defined in Section 0.5; each set $S \in \operatorname{Ob} \mathbf{K}$ can be viewed as a random variable X_S. Suppose moreover that each space $(E_i, \mathfrak{B}_i)$ is second-countable, in such a way that each $\mathfrak{B}(E_A)$ equals $\bigotimes_{i \in A} \mathfrak{B}_i$, and the choice of a σ-finite reference measure μ_i for each $\{i\} \in \operatorname{Ob} \mathbf{K}$ induces a product

[26] Also because $p(x) = \rho(\{x\})$, which does not make sense in the continuous case.

measure $\mu_A = \bigotimes_{i\in A} \mu_i$ on $(E_A, \mathfrak{B}(E_A))$ for any other $A \in \mathrm{Ob}\,\mathbf{K}$. This example is general enough to cover the discrete systems already treated (when each E_i is a finite set and μ_i is the counting measure), and also new examples: Euclidean spaces with their Lebesgue measure, and more generally products of locally compact Hausdorff topological groups endowed with a Haar measure.

We consider measures $\rho \in \Pi(S)$ that are absolutely continuous with respect to the corresponding μ_S, with density f_ρ. This gives the well known formulae for marginalizations under $\pi_{T,S} : E_S \to E_T$ by partial integration of the densities,

$$f_\rho \mapsto f_{\pi_{T,S}\rho} = \int_{E_{S\setminus T}} f_S(x_T, x_{S\setminus T})\,\mathrm{d}\mu_{S\setminus T}(x_{S\setminus T}).$$

However, conditioning introduces several complications, since $\rho|_{X_T=t}$ is a law on E_S supported on the hyperplane $\{X_T = t\}$:

$$\rho|_{X_T=t} = \frac{f_\rho}{\int_{E_{S\setminus T}} f_\rho(x,t)\,\mathrm{d}\mu_{S\setminus T}(x)} \mu_{S'\setminus T} \otimes \delta_{T=t}, \tag{0.55}$$

This forces the consideration of a larger class of measures closed under conditioning, which requires a careful study of disintegration carried out in Section 9.2.

We introduce a functor $\mathscr{Q}$ that associates to each $S \in \mathrm{Ob}\,\mathbf{K}$ the set of probability measures absolutely continuous with respect to a given reference measure,

$$\begin{aligned}\mathscr{Q}(S) = \{\, (\mu, \rho)) \mid \rho \in \Pi(S),\ \mu = \mu_{S'} \otimes \delta_{S''=a} \text{ for some } \\ S', S'' \text{disjoint such that } S = S' \cup S'', \text{and } \rho \ll \mu \,\},\end{aligned} \tag{0.56}$$

and then sets $\mathscr{F}_+(S)$ of measurable nonnegative functions $\varphi : \mathscr{Q}(S) \to \mathbb{R}$. For every $T \in \mathbf{S}_S$ and $\varphi \in \mathscr{F}_+(S)$, set

$$(T.\varphi)(\mu_{S'} \otimes \delta_{S''=s''}, \rho) := \int_{E_T} \varphi(\mu_{S'\setminus T} \otimes \delta_{T'=t'} \otimes \delta_{S''=s''}, \rho_t)\,\mathrm{d}T_*\rho(t), \tag{0.57}$$

where ρ_t is the conditional measure, supported on $\{T = t\}$; it has a density with respect to the reference measure $\mu_{S'\setminus T} \otimes \delta_{T'=t'} \otimes \delta_{S''=s''}$.

This equation defines an action of the corresponding monoid $\mathscr{S}_X$ on the set $\mathscr{F}_+(S)$. To treat signed functions, additional conditions must be imposed to guarantee the convergence of the integral. For instance, this can be accomplished restricting $\mathscr{Q}(S)$ to gaussian laws $\mathscr{Q}_{\text{Gauss}}(S)$ and using the preferred basis of each E_S to parametrize these probabilities by their mean and variance (m, Σ). Define $\mathscr{F}_{\text{Gauss}}(S)$ as the additive abelian group of measurable functions $\varphi : \mathscr{Q}_{\text{Gauss}}(S) \to \mathbb{R}$ that grow at most polynomially in the variable m; this implies that the integral in (0.57) converges and determines a functorial action of $\mathscr{S}_S$ and $\mathscr{A}_S$ on $\mathscr{F}_{\text{Gauss}}(S)$.

In Chapter 10 we compute the information cohomology of these simplicial structures, restricting to modules $\mathscr{F}(S)$ under the action (0.57), for instance $\mathscr{F}_{\text{Gauss}}(S)$. We can prove in general that 0-cochains must be constants and they all satisfy the 0-cocycle equation. We then find an alternative characterization of 1-cochains, as collections $\Phi = \{\phi^S\}_{S\in \mathrm{Ob}\,\mathbf{K}}$ such that each ϕ^S is a real-valued function of the probabilities ρ on E_S absolutely continuos with respect to the measure μ_S, subject to a simplified 1-cocycle condition. The general properties of disintegrations imply that $\phi^S(\rho) = \mathbb{E}_\rho\left(-\ln \frac{\mathrm{d}\rho}{\mathrm{d}\mu_S}\right)$ defines a 1-cocycle.

The case $\mathscr{F}_{\text{Gauss}}(S)$ is explicit enough to characterize all the 1-cocycles. First we determine the 1-cocyles that depend only on the variance Σ.

Proposition 0.10. *Suppose that* $\mathbf{K}$ *is connected and all its vertices belongs to a* 1*-cell. A collection of* C^2 *functions*[27] $\Phi = \{\phi^S : PD(S) \to \mathbb{R}\}_{S\in\mathbf{S}}$ *defines a* 1*-cocycle if and only if there exist real constants* a *and* $\{k_i\}_{i\in I}$ *such that, for every* $S \in \mathbf{S}$,

$$\phi^S(\Sigma) = a\ln(|\Sigma|) + \sum_{i\in S} k_i. \tag{0.58}$$

It is well known that the differential entropy of a gaussian law ρ on E_S with mean m and covariance Σ is

$$h^S(\rho) = d\left(\frac{\log_b(2\pi e)}{2}\right) + \frac{1}{2}\log_b(|\Sigma|), \tag{0.59}$$

which is a particular case of (0.58).

The characterization of general 1-cocycles is much more involved and requires several results from distribution theory. We prove that they also include the moments and their generalizations. We say that $\varphi(m,\sigma)$ is a *generalized moment function (gmf)* associated to the family $g = \{g_\varepsilon : \mathbb{R} \to \mathbb{C}\}_{\varepsilon>0}$ of locally integrable functions if $g_\varepsilon(x)\exp(-ax^2)$ is integrable for every $a > 0$, and if

$$\varphi(m,\sigma) = \frac{1}{\sqrt{2\pi(\sigma-\varepsilon)}}\int_{\mathbb{R}} g_\varepsilon(x)\exp\left(-\frac{(x-m)^2}{2(\sigma-\varepsilon)}\right)\mathrm{d}x \tag{0.60}$$

whenever $\sigma > \varepsilon$. We write $\varphi(g)$ to highlight the dependency on the family g. Usual moments of a univariate normal of parameters (m,ε) are an example: $g_{k,\varepsilon}(m) := M_k(m,\varepsilon) = \frac{1}{\sqrt{2\pi\varepsilon}}\int_{\mathbb{R}} z^k e^{-\frac{1}{2}\frac{(z-m)^2}{\varepsilon}}\,\mathrm{d}z$ and $\varphi(m,\sigma) := g_k(m,\sigma)$. Remark that M_0 is a constant.

Theorem 0.11 (Structure theorem of 1-cocycles, simplicial case)**.** *Suppose that every* 0*-cell of* $\mathbf{K}$ *belongs to a* 1*-cell. If* ϕ^S *is a* 1*-cocycle, then there exist generalized moment functions* $\{\varphi(f^i)\}_{i\in I}$, *and a constant* $a \in \mathbb{R}$ *such that*

$$\phi^S(m,\Sigma) = \sum_{i\in S}\phi(m_i,\sigma_{ii}|f^i) + a\ln(|\Sigma|), \tag{0.61}$$

Hence we obtain an infinite number of 1-cocycles. This can be explained by the very particular role played by the coordinate axes used to define the simplicial information structure. Since this basis was introduced just for convenience, the result motivates the introduction of more general examples of information structures that we call *grassmanian categories*.

A Grassmannian information structure is defined by a poset $\mathbf{S}$ of subspaces of a vector space E, ordered by inclusion, that is supposed to contain E and be closed

[27]The positive definite matrices $PD(S) \subset M_{|S|}(\mathbb{R})$ are supposed to have the standard differential structure.

under conditional intersection i.e. if $V, W \in \mathrm{Ob}\, \mathbf{S}$ and there exists $Z \in \mathrm{Ob}\, \mathbf{S}$ such that $Z \subset V$ and $Z \subset W$, then $V \cap W$ is also an object of $\mathbf{S}$. We also introduce a functor $\mathscr{E} : \mathbf{S} \to \mathbf{VectSpaces}$ that associates to each vector subspace V the quotient space $E_V := E/V$ and to every arrow $V \to W$ (i.e. $V \subset W$) the canonical projection $\pi^{WV} : E_V \to E_W$. The pair $(\mathbf{S}, \mathscr{E})$ is an information structure (each vector space is supposed to come with its Borel σ-algebra).

To every $V \in \mathrm{Ob}\, \mathbf{S}$ we can associate the set $\mathscr{S}_V := \{ W \in \mathrm{Ob}\, \mathbf{S} \mid V \to W \}$, that is a monoid for the intersection $(W, W') \mapsto W \cap W'$. As before, let $\mathscr{A}_V$ denote its induced real algebra; an arrow $V \to W$ gives an inclusion $\mathscr{A}_W \hookrightarrow \mathscr{A}_V$ i.e. a presheaf. The pair $(\mathbf{S}, \mathscr{A})$ is a ringed site.

We introduce a precosheaf that associates to every $V \in \mathrm{Ob}\, \mathbf{S}$ the set $\mathscr{M}_V$ of affine subspaces of E_V or more generally a subset $\mathscr{N}_V \subset \mathscr{M}_V$ that represent supports of probability laws; the corresponding morphisms $\mathscr{N}(\pi^{WV}) : \mathscr{N}_V \to \mathscr{N}_W$ are induced by projections. In view of conditioning, we suppose that $\mathscr{N}$ is closed under projections, and also that it contains the fibers of the projections and every nonempty intersection of its elements. We also introduce a functor $\mathscr{L}$ that combines these supports with a possible choice of Lebesgue measure: $\mathscr{L}_V$ contains pairs (A, λ_A) with $A \in \mathscr{M}_V$ and λ_A Lebesgue measure on A; this has the structure of a principal bundle for the group $(\mathbb{R}_+^*, \times)$, that acts on the measures by multiplication.

Section 11.2 studies the moments of Gaussian laws in a basis-free manner, which differs from standard presentations. A Gaussian law ρ is a probability law supported on an affine space $A(\rho)$, absolutely continuous with respect to a Lebesgue measure λ; the density $G(\rho, \lambda)$ is such that $-D_X^2 \ln G(\rho, \lambda)$ is a non degenerate symmetric positive bilinear form B. The covariance Σ if defined to be the inverse of B; it is proved to be independent of the choice of λ. The mean is defined as usual, $M(\rho) = \mathbb{E}_\rho\,(X)$.

Additional choices are needed to introduce the trace and the determinant. A choice of Lebesgue measure induces a determinant $\det_\lambda$. The trace requires an isomorphism between the tangent space of A, denoted $T(A)$, and its dual. This implies that the moment of order two of a gaussian law on E_V depends on the choice of a metric on this space.

We introduce a sheaf $\mathscr{F}$ such that $\mathscr{F}_V$ are the measurable functions on the probability laws $\mathscr{P}_V$ supported on some space of $\mathscr{N}_V$, that grow at most polynomially in the mean $M(\rho)$. The action introduced in the simplicial case is generalized straightforwardly to this setting. Proposition 11.24 establishes $\rho \mapsto \dim A(\rho)$ defines a cocycle for this module of coefficients; this is implied by the rank theorem.

For every $V \in \mathrm{Ob}\, \mathbf{S}$, the differential entropy S_V is a function of a probability ρ and a reference measure λ on its support given by the formula $S_V(\rho, \lambda) = -\int \log \frac{\mathrm{d}\rho}{\mathrm{d}\lambda}\, \mathrm{d}\rho$. It does not belong to $\mathscr{F}_V$, since it depends on the choice of reference measure λ: in fact, $S(\rho, C\lambda) = S(\rho, \lambda) + \log C$. To formalize this variation in a functorial way, remark that the choice of an euclidean metric on E induces identifications $E/V \cong V^\perp$ and inclusions of every affine subspace $A \subset E_V$ in E, hence a choice of Lebesgue measure $\lambda_Q(A)$ on every support that we call a *metric trivialization* of $\mathscr{L}$. We introduce the vector space $\mathscr{X}$ of functions of a probability ρ on E_V and a metric trivialization λ_Q of $\mathscr{L}_V$ such that

$$\forall \rho \in \mathscr{P}_V, \forall Q, Q' \text{ euclidean metrics on } E,$$
$$\phi(\rho, \lambda_{Q'}) = \phi(\rho, \lambda_Q) + \ln D(T(A(\rho)); Q, Q'), \quad (0.62)$$

where D is a "discriminant" function that satisfies

$$D(B;Q,Q')D(B;Q',Q'')=D(B;Q,Q'')$$

for any triple of metrics Q, Q', Q'' and any vector space $B \subset E_V$. For any $\iota : V \to W$ in $\mathbf{S}$ and $\phi \in \mathscr{X}_V$, the equation $(W.\phi)(\rho,\lambda_Q) = \int_{A(\iota_*\rho)} \phi(\rho|_{W=w},\lambda_Q)\,\mathrm{d}\iota_*\rho(w)$ defines a natural action of $\mathscr{S}_V$ on $\mathscr{X}_V$.

The collection of function $\{S_V\}_{V\in\mathrm{Ob}\,\mathbf{S}}$ determines a 1-cochain in information cohomology with coefficients in $\mathscr{X}$, that in fact is a 1-cocycle because it obeys the continuous version of the chain rule. In sufficiently non-degenerate situations the dimension and the entropy are the only 1-cocycles.

Theorem 0.12. *The cohomology $H^1(\mathbf{S},\mathscr{X})$ over a sufficiently rich grassmannian information structure is the space of functions*

$$\Phi_V(\rho) = -aS(\rho) + b\dim(A(\rho)), \tag{0.63}$$

where a and b are arbitrary real constants.

For gaussian probabilities, the fact that differential entropy is a 1-cocycle is equivalent Schur's determinantal formula

$$\det\begin{pmatrix} A & B \\ C & D\end{pmatrix} = \det(A)\det(D-BA^{-1}C); \tag{0.64}$$

this recurrence relation is thus comparable to the chain rule for entropy or the multiplicative relations for multinomials, discussed in Sections 0.2 and 0.3.

The moments, in turn, appear as certain locally invariant natural transformations.

Part I

Foundations

Chapter 1

Information structures

1.1 Random variables and probabilities

This first section contains some definitions and notations coming from elementary probability theory. For the sake of completeness, we also recall the definition of *information structure* given in [10], although we work with a generalization, introduced in Section 1.2. some context and elementary definitions; it also fixes some notations. It can be skipped by a reader already familiarized with probability theory.

An ***algebra of sets*** $\mathfrak{F}$ over a set Ω is a collection of subsets of Ω such that:

1. $\emptyset$ and Ω are in $\mathfrak{F}$;
2. $\mathfrak{F}$ is closed complementation: if $A \in \mathfrak{F}$, then $A^c := \Omega \setminus A \in \mathfrak{F}$;
3. $\mathfrak{F}$ is closed under finite unions: given any $A, B \in \mathfrak{F}$, one has $A \cup B \in \mathfrak{F}$.

It is called a ***σ-algebra*** if it is also closed under countable unions, and the pair $(\Omega, \mathfrak{F})$ is then called a measurable space. When the algebra is finite, its elements are all the possible unions of its minimal sets (in the sense of inclusion), called atoms; the atoms are said to generate the algebra. Given any collection $\mathfrak{C}$ of subsets of Ω, the σ-algebra generated by $\mathfrak{C}$—denoted by $\sigma(\mathfrak{C})$—is the smallest σ-algebra that contains $\mathfrak{C}$.

Let Ω be a set representing the collection of all possible 'elementary events' of a given experience. For us, a random variable is a function X on Ω taking values in a measurable space $(E_X, \mathfrak{E}_X)$, that corresponds to the possible outputs of a measurement; in applications, the codomain is usually $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n))$ or a finite set E with the σ-algebra $\mathfrak{P}(E)$ of all its subsets.[1] Random variables are also called observables.

Every random variable X defines a σ-algebra of subsets of Ω, given by $X^{-1}(\mathfrak{E}_X)$. This is usually called the algebra *induced* by X, we shall denote it $\sigma(X)$. When $\sigma(Y) \subset \sigma(X)$, we say that $\sigma(Y)$ is coarser than $\sigma(X)$, or even that Y is coarser than X; alternatively, $\sigma(X)$ is finer than $\sigma(Y)$ or refines $\sigma(Y)$.

In Part II, we suppose that each random variable takes a finite number of different values, $E_X = \{x_1, ..., x_n\}$; to emphasize this we talk sometimes about *finite* random variables. Accordingly, we set $(E_X, \mathfrak{P}(E_X))$ as codomain of X, and we drop $\mathfrak{P}(E_X)$ from the notation. Continuous variables reappear in Part IV.

[1]It is common to fix a σ-algebra $\mathfrak{F}$ over Ω and define a random variable as a function $X : (\Omega, \mathfrak{F}) \to (E_X, \mathfrak{E})$ that is $\mathfrak{F}/\mathfrak{E}$-measurable, i.e. for all $S \in \mathfrak{E}$, $X^{-1}(S) \in \mathfrak{F}$. Here, we take a different point of view: an arbitrary function X to a measurable space *defines* a σ-algebra $\sigma(X)$ on Ω.

When the variables are finite, the subsets $\{\,\omega \in \Omega \,|\, X(\omega) = x\,\} \subset \Omega$, for $x \in E_X$, are the atoms of $\sigma(X)$, and they form a partition of the space Ω.[2] Conversely, to any finite partition $\{\Omega_1, ..., \Omega_n\}$ of Ω, one can associate a random variable $\sum_{i=1}^n a_i \chi_{\Omega_i}$, where χ_{Ω_i} denotes the indicator function and the numbers $\{a_1, ..., a_n\} \subset \mathbb{R}$ (or any ring) are all different. Two such random variables can differ in the values they take, but both are equally good to "discriminate" between elementary outcomes. Two random variables are equivalent for us if they define the same partition of Ω. The terms 'random variable' and 'partition' will be used interchangeably throughout Part I and II.

We define now a category $\mathbf{Obs}_{\text{fin}}(\Omega)$ of *finite observables*, whose objects are all the finite partitions of Ω. There is an arrow between two objects X and Y, given by a surjection $\pi : X \to Y$, whenever Y is a refinement of X (this means, $\sigma(Y) \subset \sigma(X)$); each subset $B \in Y$ equals $\cup_{A \in \pi^{-1}(B)} A$. In this case, X discriminates better between elementary outcomes. The category $\mathbf{Obs}_{\text{fin}}(\Omega)$ has a terminal element: the trivial partition $\mathbf{1} := \{\Omega\}$. When Ω is finite, it also has an initial element: the partition by points, that we denote by $\mathbf{0}$. The categorical product $X \times Y$ of two partitions X and Y is the coarsest partition that refines both. This product is commutative and associative. Moreover, given any element X, we have $XX = X$ (idempotency), $\mathbf{0}X = \mathbf{0}$ and $\mathbf{1}X = X$.

Definition 1.1. A ***classical, concrete and finite information structure***[3] $\mathbf{S}$ is a *full* subcategory of $\mathbf{Obs}_{\text{fin}}(\Omega)$ that satisfies the following properties:

1. The partition $\mathbf{1}$ is in $\operatorname{Ob}\mathbf{S}$.
2. for every X, Y, Z in $\operatorname{Ob}\mathbf{S}$, if $X \to Y$ and $X \to Z$, then YZ belongs to $\mathbf{S}$.[4]

A big family of examples can be obtained as follows: let Ω be a set and $\Sigma = \{\, S_i : \Omega \to E_i \,|\, 1 \leq i \leq n\,\}$ an arbitrary collection of finite variables. Any subset $I := \{i_1, ..., i_k\}$ of $[n] = \{1, ..., n\}$ defines a new partition by means of the product already described, $S_I := S_{i_1} \cdots S_{i_k}$; by convention, $S_\emptyset := \mathbf{1}$. Let $W(\Sigma)$ be the full subcategory $\mathbf{Obs}_{\text{fin}}(\Omega)$ with objects $\{S_I \mid I \subset [n]\}$. Since $W(\Sigma)$ contains all the products by construction, it is an information structure. Algebraically, $W(\Sigma)$ has the structure of a commutative idempotent monoid, with identity $\mathbf{1}$.

Suppose now that $\Omega = \prod_{i \in [n]} E_i$, where $|E_i| \geq 2$ for all i, and $S_i : \Omega \to E_i$ is the i-th canonical projection ($i = 1, ..., n$). Under these assumptions, $S_I \neq S_J$, as partitions, whenever $I \neq J$, and $S_I \to S_J$ implies that $J \subset I$.[5] In consequence, there is an injection $\iota : \Delta([n]) \to \operatorname{Ob}(W(\Sigma))$, $I \mapsto S_I$, where $\Delta([n])$ denotes the abstract simplex of dimension $n-1$ (see Appendix B). Let K be a simplicial subcomplex of $\Delta([n])$; by ι, it determines a full subcategory of $W(\Sigma)$, to which we add $S_\emptyset$ as a terminal object, constructing this way a new category $\mathbf{S}(K)$, that is an information

[2] In the sequel, $\{\,\omega \in \Omega \,|\, X(\omega) = x\,\}$ is simply written $\{X = x\}$.

[3] We simplify the name to 'concrete structures.'

[4] It would be simpler to take $\mathbf{S}$ cartesian. But we already know that, in quantum mechanics, some joint measurements are incompatible. We would like to describe the classical and quantum cases with the same axioms; in the classical case, this just adds flexibility.

[5] Proof: for any $I \subset [n]$,

$$\sigma(S_I) = \{S_I^{-1}(A) \mid A \in \bigotimes_{i \in I} \mathfrak{E}_i\}$$

and every set $S_I^{-1}(A)$ has the form $\prod_{i \in [n]} F_i$ with $F_i = E_i$ whenever $i \notin I$. If $J \not\subset I$, say $j^* \in J \setminus I$, then in general $S_{j^*}(S_J^{-1}(A))$, with $A \in \bigotimes_{j \in J} \mathfrak{E}_i$, will differ from E_i, proving that $S_J^{-1}(A)$ is not in $\sigma(S_I)$.

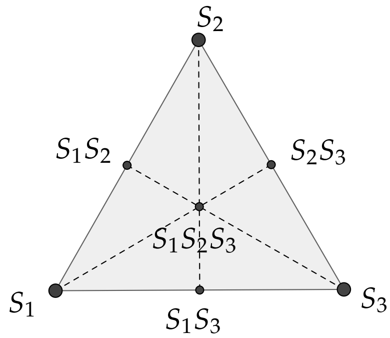

Figure 1.1: Identification of $\Delta([3])$ as the nontrivial part of $W(S_1, S_2, S_3)$. We have depicted also the barycentric subdivision, that has one point for each variable.

structure too. In fact, the diagram

$$S_J \leftarrow S_I \rightarrow S_L,$$

means that J and L are faces of I (for this, we need $|E_i| \geq 2$; see Remark 1.3 bellow); therefore, $J \cup L$ is also a face of I, that belongs to K by the simplicial condition; $J \cup L \in K$ implies that $S_{J\cup L} = S_J S_L \in \mathrm{Ob}(\mathbf{S}(K))$.

Example 1.2. Take $\Omega = \{0,1\}^3$, and consider the projections $S_i : \Omega \to \{0,1\}$ such that $(x_1, x_2, x_3) \mapsto x_i$, for $i \in \{1,2,3\}$. Taking all the possible joint variables, we obtain the monoid depicted in Figure 1.1. However, if we forbid the maximal face $S_1S_2S_3$, we obtain a new information structure, which is not a monoid. This could be linked to physical constraints related to measurements.

Remark 1.3. Bennequin and Baudot [10] define the structures $\mathbf{S}(K)$ for *any* collection of finite variables $\Sigma := \{\, S_i : \Omega \to E_i \,|\, 1 \leq i \leq n \,\}$. However, these structures do not satisfy in general the axiom (2) above: for instance, let Ω be $\{0,1\}^2$; X_i, the projection on the i-th component ($i = 1,2$), and $X_3 = \{\{(0,0)\}, \{(0,0)\}^c\}$. Define K as the simplicial subcomplex of $\Delta([3])$ with maximal faces $\{\{1,2\},\{3\}\}$. The product X_1X_2 is the atomic partition, that refines all the others, while some products (like X_1X_3) are not in $\mathbf{S}(K)$.

A probability law on general measurable space $(\Omega, \mathfrak{F})$ is a function $P : \mathfrak{F} \to [0,1]$ such that:

1. $P(\Omega) = 1$.
2. Given a collection of pair-wise disjoint sets $\{A_i\}_{i\in\mathbb{N}} \subset \mathfrak{F}$,

$$P\left(\bigcup_{i_1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i). \tag{1.1}$$

This property is called σ-additivity.

When $\mathfrak{F}$ is finite, we denote by $\Pi(\mathfrak{F})$ the set of all possible laws on $(\Omega, \mathfrak{F})$. If $\mathfrak{F}$ has N atoms, say $\{a_1, ..., a_N\}$, then the probability laws on $(\Omega, \mathfrak{F})$ can be identified with functions p defined on the atoms, such that $\sum_{i=1}^{N} p(a_i) = 1$. In this sense, $\Pi(\mathfrak{F})$ is a simplex embedded in $\mathbb{R}^N$; as such, can be considered a measurable space in itself (this is important in Section 3.1). For convenience, we identify each vertex δ_i of $\Pi(\mathfrak{F})$ with the corresponding atom a_i, such that $\delta_i(a_i) = 1$. The measure δ_i is called the Dirac (δ-)measure on a_i. More generally, we shall consider a simplicial subcomplex of $\Pi(\mathfrak{F})$, denoted by $\mathcal{Q}_{\mathfrak{F}}$.

Classically, the law of a random variable $X : (\Omega, \mathfrak{F}, \mathbb{P}) \to (E_X, \mathfrak{E})$ is the image measure $X[\mathbb{P}] = \mathbb{P} \circ X^{-1} : \mathfrak{E} \to [0, 1]$. Since a law is completely determined by the restriction of $\mathbb{P}$ to $\sigma(X)$, the set $\Pi(\sigma(X))$ contains all possible laws for X. Given $\pi : X \to Y$ in $\mathbf{Obs}_{\text{fin}}(\Omega)$, define the application $\pi_* : \Pi(\sigma(X)) \to \Pi(\sigma(Y))$ that associates to any law P on $(\Omega, \sigma(X))$ a new law π_*P on $(\Omega, \sigma(Y))$ given by

$$\pi_*P(B_Y) = P(B_Y), \qquad \text{for all } B_Y \in \sigma(Y) \subset \sigma(X). \tag{1.2}$$

This is called ***marginalization*** by Y. We write Y_* instead of π_* when π is clear from the context. Explicitely, for each $y \in Y$,

$$\pi_*P(y) = \sum_{x \in \pi^{-1}(y)} P(x). \tag{1.3}$$

Proposition 1.4. *Let X, Y be variables on Ω, $\pi : X \to Y$. If $\mathcal{Q}_X$ is a simplicial subcomplex of $\Pi(\sigma(X))$, then $\pi_*\mathcal{Q}_X$ is a simplicial subcomplex of $\Pi(\sigma(Y))$.*

Proof. Remark that π_* maps the convex combination $P = \sum_{x \in E_X} \lambda(x)\delta_x$ to the convex combination $\pi_*P = \sum_{y \in Y} \left(\sum_{x \in \pi^{-1}(y)} \lambda(x) \right) \delta_y$. Let σ be a simplex of $\mathcal{Q}_X$, with vertices $\{\delta_{x_1}, ..., \delta_{x_n}\}$. Each x_i refines the corresponding atom $\pi(x_i) \in Y$; clearly, $\pi_*(\delta_{x_i}) = \delta_{\pi(x_i)}$. Let $\{y_1, ..., y_m\}$ be the set of images of $\{x_1, ..., x_k\}$ under π. The map π_* sends convex combinations of $\{\delta_{x_1}, ..., \delta_{x_n}\}$ to convex combinations of $\{\delta_{y_1}, ..., \delta_{y_k}\}$, and hence the simplex $\sigma = [\delta_{x_1}, ..., \delta_{x_n}]$ to the simplex $[\delta_{y_1}, ..., \delta_{y_k}]$. Given a face ρ with vertices $V \subset \{\delta_{y_1}, ..., \delta_{y_k}\}$, there is a corresponding face σ' of σ with vertices $\pi^{-1}(V) \cap \{\delta_{x_1}, ..., \delta_{x_n}\}$, that is necessarily in $\mathcal{Q}_X$ by the definition of simplicial complex, and $\pi_*(\sigma') = \rho$, showing that ρ is in $\pi_*\mathcal{Q}_X$. □

Given an information structure $\mathbf{S}$, a ***probability functor*** $\mathcal{Q} : \mathbf{S} \to \mathbf{Sets}$ is a rule that assigns to each variable $X \in \operatorname{Ob} \mathbf{S}$ a simplicial subcomplex of $\Pi(\sigma(X))$, denoted simply $\mathcal{Q}_X$, and to each arrow of refinement $\pi : X \to Y$, the simplicial mapping $\mathcal{Q}(\pi) : \mathcal{Q}_X \to \mathcal{Q}_Y$ given by the marginalization $P \mapsto \pi_*P$.

Proposition 1.4 shows that we can obtain examples in a standard way: given a measurable space $(\Omega, \mathfrak{F})$ with $\mathfrak{F}$ finite, and a family of measurable functions on it (forming an information structure), fix initially a subcomplex $\mathcal{Q}_{\mathfrak{F}}$ of $\Pi(\mathfrak{F})$ and, for each variable X, define $\mathcal{Q}_X := X_*\mathcal{Q}_{\mathfrak{F}} \subset \Pi(\sigma(X))$.

Another fundamental operation is conditioning. Let $X : (\Omega, \mathfrak{F}) \to E_X$ be a random variable, P a law of $\Pi(\mathfrak{F})$ and $P(X = x) \neq 0$ for certain $x \in E_X$. Then, it is possible to define a new probability law $P|_{X=x}$ on $(\Omega, \mathfrak{F})$, called ***conditional law*** and given by

$$P|_{X=x}(B) \equiv P(B|X = x) := \frac{P(B \cap \{X = x\})}{P(X = x)}. \tag{1.4}$$

Proposition 1.5. *With the previous notation, if P belongs to a simplicial subcomplex $\mathcal{Q}$ of $\Pi(\mathfrak{F})$ and $P(X = x) > 0$, the law $P|_{X=x}$ also belongs to $\mathcal{Q}$.*

Proof. Suppose that the minimal face σ that contains P has vertices $V = \{\delta_{a_1}, ..., \delta_{a_k}\}$ (maybe just one): this means that $P(a) > 0$ for all $a \in V$ and $P(a') = 0$ for all other atom a'. The set $V' := \{a \text{ atom of } \mathfrak{F} \mid P(a \cap \{X = x\}) > 0\}$ is contained in V, because $P(a \cap \{X = x\}) \leq P(a)$. The minimal face that contains $P|_{X=x}$ is that of vertices V', which is a face of σ, and therefore contained in $\mathscr{Q}$. □

A probability family $\mathscr{Q}$ and an information structure $\mathbf{S}$ are ***mutually adapted***, if the conditioning of any law in $\mathscr{Q}$ by an element of $\mathbf{S}$ belongs to $\mathscr{Q}$. In particular, simplicial families are adapted: if X is any variable coarser than Y and $\mathscr{Q}_Y$ is simplicial complex, Proposition 1.4 (applied to $\mathfrak{F} = \sigma(Y)$ and $\mathscr{Q} = \mathscr{Q}_Y$) implies that $P|_{X=x}$ belongs to $\mathscr{Q}_Y$ (when it is well defined).

1.2 Category of information structures

We introduce here the general notion of statistical space motivated in the Introduction. It is flexible enough to cover all the situations where information cohomology has been studied: discrete and continuous classical random variables as well as finite dimensional quantum systems.

Definition 1.6. An ***information structure*** is a pair $(\mathbf{S}, \mathscr{E})$, where $\mathbf{S}$ ('the variables') is a small category such that

1. $\mathbf{S}$ has a terminal object, denoted $\mathbf{1}$;
2. $\mathbf{S}$ is a skeletal partially ordered set (poset).[6]
3. for objects $X, Y, Z \in \operatorname{Ob} S$, if $Z \to X$ and $Z \to Y$, then the categorical product $X \wedge Y$ exists;[7]

and $\mathscr{E} : X \mapsto (E_X, \mathfrak{B}_X)$ is a conservative[8,9] covariant functor ('the possible values') from $\mathbf{S}$ into the category **Meas** of measurable spaces, that satisfies

4. $E_{\mathbf{1}} \cong \{*\}$, with the trivial σ-algebra $\mathfrak{B}_{\mathbf{1}} = \{\emptyset, E_{\mathbf{1}}\}$;
5. for every $X \in \operatorname{Ob} \mathbf{S}$ and any $x \in E_X$, the σ-algebra $\mathfrak{B}_X$ contains the singleton $\{x\}$;
6. for every diagram $X \xleftarrow{\pi} X \wedge Y \xrightarrow{\sigma} Y$ the measurable map $E_{X \wedge Y} \hookrightarrow E_X \times E_Y, z \mapsto (x(z), y(z)) := (\pi_*(z), \sigma_*(z))$ is an injection.

To simplify notation, we usually write π_* or even π instead of $\mathscr{E}(\pi)$.

Even if the axiom (3) in Definition 1.6 is the obvious analogue of the conditional existence of products imposed in Section 1.1, only (6) allows us to recover the good properties of the product of partitions. See the proof of Proposition 3.1.

The structure $(\mathbf{S}, \mathscr{E})$ is said to be ***bounded*** if the nerve of $\mathbf{S}$ has finite dimension. Unbounded structures appear, for instance, in the study of Markov chains, as projective systems of measurable spaces, see [71, Ch. 8].

The structure is said to be ***finite*** if all the sets E_X are finite. In this case, E_X corresponds to the atoms of $\mathfrak{B}_X$, and the algebra can be omitted from the description,

[6]Being a poset means that, for any objects A and B, $\operatorname{Hom}(A, B)$ has at most one element. The poset is skeletal if it has no isomorphisms different from the identities: if $A \neq B$ and $A \to B$, then $B \not\to A$.

[7]This could be called "*conditional* meet semi-lattice".

[8]Given a functor $\mathscr{F} : \mathbf{S} \to \mathbf{Sets}$, we denote its value at $X \in \operatorname{Ob} \mathbf{S}$ by $\mathscr{F}(X)$ or $\mathscr{F}_X$.

[9]Conservative means that, if $\mathscr{E}(f)$ is an isomorphism, then f is an isomorphism. Since $\mathbf{S}$ is skeletal, this condition implies that, for every arrow $\pi : X \to Y$ such that $X \neq Y$, the measurable map $\pi_* := \mathscr{E}(\pi) : E(X) \to E(Y)$ is not a bijection.

and we denote the structure by $(\mathbf{S}, E)$. For example, probabilities on $\mathfrak{B}_X$ are in bijective correspondence with maps $p : E_X \to [0, 1]$ such that $\sum_{x \in E_X} p(x) = 1$, etc.

All concrete information structures $\mathbf{S}$, as defined in Section 1.1, are examples of *finite* generalized information structures, taking E to be the identity functor. In the general case, we still call $X \in \operatorname{Ob} \mathbf{S}$ a partition (also: variable, observable) and the elements in E_X (denoted also $E(X)$) are interpreted as the elements of this partition (or the possible values of the variable). In general, we can transfer to general information structures all the notations and notions from the previous section. For example, the set-theoretical notation $\{X = x\}$ simply means "the element x contained in E_X" and $\{X = x, Y = y\}$ should be interpreted as *the* element z of $E_{X \wedge Y}$ mapped to x by $E_{X \wedge Y} \to E_X$ and to y by $E_{X \wedge Y} \to E_Y$ (if such z does not exist, write $\{X = x, Y = y\} = \emptyset$); the uniqueness of z is guaranteed by axiom (6). As before, we write $XY := X \wedge Y$ and refer to this as the product of observables.

In the concrete case, for each $f : X \to Y$ in $\mathbf{S}$, the map map $\mathscr{E}(f)$ is a strict surjection, but we do not suppose this in general. Sums over the empty set equal zero, and products over the empty set equal one; with this conventions, expressions like $\sum_{x \in \mathscr{E}(f)^{-1}(y)} a(x)$ give the expected results.

The definitions and propositions in this section are also valid for infinite structures, but a more detailed study of these is postponed to Part IV.

The definition of $W(\Sigma)$ introduced in Section 1.1 is more natural in this context. Let I be a finite set, and $\mathbf{\Delta(I)}$ be the category of subsets of I, with arrows $I \to J$ whenever $J \subset I$. Set $\mathbf{S} = \mathbf{\Delta(I)}$. Let $\mathscr{E}_i = (E_i, \mathfrak{B}_i)$ be arbitrary measurable sets such that $|E_i| \geq 2$, and associate to $I = \bigwedge_{i \in I} \{i\}$ the product measurable space $\prod_{i \in I} \mathscr{E}_i$; the maps $\mathscr{E}(\pi) : \mathscr{E}(I) \to \mathscr{E}(J)$, for each $\pi : I \to J$, are the canonical projectors. There is no need to consider all the abstract simplicial complex $\mathbf{\Delta(I)}$, $\mathbf{S}$ could be a simplicial subcomplex $\mathbf{K}$ of $\mathbf{\Delta(I)}$, and $\mathscr{E}$ the restriction of the functor just defined; we obtain in this way a ***simplicial information structure*** $\mathbf{S(K)}$.

Information structures form a category.

Definition 1.7. Given two structures $(\mathbf{S}, \mathscr{E}), (\mathbf{S}', \mathscr{E}')$, a morphism $\phi = (\phi_0, \phi^{\#}) : (\mathbf{S}, \mathscr{E}) \to (\mathbf{S}', \mathscr{E}')$ between them is a functor $\phi_0 : \mathbf{S} \to \mathbf{S}'$ and a natural transformation $\phi^{\#} : \mathscr{E} \Rightarrow \mathscr{E}' \circ \phi_0$, such that

1. $\phi_0(\mathbf{1}) = \mathbf{1}$;
2. if $X \wedge Y$ exists, then $\phi_0(X \wedge Y) = \phi_0(X) \wedge \phi_0(Y)$;
3. for each $X \in \operatorname{Ob} S$, the component $\phi^{\#}_X : \mathscr{E}(X) \to \mathscr{E}'(\phi_0(X))$ is a measurable map.

Given $\phi : (\mathbf{S}, \mathscr{E}) \to (\mathbf{S}', \mathscr{E}')$ and $\psi : (\mathbf{S}', \mathscr{E}') \to (\mathbf{S}'', \mathscr{E}'')$, their composition $\psi \circ \phi$ is defined as $(\psi_0 \circ \phi_0, \psi^{\#} \circ \phi^{\#} : \mathscr{E} \Rightarrow \mathscr{E}'' \circ \psi_0 \circ \phi_0)$ (it is easy to verify that $\psi \circ \phi$ is also a morphism). If there is no risk of ambiguity, we write ϕ instead of ϕ_0.

We denote by **InfoStr** the category of information structures and its morphisms in the sense just defined.

Note that, if $X \wedge Y$ exists, then $\phi_0(X \wedge Y) \to \phi_0(X)$ and $\phi_0(X \wedge Y) \to \phi_0(X)$, and thus the product $\phi_0(X) \wedge \phi_0(Y)$ exists too, in virtue of Definition 1.6-(3).

This simple definition of a morphism between information structures and the corresponding construction of products and coproducts is one of the main motivations for this generalized setting.

Proposition 1.8. *The category* **InfoStr** *has finite products and coproducts.*

Proof.
Products: Given information structures $(\mathbf{S}_1, \mathscr{E}_1)$ and $(\mathbf{S}_2, \mathscr{E}_2)$, we introduce first the ordinary categorical product $\mathbf{S} = \mathbf{S}_1 \times \mathbf{S}_2$: its objects are all the pairs $\langle X_1, X_2\rangle$ with $X_i \in \operatorname{Ob}\mathbf{S_i}$ $(i = 1, 2)$; there is an arrow $\langle \pi_1, \pi_2\rangle : \langle X_1, X_2\rangle \to \langle Y_1, Y_2\rangle$ whenever $\pi_i : X_i \to Y_i$ in $\mathbf{S}_i$ $(i = 1, 2)$. Let $\mathscr{E} : \mathbf{S} \to \mathbf{Sets}$ be a functor $X \mapsto \mathscr{E}(X) = (E(X), \mathfrak{B}(X))$ defined by $E(\langle X_1, X_2\rangle) = E_1(X_1) \times E_2(X_2)$ and $\mathfrak{B}(\langle X_1, X_2\rangle) = \mathfrak{B}(X_1) \otimes \mathfrak{B}(X_2)$, the product σ-algebra (see [19, Sec. 5.1]). The projections are simply given by $\mathscr{E}(\langle \pi_1, \pi_2\rangle) = \mathscr{E}_1(\pi_1) \times \mathscr{E}_2(\pi_2)$, which comes from the product in **Sets**. The pair $(\mathbf{S}, \mathscr{E})$ is an information structure:

- $\mathbf{S}$ is a small category, with terminal object $\langle \mathbf{1}_{\mathbf{S}_1}, \mathbf{1}_{\mathbf{S}_2}\rangle$;
- $\mathbf{S}$ is a skeletal poset: for $\langle X_1, X_2\rangle \neq \langle Y_1, Y_2\rangle$,

$$\begin{aligned}\operatorname{Hom}(\langle X_1, X_2\rangle, \langle Y_1, Y_2\rangle) \neq \emptyset &\Leftrightarrow X_1 \to Y_1 \text{ and } X_2 \to Y_2\\ &\Leftrightarrow Y_1 \not\to X_1 \text{ or } Y_2 \not\to X_2\end{aligned}$$

The last equivalence, because both arrows cannot be identities. Therefore,

$$\operatorname{Hom}(\langle Y_1, Y_2\rangle, \langle X_1, X_2\rangle) = \emptyset;$$

- Given $\langle X_1, X_2\rangle, \langle Y_1, Y_2\rangle$ and $\langle Z_1, Z_2\rangle$ such that $\langle X_1, X_2\rangle \to \langle Y_1, Y_2\rangle$ and $\langle X_1, X_2\rangle \to \langle Z_1, Z_2\rangle$, then $Y_i \xleftarrow{\pi_{Y_i}} X_i \xrightarrow{\pi_{Z_i}} Z_i$ in $\mathbf{S}_i$ $(i = 1, 2)$. By the conditional existence of products in $\mathbf{S}_i$, $Y_i \wedge Z_i$ exists $(i = 1, 2)$ and evidently $\langle Y_1 \wedge Z_1, Y_2 \wedge Z_2\rangle$ is the infimum of $\langle Y_1, Y_2\rangle$ and $\langle Z_1, Z_2\rangle$ in $\mathbf{S}$,

$$\langle Y_1, Y_2\rangle \wedge \langle Z_1, Z_2\rangle = \langle Y_1 \wedge Z_1, Y_2 \wedge Z_2\rangle.$$

- $E(\langle \mathbf{1}_{\mathbf{S}_1}, \mathbf{1}_{\mathbf{S}_2}\rangle) = \{(*, *)\}$, with trivial algebra;
- For $i = 1, 2$ and any $x_i \in E(X_i)$, the singleton $\{x_i\}$ is an element of $\mathfrak{B}(X_i)$; so the product $\{x_1\} \times \{x_2\} = \{(x_1, x_2)\}$ belongs to $\mathfrak{B}(\langle X_1, X_2\rangle)$.
- The argument above gives the following diagram in $\mathbf{S}$:

$$\langle Y_1, Y_2\rangle \xleftarrow{\langle \pi_{Y_1}, \pi_{Y_2}\rangle} \langle Y_1 \wedge Z_1, Y_2 \wedge Z_2\rangle \xrightarrow{\langle \pi_{Z_1}, \pi_{Z_2}\rangle} \langle Z_1, Z_2\rangle,$$

where $Y_i \xleftarrow{\pi_{Y_i}} Y_i \wedge Z_i \xrightarrow{\pi_{Z_i}} Z_i$ is the diagram of the product in $\mathbf{S}_i$ $(i = 1, 2)$. Given $(y_1, y_2) \in E(\langle Y_1, Y_2\rangle)$ and $(z_1, z_2) \in E(\langle Z_1, Z_2\rangle)$,

$$\begin{aligned}&\langle \pi_{Y_1}, \pi_{Y_2}\rangle_*^{-1}(y_1, y_2) \cap \langle \pi_{Z_1}, \pi_{Z_2}\rangle_*^{-1}(z_1, z_2)\\ &\qquad = \left\{\pi_{Y_1 *}^{-1}(y_1) \times \pi_{Y_2 *}^{-1}(y_2)\right\} \cap \left\{\pi_{Z_1 *}^{-1}(z_1) \times \pi_{Z_2 *}^{-1}(z_2)\right\}\\ &\qquad = \left\{\pi_{Y_1 *}^{-1}(y_1) \cap \pi_{Z_1 *}^{-1}(z_1)\right\} \times \left\{\pi_{Y_2 *}^{-1}(y_2) \cap \pi_{Z_2 *}^{-1}(z_2)\right\}.\end{aligned}$$

Thus $|\langle \pi_{Y_1}, \pi_{Y_2}\rangle_*^{-1}(y_1, y_2) \cap \langle \pi_{Z_1}, \pi_{Z_2}\rangle_*^{-1}(z_1, z_2))| \leq 1$.

For each $i \in \{1, 2\}$, we define functors $\pi_{\mathbf{S}_i} : (\mathbf{S}, \mathscr{E}) \to (\mathbf{S}_i, \mathscr{E}_i)$ such that $\pi_{\mathbf{S}_i 0}$ maps each $\langle X_1, X_2\rangle$ to X_i, and each morphism $\langle f_1, f_2\rangle$ to f_i. At the level of $\mathscr{E}$, let $\pi^{\#}_{\mathbf{S}_i} : \mathscr{E}_1(X_1) \times \mathscr{E}_2(X_2) \to \mathscr{E}_i(X_i)$ be the canonical projection. These formulae define morphisms of information structures. We claim that $\mathbf{S}$, with the projections

$$(\mathbf{S}_1, \mathscr{E}_1) \xleftarrow{\pi_{\mathbf{S}_1}} (\mathbf{S}, \mathscr{E}) \xrightarrow{\pi_{\mathbf{S}_2}} (\mathbf{S}_2, \mathscr{E}_2),$$

is the product of $(\mathbf{S}_1, \mathscr{E}_1)$ and $(\mathbf{S}_2, \mathscr{E}_2)$ in **InfoStr**, written $(\mathbf{S}_1, \mathscr{E}_1) \times (\mathbf{S}_2, \mathscr{E}_2)$, unique up to unique isomorphism. In fact, given $(\mathbf{S}_1, \mathscr{E}_1) \xleftarrow{f_1} (\mathbf{R}, \mathscr{F}) \xrightarrow{f_2} (\mathbf{S}_2, \mathscr{E}_2)$, define $\langle f_1, f_2 \rangle : (\mathbf{R}, \mathscr{F}) \to (\mathbf{S}, \mathscr{E})$ as

$$\begin{aligned} \langle f_1, f_2 \rangle_0 :& \mathbf{R} \to \mathbf{S} \\ & R \mapsto \langle f_1(R), f_2(R) \rangle \end{aligned}$$

for any object or morphism R; given an variable $X \in \mathbf{R}$, the map $\phi^\#_X : \mathscr{F}(X) \to \mathscr{E}(\langle f_1(X), f_2(X) \rangle) = \mathscr{E}_1(f_1(X)) \times \mathscr{E}_2(f_2(X))$ is the product $f_1{}^\#_X \times f_2{}^\#_X$ of the maps $f_i{}^\#_X : \mathscr{F}(X) \to \mathscr{E}_i(f_i(X))$. Evidently, $\pi_{\mathbf{S}_i} \circ \langle f_1, f_2 \rangle = f_i$ for $i = 1, 2$.

Coproducts: Given information structures $(\mathbf{S}_1, \mathscr{E}_1)$ and $(\mathbf{S}_2, \mathscr{E}_2)$, define a category $\mathbf{S}$ such that $\operatorname{Ob} \mathbf{S} = \operatorname{Ob} \mathbf{S_1} \sqcup \operatorname{Ob} \mathbf{S_2} / \mathbf{1}_{\mathbf{S}_1} \sim \mathbf{1}_{\mathbf{S}_2}$ and $A \to B$ in $\mathbf{S}$ if and only if $A \to B$ in $\mathbf{S}_1$ or in $\mathbf{S}_2$. Define a functor $\mathscr{E} : \mathbf{S} \to \mathbf{Sets}$ such that $\mathscr{E}(X) = \mathscr{E}_i(X)$ if $X \in \operatorname{Ob} \mathbf{S_i}$. The pair $(\mathbf{S}, \mathscr{E})$ is an information structure: axioms (2), (3) and (6) in Definition 1.6 are verified locally in $\mathbf{S}_1$ or $\mathbf{S}_2$.

Injections $\iota_i : \mathbf{S}_i \to \mathbf{S}$ are defined in the obvious way: $\iota_i(A) = A$ for $A \in \operatorname{Ob} \mathbf{S_i}$ or $A \in \operatorname{Hom}(\mathbf{S}_i)$; the corresponding maps $\iota_i{}^\#_X$ are identities. If

$$(\mathbf{S}_1, \mathscr{E}_1) \xrightarrow{f_1} (\mathbf{R}, \mathscr{F}) \xleftarrow{f_2} (\mathbf{S}_2, \mathscr{E}_2) \ ,$$

define

$$\begin{aligned} \langle f_1, f_2 \rangle_0 :& \mathbf{S} \to \mathbf{R} \\ & A \mapsto \begin{cases} f_1(A) & \text{if } A \in \operatorname{Ob} \mathbf{S_1} \text{ or } A \in \operatorname{Hom}(\mathbf{S}_1) \\ f_2(A) & \text{otherwise} \end{cases} . \end{aligned}$$

and, if $X \in \operatorname{Ob} \mathbf{S_i}$, set $\langle f_1, f_2 \rangle^\#_X = f_i{}^\#_X$. By construction, $\langle f_1, f_2 \rangle \circ \iota_i = f_i$. Therefore, $(\mathbf{S}, \mathscr{E})$ is the coproduct of $(\mathbf{S}_1, \mathscr{E}_1)$ and $(\mathbf{S}_2, \mathscr{E}_2)$ in **InfoStr**, denoted $(\mathbf{S}_1, \mathscr{E}_1) \coprod (\mathbf{S}_2, \mathscr{E}_2)$, which is unique up to unique isomorphism. □

Remark 1.9. If $(\mathbf{S}_1, \mathscr{E}_1)$ and $(\mathbf{S}_2, \mathscr{E}_2)$ are bounded structures, their product and coproduct are bounded too. In fact, if the dimension of $\operatorname{Nerve}(\mathbf{S}_i)$ is N_i ($i = 1, 2$), then the dimension of $\operatorname{Nerve}(\mathbf{S}_1 \times \mathbf{S}_2)$ is $N_1 + N_2$ and that of $\operatorname{Nerve}(\mathbf{S}_1 \coprod \mathbf{S}_2)$ equals $\max(N_1, N_2)$. Similarly, if both are finite, their product and coproduct is finite too.

Remark 1.10. If each measurable space $(E(X), \mathfrak{B}(X))$ appearing in $\mathbf{S}_1$ and $\mathbf{S}_2$ verifies that $E(X)$ is second countable topological space and $\mathfrak{B}(X)$ is its Borel σ-algebra, then each algebra $\mathfrak{B}(X_1) \otimes \mathfrak{B}(X_2)$ on $E(X_1) \times E(X_2)$ equals the Borel σ-algebra on this space [93, Prop. 9.1].

Example 1.11.

$$\begin{tikzcd} \mathbf{1} & & \mathbf{1} & & & (\mathbf{1},\mathbf{1}) & \\ \mathbf{0} \arrow[u] & \times & \mathbf{0} \arrow[u] & = & (\mathbf{1},\mathbf{0}) \arrow[ur] & & (\mathbf{0},\mathbf{1}) \arrow[ul] \\ & & & & & (\mathbf{0},\mathbf{0}) \arrow[ul] \arrow[ur] & \end{tikzcd}$$

and

$$\begin{array}{c}\mathbf{1}\\ \uparrow \\ \mathbf{0}\end{array} \coprod \begin{array}{c}\mathbf{1}\\ \uparrow \\ \mathbf{0}\end{array} = \begin{tikzcd} & \mathbf{1} & \\ \mathbf{0} \arrow[ur] & & \mathbf{0} \arrow[ul] \end{tikzcd}$$

1.3 Probabilities on finite structures

We introduce now probabilities on *finite* information structures. Continuous random variables are treated in Part IV.

Let $\mathbf{S}$ be a finite information structure and $\Pi : \mathbf{S} \to \mathbf{Sets}$ a functor that associates to each $X \in \operatorname{Ob}\mathbf{S}$ the set

$$\Pi(X) := \{\, p : E_X \to [0,1] \mid \sum_{x \in E_X} p(x) = 1 \,\}, \tag{1.5}$$

of probability laws for X, and to each arrow $\pi : X \to Y$ the natural marginalization

$$\forall P \in \mathcal{Q}_X,\ \forall y \in E_Y \quad \Pi(\pi)(P)(y) = \sum_{x \in \pi_*^{-1}(y)} P(x). \tag{1.6}$$

More generally, a probability functor $\mathcal{Q}$ on an information structure $(\mathbf{S}, E)$ is a functor $\mathcal{Q} : \mathbf{S} \to \mathbf{Sets}$ such that, for every $X \in \operatorname{Ob}\mathbf{S}$, the set $\mathcal{Q}_X$ is a subcomplex of $\Pi(X)$ and each arrow $\pi : X \to Y$ is sent to $\mathcal{Q}(\pi) = \Pi(\pi)|_{\mathcal{Q}_X}$, written simply π_* or even Y_* to simplify the notation.

We adopt the probabilistic notation, in the following sense: if $X, Y \in \operatorname{Ob}\mathbf{S}$, $\pi_{YX} : X \to Y$ in $\mathbf{S}$, $P_X \in \mathcal{Q}_X$, and for $y \in E(y)$, the notation $P_X(Y = y)$ means $P(\pi_{YX*}^{-1}(y)) = \pi_{YX*}P(y)$; similarly, if $Y \xleftarrow{\pi_{YX}} X \xrightarrow{\pi_{ZX}} Z$ is a diagram in $\mathbf{S}$, the notation $P_X(Y = y, Z = z) \equiv P_X(\{Y = y\} \cap \{Z = z\})$ means $P_X(\pi_{YX*}^{-1}(y) \cap \pi_{ZX*}^{-1}(z))$, which equals $P_X(\langle \pi_{YX}, \pi_{ZX}\rangle_*^{-1}(w(y,z)))$ for the unique $w(y,z)$ sent to (y,z) by the injection in Definition 1.6-(6).

Given an arrow $\pi_{ZX} : X \to Z$ and a law $P \in \Pi(X)$, the conditional law $P|_{Z=z}$ is defined by

$$P|_{Z=z}(x) := \frac{P(x \cap \{Z = z\})}{P(Z = z)} = \frac{P(x \cap \pi_{ZX}^{-1}(z))}{P(Z = z)}. \tag{1.7}$$

We say that the functor $\mathcal{Q}$ is adapted if it is stable by conditioning: for every arrow $X \to Z$ in $\mathbf{S}$, every law $P \in \mathcal{Q}_X$, and every $z \in E_Z$, $P|_{Z=z} \in \mathcal{Q}_X$.

Conditioning commutes with marginalizations: given arrows $\pi_{YX} : X \to Y$ and $\pi_{ZY} : Y \to Z$,

$$\begin{aligned}
\pi_*^{YX}(P|_{Z=z})(y) &= \sum_{x \in \pi_{YX}^{-1}(y)} \frac{P(\{x\} \cap \pi_{ZX}^{-1}(z))}{P(Z = z)} = \frac{\sum_{x \in \pi_{YX}^{-1}(y)} P(\{x\} \cap \pi_{ZX}^{-1}(z))}{P(Z = z)} \\
&= \frac{P(\pi_{YX}^{-1}(y) \cap \pi_{YX}^{-1}(\pi_{YZ}^{-1}(z)))}{P(Z = z)} = \frac{\pi_*^{YX} P(y \cap \pi_{YZ}^{-1}(z))}{\pi_*^{YX} P(Z = z)} \\
&= (\pi_*^{YX} P)|_{Z=z}(y).
\end{aligned}$$

Let $\mathcal{Q}_i$ be a probability functor on $(\mathbf{S}_i, E_i)$ $(i = 1, 2)$. We define:

1. $\mathcal{Q}_1 \times \mathcal{Q}_2 : \mathbf{S}_1 \times \mathbf{S}_2 \to \mathbf{Sets}$ as a functor that associates to each object $\langle X_1, X_2 \rangle \in \operatorname{Ob} \mathbf{S_1} \times \mathbf{S_2}$ the set of laws:

$$\mathcal{Q}_1 \times \mathcal{Q}_2(\langle X_1, X_2 \rangle) := \{ P : E_1(X_1) \times E(X_2) \to [0,1] \mid \exists P_1 \in \mathcal{Q}_1(X_1), \exists P_2 \in \mathcal{Q}_2(X_2) \text{ such that } P(x_1, x_2) = P_1(x_1) P_2(x_2) \}. \tag{1.8}$$

If two pairs (P_1, P_2), and (P_1', P_2') correspond to the same law $P : E_1(X_1) \times E(X_2) \to [0,1]$, we can marginalize one of the components under the arrow $\pi_1 : X_1 \to \mathbf{1}_{\mathbf{S}_1}$ to conclude that $P_2 = P_2'$; analogously, $P_1 = P_1'$. Therefore, what we call $\mathcal{Q}_1 \times \mathcal{Q}_2(\langle X_1, X_2 \rangle)$ is in bijection with the usual product of sets $\mathcal{Q}_1(X_1) \times \mathcal{Q}_2(X_2)$; we write $P = (P_1, P_2)$.

For each morphism $\langle X_1, X_2 \rangle \overset{\langle \pi_1, \pi_2 \rangle}{\to} \langle Y_1, Y_2 \rangle$ the induced map

$$\mathcal{Q}_1 \times \mathcal{Q}_2(\langle X_1, X_2 \rangle) \overset{\mathcal{Q}(\langle \pi_1, \pi_2 \rangle)}{\to} \mathcal{Q}_1 \times \mathcal{Q}_2(\langle Y_1, Y_2 \rangle)$$

(see Equation (1.6)) is compatible with marginalizations: for every $(y_1, y_2) \in E(\langle Y_1, Y_2 \rangle)$,

$$\begin{aligned}
[\mathcal{Q}(\langle \pi_1, \pi_2 \rangle)(P_1, P_2)](y_1, y_2) &\overset{\text{(def)}}{=} \sum_{(x_1, x_2) \in \langle \pi_1, \pi_2 \rangle_*^{-1}(y_1, y_2)} (P_1, P_2)(x_1, x_2) \\
&= \sum_{x_1 \in \pi_{1*}^{-1}(y_1)} P_1(x_1) \sum_{x_2 \in \pi_{2*}^{-1}(y_2)} P_2(x_2) \\
&= [\mathcal{Q}(\pi)(P_1)](x_1)[\mathcal{Q}(\pi_2)(P_2)](x_2).
\end{aligned}$$

We summarize this with the formula

$$\mathcal{Q}(\langle \pi_1, \pi_2 \rangle)(P_1, P_2) = (\mathcal{Q}(\pi_1)(P_1), \mathcal{Q}(\pi_2)(P_2)). \tag{1.9}$$

2. $\mathcal{Q}_1 \coprod \mathcal{Q}_2 : \mathbf{S}_1 \coprod \mathbf{S}_2 \to \mathbf{Sets}$, a functor that coincides with $\mathcal{Q}_1$ on the $\mathbf{S}_1$ and with $\mathcal{Q}_2$ on $\mathbf{S}_2$.

Example 1.12. Let $X : \Omega \to E_X$ be a finite random variable; it defines a concrete structure $\mathbf{S}$ given by $X \to \mathbf{1}$. Let $\mathcal{Q}_X$ be collection of probability laws on X. The product $\mathbf{S}^{\times n} := \mathbf{S} \times \ldots \times \mathbf{S}$ (n times) represents n independent trials, not necessarily identically distributed. In fact, an element $P \in \mathcal{Q}_X^{\times n}$ is a probability law $P : E_X^n \to [0,1]$ which can be factored as $P_1 P_2 \cdots P_n$, where each $P_i \in \mathcal{Q}_X$.

1.4 Classical models of finite structures

In this section, we formalize the relation between finite information structures and usual probability spaces. A finite information structure (S, E) is said to be ***quasi-concrete*** if for each nonidentity arrow $f : X \to Y$ in $\mathbf{S}$, the map $\mathcal{E}(f)$ is a strict surjection. Concrete information structures are quasi-concrete, but the converse is not always true, as explained in this section.

Recall that $\mathbf{Obs}_{\text{fin}}(\Omega)$ denotes the poset of finite partitions of a set Ω, ordered by the relation of refinement, with arrows implementing the corresponding surjections (see Section 1.1).

Definition 1.13. A ***classical model*** of a quasi-concrete information structure $(\mathbf{S}, E)$ is a triple $(\Omega, \rho_0, \rho^\#)$, where Ω is a set, $\rho_0 : \mathbf{S} \to \mathbf{Obs}_{\text{fin}}(\Omega)$ is a functor, and $\rho^\# : E \to \rho_0$ is a natural transformation such that:

1. ρ_0 is injective on objects;
2. For each $X \in \operatorname{Ob}\mathbf{S}$, the component $\rho^\#_X : E(X) \to \rho_0(X)$ is a bijection;
3. If $X \wedge Y$ exists, $\rho_0(X \wedge Y) = \rho_0(X) \times \rho_0(Y)$.

We also refers to (Ω, ρ) as a representation of $(\mathbf{S}, E)$.

If $(\Omega, \rho_0, \rho^\#)$ is a classical model of $\mathbf{S}$, each observable X in $\mathbf{S}$ can be associated (not uniquely) to a function $\tilde{X}$ on Ω, in such a way that $\rho(X)$ is the partition induced by $\tilde{X}$. Under this representation as functions, all observables commute. It is also possible to introduce quantum models, which respect the noncommutativity of quantum observables, see Section 1.5.

A concrete information structure, as defined in Section 1.1, can be seen as a classical model of an underlying generalized information structure. We now show that, in certain cases, $\lim_{\mathbf{S}} E$ provides a model for a structure $(\mathbf{S}, E)$. In general, if we begin with a concrete structure $\mathbf{S} \subset \mathbf{Obs}_{\text{fin}}(\Omega)$ and forget Ω to obtain a generalized structure $(\Box\mathbf{S}, E = \text{id})$, the set $\lim_{\mathbf{S}} E$ is different from Ω. See Example 1.20.

When $\rho_0(\mathbf{S})$ is an information structure, $(\rho_0, \rho^\#)$ is a morphism in **InfoStr**, but the following example shows that this is not always the case.

Example 1.14. Let $\mathbf{S}$ be the simplicial subcomplex of $\mathbf{\Delta}([\mathbf{3}])$ with maximal faces $\{1, 2\}$ and $\{3\}$; suppose E is such that $E_1 = E_2 = E_3 = \{0, 1\}$, $E_{\{1,2\}} = E_1 \times E_2$, and $E_{\{1,2\}} \to E_i$, $i = 1, 2$ are the canonical projections. The pair $(\mathbf{S}, E)$ is a finite information structure, that can be represented on $\Omega = \{0, 1\}^2$ mapping X_i to the partition induced by the projection $pr_i : \{0, 1\}^2 \to \{0, 1\}$, when $i = 1, 2$, and X_3 to $\{\{(0, 0)\}, \{(0, 0)\}^c\}$. As we established in Remark 1.3, $\rho_0(\mathbf{S})$ is not a concrete information structure. This illustrates the difference between concrete and generalized structures.

Given a structure $(\mathbf{S}, E)$, the limit of E corresponds to

$$\lim_{\mathbf{S}} E := \operatorname{Hom}_{[\mathbf{S},\mathbf{Sets}]}(*, E), \tag{1.10}$$

where $[\mathbf{S}, \mathbf{Sets}]$ is the category of functors from $\mathbf{S}$ to sets and $*$ is the functor that associates to each object a one-point set; equivalently

$$\lim_{\mathbf{S}} E \cong \left\{ (s_Z)_{Z \in \operatorname{Ob}\mathbf{S}} \in \prod_{Z \in \operatorname{Ob}\mathbf{S}} E(Z) \,\middle|\, E(\pi_{YX})(s_X) = s_Y \text{ for all } \pi_{YX} : X \to Y \right\}, \tag{1.11}$$

where s_Z denotes $\varphi(*)$ for any $\varphi \in \operatorname{Hom}_{[\mathbf{S},\mathbf{Sets}]}(*, E)$. The requirements imposed on $(s_Z)_{Z \in \operatorname{Ob}\mathbf{S}}$ in (1.11) are referred hereafter as 'compatibility conditions'. We denote the restriction of each projection $\pi_{E(X)} : \prod_{Z \in \operatorname{Ob}\mathbf{S}} E(Z) \to E(X)$ to $\lim_{\mathbf{S}} E$ by the same symbol. We interpret the limit as all possible combinations of *compatible* measurements.

Proposition 1.15. *If a quasi-concrete structure* $(\mathbf{S}, E)$ *has a classical model* $(\Omega, \rho_0, \rho^\#)$*, then, irrespective of the choice of* $x \in E(X)$*, with* $X \in \operatorname{Ob}\mathbf{S}$*, there exists an element* $s(x) \in \lim_{\mathbf{S}} E$ *such that* $\pi_{E(X)}(s(x)) = x$*.*

Proof. Use the bijection $\rho_X^\# : E(X) \to \rho_0(X)$, to identify $x \in E(X)$ with certain subset $\rho_X^\#(x)$ of Ω. Take any $\omega \in \rho_X^\#(x)$ and then define $s(x)_Z$ as the part $z \in E_Z$ such that $\omega \in \rho_Z^\#(z)$. This section has the desired property. □

Remark 1.16. We could have introduced this necessary condition for representability in the definition of information structure, but there are two good reasons to avoid this. First, it is completely irrelevant for the cohomological computations. Second, and more importantly: information structures are sufficiently flexible to model contextual situations, that arise when data is locally consistent, but globally inconsistent. This happens in different domains, notably in quantum mechanics and in database theory. In the terminology of [1, Sec. 3], a structure is said to be logically contextual at a value $x \in E_X$ if x belongs to no compatible family of measurements (there is no section $s(x) \in \lim_{\mathbf{S}} E$ such that $\pi_{E(X)}(s(x)) = x$), and strongly contextual if E does not accept any global section, i.e. $\lim_{\mathbf{S}} E = \emptyset$.

Definition 1.17. An (arbitrary) information structure is ***noncontextual*** if, for all $X \in \operatorname{Ob} \mathbf{S}$ and all $x \in E(X)$, there exists an element $s(x) \in \lim_{\mathbf{S}} E$ such that $\pi_{E(X)}(s(x)) = x$.

Proposition 1.18. *The product and the coproduct of two noncontextual structures is noncontextual.*

Proof. Let $\mathbf{S}_1$ and $\mathbf{S}_2$ be noncontextual structures. We use the notations in the proof of Proposition 1.8.

Products: Consider a point $(x_1, x_2) \in E(\langle X_1, X_2\rangle)$. There exist sections

$$s^i(x_i) = (s^i_Z(x_i))_{Z \in \operatorname{Ob} \mathbf{S_i}} \in \lim_{\mathbf{S}} E_i \subset \prod_{Z \in \operatorname{Ob} \mathbf{S_i}} E(Z),$$

such $\pi_{E_i(X_i)}(s(x_i)) = x_i$ (for $i = 1, 2$). Note that the vector

$$s(x_1, x_2) := (s^1_{Z_1}(x_1), s^2_{Z_2}(x_2))_{\langle Z_1, Z_2\rangle \in \operatorname{Ob} \mathbf{S}} \in \prod_{\langle Z_1, Z_2\rangle \in \operatorname{Ob} \mathbf{S}} E(\langle Z_1, Z_2\rangle)$$

satisfies all the compatibility conditions and is therefore in $\lim_{\mathbf{S}} E$. By definition, $\pi_{E(\langle X_1, X_2\rangle)}(s(x_1, x_2)) = (x_1, x_2)$.

Coproducts: given $X \in \mathbf{S}_1$, $x \in E(X)$, there exists $s^1(x) = (s_Z(x))_{Z \in \operatorname{Ob} \mathbf{S_1}} \in \lim_{\mathbf{S}_1} E_1$ satisfying $\pi_{E(X)}(s^1(x)) = x$, and similarly for E_2; we can build a new vector $(s_Z(x))_{Z \in \operatorname{Ob} \mathbf{S}} \in \operatorname{Ob} \mathbf{S_1} \sqcup \operatorname{Ob} \mathbf{S_2}$ such that $s_Z = s^i_Z$ if $Z \in \operatorname{Ob} \mathbf{S_i}$; luckily, for $\mathbf{1}$ there is no choice. □

Define $\tilde{\rho}_0 : \mathbf{S} \to \lim_{\mathbf{S}} E$ as follows: associate to $X \in \operatorname{Ob} \mathbf{S}$ the collection $\tilde{\rho}_0(X) := \{\pi^{-1}_{E(X)}(x)\}_{x \in E(X)}$, which is a partition of $\lim_{\mathbf{S}} E$, and none of the parts is $\emptyset$ as long as $\mathbf{S}$ is noncontextual; $\tilde{\rho}^\#_X$ maps $x \in E(X)$ to $\pi^{-1}_{E(X)}(x)$. Given $\pi_{YX} : X \to Y$, there is a corresponding arrow $\tilde{\rho}(X) \to \tilde{\rho}(Y)$ in $\mathbf{Obs}_{\text{fin}}(\Omega)$, which is equivalent to $\tilde{\rho}(Y) \subset \tilde{\rho}(X)$. The existence of such arrow is ensured by the equality

$$\pi^{-1}_{E(Y)}(y) = \bigcup_{x \in E(\pi_{YX})^{-1}(y)} \pi^{-1}_{E(X)}(x). \tag{1.12}$$

It is proved as follows: if $x \in E(\pi_{YX})^{-1}(y)$ and $s \in \pi^{-1}_{E(X)}(x)$, then

$$\pi_{E(Y)}(s) = E(\pi_{YX})(\pi_{E(X)}(s)) = E(\pi_{YX})(x) = y,$$

which means $\cup_{x\in E(\pi_{YX})^{-1}(y)}\pi^{-1}_{E(X)}(x) \subset \pi^{-1}_{E(Y)}(y)$; to prove the other inclusion, take $s = (s_Z)_{Z\in \mathrm{Ob}\,\mathbf{S}} \in \pi^{-1}_{E(Y)}(y)$ and note that s_X must satisfy—by definition—the compatibility condition $E(\pi_{YX})(s_X) = s_Y = y$, thus $s_X \in E(\pi_{YX})^{-1}(y)$ and s itself belong to $\cup_{x\in E(\pi_{YX})^{-1}(y)}\pi^{-1}_{E(X)}(x)$.

Proposition 1.19. *Let* $(\mathbf{S}, E)$ *be a quasi-concrete, noncontextual information structure.*

1. *The pair* $(\lim_{\mathbf{S}} E, \tilde{\rho}_0, \tilde{\rho}^{\#})$ *is a classical model of* $(\mathbf{S}, E)$ *if, and only if, for every pair of variables* X, Y *such that* $X \wedge Y$ *does not exist,* $\tilde{\rho}(X) \neq \tilde{\rho}(Y)$.
2. *If* $(\mathbf{S}, E)$ *has a model* (Ω, ρ, ρ^*), *then* $(\lim_{\mathbf{S}} E, \tilde{\rho}_0, \tilde{\rho}^{\#})$ *is also a model (maybe the same).*

Proof. Claim 1: The "only if" part is straightforward from the definitions of classical model (injectivity of ρ_0). We simply prove sufficiency.

Many properties of models are always verified by $(\lim_{\mathbf{S}} E, \tilde{\rho})$. The noncontextuality implies that each $\pi^{-1}_{E(X)}(x) \neq \emptyset$; we obtain in this way the desired bijection $E(X) \simeq \rho(X)$. To prove property (3) in Definition 1.13, take a diagram $X \leftarrow X\wedge Y \rightarrow Y$, and an arbitrary partition W of $\lim_{\mathbf{S}} E$ that refines $\tilde{\rho}(X)$ and $\tilde{\rho}(Y)$. We have to show that W also refines $\tilde{\rho}(X \wedge Y)$. If W refines $\tilde{\rho}(X)$, each $w \in W$ (w is a subset of $\lim_{\mathbf{S}} E$) is mapped to certain x_w by $\pi_{E(X)}$; analogously, $\pi_{E(Y)}(w) = \{y_w\}$. This means that $\pi_{E(X\wedge Y)}(w) = \{z_w\}$, where z_w is the only point of $E(X \wedge Y)$ that satisfies $E(\pi_{X(X\wedge Y)})(z_w) = x_w$, $E(\pi_{X(X\wedge Y)})(z_w) = y_w$, which means that $w \subset \pi^{-1}_{E(X\wedge Y)}(z_w)$. Thus, W refines $\tilde{\rho}(X \wedge Y)$.

To prove the property 1 in Definition 1.13, consider to variables X, Y such that $X \neq Y$. If their infimum exists, $X \leftarrow X \wedge Y \rightarrow Y$ in $\mathbf{S}$, then $\tilde{\rho}(X) \neq \tilde{\rho}(Y)$; we prove it by contradiction. Each point in $E(X \wedge Y)$ is indexed by a pair $(x, y) \in E(X) \times E(Y)$; a point $w \in E(\pi_{X\wedge Y})^{-1}(x, y) \subset \lim_{\mathbf{S}} E$ goes to x under $\pi_{E(X)}$ and to y under $\pi_{E(Y)}$. If $\tilde{\rho}(Y) = \tilde{\rho}(Y)$, there is a bijection $y : E(X) \rightarrow E(Y), x \mapsto y(x)$ in such a way that $\pi^{-1}_{E(X)}(x) = \pi^{-1}_{E(Y)}(y(x))$. Therefore, the points of $X \wedge Y$ would be indexed by $(x, y(x))$, with $x \in X$, in contradiction with $\pi_{X(X\wedge Y)}$ being a strict surjection. If moreover we suppose that for each pair of variables such that $X \wedge Y$ does not exist $\tilde{\rho}(X) \neq \tilde{\rho}(Y)$, then $\tilde{\rho}$ is injective on objects.

Claim 2: Here, we denote $\{X = x\}$ the image of $x \in E(X)$ under the bijection $E(X) \xrightarrow{\sim} \rho(X)$ given by property (2) in Definition 1.13. Each element $\omega \in \Omega$ defines a section $s(\omega) = (s(\omega)_X)_{X\in \mathrm{Ob}\,\mathbf{S}} \in \lim_{\mathbf{S}} E$, such that $s(\omega)_X = x$ iff $\omega \in \{X = x\}$. It is clear that several ω could give the same section. Suppose now that $\tilde{\rho}(X) \xrightarrow{\sim} \tilde{\rho}(Y), x \mapsto y(x)$. If $\omega \in \{X = x\} \subset \Omega$, then $s(\omega) \in \pi^{-1}_{E(X)}(x) = \pi^{-1}_{E(Y)}(y(x))$. We conclude that $\omega \in \{Y = y(x)\}$, and therefore $\rho(X)$ and $\rho(Y)$ are the same partition, only with different labels. For ρ is injective on objects, $X = Y$. Use the first part to conclude. □

Example 1.20. Let $\Omega = \{1, 2, 3, 4\}$. Define the partitions $X_i = \{\{i\}, \Omega \setminus \{i\}\}$, for $i = 1, ..., 4$, and $\mathbf{S}$ as let $\mathbf{S}$ be the concrete information structure that includes only the partitions X_1, X_2, X_3, X_1X_2, and X_2X_3. The corresponding general information structure has as variables the free category $\square\mathbf{S}$ generated by the graph

$$\begin{tikzcd}
 & & \mathbf{1} & & \\
X_1 \arrow[urr] & & X_2 \arrow[u] & & X_3 \arrow[ull] \\
 & X_1X_2 \arrow[ul] \arrow[ur] & & X_2X_3 \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

and the corresponding functor E can be represented by the diagram

$$\begin{tikzcd}
 & & \{x_{\{1,2,3,4\}}\} & & \\
\{x_{\{1\}}, x_{\{2,3,4\}}\} \arrow[urr] & & \{x_{\{2\}}, x_{\{1,3,4\}}\} \arrow[u] & & \{x_{\{3\}}, x_{\{1,2,4\}}\} \arrow[ull] \\
 & \{x_{\{1\}}, x_{\{2\}}, x_{\{3,4\}}\} \arrow[ul] \arrow[ur] & & \{x_{\{2\}}, x_{\{3\}}, x_{\{1,4\}}\} \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

Each arrow corresponds to a surjection of finite sets, that sends x_I to x_J when $I \subset J$. These are just the surjections of partitions in the original $\mathbf{S}$. In this case, $\lim_{\mathbf{S}} E \subset \{*\} \times E(X_1) \times E(X_2) \times E(X_3) \times E(X_1X_2) \times E(X_2X_3)$ corresponds to the set

$$\begin{aligned}
\lim_{\mathbf{S}} E = \{&(x_{\{1,2,3,4\}}, x_{\{1\}}, x_{\{1,3,4\}}, x_{\{3\}}, x_{\{1\}}, x_{\{3\}}), (x_{\{1,2,3,4\}}, x_{\{1\}}, x_{\{1,3,4\}}, x_{\{1,2,4\}}, x_{\{1\}}, x_{\{1,4\}}), \\
&(x_{\{1,2,3,4\}}, x_{\{2,3,4\}}, x_{\{2\}}, x_{\{1,2,4\}}, x_{\{2\}}, x_{\{2\}}), (x_{\{1,2,3,4\}}, x_{\{2,3,4\}}, x_{\{1,3,4\}}, x_{\{3\}}, x_{\{3,4\}}, x_{\{3\}}), \\
&(x_{\{1,2,3,4\}}, x_{\{2,3,4\}}, x_{\{1,3,4\}}, x_{\{1,2,4\}}, x_{\{3,4\}}, x_{\{1,4\}})\}.
\end{aligned}$$

The difference between Ω and $\lim_{\mathbf{S}} E$ is explained by the presence of

$$(x_{\{1,2,3,4\}}, x_{\{1\}}, x_{\{1,3,4\}}, x_{\{3\}}, x_{\{1\}}, x_{\{3\}});$$

this measurement (where $X_1 = x_{\{1\}}$, $X_3 = x_{\{3\}}$) is impossible in the concrete structure $\mathbf{S} \subset \mathbf{Obs}_{\text{fin}}(\Omega)$, but the observables in $(\square\mathbf{S}, E)$ cannot distinguish between the points 1 and 3, a sort of nonseparability. In fact, if we also include X_1X_3 at the beginning, we obtain $\Omega \cong \lim_{\mathbf{S}} E$.

Example 1.21. Consider the information structure given by

$$\begin{tikzcd}
 & \{*\} & \\
\{x_{\{00,01\}}, x_{\{10,11\}}\} \arrow[ur] & & \{x_{\{00,10\}}, x_{\{01,11\}}\} \arrow[ul] \\
\{x_{\{00\}}, x_{\{10\}}, x_{\{1,0\}}, x_{\{1,1\}}\} \arrow[u] \arrow[urr] & & \{x_{\{00\}}, x_{\{10\}}, x_{\{1,0\}}, x_{\{1,1\}}\} \arrow[u] \arrow[ull]
\end{tikzcd}$$

where we suppose again that $x_I \mapsto x_J$ when $I \subset J$. Such structure cannot be modeled by its inverse limit, since the two minimal variables induce the same partition.

We study now the models associated to products and coproducts of structures that already accept a model.

Let Ω_1 and Ω_2 be sets. Given collections $\mathfrak{A} = \{A_i\}_i$ of subsets of Ω_1 and $\mathfrak{B} = \{B_j\}_j$ of subsets of Ω_2, denote by $\mathfrak{A} \times \mathfrak{B}$ the collection $\{\, A_i \times B_j \mid A_i \in \mathfrak{A} \text{ and } B_j \in \mathfrak{B} \,\}$ of subsets of $\Omega_1 \times \Omega_2$. If $\mathfrak{A}$ and $\mathfrak{B}$ are partitions, then $\mathfrak{A} \times \mathfrak{B}$ is a partition too.

Let $(\Omega_i, \rho_i^0, \rho_i^\#)$ be a model of $(\mathbf{S}_i, E_i)$, for $i = 1, 2$. Associate to each variable $\langle X_1, X_2\rangle \in \operatorname{Ob} \mathbf{S_1} \times \mathbf{S_2}$ the partition of $\Omega_1 \times \Omega_2$ given by

$$\rho_\times^0(\langle X_1, X_2\rangle) = \rho_1^0(X_1) \times \rho_2^0(X_2). \tag{1.13}$$

The map $\rho_\times^\# : E(\langle X_1, X_2\rangle) \to \rho_\times(\langle X_1, X_2\rangle)$, where $E(\langle X_1, X_2\rangle) = E(X_1) \times E(X_2)$, is $(x_1, x_2) \mapsto \rho_1^\#(x_1) \times \rho_2^\#(x_2)$.

Analogously, for each $X \neq \mathbf{1}$ in $\operatorname{Ob} \mathbf{S_1} \coprod \mathbf{S_2}$, let us define the partition of $\Omega_1 \times \Omega_2$ given by

$$\rho_\coprod^0(X) = \begin{cases} \rho_1^0(X) \times \{\Omega_2\} & \text{if } X \in \operatorname{Ob} \mathbf{S_1} \\ \{\Omega_1\} \times \rho_2^0(X) & \text{if } X \in \operatorname{Ob} \mathbf{S_2} \end{cases}. \tag{1.14}$$

The map $\rho_\coprod$ is $x \mapsto \rho_1^\#(x) \times \{\Omega_2\}$ or $x \mapsto \{\Omega_1\} \times \rho_2^\#(x)$ accordingly. By convention, $\rho_\coprod^0(\mathbf{1}) = \{\Omega_1 \times \Omega_2\}$.

Proposition 1.22. *Let (Ω_i, ρ_i) be a classical model of $(\mathbf{S}_i, E_i)$, for $i = 1, 2$. Then*

1. *$(\Omega_1 \times \Omega_2, \rho_\times)$ is a classical model of $(\mathbf{S}_1, E_1) \times (\mathbf{S}_2, E_2)$;*
2. *$(\Omega_1 \times \Omega_2, \rho_\coprod)$ is a classical model of $(\mathbf{S}_1, E_1) \coprod (\mathbf{S}_2, E_2)$.*

It depends on the following lemma.

Lemma 1.23.
1. *If $\mathfrak{A} = \{A_i\}_i$ and $\mathfrak{A}' = \{A'_j\}_j$ are finite partitions of a set Ω, then $\sigma(\mathfrak{A}, \mathfrak{A}') = \sigma(\{A_i \cap A'_j\}_{i,j})$, and $\{A_i \cap A'_j\}_{i,j}$ are the atoms of $\sigma(\mathfrak{A}, \mathfrak{A}')$.*
2. *If $\mathfrak{A} = \{A_i\}_i$, $\mathfrak{A}' = \{A'_j\}_j$ are two finite partitions of Ω_1 and $\mathfrak{B} = \{B_l\}_l$, $\mathfrak{B}' = \{B_m\}_m$ two finite partitions of Ω_2, then $(\mathfrak{A} \times \mathfrak{B})(\mathfrak{A}' \times \mathfrak{B}') = \mathfrak{A}\mathfrak{A}' \times \mathfrak{B}\mathfrak{B}'$, where juxtaposition of partitions denotes their product in $\mathbf{Obs}_{\text{fin}}(\Omega)$, as introduced in Section 1.1.*

Proof.
1. On one hand, note that each set $A_i \cap A'_j$ is contained in $\sigma(\mathfrak{A}, \mathfrak{A}')$, therefore $\sigma(\{A_i \cap A'_j\}_{i,j}) \subset \sigma(\mathfrak{A}, \mathfrak{A}')$. On the other, each generator $A_i \in \mathfrak{A}$ of $\sigma(\mathfrak{A}, \mathfrak{A}')$ can be written as
$$A_i = A_i \cap \Omega = A_i \cap \left(\bigcup_j A'_j \right) = \bigcup_j (A_i \cap A'_j),$$
and similarly for the generators $A'_j \in \mathfrak{A}'$, which implies that $\sigma(\mathfrak{A}, \mathfrak{A}') \subset \sigma(\{A_i \cap A'_j\}_{i,j})$. The reader can verify that $\{A_i \cap A'_j\}_{i,j}$ are atoms.
2. The previous result can be read as $\mathfrak{A}\mathfrak{A}' = \{A_i \cap A'_j\}_{i,j}$. The set-theoretical identity
$$(A_i \times B_l) \cap (A'_j \times B'_m) = (A_i \cap A'_j) \times (B_l \cap B'_m), \tag{1.15}$$
implies that the atoms of $(\mathfrak{A} \times \mathfrak{B})(\mathfrak{A}' \times \mathfrak{B}')$ and $\mathfrak{A}\mathfrak{A}' \times \mathfrak{B}\mathfrak{B}'$ coincide.

□

Proof of Proposition 1.22. Most verifications are almost immediate from the definitions. We simply prove that $\rho^0_\times(\langle X_1, X_2\rangle \wedge \langle Y_1, Y_2\rangle) = \rho^0_\times(\langle X_1, X_2\rangle)\rho^0_\times(\langle Y_1, Y_2\rangle)$. Note that

$$\begin{aligned}
\rho^0_\times(\langle X_1, X_2\rangle \wedge \langle Y_1, Y_2\rangle) &= \rho^0_\times(\langle X_1 \wedge Y_1, X_2 \wedge Y_2\rangle)\\
&= \rho^0_1(X_1 \wedge Y_1) \times \rho^0_2(X_2 \wedge Y_2)\\
&= \rho^0_1(X_1)\rho^0_1(Y_1) \times \rho^0_2(X_2)\rho^0_2(Y_2)\\
&= (\rho^0_1(X_1) \times \rho^0_2(X_2))(\rho^0_1(Y_1) \times \rho^0_2(Y_2))\\
&= \rho^0_\times(\langle X_1, X_2\rangle)\rho^0_\times(\langle Y_1, Y_2\rangle)
\end{aligned}$$

The first equality comes from the construction of $\mathbf{S}_1 \times \mathbf{S}_2$; the second, from the definition of $\rho^0_\times$; the third, from the fact that ρ^0_1 and ρ^0_2 are models; the fourth equality is just a consequence of Lemma 1.23, and the fifth is just a rewriting of the previous one. □

The partitions of $\Omega_1 \times \Omega_2$ in the image of $\rho^0_\coprod$ are also in the image of $\rho^0_\times$. This is consistent with the existence of a morphism of structures $\phi : (\mathbf{S}_1, E_1) \coprod (\mathbf{S}_2, E_2) \to (\mathbf{S}_1, E_1) \times (\mathbf{S}_2, E_2)$, with ϕ_0 given at the level of objects by the injection

$$X \mapsto \begin{cases} \mathbf{1}_{\mathbf{S}_1\times\mathbf{S}_2} & \text{if } X = \mathbf{1}_{\mathbf{S}_1 \coprod \mathbf{S}_2} \\ \langle X, \mathbf{1}_{\mathbf{S}_2}\rangle & \text{if } X \in \operatorname{Ob}\mathbf{S_1} \\ \langle \mathbf{1}_{\mathbf{S}_1}, X\rangle & \text{if } X \in \operatorname{Ob}\mathbf{S_2} \end{cases}, \tag{1.16}$$

and the corresponding components $\phi^\#_X$ being the obvious bijections: $E_1(X) \to E_1(X) \times \{*\}$ when $X \in \operatorname{Ob}\mathbf{S_1}$ or $E_2(X) \to \{*\} \times E_2(X)$ when $X \in \operatorname{Ob}\mathbf{S_2}$. The model $(\Omega_1 \times \Omega_2, \rho^0_\times, \rho^\#_\times)$ on $(\mathbf{S}_1, E_1) \times (\mathbf{S}_2, E_2)$ restricts then to a model $(\Omega_1 \times \Omega_2, \rho^0_\times \circ \phi_0)$ on $(\mathbf{S}_1, E_1) \coprod (\mathbf{S}_2, E_2)$, that coincides with $(\Omega_1 \times \Omega_2, \rho^0_\coprod)$. This is clearly a particular example of a more general procedure to restrict models, valid for any morphism of structures $\phi = (\phi_0, \phi^\#)$ such that ϕ_0 is injective on objects and each $\phi^\#_X$ is a bijection; therefore, it makes sense to call these morphims ***embeddings***.

1.5 Quantum probability and quantum models

Let V be a finite dimensional Hilbert space: a complex vector space with a positive definite hermitian form $\langle\cdot,\cdot\rangle$. In the quantum setting, random variables are generalized by endomorphisms of V (operators). An operator H is called hermitian if for all u, $v \in V$, one has $\langle u, Hv\rangle = \langle Hu, v\rangle$. A quantum observable is a Hermitian operator: the result of a quantum experiment is supposed to be an eigenvalue of such operator, that is always a real number.

A fundamental result of linear algebra, the Spectral Theorem [36, Sec. 79], says that each hermitian operator Z can be decomposed as weighted sum of positive hermitian projectors $Z = \sum_{j=1}^K z_j V_j$ where $z_1, ..., z_K$ are the (pairwise distinct) *real* eigenvalues of Z. Each V_j is the projector on the eigenspace spanned by the eigenvectors of z_j; the dimension of this subspace equals the multiplicity of z_j as eigenvalue. As hermitian projectors, they satisfy the equation $V_j^2 = V_j$ and $V_j^* = V_j$. They are also mutually orthogonal ($V_jV_k = 0$ for integers j, k), and their sum equals the identity,

$\sum_{1\leq j\leq K} V_j = Id_V$. This decomposition of Z is not necessarily compatible with the preferred basis of V (that diagonalizes its hermitian product).

In analogy to the classical case, we consider as equivalent two hermitian operators that define the same orthogonal decomposition $\{V_j\}_j$ of V by means of the Spectral Theorem, ignoring the particular eigenvalues. For us, observable and orthogonal decomposition (sometimes just 'decomposition', for brevity) are then interchangeable terms. In what follows, we denote by V_A both the subspace of V and the orthogonal projector on it. A decomposition $\{V_\alpha\}_{\alpha\in A}$ is said to refine $\{V'_\beta\}_{\beta\in B}$ if each V'_β can be expressed as sum of subspaces $\{V_\alpha\}_{\alpha\in A_\beta}$, for certain $A_\beta \subseteq A$. In that case we say also that $\{V_\alpha\}_{\alpha\in A}$ divides $\{V'_\beta\}_{\beta\in B}$, and we write $\{V_\alpha\}_{\alpha\in A} \to \{V'_\beta\}_{\beta\in B}$. With this arrows, direct sums decompositions form a category called $\mathbf{Orth}(V)$.

Definition 1.24. A ***quantum model*** of an information structure $\mathbf{S}$ is a triple $(V, \rho_0, \rho^\#)$, where V is a finite dimensional Hilbert space and $\rho : \mathbf{S} \to \mathbf{Orth}(V)$ is a functor, and $\rho^\# : E \to \rho_0$ is a natural transformation such that:

1. ρ_0 is injective on objects;
2. for each $X \in \mathrm{Ob}\,\mathbf{S}$, the component $\rho^\#_X : E(X) \to \rho_0(X)$ is a bijection;
3. if $X \wedge Y$ exists, $\rho(X \wedge Y) = \rho(X) \times \rho(Y)$.

A quantum model gives rise to a quantum information structure as defined in [10]. All the cohomological computations in this thesis concern classical probabilities, but the general constructions in Chapter 2 only depend on the abstract structure and are equally valid in the quantum case.

Chapter 2

Topoi and cohomology

We shall use the information structure and its associated functors to construct a Grothendieck topos, where cohomology can be defined. Section 2.1 explains the basic notions related to this approach, abelian categories and derived functors, in order to give a general definition of cohomology. Section 2.2 introduces the main definition of the thesis, *information cohomology*, while the rest of the chapter develops tools to compute it.

2.1 Preliminaries on homological algebra

In what follows, we denote sometimes the monomorphisms by $\rightarrowtail$ and epimorphisms by $\twoheadrightarrow$. Some important definitions of category theory can be found in Appendix A.

Nothing in this section is original, and it can be skipped by any reader already familiarized with homological algebra.

Abelian categories and ∂-functors where defined by Grothedieck in [35]. For more details about the history of these concepts see Mac Lane [60, p. 257]. Sections 2.1.1, 2.1.2 and 2.1.3 are based on the translations of Tamme [83]. The case of sheaves on topological spaces is developed in [61, Ch. II]. Section 2.1.4 is based on [81].

2.1.1 Additive categories

We first examine categories in which suitable pairs of morphisms can be added.

An ***additive category*** $\mathbf{C}$ is a class of objects A, B, C... (denoted $\mathrm{Ob}\,\mathbf{C}$) together with

1. A familly of disjoint abelian groups $\mathrm{Hom}(A,B)$, one for each pair of objects. We write $\alpha : A \to B$ for $\alpha \in \mathrm{Hom}(A,B)$ and call α a morphism of $\mathbf{C}$.
2. To each triple of objects A, B and C, a homomorphism
$$\mathrm{Hom}(B,C) \otimes \mathrm{Hom}(A,B) \to \mathrm{Hom}(A,C)$$
of abelian groups, called composition. The image of $\beta \otimes \alpha$ under composition is written $\beta\alpha$, and called its composite.
3. To each object A a morphism $1_A : A \to A$, called the identity of A.

These data are subject to the following four axioms:

- *Associativity*: If $\alpha : A \to B$, $\beta : B \to C$ and $\gamma : C \to D$, then
$$\gamma(\beta\alpha) = (\gamma\beta)\alpha. \tag{2.1}$$

- *Identities*: If $\alpha : A \to B$, then

$$\alpha 1_A = \alpha = 1_B \alpha. \tag{2.2}$$

- *Zero Object*: There is an object $0'$ such that $\mathrm{Hom}(0', 0')$ is the zero group.
- *Finite Direct Sums*: To each pair of objects A_1, A_2 there exists an object B and four morphisms forming a diagram

$$A_1 \underset{\iota_1}{\overset{\pi_1}{\leftrightarrows}} B \underset{\iota_2}{\overset{\pi_2}{\rightleftarrows}} A_2 \tag{2.3}$$

with $\pi_1\iota_1 = 1_{A_1}$, $\pi_2\iota_2 = 1_{A_2}$ and $\iota_1\pi_1 + \iota_2\pi_2 = 1_B$.
The diagram (2.3) determines B up to equivalence.

This definition is very similar to the standard one of category, but assuming also the existence of zero objects and direct sums/products, and the possibility to add morphisms; composition is required to be bilinear in both arguments. [1]

A functor $F : \mathbf{C} \to \mathbf{C}'$ between additives categories is called additive if, given any $A, B \in \mathrm{Ob}\,\mathbf{C}$ and $\alpha, \beta \in \mathrm{Hom}(A, B)$, the equality $F(\alpha + \beta) = F(\alpha) + F(\beta)$ holds; there is an analogous definition for multi-functors. As in the standard case of modules, $\mathrm{Hom}(-, -)$ is an additive bifunctor from $\mathbf{C}$ to the category of abelian groups. It is contravariant in the first component and covariant in the second.

Let $\mathbf{C}$ be an additive category and $\alpha : A \to B$ a morphism in $\mathbf{C}$. Recall that, by definition, a morphism μ is a monomorphism (or injective) if the induced application $\mu_* : \mathrm{Hom}(C, A) \to \mathrm{Hom}(C, B)$, given by $\nu \mapsto \mu\nu$, is injective for all $C \in \mathrm{Ob}\,\mathbf{C}$. Therefore, μ is a monomorphism if and only if there is no morphism $\xi \neq 0$, such that $\mu\xi = 0$. We call generalized kernel of α any monomorphism $\iota : A' \to A$ such that any $\xi : C \to A$ satisfying $\alpha\xi = 0$ can be factorized as $C \to A' \xrightarrow{\iota} A$. This morphism is defined up to equivalence (see Section A.2); hence, between the generalized kernels (if there is any), there is exactly one subobject of A: we call it ***kernel of*** α and denote it by $\ker \alpha$. A ***cokernel of*** α can be defined dually; it is a quotient of B (if it exists). We call ***image*** of α ($\mathrm{im}\,\alpha$) the kernel of its cokernel; it is a subobject of B. The ***coimage*** of α ($\mathrm{coim}\,\alpha$)is the cokernel of its kernel; it is a quotient of A. If α admits an image and a coimage, then there exists a unique morphism $\bar{\alpha} : \mathrm{coim}\,\alpha \to \mathrm{im}\,\alpha$ such that α equals the composition $A \to \mathrm{coim}\,\alpha \to \mathrm{im}\,\alpha \to B$, where the extreme morphisms are canonic (see [35]).

2.1.2 Abelian categories

An ***abelian category*** is an additive category with the following two properties:

(AB1) Each morphism in $\mathbf{C}$ has a kernel and a cokernel.

(AB2) For each morphism α in $\mathbf{C}$ the canonical morphism $\bar{\alpha} : \mathrm{coim}(\alpha) \to \mathrm{im}(\alpha)$ is an isomorphism.

As a consequence, in an abelian category each bijective morphism is an isomorphism. The most basic example of an abelian category is $\mathbf{Ab}$, the category of abelian groups; here all this notions reduce to the classical ones.

---

[1]For foundational reasons, it is convenient to add as axiom that for each object A in the category, the subobjects of A form a set, in opposition to a proper class. The same is supposed for equivalence classes of n-fold extensions from A to C, see [60, p. 253].

By definition, a sequence $A \xrightarrow{\alpha} B \xrightarrow{\beta} C$ of morphisms in an abelian category is ***exact*** if $\ker(\beta) = \operatorname{im}(\alpha)$. A sequence $0 \to A \xrightarrow{\alpha} B \xrightarrow{\beta} C \to 0$ is called short exact if α is a monomorphism, β is an epimorphism and $\ker(\beta) = \operatorname{im}(\alpha)$; it can be also denoted $A \rightarrowtail B \twoheadrightarrow C$.

Under these axioms, each morphism α fits into a commutative diagram

$$\begin{tikzcd}
\ker\alpha \arrow[r, tail] & A \arrow[r, two heads] \arrow[drr, "\alpha"'] & \operatorname{coim}\alpha \arrow[r, "\simeq"] & \operatorname{im}\alpha \arrow[d, tail] \\
 & & & B \arrow[d, two heads] \\
 & & & \operatorname{coker}\alpha
\end{tikzcd} \tag{2.4}$$

where the row and the column are short exact sequences (the dots designate unnamed objects). Here, "$\ker\alpha$" stands for an equivalence class of morphisms, and similarly with the other arrows. Therefore, α admits an standard factorization $\alpha = \lambda\sigma$, where σ is an epimorphism and λ a monomorphism; this factorization is uniquely determined up to equivalences.

A covariant functor $F : \mathbf{C} \to \mathbf{C}'$ between abelian categories is called ***left exact*** (resp. ***right exact***) if for each exact sequence $0 \to A' \to A \to A'' \to 0$ in $\mathbf{C}$, the sequence $0 \to F(A') \to F(A) \to F(A'')$ (resp. the sequence $F(A') \to F(A) \to F(A'') \to 0$) is exact in $\mathbf{C}'$. The notion extend to contravariant functors: for example, if the result of applying G is $0 \to G(A'') \to G(A) \to G(A')$, then G is called left exact.

Proposition 2.1 (cf. [35]). *If the sequence $A \rightarrowtail B \twoheadrightarrow C$ in $\mathbf{C}$ is exact, then the sequence*

$$0 \to \operatorname{Hom}(X, A) \to \operatorname{Hom}(X, B) \to \operatorname{Hom}(X, C) \tag{2.5}$$

of abelian groups is exact, for each $X \in \operatorname{Ob}\mathbf{C}$.

In the context of abelian categories, I is called a ***injective object*** if all diagrams of the form

$$\begin{tikzcd}
A \arrow[r, tail] \arrow[d] & B \\
I &
\end{tikzcd}$$

admit an extension

$$\begin{tikzcd}
A \arrow[r, tail] \arrow[d] & B \arrow[dl, dashed] \\
I &
\end{tikzcd}$$

An abelian category $\mathbf{C}$ is said to have ***enough injectives*** if for each object $A \in \operatorname{Ob}\mathbf{C}$ there exists a monomorphism from A into an injective object of $\mathbf{C}$.

Proposition 2.2 (cf. [60] or [95]). *If $\mathbf{C}$ is an abelian category and $I \in \operatorname{Ob}\mathbf{C}$, the following statements are equivalent:*

1. *I is an injective object;*
2. *the left exact functor $\operatorname{Hom}(-, I)$ is exact.*

Analogously, an object P is called projective is each diagram of the form

$$\begin{tikzcd} & P \arrow[d] \\ A \arrow[r, two heads] & B \end{tikzcd}$$

can be extended to

$$\begin{tikzcd} & P \arrow[d] \arrow[dl, dashed] \\ A \arrow[r, two heads] & B \end{tikzcd}$$

There is a correspondent notion of enough projectives.

2.1.3 Derived functors

Let **C** be an abelian category and **C**′ an additive category. A covariant ***∂-functor*** from **C** to **C**′ is a system $T = (T^i)_{i\geq 0}$ of covariant additive functors $T^i : \mathbf{C} \to \mathbf{C}'$ together with a connecting morphism $\partial : T^i(A'') \to T^{i+1}(A')$ defined for each $i \geq 0$ and each short exact sequence $0 \to A' \to A \to A'' \to 0$, satisfying the following properties:

1. Given a commutative diagram with exact rows

$$\begin{tikzcd} 0 \arrow[r] & A' \arrow[r] \arrow[d] & A \arrow[r] \arrow[d] & A'' \arrow[r] \arrow[d] & 0 \\ 0 \arrow[r] & B' \arrow[r] & B \arrow[r] & B'' \arrow[r] & 0 \end{tikzcd}$$

   in **C**, the diagram

$$\begin{tikzcd} T^i(A'') \arrow[r, "\partial"] \arrow[d] & T^{i+1}(A') \arrow[d] \\ T^i(B'') \arrow[r, "\partial"] & T^{i+1}(B') \end{tikzcd}$$

   is commutative for all $i \geq 0$.
2. Given an exact sequence $0 \to A' \to A \to A'' \to 0$ in **C**, the long sequence

$$0 \to T^0(A') \to T^0(A) \to T^0(A'') \to T^1(A') \to T^1(A) \to \dots \qquad (2.6)$$

   is a complex in **C**′ (the compositions of two arrows gives 0).

In case **C**′ is abelian too, the ∂-functor T is called ***exact*** if for every exact sequence $0 \to A' \to A \to A'' \to 0$ in **C**, the long sequence (2.6) is exact. The exact ∂-functors are called ***cohomological functors***. The reason is clear if one compares the definition above with the classic Eilenberg–Steenrod axioms for cohomology of topological spaces. The main computational features of both theories are the presence of connecting morphisms and long exact sequences.

Given two ∂-functors $T, T' : \mathbf{C} \to \mathbf{C}'$, a morphism from T to T' is a system $f = (f^i)_{i\geq 0}$ of functorial morphisms (natural transformations) $f^i : T^i \to T'^i$ which commute naturally with ∂. This means that, for any exact sequence $0 \to A' \to A \to$

$A'' \to 0$ in $\mathbf{C}$, the following diagram is commutative:

$$\begin{array}{ccc} T^i(A'') & \xrightarrow{\partial} & T^{i+1}(A') \\ {\scriptstyle f^i(A'')}\downarrow & & \downarrow{\scriptstyle f^{i+1}(A')} \\ T'^i(A'') & \xrightarrow{\partial} & T'^{i+1}(A') \end{array}$$

A ∂-functor $T = (T^i)_{i\geq 0}$ from $\mathbf{C}$ to $\mathbf{C}'$ is called ***universal*** if each natural transformation $f^0 : T^0 \to T'^0$ has one and only one extension to a morphism $f : T \to T'$.

By the very definition, given a left exact and additive covariant functor $F : \mathbf{C} \to \mathbf{C}'$, there is a unique universal ∂-functor from $\mathbf{C}$ to $\mathbf{C}'$ extending F; it is called the right derived functor of F and denoted by $(R^iF)_{i\geq 0}$.

Theorem 2.3. *Let $\mathbf{C}$ be an abelian category with enough injectives, and let $\mathbf{C}'$ be an abelian category. Then for each left exact additive covariant functor $F : \mathbf{C} \to \mathbf{C}'$ the right derived functor $(R^iF)_{i\geq 0}$ exists.*

Proof. For the proof, see [83, p. 11]. We just sketch here the main points. As $\mathbf{C}$ has enough injective objects, each object $A \in \mathrm{Ob}\,\mathbf{C}$ has an injective resolution. This means that there is a exact sequence

$$I^*(A): \quad 0 \to A \to I^0 \to I^1 \to I^2 \to \dots$$

where each I^i is an injective object of $\mathbf{C}$. This resolution I is functorial.

We can apply the functor F to this injective resolution of A, and define

$$\begin{aligned} R^0F(A) &= \ker(F(I^0) \to F(I^1)) \\ R^iF(A) &= \frac{\ker(F(I^i) \to F(I^{i+1}))}{\mathrm{im}(F(I^{i-1}) - F(I^i))}, \quad \text{for } i \geq 1 \end{aligned}$$

One shows that R^i is independent of the injective resolution, and functorial on $\mathbf{C}$. It has also the required properties of ∂-functors. As F is left exact, $R^0F = F$. □

Remark that $\mathrm{Hom}(A, -)$ is an additive covariant functor, and left exact in virtue of 2.5. The corresponding right derived functor is called $\mathrm{Ext}^n(A, -)$. This is the cohomological functor of our interest.

The theorem above has an analogous version for projective resolutions.

Theorem 2.4. *Let $\mathbf{C}$ be an abelian category with enough projectives, and let $\mathbf{C}'$ be an abelian category. Then for each left exact additive contravariant functor $F : \mathbf{C} \to \mathbf{C}'$ the right derived functor $(R^iF)_{i\geq 0}$ exists.*

2.1.4 Sheaves of modules

We have a general setting for homological algebra, given by abelian categories and cohomological functors. In this section, we develop an important example of abelian category: sheaves of modules. We shall see later that our information-theoretical constructions are naturally related to them.

Let $\mathbf{C}$ be a category. A ***presheaf of sets*** is any contravariant functor $\mathscr{F}$ from $\mathbf{C}$ to **Sets**, the category of sets. A morphism of presheaves $\phi : \mathscr{F} \to \mathscr{G}$ is a natural transformation of functors. Presheaves of sets and their morphisms form a new category, denoted by $\mathbf{PSh}(\mathbf{C})$. By definition, we say the ϕ is injective (resp. surjective) if for every $X \in \operatorname{Ob}\mathbf{C}$, the map $\phi(X) : \mathscr{F}(X) \to \mathscr{G}(X)$ is injective (resp. surjective).

Proposition 2.5. *The injective morphisms defined above are exactly the monomorphisms of* $\mathbf{PSh}(\mathbf{C})$. *The surjective morphisms are exactly the epimorphisms of* $\mathbf{PSh}(\mathbf{C})$.

It is possible to define a topology on a category, obtaining a site. Presheaves that are 'well-behaved' for this topology are called sheaves. Moreover, every category admits a trivial topology, such that every presheaf is a sheaf. As we shall use the trivial topology over our information structure $\mathbf{S}$, the general definitions of site and sheaf will not play a special role in the theory, and we omit them. For details, see [32, Ch. 0]. If $\mathbf{C}$ is a site, we can consider the full subcategory of $\mathbf{PSh}(\mathbf{C})$, whose objects are the sheaves; this category is denoted by $\mathbf{Sh}(\mathbf{C})$.

Abelian presheaves are presheaves that take values in abelian groups. They form an abelian category (for a proof, see [60, Ch. 9, Prop. 3.1]). A morphism of abelian presheaves $\phi : \mathscr{F} \to \mathscr{G}$ is a natural transformation between $\mathscr{F}$ and $\mathscr{G}$ that induces a homomorphism of abelian groups $\phi(X) : \mathscr{F}(X) \to \mathscr{G}(X)$ on every $X \in \operatorname{Ob}\mathbf{C}$. Given a morphism $\phi : \mathscr{F} \to \mathscr{G}$, the kernel of ϕ is the abelian presheaf $X \mapsto \ker\{\phi : \mathscr{F}(X) \to \mathscr{G}(X)\}$ and its cokernel is $X \mapsto \operatorname{coker}\{\phi : \mathscr{F}(X) \to \mathscr{G}(X)\}$. One has coim = im, because it holds over each $X \in \operatorname{Ob}\mathbf{C}$. Moreover, a sequence of presheaves $\mathscr{F}_1 \to \mathscr{F}_2 \to \mathscr{F}_3$ is exact if $\mathscr{F}_1(X) \to \mathscr{F}_2(X) \to \mathscr{F}_3(X)$ is exact as a sequence of groups over every $X \in \operatorname{Ob}\mathbf{C}$. Given a site $\mathbf{C}$, the category of abelian sheaves (denoted by $\mathbf{Ab}(\mathbf{C})$) is the full subcategory of $\mathbf{PAb}(\mathbf{C})$ of those abelian presheaves whose underlying presheaves of sets are sheaves.

If $\mathbf{C}$ is a site and $\mathscr{O}$ is a sheaf of rings on $\mathbf{C}$, the pair $(\mathbf{C}, \mathscr{O})$ is called a ***ringed site*** and $\mathscr{O}$, the structure ring. The pair $(\mathbf{Sh}(\mathbf{C}), \mathscr{O})$ is called a ***ringed topos***. There exist appropriate notions of morphisms between ringed sites or ringed topos, cf. [81, Modules on sites, Secs. 6, 7].

Given a ringed site $(\mathbf{C}, \mathscr{O})$, a ***sheaf of*** $\mathscr{O}$***-modules*** is given by an abelian sheaf $\mathscr{F}$ together with a map of presheaves of sets $\mathscr{O} \times \mathscr{F} \to \mathscr{F}$, such that for every $X \in \operatorname{Ob}\mathbf{C}$, the map $\mathscr{O}(X) \times \mathscr{F}(X) \to \mathscr{F}(X)$ defined a structure of $\mathscr{O}(X)$-module on the abelian group $\mathscr{F}(X)$. A morphism $\phi : \mathscr{F} \to \mathscr{G}$ between sheaves of $\mathscr{O}$-modules is a morphism of abelian presheaves $\phi : \mathscr{F} \to \mathscr{G}$ such that

$$\begin{tikzcd}
\mathscr{O} \times \mathscr{F} \arrow[r] \arrow[d, "1\times\phi"] & \mathscr{F} \arrow[d, "\phi"] \\
\mathscr{O} \times \mathscr{G} \arrow[r] & \mathscr{G}
\end{tikzcd}$$

The set of $\mathscr{O}$-module morphisms from $\mathscr{F}$ to $\mathscr{G}$ is denoted by $\operatorname{Hom}_{\mathscr{O}}(\mathscr{F}, \mathscr{G})$. Sheaves of $\mathscr{O}$-modules and its morphisms form the category $\mathbf{Mod}(\mathscr{O})$. We quote a important result in the context of our work.

Proposition 2.6. *Let* $(\mathbf{Sh}(\mathbf{C}), \mathscr{O})$ *be a ringed topos. The category* $\mathbf{Mod}(\mathscr{O})$ *is abelian. Moreover, it has enough injective objects.*

Proof. For the first assertion, see [81, Ch. 18, Lem. 4.1]. For the second, [81, Ch. 19, Lem. 5.1]. □

This result is due to Grothendieck [35]. Gabriel proves in [31] that there is a minimal injective object containing a given object, called its injective envelope.

Finally, given a (trivial) ringed site $(\mathbf{C}, \mathscr{O})$ and E in $\mathbf{PSh}(\mathbf{C})$, it is possible to define a presheaf of $\mathscr{O}$-modules, denoted by $\mathscr{O}[E]$, that associates to each $X \in \operatorname{Ob}\mathbf{C}$ the free $\mathscr{O}(X)$-module on generators $E(X)$. There is an adjunction

$$\operatorname{Mor}_{\mathbf{Mod}(\mathscr{O})}(\mathscr{O}[E], \mathscr{F}) \cong \operatorname{Mor}_{\mathbf{PSh}(\mathbf{C})}(E, \square\mathscr{F}) \tag{2.7}$$

where $\square$ denotes the forgetful functor.

2.2 Information cohomology

Let $\mathbf{S}$ be the poset of variables of an information structure $(\mathbf{S}, \mathscr{E})$. We view it as a site with the trivial topology (called *topologie grossière* or *chaotique* in [4, II.1.1.4]), such that every presheaf is a sheaf. For each $X \in \operatorname{Ob}\mathbf{S}$, set $\mathscr{S}_X := \{Y \in \operatorname{Ob}\mathbf{S} \mid X \to Y\}$, with the monoid structure given by the product of observables in $\mathbf{S}$: $(Z, Y) \mapsto ZY := Z \wedge Y$. Let $\mathscr{A}_X := \mathbb{R}[\mathscr{S}_X]$ be the corresponding monoid algebra. The contravariant functor $X \mapsto \mathscr{A}_X$ is a sheaf of rings; we denote it by $\mathscr{A}$. The pair $(\mathbf{S}, \mathscr{A})$ is a ringed site.

For a fixed object $\mathscr{G}$ of $\mathbf{Mod}(\mathscr{A})$, the covariant functor $\operatorname{Hom}(\mathscr{G}, -)$ is always additive and left exact. As $\mathbf{Mod}(\mathscr{A})$ has enough injective objects, it is possible to define the right derived functors associated to any left exact additive covariant functor. In the case of $\operatorname{Hom}(A, -)$, the associated right derived functors are called $\operatorname{Ext}^n(\mathscr{G}, -)$, for $n \geq 0$.

Let $\mathbb{R}_{\mathbf{S}}(X)$ be the $\mathscr{A}_X$-module defined by the trivial action of $\mathscr{A}_X$ on the abelian group $(\mathbb{R}, +)$ (for $s \in \mathscr{S}_X$ and $r \in \mathbb{R}$, take $s \cdot r = r$). The presheaf that associates to each $X \in \operatorname{Ob}\mathbf{S}$ the module $\mathbb{R}_{\mathbf{S}}(X)$, and to each arrow the identity map is denoted $\mathbb{R}_{\mathbf{S}}$.

Definition 2.7. The ***information cohomology*** associated to the poset of variables $\mathbf{S}$, with coefficients in the $\mathscr{A}$-module $\mathscr{F}$, is

$$H^\bullet(\mathbf{S}, \mathscr{F}) := \operatorname{Ext}^\bullet(\mathbb{R}_{\mathbf{S}}, \mathscr{F}). \tag{2.8}$$

The definition of information cohomology is formally analogous to that of group cohomology . In this case, one begins with a multiplicative group G and constructs the free abelian group $\mathbb{Z}[G]$, whose elements are finite sums $\sum m_g g$, with $g \in G$ and $m_g \in \mathbb{Z}$. The product of G induces a product between two such elements, and makes $\mathbb{Z}[G]$ a ring, called the integral group ring of G. The category of $\mathbb{Z}[G]$-modules is abelian and has enough injective objects. The cohomology groups of G with coefficients in a $\mathbb{Z}[G]$-module A are defined by

$$H^n(G, A) = \operatorname{Ext}^n(\mathbb{Z}, A), \tag{2.9}$$

where $\mathbb{Z}$ is the trivial module.

Finally, we make some observations concerning the computation of cohomology. Let $\mathbf{C}$ be an abelian category with enough injectives, like $\mathbf{Mod}(\mathscr{A})$, and suppose that we are interested in computing the groups $\{\operatorname{Ext}^n(A, B)\}_{n\geq 0}$ for certain fixed objects A and B. In addition, we assume that A has a *projective* resolution $0 \leftarrow A \leftarrow P_0 \leftarrow P_1 \leftarrow ....$ Then, Theorem 4.6.10 in [74] implies that, for all $n \geq 0$,

$$(R^n \operatorname{Hom}_{\mathbf{C}}(A, -))(B) \simeq (R^n \operatorname{Hom}_{\mathbf{C}}(-, B))(A). \tag{2.10}$$

We denote $(R^n \operatorname{Hom}_{\mathbf{C}}(-,B))(A)$ by $\underline{\operatorname{Ext}}^n(A,B)$. They are given by the formulas:

$$\underline{\operatorname{Ext}}^0(A,B) = \ker(\operatorname{Hom}(P_0,B) \to \operatorname{Hom}(P_1,B)), \tag{2.11}$$

$$\underline{\operatorname{Ext}}^i(A,B) = \frac{\ker(\operatorname{Hom}(P_i,B) \to \operatorname{Hom}(P_{i+1},B))}{\operatorname{im}(\operatorname{Hom}(P_{i-1},B) \to \operatorname{Hom}(P_i,B))}, \quad \text{for } i \geq 1. \tag{2.12}$$

2.3 Relative homological algebra

2.3.1 General results

In this subsection, we summarize some results from [60, Ch. IX]. The purpose is to find the analogous of a free resolution of modules, but in the general context of abelian categories. Capital Latin letters $A, B, C...$ denote objects and Greek letters $\alpha, \beta...$ morphisms.

A ***relative abelian category*** is a pair of abelian categories $\mathbf{A}$ and $\mathbf{M}$ and a convariant functor $\square : \mathbf{A} \to \mathbf{M}$ which is additive, exact and faithful (we write $\square(X) = X_\square$, for objects and morphisms). Additivity implies that $(A \oplus B)_\square = A_\square \oplus B_\square$; by exactness, $\square$ carries exact sequences into exact sequences; as $\square$ is faithful, $\alpha_\square = 0$ implies $\alpha = 0$, therefore $A_\square = 0$ entail $A = 0$.

Example 2.8. The simple example to have in mind are R-modules and S-modules, when S is a subring of R with the same unity (write $\iota : S \to R$ for the injection). In this case, every R-module A can be seen as an S-module ${}_\iota A$ by *restriction of scalars*: denote by $\bar{A}$ the underlying abelian group and by $\Lambda : R \to \operatorname{End}(\bar{A})$ the action of R over $\bar{A}$, then define the action $\Lambda' : S \to \operatorname{End}(\bar{A})$ by $\Lambda' = \Lambda \circ \iota$. Every R-module morphism $\alpha : A \to B$ is also a S-module morphism ${}_\iota\alpha : {}_\iota A \to {}_\iota B$. Therefore, $\square A := {}_\iota A$ and $\square\alpha := {}_\iota\alpha$ defines a functor from the category $\mathbf{A}$ of left R-modules to the category $\mathbf{M}$ of left S-modules that forgets part of the structure. This functor is exact, additive and faithful.

A short exact sequence $\chi \| \sigma$ in $\mathbf{A}$ is relatively split ($\square$-split) if $\chi_\square \| \sigma_\square$ splits in $\mathbf{M}$, this means that $\sigma_\square$ has a right inverse k or, equivalently, $\chi_\square$ has a left inverse t. We obtain a direct sum diagram in $\mathbf{M}$,

$$A \underset{\chi_\square}{\overset{t}{\leftrightarrows}} B \underset{k}{\overset{\sigma_\square}{\rightleftarrows}} C. \tag{2.13}$$

This class of short exact sequences is also called $\square$-allowable, or simply allowable (see [60, Ch. IX, Sec. 4]). A monomorphism χ is called allowable if $\chi \| \sigma$ is $\square$-split for some σ; this is the case if and only if $\chi \| (\operatorname{coker} \chi)$ is $\square$-split. Dually, an epimorphism is called allowable if $(\ker \sigma) \| \sigma$ is $\square$-split. Therefore, the class of allowable short exact sequences is determined by the allowable monomorphisms or the allowable epimorphisms.

The following conditions on a morphism α are equivalent:

1. $\operatorname{im} \alpha$ is an allowable monomorphism and $\operatorname{coim} \alpha$ is an allowable epimorphism;
2. $\ker \alpha$ is an allowable monomorphism and $\operatorname{coker} \alpha$ is an allowable epimorphism;

A morphism is called allowable when it satisfies any of these conditions (see [60, p. 264]).

A ***relative projective object*** P is any object of $\mathbf{A}$ such that, for every *allowable* epimorphism $\sigma : B \to C$, each morphism $\varepsilon : P \to C$ of $\mathbf{A}$ can be factored through σ as $\varepsilon = \sigma\varepsilon'$ for some $\varepsilon' : P \to A$.

In order to construct enough relative projectives, we consider the following definition. A ***resolvent pair*** is a relative abelian category $\square : \mathbf{A} \to \mathbf{M}$ together with a covariant functor $F : \mathbf{M} \to \mathbf{A}$ left adjoint to $\square$. This means that there exist a isomorphism φ,

$$\varphi : \mathrm{Hom}_{\mathbf{A}}(FM, A) \xrightarrow{\sim} \mathrm{Hom}_{\mathbf{M}}(M, \square A), \tag{2.14}$$

natural in both arguments. We can think of $\square$ as a forgetful functor, and F as the corresponding "free" functor.

Proposition 2.9. *Let $\square : \mathbf{A} \to \mathbf{M}$ be a relative abelian category. The following conditions are equivalent:*

1. *there exists a covariant functor $F : \mathbf{M} \to \mathbf{A}$ left adjoint to $\square$;*
2. *there exist a covariant functor $F : \mathbf{M} \to \mathbf{A}$, and a natural transformation $e : 1_{\mathbf{M}} \to \square F$ (where $1_{\mathbf{M}}$ is the identity functor), such that every $u : M \to A_\square$ in $\mathbf{M}$ has a factorization $u = \alpha_\square e_M$, with $\alpha : F(M) \to A$ unique.*

Proof. Suppose (1); taking $A = FM$ in (2.14), define e_M as $\varphi(1_{FM})$. Now, for arbitrary $A \in \mathrm{Ob}\,\mathbf{A}$, take $\alpha := \varphi^{-1}(u)$; the naturality of φ implies $u = \alpha_\square e_M$. The implication (2)$\Rightarrow$ (1) follows immediately taking $\varphi^{-1}(u) := \alpha$. More details can be found in [60, p. 266]. $\square$

Example 2.10 (continuation of 2.8). Take $F(M) = R \otimes_S M$ and $e_M = 1 \otimes m \in F(M)$. Given a map of S-modules $u : M \to A_\square$, define $\alpha : FM \to A$ by $\alpha(1 \otimes m) = u(m)$.

The following proposition exploits the properties of the allowable morphisms that we are studying ($\square$-split), and give us "free" objects, as suggested by the notation above.

A complex $\varepsilon : X \to A$ over A (in $\mathbf{A}$) is a sequence of $\mathbf{A}$-objects and $\mathbf{A}$-morphisms $\ldots X_n \to X_{n-1} \to \ldots \to X_1 \to X_0 \xrightarrow{\varepsilon} A \to 0$, such that the composite of any two succesive morphisms is zero. This complex is called:

1. a ***resolution*** of C, if the sequence is exact;
2. ***relatively free*** if each X_n has the form $F(M_n)$ for certain M_n in $\mathbf{M}$ (we write e_n instead of $e_{M_n} : M_n \to X_n$);
3. ***allowable*** if all its morphisms are allowable.

Each object C of $\mathbf{A}$ has a canonical relatively free resolution. Writing $\tilde{F}C$ for $F\square C$, and $\tilde{F}^n$ for its n-fold iteration, construct the objects

$$B_n(C) = \tilde{F}^{n+1}C, \quad n \in \mathbb{N}. \tag{2.15}$$

Define $\mathbf{M}$-morphisms $s_\bullet$ between the corresponding objects

$$\square C \xrightarrow{s_{-1}} \square B_0(C) \xrightarrow{s_0} \square B_1(C) \xrightarrow{s_1} \square B_2(C) \xrightarrow{s_2} \ldots \tag{2.16}$$

as $s_{-1} := e(\square C)$ and $s_n := e(\square B_n(C))$ (here e is the natural transformation in Proposition 2.9).

Proposition 2.11 ([60, p. 268]). *There are unique* **A**-*morphisms*

$$\varepsilon : B_0(C) \to C, \quad \partial_{n+1} : B_{n+1}(C) \to B_n(C) \quad \text{for } n \in \mathbb{N},$$

which make $B(C) := \{B_n(C)\}_n$ *a relatively free allowable resolution of* C *with* s *as contracting homotopy in* **M**. *This resolution, with its contracting homotopy, is a covariant functor of* C.

Proof. We simply quote here the construction of ε and ∂_n. They form the following diagram (solid arrows belong to **A**, and dashed arrows belong to **M**):

$$\begin{tikzcd} 0 & C \arrow[l] \arrow[r, dashed, shift right, "s_{-1}"'] & B_0(C) \arrow[l, shift right, "\varepsilon"'] \arrow[r, dashed, shift right, "s_0"'] & B_1(C) \arrow[l, shift right, "\partial_1"'] \arrow[r, dashed, shift right, "s_1"'] & B_2(C) \arrow[l, shift right, "\partial_2"'] \arrow[r, dashed, shift right, "s_2"'] & \cdots \arrow[l, shift right, "\partial"'] \end{tikzcd} \tag{2.17}$$

By Proposition 2.9, $1_{\square C}$ factors through a unique $\varepsilon : B_0(C) \to C$; the formula $1_{\square C} = \varepsilon_\square e_C$ shows that ε is allowable (note that ε is an epimorphism). Boundary operators are defined by recursion so that s will be a contracting homotopy. Given ε, the morphism $1_{\square B_0} - s_{-1}\varepsilon_\square$ factors uniquely as $\partial_{1\square}s_0$, for some $\partial_1 : B_1(C) \to B_0(C)$. Similarly, $1_{\square B_n} - s_{n-1}\partial_{n\square} : \square B_n(C) \to \square B_n(C)$ determines ∂_{n+1} given ∂_n, as the unique **A**-morphism such that $\partial_{n+1\square}s_n = 1_{\square B_n} - s_{n-1}\partial_{n\square}$.

$$\begin{tikzcd} \square B_{n+1} \arrow[r, "\partial_{1\square}"] & \square B_n \\ \square B_n \arrow[u, "s_n"] \arrow[ur, "1_{\square B_n}"'] & \end{tikzcd} \tag{2.18}$$

□

The resolution $B(C)$ is called the (unnormalized) bar resolution. A ***relative*** Ext ***bifunctor*** may be defined by

$$\mathrm{Ext}^n_\square(C, A) := H^n(\mathrm{Hom}_{\mathbf{A}}(B(C), A)). \tag{2.19}$$

A "relative" version of the comparison theorem (see [60, Ch. IX, Th. 6.2]) shows that one can use any other allowable and relatively projective resolution $\varepsilon : X \to C$ to compute $\mathrm{Ext}^n_\square$ as

$$\mathrm{Ext}^n_\square(C, A) \cong H^n(\mathrm{Hom}_{\mathbf{A}}(X, A)). \tag{2.20}$$

Note that $\mathrm{Hom}_{\mathbf{A}}(X, A)$ stands for *all* the **A**-morphisms, not just the allowable ones. It is clear that $\mathrm{Ext}^0_\square(C, A) \cong \mathrm{Ext}^0(C, A)$, but in general the groups $\mathrm{Ext}^n_\square(C, A)$ depend on $\square$.

2.3.2 Example: Presheaves of modules

We develop now the particular case relevant to our theory. Let **S** be a category, and $\mathscr{R}, \mathscr{T} : \mathbf{S}^{\mathrm{op}} \to \mathbf{Rings}$ presheaves, such that $\mathscr{T}_X$ is a subring of $\mathscr{R}_X$ with the same unity, for every $X \in \mathrm{Ob}\, T$. Take $\mathbf{A} = \mathbf{Mod}(\mathscr{R})$, the category of presheaves of $\mathscr{R}$-modules, and $\mathbf{M} = \mathbf{Mod}(\mathscr{T})$, the category of presheaves of $\mathscr{T}$-modules. A relative abelian category is obtained when $\square : \mathbf{A} \to \mathbf{M}$ is the forgetful functor over each X, as defined in Example 2.8. The functor $F : \mathbf{M} \to \mathbf{A}$ sends a presheaf $\mathscr{P}$ to the new presheaf $X \mapsto \mathscr{R}_X \otimes_{\mathscr{T}_X} \mathscr{P}_X$,[2] and each morphism of $\mathscr{T}$-presheaves (abbreviated to $\mathscr{T}$-morphism) $f : M \to N$ to the $\mathscr{R}$-morphism defined by

$$\forall X \in \mathrm{Ob}\,\mathbf{S}, \forall m \in M(X), \quad Ff(X)(1 \otimes m) = 1 \otimes f(m), \quad \text{for } X \in \mathbf{S}. \tag{2.21}$$

The natural transformation e mentioned in Proposition 2.9 corresponds to a collection of $\mathscr{T}$-morphisms $e_{\mathscr{P}} : \mathscr{P} \to \square F(\mathscr{P})$, one for each presheaf $\mathscr{P}$ of $\mathscr{T}$-modules; given X in $\mathbf{S}$, we define $e_{\mathscr{P}}(X)(m) = 1 \otimes m$ for each $m \in M(X)$. [3]

Fix now a presheaf $\mathscr{C}$ in $\mathbf{Mod}(\mathscr{R})$. We denote by X a generic element in $\mathrm{Ob}\,\mathbf{S}$. Then, $B_0\mathscr{C}(X) := F\square\mathscr{C}(X) = \mathscr{R}_X \otimes_{\mathscr{T}_X} (\square\mathscr{C}(X))$; this $\mathscr{R}_X$-module is formed by finite $\mathscr{R}_X$-linear combinations of tensors $1 \otimes c$, with $c \in \mathscr{C}$. Generally, an element of $B_n\mathscr{C}(X) = \mathscr{R}_X \otimes \square B_{n-1}\mathscr{C}(X)$, for $n \geq 1$, is a finite $\mathscr{R}_X$-linear combination of tensors $1 \otimes r_1 \otimes r_2 \otimes ... \otimes r_n \otimes c$. The ring $\mathscr{R}_X$ acts on $B_n\mathscr{C}(X)$ by multiplication on the first factor of the tensor product; to highlight this fact, people usually write $r[r_1|r_2|...|r_n|c]$ instead of $r \otimes r_1 \otimes r_2 \otimes ... \otimes r_n \otimes c$. This notation explains the name "bar resolution" adopted above. The definition of e implies that

$$s^X_{-1} : \square\mathscr{C}(X) \to \square B_0\mathscr{C}(X), \quad c \mapsto 1 \otimes c = [c], \tag{2.22}$$

and

$$s^X_n : \square B_n\mathscr{C}(X) \to \square B_{n+1}\mathscr{C}(X), \quad r[r_1|r_2|...|r_n|c] \mapsto [r|r_1|r_2|...|r_n|c] \quad \text{for } n \in \mathbb{N}. \tag{2.23}$$

These equalities determine $s_\bullet$, since these functions are $\mathscr{T}_X$-linear.

Now ε is the *unique* $\mathscr{R}$-morphism such that $1_{\square\mathscr{C}} = \varepsilon_\square e_{\mathscr{C}}$; this is clearly the case if $\varepsilon^X([c]) = c$. Similarly, ∂_1 is the unique $\mathscr{R}$-morphism from $B_1\mathscr{C}$ to $B_0\mathscr{C}$ that satisfies

$$\partial_{1\square}e_0 \equiv \partial_{1\square}s_0 = 1 - s^X_{-1}\varepsilon_\square \tag{2.24}$$

Since $B_1\mathscr{C}(X)$ is generated as a $\mathscr{R}_X$-module by the elements $[r|c]$, and $s^X_0(r[c]) = [r|c]$, the equation (2.24) defines ∂_1 completely. Just remark that $\varepsilon(r[c]) = r\varepsilon([c]) = r\varepsilon(1 \otimes c) = rc$ and $s_1(rc) = [rc]$. We conclude that

$$\partial_1([r|c]) = r[c] - [rc]. \tag{2.25}$$

It can be proved by recursion that (cf. [60, p. 281])

$$\partial[r_1|...|r_n|c] = r_1[r_2|...|r_n|c] + \sum_{k=1}^{n-1}(-1)^k[r_1|...|r_kr_{k+1}|...|r_n|c] + (-1)^n[r_1|...|r_{n-1}|r_nc]. \tag{2.26}$$

In virtue of Proposition 2.11, we obtain in this way a free allowable resolution of $\mathscr{C}$.

[2]This is a left $\mathscr{R}$-module with action defined by $r(r' \otimes p) = (rr') \otimes g$. For iterated tensor products, this definition is not canonical; for example, when considering $\mathscr{R}_X \otimes \mathscr{R}_X \otimes \mathscr{P}_X$, the element $(sr) \otimes r' \otimes g$ does not equal $r \otimes (sr') \otimes g$ (for $s \in \mathscr{T}$), unless $\mathscr{T}$ is in the center of $\mathscr{R}$. As in this work we only use commutative rings and algebras, these differences do not pose any problem.

[3]Of course, one has to prove that e is in fact a natural transformation and satisfies the properties required by Proposition 2.9. This proof is rather trivial but complicated to write, and we omit it.

2.4 Nonhomogeneous bar resolution

In this section, we introduce a projective resolution of the sheaf of $\mathscr{A}$-modules $\mathbb{R}_{\mathbf{S}}$: a long right exact sequence

$$0 \longleftarrow \mathbb{R}_{\mathbf{S}} \xleftarrow{\varepsilon} \mathscr{B}_0 \xleftarrow{\partial_1} \mathscr{B}_1 \xleftarrow{\partial_2} \mathscr{B}_2 \xleftarrow{\partial_3} \dots \tag{2.27}$$

that will allow us to compute the information cohomology.

Remember that $\mathscr{A}_X$ is the algebra over $\mathbb{R}$ generated by the monoid $\mathscr{S}_X$. Let $\mathscr{B}_n(X)$ be the free $\mathscr{A}_X$ module generated by the symbols $[X_1|...|X_n]$, where $\{X_1,...,X_n\} \subset \mathscr{S}_X$. Remark that $\mathscr{B}_0(X)$ is the free module on one generator $[\,]$.

We introduce now $\mathscr{A}_X$-module morphisms $\varepsilon_X : \mathscr{B}_0(X) \to \mathbb{R}_{\mathbf{S}}(X)$, from $\mathscr{B}_0(X)$ to the trivial $\mathscr{A}_X$-module $\mathbb{R}_{\mathbf{S}}(X)$, given by the equation $\varepsilon([\,]) = 1$, and boundary morphisms $\partial : \mathscr{B}_n(X) \to \mathscr{B}_{n-1}(X)$, given by

$$\partial([X_1|...|X_n]) = X_1[X_2|...|X_n] + \sum_{k=1}^{n-1}(-1)^k[X_1|...|X_kX_{k+1}|...|X_n] + (-1)^n[X_1|...|X_{n-1}]. \tag{2.28}$$

These morphisms are natural in X.

Proposition 2.12. *The complex* (2.27) *is a resolution of the sheaf* $\mathbb{R}_{\mathbf{S}}$.

Proof. The construction corresponds to the relatively projective bar resolution [60, Ch. IX], more especifically to the example developed at the end of Appendix 2.3, setting $\mathscr{R}$ and $\mathscr{T}$ there equal to $\mathscr{S}$ and $\mathbb{R}_{\mathbf{S}}$, respectively. The resolution $\mathscr{B}_\bullet$ introduced above is $B_\bullet\mathscr{C}$, for $\mathscr{C} = \mathbb{R}_{\mathbf{S}}$. The notation can be simplified, because $\mathscr{C}(X)$ is generated by 1 as an $\mathscr{A}_X$-module (and also as a vector space over $\mathbb{R}$). Therefore, $B_0\mathscr{C}$ is generated over $\mathscr{A}_X$ by the symbol [1], written simply as $[\,]$. In general, $B_n\mathscr{C}(X)$ is generated over $\mathscr{A}_X$ by the symbols $[X_1|...|X_n|1]$, or simply $[X_1|...|X_n]$ if we omit the 1. □

Thus far we have a resolution with relatively free objects, that in general need not be projective. However, the special properties of **S** allow us to improve the result.

Proposition 2.13. *For each $n \geq 0$, the sheaf $\mathscr{B}_n$ is a projective object in* **Mod**$(\mathscr{A})$.

Proof. Let $\mathscr{T}$ be the presheaf of sets defined by $\mathscr{T}(X) = \{\,[X_1|...|X_n] \mid X_i \in \mathscr{S}_X\,\}$, for $X \in \operatorname{Ob} S$. We have $\mathscr{B}_n = \mathscr{A}[\mathscr{T}]$.

Consider an epimorphism $\sigma : \mathscr{M} \twoheadrightarrow \mathscr{N}$ and a morphism $\varepsilon : \mathscr{A}[\mathscr{T}] \to \mathscr{N}$, both in **Mod**$(\mathscr{A})$. By the adjunction $\operatorname{Hom}_{\mathbf{Mod}(\mathscr{A})}(\mathscr{A}[\mathscr{T}],\mathscr{G}) \cong \operatorname{Hom}_{\mathbf{PSh(S)}}(\mathscr{F},\mathscr{G})$, ε determines a unique morphism $\tilde{\varepsilon} : \mathscr{T} \to \mathscr{N}$ in **PSh**(**S**). To show that $\mathscr{B}_n$ is projective, it suffices to show that there exists $\tilde{\varepsilon}' : \mathscr{F} \to \mathscr{M}$ such that $\tilde{\varepsilon} = \sigma\tilde{\varepsilon}'$, since by the adjunction this determines a morphism of $\mathscr{A}$-modules.

To define $\tilde{\varepsilon}'$, one has to determine the image of every symbol $[X_1|...|X_n]$, each time it appears in a set $\mathscr{T}(X)$. Remark that

$$[X_1|...|X_n] \in \mathscr{T}(X) \Leftrightarrow (\forall i)(X \to X_i) \Leftrightarrow X \to X_1 \cdots X_n = \Pi_{i=1}^n X_i$$

The last equivalence is true due to the definition of **S**. Therefore, the symbol $[X_1|...|X_n]$ just appears in the sets $\mathscr{T}(X)$ where $X \to \Pi_{i=1}^n X_i$; the full subcategory of **S** determined by these objects X has a terminal object, $\Pi_{i=1}^n X_i$ itself. To solve the lifting problem, it is enough to pick $m \in \sigma^{-1}_{\Pi_{i=1}^n X_i}(\tilde{\varepsilon}([X_1|...|X_n]))$, and define $\tilde{\varepsilon}'_{\Pi_{i=1}^n X_i}([X_1|...|X_n]) := m$. This choice gives, by funtoriality, a well defined value $\tilde{\varepsilon}_X([X_1|...|X_n]) = \mathscr{M}(\pi)(m)$ over each X such that $\pi : X \to \Pi_{i=1}^n X_i$ in $\mathbf{S}$. □

The existence of this projective resolution just depends on the definition of an abstract information structure (Definition 1.6). It appears in the computation of classical and quantum information cohomology: the difference between this cases lies in the coefficients.

2.5 Description of cocycles

We have built a projective resolution (2.27) of $\mathbb{R}_\mathbf{S}$ in $\mathbf{Mod}(\mathscr{A})$. For every $\mathscr{A}$-module $\mathscr{F}$, the information cohomology $H^\bullet(\mathbf{S},\mathscr{F})$ can be computed as $\underline{\mathrm{Ext}}^n(\mathbb{R}_\mathbf{S},\mathscr{F})$, defined in formulas (2.11) and (2.12) i.e. we deal with the cohomology of the differential complex $\{C^n(\mathbf{S},\mathscr{F}),\delta\}$, where

$$C^n(\mathbf{S},\mathscr{F}) := \mathrm{Hom}_\mathscr{A}(\mathscr{B}_n(\mathbb{R}_\mathbf{S}),\mathscr{F})\}_{n\geq 0}$$

and δ is given by (2.29) bellow. A morphism f in $C^n(\mathbf{S},\mathscr{F})$ is called n-cochain. More explicitly, an n-cochain f consists of a collection of morphisms $f_X \in \mathrm{Hom}_{\mathscr{A}_X}(\mathscr{B}_n(X),\mathscr{F}_X)$ that satisfies the following conditions:

1. f is a natural transformation (a functor of presheaves): given $\pi : X \to Y$, the diagram

$$\begin{array}{ccc} \mathscr{B}_n(Y) & \xrightarrow{f_Y} & \mathscr{F}_Y \\ \downarrow & & \downarrow{\scriptstyle \mathscr{F}(\pi)} \\ \mathscr{B}_n(X) & \xrightarrow{f_X} & \mathscr{F}_X \end{array}$$

   commutes. We refer to this property as ***(joint) locality***, for reasons that become evident in the following chapters.
2. f is compatible with the action of $\mathscr{A}$: for every $X \in \mathrm{Ob}\,\mathbf{S}$, the diagram

$$\begin{array}{ccc} \mathscr{A}_X \times \mathscr{B}_n(X) & \longrightarrow & \mathscr{B}_n(X) \\ \downarrow{\scriptstyle 1\times f_X} & & \downarrow{\scriptstyle f_X} \\ \mathscr{A}_X \times \mathscr{F}_X & \longrightarrow & \mathscr{F}_X \end{array}$$

   commutes. This means that f_X is ***equivariant***; in particular, $f_X(Y[Z]) = Y.f_X[Z]$ whenever $Y \in \mathscr{S}_X$.

Since $\mathscr{B}_n(X)$ is a free module, f_X is determined by the values on the generators $[X_1|...|X_n]$. Just to simplify notation, we write $f_X[X_1|...|X_n]$ instead of $f_X([X_1|...|X_n])$.

The coboundary of $f \in C^n(\mathbf{S},\mathscr{F})$ is the $(n+1)$-cochain $\delta f = f\partial : \mathscr{B}^{n+1} \to \mathscr{F}$. More explicitly,

$$\delta f[X_1|...|X_{n+1}] = X_1.f[X_2|...|X_{n+1}] + \sum_{k=1}^{n}(-1)^k f[X_1|...|X_k X_{k+1}|...|X_n] + (-1)^{n+1} f[X_1|...|X_n] \quad (2.29)$$

As customary, a cochain $f \in C^n(\mathbf{S}, \mathscr{F})$ is called an ***n-cocycle*** when $\delta f = 0$; the submodule of all n-cocycles is denoted by $Z^n(\mathbf{S}, \mathscr{F})$. The image under δ of C^{n-1} is another submodule of $C^n(\mathbf{S}, \mathscr{F})$, denoted $\delta C^{n-1}(\mathbf{S}, \mathscr{F})$; its elements are called ***n-coboundaries***. By definition, $\delta C^{-1}(\mathbf{S}, \mathscr{F}) = \langle 0 \rangle$, the trivial module. Since $\delta^2 = 0$, δC^{n-1} is a submodule of Z^n. With this notation, $H^n(\mathbf{S}, \mathscr{F}) = Z^n(\mathbf{S}, \mathscr{F})/\delta C^{n-1}(\mathbf{S}, \mathscr{F})$, for every $n \geq 0$.

Part II

Information cohomology of discrete random variables

Chapter 3

Probabilistic information cohomology

In this chapter, all information structures are supposed to be finite; they are denoted $(\mathbf{S}, E)$. We compute the information cohomology when the coefficients are functionals of probability laws; Shannon entropy and Tsallis α-entropies appear as 1-cocycles.

3.1 Functional module

Let $(\mathbf{S}, E)$ be an information structure, and $\mathscr{Q}$ an adapted probability functor. Information theory uses some functions defined on each set $\mathscr{Q}_X$ to measure the amount of information associated to the variable X. For example, given a variable X and a probability $P \in \mathscr{Q}_X$, the ***Gibbs-Shannon entropy***

$$S_1[X](P) := -\sum_{x \in E_X} P(x) \log P(x) \tag{3.1}$$

was proposed by Shannon [78] as a measure of uncertainty. Other example is given by the ***structural*** α***-entropy***, defined as

$$S_\alpha[X](P) = \frac{1}{1-\alpha}\left(\sum_{x \in E_X} P(x)^\alpha - 1\right), \tag{3.2}$$

for $\alpha > 0$, $\alpha \neq 1$.

In view of these considerations, let us introduce, for each $X \in \operatorname{Ob}\mathbf{S}$, the real vector space $\mathscr{F}(\mathscr{Q}_X)$ of measurable functions on $\mathscr{Q}_X$; we call it functional space. For each arrow $\pi : X \to Y$ in $\mathbf{S}$, there is a morphism $\pi^* : \mathscr{F}(\mathscr{Q}_Y) \to \mathscr{F}(\mathscr{Q}_X)$ defined by

$$\pi^* f(P_X) = f(\pi_* P_X).$$

Therefore, $\mathscr{F}(\mathscr{Q})$ is a contravariant functor from $\mathbf{S}$ to the category of real vector spaces.

Whenever $\mathscr{Q}$ is adapted to $\mathbf{S}$, the functional space $\mathscr{F}(\mathscr{Q}_X)$ admits an action of the monoid $\mathscr{S}_X$ (parameterized by $\alpha > 0$): for $Y \in \mathscr{S}_X$, and $f \in \mathscr{F}(\mathscr{Q}_X)$, the new function $Y.f$ is given by

$$(Y.f)(P_X) = \sum_{\substack{y\in E_Y\\ Y_*P_X(y)\neq 0}} (Y_*P_X(y))^\alpha f(P_X|_{Y=y}). \tag{3.3}$$

By Proposition 3.1, there is a morphism of monoids $\mathscr{S}_X \to \operatorname{End}(\mathscr{F}(\mathscr{Q}_X))$, given by Equation (3.3), that extends by linearity to a morphism of rings $\Lambda_\alpha(X) : \mathscr{A}_X \to \operatorname{End}(\mathscr{F}(\mathscr{Q}_X))$. This means that, for each $\alpha > 0$, $\mathscr{F}(\mathscr{Q}_X)$ has the structure of a $\mathscr{A}_X$-module, denoted $\mathscr{F}_\alpha(\mathscr{Q}_X)$. [1]

Proposition 3.1. *Given any $X \in \operatorname{Ob}\mathbf{S}$, observables Y and Z in $\mathscr{S}_X$, and $f \in \mathscr{F}(\mathscr{Q}_X)$:*

$$(ZY).f = Z.(Y.f).$$

Proof. The universal property of products gives the commutative diagram:

$$\begin{tikzcd} & X \arrow[dl, "\rho_Y"'] \arrow[d, "\langle \rho_Y,\rho_Z\rangle"] \arrow[dr, "\rho_Z"] & \\ Y & YZ \arrow[l, "\pi_Y"] \arrow[r, "\pi_Z"'] & Z \end{tikzcd}$$

For $P \in \mathscr{Q}_X$,

$$\begin{aligned} Z.(Y.f)(P) &= \sum_{\substack{z\in E_Z\\ Z_*P(z)\neq 0}} P(Z=z)^\alpha \sum_{\substack{y\in E_Y\\ Y_*P|_{Z=z}(y)\neq 0}} (P|_{Z=z}(Y=y))^\alpha f((P|_{Z=z})|_{Y=y}) \\ &= \sum_{\substack{z\in E_Z\\ Z_*P(z)\neq 0}} \sum_{\substack{y\in E_Y\\ Y_*P|_{Z=z}(y)\neq 0}} P(\{Y=y\}\cap\{Z=z\})^\alpha f((P|_{Z=z})|_{Y=y}) \end{aligned}$$

The equality $P(Z=z)P|_{Z=z}(Y=y) = P(\{Y=y\}\cap\{Z=z\})$ simply corresponds to the definition of conditional probabilities. The pairs (y,z) that appear in the sum are such that $P(\{Y=y\}\cap\{Z=z\}) \neq 0$, so $P(Y=y)$ and $P(Z=z)$ are different from zero; in this case, the equality

$$(P|_{Z=z})|_{Y=y}(B) = \frac{P|_{Z=z}(B\cap\{Y=y\})}{P|_{Z=z}(Y=y)} = \frac{P(B\cap\{Y=y\}\cap\{Z=z\})}{P(\{Y=y\}\cap\{Z=z\})} = P|_{Z=z,X=x}(B)$$

holds for every $B \subset X$. By (6), the nonempty sets $\{Y=y\}\cap\{Z=z\} = \rho_{Y_*}^{-1}(y) \cap \rho_{Z_*}^{-1}(z) \subset E_X$ are the preimage by $\langle\rho_Y,\rho_Z\rangle_*$ of a *unique* element $w(y,z) \in E(YZ)$; moreover, for every element $w \in E(YZ)$ we find such set. Remark that

$$P(\{Y=y\}\cap\{Z=z\}) = P(\rho_{Y_*}^{-1}(y)\cap\rho_{Z_*}^{-1}(z)) = P(\langle\rho_Y,\rho_Z\rangle_*^{-1} w(y,z)).$$

Therefore,

$$Z.(Y.f)(P) = \sum_{\substack{w(y,z)\in E_{YZ}\\ YZ_*P(w)\neq 0}} P(\{Y=x\}\cap\{Z=z\})^\alpha f(P|_{Z=z,X=x}) = (ZY).f(P).$$

□

[1] As $\mathscr{A}_X$ is a $\mathbb{R}$-algebra, it comes with an inclusion $f_X : \mathbb{R} \to \mathscr{A}_X$, $r \mapsto r\mathbf{1}_\mathbf{S}$. The composite $\Lambda_\alpha(X)\circ f_X$ gives an action of $\mathbb{R}$ over $\mathscr{F}(\mathscr{Q}_X)$, that coincides with the usual multiplication of functions by scalars.

The next proposition shows that this action is compatible with the morphisms between functional modules. Hence, the sheaf $\mathscr{F}_\alpha(\mathscr{Q})$ belongs to $\mathbf{Mod}(\mathscr{A})$, and can be used as coefficients in information cohomology.

Proposition 3.2. *Given $\pi_{YX} : X \to Y$ and $\pi_{ZY} : Y \to Z$, the action of Z makes the following diagram commute*

$$\begin{array}{ccc}
\mathscr{F}(\mathscr{Q}_Y) & \xrightarrow{Z} & \mathscr{F}(\mathscr{Q}_Y) \\
\Big\downarrow{\scriptstyle \pi^*_{YX}} & & \Big\downarrow{\scriptstyle \pi^*_{YX}} \\
\mathscr{F}(\mathscr{Q}_X) & \xrightarrow{Z} & \mathscr{F}(\mathscr{Q}_X)
\end{array}$$

Proof. We must prove that, for all $f_Y \in \mathscr{F}(\mathscr{Q}_Y), P \in \mathscr{Q}_X$, the equality $(Z.f_Y)(\pi_*^{YX}P) = Z.(f_Y \circ \pi_*^{YX})(P)$. On one hand,

$$(Z.f_Y)(\pi_*^{YX}P) = \sum_{\substack{z \in E_Z \\ \pi_*^{ZY}\pi_*^{YX}P(z) \neq 0}} \pi_*^{ZY}\pi_*^{YX}P(z) f_Y((\pi_*^{YX}P)|_{Z=z}), \tag{3.4}$$

and on the other,

$$Z.(f_Y \circ \pi_*^{YX})(P) = \sum_{\substack{z \in E_Z \\ \pi_*^{ZX}P(z) \neq 0}} \pi_*^{ZX}P(z) f_Y(\pi_*^{YX}(P|_{Z=z})). \tag{3.5}$$

The two expressions coincide since marginalizations are functorial, $\pi_*^{ZY}\pi_*^{YX} = \pi_*^{ZX}$, and commute with conditioning (cf. Section 1.3). □

3.2 Functoriality

In this and the following sections we study information cohomology with coeffients in $\mathscr{F}_\alpha(\mathscr{Q})$. This cohomology and its generalizations in Part IV are called ***probabilistic information cohomology***.

Let $\phi : (\mathbf{S}, E) \to (\mathbf{S}', E')$ be a morphism between finite information structures, and let $\mathscr{Q}$ be a probability functor on $\mathbf{S}$ and $\mathscr{Q}'$, a probability functor on $\mathbf{S}'$. Given a $X \in \operatorname{Ob}\mathbf{S}$ and a law $P \in \mathscr{Q}_X$, define a law $m_X(P)$ on $E'_{\phi(X)}$ by the equation

$$\forall x' \in E_{\phi(X)}, \quad (m_X(P))(x') = \sum_{x \in \phi_X^{\#\,-1}(x')} P(x). \tag{3.6}$$

We suppose that, for all $X \in \operatorname{Ob}\mathbf{S}$ and all $P \in \mathscr{Q}_X$, the law $m_X(P)$ belongs to $\mathscr{Q}'_{\phi(X)}$. Then $m_\bullet : \mathscr{Q} \to \mathscr{Q}' \circ \phi$ is a natural transformation. In fact, for every arrow $\pi : X \to Y$

and $y' \in E_{\phi(Y)}$,

$$\begin{aligned}
(\mathcal{Q}'(\pi)(m_X(P)))(y') &= \sum_{x'\in E'(\phi(\pi))^{-1}(y')} (m_X(P))(x') \\
&= \sum_{x'\in E'(\phi(\pi))^{-1}(y')} \sum_{x\in \phi_X^{\#\,-1}(x')} P(x) \\
&= \sum_{x\in (E(\phi(\pi))\circ\phi_X^{\#})^{-1}(y')} P(x) \\
&= \sum_{x\in (\phi_Y^{\#}\circ E(\pi))^{-1}(y')} P(x) \\
&= \sum_{y\in \phi_Y^{\#\,-1}(y')} \sum_{x\in E(\pi)^{-1}(y)} P(x) \\
&= m_Y(\mathcal{Q}(\pi)(P))
\end{aligned}$$

The forth equality comes from the naturality of $\phi^{\#}_{\bullet}$, as stated in Definition 1.7.

We construct now a functor between cohomology groups.

Proposition 3.3. *Let $\phi : (\mathbf{S}, E) \to (\mathbf{S}', E')$ be a morphism of information structures; let $\mathcal{Q}$ (resp. $\mathcal{Q}'$) be an adapted probability functor on $\mathbf{S}$ (resp. $\mathbf{S}'$). Suppose that*

1. *for all $X \in \operatorname{Ob}\mathbf{S}$, the map $\phi^{\#}_X$ is a bijection, and*
2. *for all $X \in \operatorname{Ob}\mathbf{S}$ and all $P \in \mathcal{Q}_X$, the law $m_X(P)$ belongs to $\mathcal{Q}'_{\phi(X)}$.*

Then, there exist a cochain map

$$\phi^*_{\bullet} : (C^{\bullet}(\mathcal{F}_\alpha(\mathcal{Q}'), \delta) \to (C^{\bullet}(\mathcal{F}_\alpha(\mathcal{Q}), \delta), \tag{3.7}$$

given by the formula

$$(\phi^*_n f)_Y[X_1|...|X_n](P) := f_{\phi(Y)}[\phi(X_1)|...|\phi(X_n)](m_Y(P)). \tag{3.8}$$

The chain map induces a morphism of graded vector spaces in cohomology

$$\phi^*_{\bullet} : H^{\bullet}(\mathbf{S}', \mathcal{F}_\alpha(\mathcal{Q}')) \to H^{\bullet}(\mathbf{S}, \mathcal{F}_\alpha(\mathcal{Q})). \tag{3.9}$$

Proof. First, we prove that $\phi^* f$ is jointly local. For f is jointly local,

$$f_{\phi(Y)}[\phi(X_1)|...|\phi(X_n)](m_Y(P))$$

only depends on

$$(\phi(X_1)\cdots\phi(X_n))_* m_Y(P) = (\phi(X_1\cdots X_n))_* m_Y(P).$$

Let $\pi : Y \to X_1 \cdots X_n$ be the corresponding refinement. Since $m_\bullet$ is a natural transformation, $m_{X_1\cdots X_n} \circ \mathcal{Q}(\pi) = \mathcal{Q}'(\phi(\pi)) \circ m_Y$; this means that

$$(\phi(X_1\cdots X_n))_* m_Y(P) = m_{X_1\cdots X_n}(\mathcal{Q}(\pi)(P)) = m_{X_1\cdots X_n}((X_1\cdots X_n)_* P). \tag{3.10}$$

We conclude that $\phi^* f$ depends only on $(X_1\cdots X_n)_* P$ and its therefore a cocycle.

We show now that ϕ^* commutes with δ. For simplicity, we write the formulas for $n = 2$; the argument works in general. Note that

$$(\phi^*(\delta f))_Y[X_1|X_2] = \delta f_{\phi(Y)}[\phi(X_1)|\phi(X_2)] \tag{3.11}$$
$$= \phi(X_1).f_{\phi(Y)}[\phi(X_2)] - f_{\phi(Y)}[\phi(X_1)\phi(X_2)] + f_{\phi(Y)}[\phi(X_1)]. \tag{3.12}$$

By definition, $f_{\phi(Y)}[\phi(X_1)] = \phi^* f_Y[X_1]$ and similarly $f_{\phi(Y)}[\phi(X_1)\phi(X_2)] = \phi^* f_Y[X_1X_2]$, since $\phi(X_1)\phi(X_2) = \phi(X_1X_2)$. The remaining term $(\phi(X_1).f_\phi(Y))[\phi(X_2)](m_Y(P))$ equals

$$\sum_{x_1'\in E'(\phi(X_1))} \{(\phi(\pi_1)_* \circ m_Y)(P)\}(x_1')f_{\phi(Y)}[\phi(X_2)]((m_Y(P))|_{\phi(X_1)=x_1'}) \tag{3.13}$$

where $\pi_1 : Y \to X_1$. We write $\phi(\pi_1)_*$ instead of $\mathscr{Q}'(\phi(\pi_1))$ and π_{1*} instead of $\mathscr{Q}(\pi_1)$. Set $x_1 = \phi^{\#\,-1}_{X_1}(x_1')$. The naturality of $m_\bullet$ implies that

$$\{(\phi(\pi_1)_* \circ m_Y)(P)\}(x_1') = \{(m_{X_1} \circ \pi_*)(P)\}(x_1') = \sum_{x\in\phi^{\#\,-1}_{X_1}(x_1')} \pi_*P(x) = \pi_*P(x_1). \tag{3.14}$$

Finally, for every $y' \in \phi(Y)$,

$$\begin{aligned}(m_Y(P))|_{\phi(X_1)=x_1'}(y') &= \frac{m_Y(P)(\{y'\}\cap\{\phi(X_1)=x_1'\})}{m_Y(P)(\phi(X_1)=x_1')}\\ &= \frac{\sum_{z\in\phi^{\#\,-1}_Y(\{y'\})\cap\phi^{\#\,-1}_Y(\{\phi(X_1)=x_1'\})} P(z)}{\sum_{z\in\phi^{\#\,-1}_Y(\{\phi(X_1)=x_1'\})} P(z)}\\ &= \frac{\sum_{z\in\phi^{\#\,-1}_Y(\{y'\})} P(\{z\}\cap\{X_1=x_1\})}{P(X=x_1)}\\ &= m_Y(P|_{X=x_1}).\end{aligned}$$

The first equality comes from the definition of conditioning and the second from that of m_Y. The third is a consequence of

$$\begin{aligned}\phi^{\#\,-1}_Y(\{\phi(X_1)=x_1'\}) &= \{\, z\in Y \mid \phi(\pi)\circ\phi^\#_Y(z) = x_1' \,\}\\ &= \{\, z\in Y \mid \phi^\#_{X_1}\circ\pi(z) = x_1' \,\} = \{\, z\in Y \mid \pi(z) = x_1 \,\}.\end{aligned}$$

that depends on $\phi^\#_{X_1}$ being a bijection. Therefore,

$$f_{\phi(Y)}[\phi(X_2)]((m_Y(P))|_{\phi(X_1)=x_1'}) = f_{\phi(Y)}[\phi(X_2)](m_Y(P|_{X=x_1})) = (\phi^*f)_Y[X_2](P|_{X_1=x_1}),$$

hence $(\phi(X_1).f_{\phi(Y)})[\phi(X_2)](m_Y(P)) = (X_1.\phi^*f_Y)(P)$. In consequence, δ commutes with ϕ^*, as we wanted to prove. □

Corollary 3.4. *If $\phi : \mathbf{S} \to \mathbf{S}'$ is an isomorphism of information structures, and $\mathscr{Q}'(\phi(X)) = m_X(\mathscr{Q}_X)$ for every $X \in \operatorname{Ob}\mathbf{S}$, then $\phi^* : H^\bullet(\mathbf{S}', \mathscr{F}_\alpha(\mathscr{Q}')) \to H^\bullet(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q}))$ is an isomorphism too.*

Proof. Let $\psi : \mathbf{S}' \to \mathbf{S}$ be the inverse of ϕ. It is clear from the definitions that ϕ_0 and ψ_0. Moreover, for every $X \in \operatorname{Ob}\mathbf{S}$, we have $\psi^{\#}_{\phi(X)} \circ \phi^{\#}_X = \mathrm{id}_X$, and similarly for $\phi(X)$; this implies that $\phi^{\#}_X$ and $\psi^{\#}_{\phi(X)}$ are bijections. Proposition 3.3 ensures the existence of $\phi^* : H^\bullet(\mathbf{S}', \mathscr{F}_\alpha(\mathscr{Q}')) \to H^\bullet(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q}))$ and $\psi^* : H^\bullet(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q})) \to H^\bullet(\mathbf{S}', \mathscr{F}_\alpha(\mathscr{Q}'))$; and $\psi^* \circ \phi^* = \mathrm{id}$, $\phi^* \circ \psi^* = \mathrm{id}$ by formula (3.8). □

We also recover from Proposition 3.3 two functorial properties for concrete information structures stated in [10].

Proposition 3.5. *Consider concrete information structures* $\mathbf{S}$, $\mathbf{S}'$ *associated to a measurable spaces* $(\Omega, \mathfrak{B})$ *and* $(\Omega', \mathfrak{B}')$*, respectively. Let* $\mathscr{Q}$ *(resp.* $\mathscr{Q}'$*) be a probability functor defined on* $\mathbf{S}$ *(resp.* $\mathbf{S}'$*). Let* $\sigma : (\Omega, \mathfrak{B}) \to (\Omega', \mathfrak{B}')$ *be a* surjective *measurable function, such that*

1. *for all* $X \in \operatorname{Ob}\mathbf{S}$*, there exists* $\phi(X) \in \operatorname{Ob}\mathbf{S}'$ *such that* σ *descends to a bijection* $\sigma_X : \Omega/X \xrightarrow{\sim} \Omega'/\phi(X)$*;*[2]
2. *for every* $X \in \operatorname{Ob}\mathbf{S}$ *and* $P \in \mathscr{Q}_X$*, the marginalization* $\sigma_{X*}P$ *is in* $\mathscr{Q}'_{\phi(X)}$*;*

Then, there exists a natural morphism of graded vector spaces

$$\sigma^* : H^\bullet(\mathbf{S}', \mathscr{F}_\alpha(\mathscr{Q}')) \to H^\bullet(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q})), \tag{3.15}$$

defined at the level of cochains by

$$(\sigma^* f)_Y[X_1|...|X_n](P) = f_{\phi(Y)}[\phi(X_1)|...|\phi(X_n)](\phi_* P), \tag{3.16}$$

where $X_j = X'_j \circ \phi$*, for each index* i*.*

Proof. The correspondence $X \mapsto \phi(X)$ defines a functor from $\phi_0 : \mathbf{S} \to \mathbf{S}'$; in fact, if $\pi : X \to Y$, there is a corresponding surjection $\pi_* : \Omega/X \to \Omega/Y$ and $\sigma_Y \circ \pi_* \circ \sigma_X^{-1} : \Omega'/\phi(X) \to \Omega'/\phi(Y)$ is also a surjection, that gives a morphism $\phi(X) \to \phi(Y)$ in $\mathbf{S}'$. We take as $\phi^{\#}_X : X \to \phi(X)$ the bijection of partitions induced by σ_X (recall that, for concrete structures, the functor of values is the identity). The condition (2) implies that $m_\bullet : \mathscr{Q} \to \mathscr{Q}' \circ \phi$ is a natural transformation. Proposition 3.3 entails the existence of $\phi^* =: \sigma^*$. □

Proposition 3.6. *Consider concrete information structures* $\mathbf{S}$, $\mathbf{S}'$ *associated to a measurable spaces* $(\Omega, \mathfrak{B})$ *and* $(\Omega', \mathfrak{B}')$*, respectively. Let* $\mathscr{Q}$ *(resp.* $\mathscr{Q}'$*) be a probability functor defined on* $\mathbf{S}$ *(resp.* $\mathbf{S}'$*). Let* $\eta : (\Omega, \mathfrak{B}) \to (\Omega', \mathfrak{B}')$ *be a measurable function, such that*

1. *for all* $X' \in \operatorname{Ob}\mathbf{S}'$*, there exists* $\phi(X') \in \operatorname{Ob}\mathbf{S}$ *such that* η *descends to a bijection* $\eta_X : \Omega/\phi(X') \to \Omega/X'$*;*
2. *for all* $X' \in \operatorname{Ob}\mathbf{S}'$ *and* $P' \in \mathscr{Q}'_{X'}$*, there exists* $P \in \mathscr{Q}_{\phi(X')}$ *with* $P' = \eta_{X*}P$*.*

Then, there exists a natural morphism of graded vector spaces

$$\eta_* : H^m(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q})) \to H^m(\mathbf{S}', \mathscr{F}_\alpha(\mathscr{Q}')), \tag{3.17}$$

defined at the level of cochains by

$$(\sigma_* f)_Y[X'_1|...|X'_n](P') = f_{\phi(Y)}[\phi(X'_1)|...|\phi(X'_n)](P), \tag{3.18}$$

where $P' = \eta_{Y*}P$*.*

Proof. The correspondence $X' \mapsto \phi(X')$ defines a functor $\phi_0 : \mathbf{S}' \to \mathbf{S}$ (see the proof of Proposition 3.5). We take as $\phi^{\#}_X : X' \to \phi(X')$ the bijection induced by η_X^{-1}. Assumption (2) implies that $m_\bullet : \mathscr{Q}' \to \mathscr{Q} \circ \phi$ is a natural transformation and we can use Proposition 3.3 to conclude. □

[2] Every partition X defines an equivalence relation and Ω/X denotes the corresponding quotient.

3.3 Determination of H^0

Each 0-cochain $f[\,] \equiv f$ corresponds to a collection of functions $f_X(P_X) \in \mathscr{F}_\alpha(\mathscr{Q}_X)$, for each $X \in \operatorname{Ob}\mathbf{S}$, that satisfy $f_Y(Y_*P_X) = f_X(P_X)$ for any arrow $X \to Y$ in $\mathbf{S}$. As we assume that $\mathbf{1} \in \operatorname{Ob}\mathbf{S}$, this means that f is constant. Given an arrow $X \to Y$, and a 0-cochain f such that $f_X(P) = K$,

$$(\delta f)_X[Y](P) = Y.f_X(P) - f_X(P) = \sum_{y \in E_Y} P(Y=y)^\alpha f(P|_{Y=y}) - f(P)$$

$$= K\left(\sum_{y \in E_Y} P(Y=y)^\alpha - 1\right) = 0.$$

This means that $Z^0(\mathbf{S}, \mathscr{F}_1(\mathscr{Q})) = C^0(\mathbf{S}, \mathscr{F}_1(\mathscr{Q})) \cong \mathbb{R}$ and $Z^0(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q})) = \langle 0 \rangle$ when $\alpha \neq 1$ (as long as some $\mathscr{Q}_Y$ contains a nonatomic probability). Equivalently, $H^0(\mathbf{S}, \mathscr{F}_1(\mathscr{Q})) \cong \mathbb{R}$, and $H^0(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q})) \cong \langle 0 \rangle$ when $\alpha \neq 1$.

3.4 Local structure of 1-cocycles

Now we turn to $C^1(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q}))$. The 1-cochains are families $\{f_X[Y] \mid X \in \operatorname{Ob}\mathbf{S}\}$ such that for all $Z \to X \to Y$, the equality $f_X[Y](X_*P_Z) = f_Z[Y](P_Z)$ holds. This means that it is sufficient to know $f_Y[Y](Y_*P)$ to recover $f_X[Y](P)$, for any $X \to Y$; in this sense, we usually omit the subindex and just write $f[Y]$. The computation above implies that $\delta C^0(\mathbf{S}, \mathscr{F}_1(\mathscr{Q})) = \langle 0 \rangle$. On the other hand, $\delta C^0(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q})) \cong \mathbb{R}$ when $\alpha \neq 1$, and 1-coboundaries are multiples of the section of $\mathscr{F}_\alpha(\mathscr{Q})$ given by $X \mapsto S_\alpha[X]$; we write $\delta C^0(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q})) \cong \mathbb{R} \cdot S_\alpha$.

By equation (2.29) and commutativity of the product, every 1-cycle ($\delta f = 0$) must satisfy the following symmetric relation:

$$f[XY] = f[Y] + Y.f[X] = f[X] + X.f[Y]. \tag{3.19}$$

Proposition 3.7. *Let f be a* 1*-cocycle. Then*

1. $f[\mathbf{1}] \equiv 0$
2. *For every $X \in \operatorname{Ob}\mathbf{S}$ and $x \in E_X$, the equality $f[X](\delta_x) = 0$ holds.*

Proof. Statement (1) is a particular case of (2); we prove the later. From $f[XX] = f[X] + X.f[X]$, we conclude that $X.f[X] = \sum_{x \in E_X | P(x) \neq 0} P(x)^\alpha f[X](P|_{X=x}) = 0$. For $P = \delta_x$, one obtains $f[X](\delta_x) = 0$. □

Example 3.8. We compute now $H^1(\mathbf{S}, \mathscr{F}_\alpha(\Pi))$, taking $\mathbf{S}$ equal to $\mathbf{0} \to \mathbf{1}$, and $E(\mathbf{0}) = \{a, b\}$. Proposition Proposition 3.7 implies that $f[\mathbf{0}](1,0) = f[\mathbf{0}](0,1) = 0$, as a consequence of $f[\mathbf{0}] = f[\mathbf{0}] + \mathbf{0}.f[\mathbf{0}]$. All the other relations derived from the cocycle condition (3.19) become tautological. Therefore, 1-cocycles are in correspondence with measurable functions f on arguments (p_a, p_b) such that $f(1,0) = f(0,1) = 0$. We conclude that $H^1(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q}))$ has infinite dimension. For a more general condition under which $\dim H^1$ diverges, see Proposition 3.16.

The functions $S_\alpha[X]$ introduced in (3.1) and (3.2) are local, since they only depend on X_*P. The following proposition establishes that they correspond to a 1-cocycles.

Proposition 3.9. *Let $(\mathbf{S}, E)$ be an information structure, $\mathscr{Q}$ an adapted probability functor, X an element of* $\operatorname{Ob}\mathbf{S}$*, and $Y, Z \in \operatorname{Ob}\mathbf{S}$ two variables refined by X. Then, for all $\alpha > 0$, the section S_α of $\mathscr{F}_\alpha(\mathscr{Q})$ satisfy the relation*

$$(S_\alpha)_X[YZ] = (S_\alpha)_X[Y] + (Y.S_\alpha)_X[Z]. \tag{3.20}$$

This means that S_α belongs to $Z^1(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q}))$.

Proof. Let P be a probability in $\mathscr{Q}_X$. We further simplify the notation, writing $P(y)$ instead of $P(Y = y) = Y_*P(y)$, and $P(z|y)$ in place of $P(Z = z|Y = y)$. We label the points in $E(YZ)$ by their image under the injection $\iota : E(YZ) \to E(Y) \times E(Z)$, writing $w(y, z) \in E(YZ)$.

1. Case $\alpha = 1$: by definition

$$-S_1[YZ](P) = \sum_{w(y,z)\in E(YZ)} P(y, z) \log P(y, z)$$

   and in fact we can extend this to a sum over the whole set $E(X) \times E(Y)$, setting $P(y, z) = 0$ whenever $(y, z) \notin \operatorname{im}\iota$ (recall the convention $0 \log 0 = 0$). We rewrite the previous expression using the conditional probabilities

$$\begin{aligned}-S_1[YZ](P) &= \sum_{z\in E_Z}\sum_{y\in E_Y} P(z|y)P(y)(\log P(y) + \log P(z|y))\\ &= \sum_{y\in E_Y} P(y)\log P(y) \sum_{z\in E_Z} P(z|y) + \sum_{y\in E_Y} P(y) \sum_{z\in E_Z} P(z|y)\log P(z|y).\end{aligned}$$

   This gives the result, because $\sum_{z\in E_Z} P(z|y) = 1$, and $\sum_{z\in E_Z} P(z|y)\log P(z|y) = S_1[Z](P|_{Y=y})$. Cf. [49].

2. Case $\alpha \neq 1$: The result is a consequence of $\delta^2 = 0$, but can be proved by a direct computation.

$$\begin{aligned}(1-\alpha)(S[Y] + X.S[Y]) &= \left(\sum_{y\in E_Y} P(y)^\alpha - 1\right) + \sum_{y\in E_Y} P(y)^\alpha \left(\sum_{z\in E_Z} P(z|y)^\alpha - 1\right)\\ &= \sum_{y\in Y}\sum_{z\in Z} P(z|y)^\alpha P(y)^\alpha - 1\\ &= (1-\alpha)S[XY].\end{aligned}$$

   The last equality comes from $P(z|y)P(y) = P(z, y)$, and the fact that we can restrict the sum to $E(YZ)$, neglecting terms that vanish.

□

We shall see that any nontrivial 1-cocycle of type α is locally a multiple of S_α; we still have to formalize this notion of locality. Proposition 3.10 present the solution to a functional equation that comes from the cocycle condition. Then, Proposition 3.13 determine the local form of a cocycle. Finally, Theorem 3.14 determine H^1 under appropriate nondegeneracy hypotheses on the information structure $\mathbf{S}$ and the probability functor $\mathscr{Q}$.

For convenience, we introduce the functions

$$s_1(p) := -p\log p - (1-p)\log(1-p); \tag{3.21}$$

$$s_\alpha(p) := \frac{1}{1-\alpha}(p^\alpha + (1-p)^\alpha - 1) \qquad (\text{for } \alpha \neq 1), \tag{3.22}$$

both defined for $p \in [0,1]$.

Theorem 3.10 (Generalized FEITH). *Let $f_1, f_2 : \Delta^2 \to \mathbb{R}$ be two unknown measurable functions satisfying*

1. *$f_i(0,1) = f_i(1,0) = 0$ for $i = 1, 2$.*
2. *for all $(p_0, p_1, p_2) \in \Delta^2$,*

$$(1-p_2)^\alpha f_1\left(\frac{p_0}{1-p_2}, \frac{p_1}{1-p_2}\right) - f_1(1-p_1, p_1) \\ = (1-p_1)^\alpha f_2\left(\frac{p_0}{1-p_1}, \frac{p_2}{1-p_1}\right) - f_2(1-p_2, p_2). \tag{3.23}$$

Then, $f_1 = f_2$ and there exists $\lambda \in \mathbb{R}$ such that $f_1(p) = \lambda s_\alpha(p)$.

Proof. The restriction to $p_0 = 0$ (with $p_1 = x, p_2 = 1-x$) implies $f_2(x, 1-x) = f_1(1-x, x)$. We eliminate f_2 in (3.23) and set $u(x) := f_1(x, 1-x)$. Setting $p_1 = x$, $p_2 = y$ and $p_0 = 1-x-y$, we obtain the functional equation

$$u(1-x) + (1-x)^\alpha u\left(\frac{y}{1-x}\right) = u(y) + (1-y)^\alpha u\left(\frac{1-x-y}{1-y}\right). \tag{3.24}$$

This functional equation is related to the so-called "fundamental equation of information theory", which first appeared in the work of Tverberg [89]. In [47], Kannappan and Ng show that every measurable solution of (3.24) with $\alpha = 1$ has the form $u(x) = \lambda s_1(x)$, with $\lambda \in \mathbb{R}$. Analogously, we show in Chapter 5 that the general solution in the case $\alpha \neq 1$ is $u(x) = \lambda s_\alpha(x)$, with $\lambda \in \mathbb{R}$; this is directly connected to a generalization of the fundamental equation introduced by Daróczy in [25]. □

Example 3.11. Let **S** be the free category generated by

$$\begin{tikzcd} & \mathbf{1} & \\ X_1 \arrow[ur] & & X_2 \arrow[ul] \\ & X_1X_2 \arrow[ul] \arrow[ur] & \end{tikzcd} \tag{3.25}$$

and E be the functor defined at the level of objects by $E(X_1) = \{x_{\{1\}}, x_{\{0,2\}}\}$, $E(X_2) = \{x_{\{2\}}, x_{\{0,1\}}\}$, and $E(X_1X_2) = \{x_{\{1\}}, x_{\{2\}}, x_{\{3\}}\}$; for each arrow $\pi : X \to Y$, the map $\pi_* : E(X) \to E(Y)$ sends $x_I \to x_J$ iff $I \subset J$. The pair $(\mathbf{S}, E)$ is an information structure (in fact, it comes from a concrete one). Consider $f \in Z^1(\mathscr{F}_\alpha(\Pi))$: the cocyle condition means that, as functions on $\Pi(X_1X_2)$,

$$f[X_1X_2] = X_1.f[X_2] + f[X_1] \quad \text{and} \quad f[X_1X_2] = X_2.f[X_1] + f[X_2]. \tag{3.26}$$

Write $f[X_1] = f_1$ and $f[X_2] = f_2$. Clearly, the determination of f_1 and f_2 such that $X_1.f_2 + f_1 = X_2.f_1 + f_2$ fix f completely. In terms of a probability (p_0, p_1, p_2) in $\Pi(X_1X_2)$ this equation is exactly (3.23). We conclude that every cocycle is a multiple of the corresponding α-entropy: there exists a unique constant $\lambda \in \mathbb{R}$ such that $f[Z](P) = \lambda S_\alpha[Z](P)$, for every variable $Z \in \operatorname{Ob}\mathbf{S}$ and every probability law $P \in \Pi(X_1X_2)$. This establishes that $Z^1(\mathbf{S}, \mathscr{F}_\alpha(\Pi)) \cong \mathbb{R}$. Hence $H^1(\mathbf{S}, \mathscr{F}_1(\Pi)) \cong \mathbb{R}$, and $H^1(\mathbf{S}, \mathscr{F}_\alpha(\Pi)) \cong \langle 0\rangle$. The hypotheses are minimal: on one hand, if we remove X_1 or X_2, Proposition 3.16 shows that $\dim H^1 = \infty$; on the other, if $\mathscr{Q}_{X_1X_2}$ does not contain the interior of Δ^2, the cocycle equations accept an infinite number of solutions, because one cannot obtain (3.23).

We want to extend this result to more general triples $(\mathbf{S}, E, \mathscr{Q})$. The strategy is to reduce the problem to the equations in Proposition 3.10, as we did in the previous example: first, considering a product XY with good properties, and then deducing some functional equations for $f[X]$ and $f[Y]$. This is carried out in Proposition 3.13, that you can read directly if you suppose that $E_{XY} = E_X \times E_Y$ and $\mathscr{Q} = \Pi$. We introduce now the notion of *nondegenerate product*, that states precisely what is needed for the proof. The condition should be thought as some kind of "transversality" between X and Y. If many events $\{X = x, Y = y\}$ are impossible, then the product XY degenerates.

Remarks on notation: to avoid confusion with subindexes, we denote the (i, j) component of a matrix M by $M[i, j]$. Recall that each set $\mathscr{Q}_X$ is a simplicial subcomplex of $\Pi(X)$. For each $S \subset E_X$, we denote $[S]$ the face of $\Pi(X)$ generated by the Dirac laws $\{\, \delta_x \mid x \in S \,\}$, which implies that $\mathscr{Q}_X \cap [S]$ is the set of probabilities in $\mathscr{Q}_X$ whose support is contained in S.

Definition 3.12. Given two partitions X and Y, such that $|E_X| = k$ and $|E_Y| = l$, we call its product XY ***nondegenerate*** if there exist enumerations $\{x_1, ..., x_k\}$ of E_X and $\{y_1, ..., y_l\}$ of E_Y, and a North-East (NE) lattice path[3] $(\gamma_i)_{i=1}^m$ on $\mathbb{Z}^2$ going from $(1, 1)$ to (k, l) such that

1. For each $\gamma_i = (a, b)$, the simplicial complex
$$\mathscr{Q}_{XY} \cap [\iota^{-1}\{\, (x_i, y_j) \mid a \le i \le a + 1 \text{ and } b \le j \le b + 1 \,\}]$$
has at least one 2-dimensional cell. Here ι denotes the injection $E_{XY} \hookrightarrow E_X \times E_Y$.
2. If $\gamma_i = (a, b)$ and $\gamma_{i+1} - \gamma_i = (1, 0)$, we ask that for every law p in $\mathscr{Q}_X \cap [\{\, x_i \mid a \le i \le k \,\}]$ there exists a law $\tilde{p}$ in the intersection of $\mathscr{Q}_{XY}$ and
$$[\iota^{-1}(\{(x_a, y_{b+1})\} \cup \{\, (x_i, y_b) \mid a{+}1 \le i \le k \,\})] \cup [\iota^{-1}(\{(x_a, y_b)\} \cup \{\, (x_i, y_{b+1}) \mid a{+}1 \le i \le k \,\})]$$
such that $p = X_*\tilde{p}$.
Analogously, if $\gamma_{i+1} - \gamma_i = (0, 1)$, we ask that every law p in $\mathscr{Q}_Y \cap [\{\, y_i \mid b \le i \le l \,\}]$ there exists a law $\tilde{p}$ in the intersection of $\mathscr{Q}_{XY}$ with
$$[\iota^{-1}(\{(x_{a+1}, y_b)\} \cup \{\, (x_a, y_j) \mid b{+}1 \le j \le l \,\})] \cup [\iota^{-1}(\{(x_a, y_b)\} \cup \{\, (x_{a+1}, y_j) \mid b{+}1 \le j \le k \,\})]$$
such that $p = Y_*\tilde{p}$.

[3] A North-East (NE) lattice path on $\mathbb{Z}^2$ is a sequence of points $(\gamma_i)_{i=1}^m \subset \mathbb{Z}^2$ such that $\gamma_{i+1} - \gamma_i \in \{(1, 0), (0, 1)\}$ for every $i \in \{1, ..., m - 1\}$.

Remark that the product of a variable X with itself is always degenerate, because it only accepts nontrivial probabilities for pairs $(x,x) \in E_X^2$.

Proposition 3.13. *Let* $(\mathbf{S}, E)$ *be a finite information structure,* $\mathcal{Q}$ *an adapted probability functor, and* X, Y *two different variables in* $\operatorname{Ob}\mathbf{S}$ *such that* $XY \in \operatorname{Ob}\mathbf{S}$*. Let* f *be a* 1*-cocycle of type* α*, i.e. an element of* $Z^1(\mathbf{S}, \mathscr{F}_\alpha(\mathcal{Q}))$*. If* XY *is nondegenerate, there exists* $\lambda \in \mathbb{R}$ *such that*

$$f[X] = \lambda S_\alpha[X], \quad f[Y] = \lambda S_\alpha[Y], \quad f[XY] = \lambda S_\alpha[XY].$$

Proof. As f is a 1-cocycle, it satisfies the two equations derived from (2.29)

$$Y.f[X] = f[XY] - f[Y], \tag{3.27}$$

$$X.f[Y] = f[XY] - f[X]. \tag{3.28}$$

and therefore the symmetric equation

$$X.f[Y] - f[Y] = Y.f[X] - f[X]. \tag{3.29}$$

For a law P, we write

$$\begin{pmatrix} s & t & u & \dots \\ p & q & r & \dots \end{pmatrix}$$

if $P(s) = p$, $P(t) = q$, $P(u) = r$, etc. and the probabilities of the unwritten parts are zero.

Fix enumerations $(x_1, ..., x_k)$ and $(y_1, ..., y_l)$ that satisfy the definition of nondegenerate product, and let $\{\gamma_i\}_{i=1}^m$ be the corresponding NE path. Write $\gamma_i = (a, b)$. If $\gamma_{i+1} - \gamma_i = (1, 0)$, we shall show that the following recursive formula holds:

$$f[X]\begin{pmatrix} x_a & \dots & x_k \\ \mu_a & \dots & \mu_k \end{pmatrix} = (1-\mu_a)^\alpha f[X]\begin{pmatrix} x_{a+1} & \dots & x_k \\ \mu_{a+1}/(1-\mu_a) & \dots & \mu_k/(1-\mu_a) \end{pmatrix} + f[X]\begin{pmatrix} x_a & x_{a+1} \\ \mu_a & 1-\mu_a \end{pmatrix}. \tag{3.30}$$

Analogously, if $\gamma_{i+1} - \gamma_i = (0, 1)$,

$$f[Y]\begin{pmatrix} y_b & \dots & y_l \\ \nu_b & \dots & \nu_l \end{pmatrix} = (1-\nu_b)^\alpha f[Y]\begin{pmatrix} y_{b+1} & \dots & y_l \\ \nu_{b+1}/(1-\nu_c) & \dots & \nu_l/(1-\nu_b) \end{pmatrix} + f[Y]\begin{pmatrix} y_b & y_{b+1} \\ \nu_b & 1-\nu_b \end{pmatrix}. \tag{3.31}$$

Suppose that $\gamma_{i+1} - \gamma_i = (1, 0)$. Let

$$p = \begin{pmatrix} x_a & \dots & x_k \\ \mu_a & \dots & \mu_k \end{pmatrix}$$

be a probability in $\mathcal{Q}_X \cap [\{\, x_i \,|\, a \le i \le k \,\}]$. We know it has a preimage $\tilde{p}$ under marginalization X_* as in Definition 3.12-(2): for such law, knowledge of X implies knowledge of Y with certainty, therefore $X.f[Y](\tilde{p}) = 0$; by equation (3.28),

$f[XY](\tilde{p}) = f[X](X_*\tilde{p}) = f[X](p)$. Equation (3.27) reads

$$(1-\mu_a)^\alpha f[X]\begin{pmatrix} x_{a+1} & \dots & x_k \\ \mu_{a+1}/(1-\mu_a) & \dots & \mu_k/(1-\mu_a)\end{pmatrix} = \\ f[X]\begin{pmatrix} x_a & \dots & x_k \\ \mu_a & \dots & \mu_k \end{pmatrix} - f[Y]\circ\tau\begin{pmatrix} y_b & y_{b+1} \\ 1-\mu_a & \mu_a\end{pmatrix}, \tag{3.32}$$

where τ is the identity or the transposition of the arguments. In any case, setting $\mu_{a+1} = 1-\mu_a$ and $\mu_{a+2} = \dots = \mu_k = 0$, we conclude that

$$f[X]\begin{pmatrix} x_a & x_{a+1} \\ \mu_a & 1-\mu_a \end{pmatrix} = f[Y]\circ\tau\begin{pmatrix} y_b & y_{b+1} \\ 1-\mu_a & \mu_a\end{pmatrix}, \tag{3.33}$$

which combined with (3.32) implies (3.30). The identity (3.31) can be obtained analogously.

We proceed to the determination of

$$\phi_a(z) := f[X]\begin{pmatrix} x_a & x_{a+1} \\ z & 1-z \end{pmatrix} \quad\text{and}\quad \psi_b(z) := f[Y]\begin{pmatrix} y_b & y_{b+1} \\ z & 1-z \end{pmatrix}, \quad \text{for } z\in[0,1].$$

Let b_1, b_2, b_3 be the three elements of $E_{XY}\subset E_X\times E_Y$ such that $[\delta_{b_1},\delta_{b_2},\delta_{b_3}]$ is the 2-cell given by Definition 3.12-(1) when $\gamma_i = (a,b)$. For any $\mu = (\mu_0,\mu_X,\mu_Y)\in\Delta^2$, set $P(b_i) := \mu_Y$, where b_i is the component that differ from the others on the y-coordinate (in such a way that $Y_*\mu$ is $(\mu_Y, 1-\mu_Y)$). Similarly, set $P(b_j) := \mu_X$, where b_j is the component that differ from the others on the x-coordinate. With this assignment, the equation (3.29) gives:

$$(1-\mu_X)^\alpha f[Y]\circ\sigma\begin{pmatrix} y_b & y_{b+1} \\ \mu_0/(1-\mu_X) & \mu_Y/(1-\mu_X)\end{pmatrix} - f[Y]\circ\sigma\begin{pmatrix} y_b & y_{b+1} \\ 1-\mu_Y & \mu_Y\end{pmatrix} \\ = (1-\mu_Y)^\alpha f[X]\circ\tau\begin{pmatrix} x_a & x_{a+1} \\ \mu_0/(1-\mu_Y) & \mu_Y/(1-\mu_Y)\end{pmatrix} - f[X]\circ\tau\begin{pmatrix} x_a & x_{a+1} \\ 1-\mu_X & \mu_X\end{pmatrix}, \tag{3.34}$$

where σ, τ are the identity or the transposition of both nontrivial arguments. In any case, this leads to the functional equation in Proposition 3.10, which imply that $\phi_a(z) = \psi_b(z) = \lambda s_\alpha(z)$ for certain $\lambda\in\mathbb{R}$ (the solution is symmetric in the arguments).

When considering γ_{i+1}, one finds the functions ϕ_a and ψ_{b+1}, or the functions ϕ_{a+1} and ψ_b, since the difference $\gamma_{i+1}-\gamma_i$ is either $(0,1)$ or $(1,0)$. This ensures that the constant λ that appears for each γ_i is always the same.

Repeat the process above with every γ_i ($1\le i\le m$). The collection of equations (3.30) obtained in this way, together with the functions already determined $f[X]\begin{pmatrix} x_a & x_{a+1} \\ z & 1-z \end{pmatrix}$, $1\le a\le k-1$, entails that

$$f[X](\mu_1,\dots,\mu_k) = \lambda\sum_{i=0}^{k-1}\left(1-\sum_{j=1}^{i}\mu_j\right)^\alpha s_\alpha\left(\frac{\mu_{i+1}}{\left(1-\sum_{j=1}^{i}\mu_j\right)}\right). \tag{3.35}$$

Set $T_i := 1-\sum_{j=1}^i \mu_j$. A direct computation shows that, when $\alpha = 1$,

$$\sum_{i=0}^{k-1}(1-T_i)\, s_1\left(\frac{\mu_{i+1}}{(1-T_i)}\right) = \sum_{i=1}^{k}\mu_i\log\mu_i, \tag{3.36}$$

and when $\alpha \neq 1$,

$$\sum_{i=0}^{k-1} (1 - T_i)^\alpha s_\alpha \left(\frac{\mu_{i+1}}{(1 - T_i)} \right) = \sum_{i=1}^{k} \mu_i^\alpha - 1. \tag{3.37}$$

Therefore, for any $\alpha > 0$, we have $f[X] = \lambda S_\alpha[X]$. Analogously, $f[Y] = \lambda S_\alpha[Y]$. □

3.5 Determination of H^1

In this section, we shall specify conditions on $(\mathbf{S}, E, Q)$ that allows us to determine $H^1(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q}))$.

We call a variable Z ***reducible*** if there exist $X, Y \in \operatorname{Ob} \mathbf{S} \setminus \{1, Z\}$ such that $Z = XY$, and ***irreducible*** otherwise.

We denote by $\min(\mathbf{S})$ the set of minimal objects in $\mathbf{S}$: these are the $Y \in \operatorname{Ob} \mathbf{S}$ such that $\nexists X \in \operatorname{Ob} \mathbf{S}$ with $X \to Y$.

Theorem 3.14. *Let* $(\mathbf{S}, E)$ *be a bounded, finite information structure, and* $\mathscr{Q}$ *an adapted probability functor. Denote by* $\mathbf{S}^*$ *the full subcategory of* $\mathbf{S}$ *generated by* $\operatorname{Ob} \mathbf{S} \setminus \{\mathbf{1}\}$*. Suppose that every minimal object can be factored as a nondegenerate product. Then,*

$$H^1(\mathbf{S}, \mathscr{F}_1(\mathscr{Q})) = \prod_{[\mathbf{C}] \in \pi_0(\mathbf{S}^*)} \mathbb{R} \cdot S_1^{\mathbf{C}} \tag{3.38}$$

and, when $\alpha \neq 1$,

$$H^1(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q})) = \left(\prod_{[\mathbf{C}] \in \pi_0(\mathbf{S}^*)} \mathbb{R} \cdot S_\alpha^{\mathbf{C}} \right) / \mathbb{R} \cdot S_\alpha \tag{3.39}$$

In the formulae above, $\mathbf{C}$ *represents a connected component of* $\mathbf{S}^*$*, and*

$$S_\alpha^{\mathbf{C}}[X] = \begin{cases} S_\alpha[X] & \text{if } X \in \operatorname{Ob} \mathbf{C} \\ 0 & \text{if } X \notin \operatorname{Ob} \mathbf{C} \end{cases}$$

Proof. Let f be an element of $Z^1(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q}))$. If every $M \in \min(\mathbf{S})$ can be factorized as a nondegenerate product, Proposition 3.13 implies that $f[M] = \lambda_M S_\alpha[M]$, where λ_M is a constant that depends *a priori* on M. If $M \to Z$ in $\mathbf{S}$,

$$f[Z] = f[M] - Z.f[M] = \lambda_M S_\alpha[M] - Z.(\lambda_M S_\alpha[M]) = \lambda_M S_\alpha[Z]. \tag{3.40}$$

If Z is refined by two variables $M, N \in \min(\mathbf{S})$, we can apply the previous formula twice to conclude that $f[Z] = \lambda_M S_\alpha[Z] = \lambda_N S_\alpha[Z]$, and therefore $\lambda_M = \lambda_N$.

Let M, N be two elements of $\min(\mathbf{S})$. If they belong to the same connected component of $\mathbf{S}^*$, there is a zig-zag diagram in $\mathbf{S}^*$ of the form

$$M \to X_1 \leftarrow M_1 \to X_2 \leftarrow M_2 \to \cdots \leftarrow M_k \to X_{k+1} \leftarrow N$$

for certain $k \in \mathbb{N}$, where $M_i \in \min(\mathbf{S})$ for all i. The repeated application of the argument in the previous paragraph implies that $\lambda_M = \lambda_{M_1} = \cdots \lambda_N$.

On the other hand, if $\mathbf{C}$ and $\mathbf{C}'$ are different components of $\mathbf{S}^*$, there is no cocycle equation that relates $f[X]$ and $f[Y]$, for any variables $X \in \operatorname{Ob} \mathbf{C}$ and $Y \in \operatorname{Ob} \mathbf{C}'$. In fact, such a cocycle equation only is possible if there is a third nontrivial variable Z

such that $X, Y \in \mathscr{S}_Z$, and therefore $[X], [Y]$ appear as generators of $\mathscr{B}_1(Z)$; but this would mean that $X \leftarrow Z \rightarrow Y$ in $\mathbf{S}^*$.

The previous argument proves that $Z^1(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q})) \cong \prod_{\mathbf{C}\in\pi_0(\mathbf{S}^*)} \mathbb{R}\cdot S_1^{\mathbf{C}}$, and we saw in Section 3.4 that $\delta C^0(\mathbf{S}, \mathscr{F}_1(\mathscr{Q})) \cong \langle 0\rangle$ and $\delta C^0(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q})) \cong \mathbb{R}\cdot S_\alpha$. □

As a byproduct of the previous proof, we also obtain the following proposition.

Proposition 3.15. *Let $\{(\mathbf{S}_i, E_i, \mathscr{Q}_i)\}_{i\in I}$ be a collection of triples that satisfy separately the hypotheses stated in Theorem 3.14. Then,*

$$Z^1\left(\coprod_{i\in I} \mathbf{S}, \mathscr{F}(\coprod_{i\in I} \mathscr{Q})\right) \cong \prod_{i\in I} Z^1(\mathbf{S}_i, \mathscr{F}(\mathscr{Q}_i)).$$

Proof. The category $(\coprod_{i=1}^n \mathbf{S})^*$ is the disjoint union of the categories $\mathbf{S}_i^*$, for $i \in I$. □

The cases uncovered by Theorem 3.14 can be classified in two families:

1. There is an irreducible minimal object;
2. All minimal objects are reducible, but some of them cannot be written as nondegenerate products.

In the latter case, all kinds of behaviours are possible, as the examples at the end this section show.

In the Example 3.8, we proved that $H^1(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q}))$ has infinite dimension when $\mathbf{S} \cong \mathbf{Obs}_{\text{fin}}(\{0,1\})$; in this case, there is only one nontrivial variable and it is obviously irreducible. Now we proceed to the generalization of this result.

Proposition 3.16. *Let $(\mathbf{S}, E)$ be an information structure and $\mathscr{Q}$ an adapted probability functor. Let $M \in \min(\mathbf{S})$ be an irreducible minimal object, $X_1, ..., X_n$ all the variables coarser than M and suppose that*

$$M \to X_1 \to ... \to X_n \to \mathbf{1},$$

in $\mathbf{S}$. Moreover, suppose that there exists $k \in \{1, ..., n\}$ such that, for some $x \in E(X_k)$, the set $\mathscr{Q}_M^x := \{\, P|_{X_k=x} \mid P \in \mathscr{Q}_M \,\}$ contains at least one nonatomic law.[4] *Then,* $\dim H^1(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q})) = \infty$.

Proof. Let $\tilde{k}$ be the smallest k that satisfies the stated hypothesis. Set $f[X_i] = 0$ for $i = \tilde{k}, ..., n$. We shall show that $f[M]$ can be chosen arbitrarily.

The cocycle equations are

$$0 = X_i.f[M] - f[X_iM] + f[X_i], \quad i = 1, ..., n.$$

For $i < \tilde{k}$, the term $X_i.f[M]$ vanishes because all conditioned laws $P|_{X_i=x}$ give δ-laws. This implies that $f[X_i] = f[M]$. For $i = \tilde{k}$, we obtain $f[M] = X_{\tilde{k}}.f[M]$. Given this one, the others equations become redundant, since

$$X_j.f[M] = X_j.(X_{\tilde{k}}.f[M]) = (X_jX_{\tilde{k}}).f[M] = X_{\tilde{k}}.f[M], \quad \text{for } j > \tilde{k}.$$

Let $P \in \mathscr{Q}_M$ and $\tilde{P} = (X_{\tilde{k}})_* P \in \mathscr{Q}_{X_{\tilde{k}}}$. The equation $f[M] = X_{\tilde{k}}.f[M]$ reads

$$f[M](P) = \sum_{x\in X_{\tilde{k}}} \tilde{P}(x) f[M](P|_{X_{\tilde{k}}=x}).$$

[4] The law $P \in \mathscr{Q}_M$ is atomic if $P = \delta_m$ for some $m \in M$.

If $P|_{X_{\tilde{k}}=x} = \delta_m$ for some $m \in M$ (atomic law), then $f[M](P|_{X_{\tilde{k}}=x}) = 0$; otherwise, no condition determines $f[M](P|_{X_{\tilde{k}}=x})$. This means that, for each set $\mathscr{Q}_M^x$ that contains nonatomic laws, we can introduce an arbitrary function. $\square$

We illustrate the proof with an example. Consider $\Omega = \{0, 1, 2\}$ and the concrete structure $\mathbf{S}$ given by

$$M = \{\{0\}, \{1\}, \{2\}\} \to X_1 = \{\{0, 1\}, \{2\}\} \to \mathbf{1}_\Omega$$

In this case, there is an infinite family of cocycles given by $f[X_1] \equiv 0$, $f[\mathbf{1}] \equiv 0$ and

$$f[M](p_0, p_1, p_2) = (p_0+p_1)f[M]\left(\frac{p_0}{p_0+p_1}, \frac{p_1}{p_0+p_1}, 0\right) = (p_0+p_1)g\left(\frac{p_0}{p_0+p_1}, \frac{p_1}{p_0+p_1}\right),$$

where $g : \Delta^1 \to \mathbb{R}$ is an arbitrary measurable function such that $g(1, 0) = g(0, 1) = 0$.

To close this section, we make some remarks about case of reducible minimal objects that cannot be written as nondegenerate products. If the product is degenerate, multiple constants can appear or the dimension of $H^1(\mathbf{S}, \mathscr{F}_\alpha(\mathscr{Q}))$ can explode to infinity, as the following examples show.

Example 3.17. Consider the information structure $(\mathbf{S}, E)$ given by the poset $\mathbf{S}$ represented by

$$\begin{tikzcd} & \mathbf{1} & \\ X \arrow[ur] & & Y \arrow[ul] \\ & XY \arrow[ul] \arrow[ur] & \end{tikzcd}$$

and the assignment $E(X) = \{x_1, x_2, x_3, x_4\}$, $E(Y) = \{y_1, y_2, y_3, y_4\}$, and $E(XY) = E(X) \times E(Y)$; the arrows are mapped to the terminal maps and the canonical projectors. Recall the notation $[S]$ introduced before Definition 3.12. Let

$$\mathscr{Q}_{XY} = \Pi(XY) \cap ([\{x_1, x_2\} \times \{y_1, y_2\}] \cup [\{x_3, x_4\} \times \{y_3, y_4\}]). \tag{3.41}$$

Therefore, we just need to determine

$$f[X]\begin{pmatrix} x_1 & x_2 \\ p & 1-p \end{pmatrix}, f[X]\begin{pmatrix} x_3 & x_4 \\ p & 1-p \end{pmatrix}, f[Y]\begin{pmatrix} y_1 & y_2 \\ p & 1-p \end{pmatrix} \text{ and } f[Y]\begin{pmatrix} y_3 & y_4 \\ p & 1-p \end{pmatrix},$$

for arbitrary $p \in [0, 1]$. Proposition 3.10 allows us to conclude that

$$f[X]\begin{pmatrix} x_1 & x_2 \\ p & 1-p \end{pmatrix} = \lambda_1 s_\alpha(p), \quad f[Y]\begin{pmatrix} y_1 & y_2 \\ p & 1-p \end{pmatrix} = \lambda_1 s_\alpha(p). \tag{3.42}$$

We can use this proposition a second time to show that

$$f[X]\begin{pmatrix} x_3 & x_4 \\ p & 1-p \end{pmatrix} = \lambda_2 s_\alpha(p), \quad f[Y]\begin{pmatrix} y_3 & y_4 \\ p & 1-p \end{pmatrix} = \lambda_2 s_\alpha(p). \tag{3.43}$$

However, from Equations (3.27) and (3.28) is impossible to find a relation between λ_1 and λ_2 when $P \in \mathscr{Q}_{XY}$. We conclude that $Z^1(\mathscr{F}_\alpha(\mathscr{Q})) \cong \mathbb{R}^2$.

Example 3.18. Let $(\mathbf{S}, E)$ be the information structure defined in the previous example and

$$\mathscr{Q}_{XY} = \Pi(XY) \cap ([\{x_1, x_2\} \times \{y_1, y_2\}] \cup [\{(x_3, y_3), (x_4, y_4)\}]). \tag{3.44}$$

As before, we conclude that

$$f[X]\begin{pmatrix} x_1 & x_2 \\ p & 1-p \end{pmatrix} = \lambda_1 s_\alpha(p), \quad f[Y]\begin{pmatrix} y_1 & y_2 \\ p & 1-p \end{pmatrix} = \lambda_1 s_\alpha(p). \tag{3.45}$$

Equations (3.27) and (3.28) imply that

$$f[XY]\begin{pmatrix} (x_3, y_3) & (x_4, y_4) \\ p & 1-p \end{pmatrix} = f[X]\begin{pmatrix} x_3 & y_3 \\ p & 1-p \end{pmatrix} = f[Y]\begin{pmatrix} y_3 & y_4 \\ p & 1-p \end{pmatrix}. \tag{3.46}$$

and these are the only relations between these functions. Any measurable $g : \Delta^1 \to \mathbb{R}$ such that $g(0,1) = g(1,0) = 0$ solves these equations. This means that $\dim Z^1(\mathscr{F}_\alpha(\mathscr{Q})) = \infty$.

3.6 Functorial extensions of algebras

The cocycle equation has a meaning in the context of extensions of algebras. We introduce first some general definitions and results from [95]; what is said there for algebras remains valid with *pre*sheaves of algebras.

Let Λ be a presheaf of $\mathbb{R}$-algebras on $\mathbf{S}$. An extension of Λ is an epimorphism $\sigma : \Gamma \to \Lambda$. The extension is called ***singular*** (or square zero) if $\ker(\sigma_X)^2 = 0$ for all $X \in \operatorname{Ob} S$ (in this case, $\ker(\sigma_X)$ can be regarded as a Λ_X-bimodule). It is called ***cleft*** if there exists a morphism $\phi : \Lambda \to \Gamma$, ϕ_X morphism of algebras, such that $\sigma \circ \phi = 1_\Lambda$. Given a Λ-bimodule M, a singular extension of Λ by M is a short exact sequence

$$0 \longrightarrow M \xrightarrow{\ \xi\ } \Gamma \xrightarrow{\ \sigma\ } \Lambda \longrightarrow 0 \tag{3.47}$$

where ξ is a morphism of Γ-bimodules (M is Γ-bimodule by $\gamma.m = \sigma(\gamma).m$, etc.). Two extensions are called ***congruent*** if there is an algebra morphism $\gamma : \Gamma \to \Gamma'$ making

$$\begin{tikzcd}
 & & \Gamma \arrow[dd, "\gamma"] \arrow[dr, "\sigma"] & & \\
0 \arrow[r] & M \arrow[ur, "\xi"] \arrow[dr, "\xi'"'] & & \Lambda \arrow[r] & 0 \\
 & & \Gamma' \arrow[ur, "\sigma'"'] & &
\end{tikzcd}$$

commute.

A particular singular cleft extension of Λ by M is given by the ***semidirect sum***, defined to be the presheaf of vector spaces $M \oplus \Lambda$ with product defined by $(m_1, \lambda_1) \bullet (m_2, \lambda_2) = (m_1\lambda_2 + \lambda_1 m_2, \lambda_1\lambda_2)$; with $\xi(m) = (m, 0)$ and $\sigma(m, \lambda) = \lambda$. The following Proposition is a well known result.

Proposition 3.19. *Any singular cleft extension is congruent to $M \rtimes \Lambda$.*

In our case, $\Lambda = \mathscr{A}$ and $M = \mathscr{F}_\alpha := \mathscr{F}_\alpha(\mathscr{Q})$, turned into a presheaf of $\mathscr{A} - \mathscr{A}$-bimodules with trivial right action: this means that each variable acts as the identity endomorphism. If Γ is a singular cleft extension of $\mathscr{A}$ by $\mathscr{F}_\alpha$, it is isomorphic to $\mathscr{F}_\alpha \rtimes \mathscr{A}$. What are the possible morphisms $\phi : \mathscr{A} \to \mathscr{F}_\alpha \rtimes \mathscr{A}$ that implement this splitting? Set $\phi(X) = (d[X], X)$; since ϕ_X is a morphism of algebras,

$$(d[Y], Y) \bullet (d[X], X) = (d[YX], YX) \Leftrightarrow (d[Y] + Y.d[X], YX) = (d[YX], YX).$$

Thus d must be a 1-cocycle (also called derivation in this context). In this chapter, we have proved that in general there is no choice, one must take the entropy. Therefore, we can say that the entropy is the unique derivation that transforms a multiplicative operation on partitions into an additive operation on functions, introducing an appropriate 'twist'.

Note that the definition of $\mathscr{F}_\alpha$ guarantees that $d[X](P)$ depends only on X_*P. This turns out to be the appropriate notion of locality and justifies the introduction of presheaves.

The extensions that are singular and $\mathbb{R}$-split (instead of cleft) are classified by $H^2(\mathbf{S}, \mathscr{F}_\alpha)$: the morphism $\phi_X : \mathscr{A}_X \to \Gamma_X$ gives a natural vector space decomposition $\Gamma_X \simeq \mathscr{A}_X \oplus \mathscr{F}_\alpha(X)$, with product given by $(X, f) \bullet (Y, g) = (XY, f + X.g + a(X, Y))$. The function $a : \mathscr{A} \otimes_{\mathbb{R}} \mathscr{A} \to \mathbb{R}$ is called the factor set of the extension and the associativity of the product in Γ entails that a is a 2-cocycle.

3.7 Product structures and divergence

We prove in this section that the Kullback-Leibler (KL) divergence and the cross-entropy are cocycles for an adapted module of coefficients. Both quantities measure the relation between two probability laws; the KL divergence gives a nonsymmetric notion of distance. The proof is elementary and sheds some light on the meaning of these cocycle equations from the probabilistic viewpoint.

Let (X, Y) be a joint random variable taking values in a set $E_{XY} \subset E_X \times E_Y$, with certain probability law $P = \{P(x, y)\}_{(x,y)\in E_{XY}}$. Suppose n measurements of this variable are performed, obtaining in this way a realization $\mathbf{z} = (z_1, \dots, z_n) \in E_{XY}^n$; define then the ***empirical distribution*** Q by the formula

$$\forall (a, b) \in E_{XY}, \quad Q(a, b) = \frac{|\{ (x, y) \in E_{XY} \mid x = a \text{ and } y = b \}|}{n}. \tag{3.48}$$

The probability of the realization $\mathbf{z}$ is (cf. [23, § 2])

$$\begin{aligned}
P^{\otimes n}(\mathbf{z}) &= \prod_{(x,y)\in E_{XY}} P(x, y)^{Q(x,y)n} && (3.49)\\
&= \exp\left(n \sum_{(x,y)\in E_{XY}} Q(x, y) \log P(x, y) \right) && (3.50)\\
&= \exp\left(n \sum_{(x,y)\in E_{XY}} Q(x, y) \log \frac{P(x, y)}{Q(x, y)} + n \sum_{(x,y)\in E_{XY}} Q(x, y) \log Q(x, y) \right) && (3.51)\\
&= \exp(-n\{D[XY](Q||P) + H[XY](Q)\}) && (3.52)
\end{aligned}$$

where we have used the definition of the ***Kullback-Leibler (KL) divergence***

$$\forall P_1, P_2 \in \Pi(X), \quad D[X](P_1||P_2) := -\sum_{x\in E_X} P_1(x) \ln \frac{P_2(x)}{P_1(x)}. \tag{3.53}$$

Whenever $Q(x) := X_* Q(x) \neq 0$, we have the conditional probability

$$Q(y|x) = \frac{Q(x,y)}{Q(x)}. \tag{3.54}$$

The case $Q(x) = 0$ is not relevant for our computation for it forces $Q(x,y)$ to vanish for any y and the corresponding factor in (3.49) is $P(x,y)^{Q(x,y)n} = 1$ i.e. it does not contribute to the product, that can be restricted to $Q(x,y) \neq 0$, and similarly the sums.

Writing $P(y|x)p(x)$ instead of $P(x,y)$, we also have that

$$\begin{aligned}
P^{\otimes n}(\mathbf{z}) &= \prod_{\substack{(x,y)\in E_{XY}\\ Q(x,y)\neq 0}} (P(y|x)P(x))^{Q(x,y)n} && (3.55)\\
&= \prod_{\substack{(x,y)\in E_{XY}\\ Q(x,y)\neq 0}} P(x)^{Q(x,y)n} \prod_{\substack{(x,y)\in E_{XY}\\ Q(x,y)\neq 0}} P(y|x)^{Q(x,y)n} && (3.56)\\
&= \prod_{\substack{x\in E_X\\ Q(x)\neq 0}} P(x)^{Q(x)n} \prod_{\substack{(x,y)\in E_{XY}\\ Q(x,y)\neq 0}} P(y|x)^{Q(y|x)Q(x)n} && (3.57)\\
&= \exp\left(n \sum_{x\in E_X} Q(x) \ln P(x)\right) \prod_{\substack{x\in E_X\\ Q(x)\neq 0}} \exp\left(nQ(x) \sum_{y\in E_Y} Q(y|x) \ln P(y|x)\right). && (3.58)
\end{aligned}$$

For convenience, let us introduce the cross-information

$$C[X](Q:P) = -\sum_{x\in E_X} Q(x) \ln P(x). \tag{3.59}$$

Note that $D[X](Q||P) = C[X](Q:P) - H[X](Q)$ and $C[X](Q:Q) = H[X](Q)$.

A comparison of the exponents in equations (3.50) and (3.58) shows that

$$C[XY](Q:P) = C[X](X_*Q : X_*P) + \sum_{y\in E_Y} Q(x) C[Y](Q|_{X=x} : P|_{X=x}). \tag{3.60}$$

This is the 1-cocycle condition of information cohomology with coefficients in the following ***bivariate module*** (called 'product structure' in [10, Sec. 5]): for each X in Ob $\mathbf{S}$, let $\mathscr{F}^{(2)}(X)$ be the vector space of measurable functions $f : \mathscr{Q}_X \times \mathscr{Q}_X \to \mathbb{R}$, for a given functor of probabilities $\mathscr{Q}$, and define the action of $\mathscr{S}_X$ on $\mathscr{F}^{(2)}(X)$ by the formula:

$$\forall Y \in \mathscr{S}_X, \forall f \in \mathscr{F}^{(2)}(X), \forall P, Q \in Q_X, \quad (Y.f)(Q,P) = \sum_{y\in E_Y} Q(y) f(Q|_{Y=y}, P|_{Y=y}). \tag{3.61}$$

This gives a functorial action of monoids; the proof is just a modification of those in Section 3.1. Therefore $\mathscr{F}^{(2)}$ has the structure of $\mathscr{A}$-module, that we denote $\mathscr{F}_1^{(2)}$. Then (3.60) is the cocycle condition. By its frequentist nature, the proof above only works for rational Q, but (3.60) must be valid everywhere by continuity.

In the particular case $P = Q$, this shows that H is a 1-cocycle. By additivity, the KL is a cocycle too. Cf. The proof of Proposition 4 in [10].

The KL divergence also accepts an α-deformation. It can be formulated very naturally by introducing first the α-logarithm:

$$\ln_\alpha(x) := \int_1^x \frac{1}{t^\alpha}\,\mathrm{d}t = \frac{x^{1-\alpha}-1}{1-\alpha}. \tag{3.62}$$

Remark that $\ln_\alpha \to \ln_1 = \ln$ when $\alpha \to 1$. Shannon entropy $S_1[X](P)$ equals $\mathbb{E}_P\left(-\ln P(X)\right)$, whereas Tsallis α-entropy $S_\alpha[X](P)$ is $\mathbb{E}_P\left(-\ln_\alpha P(X)\right)$. The α-logarithm satisfies [88, p. 38]

$$\forall x, y > 0, \quad \ln(xy) = \ln(x) + \ln(y) - (\alpha-1)\ln(x)\ln(y). \tag{3.63}$$

The natural definition for the generalized KL divergence is, for all $\alpha > 0$,

$$D_\alpha[X](Q||P) = \sum_{x\in E_X} Q(x)\ln_\alpha\left(\frac{Q(x)}{P(x)}\right) \tag{3.64}$$

the case $\alpha = 1$ recovers (3.53).

Writing $P(y|x)P(x)$ instead of $P(x,y)$ (the same for Q) and using 3.63, we obtain

$$\begin{aligned}
&D_\alpha[XY](Q||P)\\
&= \sum_{(x,y)\in E_{XY}} Q(x,y)\ln_\alpha\left(\frac{Q(y|x)Q(x)}{P(y|x)P(x)}\right)\\
&= \sum_{(x,y)\in E_{XY}} Q(x,y)\left(\ln_\alpha\left(\frac{Q(x)}{P(x)}\right) + \ln_\alpha\left(\frac{Q(y|x)}{P(y|x)}\right) - (\alpha-1)\ln_\alpha\left(\frac{Q(x)}{P(x)}\right)\ln_\alpha\left(\frac{Q(y|x)}{P(y|x)}\right)\right)\\
&= D_\alpha[X](X_*Q||X_*P) + \sum_{(x,y)\in E_{XY}} Q(x,y)\left(\ln_\alpha\left(\frac{Q(y|x)}{P(y|x)}\right) - \left(\left(\frac{Q(x)}{P(x)}\right)^{\alpha-1} - 1\right)\ln_\alpha\left(\frac{Q(y|x)}{P(y|x)}\right)\right)\\
&= D_\alpha[X](X_*Q||X_*P) + \sum_{x\in E_X} Q(x)\left(\frac{Q(x)}{P(x)}\right)^{\alpha-1} D_\alpha[Y](Y_*Q|_{Y=y}||Y_*P|_{Y=y})
\end{aligned}$$

Since $D_\alpha[X]$ belongs to $\mathscr{F}^{(2)}(X)$, the vector space of functions introduced above, and satisfies the preceding equation, we conclude that it is also a cocycle for an $\mathscr{A}$-module $\mathscr{F}_\alpha^{(2)}$ such that the variables act as follows: for all $X \in \operatorname{Ob}\mathbf{S}$, $Y \in \mathscr{S}_X$, $f \in \mathscr{F}(X)$ and $(P,Q) \in \mathscr{Q}_X^2$,

$$\begin{aligned}
(Y.f)(Q,P) &= \sum_{y\in E_Y} Q(y)\left(\frac{Q(y)}{P(y)}\right)^{\alpha-1} f(Q|_{Y=y}, P|_{Y=y})\\
&= \sum_{y\in E_Y} Q(y)^\alpha P(x)^{1-\alpha} f(Q|_{Y=y}, P|_{Y=y}).
\end{aligned} \tag{3.65}$$

The case $\alpha = 1$ recovers the action (3.61).

Here we should take into account that P and Q are in fact densities of measures ρ and τ, respectively. Let ν denote the counting measure on E_X. In terms of these measures, the action (3.65) is

$$(Y.f)(\tau, \rho) = \int_{E_Y} f(Q|_{Y=y}, P|_{Y=y}) \left(\frac{\mathrm{d}\tau}{\mathrm{d}\nu}\right)^{\alpha} \left(\frac{\mathrm{d}\rho}{\mathrm{d}\nu}\right)^{1-\alpha} \mathrm{d}\nu \tag{3.66}$$

When $\rho = \nu$, we recover the action (3.3). This is clarified in Chapter 12: the KL divergence is in fact a particular case of relative entropy, and the usual Shannon entropy is the relative entropy with respect to the counting measure. The relative entropy has a general interpretation from the point of view of the asymptotic equipartition property: the reference measure gives the appropriate notion of volume. In this section the reference measure was P, because we were interested in the probabilities of realizations $\mathbf{z}$ according to this law.

Chapter 4

Combinatorial information cohomology

4.1 Counting functions

Let $(\mathbf{S}, E)$ be a finite information structure, and $\mathscr{C} : \mathbf{S} \to \mathbf{Sets}$ a functor that associates to each object X the set

$$\mathscr{C}_X = \left\{ \nu : E_X \to \mathbb{N} \,\Big|\, \sum_{x \in E_X} \nu(x) > 0 \right\}, \tag{4.1}$$

and to each arrow $f : X \to Y$, associated to a surjection $E(f) = \pi_{YX} : E_X \to E_Y$, the map $\mathscr{C}(f) : \mathscr{C}(X) \to \mathscr{C}(Y)$ that verifies $(\mathscr{C}(f)(\nu))(y) = \sum_{x \in \pi_{YX}^{-1}(y)} \nu(x)$. To simplify notation, we shall write $Y_* \nu$ instead of $\mathscr{C}(f)(\nu)$, whenever X is clear from context. The elements of $\mathscr{C}_X$ are called ***counting functions***. For $\nu_X \in \mathscr{C}_X$, we define its ***support*** as $\{\, x \in E_X \mid \nu_X(x) \neq 0 \,\}$, and its ***magnitude*** as the quantity $\|v\| := \sum_{x \in X} \nu(x)$.

For any subset A of X, we define

$$\nu|_A(x) := \begin{cases} \nu(x) & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}. \tag{4.2}$$

When $\|\nu|_A\| > 0$, we call $\nu|_A$ the ***restricted counting given*** $A \subset X$. Given an arrow $f : X \to Y$, the notation $\nu|_{Y=y}$ stands for $\nu|_{\pi_{YX}^{-1}(y)}$. Remark that $\nu_\emptyset = 0$ and $\left\|\nu|_{Y=y}\right\| = Y_* \nu(y)$.

Consider now the multiplicative abelian group $\mathscr{G}_X$, whose elements are $\mathbb{R}_+^*$-valued measurable functions defined on $\mathscr{C}_X$. By $\mathbb{R}_+^*$ we mean $\{\, x \in \mathbb{R} \mid x > 0 \,\}$. (The multiplicative notation is convenient, because multinomial coefficients appear directly as cocycles.) The group $\mathscr{G}_X$ becomes a real vector space if we define $(r.f)(\nu) := (f(\nu))^r$, for each $f \in \mathscr{G}_X$ and each $r \in \mathbb{R}$. [1] For each $Y \in \mathscr{S}_X$ and each $g \in \mathscr{G}_X$, set [2]

$$(Y.g)(\nu) := \prod_{\substack{y \in E_Y \\ Y_* \nu(y) \neq 0}} g(\nu|_{Y=y}). \tag{4.3}$$

[1]In principle this is a right action, but this is immaterial because $\mathbb{R}$ is commutative.

[2]In the probabilistic case, the sum $(Y.f)(P) = \sum_{y \in E_Y} Y_* P(y) f(P|_{Y=y})$ can be restricted to those $y \in E_Y$ such that $Y_* P(y) \neq 0$.

Finally, define $(aY).g := a.(Y.g) = Y.(a.g)$. As a consequence of the following proposition, these formulae give an homomorphism $\rho_X : \mathscr{A}_X \to \mathrm{End}(\mathscr{G}_X)$, that turns $\mathscr{G}_X$ into an $\mathscr{A}_X$-module.

Proposition 4.1. *Given variables $Y, Z \in \mathscr{S}_X$ and $f \in \mathscr{G}_X$,*

$$ZY.f = Z.(Y.f). \tag{4.4}$$

Proof.

$$\begin{aligned} Z.(Y.f)(\nu) &= \prod_{\substack{z \in E_Z \\ Z_*\nu(z) \neq 0}} (Y.f)(\nu|_{Z=z}) && (4.5) \\ &= \prod_{\substack{z \in E_Z \\ Z_*\nu(z) \neq 0}} \prod_{\substack{y \in E_Y \\ Y_*\nu|_{Z=z}(y) \neq 0}} f((\nu|_{Z=z})|_{Y=y}). && (4.6) \end{aligned}$$

From the definition of conditioning, we deduce that $(\nu|_{Z=z})|_{Y=y} = \nu|_{\{Z=z\}\cap\{Y=y\}} = \nu|_{A(y,z)}$, where we have set $A(y,z) := \pi_{YX}^{-1}(y) \cap \pi_{ZX}^{-1}(z)$.

Set W equal to $ZY := Z \wedge Y$, the meet in the poset **S**. Since in **S** we have the commutative diagram

$$\begin{tikzcd} & X \arrow[dl] \arrow[d] \arrow[dr] & \\ Y & W \arrow[l] \arrow[r] & Z \end{tikzcd}$$

we obtain the following commutative diagram of sets

$$\begin{tikzcd} & E(X) \arrow[dl, "\pi_{YX}"'] \arrow[d, "\pi_{WX}"] \arrow[dr, "\pi_{ZX}"] & \\ E(Y) & E(W) \arrow[l, "\pi_{YW}"] \arrow[r, "\pi_{ZW}"'] \arrow[d, hook, "\iota"] & E(Z) \\ & E(X) \times E(Y) \arrow[ul, "\pi_1"] \arrow[ur, "\pi_2"'] & \end{tikzcd}$$

where the upper triangle is explained by the functoriality of E and the lower one by the universal property of products in **Sets**; ι is an injection by definition of an information structure.

This implies that $A(y,z) := \pi_{YX}^{-1}(y) \cap \pi_{ZX}^{-1}(z) = \pi_{WX}^{-1}\iota^{-1}(\pi_1^{-1}(y) \cap \pi_2^{-1}(x))$. If $(y,z) \notin \operatorname{im}\iota$, $A(y,z)$ is empty, so $\nu|_{A(y,z)} = 0$, as well as $\left\|\nu|_{A(y,z)}\right\| = Y_*\nu|_{Z=z}(y) = 0$. Therefore, the product in (4.6) can be restricted to pairs $(y,z) \in \operatorname{im}\iota$, and the condition $Y_*\nu|_{Z=z}(y) = \left\|\nu|_{A(y,z)}\right\| \neq 0$ translates into $W_*\nu(\iota^{-1}(y,z)) = \left\|\nu|_{A(y,z)}\right\| \neq 0$. Since there is a bijection $YZ \cong \operatorname{im}\iota$, upon relabeling we obtain the desired equality. □

To any arrow $\pi : X \to Y$, we associate the map $\mathscr{G}(\pi) : \mathscr{G}_Y \to \mathscr{G}_X$ such that $\mathscr{G}(\pi)(f) = f \circ \mathscr{C}(\pi)$. Then $\mathscr{G} : \mathbf{S} \to \mathbf{Sets}$ is a contravariant functor. In fact, it is a presheaf of $\mathscr{A}$-modules: Proposition 3.2 has an obvious analogue in this setting.

4.2 Description of cocycles

Following the considerations in Section 2.5 (and the notations introduced there), we study the differential complex $(C^n(\mathscr{G}),\delta)$, with $C^n(\mathscr{G}) := \mathrm{Hom}_{\mathscr{A}}(\mathscr{B}_n,\mathscr{G})$, of ***combinatorial n-cochains***. These cochains are jointly local and equivariant.

The coboundary of $f \in C^n(\mathscr{G})$ is the $(n+1)$-cochain $\delta f : \mathscr{B}_{n+1} \to \mathscr{G}$ defined on the generators of $\mathscr{B}_{n+1}$ by

$$\delta f[X_1|...|X_{n+1}] = \\ (X_1.f[X_2|...|X_{n+1}])\left(\prod_{k=1}^{n}(f[X_1|...|X_kX_{k+1}|...|X_n])^{(-1)^k}\right)f[X_1|...|X_n]^{(-1)^{n+1}} \quad (4.7)$$

An n-cocycle is an element f in $C^n(\mathbf{S},\mathscr{G})$ that verifies $\delta f = 1$; the submodule of all n-cocycles is denoted by $Z^n(\mathscr{G})$. The image under δ of C^{n-1} forms another submodule of $C^n(\mathscr{G})$, denoted $\delta C^{n-1}(\mathscr{G})$; its elements are called n-coboundaries. By definition, $\delta C^{-1}(\mathscr{G}) = 0$. The corresponding ***combinatorial information cohomology*** corresponds to

$$H^n(\mathbb{R}_{\mathbf{S}},\mathscr{G}) = Z^n(\mathscr{G})/\delta C^{n-1}(\mathscr{G}), \quad (4.8)$$

for every $n \geq 0$.

4.3 Computation of H^0

The 0-cochains are given by a collection of functions $\{f_X\}_{X\in \mathrm{Ob}\,\mathbf{S}}$ (the image of the section [] under f). Joint locality implies that, for every $X \in \mathrm{Ob}\,\mathbf{S}$, $f_X(\nu) = f_{\mathbf{1}}(\mathbf{1}_*\nu_X) = f_{\mathbf{1}}(\|\nu_X\|)$. Hence, 0-cochains are in one-to-one correspondence with measurable real-valued functions of the magnitude, $\tilde{f} := f_{\mathbf{1}} : \mathbb{N}^* \to \mathbb{R}_+$.

A 0-cocycle f must verify, for each Y coarser than X, $(\delta f)_X[Y] = (Y.f_X)(f_X)^{-1} = 1$, which taking into account the previous remarks reads

$$\tilde{f}(\|\nu_X\|) = \prod_{\substack{y\in Y\\ Y_*\nu(y)\neq 0}} \tilde{f}(\|\nu|_{Y=y}\|). \quad (4.9)$$

Whenever $|Y| \geq 2$, this means in particular that

$$\tilde{f}(x+y) = \tilde{f}(x)\tilde{f}(y) \quad (4.10)$$

for every $x, y \in \mathbb{N}$. Setting $a := f(1) > 0$, one easily concludes by recurrence that $\tilde{f}(n) = a^n = \exp(n\ln(a))$. The function $\tilde{f}(x) = \exp(kx)$, for arbitrary $k \in \mathbb{R}$, is a general solution of (4.9), because $\|\nu_X\| = \sum_{\substack{y\in Y\\ Y_*\nu(y)\neq 0}} \|\nu|_{Y=y}\|$. We have proved the following proposition.

Proposition 4.2. *Let* $\mathrm{Exp} \in \mathrm{Hom}_{\mathscr{A}}(*,\mathscr{G})$ *be the section defined by*

$$\mathrm{Exp}_X : \mathscr{C}_X \to \mathbb{R}_+^*, \nu \mapsto \exp(\|\nu\|). \quad (4.11)$$

Then,

$$H^0(\mathbf{S},\mathscr{G}) = \langle \mathrm{Exp}\rangle_{\mathbb{R}}. \quad (4.12)$$

4.4 Computation of H^1

For any 1-cochain f, we set $f[Z] := f_Z[Z] = f_X[Z]$, the last equality being valid for any X such that $X \to Z$, by joint locality.

In order to compute the 1-cocycle, we prove first an auxiliary result.

Lemma 4.3. *Let $f \in Z^1(\mathscr{G})$. For every $X \in \operatorname{Ob}\mathbf{S}$, if $\nu \in \mathscr{C}_X$ verifies $\nu = \nu|_{X=x}$ for some $x \in E_X$, then $f[X](\nu) = 1$.*

In particular, $f[\mathbf{1}] \equiv 1$.

Proof. The cocycle condition implies $f[XX] = (X.f[X])f[X]$, that reads

$$\prod_{\substack{x\in E_X\\ \nu(x)\neq 0}} f[X](\nu|_x) = 1. \tag{4.13}$$

□

The following result will be essential for the characterization of all the 1-cocycles. It is the combinatorial analogue of Proposition 3.10. Consequently, (4.14) and (4.16) should be seen as combinatorial generalizations of the fundamental functional equation of information theory.

Theorem 4.4 (Combinatorial FEITH). *Let $f_1, f_2 : \mathbb{N}\setminus\{(0,0)\} \to \mathbb{R}_+$ be two unknown functions. The functions f_1, f_2 satisfy the conditions*

1. *for $i \in \{1,2\}$, for every $n \in \mathbb{N}^*$, $f(n,0) = f(0,n) = 1$.*
2. *for every $\nu_0, \nu_1, \nu_2 \in \mathbb{N}$ such that $\nu_0 + \nu_1 + \nu_2 \neq 0$,*

$$f_1(\nu_0+\nu_2,\nu_1)f_2(\nu_0,\nu_2) = f_2(\nu_0+\nu_1,\nu_2)f_1(\nu_0,\nu_1). \tag{4.14}$$

if, and only if, there exists a sequence of numbers $D = \{D_i\}_{i\geq 1} \subset \mathbb{R}_+$, such that $D_1 = 1$, and

$$f(\nu_1,\nu_2) = \frac{[\nu_1+\nu_2]_D!}{[\nu_1]_D![\nu_2]_D!}, \tag{4.15}$$

where $[n]_D! = D_n D_{n-1}\cdots D_1$ whenever $n > 0$, and $[0]_D! = 1$.

Proof. Setting $\nu_0 = 0$, we conclude first that $f_1(\nu_2,\nu_1) = f_2(\nu_1,\nu_2)$. Define $f(x,y) := f_1(x,y)$; it satisfies the equation

$$\frac{f(\nu_0+\nu_1,\nu_2)}{f(\nu_0,\nu_2)} = \frac{f(\nu_1,\nu_0+\nu_2)}{f(\nu_1,\nu_0)}. \tag{4.16}$$

for any $\nu_0,\nu_1,\nu_2 \in \mathbb{N}$ such that $\nu_0+\nu_1+\nu_2 \neq 0$. In particular, if $\nu_0 = t > 0$, and $\nu_1 = \nu_2 = s > 0$,

$$\frac{f(t+s,s)}{f(s,t+s)} = \frac{f(t,s)}{f(s,t)}. \tag{4.17}$$

Thus, for any $n > 1$,

$$\frac{f(n,1)}{f(1,n)} = \frac{f(n-1,1)}{f(1,n-1)} = \cdots = \frac{f(1,1)}{f(1,1)} = 1. \tag{4.18}$$

Let D_{n+1} be the common value of $f(n,1)$ and $f(1,n)$. From Equation (4.16), setting $\nu_0 = n$, $\nu_1 = 1$, and $\nu_2 = k$, we can obtain a recurrence formula for $f(n+1,k)$:

$$f(n+1,k) = \frac{D_{n+k+1}}{D_{n+1}} f(n,k). \tag{4.19}$$

By repeated application of this recurrence, we conclude that

$$f(n,k) = \frac{D_{n+k}}{D_n} \cdot \frac{D_{n+k-1}}{D_{n-1}} \cdots \frac{D_{k+1}}{D_1} f(0,k). \tag{4.20}$$

Remark that $D_1 = f(0,1) = 1$, and $f(0,k) = 1$ (Lemma 4.3). Therefore, f can be rewritten as

$$f(\nu_1,\nu_2) = \frac{[\nu_1+\nu_2]_D!}{[\nu_1]_D![\nu_2]_D!}. \tag{4.21}$$

This formula still make sense when $\nu_1 = 0$ or $\nu_2 = 0$. Conversely, for any sequence $D = \{D_i\}_{i\geq 1}$, with $D_1 = 1$, the assignment $f_1 = f_2 = f$ satisfies (4.16), and thus represents the most general solution. □

The quotients

$$\begin{Bmatrix} \nu_1+\nu_2 \\ \nu_1,\nu_2 \end{Bmatrix}_D := \frac{[\nu_1+\nu_2]_D!}{[\nu_1]_D![\nu_2]_D!} \tag{4.22}$$

were studied in detail by H. G. Gould [33], who called them ***Fontené-Ward binomial coefficients***. To our knowledge, three particular cases appear in the literature under their own name:

1. $D_n = n$ gives the usual binomial coefficients: $\left\{ {\nu_1+\nu_2 \atop \nu_1,\nu_2} \right\}_D = \binom{\nu_1+\nu_2}{\nu_1,\nu_2}$.
2. $D_n = \frac{q^n-1}{q-1}$ gives the q-binomial coefficients, also known as Gaussian binomial coefficients: $\left\{ {\nu_1+\nu_2 \atop \nu_1,\nu_2} \right\}_D = \left[{\nu_1+\nu_2 \atop \nu_1,\nu_2} \right]_q$. For more details, see Section 6.1.
3. When D is the Fibonacci sequence, the expressions $\left\{ {\nu_1+\nu_2 \atop \nu_1,\nu_2} \right\}_D$ are called Fibonomial coefficients.

Already Fontené [29] noted that $\left\{ {n \atop k} \right\}_D := \left\{ {n \atop k,n-k} \right\}_D$ verifies the *additive* recurrence formula

$$\begin{Bmatrix} n \\ k \end{Bmatrix}_D - \begin{Bmatrix} n-1 \\ k \end{Bmatrix}_D = \begin{Bmatrix} n-1 \\ k-1 \end{Bmatrix}_D \frac{D_n - D_{n-k}}{D_k}, \tag{4.23}$$

with boundary conditions $\left\{ {n \atop 0} \right\}_D = \left\{ {n \atop n} \right\}_D = 1$ for $n \geq 0$. Conversely, this recurrence formula implies that $\left\{ {n \atop k} \right\}_D$ must be given by the expression in (4.22), so also the *multiplicative* relations (4.16).

Example 4.5. Let $(\mathbf{S}, E)$ be an information structure defined as follows: the poset $\mathbf{S}$ is represented by

$$\begin{tikzcd}
 & \mathbf{1} & \\
X_1 \arrow[ur] & & X_2 \arrow[ul] \\
 & X_1X_2 \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

and E is the functor defined at the level of objects by $E(X_1) = \{x_{\{1\}}, x_{\{0,2\}}\}$, $E(X_2) = \{x_{\{2\}}, x_{\{0,1\}}\}$, and $E(X_1X_2) = \{x_{\{1\}}, x_{\{2\}}, x_{\{3\}}\}$; for each arrow $\pi : X \to Y$, the map $\pi_* : E(X) \to E(Y)$ sends $x_I \to x_J$ iff $I \subset J$.

For this structure, the cocycle condition give the equations

$$f[X_1X_2](\nu_0,\nu_1,\nu_2) = f[X_2](\nu_0+\nu_1,\nu_2)f[X_1](\nu_0,\nu_1)f[X_1](\nu_2,0), \tag{4.24}$$
$$f[X_2X_1](\nu_0,\nu_1,\nu_2) = f[X_1](\nu_0+\nu_2,\nu_1)f[X_2](\nu_0,\nu_2)f[X_2](\nu_1,0). \tag{4.25}$$

Since $X = X_1X_2 = X_2X_1$,

$$f[X_2](\nu_0+\nu_1,\nu_2)f[X_1](\nu_0,\nu_1) = f[X_1](\nu_0+\nu_2,\nu_1)f[X_2](\nu_0,\nu_2) \tag{4.26}$$

where we have taken into account that $f[X_1](\nu_2,0) = f[X_2](\nu_1,0) = 0$. This is exactly Equation (4.14), and the condition (1) in the statement is also met, therefore

$$f[X_1](\nu_0,\nu_1) = f[X_2](\nu_0,\nu_1) = \begin{Bmatrix} \nu_0+\nu_1 \\ \nu_0,\nu_1 \end{Bmatrix}_D \tag{4.27}$$

for some sequence D. From (4.24), we conclude that

$$f[X](\nu_0,\nu_1,\nu_2) = \begin{Bmatrix} \nu_0+\nu_1+\nu_2 \\ \nu_0,\nu_1,\nu_2 \end{Bmatrix}_D := \frac{[\nu_0+\nu_1+\nu_2]_D!}{[\nu_0]_D![\nu_1]_D![\nu_2]_D!}. \tag{4.28}$$

Definition 4.6. Given any sequence $D = \{D_i\}_{i\geq 1}$ verifying $D_1 = 1$ (called ***admissible sequence***), the corresponding ***Fontené-Ward multinomial coefficient*** is the 1-cochain given by

$$\forall \nu \in \mathscr{C}(X), \quad W_D[X](\nu) = \frac{[\|\nu\|]_D!}{\prod_{x\in E_X}[\nu(x)]_D!}. \tag{4.29}$$

To characterize the cocycles associated to general products XY, we introduce a definition of nondegeneracy analogous to Definition 3.12.

Definition 4.7. Given two partitions X and Y, such that $|E_X| = k$ and $|E_Y| = l$, we call its product XY ***nondegenerate*** if there exist enumerations $\{x_1,\dots,x_k\}$ of E_X and $\{y_1,\dots,y_l\}$ of E_Y, and a North-East (NE) lattice path[3] $(\gamma_i)_{i=1}^m$ on $\mathbb{Z}^2$ going from $(1,1)$ to (k,l) such that

1. For each $\gamma_i = (a,b)$, the set
$$\iota^{-1}\{\,(x_i,y_j)\,|\,a\leq i\leq a+1 \text{ and } b\leq j\leq b+1\,\}$$
contains at least three different elements. Here ι denotes the injection $E_{XY} \hookrightarrow E_X\times E_Y$.
2. If $\gamma_i = (a,b)$ and $\gamma_{i+1}-\gamma_i = (1,0)$, we ask that for every counting function $\nu \in \mathscr{C}_X$ such that $\operatorname{supp}\nu \subset \{\,x_i\,|\,a\leq i\leq k\,\}$, there exists a counting function $\tilde\nu \in \mathscr{C}_{XY}$ whose support is contained in
$$\iota^{-1}(\{(x_a,y_{b+1})\}\cup\{\,(x_i,y_b)\,|\,a+1\leq i\leq k\,\})\cup\iota^{-1}(\{(x_a,y_b)\}\cup\{\,(x_i,y_{b+1})\,|\,a+1\leq i\leq k\,\})$$
and such that $\nu = X_*\tilde\nu$.

[3] A North-East (NE) lattice path on $\mathbb{Z}^2$ is a sequence of points $(\gamma_i)_{i=1}^m \subset \mathbb{Z}^2$ such that $\gamma_{i+1}-\gamma_i \in \{(1,0),(0,1)\}$ for every $i\in\{1,\dots,m-1\}$.

Analogously, if $\gamma_{i+1} - \gamma_i = (0,1)$, we ask that every counting function $\nu \in \mathscr{C}_Y$ such that $\operatorname{supp}\nu \subset \{\, y_i \mid b \le i \le l \,\}$, there exists a counting function $\tilde{\nu} \in \mathscr{C}_{XY}$ whose support is contained in

$$\iota^{-1}(\{(x_{a+1}, y_b)\} \cup \{\, (x_a, y_j) \mid b+1 \le j \le l \,\}) \cup [\iota^{-1}(\{(x_a, y_b)\} \cup \{\, (x_{a+1}, y_j) \mid b+1 \le j \le k \,\})$$

and such that $\nu = Y_* \tilde{\nu}$.

Proposition 4.8. *Let* $(\mathbf{S}, E)$ *be an information structure and* X, Y *two different variables in* $\operatorname{Ob}\mathbf{S}$ *such that* $XY \in \operatorname{Ob}\mathbf{S}$. *Let* f *be a combinatorial* 1*-cocycle i.e. an element of* $Z^1(\mathbf{S}, \mathscr{G})$. *If* XY *is nondegenerate, there exists an admissible sequence* D, *such that*

$$f[X] = W_D[X], \quad f[Y] = W_D[Y], \quad f[XY] = W_D[XY].$$

Proof. Very similar to the proof of Proposition 3.13. As f is a 1-cocycle, it satisfies the two equations derived from (4.7)

$$Y.f[X]f[Y] = f[XY], \tag{4.30}$$
$$X.f[Y]f[X] = f[XY]. \tag{4.31}$$

and therefore the symmetric equation

$$(X.f[Y])f[Y] = (Y.f[X])f[X]. \tag{4.32}$$

For any counting function ν, we write

$$\begin{pmatrix} s & t & u & \dots \\ p & q & r & \dots \end{pmatrix}$$

if $\nu(s) = p$, $\nu(t) = q$, $\nu(u) = r$, etc. and the images of the unwritten parts are zero.

Fix an order $(x_1, ..., x_k)$ and $(y_1, ..., y_l)$ that satisfies the definition of nondegenerate product, and let $\{\gamma_i\}_{i=0}^m$ be the corresponding path. If $\gamma_i = (a, b)$ and $\gamma_{i+1} - \gamma_i = (1, 0)$, we are going to show that the following recursive formula holds:

$$f[X]\begin{pmatrix} x_a & \dots & x_k \\ \mu_a & \dots & \mu_k \end{pmatrix} = f[X]\begin{pmatrix} x_{a+1} & \dots & x_k \\ \mu_{a+1} & \dots & \mu_k \end{pmatrix} f[X]\begin{pmatrix} x_a & x_{a+1} \\ \mu_a & \|\mu\| - \mu_a \end{pmatrix}. \tag{4.33}$$

Analogously, if $\gamma_i = (a, b)$ and $\gamma_{i+1} - \gamma_i = (0, 1)$,

$$f[Y]\begin{pmatrix} y_b & \dots & y_l \\ \nu_b & \dots & \nu_l \end{pmatrix} = f[Y]\begin{pmatrix} y_{b+1} & \dots & y_l \\ \nu_{b+1} & \dots & \nu_l \end{pmatrix} f[Y]\begin{pmatrix} y_b & y_{b+1} \\ \nu_b & \|\nu\| - \nu_b \end{pmatrix}. \tag{4.34}$$

Suppose that $\gamma_i = (a, b)$ and $\gamma_{i+1} - \gamma_i = (1, 0)$. Let

$$\mu = \begin{pmatrix} x_a & \dots & x_k \\ \mu_a & \dots & \mu_k \end{pmatrix}$$

be a counting function in $\mathscr{C}_X$. By Definition 4.7-2 above, we know that μ has a preimage under marginalization $\tilde{\mu}$, whose support is such that $(X.f[Y])(\tilde{\mu}) = 1$, cf.

Lemma 4.3. Thus (4.31) becomes $f[XY](\tilde{\mu}) = f[X](X_*\tilde{\mu}) = f[X](\mu)$. Equation (4.30) then reads

$$f[X]\begin{pmatrix} x_{a+1} & \dots & x_k \\ \mu_{a+1} & \dots & \mu_k \end{pmatrix} f[Y]\circ\tau\begin{pmatrix} y_b & y_{b+1} \\ \|\mu\| - \mu_a & \mu_a \end{pmatrix} = f[X]\begin{pmatrix} x_a & \dots & x_k \\ \mu_a & \dots & \mu_k \end{pmatrix}, \quad (4.35)$$

where τ is the identity or the transposition of the nontrivial arguments of μ. In any case, setting $\mu_{a+1} = \|\mu\| - \mu_a$ and $\mu_{a+2} = \dots = \mu_k = 0$, we conclude that

$$f[X]\begin{pmatrix} x_a & x_{a+1} \\ \mu_a & \|\mu\| - \mu_a \end{pmatrix} = f[Y]\circ\tau\begin{pmatrix} y_b & y_{b+1} \\ \|\mu\| - \mu_a & \mu_a \end{pmatrix}, \quad (4.36)$$

which combined with (4.35) implies (4.33). The identity (4.34) can be obtained analogously.

To determine

$$\phi_a(n_1, n_2) := f[X]\begin{pmatrix} x_a & x_{a+1} \\ n_1 & n_2 \end{pmatrix} \quad \text{and} \quad \psi_b(n_1, n_2) := f[Y]\begin{pmatrix} y_b & y_{b+1} \\ n_1 & n_2 \end{pmatrix},$$

for $(n_1, n_2) \in \mathbb{N}^2$, consider the three different elements b_1, b_2, b_3 in $E_{XY} \subset E_X \times E_Y$ given by the property 1 of a nondegenerate product. The symmetric equation (4.32) evaluated on $\nu_1\delta_{b_1} + \nu_2\delta_{b_2} + \nu_3\delta_{b_3} \in \mathbb{N}$ gives the equation that appears in Proposition 4.4, which implies that $\phi_a(n_1, n_2) = \psi_b(n_1, n_2) = \left\{ {n_1+n_2 \atop n_1, n_2} \right\}_D$ for certain admissible sequence D (the eventual permutations of the arguments in the unknowns become irrelevant, because the solution is symmetric).

When considering γ_{i+1}, one finds the functions ϕ_a and ψ_{b+1}, or the functions ϕ_{a+1} and ψ_b, since two consecutive matrices $\gamma_{i+1} - \gamma_i$ is either $(1, 0)$ or $(0, 1)$. This ensures that the admissible sequence D that appears for each γ_i is always the same, as proved in Lemma 4.9. The recurrence relations (4.33) and (4.34) then imply the desired result. ∎

Lemma 4.9. *Let D, D' be two admissible sequences. If for all $n_1, n_2 \in \mathbb{N}^2$*

$$\left\{ {n_1 + n_2 \atop n_1, n_2} \right\}_D = \left\{ {n_1 + n_2 \atop n_1, n_2} \right\}_{D'} \quad (4.37)$$

then $D = D'$.

Proof. Just remark that

$$\left\{ {n \atop 1} \right\}_D = \left\{ {n \atop 1, n-1} \right\}_D = [n]_D! \quad (4.38)$$

so we have $[n]_D! = [n]_{D'}!$ for all $n \in \mathbb{N}$, which clearly implies the result. ∎

As in the continuous case, the number of admissible sequences that appear in the computation of the 1-cocycles $Z^1(\mathbf{S}, \mathscr{G})$ depends on the number of connected components of $\mathbf{S}^*$, that is $\mathbf{S}$ deprived of its final element: only the minimal elements are important, and when they refine a common variable (different from $\mathbf{1}$) they must share the admissible sequence D, as a consequence of Lemma 4.9. Cf. the proof of Theorem 3.14.

On the other hand, a choice of 0-cochain $g = g_{\mathbf{1}}[\,] : \mathbb{N}^* \to \mathbb{R}^*_+$ induces globally a Fontené-Ward coefficient δg for a unique admissible sequence D_g. Explicitly,

$$\delta g[Y](\nu) = \frac{(Y.g[])(\nu)}{g[](\nu)} = \frac{\prod_{y\in E_Y, Y_*\nu(y)\neq 0} g(\|\nu|_y\|)}{g(\|\nu\|)}. \tag{4.39}$$

The coboundary δg is clearly trivial when g is the exponential function. In virtue of the previous theorem, the expression (4.39), being a cocycle, must be a Fontené-Ward coefficient, which gives the existence of D_g.

Therefore, Z^1 and δC^0 are both infinite dimensional. If $\mathbf{S}^*$ is connected, the quotient is trivial; otherwise it is infinite: $|\pi_0(\mathbf{S}^*)| - 1$ admissible sequences remain arbitrary.

4.5 Asymptotic relation with probabilistic information cohomology

Proposition 4.10. *Let g be a combinatorial 1-cocycle. Suppose that, for every $X \in \operatorname{Ob}\mathbf{S}$, there exists a measurable function $f[X] : \Pi(X) \to \mathbb{R}$ with the following property: for every sequence of counting functions $\{\nu_n\}_{n\geq 1} \subset \mathscr{C}_X$ such that*

1. *$\|\nu_n\| \to \infty$, and*
2. *for every $x \in E_X$, $\nu_n(x)/\|\nu_n\| \to p(x)$ as $n \to \infty$*

the asymptotic formula

$$g[X](\nu_n) = \exp(\|\nu_n\|^\alpha f[X](p) + o(\|\nu_n\|^\alpha))$$

holds. Then f is a probabilistic 1-cocycle of type α, i.e. $f \in Z^1(\mathbf{S}, \mathscr{F}_\alpha(\Pi))$.

Proof. We must show that, whenever it makes sense,

$$f[XY](P_{XY}) = (X.f[Y])(P_{XY}) + f[X](X_*P_{XY}).$$

Let $\{\nu^n_{XY}\}_{n\geq 1}$ be a sequence of counting functions such that $\|\nu^n_{XY}\| \to \infty$ and, for every $z \in E_{XY}$, $\nu^n_{XY}(z)/\|\nu^n_{XY}\| \to P_{XY}(z)$. A sequence like this always exists: just consider a rational approximation of the values of P_{XY} with common denominator.

Since g is a 1-cocycle, $g[XY] = (X.g[Y])g[X]$. Evaluate it at ν^n_{XY}, take the logarithm and divide by $\|\nu^n_{XY}\|^\alpha$ in order to obtain

$$\frac{\ln g[XY](\nu^n_{XY})}{\|\nu^n_{XY}\|^\alpha} = \sum_{\substack{x\in E_X\\ X_*\nu^n_{XY}(x)\neq 0}} \frac{\ln g[Y](\nu^n_{XY}|_{X=x})}{\|\nu^n_{XY}\|^\alpha} + \frac{\ln g[X](\nu^n_{XY})}{\|\nu^n_{XY}\|^\alpha}. \tag{4.40}$$

Recall that, for any counting function ν, $\|\nu|_{X=x}\| = X_*\nu(x)$. Hence,

$$\frac{\ln g[Y](\nu^n_{XY}|_{X=x})}{\|n^n_{XY}\|^\alpha} = \frac{\ln g[Y](\nu^n_{XY}|_{X=x})}{\|\nu^n_{XY}|_{X=x}\|^\alpha}\frac{(X_*\nu^n_{XY}(x))^\alpha}{\|n^n_{XY}\|^\alpha}. \tag{4.41}$$

Plug this in (4.40) and take the limit as n goes to infinity to conclude. □

The proof applies almost unchanged to 0-cocycles and general n-cocycles.

Proposition 4.11. *Let g be a combinatorial n-cocycle. Suppose that, for every $X_1, \dots, X_n \in$ $\operatorname{Ob}\mathbf{S}$ such that $X_i \cdots X_n \in \operatorname{Ob}\mathbf{S}$, there exists a measurable function*

$$f[X_1|\dots|X_n] : \Pi(X_1 \cdots, X_n) \to \mathbb{R}$$

with the following property: for every sequence of counting functions $\{\nu_n\}_{n\geq 1} \subset \mathscr{C}_{X_1 \cdots X_n}$ such that

1. *$\|\nu_n\| \to \infty$, and*
2. *for every $z \in E_{X_1 \cdots X_n}$, $\nu_n(z)/\|\nu_n\| \to p(z)$ as $n \to \infty$*

the asymptotic formula

$$g[X_1|\dots|X_n](\nu_n) = \exp(\|\nu_n\|^{\alpha} f[X_1|\dots|X_n](p) + o(\|\nu_n\|^{\alpha}))$$

holds. Then f is a n-cocycle of type α, i.e. $f \in Z^n(\mathbf{S}, \mathscr{F}_{\alpha}(\Pi))$.

We discuss now some important examples:

1. The exponential $\operatorname{Exp}^k : \nu \to \exp(k\,\|\nu\|)$ is the a combinatorial 0-cocycle, and it corresponds to the constant k seen as a probabilistic 0-cocycle.
2. It is well known that
$$\binom{n}{p_1 n, \dots, p_s n} = \exp(nS_1(p_1, \dots, p_s) + o(n)) \tag{4.42}$$
This can be easily proved by using Stirling's approximation, for instance. This asymptotic formula partly explains the relevance of entropy in Shannon's communication theory: it gives an asymptotic counting of typical sequences for memoryless sources. This is the content of the Asymptotic Equipartition Property (Proposition 0.4).
3. Whereas the previous examples are not really surprising, Proposition 4.10 hints at new objects that are connected to the generalized α-entropies and have gone unnoticed until now. For example, the q-multinomial coefficients are connected asymptotically to the 2-entropy (quadratic entropy),
$$\begin{bmatrix} n \\ p_1 n, \dots, p_s n \end{bmatrix}_q = \exp(n^2 \frac{\ln q}{2} S_2(p_1, \dots, p_s) + o(n^2)) \tag{4.43}$$
These coefficients have a well-known combinatorial interpretation: when q is a prime power and $k_1, \dots, k_s$ are integers such that $\sum_{i=1}^{s} k_i = n$, the coefficient $\begin{bmatrix} n \\ k_1, \dots, k_s \end{bmatrix}_q$ counts the number of flags of vector spaces $V_1 \subset V_2 \subset \dots \subset V_n = \mathbb{F}_q^n$ such that $\dim V_i = \sum_{j=1}^{i} k_j$ (here $\mathbb{F}_q$ denotes the finite field of order q). In particular, the q-binomial coefficient $\begin{bmatrix} n \\ k \end{bmatrix}_q \equiv \begin{bmatrix} n \\ k, n-k \end{bmatrix}_q$ counts vector subspaces of dimension k in $\mathbb{F}_q^n$.
In Part III, we push this parallel between S_1 and S_2 much further: we introduce a probabilistic model that generates vector spaces and study its concentration properties, to obtain a generalization of the Asymptotic Equipartition Property that involves the quadratic entropy (Theorem 8.2).

It is quite natural to ask if, for any $\alpha > 0$, there exists a sequence $D^{\alpha} = \{D_i^{\alpha}\}_{i\geq 1}$ and $K \in \mathbb{R}$ such that

$$\begin{Bmatrix} n \\ p_1 n, \dots, p_s n \end{Bmatrix}_{D^{\alpha}} = \exp(n^{\alpha} K S_{\alpha}(p_1, \dots, p_s) + o(n^{\alpha})), \tag{4.44}$$

and the answer turns out to be yes.

Proposition 4.12. *Consider any $\alpha \in \mathbb{R}_+^* \setminus \{1\}$. If $D_n^\alpha = \exp\{K(n^{\alpha-1}-1)\}$, for any $K \in \mathbb{R}$, then*

$$\begin{Bmatrix} n \\ p_1 n, \dots, p_s n \end{Bmatrix}_{D^\alpha} = \exp\left\{ n^\alpha \frac{K}{\alpha} S_\alpha(p_1, \dots, p_s) + o(n^\alpha) \right\}.$$

Proof. Remark that $[n]_D! := \exp\{K(\sum_{i=1}^n i^{\alpha-1} - n)\}$.

Suppose first that $\alpha > 1$. In this case, $x \mapsto x^{\alpha-1}$ is strictly increasing and

$$\int_0^n x^{\alpha-1}\,\mathrm{d}x = \frac{n^\alpha}{\alpha} < \sum_{i=1}^n i^{\alpha-1} < \int_1^{n+1} x^{\alpha-1}\,\mathrm{d}x = \frac{(n+1)}{\alpha} - \frac{1}{\alpha}. \tag{4.45}$$

Hence, if $K > 0$,

$$\exp\left\{K\left(\frac{n^\alpha}{\alpha} - n\right)\right\} < [n]_D! < \exp\left\{K\left(\frac{(n+1)}{\alpha} - \frac{1}{\alpha} - n\right)\right\}. \tag{4.46}$$

This directly implies that

$$\begin{aligned} \frac{[n]_D!}{[n_1]_D! \cdots [n_s]_D!} &< \exp\left\{K\left(\frac{(n+1)^\alpha}{\alpha} - \frac{1}{\alpha} - n - \sum_{i=1}^s \left(\frac{n_i^\alpha}{\alpha} - n_i\right)\right)\right\} \\ &< \exp\left\{\frac{K}{\alpha}\left(n^\alpha\left(1 - \sum_{i=1}^s \frac{n_i^\alpha}{n^\alpha}\right) + o(n^\alpha)\right)\right\}, \end{aligned} \tag{4.47}$$

as well as

$$\begin{aligned} \frac{[n]_D!}{[n_1]_D! \cdots [n_s]_D!} &> \exp\left\{K\left(\frac{n^\alpha}{\alpha} - n - \sum_{i=1}^s \left(\frac{(n_i+1)^\alpha}{\alpha} - n_i\right)\right)\right\} \\ &> \exp\left\{\frac{K}{\alpha}\left(n^\alpha\left(1 - \sum_{i=1}^s \frac{n_i^\alpha}{n^\alpha}\right) + o(n^\alpha)\right)\right\}. \end{aligned} \tag{4.48}$$

If $K < 0$, the inequalities (4.46), (4.47) and (4.48) must be reversed, but the result is the same. Similarly, when $0 < \alpha < 1$ the argument remains valid making the necessary modifications: all inequalities are reversed, since $x \mapsto x^{\alpha-1}$ is strictly decreasing. □

It is not known if these or similar coefficients related to S_α, for $\alpha \in \mathbb{R}_+^* \setminus \{1, 2\}$, have a combinatorial or statistical interpretation.

Chapter 5

A functional equation for generalized entropies related to the modular group

Fix $\alpha > 0$. We are interested in the measurable solutions of

$$u(1-x) + (1-x)^{\alpha} u\left(\frac{y}{1-x}\right) = u(y) + (1-y)^{\alpha} u\left(\frac{1-x-y}{1-y}\right). \tag{5.1}$$

for all $x, y \in [0,1)$ such that $x + y \in [0,1]$, subject to the boundary condition $u(0) = u(1) = 0$.

In this chapter, we show that the only measurable solutions to (5.1) are multiples of the corresponding entropy s_{α}. This rests on two preliminary results.

Proposition 5.1 (Regularity). *Any measurable solution of* (5.1) *is infinitely differentiable on* $(0,1)$.

Proposition 5.2 (Symmetry). *Any solution of* (5.1) *satisfies* $u(x) = u(1-x)$ *for all* $x \in \mathbb{Q} \cap [0,1]$.

The first is proved analytically, by means of standard techniques in the field of functional equations, and the second by a geometrical argument, relating the equation to the action of the modular group on the projective line.

The propositions above imply that any measurable solution of (5.1) must satisfy $u(x) = u(1-x)$ for all $x \in [0,1]$ and therefore

$$u(x) + (1-x)^{\alpha} u\left(\frac{y}{1-x}\right) = u(y) + (1-y)^{\alpha} u\left(\frac{x}{1-y}\right), \tag{5.2}$$

with $u(1) = u(0) = 0$. By continuity, u attains a finite value on $\frac{1}{2}$, say K. For $\alpha = 1$, Kannappan and Ng [47] showed that $u(x) = Ks_1(x)$. For $\alpha \neq 1$, Daróczy [25] proved that[1]

$$u(x) = \frac{K}{2^{1-\alpha} - 1}(x^{\alpha} + (1-x)^{\alpha} - 1). \tag{5.3}$$

[1] In fact, he does the case $K = 1$, but the argument works in general.

Proof of Proposition 5.1. Lemma 3 in [47] implies that u is locally bounded on $(0,1)$ and hence locally integrable. Their proof is for $\alpha = 1$, but the argument applies to the general case with almost no modification, just replacing

$$|u(y)| = \left| u(1-x) + (1-x)u\left(\frac{y}{1-x}\right) - (1-y)u\left(\frac{1-x-y}{1-y}\right)\right| \leq 3N,$$

where x, y are such that $u(1-x) \leq N$, $u\left(\frac{y}{1-x}\right) \leq N$ and $u\left(\frac{1-x-y}{1-y}\right) \leq N$, by

$$|u(y)| = \left| u(1-x) + (1-x)^{\alpha}u\left(\frac{y}{1-x}\right) - (1-y)^{\alpha}u\left(\frac{1-x-y}{1-y}\right)\right| \leq 3N,$$

that is evidently valid too.

To prove the differentiability, we also follow the method of [47]. Let us fix an arbitrary $y_0 \in (0,1)$; then, it is possible to chose $s, t \in (0,1)$, $s < t$, such that

$$\frac{1-y-s}{1-y}, \frac{1-y-t}{1-y} \in (0,1),$$

for all y in certain neighborhood of y_0. We integrate (5.1) with respect to x, between s and t, to obtain

$$(s-t)u(y) = \int_{1-t}^{1-s} u(x)\,\mathrm{d}x + y^{1+\alpha}\int_{\frac{y}{1-s}}^{\frac{y}{1-t}} \frac{u(z)}{z^3}\,\mathrm{d}z + (1-y)^{1+\alpha}\int_{\frac{1-y-s}{1-y}}^{\frac{1-y-t}{1-y}} u(z)\,\mathrm{d}z. \quad (5.4)$$

The continuity of the RHS of (5.4) as a function of y at y_0, implies that u is continuous at y_0 and therefore on $(0,1)$. The continuity of u in the RHS of (5.4) implies that u is differentiable at y_0. An iterated application of this argument shows that u is infinitely differentiable on $(0,1)$. □

Proof of Proposition 5.2. We take $1-x = 1-y = z \in \left[\frac{1}{2}, 1\right]$ in (5.1), to obtain

$$u(z) - u(1-z) = z^{\alpha}\left[u(2-z^{-1}) - u(z^{-1}-1)\right].$$

If we define $h(z) := u(z) - u(1-z)$, the previous equation reads

$$\forall z \in \left[\frac{1}{2}, 1\right], \quad h(z) = z^{\alpha}h(2-z^{-1}), \quad (5.5)$$

and the definition directly implies that

$$\forall z \in [0,1], \quad h(z) = -h(1-z). \quad (5.6)$$

The boundary conditions are $h(0) = h(1) = 0$. From (5.5), we deduce that $h(1/2) = h(0)/2^{\alpha} = 0$. Using (5.6) to modify the right hand side of (5.5), we obtain

$$\forall x \in \left[\frac{1}{2}, 1\right], \quad h(x) = -x^{\alpha}h(x^{-1}-1). \quad (5.7)$$

In principle h is just defined on $[0,1]$, but we extend it imposing periodicity:

$$\forall x \in]-\infty, \infty[, \quad h(x+1) = h(x) \quad (5.8)$$

We establish now several results about this extended function.

Lemma 5.3.
$$\forall x \in \mathbb{R}, \quad h(x) = -h(1-x).$$

Proof. We write $x = [x] + \{x\}$, where $\{x\} := x - [x]$. Then,

$$h(x) \overset{(5.8)}{=} h(\{x\}) \overset{(5.6)}{=} -h(1-\{x\}) \overset{(5.8)}{=} -h(1-\{x\}-[x]) = -h(1-x).$$

□

Lemma 5.4.
$$\forall x \in [1,2], \quad h(x) = x^\alpha h(2-x^{-1}). \tag{5.9}$$

Proof. For h is periodic, (5.9) is equivalent to $\forall x \in [1,2]$, $h(x-1) = x^\alpha h(1-x^{-1})$, and the change of variables $u = x-1$ gives

$$\forall u \in [0,1], \quad h(u) = (u+1)^\alpha h\left(\frac{u}{u+1}\right). \tag{5.10}$$

Note that $1 - \frac{u}{u+1} = \frac{1}{u+1} \in [1/2, 1]$ whenever $u \in [0,1]$. Therefore,

$$h\left(\frac{u}{u+1}\right) \overset{\text{(Lemma 5.3)}}{=} -h\left(\frac{1}{u+1}\right) \overset{(5.7)}{=} \left(\frac{1}{u+1}\right)^\alpha h(u).$$

This establishes (5.10). □

Lemma 5.5.
$$\forall x \in [2,\infty[, \quad h(x) = x^\alpha h(2-x^{-1}). \tag{5.11}$$

Proof. If $x \in [2,\infty[$, then $1-\frac{1}{x} \in \left[\frac{1}{2}, 1\right]$ and we can apply equation (5.5) to obtain

$$h\left(1-\frac{1}{x}\right) \overset{(5.5)}{=} \left(1-\frac{1}{x}\right)^\alpha h\left(2-\left(1-\frac{1}{x}\right)^{-1}\right) = \left(\frac{x-1}{x}\right)^\alpha h\left(1-\frac{1}{x-1}\right). \tag{5.12}$$

We prove (5.11) by recurrence. The case $x \in [1,2]$ corresponds to Lemma 5.4. Suppose it is valid on $[n-1,n]$, for certain $n \geq 2$; for $x \in [n, n+1]$,

$$\begin{aligned} h(x) &\overset{(5.8)}{=} h(x-1) \overset{\text{(rec.)}}{=} (x-1)^\alpha h(2-(x-1)^{-1}) \overset{(5.8)}{=} (x-1)^\alpha h(1-(x-1)^{-1}) \\ &\overset{(5.12)}{=} x^\alpha h(1-x^{-1}) \overset{(5.8)}{=} x^\alpha h(1-x^{-1}). \end{aligned}$$

□

Lemma 5.6.
$$\forall x \in \left[0, \frac{1}{2}\right], \quad h(x) = -x^\alpha h(x^{-1}-1). \tag{5.13}$$

Proof. The previous lemma and periodicity imply that $h(x-1) = x^\alpha h(1-x^{-1})$ for all $x \geq 2$, i.e.

$$\forall u \geq 1, \quad h(u) = (u+1)^\alpha h\left(1 - \frac{1}{u+1}\right). \tag{5.14}$$

Then, for $u \geq 1$,

$$h\left(\frac{1}{u+1}\right) \overset{\text{(Lem. 5.3)}}{=} -h\left(1 - \frac{1}{u+1}\right) \overset{(5.14)}{=} -\left(\frac{1}{u+1}\right)^\alpha h(u). \tag{5.15}$$

We set $y = (u+1)^{-1} \in \left(0, \frac{1}{2}\right]$. Equation (5.15) reads

$$\forall y \in \left(0, \frac{1}{2}\right], \quad h(y) = -y^\alpha h(y^{-1} - 1). \tag{5.16}$$

Since $h(0) = 0$, the lemma is proved. □

Lemma 5.7.

$$\forall x \in \left[0, \frac{1}{2}\right], \quad h(x) = x^\alpha h(2 - x^{-1}). \tag{5.17}$$

Proof. By Lemma 5.3, $h\left(2 - x^{-1}\right) = -h(x^{-1} - 1)$. Thus,

$$\forall x \in \left[0, \frac{1}{2}\right], \quad h(x) \overset{(5.13)}{=} -x^\alpha h\left(\frac{1}{x} - 1\right) = x^\alpha h\left(2 - \frac{1}{x}\right).$$

□

Lemma 5.8.

$$\forall x \in]-\infty, 0], \quad h(x) = -x^\alpha h(2 - x^{-1}).$$

Proof. On the one hand, periodicity implies that $h(x) = h(x+1) \overset{\text{(Lem. 5.3)}}{=} -h(1-(x+1)) = -h(-x)$. On the other, for $x \leq 0$, the previous lemmas imply that $h(-x) = (-x)^\alpha h(2 - (-x)^{-1}) = |x|^\alpha h(2 - (-x)^{-1})$. Therefore,

$$h(x) = -h(-x) = -|x|^\alpha h\left(2 + \frac{1}{x}\right) \overset{\text{(Lem. 5.3)}}{=} |x|^\alpha h\left(1 - \left(2 + \frac{1}{x}\right)\right) \overset{(5.8)}{=} |x|^\alpha h\left(2 - \frac{1}{x}\right) \tag{5.18}$$

□

All these results can be summarized as follows:

Proposition 5.9. *The function h, extended periodically to $\mathbb{R}$, satisfies the equations*

$$\forall x \in \mathbb{R}, \quad h(x) = |x|^\alpha h\left(\frac{2x-1}{x}\right), \tag{5.19}$$

$$\forall x \in \mathbb{R}, \quad h(x) = -|x|^\alpha h\left(\frac{1-x}{x}\right). \tag{5.20}$$

Equation (5.20) is deduced from (5.19) using (5.6).

The group $G = SL_2(\mathbb{Z})/\{\pm I\}$ is called the ***modular group***; it is the image of $SL_2(\mathbb{Z})$ in $PGL_2(\mathbb{R})$. We keep using the matrix notation for the images in this quotient. We make G act on $P^1(\mathbb{R})$ as follows: an element $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G$ acting on $[x : y] \in P^1(\mathbb{R})$ (homogeneous coordinates) gives

$$g[x : y] = [ax + by : cx + dy].$$

Let S and T be the elements of G defined by the matrices

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{5.21}$$

The group G is generated by S and T [77, Ch. VII, Th. 2]; in fact, one can prove that $\langle S, T; S^2, (ST)^3 \rangle$ is a presentation of G.

The transformations $x \mapsto \frac{2x-1}{x}$ and $x \mapsto \frac{1-x}{x}$ in Equations (5.19) and (5.20) are homographies of the real projective line $P^1(\mathbb{R})$, that we denote respectively α and β. They correspond to elements

$$A = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix}. \tag{5.22}$$

in G, that satisfy

$$B^2 = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}, \quad BA^{-1} = \begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{5.23}$$

This last matrix corresponds to $x \mapsto 1 - x$.

Lemma 5.10. *The matrices A and B^2 generate G.*

Proof. Let

$$P = S^{-1}T^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}.$$

One has

$$PAP^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \tag{5.24}$$

and

$$PB^2P^{-1} = \begin{pmatrix} 3 & -1 \\ 1 & 0 \end{pmatrix}. \tag{5.25}$$

Therefore, $PAP^{-1} = T^{-1}$ and $S = T^{-3}PB^{-2}P^{-1}$. Inverting these relations, we obtain

$$T = PA^{-1}P^{-1}; \quad S = PA^3B^{-2}P^{-1}. \tag{5.26}$$

Let X be an arbitrary element of G. Since $Y = PXP^{-1} \in G$ and G is generated by S and T, the element Y is a word in S and T. In consequence, X is a word in $P^{-1}SP$ and $P^{-1}TP$, which in turn are words A and B^2. The Lemma is proved.

One can find explicit formulas for S and T in terms of A and B^2. Since $P = S^{-1}T^{-1}$, we deduce that $PSP^{-1} = S^{-1}T^{-1}STS$ and $PTP^{-1} = S^{-1}T^{-1}TTS = S^{-1}TS$. Hence, in virtue of (5.26),

$$\begin{aligned} S &= P^{-1}S^{-1}T^{-1}STSP \\ &= (P^{-1}S^{-1}P)(P^{-1}T^{-1}P)(P^{-1}SP)(P^{-1}TP)(P^{-1}SP) \\ &= B^2AB^{-2}A^2B^{-2} \end{aligned}$$

and

$$\begin{aligned} T &= P^{-1}S^{-1}TSP \\ &= (P^{-1}S^{-1}P)(P^{-1}TP)(P^{-1}SP) \\ &= B^2A^{-1}B^{-2}. \end{aligned}$$

□

To finish our proof of Proposition 5.2, we remark that the orbit of 0 by the action of G on $P^1(\mathbb{R})$ is $\mathbb{Q} \cup \{\infty\}$, where $\mathbb{Q} \cup \{\infty\}$ has been identified with $\{[p:q] \in P^1(\mathbb{R}) \mid p, q \in \mathbb{Z}\} \subset P^1(\mathbb{R})$. This is a consequence of Bezout's identity: for every point $[p:q] \in P^1(\mathbb{R})$ representing a reduced fraction $\frac{p}{q} \neq 0$ ($p, q \in \mathbb{Z} \setminus \{0\}$ and coprime), there are two integers x, y such that $xq - yp = 1$. Therefore

$$g' = \begin{pmatrix} x & p \\ y & q \end{pmatrix}$$

is an element of G and $g'[0:1] = [p:q]$. The case $q = 0$ is covered by

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} [0:1] = [1:0].$$

The extended equations (5.19) and (5.20) are such that $h(x) = 0$ implies $h(\alpha x) = 0$, $h(\beta x) = 0$, $h(\alpha^{-1}x) = 0$ and $h(\beta^{-1}x) = 0$. Since the orbit in $\mathbb{R}$ of 0 by the group of homographies generated by A and B^2 (i.e. G itself) contains the whole set of rational numbers $\mathbb{Q}$ and $h(0) = 0$, we conclude that $h = 0$ on $[0,1] \cap \mathbb{Q}$. □

Part III

Information theory with finite vector spaces

Chapter 6

The q-multinomial coefficients

This chapter introduces the combinatorial objects and results used later in Chapters 7 and 8. In Section 6.1, we define the q-multinomial coefficients, that are associated to the enumeration of flags of finite vector spaces. Section 6.2 studies their asymptotic behavior and establishes the connection with the quadratic entropy. Sections 6.3 and 6.4 are mutually independent and not essential to understand the rest of the paper: the former uses the asymptotic results to obtain a combinatorial explanation for the nonadditivity of Tsallis 2-entropy, and the later discuss a combinatorial justification of the maximum entropy principle with Tsallis entropy.

6.1 Definition

Let q be an indeterminate. The q-integers $\{[n]_q\}_{n\in\mathbb{N}}$ are defined by $[n]_q := (q^n - 1)/(q-1)$ and the q-factorials by $[n]_q! := [n]_q[n-1]_q\cdots[1]_q$. The q-multinomial coefficients are

$$\begin{bmatrix} n \\ k_1,\dots,k_s \end{bmatrix}_q := \frac{[n]_q!}{[k_1]_q!\cdots[k_s]_q!}. \tag{6.1}$$

defined for $(n,k_1,\dots,k_s)\in\mathbb{N}^{s+1}$ such that $\sum_{i=1}^s k_i = n$.

Throughout this paper, we shall assume that q is a fixed prime power. For such q, the q-binomial coefficient $\begin{bmatrix} n \\ k \end{bmatrix}_q \equiv \begin{bmatrix} n \\ k,n-k \end{bmatrix}_q$ counts the number of k-dimensional subspaces in $\mathbb{F}_q^n$. More generally, given a set of integers $k_1,\dots,k_s$ such that $\sum_{i=1}^s k_i = n$, the q-multinomial coefficient $\begin{bmatrix} n \\ k_1,\dots,k_s \end{bmatrix}_q$ equals the number of flags $V_1 \subset V_2 \subset \cdots \subset V_{s-1} \subset V_s = \mathbb{F}_q^n$ of vector spaces such that $\dim V_j = \sum_{i=1}^j k_i$, see [68,69]. We will say that these flags are of ***type*** $(k_1,\dots,k_s)$.

It is possible to introduce a function Γ_q as the normalized solution of a functional equation that guaranties that $[n]_q! = \Gamma_q(n+1)$, see [6]. When $q>1$ and $x>0$, this function is given by the formula [64]:

$$\Gamma_q(x) = (q^{-1};q^{-1})_\infty q^{\binom{x}{2}}(q-1)^{1-x}\sum_{n=0}^{\infty}\frac{q^{-nx}}{(q^{-1};q^{-1})_n} \tag{6.2}$$

$$= \frac{(q^{-1};q^{-1})_\infty q^{\binom{x}{2}}(q-1)^{1-x}}{(q^{-x};q^{-1})_\infty}, \tag{6.3}$$

where we have used the Pochhammer symbol

$$(a;x)_n := \prod_{k=0}^{n-1}(1-ax^k), \qquad (a;x)_0 = 1. \tag{6.4}$$

The equivalent expressions for the Γ_q-function come from the identity

$$\frac{(ax;q)_\infty}{(x;q)_\infty} = \sum_{n=0}^{\infty} \frac{(a;q)_n}{(q;q)_n} x^n \qquad (|q|<1), \tag{6.5}$$

known as ***q*-binomial theorem** [45, p. 30].

Recall [50, p. 92] that an infinite product $\prod_{i=0}^{\infty} u_i$ is said to be convergent if

1. there exists i_0 such that $u_i \neq 0$ for all $i > i_0$;
2. $\lim_{n\to\infty} u_{i_0+1}\cdots u_{i_0+n}$ exists and is different from zero.

An infinite product in the form $\prod(1+c_i)$ is said to be absolutely convergent when $\prod(1+|c_i|)$ converges. One can show that absolute convergence implies convergence. Moreover, when the terms $\gamma_i \geq 0$, the product $\prod_i(1+\gamma_i)$ is convergent if and only if the series $\sum_i \gamma_i$ converges. The convergence of $\sum_i 1/q^i$ gives then the following result, that is used without further comment throughout the thesis.

Lemma 6.1. *For every $a \in \mathbb{C}$, the product $(a;q^{-1})_\infty$ converges. Moreover, if $a \notin \{\, q^i \mid i \geq 0 \,\}$, then $(a;q^{-1})_\infty \neq 0$.*

The Γ_q function gives an alternative expression for the q-multinomial coefficients

$$\begin{bmatrix} n \\ k_1,\dots,k_s \end{bmatrix}_q = \frac{\Gamma_q(n+1)}{\Gamma_q(k_1+1)\cdots\Gamma_q(k_s+1)}, \tag{6.6}$$

which in turn extends its definition to complex arguments.

We close this subsection with a remark on the unimodality of the q-binomial coefficients.

Lemma 6.2. *For every $n \in \mathbb{N}$,*

- $\begin{bmatrix} n \\ 0 \end{bmatrix}_q < \begin{bmatrix} n \\ 1 \end{bmatrix}_q < \dots < \begin{bmatrix} n \\ \lfloor n/2 \rfloor \end{bmatrix}_q$,
- $\begin{bmatrix} n \\ \lfloor n/2 \rfloor \end{bmatrix}_q = \begin{bmatrix} n \\ \lceil n/2 \rceil \end{bmatrix}_q$,
- $\begin{bmatrix} n \\ \lceil n/2 \rceil \end{bmatrix}_q > \dots \begin{bmatrix} n \\ n-1 \end{bmatrix}_q > \begin{bmatrix} n \\ n \end{bmatrix}_q$.

Proof. Consider the quotient

$$Q(n,k) := \frac{\begin{bmatrix} n \\ k+1 \end{bmatrix}_q}{\begin{bmatrix} n \\ k \end{bmatrix}_q} = \frac{[n-k]_q}{[k+1]_q}. \tag{6.7}$$

Then, $Q(n,k) \geq 1$ iff $q^{n-k} \geq q^{k+1}$ iff $k \leq \frac{n-1}{2}$, with equality just in the case $k = \frac{n}{2} - \frac{1}{2} = \lfloor n/2 \rfloor$ (when n is odd). $\square$

6.2 Asymptotic behavior

The quadratic entropy S_2 of a probability law $(\mu_1, ..., \mu_s)$ is defined by[1]

$$S_2(\mu_1, \cdots, \mu_s) := 1 - \sum_{i=1}^{s} \mu_i^2. \tag{6.8}$$

We also use the notation $_2(x)$, for $x \in [0,1]$, as a shortcut for $S_2(x, 1-x)$.

Theorem 6.3. *For each $n \in \mathbb{N}$, let $\{k_i(n)\}_{i=1}^{s}$ be a set of positive real numbers such that $\sum_{i=0}^{s} k_i = n$ (we write k_i when n is clear from context). Suppose that, for each $i \in \{1, ..., s\}$, it is verified that $k_i(n) \to l_i \in [0, \infty]$ as $n \to \infty$. Then,*

$$\begin{bmatrix} n \\ k_1, ..., k_s \end{bmatrix}_q \sim (q^{-1}; q^{-1})_\infty^{1-s} \prod_{i=1}^{s} (q^{-(l_i+1)}; q^{-1})_\infty q^{n^2 S_2(\frac{k_1}{n}, ..., \frac{k_s}{n})/2}. \tag{6.9}$$

Recall that $f_n \sim g_n$ means $f_n / g_n \to 1$ as $n \to \infty$. By convention, $(q^{-(\infty+1)}; q^{-1})_\infty = 1$.

Proof. First, we substitute (6.2) in (6.6) (the powers of $(q-1)$ cancel):

$$\begin{bmatrix} n \\ k_1, ..., k_s \end{bmatrix}_q = (q^{-1}; q^{-1})_\infty^{1-s} q^{n^2 S_2(\frac{k_1}{n}, ..., \frac{k_s}{n})/2} \frac{\prod_{i=1}^{s} (q^{-(k_i+1)}; q^{-1})_\infty}{(q^{-(n+1)}; q^{-1})_\infty}. \tag{6.10}$$

Theorem 6.3 is a direct consequence of this equality and the following fact: for any sequence $\{t_n\}_n$ of positive numbers,

$$\lim_{n\to\infty} (q^{-(t_n+1)}; q^{-1})_\infty = 1 \tag{6.11}$$

if $t_n \to \infty$, and

$$\lim_{n\to\infty} (q^{-(t_n+1)}; q^{-1})_\infty = (q^{-(t+1)}; q^{-1})_\infty \tag{6.12}$$

if $t_n \to t \in [0, \infty)$.

To establish (6.11) and (6.12), remark first that

$$(q^{-(t_n+1)}; q^{-1})_\infty = \sum_{j=0}^{\infty} \frac{q^{-j(t_n+1)}}{(q^{-1}; q^{-1})_j}$$

can be written as $\int_{\mathbb{N}} f_n(x) \nu(dx)$, where ν denotes the counting measure and $f_n : \mathbb{N} \to [0, \infty)$ is given by

$$f_n(x) = \frac{q^{-x(t_n+1)}}{(q^{-1}; q^{-1})_x} \tag{6.13}$$

Moreover, $|f_n(x)| \le g(x) := q^{-x}/(q^{-1}; q^{-1})_x$, because $t_n \ge 0$, and $g(x)$ is integrable, $\int_{\mathbb{N}} g(x)\nu(dx) \le (q^{-1}, q^{-1})_\infty^{-1} \frac{1}{1-q^{-1}}$. Therefore, in virtue of Lebesgue's dominated convergence theorem,

$$\begin{aligned} \lim_{n\to\infty} \sum_{j=0}^{\infty} \frac{q^{-j(t_n+1)}}{(q^{-1}; q^{-1})_j} &= \lim_n \int_{\mathbb{N}} f_n(x)\nu(dx) \\ &= \int_{\mathbb{N}} \lim_n f_n(x)\nu(dx) \end{aligned}$$

[1] In this part of the thesis, we fix the constant 1 in front of $1 - \sum_{i=1}^{s} \mu_i^2$.

The point-wise limit $\lim_n f_n(x)$ is $[x = 0]$ when $t_n \to \infty$ and $\frac{q^{-x(t+1)}}{(q^{-1};q^{-1})_x}$ when $t_n \to t$. □

When f_n and g_n are positive, $f_n \sim g_n$ implies that $\lim_n \log_q f_n - \log_q g_n = 0$. For instance, we can deduce that, for any fixed $\Delta \in \mathbb{N}$,

$$\lim_n \frac{1}{n} \log_q \begin{bmatrix} n \\ n-\Delta \end{bmatrix}_q = \lim_n \frac{n}{2} S_2(\Delta/n) = \Delta, \tag{6.14}$$

where the last equality comes from a direct computation.

As an immediate application of Theorem 6.3, we obtain the following limit announced in the Introduction.

Proposition 6.4. *For each $n \in \mathbb{N}$, let $\{k_i(n)\}_{i=1}^s$ be a set of positive real numbers such that $\sum_{i=0}^s k_i = n$ (we write k_i when n is clear from context). Suppose that $k_i/n \to \mu_i \in [0,1]$ as $n \to \infty$, for all i. Then*

$$\lim_{n\to\infty} \frac{2}{n^2} \log_q \begin{bmatrix} n \\ k_1, \ldots, k_s \end{bmatrix}_q = S_2(\mu_1, \cdots, \mu_s). \tag{6.15}$$

Proof. If $f/g \to 1$, then $\log_q(f/g) \to 0$. Therefore,

$$\log_q \begin{bmatrix} n \\ k_1, \ldots, k_s \end{bmatrix}_q - \log_q \left(\frac{(q^{-1};q^{-1})_\infty^{1-s}}{\prod_{i=1}^s (q^{-(l_i+1)};q^{-1})_\infty} \right) - \frac{n^2}{2} S_2\left(\frac{k_1}{n}, \ldots, \frac{k_s}{n}\right) = o(1). \tag{6.16}$$

Multiply this by $2/n^2$ and use the continuity of S_2 to conclude. □

6.3 Combinatorial explanation for nonadditivity of Tsallis 2-entropy

Additivity corresponds to the following property of Shannon entropy: if X is an E_X-valued random variable with law $P = \{p_x\}_{x\in E_X}$ and Y an E_Y-valued variable with law $Q = \{q_y\}_{y\in E_Y}$, independent of X, then the joint variable (X, Y) has law $P \otimes Q := \{p_x q_y\}_{(x,y)\in E_X \times E_Y}$ and

$$S_1[(X, Y)](P \otimes Q) = S_1[X](P) + S_1[Y](Q). \tag{6.17}$$

For simplicity (the arguments work in general), we suppose that X, Y are binary variables, i.e. $E_X = E_Y = \{0, 1\}$. Consider the sequences counted by $\binom{N}{N_{00},N_{01},N_{10},N_{11}}$; they are the possible results of N independent trials of the variable (X, Y), under the assumption that the result (i, j) is obtained N_{ij} times, for each $(i, j) \in \{0, 1\}^2$. We treat the particular case $N_{ij} = p_i q_j N$, that correspond to the expected number of appearances of (i, j). The independence between Y and X means that, given $N_0 := N_{00}+N_{01} = p_0 N$ occurrences of $X = 0$ (resp. $N_1 := N_{10}+N_{11} = p_1 N$ occurrences of $X = 1$) in the sequences of length N counted above, there are $q_0 N_i$ occurrences of $Y = 0$ and $q_1 N_i$ occurrences of $Y = 1$ in the corresponding subsequence defined by the condition $X = i$, irrespective of the value of i. In this case, the multiplicative recurrence relation $g[XY] = (X.g[Y])g[X]$ reads

$$\binom{N}{N_{00}, N_{01}, N_{10}, N_{11}} = \binom{N}{N_0}\binom{N_0}{q_0 N_0}\binom{N_1}{q_0 N_1}. \tag{6.18}$$

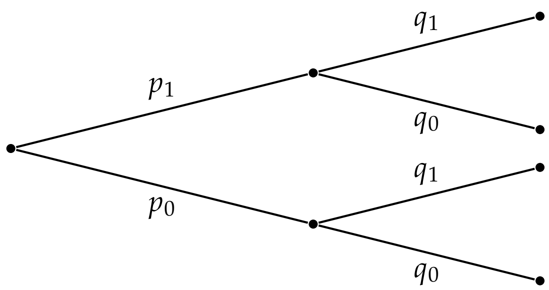

Figure 6.1: Decision tree for the recursive reasoning that leads to equations (6.18) and (6.19).

Applying $\frac{1}{N}\ln(-)$ to both sides and taking the limit $N \to \infty$, we recover (6.17).

In the q-case, $\begin{bmatrix} N \\ N_{00}, N_{01}, N_{10}, N_{11} \end{bmatrix}_q$ counts the number of flags $V_{00} \subset V_{01} \subset V_{10} \subset V_{11} = \mathbb{F}_q^n$ of type $(N_{00}, N_{01}, N_{10}, N_{11})$. When $N_{ij} = p_i q_j N$, such a flag can be determined by an iterated choice of subspaces, whose dimensions are chosen independently: pick first a subspace $V_0 \subset \mathbb{F}_q^n$ of dimension $N_0 = N_{00} + N_{01} = p_0 N$ (there are $\binom{N}{N_0}_q$ of those) and then pick a subspace of dimension $q_0 N_0 \subset V_0$ and another subspace of dimension $q_0 N_1$ in $\mathbb{F}_q^n / V_0$. This corresponds to the combinatorial identity

$$\begin{bmatrix} N \\ N_{00}, N_{01}, N_{10}, N_{11} \end{bmatrix}_q = \begin{bmatrix} N \\ N_0 \end{bmatrix}_q \begin{bmatrix} N_0 \\ q_0 N_0 \end{bmatrix}_q \begin{bmatrix} N_1 \\ q_0 N_1 \end{bmatrix}_q . \tag{6.19}$$

Applying $\frac{2}{N^2}\log_q(-)$ to both sides and taking the limit $N \to \infty$, we obtain

$$\begin{aligned} S_2(p_0q_0, p_0q_1, p_1q_0, p_1q_1) &= S_2(p_0, p_1) + p_0^2 S_2(q_0, q_1) + (1 - p_0)^2 S_2(q_0, q_1) \\ &= S_2(p_0, p_1) + S_2(q_0, q_1) - S_2(p_0, p_1) S_2(q_0, q_1). \end{aligned}$$

In both cases, the trees that represent the iterated counting are the same, see Fig. 6.1 (and compare this with Figure 6 in Shannon's paper [78]). The main difference lies in the exponential growth of the combinatorial quantity of interest and how the correspondent exponents are combined. In the q-case, even if you choose the dimensions in two independent steps, the exponents do not simply add; in fact, the counting of sequences is nongeneric in this respect. Remark also that the interpretation of probabilities as relative *frequencies* of symbols only make sense for the case of words; more generally they correspond to ratios or relative proportions.

6.4 Maximum entropy principle

In the simplest models of statistical mechanics, one assumes that the system is composed of n particles, each one in certain state from a finite set $S = \{s_1, \dots, s_m\}$ (in certain contexts, the elements of S are called *spins*). A configuration of the system is a feasible vector $\mathbf{x} \in S^n$; when all particles are independent, S^n is the sets of all configurations.

We have in mind a new type of statistical mechanics, where a configuration of the n particle system is represented by a flag of vector spaces $V_1 \subset V_2 \subset \dots \subset V_m = \mathbb{F}_q^n$.

In the classical case of independent particles, the total energy of a configuration $\mathbf{x}$ just depends on its type $(k_i)_{1 \le i \le m}$, where k_i is the number of appearances of the symbol s_i in $\mathbf{x}$. In fact, the mean (internal) energy is $\sum_{i=1}^m \frac{k_i}{n} E_i$, where $E_i \in \mathbb{R}$ is the energy associated to the spin s_i. Setting $E_{i+1} = 0$, $\tilde{E}_i = E_i - E_{i+1}$ and $r_i = \sum_{j=1}^i k_j$, one can write $\sum_{i=1}^m \frac{r_i}{n} \tilde{E}_i$ instead of $\sum_{i=1}^m \frac{k_i}{n} E_i$.

Now we plan to move beyond independence, so it is convenient to see the energy as a "global" function that depends on the type of the sequence. We assume now that the energy associated to a flag of vector spaces $V_1 \subset V_2 \subset \ldots \subset V_m = \mathbb{F}_q^n$ just depends on its type $(k_1, \ldots, k_m)$ and is of the form

$$\sum_{i=1}^m \frac{k_i}{n} E_i = \sum_{i=1}^m \frac{r_i}{n} \tilde{E}_i = \sum_{i=1}^m \frac{(\dim V_i)}{n} \tilde{E}_i \tag{6.20}$$

where $r_i = \sum_{j=1}^i k_j$, as before.

In general, if $n > 1$, the equations

$$\sum_{i=1}^n \frac{k_i}{n} E_i = \langle E \rangle \tag{6.21}$$

$$\sum_{i=1}^n k_i = n, \tag{6.22}$$

where $\langle E \rangle \in \mathbb{R}$ is a prescribed mean energy, do not suffice to determine the type $(k_1, \ldots, k_m)$ and an additional principle must be introduced to select the "best" estimate. As a solution to this inference problem, we propose an extension of the ***principle of maximum entropy*** as stated by Jaynes in [42]. Between all the types that satisfy (6.21) and (6.22), we should select the one that corresponds to the greatest number of configurations of the system. This means that we must maximize

$$W(k_1, \ldots, k_m) := \begin{bmatrix} n \\ k_1, k_2, \ldots, k_m \end{bmatrix}_q \tag{6.23}$$

under the constraints (6.21) and (6.22). The maximization of $W(k_1, \ldots, k_m)$ is equivalent to the maximization of $2 \log_q W(k_1, \ldots, k_m)/n^2$; as $n \to \infty$, the latter quantity approaches $S_2(g_1, \ldots, g_m)$, with $g_i := \lim_n k_i/n$. Our ***maximum 2-entropy principle*** says that the best estimate to $(g_1, \ldots, g_m)$ corresponds to the solution to the following problem

$$\begin{aligned} \max \quad & S_2(g_1, \ldots, g_m) \\ \text{subject to} \quad & \sum_{i=1}^m g_i E_j = \langle E \rangle \\ & \sum_{i=1}^m g_i = 1. \end{aligned}$$

This differs from usual presentations of the maximum entropy principle in the literature concerning nonextensive statistical mechanics. Usually the constraints are written in terms of escort distributions derived from $(g_1, \ldots, g_m)$; these have proven useful in several domains, e.g. the analysis of multifractals [11, 85]. However, it is not clear for us how to derive them from combinatorial facts.

Chapter 7

Grassmannian process

When q-is a prime power, the q-binomial coefficients count vector spaces. As explained in the Introduction, this motivates a generalization of information theory where messages are vector spaces in correspondence with the usual information theory for memoryless Bernoulli sources. Table 1 outline the correspondence. Sections 7.1 and 7.3 justify the last row of this table: the former describes the q-deformed version of the binomial distribution, associated to the q-binomial coefficients; the latter introduces an original stochastic model for the generation of generalized messages: a discrete-time stochastic process that gives at time n a vector subspace of $\mathbb{F}_q^n$, that we call Grassmannian process. Finally, Section 7.4 establishes some facts about the asymptotic behavior of this process.

7.1 The q-binomial distribution

Let Z be a random variable that takes the value 1 with probability $\xi \in [0,1]$ and the value 0 with probability $1-\xi$ (Bernoulli distribution). Its characteristic function is

$$\mathbb{E}(\mathrm{e}^{itZ}) = \xi \mathrm{e}^{it} + (1-\xi). \tag{7.1}$$

Let W_n be a random variable with values in $\{0,\dots,n\}$, such that k has probability $\mathrm{Bin}(k|n,\xi) := \binom{n}{k}\xi^k(1-\xi)^{n-k}$, where $\xi \in [0,1]$. The binomial theorem implies that $\mathrm{Bin}(\cdot|n,\xi)$ is a probability mass function, corresponding to the so-called binomial distribution. The theorem also implies that

$$\begin{aligned}\left(\mathbb{E}\left(\mathrm{e}^{itZ}\right)\right)^n &= (\xi\mathrm{e}^{it} + (1-\xi))^n \\ &= \sum_{k=0}^{n}\binom{n}{k}\mathrm{e}^{itk}\xi^k(1-\xi)^{n-k} \\ &= \mathbb{E}\left(\mathrm{e}^{itW_n}\right),\end{aligned} \tag{7.2}$$

which means that $W_n = Z_1 + \dots + Z_n$ (in law), where $Z_1,\dots,Z_n$ are n i.i.d. variables with the same distribution than Z [26, Ch. I, Sec. 11]. Given a collection $\{Z_i\}_{i\geq 1}$ of i.i.d. random variables such that $Z_i \sim \mathrm{Ber}(\xi)$, the process $\{W_n\}_{n\geq 1}$ defined by $W_1 = Z_1$ and $W_n = W_{n-1} + Z_n$ when $n > 1$ is an $\mathbb{N}$-valued markovian stochastic process.

There is a well known combinatorial interpretation for all this: if one generates binary sequences of length n by tossing n times a coin that gives 1 with probability ξ and 0 with probability $1-\xi$, any sequence with exactly k ones has probability $\xi^k(1-\xi)^{n-k}$ and there are $\binom{n}{k}$ of them. Therefore, if Y is the sum of the outputs of all the coins (the number of ones in the generated sequence), the probability of observing $Y=k$ is $\binom{n}{k}\xi^k(1-\xi)^{n-k}$.

There is also a q-binomial theorem, known as the Gauss binomial formula [45, Ch. 5]:

$$(x+y)(x+yq)\cdots(x+yq^{n-1}) = \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix}_q q^{\binom{k}{2}} y^k x^{n-k}. \tag{7.3}$$

Let us write $(x+y)_q^n$ instead of $(x+y)(x+yq)\cdots(x+yq^{n-1})$: the q-analog of $(x+y)^n$. Then (7.3) implies that

$$\mathrm{Bin}_q(k|n,x,y) := \begin{bmatrix} n \\ k \end{bmatrix}_q \frac{q^{\binom{k}{2}} y^k x^{n-k}}{(x+y)_q^n} \tag{7.4}$$

is a probability mass function for $k \in \{0,\dots,n\}$, with parameters $n \in \mathbb{N}$, $x \geq 0$ and $y \geq 0$. Moreover, the factorization

$$\prod_{j=0}^{n-1} \frac{(x+y\mathrm{e}^{it}q^j)}{(x+yq^j)} = \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix}_q \frac{\mathrm{e}^{itk} y^k x^{n-k} q^{\binom{k}{2}}}{(x+y)_q^n} \tag{7.5}$$

shows that a variable Y_n with law $\mathrm{Bin}_q(n,x,y)$ can be written as the sum of n independent variables $X_1,\dots,X_n$, such that X_i takes the value 0 or 1 with probability $x/(x+yq^{i-1})$ and $yq^{i-1}/(x+yq^{i-1})$, respectively.

If we begin with a collection $\{X_i\}_{i\geq 1}$ of independent variables such that $X_i \sim \mathrm{Ber}\left(\frac{yq^{i-1}}{x+yq^{i-1}}\right)$, then the process $\{Y_n\}_{n\geq 1}$ defined by $Y_n = X_1 + \cdots X_n$ is an $\mathbb{N}$-valued markovian stochastic process. When $q \to 1$, each X_i becomes a Bernoulli variable with parameter $y/(x+y)$ and Y has a $\mathrm{Bin}(n, \frac{y}{x+y})$ distribution. Equation (7.5) also implies that

$$\mathbb{E}(Y) = \sum_{j=0}^{n-1} \frac{yq^j}{x+yq^j} = n - \sum_{j=0}^{n-1} \frac{x}{x+yq^j}. \tag{7.6}$$

Provided that $x \neq 0$, one can write the mass function of the q-binomial as follows:

$$\mathrm{Bin}_q(k|n,\theta) := \begin{bmatrix} n \\ k \end{bmatrix}_q \frac{q^{\binom{k}{2}}\theta^k}{(-\theta;q)_n}, \tag{7.7}$$

where $\theta = y/x \geq 0$. We adopt here the classical notation, the q-Pochhammer symbol $(-\theta;q)_n$, instead of $(1+\theta)_q^n$, cf. Section 6.1.[1] Strictly speaking, this is the q-binomial distribution found in the literature [48]. The expectation and the variance of this

[1]The notation can be misleading, because the terms 1 and θ do not commute inside $(1+\theta)_q^n$.

simplified distribution are respectively

$$\mathbb{E}(Y) = \sum_{j=0}^{n-1} \frac{\theta q^j}{1+\theta q^j} = n - \sum_{j=0}^{n-1} \frac{1}{1+\theta q^j}, \tag{7.8}$$

$$\mathbb{V}(Y) = \sum_{j=0}^{n-1} \frac{\theta q^j}{(1+\theta q^j)^2}. \tag{7.9}$$

Set $c_n(\theta) := \sum_{j=0}^{n-1} \frac{1}{1+\theta q^j}$; this sequence is monotonic in n and convergent to certain $c(\theta)$. We do not include q in the notation, since it is fixed from the beginning.

7.2 Parameter estimation by the maximum likelihood method

Let us suppose we make n independent trials of a variable Y with distribution $\mathrm{Bin}_q(n, \theta)$, obtaining results $y_1, \dots, y_m$. The probability of this outcome is

$$P(y_1, \dots, y_m|\theta) = \prod_{i=1}^{m} \begin{bmatrix} n \\ y_i \end{bmatrix}_q \frac{\theta^{y_i} q^{y_i(y_i-1)/2}}{(-\theta; q)_n}. \tag{7.10}$$

This implies that

$$\frac{\partial \log P}{\partial \theta} = \frac{1}{\theta}\left(\sum_{i=1}^{n} y_i - m \sum_{j=0}^{n-1} \frac{\theta q^j}{(1+\theta q^j)} \right). \tag{7.11}$$

By the maximum likelihood method, the best estimate for θ, say $\hat{\theta}$, should maximize P and therefore satisfy $\left.\frac{\partial \log P}{\partial \theta}\right|_{\theta=\hat{\theta}} = 0$; in turn, this equation implies that the empirical mean

$$\bar{y} := \frac{1}{m} \sum_{i=1}^{m} y_i \tag{7.12}$$

should coincide with the theoretical mean

$$m_{q,n}(\theta) := \sum_{j=0}^{n-1} \frac{\theta q^j}{1+\theta q^j}. \tag{7.13}$$

Proposition 7.1. *The map $\theta \mapsto m_{q,n}(\theta)$ establishes a bijection between $[0, \infty)$ and $[0, n)$.*

If this correspondence is extended by $m_{q,n}(\infty) = n$—which corresponds to the case $x = 0$—the value of $\hat{\theta}$ is uniquely determined by the equation $m_{q,n}(\hat{\theta}) = \bar{y}$.

Proof. Since

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\theta q^j}{1+\theta q^j}\right) = \frac{q^j}{(1+\theta q^j)^2} > 0, \tag{7.14}$$

$m_{q,n}(\theta)$ is strictly increasing. Moreover, $m_{q,n}(0) = 0$ and $\lim_{\theta\to\infty} m_{q,n}(\theta) = n$. □

7.3 A vector-space-valued stochastic process associated to the q-binomial distribution

The vector $(Z_1, ..., Z_n)$ is a random binary sequence, but its q-deformation $(X_1, ..., X_n)$, obtained in the previous section, cannot be identified in an obvious way with a vector space. This motivates the introduction of an associated stochastic process $\{V_i\}_{i\in\mathbb{N}}$ such that, for each $n \in \mathbb{N}$, V_n is vector subspace of $\mathbb{F}_q^n$ and the law of $\{X_i\}_{i\in\mathbb{N}^*}$ can be recovered from that of $\{V_i\}_{i\in\mathbb{N}}$.

Let $\mathrm{Gr}(k, n)$ be the set of k-dimensional vector subspaces of $\mathbb{F}_q^n$ and define the total n-th Grassmannian by

$$\mathrm{Gr}(n) := \bigcup_{i=0}^{n} \mathrm{Gr}(i, n). \tag{7.15}$$

Let $\langle 0\rangle = \mathbb{F}_q^0 \hookrightarrow \mathbb{F}_q^1 \hookrightarrow \mathbb{F}_q^2 \hookrightarrow .... \hookrightarrow \mathbb{F}_q^n \hookrightarrow ...$ be a sequence of linear embeddings; note that it induces embeddings at the level of Grassmannians, that will be implicit in what follows. The $(n + 1)$***-dilationsdilation*** of w are

$$\mathrm{Dil}_{n+1}(w) := \{\, v \in \mathrm{Gr}(n + 1) \,|\, w \subset v,\ v \not\subset \mathbb{F}_q^n,\ \dim v - \dim w = 1 \,\}. \tag{7.16}$$

Definition 7.2 (Grassmannian process)**.** Define $V_0 := \mathbb{F}_q^0$, the trivial vector space; for each $n \geq 0$, let V_{n+1} be a random variable taking values in $\mathrm{Gr}(n+1)$ with law defined by[2]

$$\mathbb{P}\left(V_{n+1} = v | V_n = w, X_{n+1} = 0\right) = \delta_w(v), \tag{7.17}$$

$$\mathbb{P}\left(V_{n+1} = v | V_n = w, X_{n+1} = 1\right) = \frac{[v \in \mathrm{Dil}_{n+1}(w)]}{|\,\mathrm{Dil}_{n+1}(w)|}. \tag{7.18}$$

We refer to $\{V_n\}_{n\in\mathbb{N}}$ as the ***Grassmannian process*** associated to the q-binomial process $\{X_i\}_{i\in\mathbb{N}^*}$.

Proposition 7.3. *Let v be a subspace of $\mathbb{F}_q^n$ such that* $\dim(v) = k$*. Then,*

$$\mathbb{P}\left(V_n = v\right) = \frac{\theta^k q^{k(k-1)/2}}{(-\theta; q)_n}. \tag{7.19}$$

Proof. To shorten notation, we write in this section $\mathbb{P}_X(x)$ instead of $\mathbb{P}(X = x)$, and $\mathbb{P}_{X|Y}(x|y)$ instead of $\mathbb{P}\left(X = x|Y = y\right)$.

Our proof is by recurrence. The case $n = 1$ is straightforward; for instance,

$$\begin{aligned}
\mathbb{P}_{V_1}(\langle 0\rangle) &= \mathbb{P}_{V_1|V_0}(\langle 0\rangle|\langle 0\rangle) \\
&= \mathbb{P}_{V_1|V_0,X_1}(\langle 0\rangle|\langle 0\rangle, 0)\,\mathbb{P}_{X_1}(0)\,, \\
&= \mathbb{P}_{X_1}(0)
\end{aligned}$$

because $\langle 0\rangle$ it is not a dilation of itself.

[2] We use Iversen's convention for the characteristic function: $[p] = 1$ if p is true, and vanishes otherwise.

Suppose the formula is valid up to $n \geq 1$. Let v be a subspace of $\mathbb{F}_q^{n+1}$ of dimension k. When v is contained in $\mathbb{F}_q^n$,

$$\begin{aligned}\mathbb{P}_{V_{n+1}}(v) &= \mathbb{P}_{V_{n+1}|V_n,X_{n+1}}(v|v,0)\,\mathbb{P}_{X_{n+1}}(0)\,\mathbb{P}_{V_n}(v)\\ &= 1\cdot\frac{1}{1+\theta q^n}\frac{\theta^k q^{k(k-1)/2}}{(-\theta;q)_n} = \frac{\theta^k q^{k(k-1)/2}}{(-\theta;q)_{n+1}}.\end{aligned}$$

If $v \not\subset \mathbb{F}_q^n$,

$$\begin{aligned}\mathbb{P}_{V_{n+1}}(v) &= \sum_{w\in \mathrm{Gr}(n)} \mathbb{P}_{V_{n+1}|V_n,X_{n+1}}(v|w,1)\,\mathbb{P}_{Y_n}(w)\,\mathbb{P}_{X_{n+1}}(1)\\ &= \sum_{\substack{w\in\mathrm{Gr}(k-1,n)\\ w\subsetneq V}} \frac{1}{|\,\mathrm{Dil}_{n+1}(w)|}\left(\frac{\theta^{k-1}q^{\binom{k-1}{2}}}{(-\theta;q)_n}\right)\frac{\theta q^n}{(1+\theta q^n)}\\ &= \frac{\theta^k q^{\binom{k-1}{2}}q^n}{|\,\mathrm{Dil}_{n+1}(v\cap\mathbb{F}_q^n)|(-\theta;q)_{n+1}}.\end{aligned}$$

The formula $\dim U + \dim V = \dim(U+V) + \dim(U\cap V)$ entails that $v\cap\mathbb{F}_q^n$ has dimension $k-1$. Any $w\in\mathrm{Gr}(k-1,n)$ such that $w\subset v$ must be contained in $v\cap\mathbb{F}_q^n$ and have the same dimension, implying that $w = v\cap\mathbb{F}_q^n$; this explain the last equality above.

Finally, let w be a $k-1$ dimensional subspace in $\mathbb{F}_q^n$; to dilate it into a $v\in \mathrm{Gr}(k,n+1)\setminus\mathrm{Gr}(k,n)$, one must pick a vector x outside $\mathbb{F}_q^n$: there are $q^{n+1}-q^n$ of those. However, since $w+\langle x\rangle$ has q^k points and w just q^{k-1}, there are $q^k - q^{k-1}$ choices of x that give the same dilation v. Therefore, the number of different dilations is

$$\frac{q^{n+1}-q^n}{q^k-q^{k-1}} = q^{n-(k-1)}. \tag{7.20}$$

i.e. $|\,\mathrm{Dil}_{n+1}(v\cap\mathbb{F}_q^n)|$ equals $q^{n-(k-1)}$. □

Corollary 7.4.

$$\mathbb{P}(\dim V_n = k) = \begin{bmatrix} n\\ k\end{bmatrix}_q \frac{\theta^k q^{k(k-1)/2}}{(-\theta;q)_n}. \tag{7.21}$$

Proof. This is a consequence of Proposition 7.3 and the fact that $\begin{bmatrix} n\\ k\end{bmatrix}_q$ counts the number of k dimensional subspaces of $\mathbb{F}_q^n$. □

Proposition 7.5. *Let $\{Y_n\}_{n\in\mathbb{N}^*}$ denote a q-binomial process, $Y_n \sim \mathrm{Bin}_q(n,\theta)$, and $\{V_n\}_{n\in\mathbb{N}}$ its associated Grassmannian process. Let v be a subspace of $\mathbb{F}_q^n$ of dimension $k = n-d$, for $d\in [\![0,n]\!]$. Then,*

$$\mathbb{P}(V_n = v) = \frac{q^{-\frac{1}{2}(d-(\frac{1}{2}-\log_q\theta))^2+\frac{1}{2}(\frac{1}{2}-\log_q\theta)^2-\frac{n^2}{2}S_2(d/n)}}{(-\theta^{-1};q^{-1})_n}. \tag{7.22}$$

Proof. We shall rewrite the various factors in (7.19). In the first place,

$$(-\theta;q)_n = \prod_{i=0}^{n-1} \theta q^i (1+\frac{1}{\theta q^i}) = \theta^n q^{n(n-1)/2}(-\theta^{-1};q^{-1})_n. \tag{7.23}$$

Note also that $n^2 S_2(d/n) = n^2 - k^2 - d^2$, which implies

$$q^{\binom{k}{2}} = q^{k^2/2} q^{-k/2} = q^{(n^2-n^2S_2(d/n)-d^2)/2} q^{(d-n)/2}. \tag{7.24}$$

Finally, $\theta^k = \theta^{n-d}$. Replace all this in (7.19) and simplify to obtain

$$\mathbb{P}(V_n = v) = \frac{q^{-\frac{d^2}{2}+d(\frac{1}{2}-\log_q \theta)} q^{-\frac{n^2}{2}S_2(d/n)}}{(-\theta^{-1};q^{-1})_n}. \tag{7.25}$$

Complete the square in the exponent to conclude. □

7.4 Asymptotics

Let us define a function $\mu : \mathbb{N} \to (0,\infty)$ by

$$\mu(d) := \frac{q^{-\frac{1}{2}(d-(\frac{1}{2}-\log_q \theta))^2+\frac{1}{2}(\frac{1}{2}-\log_q \theta)^2}(q^{-(d+1)};q^{-1})_\infty}{(q^{-1};q^{-1})_\infty(-\theta^{-1};q^{-1})_\infty}, \tag{7.26}$$

and introduce the notation $\mu([\![a,b]\!]) := \sum_{d\in[\![a,b]\!]} \mu(d)$.[3]

The asymptotic formula in Theorem 6.3, combined with Proposition 7.5, implies that

$$\mathbb{P}(V_n \in \mathrm{Gr}(n-d,n)) = \begin{bmatrix} n \\ n-d \end{bmatrix}_q \mathbb{P}\left(V_n = \mathbb{F}_q^{n-d}\right) \to \mu(d), \tag{7.27}$$

for each fixed $d \in \mathbb{N}$.

Proposition 7.6.

$$\sum_{d=0}^{\infty} \mu(d) = 1. \tag{7.28}$$

Therefore, there is a well defined function $\Delta : [0,1) \to \mathbb{N}$ that associates to each $p \in [0,1)$ the smallest d such that $\mu([\![0,d]\!]) \geq p$; explicitly

$$\Delta(p) = \sum_{k=0}^{\infty} [p > \mu([\![0,k]\!])]. \tag{7.29}$$

The sum is finite for every $p \in [0,1)$. Note that Δ is *left* continuous. This function plays an important role in the proof of Theorem 8.2.

We prove first a lemma that will be useful in the proof of Proposition 7.6.

[3]We denote by $[\![a,b]\!]$ the "discrete interval" $[a,b] \cap \mathbb{Z}$.

Lemma 7.7. *For every $n \in \mathbb{N}$ and every $d \in [0, n]$,*

$$\frac{(q^{-(n-d+1)}; q^{-1})_\infty}{q^{-(n+1)}; q^{-1})_\infty} \leq 1. \tag{7.30}$$

Moreover, for every $n \in \mathbb{N}$ and every $d \in [\![0, 2\sqrt{n}]\!]$,

$$1 - c(q) q^{-(\sqrt{n}+1)^2} \leq \frac{(q^{-(n-d+1)}; q^{-1})_\infty}{q^{-(n+1)}; q^{-1})_\infty}, \tag{7.31}$$

where $c(q) = 2(q^{-1}; q^{-1})_\infty$.

Proof. In this proof we use repeatedly the q-binomial theorem (6.5). For any $k \in \mathbb{N}$, $q^{-k(n+1)} \leq q^{-k(n-d+1)}$, which in turn implies (7.30):

$$\begin{aligned} \frac{1}{(q^{-(n+1)}; q^{-1})_\infty} &= \sum_{k=0}^{\infty} \frac{q^{-k(n+1)}}{(q^{-1}; q^{-1})_k} \\ &\leq \sum_{k=0}^{\infty} \frac{q^{-k(n-d+1)}}{(q^{-1}; q^{-1})_k} \\ &= \frac{1}{(q^{-(n-d+1)}; q^{-1})_\infty}. \end{aligned}$$

To prove (7.31), first remark that

$$\begin{aligned} \frac{1}{(q^{-(n-d+1)}; q^{-1})_\infty} - \frac{1}{(q^{-(n+1)}; q^{-1})_\infty} &= \sum_{k=1}^{\infty} \frac{q^{-k(n+1)}(q^{kd} - 1)}{(q^{-1}; q^{-1})_k} \\ &\leq (q^{-1}; q^{-1})_\infty^{-1} \sum_{k=1}^{\infty} q^{-k(n+1)} q^{kd} \\ &\leq (q^{-1}; q^{-1})_\infty^{-1} \sum_{k=1}^{\infty} q^{-k(\sqrt{n}+1)^2}. \end{aligned}$$

Remark that we omit the term corresponding to $k = 0$, since it vanishes. The first of these inequalities is implied by the trivial bound $x - 1 \leq x$ and the fact that $\{(q^{-1}; q^{-1})_k\}_k$ decreases with k; the second, from $d \leq 2\sqrt{n}$. The geometric series $\sum_{k=1}^{\infty} q^{-k(\sqrt{n}+1)^2}$ equals $q^{-(\sqrt{n}+1)^2}(1 - q^{-(\sqrt{n}+1)^2})^{-1}$, that is upper-bounded by $2q^{-(\sqrt{n}+1)^2}$, because $q \geq 2$. Hence, we have

$$\begin{aligned} \frac{1}{(q^{-(n-d+1)}; q^{-1})_\infty} - \frac{1}{(q^{-(n+1)}; q^{-1})_\infty} &\leq 2(q^{-1}; q^{-1})_\infty^{-1} q^{-(\sqrt{n}+1)^2} \\ &= c(q) q^{-(\sqrt{n}+1)^2}. \end{aligned}$$

Finally, note that $\frac{1}{(q^{-(n-d+1)}; q^{-1})_\infty} = 1+$ (positive term)≥ 1, therefore it is also true that

$$\frac{1}{(q^{-(n-d+1)}; q^{-1})_\infty} - \frac{1}{(q^{-(n+1)}; q^{-1})_\infty} \leq \frac{c(q) q^{-(\sqrt{n}+1)^2}}{(q^{-(n-d+1)}; q^{-1})_\infty}. \tag{7.32}$$

□

Proof of Proposition 7.6. To simplify notation, set

$$A(d) := -\frac{1}{2}(d - (\frac{1}{2} - \log_q \theta))^2 + \frac{1}{2}(\frac{1}{2} - \log_q \theta)^2. \tag{7.33}$$

and $B_n = (-\theta^{-1}; q^{-1})_n^{-1}$. Recall from (6.10) that

$$\begin{bmatrix} n \\ n-d \end{bmatrix}_q = \frac{q^{n^2 S_2(d/n)/2}(q^{-(d+1)}; q^{-1})_\infty (q^{-(n-d+1)}; q^{-1})_\infty}{(q^{-1}; q^{-1})_\infty (q^{-(n+1)}; q^{-1})_\infty}. \tag{7.34}$$

This and (7.22) give

$$\begin{aligned} 1 &= \sum_{d=0}^{n} \mathbb{P}(V_n \in \mathrm{Gr}(n-d, n)) \\ &= B_n \sum_{d=0}^{n} \frac{q^{A(d)}(q^{-(d+1)}; q^{-1})_\infty}{(q^{-1}; q^{-1})_\infty} \frac{(q^{-(n-d+1)}; q^{-1})_\infty}{(q^{-(n+1)}; q^{-1})_\infty} \\ &\le B_n \sum_{d=0}^{n} \frac{q^{A(d)}(q^{-(d+1)}; q^{-1})_\infty}{(q^{-1}; q^{-1})_\infty}. \end{aligned} \tag{7.35}$$

At the end we have used the inequality (7.30). In turn, (7.35) implies that

$$(-\theta^{-1}; q^{-1})_\infty \le \sum_{d=0}^{\infty} \frac{q^{A(d)}(q^{-(d+1)}; q^{-1})_\infty}{(q^{-1}; q^{-1})_\infty} \tag{7.36}$$

We shall see that in fact this is an equality, as the proposition claims. Using this time (7.31), we obtain

$$\begin{aligned} 1 &\ge \sum_{d=0}^{\lfloor 2\sqrt{n} \rfloor} \mathbb{P}(V_n \in \mathrm{Gr}(n-d, n)) \\ &\ge B_n \sum_{d=0}^{\lfloor 2\sqrt{n} \rfloor} \frac{q^{A(d)}(q^{-(d+1)}; q^{-1})_\infty}{(q^{-1}; q^{-1})_\infty}(1 - c(q)q^{-(\sqrt{n}+1)^2}). \end{aligned}$$

which is equivalent to

$$\sum_{d=0}^{\lfloor 2\sqrt{n} \rfloor} \frac{q^{A(d)}(q^{-(d+1)}; q^{-1})_\infty}{(q^{-1}; q^{-1})_\infty} \le \frac{(-\theta^{-1}; q^{-1})_n}{1 - c(q)q^{-(\sqrt{n}+1)^2}}. \tag{7.37}$$

In the limit,

$$\sum_{d=0}^{\infty} \frac{q^{A(d)}(q^{-(d+1)}; q^{-1})_\infty}{(q^{-1}; q^{-1})_\infty} \le (-\theta^{-1}; q^{-1})_\infty. \tag{7.38}$$

and this finishes the proof. □

Chapter 8

Generalized information theory

In this chapter, we prove a fundamental result on measure concentration for the Grassmannian process (Theorem 8.2), that generalizes the asymptotic equipartition property to this setting. It justifies the definition of "typical subspaces". Section 8.3 applies this result to source coding.

8.1 Remarks on measure concentration and typicality

The following definition covers the different stochastic models discussed so far. We use it to clarify the correspondence between Shannon's information theory for sequences and our version for vector subspaces from the probabilistic viewpoint.

Definition 8.1 (Refinement of a law)**.** Let $\pi : (A, \mathfrak{A}) \to (B, \mathfrak{B})$ be a measurable map, and p a probability measure on $(B, \mathfrak{B})$. The law has a refinement with respect to π (or π-refinement) whenever there exists a probability distribution $\tilde{p}$ on $(A, \mathfrak{A})$ such that $\pi_* \tilde{p} = p$, where $\pi_* \tilde{p}$ denotes the image law (the push-forward of $\tilde{p}$, its marginalization).

In applications, p is the law of a $(B, \mathfrak{B})$-valued random variable X and $\tilde{p}$, the law of a $(A, \mathfrak{A})$-valued random variable Y. When $B \subset \mathbb{C}$,

$$\mathbb{E}_{\tilde{p}}(\mathrm{e}^{it\pi(Y)}) = \mathbb{E}_p(\mathrm{e}^{itX}). \tag{8.1}$$

There are four fundamental examples:

1. The probability measure $\mathrm{Ber}(\xi)^{\times n}$ on $\{0,1\}^n$, that assigns to every sequence with k ones the probability $\xi^k(1-\xi)^{n-k}$, is a refinement of the law $\mathrm{Bin}(n, \xi)$ with respect to the surjection $\pi_1 : \{0,1\}^n \to \{0, 1, ..., n\}, (x_1, ..., x_n) \mapsto \sum_i x_i$.
2. The previous example generalizes to the so-called multinomial distribution. Let $S = \{s_1, ..., s_m\}$ be a finite set and μ any probability law on S; set $p_i := \mu(\{s_i\})$. The law $\mu^{\otimes n}$ assigns to a sequence x in S^n the probability $\prod_{i=1}^m p_i^{a_i(x)}$, where $a_i(x)$ denotes the number of appearances of the symbol s_i in the sequence x. Let $T = \{ (k_1, ..., k_m) \in \mathbb{N}^m \mid \sum_{i=1}^m k_i = n \}$; there is a surjection $\pi_2 : S^n \to T$ given by $x \mapsto (a_1(x), ..., a_m(x))$. Denote by ν the marginalization of $\mu^{\otimes n}$ under this map, given explicitly by $\nu(\{(k_1, ..., k_m)\}) = \binom{n}{k_1, ..., k_n} \prod_{i=1}^m p_i^{k_i}$. Then $\mu^{\otimes n}$ is a π_2-refinement of ν.

3. The probability measure $\prod_{i=1}^{n-1} \mathrm{Ber}(\frac{\theta q^i}{1+\theta q^i})$ on $\{0,1\}^n$ is a refinement of the law $\mathrm{Bin}_q(n,\theta)$ under the application π_1 introduced above, see (7.5).
4. The probability measure on $\mathrm{Gr}(n)$ defined by (7.19), that we denote $\mathrm{Grass}(n,\theta)$, is also a refinement of $\mathrm{Bin}_q(n,\theta)$ with respect to the surjection $\pi_3 : \mathrm{Gr}(n) \to \{0,1,...,n\}, V \mapsto \dim V$.

Let us consider for a moment the binomial case 1. Since $W_n \sim \mathrm{Bin}(n,p)$, Chebyshev's inequality reads $\mathbb{P}\left(|W_n - pn| > n^{\frac{1}{2}+\xi}\right) \leq p(1-p)/n^{2\xi}$, which goes to 0 as long as $\xi > 0$. In other words, the measure $\mathrm{Bin}(n,p)$ concentrates on the interval $I_{n,\xi} = [\![np - n^{\frac{1}{2}+\xi}, np + n^{\frac{1}{2}+\xi}]\!] \cap [\![0,n]\!]$, in the sense that $\mathbb{P}(W_n \in I_n^c) \to 0$ as $n \to \infty$, and therefore the measure $\mathrm{Ber}(\xi)^{\times n}$ concentrates on $\pi_1^{-1}(I_{n,\xi})$, that can be regarded as a set of "typical sequences". Moreover, the different type classes $\pi^{-1}(t)$, for $t \in I_{n,\xi}$, have cardinality $\exp\{nH_1(p) + o(n)\}$. An analogous argument shows that the measure $\mathrm{Bin}_q(n,\theta)$ concentrates on the interval $J_{n,\xi} = [\![k_n^* - n^{\xi}, k_n^* + n^{\xi}]\!] \cap [\![0,n]\!]$ for any $\xi > 0$, and hence $\mathrm{Grass}(n,\theta)$ concentrates on $\pi_3^{-1}(J_{n,\xi})$. However, there is a difference: while $\mathrm{Bin}(k|n,p)$ goes to 0 for any value of k, and in fact on needs more than $\sqrt{n}$ different types k to accumulate asymptotically a prescribed probability $p_\varepsilon := 1-\varepsilon$, the values of $\mathrm{Grass}(k|n,\theta) = \mathbb{P}(V_n \in \mathrm{Gr}(k.n))$ tend to the constant value $\mu(d)$, independent of n. In the limit, only a finite number of different types k are necessary to accumulate probability p_ε, and the corresponding type classes differ in size (even asymptotically). Theorem 8.2 bellow reflects this particular situation.

8.2 Typical subspaces

We are ready to prove the main result of this part of the thesis, which extends Theorems 3 and 4 of Shannon's seminal article [78] to this setting.

Theorem 8.2. *Let $\{Y_n\}_{n\in\mathbb{N}^*}$ denote a q-binomial process, $Y_n \sim \mathrm{Bin}_q(n,\theta)$; $\{V_n\}_{n\in\mathbb{N}}$ its associated Grassmannian process; and $\delta \in (0,1)$ an arbitrary number. Let $\varepsilon > 0$ be such that $p_\varepsilon := 1-\varepsilon$ is a continuity point of Δ. Define $A_n = \bigcup_{k=0}^{a_n} \mathrm{Gr}(n-k,n)$ as the smallest set of the form $\bigcup_{k=0}^{m} \mathrm{Gr}(n-k,n)$ such that $\mathbb{P}(V_n \in A_n^c) \leq \varepsilon$. Then, there exists $n_0 \in \mathbb{N}$ such that, for every $n \geq n_0$,*

1. *$A_n = \bigcup_{k=0}^{\Delta(p_\varepsilon)} \mathrm{Gr}(n-k,n)$;*
2. *for any $v \in A_n$ such that $\dim v = k$,*

$$\left| \frac{\log_q(\mathbb{P}(V_n = v)^{-1})}{n} - \frac{n}{2} S_2(k/n) \right| \leq \delta. \tag{8.2}$$

The size of A_n is optimal, up to the first order in the exponential: let $s(n,\varepsilon)$ denote $\min\{\, |B_n| \mid B_n \subset \mathrm{Gr}(n)$ and $\mathbb{P}(V_n \in B_n) \geq 1-\varepsilon \,\}$; then

$$\begin{aligned} \lim_n \frac{1}{n} \log_q |A_n| &= \lim_n \frac{1}{n} \log_q s(n,\varepsilon) \\ &= \lim_n \frac{n}{2} S_2(\Delta(p_\varepsilon)/n) \\ &= \Delta(p_\varepsilon). \end{aligned} \tag{8.3}$$

The set A_n correspond to the "typical subspaces", in analogy with typical sequences.

Proof. To shorten the notation, let us write $\mathbb{P}_n(A)$ instead of $\mathbb{P}(V_n \in A)$, and G_k^n instead of $\mathrm{Gr}(k,n)$.

Given any $\eta > 0$, there exists $n(\eta) \in \mathbb{N}$ such that, for every $n \geq n(\eta)$ and every $d \in [\![0, \Delta(p_\varepsilon)]\!]$,

$$|\mathbb{P}_n\left(G_{n-d}^n\right) - \mu(d)| < \frac{\eta}{\Delta(p_\varepsilon)+1}, \tag{8.4}$$

because $\mathbb{P}_n\left(G_{n-d}^n\right) \longrightarrow \mu(d)$ for each d.

Since p_ε is a continuity point of Δ, a piece-wise constant function, there exists $\xi > 0$ such that

$$\Delta(1-\varepsilon-\xi) = \Delta(1-\varepsilon) = \Delta(1-\varepsilon+\xi).$$

Remark now that, for every $n \geq n(\xi)$,

$$\sum_{d=0}^{\Delta(p_\varepsilon)} \mathbb{P}_n\left(G_{n-d}^n\right) > \sum_{d=0}^{\Delta(p_\varepsilon)} \mu(d) - \xi \geq 1-\varepsilon, \tag{8.5}$$

because $\mu([\![0,\Delta(p_\varepsilon)]\!]) = \sum_{d=0}^{\Delta(p_\varepsilon)} \mu(d) \geq 1-\varepsilon+\xi$. This is a direct consequence of $\Delta(p_\varepsilon) = \Delta(1-\varepsilon+\xi)$.

Analogously, for each $n \geq n(\xi)$,

$$\begin{aligned}\sum_{d=0}^{\Delta(p_\varepsilon)-1} \mathbb{P}_n\left(G_{n-d}^n\right) &< \sum_{d=0}^{\Delta(p_\varepsilon)-1} \mu(d) + \frac{\Delta(p_\varepsilon)}{\Delta(p_\varepsilon)+1}\xi \\ &< 1-\varepsilon-\frac{\xi}{\Delta(p_\varepsilon)+1} \\ &< 1-\varepsilon, \end{aligned}\tag{8.6}$$

because $\mu([\![0,\Delta(p_\varepsilon)-1]\!]) < 1-\varepsilon-\xi$: if this is not the case, $\Delta(1-\varepsilon-\xi) \leq \Delta(\varepsilon)-1$. The inequalities (8.5) and (8.6) imply the part 1 of the theorem whenever $n \geq n(\xi)$.

We suppose now that $n > n(\xi)$. Let v be an element of A_n of dimension k, and set $d = n-k$. The formula in Proposition 7.5 can be stated as

$$-\frac{\log_q \mathbb{P}(V_n = v)}{n} = \frac{g(d,n)}{n} + \frac{n}{2}S_2(d/n), \tag{8.7}$$

where we have set $g(d,n) = \frac{1}{2}(d-(\frac{1}{2}-\log_q\theta))^2 - \frac{1}{2}(\frac{1}{2}-\log_q\theta)^2 + \log_q(-\theta^{-1};q^{-1})_n$. Since d belongs to the interval $[\![0,\Delta(p_\varepsilon)]\!]$, independent on n, and $(-\theta^{-1};q^{-1})_n \to (-\theta^{-1};q^{-1})_\infty$, there exists $n_0 \geq n(\xi)$ such that, for every $n \geq n_0$ and every $d \in [\![0,\Delta(p_\varepsilon)]\!]$, $g(d,n)/n < \delta$, which proves part 2 of the theorem.

For n big enough, $\Delta(p_\varepsilon)$ belongs to the interval $[n/2, n]$. The inequalities in Lemma 6.2 imply that

$$\begin{bmatrix} n \\ n-\Delta(p_\varepsilon)\end{bmatrix}_q \leq |A_n| \leq \sum_{k=0}^{\Delta(p_\varepsilon)} \begin{bmatrix} n \\ n-k\end{bmatrix}_q \leq (\Delta(p_\varepsilon)+1)\begin{bmatrix} n \\ n-\Delta(p_\varepsilon)\end{bmatrix}_q. \tag{8.8}$$

Therefore,

$$\lim_n \frac{1}{n}\log_q |A_n| = \lim_n \frac{1}{n}\log_q \begin{bmatrix} n \\ n-\Delta(p_\varepsilon)\end{bmatrix}_q = \Delta(p_\varepsilon), \tag{8.9}$$

where the second equality comes from (6.14).

For any ε, we show now how to build iteratively a set B_n of minimal cardinality such that $\mathbb{P}_n(B_n^c) \le \varepsilon$: start with $B_n = \emptyset$ and then add vector subspaces of $\mathbb{F}_q^n$ one-by-one, picking at each time any of the vector subspaces of *highest dimension* in B_n^c, until you attain $\mathbb{P}_n(B_n^c) \le \varepsilon$. Let $n - b_n$ be the dimension of the last space included in B_n. It is easy to prove that $b_n < 2\sqrt{n}$, as a consequence of Chebyshev's inequality (the interval $[n - 2\sqrt{n}, n]$ accumulates probability p_ε when n is big enough). This construction gives in fact the smallest possible set, because the function $f_n : [0, n] \to \mathbb{R}$, $x \mapsto \theta^x q^{x(x-1)/2}/(-\theta, q)_n$ is strictly convex and has attains its minimum at $x_0 = \frac{1}{2} - \log_q \theta$; therefore, all the subspaces are included in B_n in decreasing order of probability, and the probability of the last space included is bounded bellow by $\theta^{n-2\sqrt{n}} q^{(n-2\sqrt{n})(n-2\sqrt{n}-1)/2}/(-\theta, q)_n$, which is much bigger that $(-\theta, q)_n^{-1}$, the maximum of f_n on $[0, x_0]$, when n is big enough.

Two versions of B_n only differ in the particular subspaces of dimension $n - b_n$ they include, but they coincide on $\bigcup_{k=0}^{b_n-1} G_{n-k}^n$. In what follows, B_n denotes any of the possible sets. Remark also that $B_n \subset A_n$; even more, $a_n = b_n$ (a strict inequality between the two contradicts the minimality of either B_n or a_n). It is also true in general that

$$\begin{aligned} p_\varepsilon &\le \mathbb{P}_n(B_n) \\ &= \sum_{k=0}^{a_n} \mathbb{P}_n\left(B_n \cap G_{n-k}^n\right) \\ &= \mathbb{P}_n\left(B_n \cap G_{n-a_n}^n\right) + \sum_{k=0}^{a_n-1} \mathbb{P}_n\left(B_n \cap G_{n-k}^n\right). \end{aligned} \tag{8.10}$$

We restrict ourselves again to the case in which p_ε is continuity point of Δ, in such a way that $\Delta(p_\varepsilon) = a_n = b_n$. Under these hypotheses, we are able to lower-bound uniformly the term $\mathbb{P}_n\left(B_n \cap G_{n-\Delta(p_\varepsilon)}^n\right)$ by using (8.10), and deduce from this that $|B_n|$ grows like $|A_n|$, that in turn grows like $|G_{n-\Delta(p_\varepsilon)}^n|$, as shown in (8.9). In fact, we have that

$$\begin{aligned} \sum_{k=0}^{\Delta(p_\varepsilon)-1} \mathbb{P}_n\left(B_n \cap G_{n-k}^n\right) &\le \sum_{k=0}^{\Delta(p_\varepsilon)-1} \mathbb{P}_n\left(G_{n-k}^n\right) \\ &< 1 - \varepsilon - \frac{\xi}{\Delta(p_\varepsilon) + 1}, \end{aligned} \tag{8.11}$$

where we have used again the bound in (8.6). Inequalities (8.10) and (8.11) imply that

$$\frac{\xi}{\Delta(p_\varepsilon) + 1} < \mathbb{P}_n\left(B_n \cap G_{n-\Delta(p_\varepsilon)}^n\right). \tag{8.12}$$

When $n > n_0$, the part (2) entails that $\mathbb{P}_n(x) \le q^{-n^2 S_2(\Delta/n)/2 + n\delta}$ for every $x \in G_{n-\Delta(p_\varepsilon)}^n$, or equivalently $\mathbb{P}_n(x)\, q^{n^2 S_2(\Delta/n)/2 - n\delta} \leq 1$. Then,

$$\begin{aligned}
|B_n| &\geq |B_n \cap G^n_{n-\Delta(p_\varepsilon)}| \\
&\geq \sum_{x \in B_n \cap G^n_{n-\Delta(p_\varepsilon)}} \mathbb{P}_n(x)\, q^{n^2 S_2(\Delta(p_\varepsilon)/n)/2 - n\delta} \\
&\geq q^{n^2 S_2(\Delta(p_\varepsilon)/n)/2 - n\delta} \mathbb{P}_n\left(B_n \cap G^n_{n-\Delta(p_\varepsilon)}\right) \\
&> q^{n^2 S_2(\Delta(p_\varepsilon)/n)/2 - n\delta} \frac{\xi}{\Delta(p_\varepsilon)+1}.
\end{aligned} \tag{8.13}$$

We deduce that

$$\liminf_n \frac{1}{n} \log_q |B_n| \geq \lim_n \frac{n}{2} S_2(\Delta(p_\varepsilon)/n) - \delta. \tag{8.14}$$

On the other hand, since $B_n \subset A_n$, it is clear that

$$\begin{aligned}
\limsup \frac{1}{n} \log_q |B_n| &\leq \lim_n \frac{1}{n} \log_q |A_n| \\
&= \lim_n \frac{n}{2} S_2(\Delta(p_\varepsilon)/n).
\end{aligned} \tag{8.15}$$

Since $\delta > 0$ is arbitrarily small, (8.14) and (8.15) imply that $\lim_n \frac{1}{n} \log_q |B_n|$ exists and equals $\Delta(p_\varepsilon)$. The theorem is proved. □

Remark 8.3. The definition of A_n still makes sense when p_ε is a discontinuity point of Δ. In this case, there exists $\xi > 0$ such that $\Delta(p_\varepsilon)+1 = \Delta(p_\varepsilon+\xi)$ and $\Delta(p_\varepsilon) = \Delta(p_\varepsilon - \xi)$. Inequality (8.5) can be easily adapted to show that $\sum_{k=0}^{\Delta(p_\varepsilon)+1} \mathrm{Gr}(n-k,n) \geq 1-\varepsilon$, which implies that $a_n \leq \Delta(p_\varepsilon)+1$; by (8.6), $a_n \geq \Delta(p_\varepsilon)$. Of course, part 2 in the Theorem still makes sense. We also have that $B_n \subset A_n$ and $a_n = b_n$. The problems appear in the comparison of $|B_n|$ and $|A_n|$; it is possible that $\mathbb{P}_n\left(B_n \cap \mathrm{Gr}(n-\Delta(p_\varepsilon),n)\right)$ goes to zero very fast when $n \to \infty$, and (8.14) is not valid any more. However, we can still adapt the bounds in (8.13) to prove

$$\begin{aligned}
\liminf_n \frac{1}{n} \log_q |A_n| &\geq \liminf_n \frac{1}{n} \log_q |B_n| \\
&\geq \lim_n \frac{1}{n} \log_q \begin{bmatrix} n \\ n-(\Delta(p_\varepsilon)-1) \end{bmatrix}_q \\
&= \Delta(p_\varepsilon) - 1,
\end{aligned}$$

because $b_n = a_n \geq \Delta(p_\varepsilon)$ and therefore $\mathrm{Gr}(n-(\Delta(p_\varepsilon)-1),n) \subset B_n$. Analogously, $B_n \subset A_n$ and $a_n \leq \Delta(p_\varepsilon)+1$ lead to

$$\begin{aligned}
\limsup_n \frac{1}{n} \log_q |B_n| &\leq \limsup_n \frac{1}{n} \log_q |A_n| \\
&\leq \lim_n \frac{1}{n} \log_q \begin{bmatrix} n \\ n-(\Delta(p_\varepsilon)+1) \end{bmatrix}_q \\
&= \Delta(p_\varepsilon) + 1,
\end{aligned}$$

where we have used again (8.8).

Remark 8.4. In the classical case of sequences, all the typical sequences tend to be equiprobable, in the sense of (0.24). This is not valid for the process V_n: a typical space $v \in A_n$ of dimension $n-d$ satisfy asymptotically the bounds $q^{-n(\frac{n}{2}S_2(d/n)+\delta)} \leq \mathbb{P}(V_n = v) \leq q^{-n(\frac{n}{2}S_2(d/n)-\delta)}$, for any $\delta > 0$, and $\frac{n}{2}S_2(d/n) = d + O(1/n)$.

8.3 Coding

Inspired by [23], we define a generalized n-to-k q-ary block code as a pair of mappings $f : \mathrm{Gr}(n) \to \{1, ..., q\}^k$ and $\phi : \{1, ..., q\}^k \to \mathrm{Gr}(n)$. For a given stochastic process W_n, such that W_n takes values in $\mathrm{Gr}(n)$, we define the probability of error of this code as $e(f,\phi) = \mathbb{P}\left(\phi(f(W_n)) \neq W_n\right)$. Small k and small probability of error are good properties for codes, but there is a trade-off between the two. Let $k(n,\varepsilon)$ be the smallest k such that there exists a generalized n-to-k q-ary block code (f,ϕ) that satisfies $e(f,\phi) \leq \varepsilon$.

Proposition 8.5. *For the Grassmanian process V_n introduced above and for all $\varepsilon > 0$ such that $p_\varepsilon = 1-\varepsilon$ is a continuity point of Δ, one has*

$$\lim_n \frac{k(n,\varepsilon)}{n} = \Delta(p_\varepsilon). \tag{8.16}$$

Proof. The existence of an n-to-k q-ary block code (f,ϕ) such that $e(f,\phi) \leq \varepsilon$ is equivalent to the existence of a set $B_n \subset \mathrm{Gr}(n)$ such that $\mathbb{P}(V_n \in B_n) \geq 1-\varepsilon$ and $|B_n| \leq q^k$ (let B_n be the set of sequences that are reproduced correctly...). As in the main theorem, let $s(n,\varepsilon)$ denote the minimum cardinality of such a set. The statement in Proposition 8.5 is therefore equivalent to $\lim_n \frac{1}{n}\log_q s(n,\varepsilon) = \Delta(p_\varepsilon)$, which is already proved. □

In simpler terms, it is always possible to code all the typical subspaces $A_n = \bigcup_{k=0}^{\Delta(p_\varepsilon)} \mathrm{Gr}(n-k,n)$ with different code-words if one disposes of $q^{n(\Delta(p_\varepsilon)+\xi)}$ such words, for ξ positive and arbitrarily small, as long as n is big enough. In contrast, it is asymptotically impossible if one disposes of $q^{n(\Delta(p_\varepsilon)-\xi')}$ different code-words, for any $\xi' > 0$.

8.4 Further remarks

A recent paper [44] proposes the study of "exploding" phase spaces: statistical systems such that the cardinality of the space of configurations grows faster than k^n, the combination of n components that can occupy k states. The total grassmannians $\mathrm{Gr}(n) = \mathrm{Gr}(n,\mathbb{F}_q)$ are an example, since their cardinality grows like $q^{\frac{n^2}{4}+o(n^2)}$. This can be deduced from the unimodality of the q-binomial coefficients (Lemma 6.2) and our asymptotic formulae, because

$$\begin{bmatrix} n \\ \lfloor n/2 \rfloor \end{bmatrix}_q \leq |\mathrm{Gr}(n)| \leq (n+1)\begin{bmatrix} n \\ \lfloor n/2 \rfloor \end{bmatrix}_q \tag{8.17}$$

and therefore

$$\begin{aligned}\lim_n \frac{2}{n^2}\log_q|\mathrm{Gr}(n)| &= \lim_n \frac{2}{n^2}\log_q \begin{bmatrix} n \\ \lfloor n/2 \rfloor \end{bmatrix}_q \\ &= S_2\left(\frac{1}{2},\frac{1}{2}\right) = \frac{1}{2}.\end{aligned} \tag{8.18}$$

In fact, the values of $\lim_{n\to\infty}|\mathrm{Gr}(2n+1)|q^{-(2n+1)^2/4}$ and $\lim_{n\to\infty}|\mathrm{Gr}(2n)|q^{-(2n)^2/4}$ depend only on q and can be determined explicitly in terms of the Euler's generating function for the partition numbers and the Jacobi theta functions ϑ_2 and ϑ_3, see [54, Cor. 3.7]

A link between Tsallis entropy and the size of the *effective* phase space (the configurations whose probability is nonzero) was already suggested by Tsallis in [88, Sec. 3.3.4]. There, $H_{(\rho-1)/\rho}$ appears naturally as a extensive quantity when the size of the effective phase space grows *sub-exponentially* as N^ρ, for certain $\rho > 0$. Nonetheless, these growth rates are not deduced from a combinatorial model.

Part IV

Information cohomology of continuous random variables

Chapter 9

Simplicial information structures

9.1 Definition and examples

Given a set $I := \{1, \ldots, n\} \equiv [n]$, the abstract simplex $\mathbf{\Delta(I)}$ is the category whose objects are subsets of I and whose arrows correspond to inclusions: $S \to T$ iff $T \subset S$. Let $\wedge$ and $\vee$ denote respectively the product and coproduct in $\mathbf{\Delta(I)}$, and $\cup$, $\cap$ the usual operations of sets. Given sets $S, T \in \operatorname{Ob}\mathbf{\Delta(I)}$, the product $S \wedge T$ equals $S \cup T$ and the coproduct $S \vee T$ is $S \cap T$.

A full subcategory $\mathbf{K}$ of $\mathbf{\Delta(I)}$ is called a ***simplicial subcomplex*** if $S \in \operatorname{Ob}\mathbf{K}$ implies that $T \in \operatorname{Ob}\mathbf{K}$, for every T such that $S \to T$ (this is called a 'face' of S).

Given a collection $\{E_i\}_{i \in I}$ of Hausdorff topological spaces, we introduce a functor $\mathscr{E} : \mathbf{S} \to \mathbf{Meas}$ that maps each $S \in \operatorname{Ob}\mathbf{K}$ to the measurable space $(E_S, \mathfrak{B}(E_S))$, where $E_S := \prod_{i \in S} E_i$ (remark that $E_\emptyset = \{*\}$) and $\mathfrak{B}_S \equiv \mathfrak{B}(E_S)$ denotes the corresponding Borel σ-algebra; the morpisms are canonical projections. An arrow $f : S \to T$ in $\mathbf{\Delta(I)}$—an inclusion of sets $T \subset S$—induces a canonical inclusion $\iota_{T,S} : E_T \hookrightarrow E_S$ and a canonical projection $\mathscr{E}(f) := \pi_{S,T} : E_S \to E_T$ in the category of topological spaces (hence measurable). For example, if $S = \{i_1, \ldots, i_k\}$ then

$$\begin{aligned} \pi_{I,S} :&E_I \to E_S \\ &(x_1, \ldots, x_n) \mapsto (x_{i_1}, \ldots, x_{i_k}). \end{aligned} \tag{9.1}$$

Let $\mathfrak{B}_S$ denote the Borel σ-algebra of E_S; this is the σ-algebra generated by the open sets of E_S. The projection $\pi_{S,T}$ induces an inclusion $f^* : \mathfrak{B}_T \hookrightarrow \mathfrak{B}_S$ that maps $b \in \mathfrak{B}_T$ to the corresponding cylinder $\pi_{S,T}^{-1}(b)$.

The pair $(\mathbf{K}, \mathscr{E})$ is a ***simplicial information structure***; it clearly verifies all the properties stated in Definition 1.6. Sometimes we write X_S instead of $S \in \operatorname{Ob}\mathbf{K}$ if we want to emphasize its interpretation as a random variable.

There is an additional property when each space E_i is second-countable.

Proposition 9.1. *Let S, T be subsets of I. Consider the diagram $S \xleftarrow{f_S} S \wedge T \xrightarrow{f_T} Y$ in $\mathbf{\Delta(I)}$. If the spaces $\{E_i\}_{i \in I}$ are second-countable, then*

$$\sigma(f_S^*(\mathfrak{B}_S), f_T^*(\mathfrak{B}_T)) = \mathfrak{B}_{S \wedge T},$$

where $\sigma(f_S^*(\mathfrak{B}_S), f_T^*(\mathfrak{B}_T))$ denotes the sub-σ-algebra of $\mathfrak{B}_{S \wedge T}$ generated by $f_S^*(\mathfrak{B}_S) \cup f_T^*(\mathfrak{B}_T)$.

Proof. Each set E_i has a countable basis $\mathfrak{G}_i$: this means that every open set can be written as a *countable* union of open sets in the basis, which in turn implies that $\mathfrak{B}(E_i)$, by definition generated by the topology of E_i, is in fact generated by $\mathfrak{G}_i$. More generally, the rectangles $\prod_{i\in S} G_i$, with $G_i \in \mathfrak{G}_i$, are a basis of the topology of the product space E_S (this is a general topological fact); these rectangles generate $\mathfrak{B}_S$, because every open set can be written as countable union of rectangles.

We have to show that $\mathfrak{B}_{S\wedge T} \subset \sigma(f_S^*(\mathfrak{B}_S), f_T^*(\mathfrak{B}_T))$, the opposite inclusion been trivial. Since every rectangle $\prod_{i\in S} G_i$ is an element of $\mathfrak{B}_S$, the corresponding cylinder $\pi_{S,S\cup T}^{-1}(\prod_{i\in S} G_i)$ is contained in $f_S^*(\mathfrak{B}_S)$. In fact, these cylinders generate $f_S^*(\mathfrak{B}_S)$. The σ-algebra $\sigma(f_S^*(\mathfrak{B}_S), f_T^*(\mathfrak{B}_T))$ contains then the intersections

$$\pi_{S,S\cup T}^{-1}\left(\prod_{i\in S} G_i\right) \cap \pi_{T,S\cup T}^{-1}\left(\prod_{j\in T} H_j\right) \tag{9.2}$$

that can be written as rectangles $\prod_{i\in S\cup T} U_i$, where $U_i = G_i \cap H_i$ if $i \in S \cap T$, $U_i = G_i$ if $i \in S \setminus T$, and $U_i = H_i$ if $i \in T \setminus S$; it is clear that we can obtain any rectangle in $E_{S\cup T}$ this way. Hence $\sigma(f_S^*(\mathfrak{B}_S), f_T^*(\mathfrak{B}_T))$ contains the generators of $\mathfrak{B}_{S\wedge T}$, therefore whole σ-algebra too. □

In this work, we limit ourselves to the second-countable case. The choice of certain ***reference measure*** μ_i on each measurable space $(E_i, \mathfrak{B}_i)$ induces a product measure $\bigotimes_{i\in S} \mu_i$ on $(E_S, \mathfrak{B}_S)$, for any S. We suppose that each μ_i, hence every measure involved, is σ-finite (see Proposition F.6): this is crucial to apply the disintegration theorems that give regular versions of conditional probabilities and also to prove functoriality of various constructions by using Fubini's theorem.

Here are some examples of simplicial information structures:

1. **Finite spaces:** For each $i \in I$, $\mathfrak{B}_i$ is a finite nontrivial Boolean algebra, each E_i is the set of its atoms (with the discrete topology), and μ_i is the counting measure. Each product $\bigotimes_{i\in S} \mu_i$, with $S \in \operatorname{Ob} \mathbf{\Delta(I)}$, gives again the counting measure: this explains why reference measures do not appear explicitly in the treatment of information cohomology of finite structures. [1]
2. **Euclidean spaces:** Each E_i is the real line; we suppose that some Lebesgue measure μ_i has been chosen. Each space $E_S = \bigoplus_{i\in S} E_i$ can be identified with the free vector space generated by S, in such a way that the unit hypercube in E_S has Lebesgue measure $\prod_{i\in S} \lambda_i([0,1])$.
3. **Topological groups:** This is a generalization of the previous example: each E_i is a Hausdorff, locally compact topological group and each λ_i is a chosen (left) Haar measure.

Given an arbitrary information structure $(\mathbf{S}, \mathscr{E})$, one can introduce a covariant functor $\Lambda : \mathbf{S} \to \mathbf{Sets}$ that associates to each $X \in \operatorname{Ob} \mathbf{S}$ the set $\Lambda(X)$ of measures on $\mathfrak{B}_X$; given an arrow $f : X \to Y$ in S, the morphism $\Lambda(f) : \Lambda(X) \to \Lambda(Y)$

[1] Already in the foundational paper by Shannon [78], there seems to be an important difference between discrete and continuous sources: the latter are studied by means of differential entropy, that depends explicitly on a reference measure. Shannon says: "In the discrete case the entropy measures in an *absolute way* the randomness of the chance variable. In the continuous case the measurement is *relative to the coordinate system.*" (Emphasis by Shannon.) We hope this text will make clear that such absolute character is illusory: even in the finite case, one could consider any other reference measure; for example, a multiple of the counting measure, see the remarks after Proposition 12.3.

(sometimes f to simplify notation) is given by the pre-composition of each measure $\lambda : \mathfrak{B}_X \to [0,1]$ with the map $f^* : \mathfrak{B}_Y \to \mathfrak{B}_X$ i.e. $f\lambda = \lambda \circ f^*$; this operation is called ***marginalization***. Like in Part I and II, we write Y_* instead of $\Lambda(f)$ if the map involved is clear from context. Probabilities correspond to a subfunctor Π of Λ, such that $\Pi(X) = \{ \lambda \in \Lambda(X) \,|\, \lambda(E_X) = 1 \}$; it is clear that, for every arrow $f : X \to Y$ in $\mathbf{S}$, $f\lambda(E_Y) = \lambda(f^*(E_Y)) = \lambda(E_X) = 1$.

9.2 Probabilities on information structures

9.2.1 Conditional probabilities

Before studying probabilities on information structures, we review some important facts about conditional probabilities and disintegration of measures.

Kolmogorov's definition

In this section, we introduce the standard modern definition of conditional probabilities. We mainly follow the original presentation in Kolmogorov's book [53]. Let $(E, \mathfrak{B}, P)$ be a measurable space. Given an event $A \in \mathfrak{B}$ such that $P(A) > 0$, define the conditional probability P_A by

$$\forall B \in \mathfrak{B}, \quad P_A(B) := \frac{P(A \cap B)}{P(A)}. \tag{9.3}$$

Clearly, $P_A(E) = 1$.

Given a partition of $\mathfrak{A} = \{A_1, ..., A_n\}$ of E (an *experiment* in Kolmogorov's terminology), one can introduce a random variable $P_{\mathfrak{A}}(B)$ that associates to each $e \in A_i$ the value $P_{A_i}(B)$. We call $P_{\mathfrak{A}}(B)$ ***the conditional probability of the event*** $B \in \mathfrak{B}$ ***after the experiment*** $\mathfrak{A}$. The function $P_{\mathfrak{A}}(B)$ is well defined only P-almost surely, but this is enough to define its integral with respect to P.

The condition

$$\forall A \in \mathfrak{A}, \quad P(A \cap B) = \int_A P_{\mathfrak{A}}(B) \, \mathrm{d}P \tag{9.4}$$

is satisfied in this finite case and uniquely characterizes the function $P_{\mathfrak{A}}(B)$. It holds true even if $\mathfrak{A}$ is replaced by the σ-algebra generated by it. Even better: it also makes sense when $\mathfrak{A}$ is an arbitrary sub-σ-algebra of $\mathfrak{B}$, maybe of infinite cardinality.

Definition 9.2 (Kolmogorov's definition of conditional probabilities)**.** Let $(E, \mathfrak{B}, P)$ be a probability space, $\mathfrak{A}$ a sub-σ-algebra of $\mathfrak{B}$, and B an event in $\mathfrak{B}$. The conditional probability of B given $\mathfrak{A}$, denoted $P_{\mathfrak{A}}(B)$, is an $\mathfrak{A}$-measurable function that satisfies (9.4).

It turns out that $P_{\mathfrak{A}}(B)$ always exists, as a consequence of the Radon-Nikodym theorem, and its unique up to P-almost sure equivalence [53, Ch. V]. [2]

The main problem with this definition is that nothing guarantees that $B \mapsto P_{\mathfrak{A}}(B)(\omega)$ is a probability measure for each $\omega \in \Omega$. When this is the case, $P_{\mathfrak{A}}$ is called a ***regular version*** of the conditional probability [71, Ch. 5].

[2] $P_{\mathfrak{A}}(B)$ is an special case of conditional expectation, $P_{\mathfrak{A}}(B) = \mathbb{E}(1_B|\mathfrak{A})$. We have chosen to follow here the original presentation by Kolmogorov, instead of the modern presentations that introduce first general conditional expectations; they are more economical but less motivated.

Disintegrations

Under very general topological hypothesis, it is possible to build regular versions of conditional probabilities called disintegrations. We summarize in this section the main facts about them as presented in [18].

Let $(E, \mathfrak{B})$ and $(E_T, \mathfrak{B}_T)$ be measurable spaces, and $T : (E, \mathfrak{B}) \to (E_T, \mathfrak{B}_T)$ a measurable map.

When E_T is finite and $\mathfrak{B}_T$ is its algebra of subsets 2^{E_T}, we can associate to any probability P on $(E, \mathfrak{B})$ a collection of maps $P_t : \mathfrak{B} \to \mathbb{R}, B \mapsto P_{\{T=t\}}(B)$, indexed by $t \in E_T$, such that

1. Each P_t is a probability on $(E, \mathfrak{B})$;
2. P_t concentrates on $\{T = t\}$, which means that $P_t(T \neq t) = 0$;
3. For any $B \in \mathfrak{B}$,
$$P(B) = \sum_{t \in E_T} P(T = t) P_t(B).$$

These properties motivate the following generalization.

Definition 9.3. Let λ a σ-finite measure on $(E, \mathfrak{B})$, and μ a σ-finite measure on $(E_T, \mathfrak{B}_T)$. The measure λ has a disintegration $\{\lambda_t\}_{t \in E_T}$ with respect to T and μ, or a (T, μ)-disintegration, if

1. λ_t is a σ-finite measure on $\mathfrak{B}$ concentrated on $\{T = t\}$, which means that $\lambda_t(T \neq t) = 0$ μ-almost surely.
2. for each measurable nonnegative function $f : E \to \mathbb{R}$,
   (a) $t \mapsto \int_E f \, \mathrm{d}\lambda_t$ is measurable; [3]
   (b) $\int_E f \, \mathrm{d}\lambda = \int_{E_T} \left(\int_E f(x) \, \mathrm{d}\lambda_t(x) \right) \mathrm{d}\mu(t)$.

Disintegrations give regular versions of conditional expectations. Let λ be a probability measure, $\mu = T\lambda$, and $\{\lambda_t\}$ the corresponding T-disintegration. Then the function $x \in E \mapsto \int_E \chi_B(x) \, \mathrm{d}\lambda_{T(x)}$ (where χ_B is the characteristic function) is σ_T measurable and a regular version of the conditional probability $\lambda_{\sigma(T)}(B)$. To prove this, let A be any element in $\sigma(T) \subset \mathfrak{B}$; it can be written as $\{T \in A_T\}$ for $A_T \in \mathfrak{B}_T$; the last property of the disintegration says that, for any $B \in \mathfrak{B}$

$$\begin{aligned}
\int_A \chi_B \, \mathrm{d}\lambda &= \int_E \chi_{A \cap B}(x) \, \mathrm{d}\lambda(x) \\
&= \int_{E_T} \int_E \chi_A(x) \chi_B(x) \, \mathrm{d}\lambda_t(x) \, \mathrm{d}T\lambda(t)
\end{aligned}$$

Using that $x \in A$ if and only if $T(x) \in A_T$, we deduce that

$$\begin{aligned}
\int_A \chi_B \, \mathrm{d}\lambda &= \int_{E_T} \chi_{A_T}(t) \left(\int_E \chi_B(x) \, \mathrm{d}\lambda_t(x) \right) \mathrm{d}T\lambda(t) \\
&= \int_A \left(\int_E \chi_B(x) \, \mathrm{d}\lambda_{T(x)}(x) \right) \mathrm{d}\lambda(x).
\end{aligned}$$

[3] Pollard and Chang [18] uses linear-functional notation for measures. The symbol $\lambda_t f$ stands for $\int_E f \, \mathrm{d}\lambda_t$. Superscripts next to measures emphasize the variable of integration; for example, $\int_{E_T} \left(\int_E f(x) \, \mathrm{d}\lambda_t(x) \right) \mathrm{d}\mu(t)$ could be written $\mu^t(\lambda_t f)$.

This equality is precisely (9.4), that defines the conditional expectation in Kolmogorov's sense. The fact that each $\lambda_t(-)$ is a probability comes from Proposition 9.5 below.

The following theorems are taken from [18].

Proposition 9.4 (Existence of disintegrations). *Let λ be a σ-finite measure on a metric space E and let T be a measurable map from E into $(E_T, \mathfrak{B}_T)$. Let μ be a σ-finite measure on $\mathfrak{B}_T$ that dominates the image measure $T\lambda$. If $\mathfrak{B}_T$ is countably generated and contains all the singletons $\{t\}$, then λ has a (T, μ)-disintegration. The λ_t measures are uniquely determined up to an almost sure equivalence: if $\{\lambda_t^*\}$ is another (T, μ)-disintegration then $\mu\{\, t \in E_T \mid \lambda_t \neq \lambda_t^* \,\} = 0$.*

Proposition 9.5. *Let λ have a (T, μ)-disintegration $\{\lambda_t\}$, with λ and μ each σ-finite.*

1. *The image measure $T\lambda$ is absolutely continuous with respect to μ, with density $\lambda_t E$.*
2. *The measures $\{\lambda_t\}$ are finite for μ-almost all t if and only if $T\lambda$ is σ-finite.*
3. *The measures $\{\lambda_t\}$ are probabilities for μ-almost all t if and only if $\mu = T\lambda$.*
4. *If $T\lambda$ is σ-finite then $0 < \lambda_t E < \infty$ $T\lambda$-almost surely, and the measures*

$$\tilde{\lambda}_t(\cdot) = \frac{\lambda_t(\cdot)}{\lambda_t E}$$

are probabilities that give a T-disintegration of λ.

Proposition 9.6. *Let λ have a (T, μ)-disintegration $\{\lambda_t\}$ and let ρ be absolutely continuous with respect to λ with finite density $r(x)$, with each λ, μ and ρ σ-finite.*

1. *The measure ρ has a (T, μ)-disintegration $\{\tilde{\rho}_t\}$ where each $\tilde{\rho}_t$ is dominated by the corresponding λ_t, with density $r(x)$.*
2. *The image $T\rho$ is absolutely continuous with respect to μ, with density $\int_E r \,\mathrm{d}\lambda_t$.*
3. *The measures $\{\tilde{\rho}_t\}$ are finite for μ-almost all t if and only if $T\rho$ is σ-finite.*
4. *The measures $\{\tilde{\rho}_t\}$ are probabilities for μ-almost all t if and only if $\mu = T\rho$.*
5. *If $T\rho$ is σ-finite then $0 < \lambda_t r < \infty$ $T\lambda$-almost surely, and the measures $\{\rho_t\}$ given by*

$$\int_E f \,\mathrm{d}\rho_t = \frac{\int_E f r \,\mathrm{d}\lambda_t}{\int_E r \,\mathrm{d}\lambda_t}$$

are probabilities that give a T-disintegration of ρ.

Example 9.7 (Product spaces). We suppose that $(E, \mathfrak{B}, \lambda)$ is the product of two measured spaces spaces $(E_T, \mathfrak{B}_T, \mu)$ and $(E_S, \mathfrak{B}_S, \nu)$, with μ and ν both σ-finite. Let λ_t be the image of ν under the inclusion $s \mapsto (t, s)$. Then Fubini's theorem implies that λ_t is a (T, μ)-disintegration of λ. (Remark that $\mu \neq T\lambda$. In general, the measure $T\lambda$ is not even σ-finite.) If $r(t, s)$ is the density of a probability ρ on $(E, \mathfrak{B})$, then $\rho_t \ll \lambda_t$ with density $r(t, s)$—the value of t being fixed—and $\tilde{\rho}_t$ is a probability supported on $\{T = t\}$ with density

$$\frac{r(t,s)}{\int_{E_S} r(t,s)\,\mathrm{d}\nu(s)}.$$

9.2.2 Densities under conditioning and marginalization

Let $\Pi : \mathbf{S} \to \mathbf{Sets}$ be the functor of probabilities introduced at the end of Section 9.1. Whenever a probability $\rho \in \Pi(S)$ (that represents a possible 'law' of X_S) is absolutely continuous with respect to the reference measure μ_S, denoted $\rho \ll \mu_S$, the Radon-Nikodym theorem (Proposition F.1) guarantees the existence of a function $f_\rho \in L^1(E_S, \mathfrak{B}_S, \mu_S)$ such that

$$\rho(A) = \int_A f_\rho(x)\,\mathrm{d}\mu_S(x). \tag{9.5}$$

The function f_ρ is known as the probability density function (pdf) of X_S. We summarize the relation between ρ, μ_S and f_ρ in (9.5) by $\rho = f_\rho \mu_S$.

Now we describe the marginalization in terms of densities. Given $\rho \in \Pi(S)$ with $\rho = f_S \mu_S$, an arrow $\pi_{T,S} : S \to T$ in $\mathbf{S}$, and $B \in \mathfrak{B}_T$, one has

$$\pi_{T,S}\rho(B) \overset{\text{(def)}}{=} \rho(\pi^*_{T,S}(B)) \tag{9.6}$$

$$= \int_{\pi^{-1}_{T,S}(B)} f_S(x)\,\mathrm{d}\mu(x) \tag{9.7}$$

$$\overset{\text{(Fubini)}}{=} \int_B \left(\int_{E_{S\setminus T}} f_S(x_T, x_{S\setminus T})\,\mathrm{d}\mu_{S\setminus T}(x_{S\setminus T}) \right) \mathrm{d}\mu_T(x_T), \tag{9.8}$$

which means that $\pi_{T,S}\rho$ has density $\int_{E_{S\setminus T}} f_{X_S}(x_T, x_{S\setminus T})\,\mathrm{d}\mu_{S\setminus T}(x_{S\setminus T})$ with respect to the reference measure μ_T. (The use of Fubini's theorem F.7 is justified by the positivity of the densities and the fact that each measure involved is σ-finite.)

The description of ***conditioning*** is more involved. Consider sets $T \subset S \subset I$, and the corresponding surjection $\pi_{T,S} : E_S \to E_T$. The measure μ_S has a $(\pi_{T,S}, \mu_T)$-disintegration $\{\mu_{S,t}\}_{t\in E_T}$ such that each measure $\mu_{S,t}$ is the image of $\mu_{S\setminus T}$ under the inclusion $E_{S\setminus T} \hookrightarrow E_S, s' \mapsto (s', t)$, that can be identified with the product measure $\mu_{S\setminus T} \otimes \delta_{T=t}$. See Example 9.7 in Section 9.2.1.

More generally, we can start with a reference measure $\mu_{S'} \otimes \delta_{S''=s''}$ on E_S, for certain S' and S'' that form a partition of S. Set $T' = S' \cap T$, $T'' = S'' \cap T$, and denote by $\mu_{T'} \otimes \delta_{T''=\pi(s'')}$ the measure on $E_T = E_{T'} \times E_{T''}$ concentrated on $\{T'' = \pi_{T'',S''}(s'')\} := E_{T'} \times \{\pi_{T'',S''}(s'')\}$. [4]

Proposition 9.8. *The $(\pi_{T,S}, \mu_{T'} \otimes \delta_{T''=\pi(s'')})$-disintegration of $\mu_{S'} \otimes \delta_{S''=s''}$, denoted by $\{(\mu_{S'} \otimes \delta_{S''=s''})_t\}_{t\in E_T}$, verifies*

$$(\mu_{S'} \otimes \delta_{S''=s''})_t = \mu_{S'\setminus T} \otimes \delta_{T'=t'} \otimes \delta_{S''=s''} \tag{9.9}$$

whenever $t = (t', \pi_{T'',S''}(s'')) \in E_{T'} \times E_{T''} = E_T$, and this determines it up to $\mu_{T'} \otimes \delta_{T''=\pi(s'')}$-almost sure equivalence.[5]

Proof. We prove that the given collection of measures satisfy all the conditions in Definition 9.3. The disintegration is almost sure unique according to Proposition

[4] Usually we omit the subscripts of π if they are clear from context.

[5] Remark that $(\mu_{S'} \otimes \delta_{S''=s''})_t$ is well-defined only $\mu_{T'} \otimes \delta_{T''=\pi(s'')}$-almost surely. Its value on $\{T'' \neq \pi(s'')\}$ is immaterial. The same remark is relevant for the statement of Proposition 9.9.

9.4. Remark that the specific value of the disintegration on $\{T'' \neq \pi(s'')\}$ is not well-defined.

The measure $\mu_{S'\setminus T} \otimes \delta_{T'=t'} \otimes \delta_{S''=s''}$ is σ-finite, being a product of σ-finite measures (see Proposition F.6). The union bound implies its concentration on the corresponding level set:

$$\begin{aligned}
&(\mu_{S'\setminus T} \otimes \delta_{T'=t'} \otimes \delta_{S''=s''})\{T \neq (t', \pi(s''))\} \\
&\qquad\qquad = (\mu_{S'\setminus T} \otimes \delta_{T'=t'} \otimes \delta_{S''=s''})(\{T' \neq t'\} \cup \{T'' \neq \pi(s'')\}) \\
&\qquad\qquad \leq \mu_{S'\setminus T}(E_{S'\setminus T})\delta_{T'=t'}(T' \neq t')\delta_{S''\neq s''}(E_{S''}) \\
&\qquad\qquad\quad + \mu_{S'\setminus T}(E_{S'\setminus T})\delta_{T'=t'}(E_{T'})\delta_{S''\neq s''}(\pi(S'') \neq \pi(s'')) = 0.
\end{aligned}$$

We prove now the disintegration property, that says: for every $\mathfrak{B}_S$-measurable function f,

$$\int_E f \,\mathrm{d}\mu_{S'} \otimes \delta_{S''=s''} = \int_{E_T} \left(\int_E f \,\mathrm{d}(\mu_{S'} \otimes \delta_{S''=s''})_t \right) \mathrm{d}\mu_{T'} \otimes \delta_{T''=\pi(s'')} \tag{9.10}$$

According to (9.9), for $t \in E_{T'} \times \{\pi_{T'',S''}(s'')\}$,

$$\int_E f \,\mathrm{d}(\mu_{S'} \otimes \delta_{S''=s''})_t = \int_{E_{S'\setminus T}} f(z, t', s'') \,\mathrm{d}\mu_{S'\setminus T}(z). \tag{9.11}$$

Hence

$$\begin{aligned}
&\int_{E_T} \left(\int_E f \,\mathrm{d}(\mu_{S'} \otimes \delta_{S''=s''})_t \right) \mathrm{d}\mu_{T'} \otimes \delta_{T''=\pi(s'')}(t) \\
&\quad = \int_{E_{T'}\times\{\pi_{T'',S''}(s'')\}} \left(\int_{E_{S'\setminus T}} f(z, t', s'') \,\mathrm{d}\mu_{S'\setminus T}(z) \right) \mathrm{d}\mu_{T'} \otimes \delta_{T''=\pi(s'')}(t) \\
&\quad = \int_{\{\pi_{T'',S''}(s'')\}} \left(\int_{E_{T'}} \int_{E_{S'\setminus T}} f(z, t', s'') \,\mathrm{d}\mu_{S'\setminus T}(z) \,\mathrm{d}\mu_{T'}(t') \right) \delta_{T''=\pi(s'')}(t'') \\
&\quad = \int_{E_{S'}} f(w, s'') \,\mathrm{d}\mu_{S'}(w).
\end{aligned}$$

The first equality is justified by $\mu_{T'} \otimes \delta_{T''=\pi(s'')}(E_{T'} \times \{\pi_{T'',S''}(s'')\}^c) = 0$. The second and the third are a consequence of Fubini's theorem for positive functions (Proposition F.7). Recall that, according to the definition of a simplicial information structure, each μ is σ-finite, $E_{S'} = E_{S'\setminus T} \times E_{T'}$, and $\mu_{S'} = \mu_{S'\setminus T} \otimes \mu_{T'}$. This proves (9.10).

Finally, Proposition F.8 establishes the measurability of

$$(t', s'') \mapsto \int_{E_{S'\setminus T}} f(z, t', s'') \,\mathrm{d}\mu_{S'\setminus T}(z).$$

Hence $(t', \pi(s'')) \to \int_{E_{S'\setminus T}} f(z, t', s'') \,\mathrm{d}\mu_{S'\setminus T}(z)$ is measurable too, in virtue of Proposition F.5. □

Proposition 9.9. *Let ρ be a probability on $(E_S, \mathfrak{B}_S)$, absolutely continuous with respect to $\mu_{S'} \otimes \delta_{S''=s''}$ with density f_ρ. Let U, T be subsets of S. Then:*

1. *ρ has a $(\pi_{T,S}; \mu_{T'} \otimes \delta_{T''=\pi(s'')})$-disintegration $\{\tilde{\rho}_t\}_{t\in E_T}$ such that*

$$\tilde{\rho}_{(t',\pi(s''))} = f_\rho \mu_{S'\setminus T} \otimes \delta_{T'=t'} \otimes \delta_{S''=s''},$$

2. *ρ has a $\pi_{T,S}$-disintegration $\{\rho_t\}$ such that ρ_t is a probability measure $\pi_{T,S}\rho$-almost surely and*

$$\rho_{(t',\pi(s''))} = \frac{f_\rho}{\int_{E_{S'\setminus T}} f_\rho(x,t',s'')\,\mathrm{d}\mu_{S'\setminus T}(x)} \mu_{S'\setminus T} \otimes \delta_{T'=t'} \otimes \delta_{S''=s''},$$

3. *Let $\{\rho_w\}_{w\in E_{T\cup U}}$ be a $\pi_{T\cup U,S}$-disintegration of ρ. The equality*

$$\rho_w = (\rho_{t(w)})_{u(w)},$$

where $w \in E_{T\cup U}$, $t(w) := \pi_{T,T\cup U}(w)$ and $u(w) := \pi_{U,T\cup U}(w)$, holds $\mu_{(T\cup U)'} \otimes \delta_{(T\cup U)''=\pi(s'')}$-almost surely.

4. *Let φ be a nonnegative real-valued $\mathfrak{B}_T \otimes \mathfrak{B}_U$-measurable function. The equality*

$$\int_{E_T}\int_{E_U} \varphi(t,u)\,\mathrm{d}U_*\rho_t(u)\,\mathrm{d}T_*\rho(t) = \int_{E_{T\cup U}} \varphi(\iota(w))\,\mathrm{d}(T\wedge U)_*\rho(w), \tag{9.12}$$

where $\iota : E_{T\cup U} \to E_T \times E_U, w \mapsto (t(w), u(w))$ holds.

Proof.
Claim (1): Since ρ is absolutely continuous with respect to $\mu_{S'} \otimes \delta_{S''=s''}$, with density f_ρ, Propositions 9.6 and 9.8 imply that ρ has a $(\pi_{T,S}; \mu_{T'} \otimes \delta_{T''=\pi(s'')})$-disintegration $\{\tilde{\rho}_t\}_{t\in E_T}$ such that each $\rho_{(t',\pi(s''))}$ is absolutely continuous with respect to $\mu_{S'\setminus T} \otimes \delta_{T'=t'} \otimes \delta_{S''=s''}$ with Radon-Nikodym derivative f_ρ.

Claim (2): It is an immediate consequence of Proposition 9.8-(5).

Claim (3): Set $W := T \cup U$, $W' := W \cap S'$, and $W'' := W \cap S''$. Since $\mu_{W'} \otimes \delta_{W''=\pi_{W'',S''}(s'')}(E_{W'} \times \{\pi_{W'',S''}(s'')\}^c) = 0$, we assume that $w = (w', \pi(s'')) \in E_{W'} \times E_{W''} = E_W$. In virtue of (2), ρ_w equals

$$f_{\rho_w}\mu_{S'\setminus W} \otimes \delta_{W'=w'} \otimes \delta_{W''=\pi(s'')} = \frac{f_\rho}{\int_E f_\rho\,\mathrm{d}\mu_{S'\setminus W} \otimes \delta_{W''=\pi(s'')}} \mu_{S'\setminus W} \otimes \delta_{W'=w'} \otimes \delta_{W''=\pi(s'')}. \tag{9.13}$$

Define $t(w) = \pi_{T,W}(w)$, $t(w)' = \pi_{T',W'}(w)$, $t(w)'' = \pi_{T'',W''}(w)$, and similarly $u(w)$, $u(w)'$, $u(w)''$. By functoriality $t(w)'' = \pi_{T'',W''}(w'') = \pi_{T'',S''}(s'')$.

We apply now the result in (2) to the measure $\rho_{t(w)}$, coming from the $\pi_{T,S}$-disintegration of ρ. It is absolutely continuous with respect to the measure $\mu_{S'\setminus T} \otimes \delta_{T'=t(w)'} \otimes \delta_{S''=s''}$, with density

$$f_{\rho_{t(w)}} = \frac{f_\rho}{\int_{E_{S'\setminus T}}(x, t(w)', s'')\,\mathrm{d}\mu_{S'\setminus T}(x)}. \tag{9.14}$$

Remark that the measure $\mu_{S'\setminus T} \otimes \delta_{(T',S'')=(t(w)',s'')}$ can be disintegrated according Proposition 9.8. Define $\tilde{U}'' := (T' \cup S'') \cap U$, $\tilde{U}' := U \setminus \tilde{U}''$, and $\tilde{u}(w)' := \pi_{\tilde{U}',S}(w)$. We apply again the result of (2), *mutatis mutandis*, to say that the $\pi_{U,S}$-disintegration of $\rho_{t(w)}$ verifies, for $u = u(w) = (\tilde{u}(w)', \pi_{\tilde{U}'',T'\cup S''}(t(w)', s''))$,

$$(\rho_{t(w)})_{u(w)} \ll \mu_{(S'\setminus T)\setminus U} \otimes \delta_{\tilde{U}'=\tilde{u}(w)'} \otimes \delta_{(T',S'')=(t(w)',s'')}.$$

and corresponding density is

$$\frac{f_{\rho_{t(w)}}}{\int_E f_{\rho_{t(w)}} \,\mathrm{d}\mu_{(S'\setminus T)\setminus U} \otimes \delta_{\tilde{U}'=\tilde{u}(w)'} \otimes \delta_{(T',S'')=(t',s'')}} = \frac{f_\rho}{\int_E f_\rho(x, \tilde{u}(w)', t(w)', s'') \,\mathrm{d}\mu_{(S'\setminus T)\setminus U}(x)}. \tag{9.15}$$

Remark now that the disjoint union $\tilde{U}' \sqcup T'$ equals $(T \cup U) \cap S' = W'$, therefore $E_{W'} = E_{\tilde{U}'} \times E_{T'}$ and $(\tilde{u}(w)', t(w)')$ corresponds to the initial w'. The equality

$$\mu_{(S'\setminus T)\setminus U} \otimes \delta_{\tilde{U}'=\tilde{u}(w)'} \otimes \delta_{(T',S'')=(t(w)',s'')} = \mu_{S'\setminus W} \otimes \delta_{W=w'} \otimes \delta_{S''=s''},$$

allow us to rewrite (9.15) and establish that

$$(\rho_{t(w)})_{u(w)} = \frac{f_\rho}{\int_E f_\rho(x, w', s'') \,\mathrm{d}\mu_{S'\setminus W}(x)} \mu_{S'\setminus W} \otimes \delta_{W=w'} \otimes \delta_{S''=s''}$$

that is exactly the expression in (9.13).

CLAIM (4): Since $T_*\rho \ll \mu_{T'} \otimes \delta_{T''=\pi(s'')}$, we can restrict the domain of integration:

$$\int_{E_T} \int_{E_U} \varphi(t, u) \,\mathrm{d}U_*\rho_t(u) \,\mathrm{d}T_*\rho(t) = \int_{E_{T'} \times \{\pi_{T'',S''}(s'')\}} \left(\int_{E_U} \varphi(t, u) \,\mathrm{d}U_*\rho_t(u) \right) \mathrm{d}T_*\rho(t). \tag{9.16}$$

In (2) we obtained an explicit formula for ρ_t appearing in the last integral,

$$\rho_{(t',\pi(s''))} = \frac{f_\rho}{K(t', s'')} \mu_{S'\setminus T} \otimes \delta_{T'=t'} \otimes \delta_{S''=s''},$$

with $K(t', s'') := \int_{E_{S'\setminus T}} f_\rho(x, t', s'') \,\mathrm{d}\mu_{S'\setminus T}(x)$. The function $K(t', s'')$ is the density of $T_*\rho$ (at least on the probability 1 set $E_{T'} \times \{\pi(s'')\}$).

Let us write $\mu_{S'\setminus T} \otimes \delta_{T'=t'} \otimes \delta_{S''=s''}$ in the form $\mu_{\tilde{S}'} \otimes \mu_{\tilde{S}''=\pi(t',s'')}$, with $\tilde{S}' = S' \setminus T$ and $\tilde{S}'' = T' \sqcup S''$. Then $\tilde{U} = \tilde{S}' \cap U$ and $\tilde{U}'' = \tilde{S}'' \cap U$ coincide with the sets already introduced in the proof of (3). The measure $U_*\rho_t$ is absolutely continuous with respect to $\mu_{\tilde{U}'} \otimes \delta_{\tilde{U}''=\pi(t',s'')}$, where $\pi(t', s'') := \pi_{\tilde{U}'',T'\sqcup S''}(t', s'')$. Moreover,

$$\begin{aligned}
&U_*\rho_t(B) \\
&= \int_{\pi^{-1}(B)} \frac{f_\rho}{K(t', s'')} \,\mathrm{d}\mu_{\tilde{S}'} \otimes \delta_{\tilde{S}''=(t',s'')} \\
&= \int_B \left(\int_{E_{\tilde{S}'\setminus U} \times E_{\tilde{S}''\setminus U}} \frac{f_\rho}{K(t', s'')} \,\mathrm{d}\mu_{\tilde{S}'\setminus U} \otimes \delta_{\tilde{S}''\setminus U=\pi_{\tilde{S}''\setminus U,\tilde{S}''}(t',s'')} \right) \mathrm{d}\mu_{\tilde{U}'} \otimes \delta_{\tilde{U}''=\pi(t',s'')}(\tilde{u}) \\
&= \int_B \left(\int_{E_{S'\setminus W}} \frac{f_\rho(x, \tilde{u}, \pi_{\tilde{S}''\setminus U,\tilde{S}''}(t', s''))}{K(t', s'')} \,\mathrm{d}\mu_{S'\setminus W}(x) \right) \mathrm{d}\mu_{\tilde{U}'} \otimes \delta_{\tilde{U}''=\pi(t',s'')}(\tilde{u}).
\end{aligned}$$

We conclude that the density of $U_*\rho_t$ w.r.t. $\mu_{\tilde{U}'} \otimes \delta_{\tilde{U}''=\pi(t',s'')}$ is

$$\int_{E_{S'\setminus W}} \frac{f_\rho(x, \tilde{u}, \pi_{\tilde{S}''\setminus U, \tilde{S}''}(t', s''))}{K(t', s'')} \, \mathrm{d}\mu_{S'\setminus W}(x). \tag{9.17}$$

The function

$$g(\tilde{u}|t') := \int_{E_{S'\setminus W}} f_\rho(x, \tilde{u}, \pi_{\tilde{S}''\setminus U, \tilde{S}''}(t', s''))\, \mathrm{d}\mu_{S'\setminus W}(x),$$

correspond to the product of the density of $U_*\rho_t$ | in (9.17) | and the density of $T_*\rho$.
Hence the integral in (9.16) equals

$$\int_{E_{T'}\times\{\pi_{T'',S''}(s'')\}} \left(\int_{E_{\tilde{U}'}\times E_{\tilde{U}''}} \varphi(t, \tilde{u}', \tilde{u}'') g(\tilde{u}', \tilde{u}''|t')\, \mathrm{d}\mu_{\tilde{U}'} \otimes \delta_{\tilde{U}''=\pi(t',s'')}(\tilde{u}', \tilde{u}'') \right) \mathrm{d}\mu_{T'} \otimes \delta_{T''=\pi(s'')}(t', t'')$$
$$= \int_{E_{T'}\times\{\pi_{T'',S''}(s'')\}} \left(\int_{E_{\tilde{U}'}} \varphi(t, \tilde{u}', \pi_{\tilde{U}'',\tilde{S}''}(t', s''))g(\tilde{u}', \pi_{\tilde{U}'',\tilde{S}''}(t', s'')|t')\, \mathrm{d}\mu_{\tilde{U}'} \right) \mathrm{d}\mu_{T'} \otimes \delta_{T''=\pi(s'')}(t', t'')$$
$$= \int_{E_{T'}} \int_{E_{\tilde{U}'}} \varphi(t', \pi_{T''}(s''), \tilde{u}', \pi_{T'\cap U}(t'), \pi_{U''}(s''))g(\tilde{u}', \pi_{T'\cap U}(t'), \pi_{U''}(s'')|t')\, \mathrm{d}\mu_{\tilde{U}'}(\tilde{u}')\, \mathrm{d}\mu_{T'}(t')$$

where we have used Fubini's theorem two times, first for the inner integral and then for the outer one. Recall now that $W' = (T \cup U) \cap S' = T' \sqcup \tilde{U}'$, $E_{W'} \cong E_{T'} \times E_{\tilde{U}'}$ and $\mu_{W'} = \mu_{T'} \otimes \mu_{\tilde{U}'}$. Therefore Fubini again allows us to write

$$\int_{E_{W'}} \varphi(\pi_{T'}(w')\pi_{T''}(s''), \pi_{\tilde{U}'}(w'), \pi_{U''}(s''))g(\pi_{\tilde{U}'}(w'), \pi_{U\cap T'}(w'), \pi_{U''}(s'')|\pi_{T'}(w'))\, \mathrm{d}\mu_{W'}(w')$$
$$= \int_{E_{W'}} \varphi(\iota(w))\, \mathrm{d}(T \cup U)_*\rho(w)$$

where the last equality is justified by the fact that

$$g(\pi_{\tilde{U}'}(w'), \pi_{U\cap T'}(w'), \pi_{U''}(s'')|\pi_{T'}(w')) = \int_{S'\setminus W} f_\rho(x, w', \pi_{W''}(s''))\, \mathrm{d}\mu_{S'\setminus W}(x)$$

is the density of $(T \cup U)_*\rho(w)$ with respect $\mu_{W'} \otimes \delta_{W''=\pi(s'')}$. □

9.3 Probabilistic functionals

For given S, let $\Pi(S, \nu)$ denote the set of probability laws on $(E_S, \mathfrak{B}_S)$ absolutely continuous with respect to the measure ν, and $\mathscr{R}(S)$ the set of ***simplicial reference measures***,

$$\mathscr{R}(S) = \{\, \mu \in \Lambda(S) \mid \mu = \mu_{S'} \otimes \delta_{S''=a} \text{ for some } S', S'' \text{disjoint such that } S = S' \cup S'' \,\}. \tag{9.18}$$

and finally $\mathscr{Q}(S)$ the set of probability measures absolutely continuous with respect to a given reference measure,

$$\mathscr{Q}(S) = \{\, (\mu, \rho) \in \Lambda(S) \times \Pi(S) \mid \mu = \mu_{S'} \otimes \delta_{S''=a} \text{ for some } S', S'' \text{disjoint such that } S = S' \cup S'', \text{and } \rho \ll \mu \,\}. \tag{9.19}$$

Equivalently,

$$\mathscr{Q}(S) := \bigcup_{\mu \in \mathscr{R}(S)} \{\mu\} \times \Pi(S, \mu) \tag{9.20}$$

$\mathscr{Q}$ is a functor. An arrow $\pi_{T,S} : S \to T$, with the corresponding surjection $\pi_{T,S} : E_S \to E_T$, induces a map $\mathscr{Q}(\pi_{T,S}) : \mathscr{Q}(S) \to \mathscr{Q}(T)$ that sends $(\mu_{S'} \otimes \delta_{S''=s''}, \rho)$ to $(\mu_{T'} \otimes \delta_{T''=\pi(a)}, \pi_{T,S*}\rho)$, where $T' = S' \cap T$, $T'' = S'' \cap T$, and $\pi(a) := \pi_{T'',S''}(s'')$. It is important to remark that, for any $A \in \mathfrak{B}_T$, if $(\mu_{T'} \otimes \delta_{T''=\pi(s'')})(A) = 0$, then $(\mu_{S'} \otimes \delta_{S''=s''})(\pi_{T,S}^{-1}(A)) = 0$ in virtue of the disintegration formula; therefore $\rho(\pi_{T,S}^{-1}(A)) = 0$ as a consequence of the absolute continuity imposed on the definition of $\mathscr{Q}$, and $\pi_{T,S*}\rho(A)$ is also zero. This means that $(\mu_{T'} \otimes \delta_{T''=t(a)}, \pi_{T,S*}\rho)$ effectively lies in $\mathscr{Q}(T)$.

Let $\mathscr{F}_+(S)$ be the set of measurable nonnegative functions $\varphi : \mathscr{Q}(S) \to \mathbb{R}$. For every $T \in \mathscr{S}_S$ and $\varphi \in \mathscr{F}_+(S)$, set

$$(T.\varphi)(\mu_{S'} \otimes \delta_{S''=s''}, \rho) := \int_{E_T} \varphi(\mu_{S'\setminus T} \otimes \delta_{T'=t'} \otimes \delta_{S''=s''}, \rho_t) \,\mathrm{d}T_*\rho(t), \tag{9.21}$$

where $t = (t', t'') \in E_T$, $T_*\rho$ is a probability on E_T absolutely continuous with respect to $\mu_{T'} \otimes \delta_{T''=\pi_{T'',S''}(s'')}$, with density $\int_{E_{S'\setminus T}} f_P(z, y', x'')\mu_{S'\setminus T}(z)$, and $\{\rho_t\}$ is a $\pi_{T,S}$-disintegration of ρ.

Proposition 9.10. *For any $T, U \in \mathscr{S}_S$ and $\varphi \in \mathscr{F}_+(S)$,*

$$T.(U.\varphi) = (T \cup U).\varphi. \tag{9.22}$$

Proof. The iterated application of the definition gives

$$\begin{aligned}
&(T.(U.\varphi))(\mu_{S'} \otimes \delta_{S''=s''}, \rho) \\
&\quad = \int_{E_T} (U.\varphi)(\mu_{S'\setminus T} \otimes \delta_{T'=t'} \otimes \delta_{S''=s''}, \rho_t) \,\mathrm{d}T_*\rho(t) \\
&\quad = \int_{E_T} \left(\int_{E_U} \varphi(\mu_{(S'\setminus T)\setminus U} \otimes \delta_{\tilde{U}'=\tilde{u}'} \otimes \delta_{(T',S'')=(t',s'')}, (\rho_t)_u) \,\mathrm{d}U_*\rho_t(u) \right) \mathrm{d}T_*\rho(t)
\end{aligned}$$

In virtue of Proposition 9.9-(4), the last integral equals

$$\int_{E_W} \varphi(\mu_{(S'\setminus T)\setminus U} \otimes \delta_{\tilde{U}'=\tilde{u}(w)'} \otimes \delta_{(T',S'')=(t(w)',s'')}, (\rho_{t(w)})_{u(w)}) \,\mathrm{d}(T \cup U)_*\rho(w).$$

But $(S' \setminus T) \setminus U = S' \setminus W$, $\tilde{U}' \sqcup T' = W'$, and $(\rho_{t(w)})_{u(w)} = \rho_w$ almost surely (Proposition 9.9), therefore

$$(T.(U.\varphi))(\mu_{S'} \otimes \delta_{S''=s''}, \rho) = \int_{E_W} \varphi(\mu_{S'\setminus W} \otimes \delta_{W'=w'} \otimes \delta_{S''=s''}, \rho_w) \,\mathrm{d}(T \cup U)_*\rho(w),$$

as we wanted to prove. □

In other words: equation (9.21) defines a monoid action of $\mathscr{S}_S$ on $\mathscr{F}_+(S)$.

9.4 Restriction to Gaussian laws; functional module

We make now the following hypotheses: each $E_i = \mathbb{R}$, in such a way that $E_S \cong \mathbb{R}^{|S|}$ taking S itself as the canonical basis. Each reference measure μ_S is the corresponding Lebesgue measure on E_S, normalized to assign unit measure to the hypercube. Moreover, let us introduce a sub-sheaf of $\mathcal{Q}$ defined by

$$\begin{aligned}\mathcal{Q}_{\text{Gauss}}(S) := \{\, (\mu,\rho) \in \Lambda(S)\times\Pi(S) \mid \mu = \mu_{S'}\otimes\delta_{S''=a} \text{ for some}\\ S', S'' \text{disjoint such that } S = S'\cup S'',\ \rho \ll \mu, \text{ and } \rho \text{ gaussian}\,\}.\end{aligned} \tag{9.23}$$

A gaussian (or normal) probability distribution is defined as a probability on E_S such that, for every linear functional $f : E_S \to \mathbb{R}$, the push-forward $f_*\rho$ is a univariate normal. Denote by $(e_1,\dots,e_n)$ the base of E_S and let $(e_1^*,\dots,e_n^*)$ be the dual basis. Each ρ is uniquely characterized by a mean vector $m \in E_S$, whose components are $m_i = \int x\, \mathrm{d}e_i^*\rho(x)$, and a covariance matrix $\Sigma = (\sigma_{ij})_{1\le i,j\le |S|}$ such that $\sigma_{ij} = \int\int (x-\mu_i)(y-\mu_j)\,\mathrm{d}e_i^*\rho(x)\,\mathrm{d}e_j^*\rho(y)$. See Appendix D . As a consequence we also denote the elements of $\mathcal{Q}_{\text{Gauss}}(S)$ by triplets (μ, m, Σ).[6]

Given a reference measure $\mu_{S'}\otimes\delta_{S''=s''}$, we say that the covariance matrix is *admissible* if the eigenvectors associated to nonzero eigenvalues span $E_{S'}\subset E_S$. If this is the case, a gaussian law with covariance Σ is absolutely continuous with respect to $\mu_{S'}$, cf. Proposition D.7.

Let $\mathcal{F}_{\text{Gauss}}(S)$ be the additive abelian group of measurable functions $\varphi : \mathcal{Q}_{\text{Gauss}}(S) \to \mathbb{R}$ that verify the following ***polynomial-growth condition***:

for every reference measure $\mu_{S'}\otimes\delta_{S''=s''}$, every admissible covariance matrix Σ, and every variable T coarser than S, there exist $C > 0$ and $\gamma > 0$ such that (9.24)

$$|\varphi(\mu_{S'\setminus T}\otimes\delta_{T'=t'}\otimes\delta_{S''=s''}, m, \Sigma)| \le C(1+\|t\|+\|m\|)^{\gamma}.$$

Proposition 9.11. *Equation* (9.21) *defines an action of the monoid* $\mathcal{S}_S$ *on the vector space* $\mathcal{F}_{\text{Gauss}}(S)$*: for every* $T, U \in \mathcal{S}_S$ *and every* $\varphi \in \mathcal{F}_{\text{Gauss}}(S)$,

$$U.(T.\varphi) = TU.\varphi.$$

It extends linearly to an action of $\mathbb{R}[\mathcal{S}_S]$, *the monoid algebra.*

Proof. Decompose φ into its positive and negative part, in such a way that $\varphi = \varphi^+ - \varphi^-$, and $|\varphi| = |\varphi^+| + |\varphi^-|$. The functions φ^+ and φ^- belong to $\mathcal{F}_+(S)$ and verify the condition (9.24).

Consider $T \in \mathcal{S}_S$. The random variable T corresponds to certain components of the variable S, and its marginal law $T_*\rho$ is absolutely continuous with respect to $\mu_{T'}\otimes\delta_{T''=\pi(s'')}$. If $\rho = N_{|S|}(m,\Sigma)$, then $T_*\rho = N_{|T|}(m_T,\Sigma_T)$, where m_T, Σ_T are specified in Proposition D.6. Moreover, $T = T'\sqcup T''$ translates into $E_T = E_{T'}\times E_{T''}$, which in turn induces splittings $m_T = (m_{T'}, m_{T''})$ and

$$\Sigma_T = \begin{pmatrix} \Sigma_{T'} & \Sigma_{T',T''} \\ \Sigma_{T'',T'} & \Sigma_{T''} \end{pmatrix}.$$

[6]It is characteristic of this *simplicial case* that every space E_S has a preferred basis, such that, for every $T \in \mathcal{S}_S$, E_T is naturally included in E_S and corresponds the span of the basis elements T. In a nonsimplicial case, there could exist $S \in \mathbf{K}$ and collections $\{U_1,\dots,U_n\}\subset\mathcal{S}_S$ and $\{T_1,\dots,T_m\}\in\mathcal{S}_S$ such that $E_S = \bigoplus_{1\le i\le n} E_{U_i} = \bigoplus_{1\le j\le m} E_{T_j}$, and $E_{U_i}\cap E_{T_j} = \{0\}$ for every i,j.

Since $T'' = \pi(s'')$ almost surely, the covariance matrices $\Sigma_{T''}$ and $\Sigma_{T',T''} = \Sigma'_{T'',T'}$ vanish. The matrix $\Sigma_{T'}$ is positive definite, because the law of T' is absolutely continuous with respect to $\mu_{T'}$ (see Proposition D.7). We have,

$$\frac{dT_*\rho}{d\mu_{T'} \otimes \delta_{T''=\pi(s'')}}(t', t'') = \frac{1}{(\det(2\pi\Sigma))^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(t' - m_{T'})^{\mathrm{tr}}\Sigma_{T'}^{-1}(t' - m_T)\right). \tag{9.25}$$

Consider now the action as defined in (9.21).

$$(T.\varphi)(\mu_{S'} \otimes \delta_{S''=s''}, m, \Sigma) = \int_{E_T} \varphi(\mu_{S'\setminus T} \otimes \delta_{T'=t'} \otimes \delta_{S''=s''}, \bar{m}(t), \bar{\Sigma})\, \mathrm{d}T_*\rho(t),$$

Proposition D.6 states that

$$\bar{m}(t) = \begin{pmatrix} m_{S\setminus T} + \Sigma_{S\setminus T,T}\Sigma_T^-(t - \mu_T) \\ t \end{pmatrix} \in E_{S\setminus T} \times E_T, \tag{9.26}$$

and

$$\bar{\Sigma} = \begin{pmatrix} \Sigma_{S\setminus T} - \Sigma_{S\setminus T,T}\Sigma_T^-\Sigma_{T,S\setminus T} & 0 \\ 0 & 0 \end{pmatrix}. \tag{9.27}$$

Since φ^+ and φ^- satisfy (9.24), the expressions $\varphi^\pm(\mu_{S'\setminus T} \otimes \delta_{T'=t'} \otimes \delta_{S''=s''}, \bar{m}(t), \bar{\Sigma})$ grow slower than certain polynomial in t' when $\|t'\| \to \infty$ (remark that (9.26) is linear in $t = (t', t'')$), and therefore their integrals against the density (9.25) converge. In this case,

$$T.\varphi = T.\varphi^+ - T.\varphi^-.$$

Proposition 9.11 ensures that, for any pair of variables $T, U \in \mathscr{S}_S$, $U.(T.\varphi^\pm) = UT.\varphi^\pm$. Thus, $U.(T.\varphi) = TU.\varphi^+ - TU.\varphi^- = TU.\varphi$. □

Chapter 10

Probabilistic information cohomology on simplicial structures

Given any information structure $(\mathbf{S}, \mathscr{E})$, the set $\mathscr{S}_S := \{T \mid S \to T\}$ is a monoid with product $(T, U) \mapsto T \cup U =: T \wedge U$. Given $S \to T$ there is a natural inclusion $\mathscr{S}_T \hookrightarrow \mathscr{S}_S$. Therefore $X \mapsto \mathscr{S}_X$ is a presheaf of monoids, and $X \mapsto \mathscr{A}_X := \mathbb{R}[\mathscr{S}_X]$ is a presheaf of algebras.

The category of $\mathscr{A}$-modules, $\mathbf{Mod}(\mathscr{A})$, is abelian and has enough injectives. Information cohomology is $H^\bullet(\mathbf{S}, \mathscr{E}) := \operatorname{Ext}^\bullet(\mathbf{S}, \mathscr{E})$, as defined in Section 2.2, and we use the bar-resolution in order to characterize the cocycles as in Section 2.5.

In this chapter, we study information cohomology on a simplicial structure $(\mathbf{K}, \mathscr{E})$. We take as coefficients a module $\mathscr{F}$, such that each $\mathscr{F}(S)$ correspond to measurable real-valued functionals defined on a subset of $\mathscr{Q}(S)$ stable by conditioning and such that, for every $\varphi \in \mathscr{F}(S)$ and $Y \in \mathscr{S}_S$, the integrals $Y.\varphi^+$ and $Y.\varphi^-$ are finite, in which case $Y.\varphi$ is well defined and the monoid $\mathscr{S}_S$ acts on $\mathscr{F}(S)$. An example is the module $\mathscr{F}_{\text{Gauss}}$ introduced in Section 9.4.

10.1 Computation of H^0

A 0-cocycle is a collection of functions $\varphi_S[] : \mathscr{Q}(S) \to \mathbb{R}$ such that

$$\varphi_S[](\mu, \rho) = \varphi_\emptyset(\delta_0, \delta_0) \in \mathbb{R}, \tag{10.1}$$

thus they correspond to real constants.

A 0-coboundary φ must satisfy $0 = (\delta\varphi)_S[T]$. Supposing that $\varphi_S[] = C$, one has

$$\begin{aligned}
&(\delta\varphi)_S[T](\mu_{S'} \otimes \delta_{S''=s''}, \rho) \\
&= (T.\varphi)_S[](\mu_{S'} \otimes \delta_{S''=s''}, \rho) - \varphi_S[](\mu_{S'} \otimes \delta_{S''=s''}, \rho) \\
&= \int_{E_T} \varphi_S[](\mu_{S' \setminus T} \otimes \delta_{T'=t'} \otimes \delta_{S''=s''}, \rho_t) \, \mathrm{d}T_*\rho(t) - \varphi_S[](\mu_{S'} \otimes \delta_{S''=s''}, \rho) \\
&= C - C = 0
\end{aligned}$$

We conclude that $H^0(\mathbf{K}, \mathscr{F}) \cong \mathbb{R}$.

10.2 General properties of 1-cocycles

Remark first that $\delta C^0 = \{0\}$, so we only need to compute cocycles:

$$H^1(\mathbf{K}, \mathscr{F}) = \ker(\delta : C^1 \to C^2). \tag{10.2}$$

Recall that $C^1 = \mathrm{Hom}_{\mathscr{A}}(\mathscr{B}_1, \mathscr{F})$ and a 1-cochain $\varphi \in C^1$ corresponds to a collection of morphisms $\varphi_S \in \mathrm{Hom}_{\mathscr{A}_S}(\mathscr{B}_1(S), \mathscr{F}(S))$ that verifies two conditions:

1. Equivariance: $\varphi_S(T[U]) = T.\varphi([U])$.
2. Joint locality: whenever $S \to T$,

$$\varphi_S[T](\mu_{S'} \otimes \delta_{S''=s''}, \rho) = \varphi_T[T](\mu_{T'} \otimes \delta_{T''=\pi(s'')}, T_*\rho)$$

We write $\varphi[T]$ instead of $\varphi_T[T]$, and even of $\varphi_S[T]$ if S is clear from context (its evaluation is related to $\varphi_T[T]$ by joint locality).

A 1-cochain φ is a 1-cocycle if additionally it verifies, for every $S \in \mathrm{Ob}\,\mathbf{K}$ and every $T, U \in \mathscr{S}_S$,

$$0 = T.\varphi_S[U] - \varphi_S[TU] + \varphi_S[T] \tag{10.3}$$

as functionals in $\mathscr{F}(S)$. This and the commutativity of the product give the symmetric equation

$$T.\varphi_S[U] + \varphi_S[T] = U.\varphi_S[T] + \varphi_S[U]. \tag{10.4}$$

The general properties of disintegrations of measures imply that entropy is a 1-cocycle, cf. Proposition 11.35.

Proposition 10.1. *If φ is a 1-cocycle, then:*

1. *For every $S \in \mathrm{Ob}\,\mathbf{K}$, $\varphi[S](\delta_s, \delta_s) = 0$.*
2. *For every $S \in \mathrm{Ob}\,\mathbf{K}$, and any decomposition $S = S' \sqcup S''$,*

$$\varphi_S[S](\mu_{S'} \otimes \delta_{S''=s''}, \rho) = \varphi[S'](\mu_{S'}, S'_*\rho).$$

Proof.
Claim (1): Setting $S = T = U$ in (10.3) and evaluating it on (δ_s, δ_s), we obtain

$$0 = (S.\varphi)_S[S](\delta_s, \delta_s) = \varphi_S[S](\delta_s, \delta_s).$$

Claim (2): Set $T = S'$ and $U = S''$ in (10.3), and evaluate the expression on $(\mu_{S'} \otimes \delta_{S''}, \rho)$. The conditional term vanishes, because

$$(S'.\varphi[S''])_S(\mu_{S'} \otimes \delta_{S''=s''}, \rho) = \int_{E_{S'}} \varphi[S''](\delta_{(s',s'')}, \delta_{(s',s'')})\, \mathrm{d}S'_*\rho(s').$$

The claim becomes a consequence of joint locality, which implies that

$$\varphi_S[S'](\mu_{S'} \otimes \delta_{S''=s''}, \rho) = \varphi[S'](\mu_{S'}, S'_*\rho).$$

□

Proposition 10.1-(2) implies that we only need to characterize each functional $\varphi[S]$ evaluated on the corresponding nondegenerate laws (μ_S, ρ). This simplifies greatly the computation of cocycles, as expressed in the following proposition. (Recall that $\Pi(S, \mu_S)$ are the measures on $(E_S, \mathfrak{B}(E_S))$ absolutely continuous with respect to μ_S.)

Proposition 10.2. *Let φ be a 1-cochain and $\Phi = \{\phi^S\}_{S\in \mathrm{Ob}\,\mathbf{K}}$ a collection of measurable functionals $\phi^S : \Pi(S, \mu_S) \to \mathbb{R}$. The following conditions are equivalent:*

1. *φ is a 1-cocycle, and*
$$\varphi_S[S](\mu_S, \cdot) \equiv \phi^S(\cdot). \tag{10.5}$$
2. *For every $S \in \mathrm{Ob}\,\mathbf{K}$, every $U \subset S$, and every $\rho \in \Pi(S, \mu_S)$,*
$$\phi^S(\rho) = \int_{E_U} \phi^{S\setminus U}((S \setminus U)_*\rho_u)\,\mathrm{d}U_*\rho(u) + \phi^U(U_*\rho). \tag{10.6}$$

Besides, for every $S \in \mathrm{Ob}\,\mathbf{K}$ and $U \subset S$,
$$\varphi_S[U](\mu_{S'} \otimes \delta_{S''=s''}, \cdot) \equiv \phi^{U\cap S'}((U \cap S')_*(\cdot)). \tag{10.7}$$

From now on, we use the collection $\Phi = \{\phi^S\}_{S\in \mathrm{Ob}\,\mathbf{K}}$ that verifies $\varphi_S[U](\mu_{S'} \otimes \delta_{S''=s''}, \cdot) \equiv \phi^{U\cap S'}((U\cap S')_*(\cdot))$ as a simplified description of a 1-cochain; such cochain is a 1-cocycle if and only if it satisfies the cocycle equation (10.6).

Proof. First, let us prove that (1)⇒(2): Since φ is a 1-cocycle, for every $S \in \mathrm{Ob}\,\mathbf{K}$, every $U \subset S$, and $(\mu_{S'} \otimes \delta_{S''=s''}, \rho) \in \mathcal{Q}(S)$,
$$\varphi_S[U](\mu_{S'} \otimes \delta_{S''=s''}, \rho) = \varphi_U[U](\mu_{U\cap S'} \otimes \delta_{S''\cap U=\pi(s'')}, U_*\rho) \tag{10.8}$$
$$= \varphi_{U\cap S'}[U \cap S'](\mu_{U\cap S'}, (U \cap S')_*\rho) \tag{10.9}$$
$$= \phi^{U\cap S'}((U \cap S')_*\rho), \tag{10.10}$$
where (10.8) corresponds to joint locality, and (10.9) comes from Proposition 10.1-(2) along with the functoriality of marginalizations. The last equality takes (10.5) into account. This establishes (10.7).

Consider now an arrow $S \to U$ in $\mathbf{K}$. The cocycle condition (10.3) implies in particular that
$$\varphi_S[S](\mu_S, \rho) = \int_{E_U} \varphi_S[S \setminus U](\mu_{S\setminus U} \otimes \delta_{U=u}, \rho_u)\,\mathrm{d}U_*\rho(u) + \varphi_S[U](\mu_S, \rho). \tag{10.11}$$
Because of (10.7),
$$\varphi_S[S \setminus U](\mu_{S\setminus U} \otimes \delta_{U=u}, \rho_u) = \phi^{S\setminus U}((S \setminus U)_*\rho_u). \tag{10.12}$$
Joint locality also entails that
$$\varphi_S[U](\mu_S, \rho) = \varphi_U[U](\mu_U, U_*\rho) = \phi^U(U_*\rho). \tag{10.13}$$
The equalities (10.12) and (10.13) show that (10.11) corresponds exactly to (10.6).

Now we prove (2)⇒(1): That φ is a 1-cocycle means that, for any $S \in \mathrm{Ob}\,\mathbf{K}$ and $T, U \in \mathscr{S}_S$,
$$\varphi_S[T \cup U](\mu_{S'} \otimes \delta_{S''=s''}, \rho) \overset{(!)}{=}$$
$$\int_{E_U} \varphi_S[T](\mu_{S'\setminus U} \otimes \delta_{(U,S'')=(u,s'')}, \rho_u)\,\mathrm{d}U_*\rho(u) + \varphi_S[U](\mu_{U'} \otimes \delta_{U''=\pi(s'')}, U_*\rho), \tag{10.14}$$
where $U' = S' \cap U$ and $U'' = S'' \cap U$ as usual.

Consider each term separately. The identity (10.7) says that

$$\varphi_S[W](\mu_{S'} \otimes \delta_{S''=s''}, \rho) = \phi^{W'}(W'_*\rho), \tag{10.15}$$

where we have written again W instead of $T \cup U$, and $W' := W \cap S' = T' \cup U'$. Similarly,

$$\varphi_S[T](\mu_{S'\setminus U} \otimes \delta_{(U,S'')=(u,s'')}, \rho_u) = \phi^{T'\setminus U'}((T' \setminus U')_*\rho_u), \tag{10.16}$$

because $T \cap (S' \setminus U) = (T \cap S') \setminus (S' \cap U) = T' \setminus U'$. Finally,

$$\varphi_S[U](\mu_{U'} \otimes \delta_{U''=\pi(s'')}, U_*\rho) = \phi^{U'}(U_*\rho). \tag{10.17}$$

Replace (10.15), (10.16) and (10.17) in (10.14) to obtain the equivalent form

$$\phi^{T'\cup U'}((T' \cup U')_*\rho) = \int_{E_U} \phi^{T'\setminus U'}((T' \setminus U)_*\rho_u)\, \mathrm{d}U_*\rho(u) + \phi^{U'}(U_*\rho), \tag{10.18}$$

which corresponds precisely to the identity (10.6) if one takes into account the funtoriality of marginalizations ($\phi^{U'}(U_*\rho) = \phi^{U'}(U_*(T' \cup U')_*\rho)$, etc.). □

Remark 10.3. If each $\varphi_T[T]$ is just a function of the probabilities on $(E_T, \mathfrak{B}_T)$, then

$$S.\varphi_U[T](\rho) = \int_{E_T} \varphi_T[T](\delta_{\pi(u)})\, \mathrm{d}U_*\rho(u),$$

whenever $T \subset U \subset S$. If $\phi_T[T](\rho)$ is the differential entropy when $\rho \ll \mu_T$, the quantity $\varphi_T[T](\delta_{\pi(u)})$ is expected to diverge to $-\infty$ under any sensible definition (cf. Section 12.2). In general, divergences are immediately introduced with conditioning if one does not update the pertinent reference measure. The sheaf $\mathscr{F}$ allows us to keep track of this. The relevance of the reference measure was already remarked by Shannon (see the Introduction) and later by Csiszár, who defined the *generalized entropy* as

$$S_\mu(\rho) = -\int \log\left(\frac{d\rho}{\mathrm{d}\mu}\right) \mathrm{d}\rho,$$

when λ is an arbitrary reference measure and μ a probability such that $\mu \ll \lambda$. This is further developed in Chapter 12.

10.3 Computation of H^1: Gaussian case

In this case each reference measure μ_S is the Lebesgue measure on $E_S \cong \mathbb{R}^{|S|}$, and each functional ϕ^S is defined on the set of nondegenerate $|S|$-variate normal (gaussian) laws on E_S. Since E_S comes with a basis, S itself, these laws are in bijective correspondence with pairs $(m, \Sigma) \in E_S \times PD(S)$,[1] that correspond to the mean and covariance, respectively.[2] From now on, we write $\phi^S(m, \Sigma)$.

[1] $PD(S)$ denotes the positive definite matrices in $M_{|S|}(\mathbb{R})$.

[2] More precisely, a basis establishes a bijection $B : E_S \times PD(S) \xrightarrow{\sim} \Pi_{\text{Gauss}}(S, \mu_S)$ and precomposition with B defined the functional $\phi^{S,B} := \phi^S \circ B : E_S \times PD(S) \to \mathbb{R}$, that we also denote by ϕ^S since the basis is fixed. In a nonsimplicial case, different bases should be taken into account in such a way that, for every $G \in GL(E_S)$, $\phi^{S,B}(m, \Sigma) = \phi^{S,GB}(Gm, G^T\Sigma G)$.

Consider a nondegenerate gaussian law ρ with mean m and covariance Σ. Following the conventions in Section D, the splitting of the variable, $S = S_1 \sqcup S_2$ (i.e $X_S = (X_{S_1}, X_{S_2}) \in E_{S_1} \times E_{S_2}$) induces an splitting of the parameters: $m = (m_1, m_2)$, and

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

According to Proposition D.6, the marginal law $U_*\rho$ is a $|U|$-variate normal distribution with mean m_1 and covariance Σ_{11}, and $(S \setminus U)_*\rho_{x_1}$ is a $|S \setminus U|$-variate normal with mean

$$\bar{m}_2(x_1) := m_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - m_1),$$

and covariance

$$\bar{\Sigma}_{22} := \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}.$$

Since Σ is supposed to be positive definite,[3] Σ_{11} and Σ_{22} are positive definite too, hence invertible.

The functionals in Φ are subject to the condition (10.6): for any $S \in \mathbf{K}$ and any decomposition $S = S_1 \sqcup S_2$,

$$\phi^S(m, \Sigma) = \int_{E_{S_1}} \phi^{S_2}(\bar{m}_2(x_1), \bar{\Sigma}_{22}) \frac{\exp(-\frac{1}{2}(x_1 - m_1)^{\mathrm{tr}}\Sigma_{11}^{-1}(x - m_1))}{\sqrt{|2\pi\Sigma_{11}|}} \,\mathrm{d}x_1 + \phi^{S_1}(m_1, \Sigma_{11}). \quad (10.19)$$

where we have followed the conventions in the previous paragraph. As a consequence, they also fulfill the symmetric equation (10.4):

$$\int_{E_{S_1}} \phi^{S_2}(\bar{m}_2(x_1), \bar{\Sigma}_{22}) \frac{\exp\left(-\frac{1}{2}(x_1 - m_1)^{\mathrm{tr}}\Sigma_{11}^{-1}(x_1 - m_1)\right)}{\sqrt{|2\pi\Sigma_{11}|}} \,\mathrm{d}x_1 - \phi^{S_2}(m_2, \Sigma_{22}) =$$
$$\int_{E_{S_2}} \phi^{S_1}(\bar{m}_1(x_2), \bar{\Sigma}_{11}) \frac{\exp\left(-\frac{1}{2}(x_2 - m_2)^{\mathrm{tr}}\Sigma_{22}^{-1}(x_2 - m_2)\right)}{\sqrt{|2\pi\Sigma_{22}|}} \,\mathrm{d}x_2 - \phi^{S_1}(m_1, \Sigma_{11}). \quad (10.20)$$

10.3.1 1-cocycles that depend only on the covariance matrix

First, we compute the cocycles that depend only on the covariance matrix, $\phi^S(m, \Sigma) = \phi^S(\Sigma)$. The domain of ϕ^S are the positive definite matrices in $M_{|S|}(\mathbb{R})$, that we denote $PD(S)$.[4]

Proposition 10.4. *Suppose that* $\mathbf{K}$ *is connected and all its vertices belongs to a* 1*-cell. A collection of* C^2 *functions*[5] $\Phi = \{\phi^S : PD(S) \to \mathbb{R}\}_{S \in \mathbf{K}}$ *satisfies the cocycle condition* (10.19) *if and only if there exist real constants a and* $\{k_i\}_{i \in I}$ *such that, for every* $S \in \mathbf{K}$,

$$\phi^S(\Sigma) = a \ln(|\Sigma|) + \sum_{i \in S} k_i. \quad (10.21)$$

[3]The measure ρ is a $|S|$-variate normal distribution absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{|S|}$. This is equivalent to $\Sigma \gg 0$, as shown in Proposition D.7.

[4]The basis establishes a bijection $B : E_S \times PD(S) \xrightarrow{\sim} \Pi_{\mathrm{Gauss}}(S, \mu_S)$ and the functional $\phi^{S,B} := \phi^S \circ B : E_S \times PD(S) \to \mathbb{R}$ is required to factor through the quotient $E_S \times PD(S) \to PD(S)$. We use the same symbol ϕ^S to denote the factor, since there is no risk of confusion.

[5]$PD(S) \subset M_{|S|}(\mathbb{R})$ is supposed to have the standard differential structure.

Proof. For any $i \in I$, there exists $j \in I$ such that $S := \{i, j\} \in \mathrm{Ob}\,\mathbf{K}$. Set $S_1 = \{i\}$ and $S_2 = \{j\}$. Let $\Sigma = \left[\begin{smallmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{smallmatrix}\right]$ be the covariance matrix of (S_1, S_2). Under these circumstances, (10.20) says that

$$\phi^i(\sigma_{11} - \sigma_{12}^2\sigma_{22}) - \phi^i(\sigma_{11}) = \phi^j(\sigma_{22} - \sigma_{12}^2\sigma_{11}) - \phi^j(\sigma_{22}), \tag{10.22}$$

where we have written ϕ^i instead of $\phi^{\{i\}}$. The derivative with respect to σ_{11} is

$$\dot\phi^i(\sigma_{11} - \sigma_{12}^2\sigma_{22}) - \dot\phi^i(\sigma_{11}) = \dot\phi^j(\sigma_{22} - \sigma_{12}^2\sigma_{11})\frac{\sigma_{12}^2}{\sigma_{11}^2}, \tag{10.23}$$

that can be derived with respect to σ_{22} to obtain

$$\frac{\ddot\phi^i(\sigma_{11} - \sigma_{12}^2\sigma_{22}^{-1})}{\sigma_{22}^2} = \frac{\ddot\phi^j(\sigma_{22} - \sigma_{12}^2\sigma_{11}^{-1})}{\sigma_{11}^2}. \tag{10.24}$$

Or equivalently,

$$\ddot\phi^i(\sigma_{11} - \sigma_{12}^2\sigma_{22}^{-1})\frac{(\sigma_{11}\sigma_{22} - \sigma_{12}^2)^2}{\sigma_{22}^2} = \ddot\phi^j(\sigma_{22} - \sigma_{12}^2\sigma_{11}^{-1})\frac{(\sigma_{11}\sigma_{22} - \sigma_{12}^2)^2}{\sigma_{11}^2}. \tag{10.25}$$

Set $u = \sigma_{11} - \sigma_{12}^2\sigma_{22}^{-1} \in (0, \infty)$, $v = \sigma_{22} - \sigma_{12}^2\sigma_{11}^{-1} \in (0, \infty)$, to obtain

$$\ddot\phi^i(u)u^2 = \ddot\phi^j(v)v^2 = \text{constant} =: -a, \tag{10.26}$$

which means that $\dot\phi^i(u) = a/u + b_i$, and $\phi_i(u) = a\ln(u) + b_i u + k_i$, with $a, b_i, k_i \in \mathbb{R}$. But the functions $u \mapsto B_i u$ do not solve (10.22) unless both B_i vanish, while $u \mapsto a\ln(u) + k_i$ is a solution. Therefore,

$$\phi^i(\sigma) = a\ln(\sigma) + k_i \quad \text{and} \quad \phi^j(\sigma) = a\ln(\sigma) + k_j$$

for arbitrary real constants a, k_i, k_j. Since the complex $\mathbf{K}$ is connected, a must be common to all vertices. Remark that in this case,

$$\phi^S(\Sigma) = \phi^j(|\Sigma|/\sigma_{11}) + \phi^i(\sigma_{11}) = a\ln(|\Sigma|) + k_1 + k_2. \tag{10.27}$$

The general form (10.21) is obtained by induction: consider $S = S_1 \sqcup \{m\}$,

$$\Sigma_S = \begin{pmatrix} \Sigma_S & B \\ C & \sigma_{mm} \end{pmatrix}$$

and denote by Σ_S/σ_{mm} the Schur complement of σ_{mm} in Σ_S. Then

$$\begin{aligned} \phi^S(\Sigma_S) &= \phi^{S_1}(\Sigma_S/\sigma_{mm}) + \phi^m(\sigma_{mm}) && (10.28)\\ &= a\ln(|\Sigma_S/\sigma_{mm}||\sigma_{mm}|) + \sum_{i\in S_1} k_i + k_m && (10.29)\\ &= a\ln(|\Sigma_S|) + \sum_{i\in S} k_i. && (10.30) \end{aligned}$$

In the last step, we used Schur's determinantal identity (Proposition C.1).

It is easy to show that these ϕ^S just introduced satisfy all the cocycle equations (10.19). In fact, by linearity, this can be verified separately for $\phi_1^S(m, \Sigma) = \sum_{i\in S} k_i$ and $\phi_2^S(m, \Sigma) = a\ln(|\Sigma|)$. In the case of $\phi_1^\bullet$, this is immediate. For $\phi_2^\bullet$, (10.19) is equivalent to Schur's determinantal identity. $\square$

Remark 10.5. Set $d := |S|$. Let ρ be a nondegenerate gaussian law on E_S, with mean m and covariance Σ. According to the traditional definition, the differential entropy of ρ (with respect to the Lebesgue measure λ_S on E_S) is

$$\begin{aligned}
h^S(\rho) &:= -\int_{E_S} \log_b\left(\frac{\mathrm{d}\rho}{\mathrm{d}\lambda_S}\right)\mathrm{d}\rho \\
&= -\int_{E_S} \log_b\left(\frac{\exp\left(-\frac{1}{2}(x-m)^{\mathrm{tr}}\Sigma^{-1}(x-m)\right)}{|2\pi\Sigma|^{1/2}}\right)\mathrm{d}\rho \\
&= \frac{\log_b(e)}{2}\int_{E_S}\left((x-m)^{\mathrm{tr}}\Sigma^{-1}(x-m)\right)\frac{e^{-\frac{1}{2}(x-m)^{\mathrm{tr}}\Sigma^{-1}(x-m)}}{|2\pi\Sigma|^{1/2}}\,\mathrm{d}x + \frac{1}{2}\log_b(|2\pi\Sigma|)
\end{aligned}$$

The change of variables[6] $y = \Sigma^{-1/2}(x-m)$ gives

$$\int_{E_S}\left(-\frac{1}{2}(x-m)^{\mathrm{tr}}\Sigma^{-1}(x-m)\right)\frac{\exp(-\frac{1}{2}(x-m)^{\mathrm{tr}}\Sigma^{-1}(x-m))}{|2\pi\Sigma|^{1/2}}\,\mathrm{d}x = \int_{E_S}\frac{(-y^2)}{2}\frac{\exp(-\frac{y^2}{2})}{|2\pi I|^{1/2}}\,\mathrm{d}y = -d.$$

We conclude that

$$h^S(\rho) = d\left(\frac{\log_b(2\pi e)}{2}\right) + \frac{1}{2}\log_b(|\Sigma|). \tag{10.31}$$

This is a particular case of the general form in Proposition 10.4, where $a = (2\ln(b))^{-1}$ and $k_i = \log_b(2\pi e)/2$.

10.3.2 Decomposition of ϕ^S as a sum; Convolutions

In this and the following sections, we use the theory of distributions, because the conditional term in the cocycle equations can be written as a convolution and its analysis is naturally related to the Fourier transform. The main definitions and results of this theory that are used in this thesis are summarized in Appendix E.

Remark that ϕ^{S_1} defines a distribution in $\mathcal{S}'(E_{S_1})$, the space of tempered distributions, through the formula $f \in \mathcal{S} \mapsto \int \phi^{S_1} f$; the integral converges in virtue of the polynomial-growth condition (9.24).

Given a distribution $T \in \mathcal{S}'(\mathbb{R}^d)$, its convolution with an element $f \in \mathcal{S}(\mathbb{R}^d)$, denoted $T * f$, corresponds to the *function* $x \mapsto \langle T, \tau_x \check{f}\rangle$ (the operator τ_x is a translation, $\tau_x f(y) := f(x+y)$, and $\check{f}(x) := f(-x)$). The following proposition shows that sometimes the conditional term in the cocycle equations can be written as a convolution.

Proposition 10.6. *1. The integral*

$$\int_{E_{S_2}} \phi^{S_1}(\bar{m}_1(x_2), \bar{\Sigma}_{11})\frac{\exp\left(-\frac{1}{2}(x_2-m_2)^{\mathrm{tr}}\Sigma_{22}^{-1}(x_2-m_2)\right)}{\sqrt{|2\pi\Sigma_{22}|}}\,\mathrm{d}x_2 \tag{10.32}$$

equals $\mathbb{E}\left(\phi^{S_1}(m_1 - Y_1, \Sigma_{11} - \Sigma'_{11})\right)$, *where* Y_1 *has a* $|S_1|$*-variate normal distribution with mean* m_1 *and covariance* $\Sigma'_{11} := \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

---

[6]We denote by $\Sigma^{-1/2}$ or $\sqrt{\Sigma}$ the *principal* square root of the positive definite matrix Σ: its eigenvalues are the positive square roots of the eigenvalues of Σ.

2. *There expressions*

$$C(S_1,S_2,\Sigma) := \mathbb{E}\left(\phi^{S_1}(m_1 - Y_1, \Sigma_{11} - \Sigma'_{11})\right) - \phi^{S_1}(m_1, \Sigma_{11}) \tag{10.33}$$

and

$$C(S_2,S_1,\Sigma) := \mathbb{E}\left(\phi^{S_2}(m_2 - Y_2, \Sigma_{22} - \Sigma'_{22})\right) - \phi^{S_1}(m_2, \Sigma_{22}) \tag{10.34}$$

are independent of m, *and* $C(S_1,S_2,\Sigma) = C(S_2,S_1,\Sigma)$.

3. *If* Σ_{12} *is surjective, then the matrix* Σ'_{11} *is positive definite and*

$$\mathbb{E}\left(\phi^{S_1}(m_1 - Y_1, \Sigma_{11} - \Sigma'_{11})\right) = (\phi^{S_1}_{\Sigma_{11}-\Sigma'_{11}} * G_{\Sigma'_{11}})(m_1), \tag{10.35}$$

where $\phi^{S_1}_A$ *is the map* $x \mapsto \phi^{S_1}(x,A)$, *and* G_B *is* $x \mapsto e^{-\frac{1}{2}x^{\mathrm{tr}}B^{-1}x}/\sqrt{\det(2\pi B)}$.

Proof.
Claim (1): The integral (10.32) equals $\mathbb{E}\left(\phi^{S_1}(m_1 - \Sigma_{12}\Sigma_{22}^{-1}(X - m_2), \bar{\Sigma}_{11})\right)$, where $X \sim N_{|S_2|}(m_2, \Sigma_{22})$. According to Proposition D.4,

$$Y_1 := \Sigma_{12}\Sigma_{22}^{-1}(X - m_2) \sim N_{|S_1|}(0, (\Sigma_{12}\Sigma_{22}^{-1})\Sigma_{22}(\Sigma_{12}\Sigma_{22}^{-1})^{\mathrm{tr}}) = N_{|S_1|}(0, \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}).$$

The definitions imply that $\bar{\Sigma}_{11} = \Sigma_{11} - \Sigma'_{11}$.

Claim (2): Taking into account part (1), the symmetric equation (10.20) can be rewritten as

$$\mathbb{E}\left(\phi^{S_1}(m_1 - Y_1, \Sigma_{11} - \Sigma'_{11})\right) - \phi^{S_1}(m_1, \Sigma_{11}) = \\ \mathbb{E}\left(\phi^{S_2}(m_2 - Y_2, \Sigma_{22} - \Sigma'_{22})\right) - \phi^{S_1}(m_2, \Sigma_{22}). \tag{10.36}$$

with $Y_1 \sim N_d(0, \Sigma'_{11})$ and $Y_1 \sim N_d(0, \Sigma'_{22})$. The expression on the left depends only on m_1 and Σ, not on m_2; similarly, the expression on the right does not depend on m_1. We conclude that both expressions equal certain "constant" $C(S_1,S_2,\Sigma) = C(S_2,S_1,\Sigma)$.

Claim (3): Consider $x \in \mathbb{R}^n \setminus \{0\}$. Since Σ_{12} surjective, its transpose its injective and $\Sigma_{21}x = \Sigma_{12}^T x \neq 0$. Then

$$x^T(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})x = (\Sigma_{21}x)^T\Sigma_{22}^{-1}(\Sigma_{21}x) > 0,$$

because Σ_{22}^{-1} is positive definite. This proves that Σ'_{11} is positive definite, which implies—according to Proposition D.7—that Y_1 has a density with respect to μ_{S_1}, that is precisely $G_{\Sigma'_{11}}$. Therefore,

$$\begin{aligned}\mathbb{E}\left(\phi^{S_1}(m_1 - Y_1, \Sigma_{11} - \Sigma'_{11})\right) &= \int_{E_{S_1}} \phi^{S_1}_{\Sigma_{11}-\Sigma'_{11}}(m_1 - y_1)G_{\Sigma'_{11}}(y)\,\mathrm{d}y &\text{(10.37)}\\ &= (\phi^{S_1}_{\Sigma_{11}-\Sigma'_{11}} * G_{\Sigma'_{11}})(m_1). &\text{(10.38)}\end{aligned}$$

□

Remark 10.7. The function $C(S_1,S_2,\Sigma)$ is nontrivial. In fact, when $\phi^\bullet$ is given by the differential entropy in (10.31), $C(S_1,S_2,\Sigma)$ equals the usual mutual information $I(S_1,S_2)$.

Proposition 10.8. *For any set $S \in \operatorname{Ob} \mathbf{K}$ and any flag of sets $S_1 \subset S_2 \subset \ldots \subset S_{|S|} \subset S_{|S|} = S$ such that $|S_k| = k$, the equality*

$$\phi^S(m, \Sigma) = \sum_{i \in S} \phi^i(m_i, \sigma_{ii}) + \sum_{k=2}^{|S|} C(S_{k-1}, S_k \setminus S_{k-1}, \Sigma_{S_k S_k}) \tag{10.39}$$

holds, where ϕ^i denotes $\phi^{\{i\}}$ and $\Sigma_{S_k S_k}$ is the square block of Σ corresponding to the indexes in S_k.

Proof. Since $S_k = S_{k-1} \sqcup (S_k \setminus S_{k-1})$, the cocycle equation says

$$\phi^{S_k}(m_{S_k}, \Sigma_{S_k S_k}) = \mathbb{E}\left(\phi^{S_{k-1}}(m_{S_{k-1}} - Y_{S_{k-1}}, \Sigma_{S_{k-1}S_{k-1}} - \Sigma'_{S_{k-1}S_{k-1}})\right) + \phi^{i_k}(m_{i_k}, \sigma_{i_k i_k}) \tag{10.40}$$

where we have denoted by i_k the element in $S_k \setminus S_{k-1}$. Proposition 10.6-(2) says that the expectation above equals $\phi^{S_{k-1}}(m_{S_{k-1}}, \Sigma_{S_{k-1}S_{k-1}}) + C(S_{k-1}, \{i_k\}, \Sigma_{S_k S_k})$. The result follows by induction on k. □

The quantity $\sum_{k=2}^{|S|} C(S_{k-1}, S_k \setminus S_{k-1}, \Sigma_{S_k S_k})$ turns out to be independent of the chosen flag and we call it $D_\phi(S, \Sigma)$:

$$D_\phi(S, \Sigma) := \phi^S(m, \Sigma) - \sum_{i \in S} \phi^i(m_i, \sigma_{ii}). \tag{10.41}$$

Remark 10.9. In the case of differential entropy, $C(S_1, S_2, \Sigma) = -I(S_1; S_2)$ holds, and it is known that

$$H[(S_1, S_2)] = H[S_1|S_2] + H[S_2] = H[S_1] + H[S_2] - I(S_1; S_2). \tag{10.42}$$

This implies that

$$H[(S_1, S_2, \cdots, S_n)] = \sum_{i=1}^n H[S_i] - \sum_{i=1}^n I((S_1 \cdots S_{i-1}); S_i), \tag{10.43}$$

that is an special case of Proposition 10.8. Finally, $D(S, \Sigma)$ is proportional to $\ln(|\Sigma| / \prod_{i \in S} \sigma_{ii})$.

10.3.3 General cocycles

Provided that $|S_1| \leq |S_2|$, the matrix Σ_{12} is *generically* surjective and (10.33) has the form

$$C(S_1, S_2, \Sigma) = (\phi^{S_1}_{\Sigma_{11} - \Sigma'_{11}} * G_{\Sigma'_{11}})(m_1) - \phi^{S_1}_{\Sigma_{11}}(m_1). \tag{10.44}$$

Furthermore, $G_{\Sigma'_{11}}$, being a element of $\mathcal{S}(E_{S_1})$, belongs to the space $\mathcal{O}'_C$ of distributions rapidly decreasing at infinity (convoluters). In virtue of the convolution theorem (Proposition E.5), the Fourier transform turns the last equation into

$$\widehat{\phi}^{S_1}(\xi_1, \Sigma_{11} - \Sigma'_{11}) \exp\left(-2\pi^2 \xi_1^{\mathrm{tr}} \Sigma'_{11} \xi_1\right) - \widehat{\phi}^{S_1}(\xi_1, \Sigma_{11}) = C(S_1, S_2, \Sigma)\delta_{\xi_1 = 0}, \tag{10.45}$$

where $\xi_1 = \widehat{m}_1$, and $\widehat{\phi}^{S_1}$ belongs to $\mathcal{S}'(\mathbb{R}^d)$.

Function	**Transform**
1	δ_0
δ_0	1
$\delta_0^\alpha(x)$	$(2\pi i\xi)^\alpha$
$\widehat{f}(x)$	$f(-\xi) =: \check{f}(\xi)$
$(f * g)(x)$	$f(\xi)g(\xi)$
$f(x)g(x)$	$(\widehat{f} * \widehat{g})(\xi)$
$\lvert 2\pi\Sigma\rvert^{-1/2}\exp\left(-\frac{1}{2}x^{\mathrm{tr}}\Sigma^{-1}x\right)$	$\exp\left(-2\pi^2\xi^{\mathrm{tr}}\Sigma\xi\right)$
$\partial^\alpha f(x)$	$(2\pi\xi)^\alpha\widehat{f}(x)$
$x^\alpha f(x)$	$\left(\frac{i}{2\pi}\right)^\alpha \partial^\alpha f(\xi)$

Table 10.1: Fourier transforms used in this thesis, following the definition $\widehat{f}(\xi) = \int_{\mathbb{R}^n} f(x)e^{-2\pi i x\cdot\xi}\,\mathrm{d}x$. Cf. [76, pp. 231ff], [46, App. 2].

In principle, Σ'_{11} can be *any* matrix that satisfies $0 \ll \Sigma'_{11} \ll \Sigma$. In fact, given Σ'_{11}, consider the equation $\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^{\mathrm{tr}} = \Sigma'_{11}$. Introducing the appropriate change of base to express Σ_{22} as the identity ($\Sigma_{22} \gg 0$), this reduces to $\tilde{\Sigma}_{12}\tilde{\Sigma}_{12}^{\mathrm{tr}} = \Sigma_{11}$, that has as solutions any square root of Σ'_{11}. The fact that

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^{\mathrm{tr}} & \Sigma_{22} \end{pmatrix}$$

is a valid covariance matrix comes precisely from the condition $0 \ll \Sigma'_{11} \ll \Sigma$.

In this section, we study the solutions of equation (10.45) in $\mathcal{D}'(E_{S_1})$. Each distribution $\widehat{\phi}$ can be restricted to the open set $\Omega = E_{S_1} \setminus \{0\}$, to get the equation

$$\widehat{\phi}^{S_1}(\xi_1, \Sigma_{11} - \Sigma'_{11})\exp\left(-2\pi^2\xi_1^{\mathrm{tr}}\Sigma'_{11}\xi_1\right) = \widehat{\phi}^{S_1}(\xi_1, \Sigma_{11}), \tag{10.46}$$

whose solutions are described by the following proposition.

Proposition 10.10. *Let Ω be an open set in $\mathbb{R}^d$. For a collection of distributions $\widehat{\phi}^{S_1}(\cdot, \Sigma) \in D'(\Omega)$ indexed by positive definite matrices $\Sigma \in PD(\mathbb{R}^d)$, the following conditions are equivalent:*

1. *For any $\Sigma, \Sigma' \in PD(\mathbb{R}^d)$ that satisfy $0 \ll \Sigma' \ll \Sigma$,*

$$\widehat{\psi}(\xi, \Sigma - \Sigma')\exp\left(-2\pi^2\xi^{\mathrm{tr}}\Sigma'\xi\right) = \widehat{\psi}(\xi, \Sigma). \tag{10.47}$$

2. *There exists a distribution $\widehat{\psi}(\xi, 0) \in \mathcal{D}'(\Omega)$ such that*

$$\widehat{\psi}(\xi, \Sigma) = \widehat{\psi}(\xi, 0)\exp\left(-2\pi^2\xi^{\mathrm{tr}}\Sigma\xi\right), \tag{10.48}$$

Proof. It is straightforward to verify that (10.48) solves (10.47), for any choice of $\widehat{\psi}(\xi, 0)$.

Let us prove the other implication. Set $\Sigma' = (1-\lambda)\Sigma$, for any $\lambda \in (0,1)$ to deduce from (10.47) that

$$\widehat{\psi}(\xi, \lambda\Sigma) = \widehat{\psi}(\xi, \Sigma) \exp\left(2\pi^2(1-\lambda)\xi^{\mathrm{tr}}\Sigma\xi\right). \tag{10.49}$$

The distribution on the right is a well defined element of $\mathcal{D}'(\Omega)$. [7] We finish the proof showing that, when $\lambda \to 0$, $\widehat{\psi}(\xi, \lambda\Sigma)$ tends to

$$\widehat{\psi}(\xi, 0) := \widehat{\psi}(\xi, \Sigma) \exp\left(2\pi^2\xi^{\mathrm{tr}}\Sigma\xi\right) \in \mathcal{D}'(\Omega). \tag{10.50}$$

Define $\widehat{\psi}_\lambda(\xi) := \widehat{\psi}(\xi, \lambda\Sigma)$, and $Q(\xi) := 2\pi^2\xi^{\mathrm{tr}}\Sigma\xi$. We must show that, for every test function $f \in \mathcal{D}(\Omega)$, $\langle \widehat{\psi}_\lambda, f\rangle \to \langle \widehat{\psi}_0, f\rangle$ or equivalently

$$|\langle \widehat{\psi}_1, e^Q(1-e^{-\lambda Q})f\rangle| \to 0. \tag{10.51}$$

The function f belongs to $\mathcal{D}(K)$, for certain compact set $K \subset \Omega$. The continuity of the linear functional $\widehat{\psi}_1$ implies the existence of constants C and m (dependent on K) such that, for every $g \in \mathcal{D}(K)$,

$$|\langle \widehat{\psi}_1, g\rangle| \leq C \sum_{\alpha:|\alpha|\leq m} \left\|\partial^\alpha g\right\|, \tag{10.52}$$

where $\|\cdot\|$ denotes $\|\cdot\|_{L^\infty(K)}$. Therefore,

$$|\langle \widehat{\psi}_1, e^Q(1-e^{-\lambda Q})f\rangle| \leq C \sum_{\alpha:|\alpha|\leq m} \left\|\partial^\alpha(fe^Q)(1-e^{-\lambda Q})\right\|.$$

The Leibniz rule and the triangular inequality imply that

$$\begin{aligned}\left\|\partial^\alpha(fe^Q)(1-e^{-\lambda Q})\right\| &\leq \sum_{\beta:\beta\leq\alpha} \binom{\alpha}{\beta} \left\|\partial^\beta(1-e^{-\lambda Q})\right\| \left\|\partial^{\alpha-\beta}(fe^Q)\right\| \\ &= (1-e^{-\lambda Q}) \left\|\partial^\alpha(fe^Q)\right\| + O(|\lambda|),\end{aligned}$$

thus

$$|\langle \widehat{\psi}_1, e^Q(1-e^{-\lambda Q})f\rangle| \leq C(1-e^{-\lambda Q}) \sum_{\alpha:|\alpha|\leq m} \left\|\partial^\alpha(fe^Q)\right\| + O(|\lambda|) \tag{10.53}$$

that tends to zero when $\lambda \to 0$. □

Corollary 10.11. *Suppose that $i \in I$ is contained in a 1-cell S of* $\mathbf{K}$. *Then, for every $\varepsilon > 0$ and every σ such that $0 < \varepsilon < \sigma$,*

$$\phi^i(m, \sigma) = \int_{E_i} \phi^i(x, \varepsilon) \frac{\exp\left(-\frac{(x-m)^2}{2(\sigma-\varepsilon)}\right)}{\sqrt{2\pi(\sigma-\varepsilon)}}\, \mathrm{d}x + p_i(m; \sigma), \tag{10.54}$$

where $p_i(m;\sigma)$ is a polynomial in m whose coefficients depend on σ.

[7] Given $u \in \mathcal{D}'(\Omega)$ and $f \in C^\infty(\Omega)$, the distribution $uf \in \mathcal{D}'(\Omega)$ is defined by $\langle uf, g\rangle := \langle u, fg\rangle$.

Proof. As long as each $i \in I$ is contained in a 1-cell S of $\mathbf{K}$, we can write $S = \{i\} \sqcup S_2$ and equation (10.44) makes sense. Therefore, we apply Proposition 10.10 to conclude that, on $E_i \setminus \{0\}$,

$$\begin{aligned} \widehat{\phi}^i(\xi,\sigma) &= \widehat{\phi}^i(\xi,0)\exp(-2\pi^2\sigma\xi^2) && (10.55)\\ &= \widehat{\phi}^i(\xi,0)\exp(-2\pi^2\varepsilon\xi^2)\exp(-2\pi^2(\sigma-\varepsilon)\xi^2) && (10.56)\\ &= \widehat{\phi}^i(\xi,\varepsilon)\exp(-2\pi^2(\sigma-\varepsilon)\xi^2). && (10.57)\end{aligned}$$

On the whole E_i, we must add a distribution supported on $\{0\}$: a finite linear combination of derivatives of δ_0, see Proposition E.1. We know that $\widehat{\phi}^i(\xi,\varepsilon)$ is an element of $\mathcal{S}'$, therefore we can apply the convolution theorem to conclude.

Remark that p_i does not depend on ε. In fact, one has $\widehat{\phi}^i(\xi,\sigma) = \widehat{\phi}^i(\xi,0)\exp(-2\pi^2\sigma\xi^2) + \widehat{p_i}(\xi;\sigma)$ as elements of $\mathcal{D}'(\mathbb{R})$. □

Remark 10.12. As an element of $D'(\mathbb{R})$, the function $\frac{\exp\left(-\frac{(x-m_i)^2}{2(\sigma-\varepsilon)}\right)}{\sqrt{2\pi(\sigma-\varepsilon)}}$ tends to the Dirac mass δ_m when $\varepsilon \to \sigma$. If we assume that $\varphi(x,\lambda)$ is continuous in (x,λ), then the limit of the integral in (10.54) equals $\varphi(m,\sigma)$ and therefore $p_i(m;\sigma) = 0$.

Proposition 10.8 established that

$$\phi^S(m,\Sigma) = \sum_{i\in S}\phi^i(m_i,\sigma_{ii}) + D(S,\Sigma), \tag{10.58}$$

Therefore, for every $\varepsilon > 0$, we can write

$$\phi^S(m,\Sigma) = \sum_{i\in S}\int_{E_i}\phi^i(x,\varepsilon)\frac{\exp\left(-\frac{(x-m_i)^2}{2(\sigma_{ii}-\varepsilon)}\right)}{\sqrt{2\pi(\sigma_{ii}-\varepsilon)}}\,\mathrm{d}x + \sum_{i\in S}p_i(m_i;\sigma_{ii}) + D(S,\Sigma). \tag{10.59}$$

The polynomial $p_i(m_i;\sigma_{ii})$ can be written as a linear combination of moments of $\rho(m,\sigma)$, with coefficients that depend on σ. In turn, the integrals above resemble moments, which motivates the following definition of generalized moment functions; Proposition 10.16 shows that they are cocycles.

Definition 10.13. A map $\varphi : \mathbb{R}\times(0,\infty) \to \mathbb{C}$ is said to be a ***generalized moment function (gmf)*** associated to the family $g = \{g_\varepsilon : \mathbb{R} \to \mathbb{C}\}_{\varepsilon>0}$ if

1. $g_\varepsilon(x)\exp(-ax^2)$ is integrable for every $a > 0$,
2. $\varphi(m,\sigma) = \frac{1}{\sqrt{2\pi(\sigma-\varepsilon)}}\int_{\mathbb{R}} g_\varepsilon(x)\exp\left(-\frac{(x-m)^2}{2(\sigma-\varepsilon)}\right)\mathrm{d}x$ whenever $\sigma > \varepsilon$.

Analogously, we define ***generalized moderate moment function (gmmf)*** by replacing the functions g_ε by elements of $\mathcal{S}'(\mathbb{R})$.

We write $\varphi(g)$ or $\varphi(m,\sigma|g)$ to emphasize the dependency on g.

Remark that generalized moment functions can be added to obtain a new one. Their name is justified by the following example.

Example 10.14. Let $h(x)$ be a measurable function of $x \in \mathbb{R}$ bounded by Ce^{Ax}, for certain constants $A \in \mathbb{R}$, $C \in \mathbb{R}_+^*$. For every $\varepsilon > 0$, set

$$f_\varepsilon(x) := \frac{1}{\sqrt{2\pi\varepsilon}}\int_{\mathbb{R}} h(z)e^{-\frac{1}{2}\frac{(x-z)^2}{\varepsilon}}dz. \tag{10.60}$$

Then $\varphi(m,\sigma) := f_\sigma(m)$ is a generalized moment associated to the family $\{f_\varepsilon\}_{\varepsilon>0}$.

Proof. If h belongs to L^1, we can take the Fourier transform:

$$\widehat{f_\varepsilon}(\xi) = \widehat{h}(\xi)e^{-2\pi^2\varepsilon\xi^2} \tag{10.61}$$

Therefore, if $\eta > \varepsilon > 0$,

$$\widehat{f_\eta}(\xi) = \widehat{h}(\xi)e^{-2\pi^2\eta\xi^2} = \widehat{f_\varepsilon}(\xi)e^{-2\pi^2(\eta-\varepsilon)\xi^2}; \tag{10.62}$$

which implies the desired set of equations by taking the inverse Fourier transform.

To establish the general case, we multiply h by the characteristic function of the interval $[-N, N]$, obtaining a function $h^{(N)}$ which belongs to L^1, and take the limit when N tends to ∞, in the formula

$$\begin{aligned} &\frac{1}{\sqrt{2\pi\eta}}\int_{\mathbb{R}} h^{(N)}(z)e^{-\frac{1}{2}\frac{(y-z)^2}{\eta}}\,\mathrm{d}z \\ &\quad = \frac{1}{\sqrt{2\pi(\eta-\varepsilon)}}\frac{1}{\sqrt{2\pi\varepsilon}}\int_{\mathbb{R}}\int_{\mathbb{R}} h^{(N)}(u)e^{-\frac{1}{2}\frac{(x-u)^2}{\varepsilon}}e^{-\frac{1}{2}\frac{(y-x)^2}{\eta-\varepsilon}}\,\mathrm{d}u\,\mathrm{d}x, \end{aligned} \tag{10.63}$$

which is justified by Fubini's theorem and Lebesgue's theorem of dominated convergence. □

Remark 10.15. With $h(x) = \exp(x)$, we obtain a generalized moment that is not a generalized moderate moment.

An ***axial cochain*** is a cochain $\Phi = \{\phi^S\}_{S\in\mathrm{Ob}\,\mathbf{S}}$ that verifies

$$\phi^S(m, \Sigma) = \sum_{i\in S}\varphi(m_i, \sigma_i|g^i), \tag{10.64}$$

where $\varphi(m_i, \sigma_i|g^i)$ is a gmmf associated to a family $g^i = \{g^i_\varepsilon\}_{\varepsilon>0} \subset \mathcal{S}'(\mathbb{R})$.

The following proposition implies that axial cochains are cocycles.

Proposition 10.16. *For any $\varepsilon > 0$ and any collection $\{f^i_\varepsilon\}_{i\in S}$ in $\mathcal{S}'(\mathbb{R})$, the functionals $\psi^S(m, \Sigma) = \sum_{i\in S}\psi^i(m_i, \sigma_{ii})$, with*

$$\psi^i(m_i, \sigma_{ii}) := \int_{E_i} f^i_\varepsilon(x)\exp\left(-\frac{(x-m_i)^2}{2(\sigma_{ii}-\varepsilon)}\right)(2\pi(\sigma_{ii}-\varepsilon))^{-1/2}\,\mathrm{d}x,$$

defined for $\Sigma \gg \varepsilon I$, satisfy the cocycle equation

$$\psi^S(m, \Sigma) = (S_1.\psi^{S_2})(m, \Sigma) + \psi^{S_1}(m_1, \Sigma_{11}) \tag{10.65}$$

for any $\Sigma \gg \varepsilon I$.

Proof. Let $S = S_1 \sqcup S_2$ be any partition. Equation (10.65) is equivalent to the identity

$$(S_1.\psi^{S_2})(m, \Sigma) \overset{(!)}{=} \sum_{i\in S_2}\int_{E_i} f^i_\varepsilon(x)\exp\left(-\frac{(x-m_i)^2}{2(\sigma_{ii}-\varepsilon)}\right)(2\pi(\sigma_{ii}-\varepsilon))^{-1/2}\,\mathrm{d}x, \tag{10.66}$$

that can be interpreted as invariance under conditioning.

According to Proposition 10.6, there is a variable $Y_2 \sim N_{|S_2|}(0, \Sigma'_{22})$ such that

$$(S_1.\psi^{S_2})(m,\Sigma) = \mathbb{E}\left(\psi^{S_2}(m_2 - Y_2, \Sigma_{22} - \Sigma'_{22})\right)$$
$$= \sum_{i\in S_2} \mathbb{E}\left(\psi^i(m_2[i] - Y_2[i], \Sigma_{22}[i,i] - \Sigma'_{22}[i,i])\right)$$

The symbols $[i]$, $[i,i]$ denote components. For the last equality, we have used the definition of ψ and the linearity of expectations. The marginal $Y_2[i]$ has a univariate normal distribution with mean 0 and variance $\Sigma'_{22}[i,i]$. Hence,

$$(S_1.\phi^{S_2})(m,\Sigma) = \sum_{i\in S_2} \int_{E_i} \left(\int_{E_i} f^i_\varepsilon(x) \frac{\exp\left(-\frac{(x-m_2[i]-y_2[i])^2}{\Sigma_{22}[i,i]-\Sigma'_{22}[i,i]-\varepsilon}\right)}{\sqrt{2\pi(\Sigma_{22}[i,i] - \Sigma'_{22}[i,i] - \varepsilon)}} \right) \frac{\exp\left(-\frac{(y_2[i])^2}{\Sigma'_{22}[i,i]}\right)}{\sqrt{2\pi\Sigma'_{22}[i,i]}} \,\mathrm{d}y_2. \tag{10.67}$$

Each double integral is a convolution:

$$f^i_\varepsilon *_x [G_{\Sigma_{22}[i,i]-\Sigma'_{22}[i,i]-\varepsilon} *_{y_2} G_{\Sigma'_{22}}]^{\smile}(m_2[i])$$

(where $f^{\smile}(x) = \check{f}(x) = f(-x)$). Under the Fourier transform, this convolution becomes a multiplication

$$\widehat{f^i_\varepsilon}(\xi)\exp(-2\pi^2\xi^2(\Sigma_{22}[i,i] - \Sigma'_{22}[i,i] - \varepsilon))\exp(-2\pi^2\xi^2\Sigma'_{22}[i,i])) =$$
$$\widehat{f^i_\varepsilon}(\xi)\exp(-2\pi^2\xi^2(\Sigma_{22}[i,i] - \varepsilon)), \tag{10.68}$$

which converted back to the original domain gives $\psi^i(m_2[i], \Sigma_{22}[i,i])$; the notation is such that $m_2[i] = m_i$ and $\Sigma_{22}[i,i] = \Sigma[i,i]$. So we reach the desired conclusion taking two times the Fourier transform of (10.67). $\square$

Remark now that taking $\varepsilon = 0$, $f^i_0(x) = x^k$ and $f^j_0 = 0$ when $j \neq i$, we establish that the usual moments

$$M_k(m,\sigma) = \int_{Rr} x^k \frac{\exp\left(-\frac{1}{2}\frac{(x-m)^2}{\sigma}\right)}{\sqrt{2\pi\sigma}} \,\mathrm{d}x$$

define 1-cocycles through the formula $\phi^S(m,\Sigma) = M_k(m_i, \sigma_{ii})$. Moreover, setting $f(x) = x^k$ in Example 10.14, we conclude that $f_\varepsilon(m) := M_k(m,\varepsilon)$ defines a generalized moment function.

Theorem 10.17 (Structure theorem of 1-cocycles, simplicial case)**.** *Suppose that every* 0*-cell of* **K** *belongs to a* 1*-cell. Every* 1*-cocycle is the sum of an axial cocycle and a multiple of the entropy i.e. there exist generalized moderate moment functions* $\{\varphi(g^i)\}_{i\in S}$ *and a constant* $C \in \mathbb{R}$ *such that*

$$\phi^S(m,\Sigma) = \sum_{i\in S} \varphi(m_i, \sigma_{ii}|g^i) + C\ln(|\Sigma|). \tag{10.69}$$

Proof. As we already remarked, the results in Proposition 10.8 and Corollary 10.11 imply that

$$\phi^S(m,\Sigma) = \sum_{i\in S} \varphi(m_i, \sigma_{ii}|f^i) + \sum_{i\in S} p_i(m_i;\sigma_{ii}) + D(S,\Sigma), \tag{10.70}$$

where $f^i = \{\phi^i(\cdot, \varepsilon)\}_{\varepsilon>0}$. Proposition 10.16 shows that the axial part is a cocycle, so

$$\zeta^S(m, \Sigma) = \sum_{i\in S} p_i(m_i; \sigma_{ii}) + D(S, \Sigma) \tag{10.71}$$

too.

Consider now $S = \{i\}$. The polynomial $p_i(m_i; \sigma_{ii})$ can be written in the basis given by the $M_j(m_i, \sigma_{ii})$:

$$p_i(m_i; \sigma_{ii}) = \sum_{j=0}^{k_i} c_i^j(\sigma_{ii}) M_j(m_i, \sigma_{ii}). \tag{10.72}$$

In principle the coefficients are a function of σ_{ii}, but now we use that

$$C(\{i\}, \{j\}, \Sigma) = \mathbb{E}\left(\zeta^i(m_i - Y_i, \sigma_{ii} - \sigma'_{ii})\right) - \zeta^i(m_i, \sigma_{ii}) \tag{10.73}$$

from Proposition 10.6, which gives

$$C(\{i\}, \{j\}, \Sigma) = \sum_{j=0}^{k_i} c_i^j(\sigma_{ii} - \sigma'_{ii}) \mathbb{E}\left(M_j(m_i - Y_i, \sigma_{ii} - \sigma'_{ii})\right) - \sum_{j=0}^{k_i} c_i^j(\sigma) M_j(m_i, \sigma_{ii}) \tag{10.74}$$

because D vanishes in this case. The expectation $\mathbb{E}\left(M_j(m_i - Y_1, \sigma_{11} - \sigma'_{11})\right)$ equals $M_j(m_i, \sigma_{ii})$, because M_j is itself an axial cocycle (cf. equation (10.68) in the proof of Proposition 10.16). Reading the equation

$$C(\{i\}, \{j\}, \Sigma) = \sum_{j=0}^{k_i} (c_j^i(\sigma_{11} - \sigma'_{11}) - c_j^i(\sigma_{11})) M_j(m_i, \sigma), \tag{10.75}$$

we conclude that $c_j^i(\sigma_{ii} - \sigma'_{ii}) - c_j^i(\sigma_{ii})$ vanishes for every degree $j > 0$ (the corresponding $M_j(m_i, \sigma_{ii})$ depends on m) and hence $c_i^j(\sigma_{ii})$ must be constant, say c_i^j. Thus $\phi^S(m, \Sigma)$ is a sum of an axial cocycle, a linear combination of moments, and a last term that depends only on the variance; since everything else is a cocycle, the latter must be a cocycle too, that equals $a \ln(|\Sigma|) + \sum_{i\in S} b_i$:

$$\phi^S(m, \Sigma) = \sum_{i\in S} \phi(m_i, \sigma_{ii} | \tilde{f}^i) + \sum_{i\in S} \sum_{j=1}^{k_i} c_j^i M_j(m_i, \sigma_{ii}) + \sum_{i\in S} b_i + a \ln(|\Sigma|), \tag{10.76}$$

The moments themselves are generalized moderate moment functions (and M_0 is constant), which allows us to merge the first three sums. □

Remark 10.18. These simplicial 1-cocycles give a very particular role to the coordinate axes used to define the simplicial information structure. This is not so natural, since this basis was introduced just for convenience. We are forced to consider more general structures, that are introduced in the next chapter. We shall see that the only cocycles that survive are multiples of the dimension (obtained above when all the b_i in (10.76) are equal) and the entropy. Cf. Theorem 11.30.

10.3.4 Axial cochains and the heat equation

We close this section with some remarks about axial cocycles in connection to the heat equation.

A gmf or gmmf $\varphi(m,\sigma|g)$ of $m \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+^*$ is not only measurable: it is infinitely differentiable in (m,σ) by Lebesgue's dominated convergence theorem.

Lemma 10.19. *A gmf (or gmmf) $\varphi = \varphi(g)$ satisfies the heat equation on $\mathbb{R} \times \mathbb{R}_+^*$:*

$$\frac{\partial \varphi}{\partial \sigma}(m,\sigma) = \frac{1}{2}\frac{\partial^2 \varphi}{\partial m^2}(m,\sigma) \tag{10.77}$$

Proof. By derivation with respect to m under the integral sign, we get

$$\frac{\partial \varphi}{\partial m} = \frac{1}{\sqrt{2\pi(\sigma-\varepsilon)}}\int g_\varepsilon(x)\frac{x-m}{\sigma-\varepsilon}e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma-\varepsilon}}, \tag{10.78}$$

then

$$\frac{\partial^2 \varphi}{\partial m^2} = -\frac{1}{\sqrt{2\pi(\sigma-\varepsilon)^3}}\varphi(m,\sigma) + \frac{1}{\sqrt{2\pi(\sigma-\varepsilon)^5}}\int g_\varepsilon(x)(x-m)^2 e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma-\varepsilon}}. \tag{10.79}$$

And by derivation with respect to σ we get directly the half of this sum. □

Proposition 10.20. *Every solution of the heat equation that is a differentiable function of $\sigma > 0$ to the space of tempered distributions $\mathcal{S}'$ in $m \in \mathbb{R}$, is equal to a generalized moderate moment function $\varphi(f)$, for a family $f = \{f_\varepsilon\}$ in $\mathcal{S}'(\mathbb{R})$. Moreover, if φ is a locally integrable function of m, then φ is also generalized moment function (gmf).*

Proof. Taking the Fourier transform in m we obtain

$$\frac{\partial}{\partial \sigma}\widehat{\varphi}(\xi,\sigma) = -2\pi^2\xi^2\widehat{\varphi}(\xi,\sigma). \tag{10.80}$$

For any $\varepsilon > 0$, and any element $\widehat{f_\varepsilon}$ in $\mathcal{S}'$, there exists a unique solution $\widehat{\varphi}(\xi,\sigma)$ which coincides with $\widehat{f_\varepsilon}$ for $\sigma = \varepsilon$, and it is given by

$$\forall \sigma > \varepsilon, \quad \widehat{\varphi}(\xi,\sigma) = e^{-2\pi^2(\sigma-\varepsilon)\xi^2}\widehat{f}(\xi). \tag{10.81}$$

The compatibility between different choices of $\widehat{f_\bullet}$ is: for every $\eta,\varepsilon > 0$ such that $\eta > \varepsilon > 0$,

$$\widehat{\varphi}(\xi,\eta) = \widehat{f_\eta}(\xi) = e^{-2\pi^2(\eta-\varepsilon)}\widehat{f_\varepsilon}(\xi). \tag{10.82}$$

The first statement follows by taking the inverse Fourier transform to these last two equations. The second follows from the fact that $f_\varepsilon(x) = \varphi(x,\varepsilon)$. □

Chapter 11

Grassmannian categories

11.1 Grassmannian information structures

11.1.1 Definition

Let E be a vector space over a commutative field $\mathbb{K}$. A ***grassmannian category*** $\mathbf{S}$ of E is defined at the level of objects by a subset of the full grassmannian $\mathrm{Gr}(E)$ of vector subspaces of E, which contains E and is closed by conditional intersection, i.e. $V, W \in \mathrm{Ob}\,\mathbf{S}, \exists Z \in \mathrm{Ob}\,\mathbf{S}, Z \subseteq V, Z \subseteq W$ implies $V \cap W \in \mathrm{Ob}\,\mathbf{S}$. Arrows are inclusions: $V \to W$ if and only if $V \subseteq W$. Such a category is a poset, having a maximal element and conditional finite products.

For each $V \in \mathrm{Ob}\,\mathbf{S}$, we denote by E_V the quotient space E/V. If $V \subseteq W$, we have a canonical surjection $E_V \to E_W$. This gives a covariant functor $\mathscr{E} : \mathbf{S} \to \mathbf{Meas}$ if each vector space is equipped with its Borel σ-algebra. The image $\mathbf{C} := \mathscr{E}(\mathbf{S})$ is a poset with a final element $1 = \{0\} = E/E$, and restricted products, i.e. if $E_Z \to E_V$ and $E_Z \to E_W$, meaning $Z \subseteq V$ and $Z \subseteq W$, the arrows $E_{V\cap W} \to E_V$ and $E_{V\cap W} \to E_W$ are universal factorizations of any pair of arrows to E_V and E_W from a common source.

Proposition 11.1. *The pair $(\mathbf{S}, \mathscr{E})$ is an information structure.*

Proof. Only the property 6 in Definition 1.6 is not immediate. Consider $U, V \in \mathrm{Ob}\,\mathbf{S}$ such that $W = U \wedge V \in \mathrm{Ob}\,\mathbf{S}$, and the corresponding projections $\pi^{UW} : E_W \to E_U$, $\pi^{VW} : E_W \to E_V$; we must prove that $\iota : E_W \to (E_U, E_V), [w] \mapsto (\pi^{UW}([w]), \pi^{VW}([w]))$ is a injection. An element $w \in E$ defines a class $[w] \in \ker \iota$ if and only if $w - 0_E \in U$ and $w - 0_E \in V$, thus $w \in U \cap V$ and $[w] = 0$. □

For each $V \in \mathrm{Ob}\,\mathbf{S}$, the objects W of $\mathbf{S}$ that contain V form a commutative monoid $\mathscr{S}_V$ for the intersection, whose neutral element is E. We denote by $\mathscr{A}_V$ the algebra over $\mathbb{K}$ which is generated by this monoid. If $V' \subseteq V$, we have a natural injective morphism $j_{V'V} : \mathscr{A}_V \to \mathscr{A}_{V'}$. This gives a canonical presheaf of commutative algebras $\mathscr{A}$ over $\mathbf{S}$. We are interested in the ringed topos $(\mathbf{S}, \mathscr{A})$, and the cohomology $H^\bullet_{\mathbf{S}}(\mathbb{K}; \mathscr{V})$ of certain $\mathscr{A}$-modules $\mathscr{V}$ associated to operations on probability laws supported on subsets of the vector spaces $\{E_V\}_{V\in\mathrm{Ob}\,\mathbf{S}}$, that defines topological invariants of these probability laws.

The simplicial information structures studied up to now are examples of grassmannian structures. The full grassmannian manifold itself gives an infinite grassmannian category; it can be seen as the natural linear analogue of the full simplex in

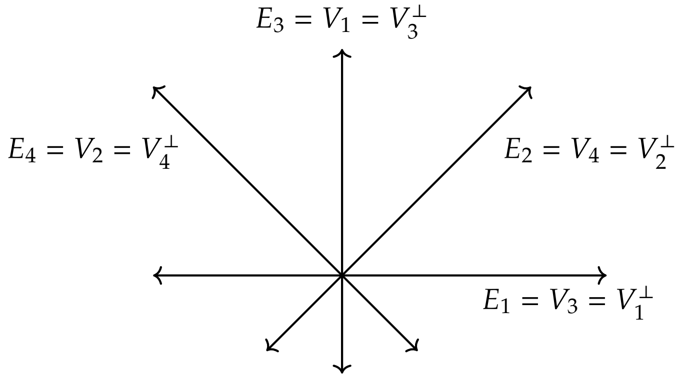

Figure 11.1: Geometrical representation of the categories $\mathbf{S}_2$ and $\mathbf{C}_2 = \mathscr{E}(\mathbf{S}_2)$ from Example 11.2.

the case of finite probabilities. Other interesting examples consist of subspaces containing (at least) a given subspace: in other terms, a pencil through given projective subspaces in the projective n-dimensional space $\mathbb{P}_n(\mathbb{K})$.

Example 11.2. Set $E = \mathbb{R}^2$ with the standard euclidean structure, and consider the Grassmannian category $\mathbf{S}_2$ made by six objects E, V_1, V_2, V_3 V_4, and $\langle 0\rangle$, where the Vs are lines through 0, such that V_4, V_1 and V_2 make respectively an angle of $\pi/4$, $\pi/2$ and $3\pi/4$ with V_3. Each quotient $E_i := E/V_i$ can be identified with $V_i^\perp$, in such a way that $\mathbf{C}_2 = \mathscr{E}(\mathbf{S}_2)$ is the same arrangement of lines and the map $\pi_i : E \to E_i$, induced by $\langle 0\rangle \to V_i$ in $\mathbf{S}$, is the orthogonal projection on E_i. The situation is depicted in Figure 11.1.

Remark also that if $\mathbb{K}$ is a finite field, $\mathbf{S}$ is a concrete category in the sense of 1.1; in fact E_V corresponds to the partition of $E = \Omega$ into affine spaces parallel to V.

A grassmannian category $\mathbf{S}$ over the field $\mathbb{R}$ of real numbers or any local field of characteristic zero, for instance $\mathbb{C}$ or $\mathbb{Q}_p$, with E finite dimensional, satisfies additionally Proposition 9.1, the σ-algebra $\mathfrak{B}_V$ associated to each E_V being the Borel algebra of the vector space E_V, which is Polish (i.e. separable completely metrizable, which implies second countable).

11.1.2 Measures

In what follows, because we want to study gaussian laws or some related probability laws, we restrict ourselves to the case where $\mathbb{K} = \mathbb{R}$ and E has finite dimension.

To study the probability laws supported on affine subspaces of E, we introduce the pre-cosheaf of affine supports $\mathscr{M}$ over $\mathbf{S}$: for $V \in \operatorname{Ob}\mathbf{S}$, the set $\mathscr{M}_V$ contains all the affine (nonempty) subspaces in E_V, which corresponds by projection $E \twoheadrightarrow E_V$ to the affine subspaces of E whose direction (tangent space) is any vector space that contains V. If $V \subseteq W$, the linear map $\pi^{WV} : E_V \to E_W$ sends $\mathscr{M}_V$ into $\mathscr{M}_W$.

It is useful to restrict this functor to subsets $\mathscr{N}_V$ of $\mathscr{M}_V$, for $V \in \operatorname{Ob}\mathbf{S}$, that are closed under the projections $\pi^{WV} : E_V \to E_W$. In the perspective of conditioning, we also require that the fibers of the projections belong to $\mathscr{N}_V$, and that each $\mathscr{N}_V$ is closed by nonempty intersection. If all these conditions are satisfied, we say that the cosheaf $\mathscr{N}$ is ***admissible*** or ***adapted***. This allows us to recover the examples treated in Chapters 9 and 10, where the only supports considered were parallel to the coordinate axes. The restriction to $\mathscr{N}_V$ in $\mathscr{M}_V$ is analog to the restriction to a set of probabilities Q_X more general than the full simplex $\Pi(X)$, for a variable X in a finite information structure, cf. Chapter 1.

Definition 11.3. We say that the admissible cosheaf $\mathscr{N}$ is ***minimal*** if for every $V \in \operatorname{Ob}\mathbf{S}$ and $A \in \mathscr{M}_V$, the space A is contained in $\mathscr{N}_V$ if and only if there exists $W \in \operatorname{Ob}\mathbf{S}$ such that $V \subseteq W$, and the direction (tangent space) $T(A)$ of A in E_V coincides with W/V.[1] (Such a W is necessarily unique.)

The minimal supports can be though as those "generated" by the spaces in $\mathbf{S}$. This hypothesis is verified by the full grassmannian manifold, with $\mathscr{M} = \mathscr{N}$; by the simplicial information structures if each $\mathscr{N}_V$ only contains spaces that are parallel to those generated by elements of the preferred basis; and the structure $(\mathbf{S}_2, \mathscr{E})$ introduced in Example 11.2 if the only affine supports considered are either the full space E, any line parallel to some V_i, or singletons.

Remark 11.4. Suppose $\mathscr{N}$ minimal, take $A' \in \mathscr{N}_V$ parallel to W'/V, and consider $W \in \operatorname{Ob}\mathbf{S}$, then the fibers in A of the restriction of the projection π^{WV} are parallel to $(W \cap W')/V$; by admissibility these fibers must belong to $\mathscr{N}_V$. This reflects the information axiom: the inclusions $V \hookrightarrow W$ and $V \hookrightarrow W'$ imply that $W \cap W'$ belongs to $\mathbf{S}$.

We also introduce the set $\mathscr{L}$ of Lebesgue measures over the pre-cosheaf of supports: for each $V \in \operatorname{Ob}\mathbf{S}$, $\mathscr{L}_V$ denotes the set of pairs (A, λ) where $A \in \mathscr{M}_V$ and λ is a positive Lebesgue measure of support exactly A. Note that the set $\mathscr{L}_V(A)$ for a given $A \in \mathscr{M}_V$ depends only of the tangent vector space $T(A)$, and coincide with the strictly positive cone $\Lambda_+(T(A))$ in the exterior power $\Lambda^{max}(T(A))$. For every $V \in \operatorname{Ob}\mathbf{S}$ we consider the set $\mathscr{L}_V$ as a bundle over $\mathscr{M}_V$, the fiber over $A \subset E_V$ is $\Lambda_+(T(A))$, thus this bundle is obtained by pullback from a cone bundle of rank one over the Grassmann manifold of E_V. This is a principal bundle for the multiplicative group $\mathbb{R}^*_+$. Everything can be restricted over a subfunctor $\mathscr{N}$ of M.

11.1.3 Orthogonal embeddings

Let $\mathbf{S}$ be any grassmannian category of E. Choose arbitrarily an euclidean metric Q on E. For any $V \in \operatorname{Ob}\mathbf{S}$ the orthogonal $V^\perp$ is a supplementary space of V in E, then we can identify E_V with the subspace $V^\perp$ of E. Also, when $V \subseteq W$, we have $W^\perp \subseteq V^\perp$, then we can identify the quotient $W/V = E_V/E_W$ with the orthogonal $W \cap V^\perp$ of V in W. Therefore, when $V \subseteq W$, the morphism $\pi^{WV} : E_V \to E_W$ can be identified with the linear projection from $V^\perp$ to $W^\perp$ parallel to $W \cap V^\perp$, which is the orthogonal projection. In this way, every affine subspace A of E_V is identified with an affine subspace of E.

In summary, given the grassmannian structure $(\mathbf{S}, \mathscr{E})$, every euclidean metric on E gives a lift of $\mathbf{C} = \mathscr{E}(\mathbf{S})$ into the category of orthogonal projections in E. We name this lift the embedding of $\mathbf{C}$ associated to Q, denote it by J_Q, and write $\mathbf{C}_Q$ its image.

Every such embedding J_Q induces a metric on every affine subspaces A of E_V, and consequently a canonical Lebesgue measure $\lambda_Q(A)$ on A, that is $\lambda = \mathrm{d}x_1 \ldots \mathrm{d}x_n$ in

[1] A point $e \in E_V$ is an affine space associated to the trivial vector space $\langle 0 \rangle$ and its tangent space $T(e)$ is taken here to be the trivial space too.

orthonormal coordinates. Therefore the bundle $\mathscr{L}$ is trivializable, but not canonically trivial.

Let us fix a metric Q on E, and consider an element $A \in \mathscr{M}_W$, we have a short exact sequence

$$0 \to T(A) \cap W/V \to A \to \pi^{WV}(A) \to 0; \tag{11.1}$$

Then, from Proposition F.9 (Weil's formula), there exists a strictly positive number $c(A; W/V)$ such that

$$\lambda_A = c(A; W/V)\lambda_{\pi^{WV}(A)}.\lambda_{T(A)\cap W/V}. \tag{11.2}$$

Changing the metric Q in another metric Q' induces the multiplication of every measure λ_A by a strictly positive number $\Delta(A; Q, Q')$; we name this function over $\mathscr{M}$ the discriminant of Q' with respect to Q.

A multiple cQ with $c > 0$ gives the same embedding J_Q and the same category $\mathbf{C}_Q$.

Definition 11.5. A grassmannian information structure $(\mathbf{S}, \mathscr{E})$ is said to be ***orthogonally closed*** with respect to the euclidean metric Q on E, if $V^\perp \in \mathbf{C}_Q$ implies $V \in \mathbf{C}_Q$, or equivalently if $V \in \operatorname{Ob}\mathbf{S}$ implies $V^\perp \in \operatorname{Ob}\mathbf{S}$.

Examples of orthogonally closed gassmannian structures:

1. The full Grassmann manifold of E. The simplicial structures of the preceding chapters, given by a basis of E. We see on this example that, given the category $\mathbf{C}_Q$, the conformal structure of the metric Q is not unique in general.
2. The category $\mathbf{C}_2$ of Section 11.3.4 is also closed with respect to an euclidean structure. In this example, the conformal structure of Q is unique.

For sufficiently general orthogonal closed categories the unicity of conformal structure is the rule.

11.2 Gaussian laws

11.2.1 Mean and covariance

Definition 11.6. The pre-cosheaf $\mathscr{P} : \mathbf{S} \to \mathbf{Sets}$ of gaussian laws over $\mathbf{S}$ maps $V \in \operatorname{Ob}\mathbf{S}$ to the set $\mathscr{P}_V$ of probability measures on $(E_V, \mathfrak{B}_V)$ that have a support A contained in $\mathscr{M}_V$ and a gaussian density with respect to a Lebesgue measure on A. The direct images give a covariant functor over $\mathbf{S}$ (affine images of gaussians are gaussians, see Appendix D).

If we work with restricted supports $\mathscr{N} \subseteq \mathscr{M}$, we restrict the probability laws accordingly.

Every element ρ of $\mathscr{P}_V$ has a support $A(\rho)$ in $\mathscr{M}_V$ and is absolutely continuous with respect to every Lebesgue measure λ such that $(A(\rho), \lambda) \in \mathscr{L}_V$. We denote by $g(\rho; \lambda)$ or $g_\lambda(\rho)$ the density $\mathrm{d}\rho/\mathrm{d}\lambda$; its integral with respect to λ equals 1. The mean $M(\rho)$ is the point of $A(\rho) \subseteq E_V$ defined by

$$M(\rho) = \int_A X g(\rho; \lambda)(X)\, \mathrm{d}\lambda(X). \tag{11.3}$$

This equation shows that the mean does not depend on any choice. The gaussian measure ρ is defined by its mean $M(\rho)$ and by a covariance Σ_ρ, which is the inverse of the bilinear map $-D^2_X \ln g(\rho;\lambda)(X)$. Choosing an euclidean metric on A, the covariance is expressed by a symmetric matrix of rank $\dim(A)$, which is positive definite.

More precisely, there exist a nondegenerate symmetric positive bilinear form B on $T := T(A)$—which defines an element element of the symmetric power $S^2(T^*)$—such that, if we choose a point N in A, there exist a constant C_N and a linear form L_N that verify, for every $x \in A$,

$$\ln g(x) = -\frac{1}{2}B(x-N, x-N) + L_N(x-N) + C_N. \tag{11.4}$$

where $g := g(\rho;\lambda)$. The rule $x \mapsto d\ln g_x$ defines a map from A to T^*, characterized by

$$\forall X \in T, \quad d\ln g(x).X = -B(x-N, X) + L_N(X). \tag{11.5}$$

It can be differentiated again in a canonical way to obtain

$$Dd\ln g_x(X,Y) = -B(Y,X) = -B(X,Y), \tag{11.6}$$

that is an element of $S^2(T^*)$. Since B is nondegenerate, the map $\beta : T \to T^*, x \mapsto B(x,\cdot)$ is invertible and the covariance is the symmetric bilinear form Σ (on T^*) induced by $\sigma = \beta^{-1} : T^* \to T$; we write $\Sigma = B^{-1}$. The previous considerations show that Σ does not depend on N, which vanishes under differentiation.

Moreover, the covariance does not depend on the choice of reference measure: if we change the Lebesgue measure λ on $T(A)$ (or A) in $\lambda' = C\lambda$ for $C > 0$, we change g in $g' = g/C$, then $\ln g' = \ln g - C$, thus $d\ln g$ is unchanged and we have $\Sigma' = \Sigma$.

Nevertheless, the trace and the determinant of Σ, which appear in many formulas of probability theory, do change. They are not invariants of a bilinear form on $T = T(A)$ but of an endomorphism of T. In fact, the only invariant of a nondegenerate symmetric positive bilinear form is the dimension of T, equal to its rank (cf. Sylvester's law of inertia).

Every linear isomorphism from T^* to T is enough to define $\mathrm{Tr}(\Sigma)$ and $\det(\Sigma)$, which explains why this problem has not arisen up to this point.

A Lebesgue measure is sufficient for defining $\det(\Sigma)$: the measure λ defines a basis of $\Lambda^{max}(T^*)$ then a dual basis of $\Lambda^{max}(T)$, and $\det_\lambda(\Sigma)$ is the matrix of $\Lambda^{max}(\Sigma)$ in these basis. Changing λ in $C\lambda$, changes the dual basis in λ^{-1}/C, then $\det_{\lambda'}(\Sigma)$ equals $C^{+2}\det(\Sigma)$. We will recover that when studying the entropy of gaussian laws.

Remark 11.7. In other terms, Σ belongs to the symmetric power $S^2(T)$, and the functor Λ^{max} gives an element $\Lambda^{max}\Sigma$ in $S^2(\Lambda^{max}T)$ which is equal to $\Lambda^{max}T \otimes \Lambda^{max}T$. Taking a basis λ^{-1} of $\Lambda^{max}T$, the determinant $\det_\lambda \Sigma$ is the coordinate of $\Lambda^{max}\Sigma$ in the basis $\lambda^{-1} \otimes \lambda^{-1}$ of $S^2(\Lambda^{max}T)$, and changing the basis λ^{-1} of $\Lambda^{max}T$ into $C^{-1}\lambda^{-1}$ multiply the coordinate in $S^2(\Lambda^{max}T)$ by C^2.

The trace is a linear form on $\mathrm{End}(E) \cong E^* \otimes E$ induced by the canonical pairing $E^* \otimes E \to \mathbb{C}, (x^*, x) \mapsto x^*(x)$ [13, Sec. II.4.3]. To introduce the trace of a bilinear form $B \in T^* \otimes T^*$, we need an isomorphism between T and T^*, that turns B into an element of $T^* \otimes T$. Such duality may come from an euclidean metric Q (a positive definite symmetric bilinear form): we choose an orthonormal basis, and take the matrix associated to Σ in this basis, then $\mathrm{Tr}_Q(\Sigma)$ is the usual trace of this matrix, that is the sum of the coefficients along the diagonal; in particular, an appropriate choice of basis diagonalizes Σ and the trace then corresponds to sum of the principal values of Σ with respect to Q [13, Sec. II.10.11].

Choosing a Lebesgue measure on the affine subspace $A \subseteq E_V$, the gaussian law ρ of support A, mean $M \in A$ and covariance $\Sigma \in S^2(T)$ is given by

$$\forall x \in A, \quad \frac{\mathrm{d}\rho}{\mathrm{d}\lambda}(x) = \rho_\lambda(x) = \frac{1}{\sqrt{\det_\lambda(2\pi\Sigma)}} e^{-\frac{1}{2}\Sigma^{-1}(x-M,x-M)}. \tag{11.7}$$

The data of the mean is equivalent to the data of the linear part L_N in the equation (11.5). The nondegeneracy of B or $\Sigma = B^{-1}$ implies the existence of a unique point M, such that, for every $x \in A$,

$$B(x-N, N-M) = L_N(x-N). \tag{11.8}$$

This implies

$$B(x-N, x-N) + L_N(x-N) + C_N = B(x-M, x-M) + C_M, \tag{11.9}$$

for the constant C_M given by

$$C_M = C_N - \frac{1}{2}B(N-M, N-M); \tag{11.10}$$

The fact that

$$C_M = -\frac{1}{2}\ln(\det_\lambda(2\pi\Sigma)), \tag{11.11}$$

follows from the celebrated Gauss formula.

11.2.2 Moments of order two

Definition 11.8. Let N be a point in $A(\rho)$ and a positive symmetric bilinear form B' on $T(A)$, the associated moment of order two of ρ is defined by

$$\begin{aligned}\Phi_2(\rho; B', N) &= \mathbb{E}_\rho(B'(x-N, x-N)) \\ &= \int_A B'(x-N, x-N) g_\lambda(\rho)(x)\, \mathrm{d}\lambda(x).\end{aligned} \tag{11.12}$$

Remark 11.9. The change of variables $Y = x - N$ identifies A with $T(A)$, without changing the Lebesgue measure; this gives the following expression for the moment:

$$\Phi_2(\rho; B', N) = \int_{T(A)} B'(Y, Y) g_{\lambda,N}(\rho)(Y)\, \mathrm{d}\lambda(Y); \tag{11.13}$$

where $g_{\lambda,N}(\rho)(Y) = g_\lambda(\rho)(Y+N)$ is the density of the image probability.

Remark 11.10. For a general probability law which is absolutely continuous with respect to λ_A, it could happen that the mean is not defined, but the moment of order two is always defined if we accept the value $+\infty$ in $\overline{\mathbb{R}}$, because the integrand is a positive function.

Proposition 11.11. *Let Q be an euclidean metric on A, and ρ the gaussian measure of support A having the mean $M \in A$ and the covariance Σ, then*

$$\Phi_2(\rho; Q, M) = \mathrm{Tr}_Q(\Sigma). \tag{11.14}$$

Proof. Let us choose an orthonormal basis with respect to Q such that Σ is represented by diagonal matrix; taking the square roots of the diagonal coefficients, we obtain a diagonal matrix that defines the unique positive square root $\Sigma^{1/2}$. Let λ be the Lebesgue measure associated to Q. Then,

$$\Phi_2(\rho; Q, M) = \frac{1}{\sqrt{\det_\lambda(2\pi\Sigma)}} \int_A Q(x-M, x-M) e^{-\frac{1}{2}\Sigma^{-1}(x-M, x-M)} \,\mathrm{d}\lambda(x). \tag{11.15}$$

If we make the change of variables

$$Y = \Sigma^{-1/2}(x - M), \tag{11.16}$$

we get

$$\Phi_2(\rho; Q, M) = (2\pi)^{-d(A)/2} \int_{T(A)} \|\Sigma^{1/2} Y\|^2 e^{-\frac{1}{2}\|Y\|^2} \,\mathrm{d}y. \tag{11.17}$$

Note $d = d(A)$ and $\sigma_1, \ldots, \sigma_d$ the spectral values of Σ counted with their multiplicity, we have

$$\|\Sigma^{1/2} Y\|^2 = \sigma_1 y_1^2 + \ldots + \sigma_d y_d^2. \tag{11.18}$$

Thus, using Fubini's theorem and the formula for the one dimensional reduced moment, we get

$$\begin{aligned}
\Phi_2(\rho; Q, M) &= \sum_{i=1}^{d} (2\pi)^{-d/2} \int_{\mathbb{R}^d} \sigma_i y_i^2 e^{-\frac{1}{2}\|Y\|^2} \,\mathrm{d}y_1 \cdots \mathrm{d}y_d \\
&= \sum_{i=1}^{d} (2\pi)^{-1/2} \int_{\mathbb{R}} \sigma_i y_i^2 e^{-\frac{1}{2}y_i^2} \,\mathrm{d}y_i \\
&= \sum_{i=1}^{d} \sigma_i = \mathrm{Tr}_Q(\Sigma);
\end{aligned}$$

which is the expected result. □

Corollary 11.12. *Let Q be an euclidean metric on A, and ρ the gaussian measure of support A, of mean $M \in A$ and covariance Σ, and let N be any point in A,*

$$\Phi_2(\rho; Q, N) = Q(M - N, M - N) + \mathrm{Tr}_Q(\Sigma). \tag{11.19}$$

Proof. Let us decompose $x - N = x - M + M - N$, then

$$\begin{aligned}
&Q(x - N, x - N) \\
&\qquad = Q(x - M, x - M) + Q(M - N, M - N) + 2Q(x - M, M - N).
\end{aligned} \tag{11.20}$$

And the linear term in $x - M$ disappears when we integrate because M is the mean. □

In our grassmannian setting, we dispose of more structure: the support A is included in a vector space E_V, and the probability law is attached to this space E_V, thus for every positive symmetric bilinear form B_V on E_V, we can define the moment of $\rho \in \mathscr{P}_V$, for B_V, without choosing N:

$$\Phi_2(\rho; B) = E_\rho(B(X,X)) = \int_A B_V(X,X) g_\lambda(X) d\lambda_A(X). \tag{11.21}$$

Corollary 11.13. *For any gaussian law $\rho \in \mathscr{P}_V$ and any euclidean metric Q_V on E_V,*

$$\Phi_2(\rho; Q) = \|M(\rho)\|_V^2 + \mathrm{Tr}_{Q|_A}(\Sigma(\rho)). \tag{11.22}$$

Proof. For $X \in A$, we have

$$\begin{aligned} Q_V(X,X) &= Q_V(X - M + M, X - M + M) &(11.23)\\ &= Q_V(X - M, X - M) + 2Q_V(M, X - M) + Q_V(M,M) &(11.24)\\ &= Q|_A(X - M, X - M) + L(X - M) + \|M(\rho)\|_V^2; &(11.25) \end{aligned}$$

where L is a linear form on $T(A)$. By applying the proposition 6, we get

$$\mathbb{E}_\rho(Q|_A(X - M, X - M) = \mathrm{Tr}_{Q|_A}(\Sigma(\rho)). \tag{11.26}$$

By the normalization of any probability law

$$\mathbb{E}_\rho(\|M(\rho)\|_V^2) = \|M(\rho)\|_V^2. \tag{11.27}$$

And by definition of the mean

$$\mathbb{E}_\rho(L(X - M)) = 0. \tag{11.28}$$

The corollary follows by addition. □

Remark 11.14. Between the moments $\Phi(\rho; Q)$, for Q defined on E_V, and the moments $\Phi_2(\rho; Q_A, N)$, where $Q|_A$ is defined on $T(A)$, the link is

$$\begin{aligned} \Phi_2(\rho; Q) &= \Phi_2(\rho; Q|_A, N) + 2\mathbb{E}_\rho(x \mapsto Q(x - N, N)) + \mathbb{E}_\rho(x \mapsto Q(N,N)) &(11.29)\\ &= \Phi_2(\rho; Q|_A, N) + 2Q(M(\rho) - N, N) + Q(N,N), &(11.30) \end{aligned}$$

thus

$$\Phi_2(\rho; Q|_A, N) = \Phi_2(\rho; Q) - 2Q(M(\rho), N) + Q(N,N). \tag{11.31}$$

11.3 Gaussian modules

In this section, we characterize the information cohomology when the coefficients are measurable functionals of gaussians laws.

11.3.1 Module of moderate functionals

Definition 11.15. The presheaf of moderate functions $\mathscr{F} : \mathbf{S} \to \mathbf{Sets}$ maps $V \in \operatorname{Ob}\mathbf{S}$ to the set $\mathscr{F}_V$ of measurable functions on $\mathscr{P}_V$ that are of moderate growth (i.e. bounded by a polynomial) in the mean variable $M(\rho)$, cf. condition 9.24.[2]

Note that an element $\Phi \in \mathscr{F}_V$ can be seen as a function $\Phi(g, \lambda)$ satisfying the equivariance relation

$$\forall C > 0, \quad \Phi(g/C, C\lambda) = \Phi(g, \lambda). \tag{11.32}$$

Let $V \subseteq W$ be a pair of objects of $\mathbf{S}$ (interpreted as random variables with values in E_V and E_W, respectively), and A an element of $\mathscr{M}_V$; the map $\pi^{WV} : E_V \to E_W$ induces an affine projection π from A onto $B = \pi^{WV}(A)$, the fiber can be identified with $K = T(A) \cap W/V$. Let us choose a Lebesgue measure λ on A, a Lebesgue measure μ on B and define the Lebesgue measure ν on K by $\lambda = \mu.\nu$, cf. Weil's formula (Proposition F.9). Consider $\rho \in \mathscr{P}_V$, described by λ of support $A(\rho)$ and the density g; for every $y \in B = \pi^{WV}(A)$, we define the conditioned measure $\rho|_{W=y}$ by its support, which is the affine subspace $A_y := \{\, x \in A \mid \pi^{WV}(x) = y \,\}$ of A, and by its density with respect to ν_y corresponding to ν, which is defined as

$$g(\rho|_{W=y}; \nu_y) = \frac{g(\rho; \lambda)|_{A_y}}{g(\pi_*^{WV}\rho; \mu)(y)}. \tag{11.33}$$

If we replace λ by $C_A\lambda$ and μ by $C_B\mu$, the measure ν is replaced by $(C_A/C_B)\nu$, the density $g(\rho; \lambda)$ changes into $g(\rho; C_A\lambda) = g(\rho; \lambda)/C_A$, and the density $g(\pi_*^{WV}\rho; \mu)(y)$ into $g(\pi_*^{WV}\rho; \mu)(y)/C_B$, hence the new conditional density is

$$\frac{g(\rho|_{W=y}; \nu_y)/C_A}{g(\pi_*^{WV}\rho; \mu)(y)/C_B} = g(\rho|_{W=y}; (C_A/C_B)\nu_y). \tag{11.34}$$

Therefore, the probability $\rho|_{W=y}$ itself is independent of the choices of the Lebesgue measures on A and B.

Due to the growth condition, the following integral is well defined:

$$(W.\Phi)(\rho) := \int_B \Phi(\rho|_{W=y})\, \mathrm{d}\pi_*^{WV}\rho(y). \tag{11.35}$$

Proposition 11.16. *Equation* (11.35) *turns the presheaf $\mathscr{F}$ into an $\mathscr{A}$-module.*

Proof. Let ρ be an element of $\mathscr{P}_V$ with support $A = A(\rho)$, and W, W' two elements of $\mathbf{S}$ containing V (we denote the corresponding random variables by the same letters). By definition,

$$W.(W'.\Phi)(\rho) = \int_B \left(\int_{B'_y} \Phi((\rho|_{W=y})|_{W'=y'})\, \mathrm{d}\pi_*^{W'V}\rho|_{W=y}(y') \right) \mathrm{d}\pi_*^{WV}\rho(y). \tag{11.36}$$

[2]Remark that every element of $\mathscr{P}_V$ is associated to a support. This must be seen as the appropriate generalizations of the pairs (λ, ρ) considered in the simplicial case: in that context, there was a preferred Lebesgue measure on each support.

The law $\rho|_{W=y}$ is supported on $A_y := \{ x \in A \,|\, \pi^{WV}(x) = y \}$, thus $(\rho|_{W=y})|_{W'=y'}$ is supported on $A_{y,y'} := \{ x \in A_y \,|\, \pi^{W'V}(x) = y' \} = A_y \cap A_{y'}$. Set $U := W \cap W'$ and let $r : E_W \times E_{W'} \to E_U$ be a section of the inclusion $\iota : E_U \to E_W \times E_{W'}$ (such that $r \circ \iota = \mathrm{id}_{E_U}$); as long as $A_{y,y'}$ is nonempty, the value of $r(y, y')$ is uniquely defined: it is the only element in $(\pi^{WU})^{-1}(y) \cap (\pi^{W'U})^{-1}(y')$ such that

$$A_{y,y'} = A_y \cap A_{y'} = (\pi^{WV})^{-1}(y) \cap (\pi^{W'V})^{-1}(y') = (\pi^{UV})^{-1}(r(y, y')). \tag{11.37}$$

We choose a measure λ on A and μ on $B = \pi^{WV}(A)$, inducing through Weil's formula a measure ν_y on the fiber A_y, parallel to $T(A) \cap W/V$; we have $\lambda = \mu.\nu_y$. Similarly, a choice of Lebesgue measure μ'_y on the affine space $B'_y = \pi^{W'V}(A_y)$ induces a measure $\theta_{y,y'}$ on $A_{y,y'}$ such that $\nu_y = \mu_y.\theta_{y,y'}$. Finally, let us choose a Lebesgue measure $\tilde{\mu}$ on $\tilde{B} = \pi^{UV}(A)$, such that $\lambda = \tilde{\mu}.\theta_{y,y'} = \tilde{\mu}.\theta_{r(y,y')}$.

By definition, the law $(\rho|_{W=y})|_{W'=y'}$ has density

$$g((\rho|_{W=y})|_{W'=y'}; \theta_{y,y'}) = \frac{g(\rho|_{W=y}; \nu_y)|_{A_{y,y'}}}{g(\pi_*^{W'V} \rho|_{W=y}; \mu'_y)(y')} \tag{11.38}$$

$$= \frac{g(\rho; \lambda)|_{A_{y,y'}}}{g(\pi_*^{WV} \rho; \mu)(y) g(\pi_*^{W'V} \rho|_{W=y}; \mu'_y)(y')}. \tag{11.39}$$

In virtue of Proposition 9.6,

$$g(\pi_*^{W'V} \rho|_{W=y}; \mu'_y)(y') = \int_{A_y} g(\rho|_{W=y}; \nu_y) \,\mathrm{d}\theta_{y,y'} = \int_{A_y} \frac{g(\rho; \lambda)|_{A_y}}{g(\pi_*^{WV} \rho; \mu)(y)} \,\mathrm{d}\theta_{y,y'} \tag{11.40}$$

which implies that

$$g(\pi_*^{WV} \rho; \mu)(y) g(\pi_*^{W'V} \rho|_{W=y}; \mu'_y)(y') = \int_{A_y \cap A_{y'}} g(\rho; \lambda) \,\mathrm{d}\theta_{y,y'} \tag{11.41}$$

$$= g(\pi_*^{UV} \rho; \tilde{\mu})(r(y, y')) \tag{11.42}$$

again by Proposition 9.6.

Recapitulating, we have that $(\rho|_{W=y})|_{W'=y'}$ has density

$$g(\rho; \lambda)|_{(\pi^{UV})^{-1}(r(y,y'))} / g(\pi_*^{UV} \rho; \tilde{\mu})(r(y, y')),$$

which means that it equals $\rho|_{U=r(y,y')}$.

Coming back to (11.36), we have

$$W.(W'.\Phi)(\rho) = \int_B \left(\int_{B'_y} \Phi(\rho|_{U=r(y,y')}) g(\pi_*^{UV} \rho; \tilde{\mu})(r(y, y')) \,\mathrm{d}\mu'_y(y') \right) \mathrm{d}\mu(y). \tag{11.43}$$

where we have derived the probability laws w.r.t. the reference measures and simplified the densities as in (11.42). By functoriality, $B = \pi^{WU}(\tilde{B})$ and if we set $\tilde{B}_y = (\pi^{WU})^{-1}(y)$, its projection under $\pi^{W'U}$ is exactly B'_y, where we have chosen μ'_y as measure. Using again Weil's formula, there exists a measure $\tilde{\nu}$ on $\tilde{B}_y$ such that

$\tilde{\mu} = \tilde{\nu}_y.\mu$, and also a measure ψ on $\{r(y,y')\}$ (i.e. a constant) such that $\tilde{\nu}_y = \mu'_y.\psi$. Thus, we can rewrite the previous integral as

$$W.(W'.\Phi)(\rho) = \int_{\tilde{B}} \Phi(\rho|_{U=u}) g(\pi_*^{UV}\rho; \tilde{\mu})(u)\psi^{-1}\,\mathrm{d}\tilde{\mu}(u) \tag{11.44}$$

$$= \int_{\tilde{B}} \Phi(\rho|_{U=u})\psi^{-1}\,\mathrm{d}\pi_*^{UV}\rho(u). \tag{11.45}$$

Finally, remark that $\Phi \equiv 1$ is always invariant under the action, so $\psi = 1$. □

The preceding result generalizes Proposition 9.21 concerning the simplex.

11.3.2 Description of cochains and cocycles

Having this $\mathscr{A}$-module $\mathscr{F}$, we introduce the information cohomology $H^\bullet(\mathbf{S},\mathscr{F}) := \mathrm{Ext}^\bullet(\mathbb{R}, F)$. Using the bar resolution $B_\bullet$ of $\mathbb{R}$, we obtain the following description of cochains, cocycles and coboundaries, as explained in Chapter 2.

The zero cochains are elements φ_V of $\mathscr{F}_V$ for V describing $\mathscr{A}$, that are natural, i.e.

$$\forall \rho \in \mathscr{P}_V \quad \varphi_W(\pi_*^{WV}\rho) = \varphi_V(\rho), \tag{11.46}$$

which implies they are constant in virtue of the final element $V = E$. They are also cocycles, i.e. invariant of the action of $\mathscr{A}$.

Since $B_1(V)$ is generated by $\{\,[W] \mid V \subset W\,\}$, the 1-cochains $\varphi \in \mathrm{Hom}(B,\mathscr{F})$ are characterized by elements $\{\varphi_V[W]\}_{V\subseteq W}$ of $\mathscr{F}_V$ such that

$$\forall \rho \in \mathscr{P}_V, \forall W \supseteq V' \supseteq V, \quad \varphi_{V'}[W](\pi_*^{V'V}\rho) = \varphi_V[W](\rho). \tag{11.47}$$

In particular, if $V \subseteq W$, for any $\rho \in \mathscr{P}_V$,

$$\varphi_V[W](\rho) = \varphi_W[W](\pi_*^{WV}\rho). \tag{11.48}$$

then the elements $\varphi_V[V] =: \Phi_V$ determine all the other elements.

The equations for the degree one cocycles are

$$\forall V \in \mathrm{Ob}\,\mathbf{S}, \forall W \supseteq V, W' \supseteq V, \quad \varphi_V[W \cap W'] = W.\varphi_V[W'] + \varphi_V[W]. \tag{11.49}$$

From the equations of naturality (11.47), this is equivalent to the smaller set of equations:

$$\begin{aligned}&\forall W, W' \in \mathscr{A}, W \cap W' =: V \in \mathscr{A}, \forall \rho \in \mathscr{P}_V,\\ &\qquad\qquad \varphi_V[V](\rho) = W.\varphi_V[W'](\rho) + \varphi_W[W](\pi_*^{WV}\rho).\end{aligned} \tag{11.50}$$

11.3.3 Dirac distributions and parallelism

The following results are crucial in all the cohomological computations.

Lemma 11.17. *Let Φ be a 1-cocycle, then for every $V \in \mathrm{Ob}\,\mathbf{S}$ and every point $a \in E_V$, we have*

$$\Phi_V(\delta_a) = 0. \tag{11.51}$$

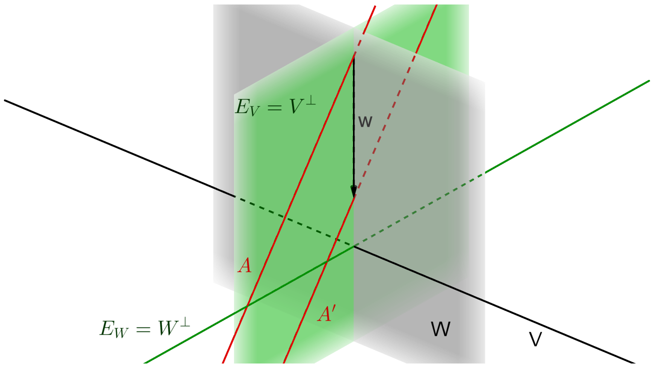

Figure 11.2: Translation of affine supports in Proposition 11.18. Each quotient E/U has been identified with the corresponding orthogonal complement $U^\perp$.

Proof. Write the cocycle equation for $W = W' = V$ and the law $\rho = \delta_a$ on E_V, in order to get

$$\Phi(\delta_a) = \Phi(\delta_a) + \Phi(\delta_a). \tag{11.52}$$

□

Proposition 11.18. *Let Φ be a 1-cocycle. If for every $V \in \mathrm{Ob}\,\mathbf{S}$ and $A \in \mathscr{M}_V$, there exists $W \in \mathrm{Ob}\,\mathbf{S}$ such that $V \subset W$ and $T(A) \cap W/V = 0$, then*

$$\Phi_V(\rho_A) = \Phi_V((T_w)_*\rho_A) = \Phi_W(\pi_*^{WV}\rho_A) \tag{11.53}$$

for any law ρ_A with support A, where T_w denotes the translation by a vector $w \in W/V$, i.e. $T_w(x) = x + w$.

In particular, if for any A' parallel to A (i.e. $T(A) = T(A')$) there exists a space W and a vector $w \in W/V$ such that $A' = A + w$, the restrictions of Φ to the sets of probability that have supports in A and A' are identified through $(T_w)_*$.

Proof. Apply the cocycle equation (11.50) with $W' = V$. We evaluate it on a probability law ρ_A supported on A and on $(T_w)_*\rho_A$ supported on $A + w \subset E_W$. Both supports are projected by $\pi := \pi^{WV}$ to the same affine subspace B of E_W. Let M_y be the unique point of A such that $\pi(M_y) = y$ for $y \in B$. The situation is depicted in Figure 11.2. The 1-cocycle condition becomes

$$\Phi_V(\rho_A) = \mathbb{E}_{\pi_*\rho}\left(y \mapsto \Phi_V(\delta_{M_y})\right) + \Phi_W(\pi_*\rho_A) = \Phi_W(\pi_*\rho_A). \tag{11.54}$$

and similarly $\Phi_V((T_w)_*\rho_A) = \Phi_W(\pi_*(T_w)_*\rho_A) = \Phi_W(\pi_*\rho_A)$, because both laws just differ by a translation of the mean that vanishes on E_W.

□

If the category C is orthogonally closed for a metric Q and if the cosheaf of supports is minimal, the hypothesis in the previous proposition is always satisfied (there is a space W' such that $T(A)$ is W'/V and its orthogonal complement gives the required W). This happened in the simplicial case.

If $\mathbf{S}$ is the full grassmannian structure, Proposition 11.18 immediately implies that cocycles are independent of the mean, since there is no preferred way to translate A

into $T(A)$. On the contrary, in the simplicial case there is a canonical way to identify the mean of ρ and its translated version.

If the condition in Proposition 11.18 is not verified, we could directly impose this invariance by translation of the cochains to develop a cohomology theory that is not less interesting.

11.3.4 Axial cocycles over $\mathbf{S}_2$

In this section, we consider the information structure $(\mathbf{S}_2, \mathscr{E})$ introduced in Example 11.2 and depicted in Figure 11.1. This structure has been embedded in $\mathbb{R}^2$ with the usual Euclidean metric. Let (e_1, e_3) and (e_2, e_4) denote orthonormal bases with coordinate axes (E_1, E_3) and (E_2, E_4), respectively. The change of coordinates between (x_1, x_3) and (x_2, x_4) is given by

$$x_2 = \frac{x_1 + x_3}{\sqrt{2}} \quad x_4 = \frac{x_3 - x_1}{\sqrt{2}}, \quad x_1 = \frac{x_2 - x_4}{\sqrt{2}}, \quad x_3 = \frac{x_2 + x_4}{\sqrt{2}}. \tag{11.55}$$

For the dual coordinates $\{\xi_i\}_{i=1,\dots,4}$, the formulas are the same, by unitarity. Then the covariance matrices change accordingly as

$$\sigma_{22} = \frac{\sigma_{11} + \sigma_{33}}{2} + \sigma_{13}, \quad \sigma_{44} = \frac{\sigma_{11} + \sigma_{33}}{2} - \sigma_{13}, \quad \sigma_{24} = \frac{\sigma_{33} - \sigma_{11}}{2}. \tag{11.56}$$

Let $\Phi = \{\phi_V \in \mathscr{F}_V\}_{V \in \mathrm{Ob}\,\mathbf{S}_2}$ be the data characterizing a 1-cochain; each ϕ_V is a function from $\mathscr{P}_V = \mathscr{P}_V(\mathscr{N}_V)$ to $\mathbb{R}$, that can be restricted to the nondegenerate laws supported on E_V to get ϕ_V^{ng} (which is analogous to ϕ^S is the previous chapter). We further simplify the notation writing $\phi := \phi_{\langle 0 \rangle}$ and $\phi_i := \phi_{V_i}$; the function ϕ_E vanishes since the associated space E_E is a singleton (cf. Lemma 11.17).

We suppose that Φ *defines a* 1*-cocycle that is axial (see Section 10.3.3) with respect to* (E_1, E_3), which means that $\phi_{\langle 0 \rangle} = \varphi_1 + \varphi_3$, with φ_1, φ_3 generalized moments defined on $\mathbb{R} \times]0, \infty[$. We shall see that φ_1 and φ_3 must be the constants and coincide. To do so, write the cocycle relation $\phi = V_4.\phi_2 + \phi_4$, in the case of a Gaussian distribution with mean $M = (m_1, m_3)$ along E_1, E_3 and a covariance matrix Σ that has principal axis E_2, E_4 with respective coefficients τ_2, τ_4:

$$\phi(\rho(M, \Sigma)) = \int \phi_2(\overline{M}_2(x_4), \overline{\Sigma}_2) G_4(x_4)\, \mathrm{d}x_4 + \phi_4(G_4(m_4, \tau_4)). \tag{11.57}$$

Due to the choice of Σ, we have $\overline{M}_2 = m_2$ and $\overline{\Sigma}_2 = \tau_2$, therefore

$$\Phi(M, \Sigma) = \phi_2(m_2, \tau_2) + \phi_4(m_4, \tau_4). \tag{11.58}$$

Substituting $m_4 = (m_3 - m_1)/\sqrt{2}$, $m_2 = (m_1 + m_3)/\sqrt{2}$, $\sigma_{11} = (\tau_2 + \tau_4)/2$ and $\sigma_{33} = (\tau_2 + \tau_4)/2$, we obtain

$$\begin{aligned} &\varphi_1(m_1, (\tau_2 + \tau_4)/2) + \varphi_3(m_3, (\tau_2 + \tau_4)/2) \\ &\qquad = \phi_2((m_1 + m_3)/\sqrt{2}, \tau_2) + \phi_4((m_3 - m_1)/\sqrt{2}, \tau_4). \end{aligned} \tag{11.59}$$

Lemma 11.19. *Let f, g, h be three differentiable functions of a real variable $s \in]0, \infty[$ satisfying $\forall s, t > 0, \quad f(s) + g(t) = h(s + t)$, then f, g and h are affine functions of the same slope.*

Proof. When differentiating with respect to s or t, we find

$$f'(s) = h'(s+t) = g'(t); \tag{11.60}$$

therefore, there exists a constant D such that $f' = g' = h' = D$, and three constants A, B, C such that $f(s) = Ds + A$, $g(s) = Ds + B$, $h(s) = Ds + C$. □

Lemma 11.20. *The only possible nonzero axial cocycles of* $(\mathbf{S}_2, \mathscr{E})$ *are linear combinations of constants, of the mean coordinates and of the moment of order two; more precisely, there exist constants* A_1*,* A_3*,* B_1*,* B_3*, and* D *such that* $\varphi_i(m, \sigma) = D(m_i^2 + \sigma_{ii}) + A_i m_i + B_i$*, for* $i = 1$ *and* 3*. The conclusion of the lemma holds also true for* ϕ_2^{ng} *and* ϕ_4^{ng}*, with the same constant* D*.*

Proof. Let us fix arbitrarily m_1 and m_3, and consider (11.59) as a functional equation of four functions in τ_2, τ_4. From the preceding lemma, we deduce the existence of four functions $C(m_1, m_3)$, $B(m_1, m_3)$, $B_2(m_1, m_3)$, and $B_4(m_1, m_3)$ such that

$$\varphi_1(m_1, \sigma) + \varphi_3(m_3, \sigma) = 2D(m_1, m_3)\sigma + B(m_1, m_3), \tag{11.61}$$

$$\phi_2^{ng}((m_1 + m_3)/\sqrt{2}, \tau) = D(m_1, m_3)\tau + B_2(m_1, m_3), \tag{11.62}$$

$$\phi_4^{ng}((m_3 - m_1)/\sqrt{2}, \tau) = D(m_1, m_3)\tau + B_4(m_1, m_3), \tag{11.63}$$

which in turn implies that

$$\frac{\partial \phi_2^{ng}((m_1 + m_3)/\sqrt{2}, \tau)}{\partial \tau} = D(m_1, m_3) = \frac{\partial \phi_4^{ng}((m_1 - m_3)/\sqrt{2}, \tau)}{\partial \tau}. \tag{11.64}$$

Therefore D is at the same time a function of $m_1 + m_3$ and a function of $m_1 - m_3$, thus it is a constant D. Consequently, (11.61) becomes

$$\frac{\partial \varphi_1(m_1, \sigma)}{\partial \sigma} + \frac{\partial \varphi_3(m_3, \sigma)}{\partial \sigma} = 2D. \tag{11.65}$$

Differentiating with respect to m_1 (resp. m_3) we obtain the existence of two constants D_1, D_3 such that

$$\frac{\partial \varphi_1(m_1, \sigma)}{\partial \sigma} = D_1, \quad \frac{\partial \varphi_3(m_3, \sigma)}{\partial \sigma} = D_3, \tag{11.66}$$

hence $D_1 + D_3 = 2D$ by 11.65. The generalized moment functions φ_1, φ_3 are solutions of the heat equation (Section 10.3.4), so we also know that

$$\frac{1}{2}\frac{\partial^2 \varphi_1(m, \sigma)}{\partial m^2} = D_1, \quad \frac{1}{2}\frac{\partial \varphi_3(m, \sigma)}{\partial m^2} = D_3. \tag{11.67}$$

Therefore φ_1 and φ_3 are both moments of order two. This gives

$$\varphi_1(m, \sigma) = D_1(m^2 + \sigma) + A_1 m + B_1, \tag{11.68}$$

$$\varphi_3(m, \sigma) = D_3(m^2 + \sigma) + A_3 m + B_3. \tag{11.69}$$

Using equation (11.59), we deduce that, for every m_2, m_4 in $\mathbb{R}$ and τ_2, τ_4 in $\mathbb{R}_+^*$,

$$\begin{aligned}
\phi_2^{ng}(m_2,\tau_2) + \phi_4^{ng}(m_4,\tau_4) &= D_1\left[\left(\frac{m_2+m_4}{\sqrt{2}}\right)^2 + \left(\frac{\tau_1+\tau_2}{2}\right)\right] + A_1\left(\frac{m_2+m_4}{\sqrt{2}}\right) + B_1 \\
&\quad + D_3\left[\left(\frac{m_2-m_4}{\sqrt{2}}\right)^2 + \left(\frac{\tau_1+\tau_2}{2}\right)\right] + A_3\left(\frac{m_2-m_4}{\sqrt{2}}\right) + B_3 \\
&= \frac{D_1+D_3}{2}(m_2^2+m_4^2+\tau_2+\tau_4) + (D_1-D_3)m_2m_4 \\
&\quad + \left(\frac{A_1+A_3}{\sqrt{2}}\right)m_2 + \left(\frac{A_1-A_3}{\sqrt{2}}\right)m_4 + B_1 + B_3. \qquad (11.70)
\end{aligned}$$

in such a way that $\partial_{m_2}\phi_2^{ng}(m_2,\tau_2)$ contains a term $(D_1 - D_3)m_4$ that must vanish, implying that $D_1 = D_3 = D$. Equation (11.70) also entails the existence of constants A_2, A_4, B_2, B_4 such that

$$\phi_2^{ng}(m_2,\tau) = D(m_2^2+\tau) + A_2m_2 + B_2, \quad \phi_4(m_4,\tau) = D(m_4^2+\tau) + A_4m_4 + B_4. \quad (11.71)$$

In particular all the moments of degree two have the same coefficient. □

Consequently, for any nondegenerate gaussian law ρ on E, with mean M of coordinates (m_1, m_3) and covariance Σ of coefficients $\{\sigma_{ij}\}_{i,j\in\{1,3\}}$ in the basis (e_1, e_3), or (m_2, m_4) and $\{\sigma_{ij}\}_{i,j\in\{2,4\}}$ in the basis (e_2, e_4),

$$\begin{aligned}
\phi^{ng}(\rho) &= D(m_1^2+m_3^2+\sigma_{11}+\sigma_{33}) + A_1m_1 + A_3m_3 + B_1 + B_3 \\
&= D(m_2^2+m_4^2+\sigma_{22}+\sigma_{44}) + A_2m_2 + A_4m_4 + B_2 + B_4. \qquad (11.72)
\end{aligned}$$

Then

$$\alpha(M) = A_1m_1 + A_3m_3 = A_2m_2 + A_4m_4, \quad (11.73)$$

defines a linear form, and we have

$$\phi^{ng}(\rho) = D\Psi_2(M,\Sigma) + \alpha(M) + \beta, \quad (11.74)$$

with

$$\Psi_2(M,\Sigma) = \|M\|^2 + \mathrm{Tr}(\Sigma) \quad (11.75)$$

and

$$\beta = B_1 + B_3 = B_2 + B_4. \quad (11.76)$$

Lemma 11.17 says that each function ϕ or ϕ_i vanishes on laws supported on points. Hence it only remains to characterize ϕ^{ng} for degenerate laws supported on lines contained in E, which is related to their projection on lines that intersect transversely their support, as stated in Proposition 11.18. The compatibility between these values is only possible if D and α vanish.

Proposition 11.21. *The only nonzero axial cocycles over $(\mathbf{S}_2, \mathscr{E})$ correspond to collections of functions $(\phi, \phi_1, \phi_2, \phi_3, \phi_4)$ from the possibly degenerate gaussian laws on the corresponding spaces E, E_1, E_2, E_3, E_4 into $\mathbb{R}$, given by*

$$\phi_i(\rho) = B\dim(A_\rho), \quad (11.77)$$

where B is any real constant, and A_ρ is the support of ρ.

Proof. A degenerate gaussian supported by the axis E_4 is the measure

$$\mathrm{d}\rho(x_2, x_4) = \delta(x_2 = 0) \otimes \frac{1}{\sqrt{2\pi\sigma_{44}}} e^{-\frac{(x_4-m_4)^2}{2\sigma_{44}}}\, \mathrm{d}x_4. \tag{11.78}$$

with $\sigma_{44} > 0$. Since both V_3 and V_4 are transversal to $E_4 = A(\rho)$ (cf. Figure 11.1), Proposition 11.18 implies that

$$\phi(\rho) = \phi_3(\pi_*^{V_3\langle 0\rangle}\rho) = \phi_4(\pi_*^{V_4\langle 0\rangle}\rho). \tag{11.79}$$

First, we compute the term $\phi_3(\pi_*^{V_3\langle 0\rangle}\rho)$. The map $\pi^{V_3\langle 0\rangle}$ is the orthogonal projection[3] $E \longrightarrow E_3 \simeq E/V_3$. We have

$$\mathrm{d}\pi_*^{V_3\langle 0\rangle}\rho(x_3) = \frac{1}{\sqrt{2\pi\sigma_{33}}} e^{-\frac{(x_3-m_3)^2}{2\sigma_{33}}}\, dx_3. \tag{11.80}$$

where $m_3 = m_4/\sqrt{2}$ under the change of coordinates (11.55), and σ_{33} is given by (11.56). We have

$$\sigma_{33} = \sigma_{11}, \quad \sigma_{44} = 2\sigma_{11}, \quad \sigma_{13} = -\sigma_{11}. \tag{11.81}$$

Since we are supposing that the cocycle is axial, $\phi_3^{ng} = \varphi_3$ as determined in the previous proposition:

$$\phi_3(\pi_*^{V_3\langle 0\rangle}\rho) = D(m_3^2 + \sigma_{33}) + A_3 m_3 + B_3. \tag{11.82}$$

In turn, the law $\pi_*^{V_4\langle 0\rangle}\rho$ is the gaussian supported on E_4 with mean m_4 and variance $\sigma_{44} > 0$. Proposition 11.20 gives the formula for ϕ_4^{ng}, so

$$\phi(\rho) = D(m_4^2 + \sigma_{44}) + A_4 m_4 + B_4 = 2D(m_3^2 + \sigma_{33}) + A_4 m_4 + B_4, \tag{11.83}$$

where we use $m_4 = m_3\sqrt{2}$.

The second equality in (11.79) reads

$$2D(m_3^2 + \sigma_{33}) + A_4 m_3\sqrt{2} + B_4 = D(m_3^2 + \sigma_{33}) + A_3 m_3 + B_3, \tag{11.84}$$

that is

$$D(m_3^2 + \sigma_{33}) = (A_3 - \sqrt{2}A_4)m_3 + B_3 - B_4, \tag{11.85}$$

which is possible if and only if $D = 0$, $\sqrt{2}A_4 = A_3$ and $B_4 = B_3$.

However, we have by coordinate changes, $\sqrt{2}A_4 = A_3 - A_1$, then $A_1 = 0$. In the same manner, turning the axis we have $A_3 = 0$, and $B_2 = B_1$. Consequently $\alpha = 0$. And by symmetry with respect to E_4, exchanging E_1 and E_3, we find $B_1 = B_3 = B_2 = B_4 = B$.

This shows that for every gaussian with linear support parallel to one of the lines E_j, the value of the cocycle is B, from (11.72) it is $2B$ for a nondegenerate law with support E, and 0 for the laws supported by a point. □

In the proposition above the fact that the angles between E_1 and E_2 (resp. between E_3 and E_4) is $\pi/4$ has no importance, it is only for simplifying the formulas, any other angle strictly between 0 and $\pi/2$ works as well.

[3]The isomorphism between E_3, a subspace of E, and the quotient E/V_3 is not canonical; it comes, however, as part of the definition of $\mathscr{E}$ (embedded in $\mathbb{R}^2$) and the advantage is that we can use the formulae for gaussians on affine subspaces as presented in Appendix D .

11.3.5 Entropy

Definition 11.22. If $\rho \in \mathscr{P}_V$ has its support equal to $A \subseteq E_V$ of dimension d, and if λ is chosen in $\mathscr{L}_A = \mathscr{L}_{T(A)}$, the entropy is defined by

$$S_V(g,\lambda) = \mathbb{E}_\rho\left(-\ln \frac{\mathrm{d}\rho}{\mathrm{d}\lambda}\right). \tag{11.86}$$

Introducing the mean and the covariance, this gives

$$S_V(g,\lambda) = \frac{(2\pi)^{-d/2}}{\sqrt{\det_\lambda \Sigma}} \int_A \frac{1}{2}\Sigma^{-1}(x-M,x-M)e^{-\frac{1}{2}\Sigma^{-1}(x-M,x-M)}\,\mathrm{d}\lambda(x) + \frac{1}{2}\ln\det\nolimits_\lambda \Sigma + \frac{d}{2}\ln 2\pi. \tag{11.87}$$

Changing the variable by $x - M = \sqrt{\Sigma} y$ shows that the first term to the right is equal to $d/2$, thus

$$S_V(g,\lambda) = \frac{1}{2}\ln\det\nolimits_\lambda \Sigma + \frac{d}{2}\ln(2\pi e). \tag{11.88}$$

Definition 11.23. For $\rho \in \mathscr{P}_V$, let us denote by $d(\rho)$ the dimension $\dim(A(\rho))$ of its support.

Proposition 11.24. *d is a 1-cocycle for the cohomology with coefficients in $\mathscr{F}$.*

Proof. This is a consequence of the rank theorem: for $V, W \in \operatorname{Ob}\mathbf{S}$, $W \supseteq V$, we consider the restriction of the projection $\pi = \pi^{WV}$ to $A(\rho)$, it induces a surjective linear map from $T(A)$ to the tangent of the support B of $\pi_*\rho$, and the kernel of this map is precisely the tangent space of the support of any one of the conditioned probabilities $\rho|_{\pi(x)=y}$. The theorem says that

$$d(\rho) = \mathbb{E}_{\pi_*\rho}(d(\rho|_{\pi(x)=y})) + d(\pi_*\rho). \tag{11.89}$$

□

By subtracting the multiple $\frac{d}{2}\ln(2\pi e)$ from S we get the normalized entropy $\overline{S}$, which is a function of $\det_\lambda \Sigma_\rho$ only. A change from λ to $\lambda' = C\lambda$ induces the addition of $\log C$ to $\bar{S}_V(\rho,\lambda)$, because $\det_{\lambda'}\Sigma = C^2 \det_\lambda \Sigma$, as we saw in Section 11.2.1. So $\bar{S}$ is not a cochain for the coefficients $\mathscr{F}$ introduced in Section 11.3.1, which motivates the following definition.

Definition 11.25. Let $(\mathbf{S},\mathscr{E})$ be a grassmannian information structure on the vector space E. Recall that an euclidean metric Q induces an identification of each quotient E/V with $V^\perp$, for each $V \in \operatorname{Ob}\mathbf{S}$. Every affine subspace $A \subset E/V$ embeds into E and inherits from Q a Lebesgue measure $\lambda_Q(A)$, see Section 11.1.3. The twisted functional space $\mathscr{X}$ is the vector space of real-valued functions ϕ of a probability measure ρ on E_V and metric trivialization λ_Q of $\mathscr{L}_V$ that verify

$$\forall \rho \in \mathscr{P}_V, \forall Q, Q' \text{ euclidean metrics on } E,\quad \phi(\rho,\lambda_{Q'}) = \phi(\rho,\lambda_Q) + \ln D(T(A_\rho);Q,Q'). \tag{11.90}$$

where $D(B;Q,Q')$ is an $\mathbb{R}_+^*$-valued function, called *generalized discriminant*, that is required to satisfy $D(B;Q,Q')D(B;Q',Q'') = D(B;Q,Q'')$ for any triplet of euclidean metrics Q,Q',Q'' and any vector space $B \subset E_V$.

For every morphism $\iota : V \to W$ in $\mathbf{S}$, there is a an induced map $\iota^* : \mathscr{X}_W \to \mathscr{X}_V$ that maps $\phi_W \in \mathscr{X}_W$ to $\phi_V = \iota^*(\phi_W)$ given by

$$\phi_V(\rho, \lambda_Q) = \phi_W(\pi_* \rho, \lambda_Q). \tag{11.91}$$

We can see that the sheaf $\mathscr{F}$ of moderate probabilistic functionals is included in $\mathscr{X}$, corresponding to $D \equiv 1$. The typical generalized discriminant $D(A;Q,Q')$ that we have in mind is a product of the discriminants $\Delta(B;Q,Q')$ introduced in Section 11.1.3 or their inverses, where B can range over certain projections of A that are independent of ρ, as is the case in formula (11.91).

Remark 11.26 (Sections of principal bundles)**.** The previous definition is also motivated by the following considerations.

For every $V \in \operatorname{Ob}\mathbf{S}$ we have defined the bundle $\mathscr{L}_V$ of Lebesgue measures on the supports in $\mathscr{M}_V$, and verified that it is a principal bundle for the multiplicative group $\mathbb{R}_+^*$ *over each stratum of the total Grassmannian*. For each $A \in \mathscr{M}_V$, the fiber is the positive cone $\Lambda_+(T(A)) \subset \Lambda^{max}(T(A))$.

Given a G-principal bundle P, and a left G action $\tau : G \times F \to F$ on a manifold F, there is G-bundle with fiber F defined as $P \times_\tau F := P \times F/\sim$, where the relation $\sim$ is given by

$$\forall (p,f) \in P \times F, \forall g \in G, \quad (p,f) \sim (p.g, \tau(g^{-1})f). \tag{11.92}$$

Let τ_a be the action of $\mathbb{R}_+^*$ on $\mathbb{R}$ given by $\mathbb{R}_+^* \times \mathbb{R} \to \mathbb{R}$, $(c,r) \mapsto a \log c + r$. The general theory of principal bundles (cf. [80, p. 39]) tells that sections of the bundle $\mathscr{L}_V \times_{\tau_a} \mathbb{R}$ are identified canonically with maps ϕ from the total space $\mathscr{L}_V$ to $\mathbb{R}$ such that

$$\forall (B,\lambda) \in \mathscr{L}_V, \forall C > 0, \quad \phi(B, C\lambda) = \phi(B,\lambda) - a \ln C. \tag{11.93}$$

A section of $\mathscr{L}_V \times_{\tau_a} \mathbb{R}$ defines, by precomposition with the function $A : \mathscr{P}_V \to \mathscr{M}_V$ that associates to each probability ρ its support $A(\rho)$, a function ϕ on $\mathscr{P}_V$ that is equivariant in the sense of (11.93). Under this point of view, $\bar{S}$ is associated to τ_{-1}.

Lemma 11.27. *The formula* (11.35) *defines a structure of $\mathscr{A}$-module on the sheaf of vector spaces $\mathscr{X}$.*

Proof. Let $\iota : V \to W$ be an arrow in $\mathbf{S}$, and ϕ an element of $\mathscr{X}_V$. We prove that $W.\phi$ also belongs to $\mathscr{X}_V$. In fact,

$$\begin{aligned}
(W.\phi)(\rho, \lambda_{Q'}) &= \int_{A(\iota_*(\rho)} \phi(\rho|_{W=w}, \lambda_{Q'}) \, \mathrm{d}\iota_*\rho(w) \\
&= \int_{A(\iota_*(\rho)} (\phi(\rho|_{W=w}, \lambda_Q) - \ln D(T(A(\rho|_{W=w})); Q, Q')) \, \mathrm{d}\iota_*\rho(w) \\
&= (W.\phi)(\rho, \lambda_Q) - \ln D(T(A) \cap W/V; Q, Q').
\end{aligned}$$

The proof of $W'.(W.\phi) = (W'W).\phi$ given for Proposition 11.16 remains valid. □

Proposition 11.28. *The normalized entropy $\bar{S}$ is* 1*-cocycle when the coefficients are $\mathscr{X}$.*

Proof. This can be proved in a more general setting, using properties of disintegrations, see Proposition 11.35. However, a direct proof shows an interesting connection with pure algebra and the beginnings of K-theory.

We have already seen that $\det_\lambda \Sigma$ gives a section of $\mathscr{X}$: when a metric Q is changed into Q', the measure $\lambda_Q(A)$ is multiplied by the discriminant $\Delta(A; Q, Q') = \Delta(T(A); Q, Q')$; the covariance Σ remains the same, but its determinant changes by a factor of $\Delta(A; Q, Q')^2$.

Only the cocycle equation has to be verified, for a given choice of Q that induces an embedding J_Q. For that we use the formulas in Appendix D, computing the covariance of the direct image and the conditioned probability of a gaussian law. We work in the space $T(A)$ and the restriction of the projection $\pi = \pi^{WV}$ (that under J_Q becomes the projection on a subspace of $T(A)$): if

$$\Sigma(\rho) = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \tag{11.94}$$

then

$$\Sigma(\pi_* \rho) = \Sigma_{11}, \tag{11.95}$$

and for every $y \in T(B)$,

$$\Sigma(\rho|_{\pi(x)=y}) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}, \tag{11.96}$$

which is the Schur's complement of Σ_{11} in $\Sigma(\rho)$. In virtue of Schur's determinantal identity (Proposition C.1),

$$\begin{aligned} \det(\Sigma(\rho)) &= \det(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) \det(\Sigma_{11}) && (11.97)\\ &= \det(\Sigma(\rho|_{\pi(x)=y}))\Sigma(\pi_*\rho), && (11.98) \end{aligned}$$

thus taking the logarithm,

$$\ln \det(\Sigma(\rho)) = \ln \det(\Sigma(\rho|_{\pi(x)=y})) + \ln \Sigma(\pi_*\rho), \tag{11.99}$$

then integrating over y with $\pi_*\rho$, and dividing by 2, we get

$$\overline{S}(\rho) = \mathbb{E}_{\pi_*\rho}\left(\overline{S}(\rho|_{\pi(x)=y})\right) + \overline{S}(\pi_*\rho). \tag{11.100}$$

which is the wanted identity. □

Definition 11.29. We say that a ***Grassmannian triple*** $(\mathbf{S}, \mathscr{E}, \mathscr{N})$, made of a grassmannian structure $(\mathbf{S}, \mathscr{E})$ and a functor $\mathscr{N}$ of supports, is ***sufficiently rich*** if it satisfies the following conditions:

1. for every $V \in \operatorname{Ob} \mathbf{S}$ and every $A \in \mathscr{N}_V$, there exists a basis $B_{V,A} = (e_1, ..., e_n)$ of E_V such that
   - for some $S' \subset [n]$ it holds that $T(A) = E_{S'}$, and
   - for every $S \subset [n]$, there exists $W_S \in \operatorname{Ob} \mathbf{S}$ such that $V \subset W_S$ and $E_S = W_S/V$,

   where $E_S := \langle e_i \rangle_{i \in S}$.
2. The sheaf $\mathscr{N}$ is adapted, and each $\mathscr{N}_V$ contains E_V.
3. For every $V \in \operatorname{Ob} \mathbf{S}$ of codimension 2 (in E) is contained in (at least) four spaces $W_1, ..., W_4 \in \operatorname{Ob} \mathbf{S}$ of codimension 1.

An immediate consequence of the axiom 2 is that $\mathscr{N}_V$ also contains all the affine spaces parallel to one of the W_S/V given by axiom 1, because they are fibers of the projection $E_V \to E_W$.

The axiom 3 mean that we can recover a situation analogous to $(\mathbf{S}_2, \mathscr{E})$: E_V is a plane, each W_i/V gives a different line; we identify $E_{W_1} = E/W_1$ with W_3/V, E_{W_2} with W_4/V, etc. and $\pi^{W_i V}$ with the projection on E_{W_i} parallel to W_i.

Taking into account all the results from this chapter and the preceding one, we get the following result.

Theorem 11.30. *Let $(\mathbf{S}, \mathscr{E}, \mathscr{N})$ be a sufficiently rich grassmannian triple. The degree one information cohomology with coefficients in the twisted module $\mathscr{X}$ is made of functions*

$$\phi_V(\rho) = -aS(\rho) + b\dim(A(\rho)), \tag{11.101}$$

where a and b are arbitrary real constants.

Proof. Let ρ be a gaussian law with support $A \subset E_V$, and $B_{V,A}$ be the basis given by axiom 1 in Def. 11.29. We suppose that $T(A) = E_S$; we clearly have $E = E_S \oplus E_{S^c}$, so there is a unique $w \in E_{S^c}$ such that $A + w = E_S$. This translation by w gives a well-defined bijection between the parameters of gaussian laws on A and E_S. Using Proposition 11.18, we have $\phi_V(\rho) = \phi_{W_{S^c}}(\pi^S_* \rho)$, where π^S_* is projection on the subspace $E_S \simeq E/W_{S^c} = E_{W_{S^c}}$ parallel to E_{S^c}; the law $\pi^S_* \rho$ is nondegenerate. The determination of $\{\phi^S := \phi_{W_{S^c}}\}_{S \subset [n]}$ restricted to nondegenerate laws was the subject of Chapter 10: it established that, for all $T \subset [n]$,

$$\phi^T(\rho) = \phi^T(m, \Sigma) = a \ln\det(\Sigma) + \sum_{i \in T} \varphi(m_i, \sigma_{ii} | g^i), \tag{11.102}$$

where ρ is a nondegenerate law on E_T with parameters (m, Σ) expressed in the basis $\{e_i\}_{i \in S} \subset B_{V,A}$, the factor a is an arbitrary real, and the $\varphi(g^i)$ are generalized moments. In particular, this gives the value of $\phi_{W_S}(\pi^S_* \rho)$. In the previous expression, the determinant is taken with respect to the standard euclidean metric associated to the basis $B_{V,A}$ (the quadratic form defined by the identity matrix in that basis).

We claim that each $\varphi(g^i)$ is a constant. In fact, suppose $S = \{i, j\} \subset [n]$ and ψ^S is an axial cocycle (the axial part of ϕ^S) i.e. E_S is a plane and $\psi^S = \varphi_i + \varphi_j$. Since there is an alternative basis to decompose ψ_{W_S} given by the axiom 3 in Def. 11.29, we can reproduce the arguments in Section 11.3.4 to establish that $\varphi_i = \varphi_j = B$ for certain $B \in \mathbb{R}$. □

11.3.6 Moments

In this section, we show that the moments appear in the theory as natural transformations invariant under the action of $\mathscr{A}$.

Covariant module of expectation

Let $(\mathbf{S}, \mathscr{E})$ be a grassmannian category on a finite dimensional vector space E over $\mathbb{R}$; $\mathscr{N}$ a compatible family of supports, and $\mathscr{P}$ an associated probability functor.

Given a finite dimensional real vector space F, let us denote by $\mathscr{H}^{(m)}(F)$ the space of locally integrable functions of moderate growth on F. It is usual to put on $\mathscr{H}^{(m)}(F)$

the structure of locally convex space, that is defined by the following semi-norms, associated to any norm $x \mapsto \|x\|$ on F,

$$\forall x \in F, k \in \mathbb{N}, \quad N_{x,k}(f) = \sup_{y \in F} \|y\|^k |f(y-x)|. \tag{11.103}$$

With these semi-norms, $\mathscr{H}^{(m)}(F)$ is complete, then a Fréchet space, and its dual, the space of continuous linear forms on $\mathscr{H}^{(m)}(F)$, is also a Fréchet space. In what follows, we denote by $\mathscr{H}(F)$ this dual space.

For each $V \in \operatorname{Ob}\mathbf{S}$, let us consider the vector space $\mathscr{H}_V^{(m)}$ of locally integrable functions of moderate growth on the vector spaces E_V, and also its dual space $\mathscr{H}_V = \mathscr{H}(E_V)$. If $V \to W$ in $\mathbf{S}$, writing π the projection $\pi^{WV} : E_V \to E_W$, we get a linear continuous map π^* from $\mathscr{H}_W^{(m)}$ to $\mathscr{H}_V^{(m)}$, thus a linear continuous map π_* from $\mathscr{H}_V$ to $\mathscr{H}_W$. This define a covariant functor $\mathscr{H}$ from $\mathbf{S}$ to the category of Fréchet vector spaces.

For $V \in \operatorname{Ob}\mathbf{S}$, denote by $\mathscr{F}_V^{(m)}$ the set of measurable maps from $\mathscr{P}_V$ to $\mathscr{H}_V$. For $V, W \in \mathbf{S}$, $W \supseteq V$, $\Psi_V \in \mathscr{F}_V^{(m)}$ and $\rho \in \mathscr{P}_V$, the formula

$$(W.\Psi)(\rho)(f) = \int_{E_W} \Psi(\rho|_{W=y}) \,\mathrm{d}\pi_*\rho(y), \tag{11.104}$$

defines another element of $\mathscr{F}_V^{(m)}$. This is an integral of continuous linear forms on a Fréchet space, that should be interpreted in the sense of Gelfand-Pettis. From the general properties of disintegrations, we expect again the equality $W'.(W.\Psi) = (W'W).\Psi$ to hold, for any $W, W' \in \mathscr{A}_V$, in such a way that $\mathscr{F}_V^{(m)}$ has the structure of a $\mathscr{A}_V$-module, that we name the ***expectation module***.

Generalized moderate moment cocycle

For $f \in \mathscr{H}_V^{(m)}$, and $\rho \in \mathscr{P}_V$, let us define

$$\Phi_V^{(m)}(\rho)(f) = \int_{E_V} f(x) \,\mathrm{d}\rho(x). \tag{11.105}$$

These integrals are always convergent, because the laws ρ are gaussians.

Proposition 11.31. *The collection of maps $\rho \mapsto \Phi_V^{(m)}(\rho)$, for $V \in \operatorname{Ob}\mathbf{S}$, defines a natural transformation from $\mathscr{P}$ to $\mathscr{H}$.*

Proof. Let $V \subseteq W$, $\rho \in \mathscr{P}_V$, $f \in \mathscr{H}_W^{(m)}$

$$\pi_*(\Phi_V^{(m)}(\rho))(f) \overset{\text{(def.)}}{=} \Phi_V^{(m)}(\rho)(\pi^* f) \overset{\text{(def.)}}{=} \int_{E_V} f(\pi(x)) \,\mathrm{d}\rho(x) = \int_{E_W} f(y) \,\mathrm{d}\pi_*\rho(y). \tag{11.106}$$

Therefore

$$\pi_*\Phi_V^{(m)}(\rho) = \Phi_W^{(m)}(\pi_*\rho). \tag{11.107}$$

□

Proposition 11.32. *The component $\Phi_V^{(m)} \in \mathscr{F}_V^{(m)}$ is invariant under the action of $\mathscr{A}_V$.*

Proof. Let $f \in \mathscr{H}_V^{(m)}$: it can also be seen as an element of $\mathscr{H}_V^*$. By definition of the action,

$$(W.\Phi_V^{(m)})(\rho)(f) = \int_{E_W} \Phi_V^{(m)}(\rho|_{W=w})(f)\, \mathrm{d}\pi_*^{WV}\rho(w). \tag{11.108}$$

According to the definition of $\Phi_V^{(m)}$, we deduce that

$$(W.\Phi_V^{(m)})(\rho)(f) = \int_{E_W} \left(\int_{E_V} f(x)\, \mathrm{d}\rho|_{W=w} \right) \mathrm{d}\pi_*^{WV}\rho(w). \tag{11.109}$$

But the definition of disintegration entails that the integral in (11.109) equals

$$\int_{E_V} f(x)\, \mathrm{d}\rho(x) = \Phi_V^{(m)}(\rho)(f).$$

□

The underlying general property of conditional expectation is that for any vectorial random variable and any σ-algebra $\mathscr{T}$,

$$\mathbb{E}(\mathbb{E}(X|\mathscr{T})) = \mathbb{E}(X). \tag{11.110}$$

Remark 11.33. The natural transformation $\Phi^{(m)}$ is an element of $\tilde{F} = \mathrm{Hom}_{[\mathbf{S},\mathbf{Meas}]}(\mathscr{P}, \mathscr{H})$; this set has the structure of an abelian group with componentwise addition: if $\phi, \psi \in \tilde{F}$ then $\rho \in \mathscr{P}_V \mapsto \phi_V(\rho) + \psi_V(\rho) \in \mathscr{H}_V$ is an element of $\tilde{F}$ too. However, there is no natural ring of variables that acts on this abelian group to turn it into a module.

Ordinary moments

For each integer $n \in \mathbb{N}$, consider the symmetric powers $S^n(E_V)$, for $V \in \mathrm{Ob}\,\mathbf{S}$; they form a covariant functor $\mathscr{E}_n$ from $\mathbf{S}$ to the category of vector spaces. For each $V \in \mathscr{A}$, the space $S^n(E_V^*)$ is a subspace of $H_V^{(m)}$, and its dual $S^n(E_V)$ is a quotient of $\mathscr{H}_V$.

Every $\rho \in \mathscr{P}_V$ defines by integration a linear function on the space of homogeneous polynomials $S^n(E_V^*)$, then an element of $S^n(E_V)$, that we name the ***generic moment of order** n **of*** ρ, and write $\Phi_V^n(\rho)$. The linear form $\Phi_V^n(\rho)$ is the restriction to $S^n(E_V^*)$ of the linear form $\Phi_V^{(m)}(\rho)$ on $H_V^{m)}$. Varying V, this gives a natural transformation from $\mathscr{P}$ to $\mathscr{E}_n$.

Proposition 11.34. *This function Φ_V^n is invariant under the action of $\mathscr{A}_V$.*

Proof. It is an immediate corollary of the Proposition 11.32. □

For $n = 0$, $S^0(E_V) = \mathbb{R}$, and the integration of a constant C is equal to C.

For $n = 1$, $S^1(E_V)$ is equal to E_V, and the moment is the mean $\rho \mapsto M_V(\rho)$. The mean vector $\{M_V\}_{V \in \mathrm{Ob}\,\mathbf{S}}$ gives a natural transformation from the covariant functor $\mathscr{P}$ to the covariant functor $\mathscr{E} : V \mapsto E_V$. The naturality is equivalent to the equations of direct images: $M_W(\pi_*\rho) = \pi(M_V(\rho))$.

11.4 Grassmannian probability modules

More generally, given a grassmannian category of E, we can consider collections $\mathscr{Q}_V$, for $V \in \mathrm{Ob}\,\mathbf{S}$ of probability laws that are absolutely continuous with respect to one of the Lebesgue measure in $\mathscr{L}_V$, corresponding to an affine subspace $A(\rho)$ in $\mathscr{M}_V$ (or more generally $\mathscr{N}_V \subseteq \mathscr{M}_V$), and we can ask that these collections are stable by push-forward $\pi_*^{WV} : \mathscr{Q}_V \to \mathscr{Q}_W$ when $V \subseteq V$, and also by conditioning $\rho|_{W=w}$, i.e. by the elements of disintegration.

We can introduce the presheaf $\mathscr{F}(\mathscr{Q})$, of universally integrable functions, that maps $V \in \mathrm{Ob}\,\mathbf{S}$ to the set $\mathscr{F}(\mathscr{Q})_V$ of measurable functions on $\mathscr{Q}_V$ such that all the integrals considered for defining the action of $\mathscr{A}$ are convergent. The general arguments of the main text show that the axioms of $\mathscr{A}$-module are satisfied.

All that also works with the bundles $\ln \mathscr{L}^a$, and the following general result holds:

Proposition 11.35. *If for every $V \in \mathrm{Ob}\,\mathbf{S}$, every $\rho \in \mathscr{Q}_V$, and any $\lambda \in \mathscr{L}_{A(\rho)}$, the function $\ln g_\lambda(\rho)$ is ρ integrable, then the multiple aS of the entropy is a cocycle of $\mathscr{F}^{-a}$.*

Proof. Working in $T(A)$, with the restriction of the projection $\pi = \pi^{WV}$, we choose Lebesgue measures adapted to the image $\mathrm{d}y$ and the kernel $\mathrm{d}x$ (cf. Proposition F.9). Let $g(x, y)$ be the joined density of (X, Y) with respect to $\mathrm{d}x\,\mathrm{d}y$; we have

$$g(y) = \int g(x, y)\,\mathrm{d}x. \tag{11.111}$$

Then

$$\int\int g(x, y) \ln g(x, y)\,\mathrm{d}x\,\mathrm{d}y$$
$$= \int \left(\int \frac{g(x, y)}{g(y)} \ln \frac{g(x, y)}{g(y)}\,\mathrm{d}x\right) g(y)\,\mathrm{d}y + \int g(y) \ln g(y)\,\mathrm{d}x. \tag{11.112}$$

□

In the same manner as for the gaussian laws, the dimension of the support $d(A(\rho))$ is a cocycle of $\mathscr{F}(\mathscr{Q})$, then we can add any multiple $c.d(A)$ to the entropy S, and obtain a class of cohomology of degree one of $\mathscr{X}^{-a}$.

In many cases it is also possible to define the mean and the moments, as natural and invariant transformations.

Chapter 12

Generalized entropy and asymptotic concentration of measure

We state the asymptotic equipartition property for a sequence of independent and identically distributed random variables in a very general form. This complements the results in the other chapters showing that, for any σ-finite measure μ and any probability law ρ that is absolutely continuous with respect to μ, there is notion of entropy:

$$S_\mu(\rho) = -\int \ln\left(\frac{\mathrm{d}\mu}{\mathrm{d}\rho}\right) \mathrm{d}\rho, \tag{12.1}$$

that is relevant from a probabilistic viewpoint. As $n \to \infty$, the iterated law $\rho^{\otimes n}$ concentrates on a typical set, whose μ-volume is approximately $\exp(nS_\mu(\rho))$.

12.1 Asymptotic equipartition property

Let $(E_X, \mathfrak{B})$ be a measurable space, supposed to be the range of some random variable X, and let μ be a σ-finite measure μ on it. In applications, several examples appear:

1. E_X a countable set, $\mathfrak{B}$ the corresponding atomic σ-algebra, and μ the counting measure;
2. E_X the real line, $\mathfrak{B}$ the Borel σ-algebra, and μ the Lebesgue measure;
3. as a generalization of the previous one, E_X a locally compact, Hausdorff topological group, $\mathfrak{B}$ its Borel σ-algebra, and μ some Haar measure.

The reference measure μ gives the relevant notion of volume.

Consider now a probability measure ρ on $\mathfrak{B}$, that is taken to be the *law* of X. By the Lebesgue decomposition theorem, there exists a *unique* decomposition $\rho = \rho_1 + \rho_2$, such that ρ_1 is absolutely continuous with respect to μ (write $\rho_1 \ll \mu$), and ρ_2 and μ are mutually singular (write $\rho_2 \perp \mu$). [1] When $\rho_2 = 0$, we say that X is nonsingular; in this case, the law ρ has a *density* f with respect to the reference measure μ, that is defined as the Radon-Nikodym derivative, $f = \mathrm{d}\rho/\mathrm{d}\mu$.

---

[1]Concretely, the decomposition is defined as follows: introduce the Radon-Nikodym derivative $f = \mathrm{d}\rho/\mathrm{d}(\mu+\rho)$, and set $B = \{\, x \in X \mid f(x) \geq 1 \,\}$. Up to a set of $(\mu+\rho)$-measure zero, this do not depend on the representative of the derivative. Then, $\rho_2(A) := \rho(A \cap B)$, and $\rho_1(A) := \rho(A \cap B^c)$.

Let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of independent and identically distributed (iid) E_X-valued random variables, each one distributed according to a probability law ρ on $(E_X, \mathfrak{B})$ that is supposed to be *non-singular* with respect to the reference measure μ, with density $f := \mathrm{d}\rho/\mathrm{d}\mu$.[2] For each $n \in \mathbb{N}$, the joint variable $(X_1, \dots, X_n)$ takes values in $(E_X^n, \mathfrak{B}^{\otimes n}, \mu^{\otimes n})$; by definition of independence, its law is $\rho^{\otimes n}$.[3] Proposition F.6 shows that $\mu^{\otimes n}$ is also σ-finite. Since $\rho^{\otimes n} \ll \mu^{\otimes n}$, we can apply the Radon-Nikodym theorem to define a joint density $f_{X_1,\dots,X_n}(x_1, \dots, x_n) := \frac{\mathrm{d}\rho^{\otimes n}}{\mathrm{d}\mu^{\otimes n}}$. The repeated application of Fubini's theorem shows that

$$\int_{B_1\times\cdots\times B_n} \frac{\mathrm{d}\rho^{\otimes n}}{\mathrm{d}\mu^{\otimes n}}\,\mathrm{d}\mu^{\otimes n} = \int_{B_1\times\cdots\times B_n} \mathrm{d}\rho^{\otimes n} \tag{12.3}$$

$$= \prod_{1\le i\le n} \int_{B_i} \mathrm{d}\rho \tag{12.4}$$

$$= \int_{B_1} \cdots \int_{B_n} \prod_i \frac{\mathrm{d}\rho}{\mathrm{d}\mu}(x_i)\,\mathrm{d}\mu(x_n)\cdots\,\mathrm{d}\mu(x_1), \tag{12.5}$$

which implies that $\rho^{\otimes n}$ has density

$$f_{X_1,\dots,X_n}(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i). \tag{12.6}$$

Proposition 12.1. *Let* $(E_X, \mathfrak{B}, \mu)$ *be a* σ*-finite measure space and* $\{X_i : (\Omega, \mathfrak{F}, \mathbb{P}) \to (E_X, \mathfrak{B}, \mu)\}_{i\in\mathbb{N}}$ *a collection of iid random variables, each following a law* ρ *with density* $f = \mathrm{d}\rho/\mathrm{d}\mu$ *with respect to* μ*. If the Lebesgue integral*

$$S_\mu(\rho) := \mathbb{E}_\rho\left(-\ln\frac{\mathrm{d}\rho}{\mathrm{d}\mu}\right) = -\int_{\operatorname{supp}\mu} f(x)\log f(x)\,\mathrm{d}\mu(x) \tag{12.7}$$

is finite, then

$$-\frac{1}{n}\log f_{X_1,\dots,X_n}(X_1, \dots, X_n) \to S_\mu(\rho), \tag{12.8}$$

$\mathbb{P}$*-almost surely.*

Proof. Remark first that

$$\rho^{\otimes n}(\{f_{X_1,\dots,X_n} = 0\}) = \int_{\{f_{X_1,\dots,X_n}=0\}} 0\,\mathrm{d}\mu = 0,$$

---

[2]In usual presentations of probability theory, first an auxiliary sample space $(\Omega, \mathfrak{F}, \mathbb{P})$ is introduced, then each X_i is defined as a measurable function $X_i : (\Omega, \mathcal{F}) \to (E_X, \mathfrak{B})$, and the law ρ corresponds to $\mathbb{P}\circ X_i^{-1}$.

[3]Denote by ρ_n the joint law. Independence implies that, for any collection $\{B_1, \dots, B_n\} \subset \mathcal{B}$,

$$\rho_n(B_1\times\cdots\times B_n) = \prod_{i=1}^n \rho(B_i) = \rho^{\otimes n}(B_1\times\cdots\times B_n). \tag{12.2}$$

For the the collection of all rectangles $B_1\times\cdots\times B_n$ constitute a π-system for the σ-algebra $\mathfrak{B}^{\otimes n}$, we conclude that $\rho_n = \rho^{\otimes n}$ on $\mathcal{B}$. See Lemma 1.6 in [96]: "if two *probability* measures agree on a π-system, then they agree on the σ-algebra generated by that π-system."

hence $\log f_{X_1,\dots,X_n}(X_1,\dots,X_n)$ is well defined $\mathbb{P}$-a.e. Moreover, $\log f_{X_1,\dots,X_n}(X_1,\dots,X_n) = \sum_{i=1}^n \log f(X_i)$, because of (12.6). The variables $\{-\log f(X_i)\}$ are iid and *real valued*, so the strong law of large numbers [96, Sec. 14.5] states that

$$-\frac{1}{n}\sum_{i=1}^n \log f(X_i) \to \mathbb{E}_\rho\left(-\log f(X_i)\right),$$

$\mathbb{P}$-almost surely. □

Remark 12.2. The strong law for an iid sequence $Y_1, Y_2, Y_3 \dots$ requires that, for all k, $\mathbb{E}(|Y_k|) < \infty$ (i.e. the function Y_k is Lebesgue integrable). In this case, $\mathbb{E}(Y_k) = \mathbb{E}\left(|\log f(X_0)|\right)$, and this must be finite in order for $S_\mu(\rho)$ to be finite.

Proposition 12.3 (Asymptotic Equipartition Property)**.** *We use the notation introduced in Proposition 12.1 and we suppose that $S_\mu(\rho)$ is finite. For every $\delta > 0$, set*

$$A_\delta^{(n)} := \left\{ (x_1,\dots,x_n) \in E_X^n \;\middle|\; \left| -\frac{1}{n}\log f_{X_1,\dots,X_n}(X_1,\dots,X_n) - S_\mu(\rho)\right| \le \delta \right\}. \tag{12.9}$$

Then,

1. *for every $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that, for all $n \ge n_0$,*
$$\mathbb{P}\left(A_\delta^{(n)}\right) > 1 - \varepsilon;$$
2. *for every $n \in \mathbb{N}$,*
$$\mu^{\otimes n}(A_\delta^{(n)}) \le \exp\{n(S_\mu(\rho) + \delta)\};$$
3. *for every $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that, for all $n \ge n_0$,*
$$\mu^{\otimes n}(A_\delta^{(n)}) \ge (1-\varepsilon)\exp\{n(S_\mu(\rho) - \delta)\}.$$

Proof. Part (1) is just a consequence of Proposition 12.1, since almost sure convergence implies convergence in probability.

For the rest, note first that, for every $(x_1,\dots,x_n) \in A_\delta^{(n)}$,

$$e^{-n(S_\mu(\rho)+\delta)} \le f(x_1,\dots,x_n) \le e^{-n(S_\mu(\rho)-\delta)}.$$

Therefore,

$$1 \ge \int_{A_\delta^{(n)}} f_{X_1,\dots,X_n}\,\mathrm{d}\mu^{\otimes n} \ge \int_{A_\delta^{(n)}} e^{-n(S_\mu(\rho)+\delta)}\,\mathrm{d}\mu^{\otimes n},$$

which implies statement (2).

For part (3): take $\varepsilon > 0$ and remark that that, if n is big enough,

$$1 - \varepsilon \le \mathbb{P}\left(A_\delta^{(n)}\right) = \int_{A_\delta^{(n)}} f_{X_1,\dots,X_n}\,\mathrm{d}\mu^{\otimes n} \le e^{-n(S_\mu(\rho)-\delta)}\mu^{\otimes n}(A_\delta^{(n)}).$$

□

When E_X is a countable set (possibly infinite), and μ the counting measure, a probability law ρ on E_X is always absolutely continuous with respect to μ, and its density is a function $p : E_X \to [0,1]$ such that $\sum_{x\in E_X} p(x) = 1$, that is usually taken as the definition of a probability law in the discrete case. Then $S_\mu(\rho)$ corresponds to the familiar expression $-\sum_{x\in E_X} p(x)\log p(x)$, that we have denoted S_1. This function was studied by Shannon in [78] for the case $|E_X| < \infty$, where he also stated a version of the asymptotic equipartition property. The idea behind this theorem apparently goes back to Boltzmann [79]. It is also possible to consider any multiple of the of the counting measure, $\nu = \alpha\mu$, for $\alpha > 0$. In this case, the condition $\int_{E_X} \frac{\mathrm{d}\rho}{\mathrm{d}\nu}\,\mathrm{d}\nu = \int_{E_X} \mathrm{d}\rho = 1$ translates into $\alpha \sum_{x\in E_X} \frac{\mathrm{d}\rho}{\mathrm{d}\nu} = 1$, i.e. a probability density that is normalized to sum α^{-1}.

If $E_X = \mathbb{R}^n$, μ is the corresponding Lebesgue measure, and ρ a probability law such that $\rho \ll \mu$, then the derivative $\mathrm{d}\rho/\mathrm{d}\mu \in L^1(\mathbb{R}^n)$ corresponds to the elementary notion of density, and the quantity $S_\mu(\rho)$ is the continuous entropy that was also introduced by Shannon in [78] in order to study continuous signals.

For any E_X, if μ is a probability law, the expression $S_\mu(\rho)$ is (the opposite of) the Kullback-Leibler divergence.

12.2 Certainty and divergence

Proposition 12.3 makes precise the relation between the entropy $S_\mu(\rho)$ and certainty.

1. Discrete case: let E_X be a countable set and μ be the counting measure. Consider an E_X-valued variable X with law $\rho = \delta_{x_0}$, for certain $x_0 \in E_X$. Then, $S_\mu(\rho) = 0$ and $A^{(n)}_\delta = \{(x_0, \dots, x_0)\}$, whose volume is evidently 1, for every n. Therefore, $\mu^{\otimes n}(A^{(n)}_\delta) = \exp(nS_\mu(\rho))$, and the inequalities in Proposition 12.3 are satisfied.
2. Suppose each X_i distributes uniformly on $B(x_0, \varepsilon) \subset \mathbb{R}^d$, which means that its density $\mathrm{d}\rho/\mathrm{d}\lambda$ is $|B(x_0,\varepsilon)|^{-1}\chi_{B(x_0,\varepsilon)}$ (the notation $|A|$ stands for $\lambda_d^{\otimes n}(A)$, the Lebesgue measure). In this case, $H_{\lambda_d}(\rho) = \log(|B(x_0,\varepsilon)|) = \log(c_d\varepsilon^d)$, where c_d is a constant characteristic of each dimension d, and $H_{\lambda_d}(\rho) \to -\infty$ as $\varepsilon \to 0$ (i.e. as X_i concentrates on a single output x_0). Part (2) in Proposition 12.3 says that
$$|A^{(n)}_\delta| \le \exp(nd\log\varepsilon + Cn), \tag{12.10}$$
which means that, for fixed n, the volume goes to zero as $\varepsilon \to 0$, as intuition would suggest. Therefore, *the divergent entropy is necessary to obtain the good volume estimates.*
3. When each X_i takes values in $\mathbb{R}^2$, according to a Gaussian law ρ with mean 0 and covariance matrix $\Sigma = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}$, with $r \in [0,1]$. Whenever $r \neq 1$, the law ρ is absolutely continuous with respect to the Lebesgue measure λ_2, and $H_{\lambda_2}(\rho) = \frac{1}{2}\ln(\det(2\pi e\Sigma)) = \ln(2\pi) + \ln(1-r^2)$. For every $n \in \mathbb{N}$,
$$|A^{(n)}_\delta| \le \exp(n\ln(1-r^2) + Cn) \to 0 \qquad \text{as } r \to 1. \tag{12.11}$$
The case $r = 1$ corresponds to a singular ρ with respect to the Lebesgue measure on $\mathbb{R}^2$: the probability is concentrated on the diagonal $\Delta = \mathrm{col}(\Sigma)$, whose Hausdorff dimension equals 1. The relevant reference volume is then a

Lebesgue measure supported on this affine subspace, that can also be seen as the corresponding Hausdorff measure.

12.3 Example: Rectifiable subspaces of $\mathbb{R}^n$

In [51], Riegler, Hlawatsch, Koliander, and Pichler introduced another family of examples for the AEP as stated in the previous section. The important technical point is that the Hausdorff measure associated to a rectifiable set, or a countable union of such sets, is a σ-finite measure.

Set $E_X = \mathbb{R}^d$, and denote by λ_d the corresponding Lebesgue measure. The diameter of a subset S of $\mathbb{R}^d$, is $\operatorname{diam}(S) = \sup\{\, |x - y| \mid x, y \in S \,\}$. For any $m \geq 0$ and any $A \subset \mathbb{R}^n$, define

$$\mathcal{H}^m(A) = \lim_{\delta \to 0} \inf_{\{S_i\}_{i \in I}} \sum_{i \in I} \alpha_m \left(\frac{\operatorname{diam}(S_i)}{2} \right)^m, \tag{12.12}$$

where α_m is a constant and the infimum taken over all countable coverings $\{S_i\}_{i \in I}$ of A such that each set S_i has diameter at most δ. This is an outer measure in $\mathbb{R}^n$, that restricts to a measure $\mathcal{H}^m$ on Borel subsets [66], called Hausdorff measure. If $\alpha_d := \pi^{d/2}/\Gamma(d/2 + 1)$, which is the volume of the ball $B(0, 1) \subset \mathbb{R}^m$ when m is an integer, the measure $\mathcal{H}^d$ coincides with λ_d.

Let S be a measurable subset of $\mathbb{R}^n$, and let m be its Hausdorff dimension:

$$m = \inf\{\, k \mid \mathcal{H}^k(S) = 0 \,\}.$$

The set S can be seen as a measurable space with the induced σ-algebra $\mathcal{A} := \{\, S \cap B \mid B \in \mathcal{B}(G) \,\}$ (that is also the Borel σ-algebra for the induced topology on S); the measure $\mathcal{H}^m$ restricts to a measure on $(S, \mathcal{A})$, that we denote λ_S (nonstandard notation): for every $A \in \mathcal{A}$, $\lambda_S(A) := \mathcal{H}^1(A)$. Remark that $\lambda_{\mathbb{R}^d} = \lambda_d$. We introduce now a particular family of subsets S such that λ_S behaves well.

Definition 12.4 (Countably rectifiable sets)**.** An $\mathcal{H}^m$ measurable subset S of $\mathbb{R}^d$ is called ***countably*** m***-rectifiable*** (for $m \leq d$) if there exist Lebesgue measurable, bounded sets $A_k \subset \mathbb{R}^m$ and Lipschitz functions $f_k : A_k \to \mathbb{R}^d$, enumerated by $k \in \mathbb{N}$, such that $\mathcal{H}^m(S \setminus \cup_k f_k(A_k)) = 0$. The set S is called 0-rectifiable if it is countable.

If S is countable m-rectifiable, the measure $\lambda_S = \mathcal{H}^m|_S$ is σ-finite [51, Lemma 4] and we can apply Proposition 12.3 to any probability measure absolutely continuous with respect to λ_S. Moreover, the measure $\lambda_S^{\otimes n}$ corresponds to the Hausdorff measure $\mathcal{H}^{nm}$ restricted to S^n [51, Lemma 27]. Reference [51] discusses some particular cases: distributions on the unit circle, and positive semidefinite rank-one random matrices.

Part V

General background material

Appendix A

Category theory

A.1 Notations

Given a category $\mathbf{C}$, we denote by $\operatorname{Ob}\mathbf{C}$ the class of its objects and $\operatorname{Hom}(\mathbf{C})$ the class of morphisms in $\mathbf{C}$. Given $A, B \in \operatorname{Ob}\mathbf{C}$, the class of morphisms between A and B is denoted by $\operatorname{Hom}_{\mathbf{C}}(A, B)$ or simply $\operatorname{Hom}(A, B)$ whenever $\mathbf{C}$ is clear from context.

A category is called ***small*** if the class $\operatorname{Ob}\mathbf{C}$ and $\operatorname{Hom}(\mathbf{C})$ are sets (instead of proper classes). One can also work with Grothendieck's universes, in which case $\operatorname{Ob}\mathbf{C}$ and $\operatorname{Hom}(\mathbf{C})$ are supposed to belong to a fixed universe. In this work, all categories are small.

A full subcategory $\mathbf{D}$ of $\mathbf{C}$ is such that $\operatorname{Ob}\mathbf{D} \subset \operatorname{Ob}\mathbf{C}$ and for each pair $A, B \in \operatorname{Ob}\mathbf{D}$,

$$\operatorname{Hom}_{\mathbf{D}}(A, B) = \operatorname{Hom}_{\mathbf{C}}(A, B).$$

A.2 Subobjects and quotients

Given two monomorphisms $u : B \to A$ and $u' : B' \to A$, one says that u ***dominates*** u', written also $u \geq u'$, if it is possible to factor u as $u'v$, where $v : B \to B'$ (and this morphism is therefore uniquely determined). This is a preorder on the class of monomorphisms with codomain A. Two morphisms u and u' are equivalent if $u \geq u'$ and $u' \geq u$, and in this case the corresponding morphisms $b \to B'$ and $B' \to B$ are inverses one of each other. Choose then a representative of each class of monomorphisms with values in A; these representatives are called ***subobjects*** of A. Therefore, a subobject is an object B equipped with a map $u : B \to A$, called canonical injection. The relation $\geq$ is an order relation on the equivalence classes. The consideration of an analogous preorder on the class of epimorphisms with domain A allow us to define the ordered class of ***quotients*** of A.

Appendix B

Abstract simplicial complexes

In this section, we recall the main definitions concerning abstract (or combinatorial) simplicial complexes, for the convenience of the reader. Most of them are taken verbatim from [56].

We define an ***(abstract) simplicial complex*** as a collection $\mathcal{K}$ of nonempty finite subsets of a finite set S, subject to the condition: if $s \in \mathcal{K}$, then every nonempty subset of s is in $\mathcal{K}$. A subcomplex $\mathcal{K}'$ of $\mathcal{K}$ is collection of subsets of S contained in $\mathcal{K}$ that also satisfies the condition above.

The finite sets that make up $\mathcal{K}$ are called abstract simplices. Given an abstract simplex $s \in \mathcal{K}$, its elements are called vertexes and its nonempty subsets are called faces. We say that $\mathcal{K}$ is a finite complex when $\mathcal{K}$ is a finite set, and locally finite if every vertex belong to a finite number of simplices. The dimension of an abstract simplex $s \in \mathcal{K}$ is $|s| - 1$. When the dimensions of the simplices of $\mathcal{K}$ are bounded above, the complex is finite dimensional and its dimension is the smallest upper bound.

The d-skeleton of $\mathcal{K}$ is the subcomplex of $\mathcal{K}$ consisting of all the simplices that have dimension at most d.

The vertex set of a complex $\mathcal{K}$ is

$$K_0 = \bigcup_{s \in \mathcal{K}} s. \tag{B.1}$$

A simplicial map $f : \mathcal{K} \to \mathcal{L}$ is given by a vertex map $f_0 : \mathcal{K}_0 \to \mathcal{L}_0$ which must satisfy the property that $f(s) := \{f_0(v_1), \dots, f_0(v_k)\} \in \mathcal{L}$ whenever $s = \{v_1, \dots, v_k\} \in \mathcal{K}$.

Example B.1. The abstract simplex $\Delta([n])$ is the simplicial complex $\mathcal{P}([n])$: its 0-dimensional simplices are the singletons, the 1-dimensional simplices are the sets of cardinality two, etc.

Appendix C

Linear algebra

C.1 Schur complements

For any block matrix

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

such that A and D are square matrices (not necessarily of the same size) and A is invertible, the Schur complement of A in M is the matrix $\bar{A} := D - BA^{-1}C$ (the standard notation is M/A). The matrix M/D is defined analogously, provided D is nonsingular. The determinant of M can be computed with a formula proposed by Schur in [75, p. 31].

Proposition C.1 (Schur's determinantal identity, [14, p. 3])**.** *Let*

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

be a block matrix such that A and D are square matrices (not necessarily of the same size). If A is invertible, then

$$|M| = |A||M/A|. \tag{C.1}$$

Similarly, if D is invertible,

$$|M| = |D||M/D|. \tag{C.2}$$

Proof. When A is invertible, the identity

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} Id_1 & -A^{-1}B \\ 0 & Id_2 \end{pmatrix} = \begin{pmatrix} A & 0 \\ C & D - CA^{-1}B \end{pmatrix} \tag{C.3}$$

implies the claim. □

Appendix D

Multivariate normal distributions

In this section, we follow closely [70, pp. 517ff.]. We also include a more explicit discussion of the effects of a change of basis.

Let E_p be a vector space of dimension p.

Definition D.1. A E_p-valued random variable U is said to have a p-variate normal distribution if and only if every linear functional of U has a univariate normal distribution.

We can restate this definition just in terms of the law of U.

Definition D.2. A law ρ on E_p is gaussian (or normal) if, for every linear functional $L : E_p \to \mathbb{R}$, the image measure $L\rho$ corresponds to a univariate normal probability measure.

Remark that these definitions are independent of any basis. If we fix a basis $\mathcal{B} = \{b_1, \dots, b_n\}$ and denote by $B^* = \{b_1^*, \dots, b_n^*\}$ the dual basis, then we can write $U = (U_1, \dots, U_n)$ and introduce the mean vector

$$m := \mathbb{E}(U) := (\mathbb{E}(U_1), \dots, \mathbb{E}(U_n)) \tag{D.1}$$

and the covariance matrix

$$D(X) := (\mathrm{Cov}(U_i, U_j))_{0 \le i,j \le n}, \tag{D.2}$$

that we denote $\Sigma = (\sigma_{ij})_{0 \le i,j \le n}$. In terms of the law ρ, the equivalent definitions are $m_i = \int_{\mathbb{R}} x \, \mathrm{d}b_i^* \rho(x)$ and $\sigma_{ij} = \int_{\mathbb{R}} \int_{\mathbb{R}} (x - m_i)(y - m_j) \, \mathrm{d}b_i^* \rho(x) \, \mathrm{d}b_j^* \rho(y)$.

If X is a p-dimensional random variable and M is any $(q \times p)$ matrix,

$$\mathbb{E}\left(M^{\mathrm{tr}} U\right) = M^{\mathrm{tr}} \mathbb{E}(X) \tag{D.3}$$

$$D(MX) = MD(X)M^{\mathrm{tr}} \tag{D.4}$$

Therefore, if $C = \{c_1, \dots, c_n\}$ is another basis and A the change-of-basis matrix ($c_i = Ab_i$), then the previous identities imply that Am and $A\Sigma A^{\mathrm{tr}}$ are the mean and variance of U expressed in the basis C.

From now on, we consider that a basis of E_p has been fixed and therefore that each normally distributed vector U (equivalently, each gaussian law ρ) has a well defined mean and variance. We write $U \sim N_p(m, \Sigma)$.

Proposition D.3 (Structure theorem for normally distributed variables, [70, p. 521])**.** *$U \sim N_p(m, \Sigma)$ with* $\operatorname{rank}\Sigma = r$ *if, and only if,*

$$U = m + BG, \quad BB' = \Sigma$$

where B is a $(p \times r)$ matrix of rank r and $G \sim N_r(0, I)$, that is, the components $G_1, \dots, G_r$ are independent and each is distributed as $N_1(0, 1)$.

In fact, since Σ is a real symmetric matrix, it can be diagonalized by an orthogonal matrix P i.e. $\Sigma = PDP'$. When Σ is positive definite (has full rank), B is simply $P\sqrt{D}$; when Σ is positive semidefinite, one can ignore the null columns of $P\sqrt{D}$, that are associated to the null eigenvalues. If $B_1, \dots, B_r$ denote the linearly independent columns of B, then for every $T \in \mathbb{R}^p$,

$$T^{\mathrm{tr}}\Sigma T = (B_1^{\mathrm{tr}}T)^2 + \dots + (B_r^{\mathrm{tr}}T)^2, \tag{D.5}$$

which corresponds to Sylvester's law of inertia.

Corollary D.4. *For every $A \in M_{p,q}(\mathbb{R})$ and $b \in \mathbb{R}^p$, if $X \sim N_q(m, \Sigma)$, then $Ax + b \sim N_p(Am + b, A\Sigma A^{\mathrm{tr}})$.*

Proof. That $Ax \sim N_p(Am, A\Sigma A^{\mathrm{tr}})$ can be verified directly from the definition: for every linear functional represented by T^{tr}, we have $T^{\mathrm{tr}}AX = (T^{\mathrm{tr}}A)X$, which is a univariate normal (since X is normal) with mean Am and covariance $A\Sigma A^{\mathrm{tr}}$. The structure theorem makes clear that a translation by b just translates the mean. □

Remark D.5. Given a matrix A, a generalized inverse A^- is any matrix that verifies $AA^-A = A$. Generalized inverses always exists, but in general they are not unique. However, they are unique when a is invertible and in this case $A^{-1} = A^-$.

Proposition D.6. *Let Y be an E_p-valued random variable with p-variate normal distribution, $Y \sim N_p(\mu, \Sigma)$. Given a decomposition $E_p = E_q \times E_s$, let us introduce the notations $Y = (Y_1, Y_2) \in E_q \times E_s$, $m_i = \mathbb{E}(Y_i)$, and $\Sigma_{ij} = \operatorname{Cov} Y_i Y_j$, in such a way that $m = (m_1, m_2)$, and*

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \sigma_{22} \end{pmatrix}.$$

Then,

1. *$Y_1 \sim N_q(m_1, \Sigma_{11})$ and $Y_2 \sim N_s(m_2, \Sigma_{22})$.*
2. *$Y_1 | Y_2 = y_2 \sim N_q(\bar{m}_1(y_2), \bar{\Sigma}_{11})$, where $\bar{m}_1(y_2) := m_1 + \Sigma_{12}\Sigma_{22}^-(y_2 - m_2)$, and $\bar{\Sigma}_{11} := \Sigma_{11} - \Sigma_{12}\Sigma_{22}^-\Sigma_{21}$.*
3. *$Y | Y_2 = y_2 \sim N_p\left(\begin{pmatrix} m_1(y_2) \\ y_2 \end{pmatrix}, \begin{pmatrix} \bar{\Sigma}_{11} & 0 \\ 0 & 0 \end{pmatrix}\right)$.*
4. *If Σ is positive definite, then Σ_{11} and Σ_{22} too. Hence $\Sigma_{ii}^- = \Sigma_{ii}^{-1}$ in the previous formulae.*

Proposition D.7. *The law of $Y \sim N_p(m, \Sigma)$ is absolutely continuous with respect to λ_p if and only if Σ is positive definite.*

Proof. We use the representation $Y = m + BG$ introduced above.

If Σ is positive definite, all its eigenvalues are strictly positive and the rank of Σ is p. Thus B is a $p \times p$ orthogonal matrix. Let $A \subset \mathbb{R}^p$ be a set of Lebesgue measure zero. Then $B^{-1}(A-m)$ also has Lebesgue measure zero and

$$\mathbb{P}(Y \in A) = \mathbb{P}\left(G \in B^{-1}(A-m)\right) = \int_{B^{-1}(A-m)} \frac{e^{-\frac{1}{2}x^2}}{(2\pi)^{d/2}}\,\mathrm{d}x = 0.$$

If Σ is not positive definite, its rank r is strictly less than p. The vector Y belongs to the affine space $m + \mathrm{col}(B)$, which has dimension r and hence p-Lebesgue measure zero. $\square$

Appendix E

Distribution theory

For all the relevant definitions, see [76] or [86].

Let us introduce first the multi-index notation. An n-dimensional multi-index is a vector $\alpha = (\alpha_1, ..., \alpha_n) \in \mathbb{N}^n$. They can be added and subtracted component-wise. We say that $\alpha \leq \beta$ if both multi-indices have the same size and $\alpha_i \leq \beta_i$ for every i. Furthermore, $\alpha! := \alpha_1! \cdots \alpha_n!$, and for any $\nu \leq \alpha$,

$$\binom{\alpha}{\nu} := \prod_{i=1}^{n} \binom{\alpha_i}{\nu_i}.$$

Given $x \in \mathbb{R}^n$, $x^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. Given a function $f \in C^\infty(\mathbb{R}^n)$, $\partial^\alpha f := \partial_{\alpha_1}\partial_{\alpha_2} \cdots \partial_{\alpha_n} f$.

The space of distributions $\mathcal{D}'(\Omega)$, the dual of smooth functions with compact support contained in an open set Ω of $\mathbb{R}^d$, is defined in [76, Ch. 3]. We use the following characterization of distributions supported on a point.

Proposition E.1 (Distributions supported in a single point, cf. [76, Thm. XXVII] and [82, Thm. 1.70])**.** *Let Ω be an open set in $\mathbb{R}^d$, $x \in \Omega$ and $u \in \mathcal{D}'(\Omega)$ with* $\operatorname{supp} u = \{x_0\}$. *Then there exist $m \in \mathbb{N}$ and, for every multi-index α that verifies $|\alpha| \leq m$, a constant $c_\alpha \in \mathbb{C}$ such that*

$$\forall \varphi \in \mathcal{D}(\Omega), \quad \langle u, \varphi \rangle = \sum_{|\alpha| \leq m} c_\alpha \partial^\alpha \varphi(x_0).$$

Definition E.2. The ***Schwartz space***, denoted by $\mathcal{S}(\mathbb{R}^d)$, is defined as the space of smooth fast-decreasing complex functions over $\mathbb{R}^d$. Explicitly,

$$\mathcal{S}(\mathbb{R}^d) := \{\varphi \in C^\infty(\mathbb{R}^d) \;\big|\; \forall \alpha, \beta \in \mathbb{N}^d \; \|x^\alpha D^\beta \varphi\|_\infty < \infty\}. \tag{E.1}$$

The space $\mathcal{S}'(\mathbb{R}^d)$ of ***tempered distributions*** is the topological dual of $\mathcal{S}(\mathbb{R}^d)$.

The topological space $\mathcal{S}(\mathbb{R}^d)$ is a complete, metrizable, and embeds continuously in $L^1(\mathbb{R}^d)$. By definition, the Fourier transform $\mathcal{F}$ associates to any $f \in L^1(\mathbb{R}^d)$ the function $\widehat{f} : (\mathbb{R}^d)^* \to \mathbb{C}$ defined by

$$\widehat{f}(\xi) = \int_{\mathbb{R}^d} f(x) \mathrm{e}^{-2\pi i \xi \cdot x} \, \mathrm{d}x. \tag{E.2}$$

One can prove that $\mathcal{F}(\mathcal{S}) \subset \mathcal{S}$. In fact, $\mathcal{F} : \mathcal{S} \to \mathcal{S}$ is an isomorphism of topological vector spaces [86, Thm. 25.1]. Its transpose is also denoted by $\mathcal{F}$ and extends the definition of the Fourier transform to the space of tempered distributions.

Definition E.3. The space of ***multiplication operators*** (cf. [86, Def. 25.3]), denoted by $\mathcal{O}_M(\mathbb{R}^d)$ is defined as the space of all complex smooth functions such that all of their derivatives of all orders are polynomially bounded. Explicitly,

$$\begin{aligned}\mathcal{O}_M := \{f \in C^\infty(\mathbb{R}^d) \mid \forall \alpha \in \mathbb{N}^d\ \exists C > 0\ \exists N \in \mathbb{N} \\ \text{such that } \forall x \in \mathbb{R}^d,\ |D^\alpha f(x)| \le C(1 + |x|^2)^N\ \}. \end{aligned} \tag{E.3}$$

If $f \in \mathcal{O}_M$ and $\varphi \in \mathcal{S}(\mathbb{R}^d)$, then $f\varphi \in \mathcal{S}(\mathbb{R}^d)$.

If $T \in \mathcal{S}'(\mathbb{R}^d)$, its ***convolution*** $T * \varphi$ with a function $\phi \in \mathcal{S}(\mathbb{R}^d)$ is defined as the function $x \in \mathbb{R}^d \to (T * \varphi)(x) := \langle T, \tau_x \check{\varphi}\rangle$, where $\tau_x\varphi(y) = \varphi(x + y)$ et $\check{\varphi}(y) = \varphi(-y)$, and this function belongs to $\mathcal{O}_M$, see [86, Thm. 30.2].

Definition E.4. A tempered distribution $T \in \mathcal{S}'(\mathbb{R}^d)$ is called a ***convoluter*** (cf. [86, Def. 30.1]) if for all $N \in \mathbb{N}$ there exists $M_N \in \mathbb{N}$ and a finite family of continuous functions $\{f_\alpha\}_{\alpha\in\mathbb{N}^d, |\alpha|\le M_N} \subset C(\mathbb{R}^d)$ such that $(1 + |x|^2)^N f_\alpha \in C_0(\mathbb{R}^d)$ for all $\alpha \in \mathbb{N}^d$ with $|\alpha| \le M_N$, and such that

$$T = \sum_{|\alpha|\le M_N} \partial^\alpha f_\alpha, \tag{E.4}$$

where the derivatives are taken in the distributional sense.

The space of convoluters of tempered distributions over $\mathbb{R}^d$ is denoted by $\mathcal{O}'_C(\mathbb{R}^d)$. The Fourier establishes a bijection between $\mathcal{O}_M$ and $\mathcal{O}'_C$.

Proposition E.5 (Convolution theorem, [86, Thm. 30.4])**.** *Let $S \in \mathcal{S}'$, $T \in \mathcal{O}'_C$ and $\alpha \in \mathcal{O}_M$ (over $\mathbb{R}^d$). Then $\mathcal{F}(S * T) = \mathcal{F}(S)\mathcal{F}(T)$ and $\mathcal{F}(\alpha S) = \mathcal{F}(\alpha) * \mathcal{F}(S)$ hold.*

Proposition E.6 (Tensor product of pure tensor, [76, p. 268])**.** *If $\phi_x \otimes \psi_y$ is a pure tensor in $\mathcal{S}'(\mathbb{R}^n) \otimes \mathcal{S}'(\mathbb{R}^m) = \mathcal{S}'(\mathbb{R}^n \times \mathbb{R}^m)$, then*

$$\mathcal{F}(\phi_x \otimes \psi_y) = \mathcal{F}(\phi_x) \otimes \mathcal{F}(\psi_y).$$

Appendix F

Measure theory

In this section, we just recall basic facts from measure theory, as stated in [39].

F.1 Radon-Nikodym derivative

Proposition F.1 (Lebesgue-Radon-Nikodym theorem, [39, Sec. 19.36]). *Let $(E, \mathfrak{B}, \mu)$ be a σ-finite measure space and let ν be a complex measure on $(E, \mathfrak{B})$ such that $\nu \ll \mu$. Then there exists a unique $f_0 \in L^1(E, \mathfrak{B}, \mu)$ such that*

1. $\int_E f \, \mathrm{d}\nu = \int_E f f_0 \, \mathrm{d}\mu$ *for all* $f \in L^1(E, \mathfrak{B}, |\nu|)$,
2. $\nu(B) = \int_B f_0 \, \mathrm{d}\mu$ *for all* $B \in \mathfrak{B}$, *and*
3. $|\nu|(B) = \int_B |f_0| \, \mathrm{d}\mu$ *for all* $B \in \mathfrak{B}$.

Remark F.2. $L^1(E, \mathfrak{B}, \mu)$ denotes the normed space of *equivalence classes* of integrable functions. In the applications, we always work with a representative of this class, but the results are independent of this choice.

Definition F.3. The essentially unique function f_0 appearing in Proposition F.1 is called the ***Radon-Nikodym derivative*** of ν with respect to μ. Usually $\frac{d\nu}{d\mu}$ is used to denote f_0. The relation between μ, ν and f_0 can also be stated with the formula $\nu = f_0 \nu$.

Proposition F.4 (Chain rule, [39, Sec. 19.40]). *Let μ_0, μ_1, and μ_2 be σ-finite measures on $(E, \mathfrak{B})$ such that $\mu_2 \ll \mu_1$ and $\mu_1 \ll \mu_0$. Then,*

1. $\mu_2 \ll \mu_0$, *and*
2. $\frac{d\mu_2}{d\mu_0} = \frac{d\mu_2}{d\mu_1} \cdot \frac{d\mu_1}{d\mu_0}$, μ_0*-almost everywhere.*

F.2 Product spaces

Let $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$ be σ-finite measure spaces. Introduce the product σ-algebra $\mathfrak{M} \otimes \mathfrak{N} := \sigma(\mathfrak{M} \times \mathfrak{N})$, the σ-algebra generated by the collection of sets

$$\mathfrak{M} \times \mathfrak{N} := \left\{ M \times N \,\middle|\, M \in \mathfrak{M} \text{ and } N \in \mathfrak{N} \right\} \subset 2^{X \times Y}.$$

The pair $(X \times Y, \mathfrak{M} \otimes \mathfrak{N})$ is a measurable space.

Proposition F.5 ([39, Sec. 21.5]). *Let $(X, \mathfrak{M})$ and $(Y, \mathfrak{N})$ be measurable spaces, and let f be an extended real-valued or complex-valued $\mathfrak{M} \otimes \mathfrak{N}$-measurable function on $X \times Y$. Then*

1. *the function $x \mapsto f(x, y)$ is $\mathfrak{M}$-measurable for all $y \in Y$, and*
2. *the function $y \mapsto f(x, y)$ is $\mathfrak{N}$-measurable for all $x \in X$.*

The product measure $\mu \otimes \nu : \mathfrak{M} \otimes \mathfrak{N} \to \mathbb{R}$ is the unique measure that satisfies $\mu \otimes \nu(M \times N) = \mu(M)\nu(N)$, see [39, Sec. 21.11].

Proposition F.6. *Let $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$ be σ-finite measure spaces and let $(X \times Y, \mathfrak{M} \otimes \mathfrak{N}, \mu \otimes \nu)$ be the product measure space. The measure $\mu \otimes \nu$ is σ-finite.*

Proof. Let $\{X_i\}_i \subset \mathfrak{M}$ (respectively, $\{Y_j\}_j \subset \mathfrak{N}$) be a countable collection of pairwise disjoint sets such that $\cup_i X_i = X$ and $\mu(X_i) < \infty$ for every i (resp. $\cup_j Y_j = Y$ and $\mu(Y_j) < \infty$ for every j). Then $\{X_i \times Y_j\}_{i,j}$ is a collection of pairwise disjoint, $\mathfrak{M} \otimes \mathfrak{N}$-measurable sets such that

$$\mu \otimes \nu(X_i \times Y_j) = \mu(X_i)\nu(Y_j) < \infty$$

for every pair (i, j), and

$$\bigcup_{i,j} X_i \times Y_j = X \times Y.$$

□

Proposition F.7 (Fubini's theorem for positive functions, [39, Sec. 21.12]). *Let $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$ be σ-finite measure spaces and let $(X \times Y, \mathfrak{M} \otimes \mathfrak{N}, \mu \otimes \nu)$ be the product measure space. If f is a nonnegative, extended real-valued, $\mathfrak{M} \otimes \mathfrak{N}$-measurable function on $X \times Y$, then*

1. *the function $x \mapsto f(x, y)$ is $\mathfrak{M}$-measurable for each $y \in Y$;*
2. *the function $y \mapsto f(x, y)$ is $\mathfrak{N}$-measurable for each $x \in X$;*
3. *the function $x \mapsto \int_Y f(x, y)\, \mathrm{d}\nu(y)$ is $\mathfrak{M}$-measurable;*
4. *the function $y \mapsto \int_X f(x, y)\, \mathrm{d}\mu(x)$ is $\mathfrak{N}$-measurable; and*
5. *the following equalities hold*

$$\begin{aligned}\int_{X\times Y} f(x, y)\, \mathrm{d}\mu \otimes \nu(x, y) &= \int_Y \int_X f(x, y)\, \mathrm{d}\mu(x)\, \mathrm{d}\nu(y) \\ &= \int_X \int_Y f(x, y)\, \mathrm{d}\nu(y)\, \mathrm{d}\mu(x).\end{aligned}$$

In virtue of Proposition F.7-(5), the finiteness of

$$\int_{X\times Y} |f(x, y)|\, \mathrm{d}\mu\otimes\nu(x, y), \quad \int_Y \int_X |f(x, y)|\, \mathrm{d}\mu(x)\, \mathrm{d}\nu(y) \text{ or } \int_X \int_Y |f(x, y)|\, \mathrm{d}\nu(y)\, \mathrm{d}\mu(x)$$

for an arbitrary measurable function f implies that the three integrals are finite. This entails a more refined result for integrable functions

Proposition F.8 (Fubini's theorem for integrable functions, [39, Sec. 21.13]). *Let $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$ be σ-finite measure spaces and let $(X \times Y, \mathfrak{M} \otimes \mathfrak{N}, \mu \otimes \nu)$ be the product measure space. Let f be a complex-valued $\mathfrak{M} \otimes \mathfrak{N}$-measurable function on $X \times Y$ and suppose that at least one of the three integrals*

$$\int_{X\times Y} |f(x, y)|\, \mathrm{d}\mu\otimes\nu(x, y), \quad \int_Y \int_X |f(x, y)|\, \mathrm{d}\mu(x)\, \mathrm{d}\nu(y), \quad \int_X \int_Y |f(x, y)|\, \mathrm{d}\nu(y)\, \mathrm{d}\mu(x)$$

is finite. Then:

1. *the function $x \mapsto f(x, y)$ is in $L^1(X, \mathfrak{M}, \mu)$ for ν-almost all $y \in Y$;*
2. *the function $y \mapsto f(x, y)$ is in $L^1(Y, \mathfrak{N}, \nu)$ for μ-almost all $x \in X$;*
3. *the function $x \mapsto \int_Y f(x, y)\, \mathrm{d}\nu(y)$ is well-defined μ-almost surely and belongs to $L^1(X, \mathfrak{M}, \mu)$;*
4. *the function $y \mapsto \int_X f(x, y)\, \mathrm{d}\mu(x)$ is well-defined ν-almost surely and belongs to $L^1(Y, \mathfrak{N}, \nu)$;*
5. *the following equalities hold*

$$\begin{aligned}\int_{X\times Y} f(x, y)\, \mathrm{d}\mu \otimes \nu(x, y) &= \int_Y \int_X f(x, y)\, \mathrm{d}\mu(x)\, \mathrm{d}\nu(y)\\ &= \int_X \int_Y f(x, y)\, \mathrm{d}\nu(y)\, \mathrm{d}\mu(x).\end{aligned}$$

F.3 Haar measures

Given a group locally compact topological group G, there is a unique left-invariant positive measure (Haar measure) up to a multiplicative constant [19, Thms. 9.2.2 & 9.2.6]. A particular choice of left Haar measure will be denoted by a Greek letter with subscript G e.g. λ_G.

Proposition F.9 (Weil's formula)**.** *Let G be a locally compact group and H a closed normal subgroup of G. Given Haar measures on two groups among G, H and G/H, there is a Haar measure on the third one such that, for any integrable function $f : G \to \mathbb{R}$,*

$$\int_G f(x) d\lambda_G(x) = \int_{G/H} \left(\int_H f(xy)\, \mathrm{d}\lambda_H(y) \right) \mathrm{d}\nu(xH). \tag{F.1}$$

The three measures are said to be in *canonical relation*, which is written $\lambda_G = \lambda_{G/H}\lambda_H$.

For a proof of Proposition F.9, see [72] p.87-88 and Theorem 3.4.6.

Bibliography

[1] Abramsky, S., Barbosa, R. S., Kishida, K., Lal, R., and Mansfield, S. Contextuality, Cohomology and Paradox. In *24th EACSL Annual Conference on Computer Science Logic (CSL 2015)* (Dagstuhl, Germany, 2015), S. Kreutzer, Ed., vol. 41 of *Leibniz International Proceedings in Informatics (LIPIcs)*, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, pp. 211–228.

[2] Abramsky, S., and Hardy, L. Logical Bell inequalities. *Phys. Rev. A 85* (Jun 2012), 062114.

[3] Aczél, J., and Daróczy, Z. *On Measures of Information and Their Characterizations*. Mathematics in Science and Engineering. Academic Press, 1975.

[4] Artin, M., Grothendieck, A., and Verdier, J.-L. *Théorie des Topos et Cohomologie Étale des Schémas: Séminaire de Géométrie Algébrique du Bois-Marie 1963/64 - SGA 4. Tome 1*. Lecture notes in mathematics. Springer-Verlag, 1972.

[5] Artin, M., Grothendieck, A., and Verdier, J.-L. *Théorie des Topos et Cohomologie Étale des Schémas: Séminaire de Géométrie Algébrique du Bois-Marie 1963/64 - SGA 4. Tome 2*. Lecture notes in mathematics. Springer-Verlag, 1972.

[6] Askey, R. The q-Gamma and q-Beta functions. *Applicable Analysis 8*, 2 (1978), 125–141.

[7] Ay, N., Jost, J., Le, H. V., and Schwachhöfer, L. *Information Geometry*, vol. 64 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics*. Springer International Publishing, 2017.

[8] Baez, J. C., and Fritz, T. A bayesian characterization of relative entropy. *Theory and Applications of Categories 29*, 16 (2014), 422–456.

[9] Barron, A. R. Entropy and the central limit theorem. *The Annals of probability* (1986), 336–342.

[10] Baudot, P., and Bennequin, D. The homological nature of entropy. *Entropy 17*, 5 (2015), 3253–3318.

[11] Beck, C., and Schögl, F. *Thermodynamics of Chaotic Systems: An Introduction*. Cambridge Nonlinear Science Series. Cambridge University Press, 1995.

[12] Bloch, S., and Esnault, H. The additive dilogarithm. *Kazuya Kato* (2003), 131.

[13] Bourbaki, N. *Algebra I. Chapters 1-3*. Springer-Verlag, 1998.

[14] Brualdi, R. A., and Schneider, H. Determinantal identities: Gauss, Schur, Cauchy, Sylvester, Kronecker, Jacobi, Binet, Laplace, Muir, and Cayley. *Linear Algebra and its applications 52* (1983), 769–791.

[15] Caramello, O. *Theories, Sites, Toposes: Relating and Studying Mathematical Theories Through Topos-theoretic 'bridges'*. Oxford University Press, 2016.

[16] Cathelineau, J. Sur l'homologie de SL_2 à coefficients dans l'action adjointe. *Mathematica Scandinavica* (1988), 51–86.

[17] Cathelineau, J.-L. Remarques sur les différentielles des polylogarithmes uniformes. In *Annales de l'institut Fourier* (1996), vol. 46, pp. 1327–1347.

[18] Chang, J. T., and Pollard, D. Conditioning as disintegration. *Statistica Neerlandica 51*, 3 (1997), 287–317.

[19] Cohn, D. *Measure Theory: Second Edition*. Birkhäuser Advanced Texts Basler Lehrbücher. Springer New York, 2013.

[20] Cover, T., and Thomas, J. *Elements of Information Theory*. A Wiley-Interscience publication. Wiley, 2006.

[21] Csiszár, I. Generalized entropy and quantization problems. In *Transactions of the 6th Praque Conference on Information Theory, Statistical Decision Functions, Random Processes (Academia, Prague)* (1973).

[22] Csiszár, I. Axiomatic characterizations of information measures. *Entropy 10*, 3 (2008), 261–273.

[23] Csiszár, I., and Körner, J. *Information theory: Coding theorems for discrete memoryless systems*. Probability and mathematical statistics. Academic Press, 1981.

[24] Curry, J. *Sheaves, cosheaves and applications*. PhD thesis, The University of Pennsylvania, 2013. arXiv:1303.3255.

[25] Daróczy, Z. Generalized information functions. *Information and control 16*, 1 (1970), 36–51.

[26] Doob, J. *Stochastic Processes*. Wiley Publications in Statistics. John Wiley & Sons, 1953.

[27] Elbaz-Vincent, P., and Gangl, H. On poly(ana)logs I. *Compositio Mathematica 130*, 2 (2002), 161–214.

[28] Elbaz-Vincent, P., and Gangl, H. Finite polylogarithms, their multiple analogues and the shannon entropy. In *International Conference on Geometric Science of Information* (2015), Springer, pp. 277–285.

[29] Fontené, G. Généralisation d'une formule connue. *Nouvelles annales de mathématiques : journal des candidats aux écoles polytechnique et normale 15* (1915), 112–112.

[30] Fritz, T., and Chaves, R. Entropic inequalities and marginal problems. *IEEE transactions on information theory 59*, 2 (2013), 803–817.

[31] Gabriel, P. Objets injectifs dans les catégories abéliennes. *Séminaire Dubreil. Algèbre et théorie des nombres 12*, 2 (1958-1959), 1–32.

[32] Giraud, J. *Cohomologie non abélienne*. Grundlehren der mathematischen Wissenschaften. Springer-Verlag, Germany, 1971.

[33] Gould, H. The bracket function and Fontené-Ward generalized binomial coefficients with application to Fibonomial coefficients. *Fibonacci Quart 7* (1969), 23–40.

[34] Gromov, M. In a search for a structure, part 1: On entropy. *Preprint available at http://www. ihes. fr/gromov* (2012).

[35] Grothendieck, A. Sur quelques points d'algèbre homologique, I. *Tohoku Mathematical Journal 9*, 2 (1957), 119–221.

[36] Halmos, P. R. *Finite-Dimensional Vector Spaces*, reprint of 2nd ed. Undergraduate Texts in Mathematics. Springer-Verlag, New York, USA, 1958.

[37] Hanel, R., and Thurner, S. When do generalized entropies apply? How phase space volume determines entropy. *EPL (Europhysics Letters) 96*, 5 (2011), 50003.

[38] Havrda, J., and Charvát, F. Quantification method of classification processes. Concept of structural a-entropy. *Kybernetika 3*, 1 (1967), 30–35.

[39] Hewitt, E., and Stromberg, K. *Real and Abstract Analysis: A modern treatment of the theory of functions of a real variable*. Springer Berlin Heidelberg, 1965.

[40] Holevo, A. S. *Probabilistic and Statistical Aspects of Quantum Theory; 2nd ed.* Publications of the Scuola Normale Superiore Monographs. Springer, Dordrecht, 2011.

[41] Jaynes, E., Jaynes, E., Bretthorst, G., and Press, C. U. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

[42] Jaynes, E. T. Information theory and statistical mechanics. *Physical Review 106*, 4 (1957), 620.

[43] Jaynes, E. T. Gibbs vs Boltzmann entropies. *American Journal of Physics 33*, 5 (1965), 391–398.

[44] Jensen, H. J., Pazuki, R. H., Pruessner, G., and Tempesta, P. Statistical mechanics of exploding phase spaces: Ontic open systems. *Journal of Physics A: Mathematical and Theoretical 51*, 37 (2018), 375002.

[45] Kac, V., and Cheung, P. *Quantum Calculus*. Universitext. Springer New York, 2001.

[46] Kammler, D. *A First Course in Fourier Analysis*. Cambridge University Press, 2007.

[47] Kannappan, P., and Ng, C. T. Measurable solutions of functional equations related to information theory. *Proceedings of the American Mathematical Society 38*, 2 (1973), pp. 303–310.

[48] Kemp, A., and Newton, J. Certain state-dependent processes for dichotomised parasite populations. *Journal of Applied Probability 27*, 2 (1990), 251–258.

[49] Khinchin, A. *Mathematical Foundations of Information Theory*. Dover Books on Mathematics Series. Dover Publications, 1957.

[50] Knopp, K. *Infinite Sequences and Series*. Dover Books on Mathematics. Dover Publications, 1956.

[51] Koliander, G., Pichler, G., Riegler, E., and Hlawatsch, F. Entropy and source coding for integer-dimensional singular random variables. *IEEE Transactions on Information Theory 62*, 11 (Nov. 2016), 6124–6154.

[52] Kolmogorov, A., and Shiryayev, A. *Selected Works of A. N. Kolmogorov. Volume III: Information Theory and the Theory of Algorithms*. Mathematics and Its Applications. Kluwer Academic Publishers, 1993.

[53] Kolmogorov, A. N. *Foundations of the Theory of Probability*, 2 ed. Chelsea Publishing Company, New York, 1956.

[54] Kousidis, S. Asymptotics of generalized Galois numbers via affine Kac-Moody algebras. *Proceedings of the American Mathematical Society 141*, 10 (2013), 3313–3326.

[55] Lawvere, F. W., and Schanuel, S. H. *Conceptual Mathematics: A First Introduction to Categories*, 2nd ed. Cambridge University Press, New York, USA, 2009.

[56] Lee, J. *Introduction to Topological Manifolds*. Graduate Texts in Mathematics. Springer New York, 2010.

[57] Lee, P. M. On the axioms of information theory. *The Annals of Mathematical Statistics 35*, 1 (03 1964), 415–418.

[58] Leinster, T. The categorical origins of entropy. Presentation at TGSI2017, Luminy. Available at url: `https://www.maths.ed.ac.uk/~tl/luminy/luminy_talk.pdf`, 2017.

[59] Linnik, J. V. An information-theoretic proof of the central limit theorem with Lindeberg conditions. *Theory of Probability & Its Applications 4*, 3 (1959), 288–299.

[60] Mac Lane, S. *Homology*, reprint of the 1975 ed. Springer-Verlag, Germany, 1994.

[61] Mac Lane, S., and Moerdijk, I. *Sheaves in Geometry and Logic: A First Introduction to Topos Theory*. Mathematical Sciences Research Institute Publications. Springer-Verlag, New York, USA, 1992.

[62] Matsuda, H. Information theoretic characterization of frustrated systems. *Physica A: Statistical Mechanics and its Applications 294*, 1-2 (2001), 180–190.

[63] Mézard, M., and Montanari, A. *Information, Physics, and Computation*. Oxford Graduate Texts. OUP Oxford, 2009.

[64] Moak, D. S. The q-analogue of Stirling's formula. *The Rocky Mountain Journal of Mathematics* (1984), 403–413.

[65] Moerdijk, I. *Classifying Spaces and Classifying Topoi*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2006.

[66] Morgan, F. *Geometric Measure Theory: A Beginner's Guide*. Elsevier Science, 2008.

[67] Pelizzola, A. Cluster variation method in statistical physics and probabilistic graphical models. *Journal of Physics A: Mathematical and General 38*, 33 (2005), R309.

[68] Prasad, A. Counting subspaces of a finite vector space — 1. *Resonance 15*, 11 (Nov 2010), 977–987.

[69] Prasad, A. Counting subspaces of a finite vector space — 2. *Resonance 15*, 12 (Dec 2010), 1074–1083.

[70] Rao, C. *Linear Statistical Inference and its Applications*. Wiley Series in Probability and Statistics. Wiley, 2009.

[71] Rao, M. *Conditional Measures and Applications*. Monographs and textbooks in pure and applied mathematics. CRC Press, 2005.

[72] Reiter, H., and Stegeman, J. D. *Classical Harmonic Analysis and Locally Compact Groups*. No. 22 in London Mathematical Society monographs, new. ser. Clarendon Press, 2000.

[73] Ruiz, G., and Tsallis, C. Emergence of q-statistical functions in a generalized binomial distribution with strong correlations. *Journal of Mathematical Physics 56*, 5 (2015), 53301.

[74] Schapira, P. Algebra and topology. *Lecture Notes from Paris VI* (2008).

[75] Schur, J. Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind. *Journal für die reine und angewandte Mathematik 147* (1917), 205–232.

[76] Schwartz, L. *Théorie des distributions*. No. v. 1-2 in Actualités scientifiques et industrielles. Hermann, 1966.

[77] Serre, J. *A Course in Arithmetic*. Graduate texts in mathematics. Springer, 1973.

[78] Shannon, C. A mathematical theory of communication. *Bell System Technical Journal 27* (1948), 379–423, 623–656.

[79] Sharp, K., and Matschinsky, F. Translation of Ludwig Boltzmann's Paper "On the Relationship between the Second Fundamental Theorem of the Mechanical Theory of Heat and Probability Calculations Regarding the Conditions for Thermal Equilibrium". *Entropy 17*, 4 (2015), 1971–2009.

[80] Sharpe, R. *Differential Geometry: Cartan's Generalization of Klein's Erlangen Program*. Graduate Texts in Mathematics. Springer New York, 1997.

[81] Stacks Project Authors. *Stacks Project*.
Available at http://stacks.math.columbia.edu/.

[82] Steinbauer, G. H. . R. Lecture notes on the theory of distributions, 2009.

[83] Tamme, G. *Introduction to Étale Cohomology*. Universitext. Springer-Verlag Berlin Heidelberg, Germany, 1994.

[84] Tao, T. *Topics in Random Matrix Theory*. American Mathematical Society, 2012.

[85] Tel, T. Fractals, multifractals, and thermodynamics. *Zeitschrift für Naturforschung A 43*, 12 (1988), 1154–1174.

[86] Treves, F. *Topological Vector Spaces, Distributions and Kernels*. Academic Press, 1967.

[87] Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *Journal of statistical physics 52*, 1-2 (1988), 479–487.

[88] Tsallis, C. *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*. Springer New York, 2009.

[89] Tverberg, H. A new derivation of the information function. *Mathematica Scandinavica 6* (1958), 297–298.

[90] Čensov, N. N. *Statistical Decision Rules and Optimal Inference*, vol. 53 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, Rhode Island, 1982.

[91] Vigneaux, J. P. Generalized information structures and their cohomology. *arXiv e-prints* (2017). arXiv:1709.07807 [cs.IT].

[92] Vigneaux, J. P. Information theory with finite vector spaces. *IEEE Transactions on Information Theory 65*, 9 (Sep. 2019), 5674–5687.

[93] Vigneaux, J. P. *Topology of Statistical Systems: A Cohomological Approach to Information Theory*. PhD thesis, Université Paris Diderot, 2019.

[94] Vontobel, P. O. Counting in graph covers: A combinatorial characterization of the bethe entropy function. *IEEE Transactions on Information Theory 59*, 9 (2013), 6018–6048.

[95] Weibel, C. A. *An introduction to homological algebra*. Cambridge University Press, Printed in USA, 1994.

[96] Williams, D. *Probability With Martingales*. Cambridge University Press, 1991.

Index