Behavior of a sample under extreme conditioning, maximum likelihood under weighted sampling

Zhansheng Cao

Université Pierre et Marie Curie

École Doctorale Paris Centre

Thèse de doctorat

Spécialité : Mathématiques

Option : Statistique

Présentée par

Zhansheng Cao

pour obtenir le grade de

Docteur de l'université Paris 6

Comportement d'un échantillon sous conditionnement extrême, maximum de vraisemblance sous échantillonnage pondéré

Soutenue le 26 novembre 2012 devant le jury composé de :

M. Michel BRONIATOWSKI	Université Pierre et Marie Curie (Paris 6)	Directeur
M. Benoît CADRE	Ecole Normale Supérieure de Cachan	Rapporteur
M. Paul DEHEUVELS	Université Pierre et Marie Curie (Paris 6)	Examinateur
M. Nikolaos LIMNIOS	Université de Technologie de Compiègne	Examinateur
M. Dietmar PFEIFER	Universität Oldenburg	Rapporteur
M. Mathieu ROSENBAUM	Université Pierre et Marie Curie (Paris 6)	Examinateur

à mes parents
à Ting

三人行，必有我师。– 孔子
"Trois personnes marchant côte à côte,
mon enseignant doit être parmi eux."
Confucius

吾生也有涯，而知也无涯。– 庄子
"Ma vie est limitée, tandis que la
connaissance est illimitée." ZhuangZi

学而不思则罔，思而不学则殆。– 孔子
"L'étude sans raisonnement mène à la
confusion ; la pensée sans apprentissage
est effort gaspillé." Confucius

Remerciements

Je tiens avant tout à exprimer ma profonde reconnaissance au Professeur Michel Broniatowski, mon directeur de thèse, pour ses conseils pertinents, pour la confiance dont il a fait preuve envers moi et ses encouragements tout au long de la thèse ainsi que ses suggestions et lectures. Je lui adresse un merci sincère pour tout ce qu'il m'apporté.

Je suis ensuite très reconnaissant envers mes rapporteurs de thèse : Benoît Cadre et Dietmar Pfeifer qui ont accepté ce travail et avoir pris le temps de le lire avec attention, dont je tiens à souligner la gentillesse ainsi que la qualité du rapport rendu. Le contenu de ce rapport constituerait sans aucun doute une bien meilleure introduction à ce manuscrit que celle que j'ai écrite.

Je remercie aussi les autres membres du jury : Paul Deheuvels, Nikolaos Limnios et Mathieu Rosenbaum qui ont accepté de venir écouter ma soutenance et juger mon travail.

Merci également à tous les doctorants du LSTA que j'ai pu croiser au cours de ma thèse, et tout spécialement Virgile Caron, Sarah Ouadah, Emmanuel Onzon, Abdullah Oueslati, Assia Boumahdaf, Batiste Gregorutti, Matthieu Clertant pour leur gentillesse et leur aide en français.

Résumé

Résumé

Dans le Chapitre 1, nous explorons le comportement joint des variables d'une marche aléatoire $(X_1, \ldots, X_n)$ lorsque leur valeur moyenne tend vers l'infini quand $n \to \infty$. Il est prouvé que toutes ces variables doivent partager la même valeur, ce qui généralise les résultats précédents, dans le cadre de grands dépassements de sommes finies de i.i.d variables aléatoires.

Dans le Chapitre 2, nous montrons un théorème de Gibbs conditionnel pour une marche aléatoire $(X_1, .., X_n)$ conditionnée à une déviation extrême de sa somme $(S_n = na_n)$ ou $(S_n > na_n)$ où $a_n \to \infty$. Il est prouvé que lorsque les opérandes ont des queues légères avec une certaine régularité supplémentaire, la distribution asymptotique conditionnelle de X_1 peut être approximée par la distribution tiltée au point a_n en norme de la variation totale, généralisant ainsi le cas classique du LDP.

Le troisième Chapitre explore le principe du maximum de vraisemblance dans les modèles paramétriques, dans le contexte du théorème de grandes déviations de Sanov. Le MLE est associé à la minimisation d'un critère spécifique de type divergence, qui se généralise au cas du bootstrap pondéré, où la divergnce est fonction de la distribution des poids. Certaines propriétés de la procédure résultante d'inférence sont présenteés ; l'efficacité de Bahadur de tests est également examinée dans ce contexte.

Mots-clefs

Déviation extrême, Principe de Gibbs, Développements d'Edgeworth, Divergence, Bootstrap pondéré

Abstract

In Chapter one, we explore the joint behaviour of the summands of a random walk when their mean value goes to infinity as its length increases. It is proved that

all the summands must share the same value, which extends previous results in the context of large exceedances of finite sums of i.i.d. random variables.

In Chapter two, we state a conditional Gibbs theorem for a random walk $(X_1, .., X_n)$ conditioned on an extreme deviation of its sum $(S_n = na_n)$ or $(S_n > na_n)$ where $a_n \to \infty$. It is proved that when the summands have light tails with some additional regulatity property, then the asymptotic conditional distribution of X_1 can be approximated by the tilted distribution at point a_n in variation norm, extending therefore the classical LDP case.

The third Chapter explores Maximum Likelihood in parametric models in the context of Sanov type Large Deviation Probabilities. MLE in parametric models under weighted sampling is shown to be associated with the minimization of a specific divergence criterion defined with respect to the distribution of the weights. Some properties of the resulting inferential procedure are presented; Bahadur efficiency of tests is also considered in this context.

Keywords

Extreme deviation, Gibbs principle, Edgeworth expansion, Divergence, Weighted bootstrap

Table des matières

Introduction

What portion of a sample makes a statistics large ?

What is done in this thesis

Consider two independent random variables (r.v's) with common standard normal distribution on $\mathbb{R}$. The r.v's $U := (X_1 + X_2)/2$ and $V := (X_1 - X_2)/2$ are independent, centered and normally distributed. Denote $T := X_1 + X_2$. Then the distribution of $(X_1/a, X_2/a)$ given $T \geq 2a$ is that of $(U/a, U/a) + (V/a, -V/a)$ given $U \geq a$. By independence of U and V the pair $(V/a, -V/a)$ goes to $(0,0)$ as t goes to infinity. Since U is gaussian, U/a converges to 1 as a tends to infinity. Thus, conditionally on $X_1 + X_2 \geq 2a$ the pair $(X_1/a, X_2/a)$ converges to $(1,1)$ as a tends to infinity.

This example shows that *for fixed* n, in some cases, the conditional distribution of $X_1^n := (X_1, ..., X_n)$ given $(T_n := X_1 + ... + X_n \geq na)$ concentrates on $(a, ..., a)$, the point in $\mathbb{R}^n$ with all coordinates equal to a as a tends to infinity. This fact has been considered in much greater generality in [4] for many kinds of statistics ; this question is of interest in statistics. In the case of a simple statistics as the sum of the sample points, it leads to the well known typology of distributions in sub and over exponential distributions. The extension of this work from fixed sample size asymptotics to extreme deviations is the starting point of the first part of this thesis.

The question to be addressed in this work can be written somewhat as follows :

We assume that the generic random variable is non negative and has a light unbounded tail, namely that its moment generating function is finite in a non void neighborhood of 0 and $\inf \{x : P(X_1 > x) = 0\} = +\infty$. We consider the statistics T_n defined above.

Assuming that for any fixed n it holds

$$\lim_{a \to \infty} P\left(X_1^n \in aB_n \middle| T_n \geq na \right) = 0 \tag{1}$$

for any Borel set B_n in $\mathbb{R}^n$ such that $(1, ..., 1) \notin B_n$.

Fix a sequence B_n and define a_n such that for any $s > a_n$

$$\sup_{s > a_n} P\left(X_1^n \in sB_n \middle| T_n \geq ns \right) \leq 1/n.$$

Such a_n surely exists, and is of interest. Obviously we could have defined a_n through some other upper bound for the above probability, and the sequence a_n depends on the sequence of sets B_n. The qualitative question raised by the existence of such a_n's is the asymptotic behaviour of the whole sample X_1^n as n increases when conditioning upon very rare events of the form $(T_n \geq na_n)$. In other words, our question is :

For which classes of distributions and for which order of magnitude of the conditioning barrier a_n do we have

$$\lim_{n\to\infty} P\left(\cap X_i \in (a_n - \varepsilon_n, a_n + \varepsilon_n) \middle| T_n \geq na_n\right) = 1.$$

In the above display the sequence ε_n is also a part of the debate : can we have

$$\lim_{n\to\infty} \varepsilon_n = 0$$

and, if yes, at which rate ?

Such questions are also of interest in many branches of physics ; the illuminating book by Sornette [75] and the paper by Frisch and Sornette [38] handle the notion of so-called "democratic localization" of a sample, which correspond precisely to (1) for fixed n and $a \to \infty$. These authors explore the consequences of (1) in fragmentation processes and in turbulence, among others ; we do not enter these considerations, out of the field of our knowledge.

Fixed sample size asymptotics have been considered by many authors ; the book by Field and Ronchetti [37] can be considered as a milestone in the area, and has given rise to the interest for these question in statistical robustness analysis ; we may quote however that the topics considered in these approaches is anyhow different from the one usually covered by robustness concepts, since small sample asymptotics have to do with sampling under the assumed sampling scheme, and has nothing to do with misspecification or outliers. The same approach is handled here. Other related works, pertaining to fixed sample size asymptotics, include Jurečková ([51], [52] and [53]), Kušnier and Mizera [57], and Broniatowski and Fuchs [17] among others.

The case when $a = a_n$ grows together with n is clearly related to the extreme deviation approach, extending therefore the Large deviation case to very rare events. The papers by Broniatowski and Mason [21] and Nagaev [54] handle these topics.

The main result of the first chapter of this thesis is a characterization of sequences a_n and ε_n according to the form of the density of X_1. By its very nature this result is strongly dependent upon regularity conditions on the upper tail of the density. These conditions are best stated in terms of notions imported from asymptotic analysis, as developed in extreme value theory, with which it bears many similarities. However we did not establish a precise link with the theory of extreme order statistics ; this link certainly exists, as could be argumented looking at Erdös-Rényi laws for the largest local slopes of a random walk. An attempt in this direction is presented shortly in the first chapter, when considering Erdös-Rényi laws for random walks conditioned on a large deviation event pertaining to its sum. The order of magnitude of the sequence a_n for which the sequence ε_n obeys $\lim_{n\to\infty} \varepsilon_n = 0$ is quite large.

Apart from arguments directly related to the proof of the result, it may also be the case that the requirement that all the summands X_i's share the same behaviour is quite strong. Besides its probabilistic content, this chapter bears a statistical question which has been a main motivation for this work ; somehow unexpectedly, robustness and loyalty concepts do not coexist, at least when it turns to statistics. This is the essence of a deep result by He, Jureskova, Koenker and Portnoy [46], although not phrased as such in their paper ; we explore their statement and provide some extension, according to our results, in the following paragraphs.

Breakdown point, loyalty and robustness

For *fixed* n, let $X_1, ..., X_n$ be independently random variables distributed with a common distribution $F(x)$, and which have a symmetric density $f(-x) = f(x)$ with $x \in \mathbb{R}$. Jereckova [52] shows the following result

Theorem 1. *Let* $T_n = X_1 + ... + X_n$*, if for some* $b > 0$ *and* $r \geq 1$

$$\lim_{a \to \infty} \frac{-\ln(1 - F(a))}{ba^r} = 1 \tag{A}$$

then

$$\lim_{a \to \infty} \frac{\ln P(|T_n| > na)}{\ln P(|X_1| > a)^n} = 1.$$

The *breakdown point* as a useful concept in robust statistics has been introduced since 1970. Hodges [47] gave its oldest definition, Hampel [44] provided a much more general but rather mathematical formulation, Donoho and Huber [33] introduced a simple finite sample version of this concept. Let $X = (x_1, ..., x_n)$ denote a finite sample of size n; we can corrupt this sample by replacing an arbitrary subset of size m of X with arbitrary values. The proportion of such "bad values" in the new sample X' is $\epsilon = m/n$. The breakdown point ϵ^* is defined as the least ϵ such that

$$\epsilon^*(X, T_n) = \inf \left\{ \epsilon \middle| \sup |T_n(X)/n - T_n(X')/n| = \infty \right\}.$$

Of course, here $T_n(X)$ could be some other statistics, but we are only interested in the sample sum. Let m^* denote the least m such that ϵ^* attains its minimum, He, Jureskova, Koenker and Portnoy [46] showed

Theorem 2. *Suppose that for any fixed* $c > 0$*, it holds*

$$\lim_{x \to \infty} \frac{-\ln(1 - F(x + c))}{-\ln(1 - F(x))} = 1, \tag{B}$$

then

$$\lim_{a \to \infty} \frac{\ln P(|T_n| > na)}{\ln P(|X_1| > a)} = n - m^* + 1.$$

Note firstly that (B) is commonly met; for example all distributions with Weibull tails satisfy (B).

Theorem is of great interest since it proves that the lack of robustness of T_n , as measured by m^*, has a strong impact on the weight of the tail behaviour of the X_i's in the large values of T_n. When $m^* = 1$, as in the case when T_n is the sample mean, then $\ln P(|T_n| > na) = \ln P^n(|X_1| > a)\,(1 + o(1))$ as $a \to \infty$, which is to say that T_n would take very large and abnormal values only when all the summands would (see the proof of Theorem 4 for an explanation of this fact). In some sense the estimator of the mean is "loyal" with respect to the sampling, but is very "weak" under contamination, which is the meaning of its lack of robustness.

At the contrary, consider a very "robust" statistics of location, the median. It is well known, and easy to verify, that $m^* = n/2$. Hence $n - m^* + 1$ is minimal, which proves that robust estimators are unfaithfull in alleagence under the sampling.

This is indeed the starting point of our research. In this thesis, we only consider the case of the empirical mean, for which $m^* = 1$ whenever n is fixed or n tends to infinity.

Evidently, for *fixed* n, we have $\epsilon^* = 1/n$ and $m^* = 1$. Hence Theorem 2 comes back to Theorem 1. What interests us is : does Theorem 2 still holds if n goes to ∞ and a depends on n? When n tends to ∞, we prefer the following definition of the *breakdown point*

Definition 3. *Let $X = (x_1, ..., x_n)$ be a* infinite *sample of size n, and corrupt this sample by replacing an arbitrary subset of size m of X with arbitrary values. Denote by X' the corrupted sample. The breakdown point ϵ^* is defined as the least ϵ such that*

$$\epsilon^*(X, T_n) = \inf\left\{\epsilon \middle| \limsup_{n\to\infty} |T_n(X')/T_n(X)| = \infty\right\}.$$

By this definition, it is straightforward that $m^* = 1$. As a result of this thesis we obtain the following expansion of Theorem 2

Theorem 4. *Let $X_1, ..., X_n$ be i.i.d. real valued random variables with common density $f(x) = c\exp(-(g(x) + q(x)))$, where $g(x)$ is some positive convex function on $\mathbb{R}^+$ and g is twice differentiable. Assume that $g(x)$ is increasing on some interval $[X, \infty)$ and satisfies*

$$\lim_{x\to\infty} g(x)/x = \infty.$$

Let $M(x)$ be some nonnegative continuous function on $\mathbb{R}^+$ for which

$$-M(x) \leq q(x) \leq M(x) \quad \text{for all positive } x$$

together with

$$M(x) = O(\log g(x))$$

as $x \to \infty$.

Let a_n be some positive sequence such that $a_n \to \infty$ and $\epsilon_n = o(a_n)$ be a positive sequence. Assume

$$\liminf_{n\to\infty} \frac{\log g(a_n)}{\log n} > 0 \tag{0.0.1}$$

$$\lim_{n\to\infty} \frac{n \log g(a_n + \epsilon_n)}{H(a_n, \epsilon_n)} = 0, \tag{0.0.2}$$

$$\lim_{n\to\infty} \frac{nG(a_n)}{H(a_n, \epsilon_n)} = 0, \tag{0.0.3}$$

where

$$H(a_n, \epsilon_n) = \min(F_{g_1}(a_n, \epsilon_n), F_{g_2}(a_n, \epsilon_n)) - ng(a_n),$$
$$G(a_n) = g\left(a_n + \frac{1}{g(a_n)}\right) - g(a_n),$$

$F_{g_1}(a_n, \epsilon_n)$ and $F_{g_2}(a_n, \epsilon_n)$ are defined as in Lemma 1.3.1.

Then

$$\lim_{n\to\infty} \frac{\ln P(|T_n| > na_n)}{\ln P(|X_1| > a_n - \epsilon_n)^n} = 1.$$

Proof :

$$\begin{aligned}&P\left(\cap X_i\in(a_n-\varepsilon_n,a_n+\varepsilon_n)\right|T_n\geq na_n)\\&\leq P\left(\cap X_i\geq a_n-\varepsilon_n\right|T_n\geq na_n)\\&\leq\frac{P(\cap X_i\geq a_n-\varepsilon_n)}{P(T_n\geq na_n)}=\frac{P(X_1\geq a_n-\varepsilon_n)^n}{P(T_n\geq na_n)},\end{aligned}$$

by Theorem 1.4.1, we have

$$\lim_{n\to\infty}\frac{P(|T_n|>na_n)}{P(|X_1|>a_n-\epsilon_n)^n}=1.$$

Hence the proof.

Evidently, condition (B) consists of a larger family of density functions than condition (A). For example, if we set $f(x)=c\exp(-g(x))$, then $g(x)$ can be any *regularly varying* function such that condition (B) is satisfied. But it is not the case for condition (A).

For applying the results of Theorems 1 and 2 to infinite sample and making a depend on n, we have to impose further conditions on $f(x)$. Roughly speaking, if $g(x)$ has the form of a power function, its exponent should be strictly larger than 1. However, our result applies to *rapidly varying functions* which do not satisfy condition (B), namely, $g(x)$ tends to ∞ faster than any power function.

Finally, for simplifying the proof, we suppose $x>0$, but it is straightforward that our results hold still for symmetric density $f(x)=f(-x)$ with $x\in\mathbb{R}$.

Extended Gibbs Principle under extreme deviation

The second chapter considers a Gibbs type result, namely a conditional limit theorem under extreme deviation in the standard i.i.d. setting. The aim of this chapter can be stated as follows :

We consider a point conditioning event of the form $(T_n=na_n)$ where the sequence a_n tends to infinity. The chapter provides an approximation of the distribution of X_1 given $(T_n=na_n)$ and some extension for the case when the conditioning event is "thick", namely $(T_n\geq na_n)$. Compared with the first chapter, it mainly quotes that such sequences a_n grow to infinity in a much more moderate way. Note however that the resulting behaviour pertaining to the summands is much weaker, since it says nothing about the joint distribution of the X_i's.

The classical Gibbs conditioning result in its local form can be stated as follows : Under $(T_n=na)$ for fixed $a\neq EX_1$ and when X_1 satisfies a Cramer type condition

$$\begin{aligned}&\phi(t):=E\exp tX_1<\infty\\&\qquad\text{for } t \text{ in a non void neighborhood of } 0\end{aligned}$$

then

$$\lim_{n\to\infty}\|P_{n,a}-\Pi^a\|_{TV}=0$$

where $\|\bullet\|_{TV}$ is the total variation norm,

$$P_{n,a}(B):=P\left(X_1\in B\right|T_n=na)$$

and Π^a is the probability measure with density π^a defined through

$$\pi^a(x) := \frac{\exp tx}{\phi(t)} p(x)$$

with t defined through $(d/dt) \log \phi(t) = a$. See Diaconis and Freedman [32].

It is of some importance to consider the case when a is substituted by $a_n \to \infty$. In some cases it can be seen that the distribution Π^{a_n} tends to concentrate around a_n with a smaller and smaller variance ; this is indeed the case when X_1 has a Weibull distribution with a shape parameter larger than 1. This phenomenon would obviously be compatible with the result of the first chapter. Therefore the aim of the second chapter is to consider the validity of

$$\lim_{n\to\infty} \|P_{n,a_n} - \Pi^{a_n}\|_{TV} = 0.$$

This result requires a sharp extension of the local Edgeworth expansion for arrays of r.v's Z_{i,k_n} of row-wise independent terms where the expectations of the independent r.v's $(Z_{i,k_n})_{i=1,\dots,k_n}$ go to infinity. This extension of Feller's proof leads to the development of a new proof for the total variation approximation in the same spirit as developed in Broniatowski and Caron [16]. The analytical form of the underlying density p of X_1 borrows some properties from Nagaev [54] and are quite natural for these problems.

What is not done in this thesis

Obviously the main question is the relation between the present approach and the theory of extreme order statistics. It is well known that the asymptotic behaviour of the maximal term of an i.i.d. sample $X_1, ..., X_n$ approaches a_n, the quantile of order $1 - 1/n$ of the distribution of X_1 when the tail of this distribution decays rapidly to 0, plus some regularity assumption. A simple sufficient condition is provided by a criterion stated by Geffroy ([39] [40] and [41]) : whenever $1-F(x) = \exp -xa(x) \log x$ for some function $a(\cdot)$ which tends to infinity as x tends to infinity, then $X_{n,n} := \max(X_1, ..., X_n)$ almost surely satisfies

$$\lim_{n\to\infty} X_{n,n} - a_n = 0. \tag{2}$$

No attempt has been performed to put forwards any relation between this profound result and the results presented in this thesis.

Also the following facts could (should ?) have been studied in connexion with the present attempt. We still consider the context of light tailed distributions. The classical Erdös-Rényi law of large numbers asserts that the maximum local slope of the trajectory of a random walk characterizes the distribution of the i.i.d. summands when evaluated on blocks of length of order $\log n$. More specifically, almost surely, for any positive C

$$\lim_{n\to\infty} \max_{1\le j\le n-[C\log n]} \frac{S_{j+[C\log n]} - S_j}{C \log n} =: \lim_{n\to\infty} \Delta_{n,C\log n} = I(C)$$

where $C \to I(C)$ characterizes the distribution of X_1. When $C \log n$ is replaced by any larger window bin (in order of magnitude, obviously), then the strong law of large number applies, so that the limit in the above display is EX_1. This interesting result can also be compared with the behavior of the local slope for blocks smaller than $C \log n$. Not surprisingly the behaviour of this slope depends sharply upon the tail behaviour of the distribution of X_1. When (2) holds the a.s. asymptotic behaviour of $\lim_{n\to\infty} \Delta_{n,R_n}$ for "small" R_n inherits from $\lim_{n\to\infty} \Delta_{n,1} = \lim_{n\to\infty} X_{n,n}$. For larger R_n (still smaller than $\log n$ in order of magnitude), then the situation may be somehow different. We refer to Broniatowski [13] and Broniatowski and Mason [21] for details. Clearly these facts may be put in relation with the present work : on the one hand, in the realm of the classical Erdös-Rényi law, the ingredient in the proof is related to a large deviation result pertaining to i.i.d. sequences, in the spirit of Cramer or, for sharper results, of Petrov [70] or Höglund [48] type. It can readily be seen that the local behaviour of a random walk conditioned on a large deviation event pertaining to its sum would require extreme deviation probabilities ; this would be a natural extension of the present work.

Other extensions should be considered in relation with physics. We quoted the papers by Frisch and Sornette [38], and the book by Sornette [75], which is in close connection with the present problems on collective behaviour under a constraint. In the same vein the model in fragmentation processes which is presented in this book raises questions in accordance with the present cases. Extensions of the Gibbs conditional principle as presented in Chapter 2 might be put in relation with classical statistical thermodynamics, for example assuming an increasing density (with respect to the size of the system under consideration) for a system of particles. The construction by Lanford [58] would then lead to the present setting (at least for zero correlation description).

Maximum likelihood, large deviations and weighted sampling

What is done in this thesis

The third chapter of the thesis is of a quite different nature and handles some topics in mathematical statistics.

It explores Maximum Likelihood paradigm in the context of sampling. It mainly quotes that inference criterion is strongly connected with the sampling scheme generating the data. Under a given model, when i.i.d. sampling is considered and some standard regularity is assumed, then the Maximum Likelihood principle loosely states that conditionally upon the observed data, resampling under the same i.i.d. scheme should resemble closely to the initial sample only when the resampling distribution is close to the initial unknown one.

Keeping the same definition it appears that under other sampling schemes, the Maximum Likelihood Principle yields a wide range of statistical procedures. Those have in common with the classical simple i.i.d. sampling case that they can be embedded in a natural class of methods based on minimization of $\phi-$divergences between the empirical measure of the data and the model. In the classical i.i.d. case the divergence is the Kullback-Leibler one, which yields the standard form of the Likelihood function. In the case of the weighted bootstrap, the divergence to be optimized is directly related to the distribution of the weights.

This chapter discusses the choice of an inference criterion in parametric setting.

We consider a wide range of commonly used statistical criterions, namely all those induced by the so-called power divergence, including therefore Maximum Likelihood, Kullback-Leibler, Chi-square, Hellinger distance, etc. The steps of the discussion are as follows.

We first insert Maximum Likelihood paradigm at the center of the scene, putting forwards its strong connection with large deviation probabilities for the empirical measure. The argument can be sketched as follows : for any putative θ in the parameter set, consider n virtual simulated r.v's $X_{i,\theta}$ with corresponding empirical measure $P_{n,\theta}$. Evaluate the probability that $P_{n,\theta}$ is close to P_n , conditionally on P_n, the empirical measure pertaining to the observed data ; such statement is refered to as a conditional Sanov theorem, and for any θ this probability is governed by the Kullback-Leibler distance between P_θ and P_{θ_T} where θ_T stands for the true value of the parameter. Estimate this probability for any θ, obviously based on the observed data. Optimize in θ; this provides the MLE, as shown in the two cases of the i.i.d. sample scheme ; our first example is the case when the observations take values in a finite set, and the second case (infinite case), helps to set the arguments to be put forwards. Introducing MLE's through Large deviations for the empirical measure is in the vein of various recent approaches ; see Grendar and Judge [43].

We next consider a generalized sampling scheme inherited from the bootstrap, which we call weighted sampling ; it amounts to introduce a family of i.i.d. weights $W_1, ..., W_n$ with mean and variance 1. The corresponding empirical measure pertaining to the data set x_1 , ..., x_n is just the weighted empirical measure. The MLE is defined through a similar procedure as just evoqued. The conditional Sanov Theorem is governed by a divergence criterion which is defined through the distribution of the weights. Hence MLE results in the optimization of a divergence measure between distributions in the model and the weighted empirical measure pertaining to the dataset. Resulting properties of the estimators are studied, together with the optimality of some weighting designs in the context of the Bahadur slope.

Optimization of $\phi-$divergences between the empirical measure of the data and the model is problematic when the support of the model is not finite. A number of authors have considered so-called dual representation formulas for divergences or, globally, for convex pseudodistances between distributions. We will make use of the one exposed in [19] ; see also [15] for an easy derivation. A somehow similar approach, using variational techniques, has been proposed by Pelletier [69].

What is not done in this thesis

Two major questions are not discussed in this chapter.

The first one is related to the role of randomly weighting a dataset. The argument may be seen as follows : randomly weighting data amounts to resampling ; indeed consider the empirical distribution function F_n pertaining to a dataset $(X_1, ..., X_n)$ obtained through i.i.d. sampling with common d.f. F. Let $W_1, ..., W_n$ be i.i.d. non negative random weights with mean 1 and F_n^W denote the function defined through

$$F_n^W(x) := \frac{1}{n}\sum_{i=1}^{n} W_i 1\left\{X_i \leq x\right\}.$$

Plugging this empirical distribution in place of the standard one in estimating equations produces a simple way to obtain realizations of estimators ; indeed this procedure is known as “wild bootstrap” or “weighted bootstrap” and the properties of the resulting estimators, obtained through the classical MLE technique, for example, have been considered in Barbe and Bertail [3] or Mammen [63]. However we conjecture that weighting the empirical measure can be of greater generality ; this results from various simulations in the field of biostatistics where this plug in has been tested in estimating equations. The resulting estimators have been calculated using the present MLE approach and the resulting estimate has been obtained averaging the MLE’s obtained on each run. The theoretical aspects have still to be achieved.

The reciprocal question pertaining to weights is as follows : for a given statistical criterion in the class of divergences, is it possible to determine weights such that this criterion would have some good properties when used under this scheme ? This question is relevant since the choice of a statistical criterion depends on the requirement about accuracy ; for example the Chi square distance measures a mean square relative error and is known to enjoy nice robustness properties. A systematic approach of this question remains to be performed.

Chapter 1

Stretched random walks and behavior of their summands

1.1 Context and scope

This paper considers the following question: Let $X, X_1, ..., X_n$ denote real valued independent random variables (r.v's) distributed as X and let $S_1^n := X_1+...+X_n$. We assume that X is unbounded upwards. Let a_n be some positive sequence satisfying

$$\lim_{n\to\infty} a_n = +\infty. \tag{1.1.1}$$

Assuming that

$$C_n := (S_1^n/n > a_n) \tag{1.1.2}$$

holds, what can be inferred on the r.v's X_i's as n goes to infinity?

Let ε_n denote a positive sequence and let

$$I_n := \cap_{i=1}^n (X_i \in (a_n - \epsilon_n, a_n + \epsilon_n)). \tag{1.1.3}$$

We consider cases when

$$\lim_{n\to\infty} P(I_n | C_n) = 1. \tag{1.1.4}$$

The relation between the various parameters in this problem is of interest and opens a variety of questions. For which distributions P_X pertaining to X is such a result valid? Which is the acceptable growth of the sequence a_n and the possible behaviours of the sequence ϵ_n such that

$$\epsilon_n = o(a_n) \tag{1.1.5}$$

and is it possible to achieve

$$\lim_{n\to\infty} \epsilon_n = 0 \tag{1.1.6}$$

under a large class of choices for P_X?

In the case when the r.v. X has light tails conditional limit theorems exploring the behavior of the summands of a random walk given its sum have been developed extensively in the range of a large deviation conditioning event, namely similar as defined by C_n with fixed a_n, hence lower-bounding S_n/n independently on n; the papers [27], or [32] together with their extension in [30] explore the asymptotic properties of a relatively small number of summands; the main result in these papers, named as Gibbs conditional principle, lies in the fact that under such C_n, the X_i's are asymptotically i.i.d. with distribution Π^a defined through $d\Pi^a(x) := (E(\exp tX))^{-1}\exp(tx)dP_X(x)$ where t satisfies $E(X\exp tX)(E(\exp tX))^{-1} = a$; in this range (1.1.6) does not hold. The joint distribution of $X_1,..,X_{k_n}$ given C_n (with fixed a_n) for large k_n (close to n) is considered in [16].

Extended large deviations results for $a_n \to \infty$ have been considered in [13], [21], in relation with versions of the Erdös-Rényi law of large numbers for the small increments of a random walk, and [54].

The case when X is heavy tailed is considered in [2] where the authors consider the support of the distribution of the whole sample $X_1,...,X_n$ when C_n holds for fixed a_n .

A closely related problem has been handled by statisticians in various contexts, exploring the number of sample observations which push a given statistics far away from its expectation, for *fixed* n. Although similar in phrasing as the so-called "breakdown point" paradigm of robust analysis, the frame of this question is quite different from the robustness point of view, since all the observations are supposed to be sampled under the distribution P_X, hence without any reference to outliers or misspecification. The question may therefore be stated as: how many sample points should be large making a given statistics large? This combines both the asymptotic behavior of the statistics (as a function defined on $\mathbb{R}^n$) and the tail properties of P_X. In the case when the statistics is S_1^n/n and X has subexponential upper tail, it is well known that, denoting

$$C_a := (S_1^n/n > a)$$

only one large value of the X_i's generates C_a for $a \to \infty$; clearly S_n/n is not a loyal statistics under this sampling. This result turns back to Darling [29]. For light tails, under C_a, all sampled values should exceed a (indeed they should be closer and closer to a as $a \to \infty$), so that S_n/n is faithful in allegiance with respect to the sample. In this case, denoting

$$I_a := \cap_{i=1}^n (X_i > a)$$

it holds

$$\lim_{a\to\infty} P(I_a|\, C_a) = 1. \tag{1.1.7}$$

Intermediate cases exist, leading to partial loyalty for a given statistics under a given sampling scheme. See [17], [8], and [4] where more general statistics than S_n/n are considered when $a \to \infty$. According to the tail behavior of the distribution of X the situation may take quite different features.

Related questions have also been considered in the realm of statistical physics. In [38] the property (1.1.7) is stated in an improved form, namely stating that when the X_i's are i.i.d. with Weibull density with shape index larger than 2 then the conditional density of $(X_1,...,X_n)$ given $(S_1^n/n = a)$ concentrates at $(a,...,a)$ as $a \to \infty$, which in the authors' words means that the X_i's are *democratically localized*. Applications of this concept in fragmentation processes, in some form of anomalous relaxation of glasses and in the study of turbulence flows are discussed.

We now come to a consequence of the present results considering the local behaviour of a random walk conditioned on its end value. Let $S_i^j := X_i + ... + X_j$ with $1 \leq i \leq j \leq n$ and $k = k_n$ denote an integer valued sequence such that

$$k_n \leq n$$

and

$$\lim_{n\to\infty} k_n = \infty.$$

Let further

$$\Delta_{j,n} := S_{j+1}^{j+k}/k$$

denote the local slope of the random walk on the interval $[j+1, j+k]$ where $1 \leq j \leq n-k$. The limit behaviour of $\max_{1\leq j\leq n-k} \Delta_{j,n}$ has been considered extensively in various cases, according to the order of magnitude of k. The case $k = C \log n$ for positive constant C defines the so-called Erdös-Rényi law of large numbers; see [35]. In the present case we consider random walks conditioned upon their end value, namely assuming that

$$S_1^n > na$$

for fixed $a > EX$. The path defined by this random walk exhibits anomalous local behavior that can be captured through the *extended democratic localization principle* stated in our results. Indeed there exist segments of length k_n (say, k_n has smaller order than $\log n$) on which the slope $\Delta_{j,n}$ tends to infinity. Obviously, when a is not fixed but goes to infinity with n then the *extended democratic localization principle* applies to the whole sample path of the random walk, and its trajectory is nearly a straight line from the origin up to its extremity. When conditioning in the range of the large deviation only, this property should hold locally; this will be studied in a future work.

This paper is organized as follows. Section 2 states the notation and hypotheses. Section 3 states the results in two cases; the first one pertains to the case when X has a log-concave density and the second case is a generalization of the former. Examples are provided. The proofs of the results are rather long and technical; they have been postponed to Section 1.5.

1.2 Notation and hypotheses

The n real valued random variables $X_1, ..., X_n$ are independent copies of a r.v. X with density p whose support is $\mathbb{R}^+$. As seen by the very nature of the problem handled in this paper, this assumption puts no restriction to the results. We write

$$p(x) := c \exp(-h(x))$$

for positive functions h which are defined and denoted according to the context, and c is some positive normalized constant. For $\mathbf{x} \in \mathbb{R}^n$ define

$$I_h(\mathbf{x}) := \sum_{1\leq i\leq n} h(x_i),$$

and for A a Borel set in $\mathbb{R}^n$ denote

$$I_h(A) = \inf_{\mathbf{x}\in A} I_h(\mathbf{x}).$$

Two cases will be considered: in the first one h is assumed to be a convex function, and in the second case h will be the sum of a convex function and a "smaller" function h in such a way that we will also handle non log-concave densities (although not too far from them). Hence we do not consider heavy tailed r.v. X.

1.3 Random walks for log-concave Density under extreme deviation

Lemma 1.3.1. *Let g be a positive convex differentiable function defined on $\mathbb{R}_+$. Assume that g is strictly increasing on some interval $[Y, \infty)$. Let (1.1.1) hold. Then*

$$I_g(I_n^c \cap C_n) = \min\Big(F_{g_1}(a_n, \epsilon_n), F_{g_2}(a_n, \epsilon_n)\Big),$$

where

$$F_{g_1}(a_n, \epsilon_n) = g(a_n + \epsilon_n) + (n-1)g\left(a_n - \frac{1}{n-1}\epsilon_n\right),$$

and

$$F_{g_2}(a_n, \epsilon_n) = g(a_n - \epsilon_n) + (n-1)g\left(a_n + \frac{1}{n-1}\epsilon_n\right).$$

Theorem 1.3.1. *Let $X_1, ..., X_n$ be i.i.d. copies of a r.v. X with density $p(x) = c\exp(-g(x))$, where $g(x)$ is a positive convex function on $\mathbb{R}^+$. Assume that g is increasing on some interval $[Y, \infty)$ and satisfies*

$$\lim_{x\to\infty} g(x)/x = \infty.$$

Let a_n satisfy

$$\liminf_{n\to\infty} \frac{\log g(a_n)}{\log n} > 0 \tag{1.3.1}$$

and that for some positive sequence ϵ_n

$$\lim_{n\to\infty} \frac{n \log g(a_n + \epsilon_n)}{H(a_n, \epsilon_n)} = 0, \tag{1.3.2}$$

$$\lim_{n\to\infty} \frac{nG(a_n)}{H(a_n, \epsilon_n)} = 0, \tag{1.3.3}$$

where

$$H(a_n, \epsilon_n) = \min\left(F_{g_1}(a_n, \epsilon_n), F_{g_2}(a_n, \epsilon_n)\right) - ng(a_n),$$
$$G(a_n) = g\left(a_n + \frac{1}{g(a_n)}\right) - g(a_n),$$

$F_{g_1}(a_n, \epsilon_n)$ and $F_{g_2}(a_n, \epsilon_n)$ are defined as in Lemma 1.3.1. Then it holds

$$\lim_{n\to\infty} P(I_n | C_n) = 1.$$

Remark 1.3.1. *Conditions (1.3.1), (1.3.2) and (1.3.3) state the relations between n, a_n and ϵ_n. Roughly speaking, under condition (C_n), in order that $X_1, ..., X_n$ concentrate in the "cube" I_n, we need to control the speed of ϵ_n such that it cannot tend to 0 too fast which is characterized by conditions (1.3.2) and (1.3.3). Intuitively, this is also correct because if the width ($2\epsilon_n$) of the "cube" I_n is too "thin", then there will be more outliers falling outside I_n.*

Example 1.3.1. *Let $g(x) := x^\beta$. For power functions, through Taylor expansion it holds*

$$g\left(a_n + \frac{1}{g(a_n)}\right) - g(a_n) = \frac{\beta}{a_n} + o\left(\frac{1}{a_n}\right) = o\left(\log g(a_n)\right)$$

hence condition (1.3.3) holds as a consequence of (1.3.2). If we assume that $\epsilon_n = o(a_n)$, by Taylor expansion we obtain

$$\min\Big(F_{g_1}(a_n, \epsilon_n), F_{g_2}(a_n, \epsilon_n)\Big) = na_n^\beta + C_\beta^2 \frac{n}{n-1} a_n^{\beta-2}\epsilon_n^2 + o(a_n^{\beta-2}\epsilon_n^2).$$

Condition (1.3.2) then becomes

$$\lim_{n\to\infty} \frac{n \log a_n}{a_n^{\beta-2}\epsilon_n^2} = 0. \tag{1.3.4}$$

Case 1: $1 < \beta \leq 2$.

To make (1.3.4) hold, if we take $n = a_n^\alpha$ with $0 < \alpha < \beta$, we need ϵ_n be large enough, specifically,

$$a_n^{1-\frac{\beta}{2}}\sqrt{n \log a_n} = o\left(\epsilon_n\right) = o\left(a_n\right)$$

which shows that $\epsilon_n \to \infty$.

Case 2: $\beta > 2$.

In this case, if we take $n = a_n^\alpha$ with $0 < \alpha < \beta - 2$, then condition (1.3.4) holds for arbitrary sequences ϵ_n bounded by below away from 0. The sequence ϵ_n may also tend to 0; indeed with $\epsilon_n = O(1/\log a_n)$, condition (1.3.4) holds. Also setting $a_n := n^\alpha$ for $\alpha > 1/(\beta - 2)$ there exist sequences ϵ_n which tend to 0 such that the conclusion in Theorem 1.3.1 holds.

Example 1.3.2. ***Let*** *$g(x) := e^x$. Through Taylor expansion*

$$g\left(a_n + \frac{1}{g(a_n)}\right) - g(a_n) = 1 + o\left(\frac{1}{a_n}\right) = o\left(\log g(a_n)\right) = o\left(a_n\right),$$

and if $\epsilon_n \to 0$, it holds

$$\min\left(F_{g_1}(a_n, \epsilon_n), F_{g_2}(a_n, \epsilon_n)\right) = ne^{a_n} + \frac{1}{2}\frac{n}{n-1}e^{a_n}\epsilon_n^2 + o(e^{a_n}\epsilon_n^2).$$

Hence condition (1.3.3) follows from condition (1.3.2); furthermore condition (1.3.2) follows from

$$\lim_{n\to\infty} \frac{na_n}{e^{a_n}\epsilon_n^2} = 0$$

if we set $a_n := n^\alpha$ where $\alpha > 0$ then condition (1.3.3) holds, and ϵ_n is rapidly decreasing to 0; indeed we may choose $\epsilon_n = o(\exp(-a_n/4))$.

Corollary 1.3.1. *Let $X_1, ..., X_n$ be independent r.v's with common Weibull density with shape parameter k and scale parameter 1,*

$$p(x) = \begin{cases} kx^{k-1}e^{-x^k} & \text{when } x > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $k > 2$. Let

$$a_n = n^{\frac{1}{\alpha}}, \tag{1.3.5}$$

for some $0 < \alpha < k - 2$ and let ϵ_n be a positive sequence tending to 0 such that

$$\lim_{n\to\infty} \frac{n^{1-(k-2)/\alpha} \log n}{\epsilon_n^2} = 0. \tag{1.3.6}$$

Then

$$\lim_{n\to\infty} P(I_n|C_n) = 1.$$

Proof. Set $g(x) = x^k - (k-1)\log x$, which is a convex function for $k > 2$. Also when $x \to \infty$, $g'(x)$ and $g''(x)$ are both infinitely small with respect to $g(x)$ as $x \to \infty$.

Both conditions (1.3.2) and (1.3.3) in Theorem 1.3.1 are satisfied. As regards to condition (1.3.3), notice firstly that, under the Weibull density by Taylor expansion

$$g(a_n + \epsilon_n) = g(a_n) + g'(a_n)\epsilon_n + \frac{1}{2}g''(a_n)\epsilon_n^2 + o\left(g''(a_n)\epsilon_n^2\right).$$

Hence it holds

$$\log g\left(a_n + \epsilon_n\right) \le \log\left(3g(a_n)\right) \le \log\left(3a_n^k\right) = \log 3 + k\log a_n. \tag{1.3.7}$$

Using Taylor expansion in $g(a_n + \epsilon_n)$ and $g\left(a_n - \frac{\epsilon_n}{n-1}\right)$, it holds

$$\begin{aligned} F_{g_1}(a_n, \epsilon_n) - ng(a_n) &= g(a_n + \epsilon_n) + (n-1)g\left(a_n - \frac{\epsilon_n}{n-1}\right) - ng(a_n) \\ &= \left(g(a_n) + g'(a_n)\epsilon_n + \frac{1}{2}g''(a_n)\epsilon_n^2 + o\left(g''(a_n)\epsilon_n^2\right)\right) \\ &\quad + \left((n-1)g(a_n) - g'(a_n)\epsilon_n + \frac{1}{2}g''(a_n)\frac{\epsilon_n^2}{n-1} + o\left(g''(a_n)\epsilon_n^2\right)\right) - ng(a_n) \\ &= \frac{1}{2}g''(a_n)\epsilon_n^2 + o\left(g''(a_n)\epsilon_n^2\right) = \frac{k(k-1)}{2}a_n^{k-2}\epsilon_n^2 + o\left(a_n^{k-2}\epsilon_n^2\right). \end{aligned}$$

In the same way, it holds when $a_n \to \infty$

$$F_{g_2}(a_n, \epsilon_n) - ng(a_n) = \frac{k(k-1)}{2}a_n^{k-2}\epsilon_n^2 + o\left(a_n^{k-2}\epsilon_n^2\right).$$

Thus we have

$$H(a_n, \epsilon_n) = \frac{k(k-1)}{2}a_n^{k-2}\epsilon_n^2 + o\left(a_n^{k-2}\epsilon_n^2\right). \tag{1.3.8}$$

Hence, when $n \to \infty$, with (1.3.7), (1.3.8), the condition (1.3.2) of Theorem (1.3.1) becomes

$$\begin{aligned}\frac{n\log g\left(a_n+\epsilon_n\right)}{H(a_n,\epsilon_n)} &\leq \frac{n\log 3+kn\log a_n}{\frac{k(k-1)}{2}a_n^{k-2}\epsilon_n^2+o\left(a_n^{k-2}\epsilon_n^2\right)}\\ &\leq \frac{2kn\log a_n}{\frac{k(k-1)}{4}a_n^{k-2}\epsilon_n^2}=\frac{8}{k-1}\frac{n\log a_n}{a_n^{k-2}\epsilon_n^2}\longrightarrow 0.\end{aligned}$$

The last step holds from conditions (1.3.5) and (1.3.6). As for condition (1.3.3) of Theorem 1.3.1, when $a_n \to \infty$, it holds

$$\begin{aligned}nG(a_n) &= ng\left(a_n+\frac{1}{g(a_n)}\right)-ng(a_n)\\ &= ng(a_n)+n\frac{g'(a_n)}{g(a_n)}+o\left(\frac{g'(a_n)}{g(a_n)}\right)-ng(a_n)\\ &= n\frac{g'(a_n)}{g(a_n)}+o\left(\frac{g'(a_n)}{g(a_n)}\right)=o(n).\end{aligned}$$

Hence under conditions (1.3.5) and (1.3.6), it holds $nG(a_n) = o(H(a_n, \epsilon_n))$, which means that condition (1.3.3) of Theorem 1.3.1 holds under conditions (1.3.5) and (1.3.6), which completes the proof.

□

1.4 Random walks for non log-concave Density under extreme deviation

In this section, we pay attention to exponential density functions whose exponents are non-convex functions. Namely, i.i.d random variables $X_1, ..., X_n$ have common density with

$$f(x) = c\exp\Big(-\left(g(x)+q(x)\right)\Big)$$

assuming that the convex function g is twice differentiable and $q(x)$ is of smaller order than $\log g(x)$ for large x.

Theorem 1.4.1. *$X_1, ..., X_n$ are i.i.d. real valued random variables with common density $f(x) = c\exp\left(-(g(x) + q(x))\right)$, where $g(x)$ is some positive convex function on $\mathbb{R}^+$ and g is twice differentiable. Assume that $g(x)$ is increasing on some interval $[Y, \infty)$ and satisfies*

$$\lim_{x\to\infty} g(x)/x = \infty.$$

Let $M(x)$ be some nonnegative continuous function on $\mathbb{R}^+$ for which

$$-M(x) \leq q(x) \leq M(x) \quad \textit{for all positive } x$$

together with

$$M(x) = O\left(\log g(x)\right) \tag{1.4.1}$$

as $x \to \infty$.

Let a_n *be some positive sequence such that* $a_n \to \infty$ *and* $\epsilon_n = o(a_n)$ *be a positive sequence. Assume*

$$\liminf_{n\to\infty} \frac{\log g(a_n)}{\log n} > 0 \tag{1.4.2}$$

$$\lim_{n\to\infty} \frac{n \log g\left(a_n + \epsilon_n\right)}{H(a_n, \epsilon_n)} = 0, \tag{1.4.3}$$

$$\lim_{n\to\infty} \frac{nG(a_n)}{H(a_n, \epsilon_n)} = 0, \tag{1.4.4}$$

where

$$H(a_n, \epsilon_n) = \min\left(F_{g_1}(a_n, \epsilon_n), F_{g_2}(a_n, \epsilon_n)\right) - ng(a_n),$$
$$G(a_n) = g\left(a_n + \frac{1}{g(a_n)}\right) - g(a_n),$$

$F_{g_1}(a_n, \epsilon_n)$ *and* $F_{g_2}(a_n, \epsilon_n)$ *are defined as in Lemma 1.3.1. Then it holds*

$$\lim_{n\to\infty} P(I_n|C_n) = 1.$$

Remark 1.4.1. *Log-concavity is a classical hypothesis in limit theorems related with convolution or concentration of measure; in relation with the present context, Jensen [49] developed a complete set of large deviation sharp results; log concave densities are closed under convolution. Nearly log-concave densities do not share the same simple features but, hopefully, their convolutions can be somehow controlled by log-concave approximating densities. Log concave measures also appear as extending the gaussian case in the concentration phenomenon; see [7] for details, outside the frame of the present work.*

We now provide examples of densities which define r.v's X_i''s for which the above Theorem 1.4.1 applies. These densities appear in a number of questions pertaining to uniformity in large deviation approximations; see [49] Ch 6.

Example 1.4.1. ***Almost Log-concave densities 1****:* p *can be written as*

$$p(x) = c(x) \exp -g(x), \qquad 0 < x < \infty$$

with g *a convex and twice differentiable function, and where for some* $x_0 > 0$ *and constants* $0 < c_1 < c_2 < \infty$, *we have*

$$c_1 < c(x) < c_2 \quad \textit{for} \quad x_0 < x < \infty,$$

and $g(x)$ *is increasing on some interval* $[Y, \infty)$ *and satisfies*

$$\lim_{x\to\infty} g(x)/x = \infty.$$

Examples of densities which satisfy the above conditions include the Normal, the hyperbolic density, etc.

Example 1.4.2. ***Almost Log-concave densities 2***: *A wide class of densities for which our results apply is when there exist constants* $x_0 > 0$, $\alpha > 0$, *and* A *such that*

$$p(x) = Ax^{\alpha-1} l(x) \exp(-g(x)) \qquad x > x_0$$

where $l(x)$ *is slowly varying at infinity,* g *a convex and twice differentiable function, increasing on some interval* $[Y, \infty)$ *and satisfies*

$$\lim_{x\to\infty} g(x)/x = \infty.$$

Remark 1.4.2. *All density functions in Examples (1.4.1) (1.4.2) satisfy the assumptions of the above Theorem 1.4.1 . Also the conditions in Theorem 1.4.1 about* a_n *and* ϵ_n *are the same as those in the convex case, so that if* $g(x)$ *is some power function with index larger than* 2, ϵ_n *can go to* 0 *more rapidly than* $O(1/\log a_n)$*(see Example 1.3.1); If* $g(x)$ *is of exponential function form,* ϵ_n *goes to* 0 *more rapidly than any power* $1/a_n$ *(see Example 1.3.2).*

1.5 Proofs

1.5.1 Proof of Lemma 1.3.1

Proof. Write $\mathbf{x} := (x_1, ..., x_n) \in \mathbb{R}_+^n$, we firstly define the following sets. Let for all k between 0 and n

$$A_k := \left\{\text{there exist } i_1, ..., i_k \text{ such that } x_{i_j} \geq a_n + \epsilon_n \text{ for all } j \text{ with } 1 \leq j \leq k\right\}$$

and

$$B_k := \left\{\text{there exist } i_1, ..., i_k \text{ such that } x_{i_j} \leq a_n - \epsilon_n \text{ for all } j \text{ with } 1 \leq j \leq k\right\}.$$

Define

$$A = \bigcup \lim_{k=1}^{n} A_k$$

and

$$B = \bigcup \lim_{k=1}^{n} B_k.$$

It then holds

$$I_n^c = A \cup B.$$

It follows that

$$\begin{aligned} I_g(I_n^c \cap C_n) = I_g\left((A \cup B) \cap C_n\right) &= \inf_{\mathbf{x} \in (A\cap C_n)\cup(B\cap C_n)} I_g(\mathbf{x}) \\ &= \min\left(I_g(A \cap C_n), I_g(B \cap C_n)\right). \end{aligned}$$

Thus we may calculate the minimum values of both $I_g(A \cap C_n)$ and $I_g(B \cap C_n)$ respectively, and finally $I_g(I_n^c \cap C_n)$.

Step 1: In this step we prove that

$$I_g(A \cap C_n) = F_{g_1}(a_n, \epsilon_n). \tag{1.5.1}$$

Without loss of generality, assume that the x_i's are ordered ascendently, $x_1 \leq ... \leq x_i \leq x_{i+1} \leq ... \leq x_n$ and let i and $k := n - i$ with $1 \leq i \leq n$ such that

$$\overbrace{x_1 \leq ... \leq x_i}^{n-k} < a_n + \epsilon_n \leq \overbrace{x_{i+1} \leq ... \leq x_n}^{k}.$$

We first claim that $k < n$. Let $\mathbf{x}_{A\cap C_n} := (x_1, ..., x_n)$ belong to $A\cap C_n$ and assume that $I_g(A \cap C_n) = I_g(\mathbf{x}_{A\cap C_n})$. Indeed let $\mathbf{y} := (y_1 = a_n - \epsilon_n, y_2 = ... = y_{n-1} = a_n + \epsilon_n)$ which clearly belongs to $A \cap C_n$. For this $\mathbf{y}$ it holds $I_g(\mathbf{y}) = (n-1)g(a_n + \epsilon_n) + g(a_n - \epsilon_n)$ which is strictly smaller than $ng(a_n + \epsilon_n) = I_g(A_n \cap C_n)$ for large n. We have proved that $\mathbf{x}_{A\cap C_n}$ does not belong to $A_n \cap C_n$.

Let $\alpha_{i+1}, ..., \alpha_n$ be nonnegative, and write $x_{i+1}, ..., x_n$ as

$$x_{i+1} = a_n + \epsilon_n + \alpha_{i+1}, ..., x_n = a_n + \epsilon_n + \alpha_n.$$

Under condition (C_n), it holds

$$\begin{aligned} x_1 + ... + x_i &\geq na_n - (x_{i+1} + ... + x_n) \\ &= na_n - k(a_n + \epsilon_n) - (\alpha_{i+1} + ... + \alpha_n). \end{aligned}$$

Applying Jensen's inequality to the convex function g, we have

$$\begin{aligned} \sum_{l=1}^{n} g(x_l) &= (g(x_{i+1}) + ... + g(x_n)) + (g(x_1) + ... + g(x_i)) \\ &\geq (g(x_{i+1}) + ... + g(x_n)) + (n-k)g(x^*), \end{aligned}$$

where equality holds when $x_1 = ... = x_i = x^*$, with

$$x^* = \frac{na_n - k(a_n + \epsilon_n) - (\alpha_{i+1} + ... + \alpha_n)}{n-k}. \tag{1.5.2}$$

Define now the function function $(\alpha_{i+1}, ..., \alpha_n, k) \to f(\alpha_{i+1}, ..., \alpha_n, k)$ through

$$\begin{aligned} f(\alpha_{i+1}, ..., \alpha_n, k) &= g(x_{i+1}) + ... + g(x_n) + (n-k)g(x^*) \\ &= g(a_n + \epsilon_n + \alpha_{i+1}) + ... + g(a_n + \epsilon_n + \alpha_n) + (n-k)g(x^*). \end{aligned}$$

Then $I_g(A \cap C_n)$ is given by

$$I_g(A \cap C_n) = \inf_{\alpha_{i+1}, ..., \alpha_n \geq 0, 1 \leq k < n} f(\alpha_{i+1}, ..., \alpha_n, k).$$

We now obtain (1.5.1) through the properties of the function f. Using (1.5.2), the first order partial derivative of $f(\alpha_{i+1}, ..., \alpha_n, k)$ with respect to α_{i+1} is

$$\frac{\partial f(\alpha_{i+1}, ..., \alpha_n, k)}{\partial \alpha_{i+1}} = g'(a_n + \epsilon_n + \alpha_{i+1}) - g'(x^*) > 0,$$

where the inequality holds since $g(x)$ is strictly convex and $a_n + \epsilon_n + \alpha_{i+1} > x^*$. Hence $f(\alpha_{i+1}, ..., \alpha_n, k)$ is an increasing function with respect to α_{i+1}. This implies that the minimum value of f is attained when $\alpha_{i+1} = 0$. In the same way, we have $\alpha_{i+1} = ... = \alpha_n = 0$. Therefore it holds

$$I_g(A \cap C_n) = \inf_{1 \leq k < n} f(\mathbf{0}, k),$$

with

$$f(\mathbf{0},k) = kg(a_n+\epsilon_n) + (n-k)g(x_0^*),$$

where

$$x_0^* = a_n - \frac{k}{n-k}\epsilon_n.$$

The function $y \to f(\mathbf{0},y)$ with $0<y<n$ is increasing with respect to y, since

$$\begin{aligned}\frac{\partial f(\mathbf{0},y)}{\partial y} &= g(a_n+\epsilon_n) - g(x_0^*) - \frac{n\epsilon_n}{n-y}g'(x_0^*)\\ &= \frac{n\epsilon_n}{n-y}\left(\frac{g(a_n+\epsilon_n)-g(x_0^*)}{a_n+\epsilon_n-x_0^*} - g'(x_0^*)\right) > 0,\end{aligned}$$

due to the convexity of $g(x)$ and $a_n+\epsilon_n > x_0^*$. Hence $f(\mathbf{0},k)$ is increasing with respect to k; the minimal value of $f(\mathbf{0},k)$ attains with $k=1$. Thus we have

$$I_g(A\cap C_n) = f(\mathbf{0},1) = F_{g_1}(a_n,\epsilon_n)$$

which proves (1.5.1).

Step 2: In this step, we follow the same proof as above and prove that

$$I_g(B\cap C_n) = F_{g_2}(a_n,\epsilon_n). \tag{1.5.3}$$

Assume that the x_i's are ranked in ascending order, with k such that $1\le k\le n$ and

$$\overbrace{x_1\le ...\le x_k}^{k} \le a_n-\epsilon_n < \overbrace{x_{k+1}\le ...\le x_n}^{n-k}$$

we obtain $k<n$, otherwise condition (C_n) won't be satisfied. Denote $x_1,...,x_k$ by

$$x_1 = a_n-\epsilon_n-\alpha_1, ..., x_k = a_n-\epsilon_n-\alpha_k,$$

where $\alpha_1,...,\alpha_k$ are nonnegative. Under condition (C_n), it holds

$$\begin{aligned}x_{k+1}+...+x_n &\ge na_n - (x_1+...+x_k)\\ &= na_n - k(a_n-\epsilon_n) + (\alpha_1+...+\alpha_k).\end{aligned}$$

Using Jensen's inequality to the convex function $g(x)$, we have

$$\begin{aligned}\sum_{l=1}^{n} g(x_l) &= (g(x_1)+...+g(x_k)) + (g(x_{k+1})+...+g(x_n))\\ &\ge (g(x_1)+...+g(x_k)) + (n-k)g(x^\sharp),\end{aligned}$$

where the equality holds when $x_{k+1} = ... = x_n = x^\sharp$, with

$$x^\sharp = \frac{na_n - k(a_n-\epsilon_n) + (\alpha_1+...+\alpha_k)}{n-k}. \tag{1.5.4}$$

Define the function $(\alpha_1,...,\alpha_k,k) \to f(\alpha_1,...,\alpha_k,k)$ through

$$\begin{aligned}f(\alpha_1,...,\alpha_k,k) &= g(x_1)+...+g(x_k) + (n-k)g(x^\sharp)\\ &= g(a_n-\epsilon_n-\alpha_1)+...+g(a_n-\epsilon_n-\alpha_k) + (n-k)g(x^\sharp),\end{aligned}$$

then $I_g(B \cap C_n)$ is given by

$$I_g(B \cap C_n) = \inf_{\alpha_1,...,\alpha_k \geq 0, 1 \leq k < n} f(\alpha_1, ..., \alpha_k, k).$$

Using (1.5.4), the first order partial derivative of $f(\alpha_1, ..., \alpha_k, k)$ with respect to α_1 is

$$\frac{\partial f(\alpha_1, ..., \alpha_k, k)}{\partial \alpha_1} = -g'(a_n - \epsilon_n - \alpha_1) + g'(x^\sharp) > 0,$$

where the inequality holds since $g(x)$ is convex and $a_n - \epsilon_n - \alpha_1 < x^\sharp$. Hence $f(\alpha_1, ..., \alpha_k, k)$ is increasing with respect to α_1. This yields

$$\alpha_1 = ... = \alpha_k = 0.$$

Therefore it holds

$$I_g(B \cap C_n) = \inf_{1 \leq k < n} f(\mathbf{0}, k),$$

with

$$f(\mathbf{0}, k) = kg(a_n - \epsilon_n) + (n-k)g(x_0^\sharp),$$

where

$$x_0^\sharp = a_n + \frac{k}{n-k}\epsilon_n.$$

The function $y \to f(\mathbf{0}, y)$ with $0 < y < n$ is increasing with respect to y, since

$$\begin{aligned}\frac{\partial f(\mathbf{0}, y)}{\partial y} &= g(a_n - \epsilon_n) - g(x_0^\sharp) + \frac{n\epsilon_n}{n-y} g'(x_0^\sharp) \\ &= \frac{n\epsilon_n}{n-y}\left(g'(x_0^\sharp) - \frac{g(x_0^\sharp) - g(a_n - \epsilon_n)}{x_0^\sharp - (a_n - \epsilon_n)}\right) > 0,\end{aligned}$$

by the convexity of g ; in the above display $x_0^\sharp > a_n - \epsilon_n$. Hence $f(\mathbf{0}, k)$ is increasing with respect to k. Thus we have

$$I_g(B \cap C_n) = f(\mathbf{0}, 1) = F_{g_2}(a_n, \epsilon_n)$$

which proves the claim.

Thus the proof is completed using (1.5.1) and (1.5.3).

□

1.5.2 Proof of Theorem 1.3.1

For $\mathbf{x} := (x_1, ..., x_n) \in \mathbb{R}_+^n$ and positive r, define

$$S_g(r) = \left\{ \mathbf{x} : \sum_{1 \leq i \leq n} g(x_i) \leq r \right\}.$$

Then for any Borel set A in $\mathbb{R}_+^n$ it holds

$$\begin{aligned}P(A) =& c^n \int_A \exp\left(-\sum_{1 \leq i \leq n} g(x_i)\right) dx_1...dx_n \\ &= c^n \exp(-I_g(A)) \int_A dx_1...dx_n \int 1_{\left[\sum_{1 \leq i \leq n} g(x_i) - I_g(A), \infty\right)}(s) e^{-s} ds \\ &= c^n \exp(-I_g(A)) \int_0^\infty Volume(A \cap S_g(I_g(A) + s)) e^{-s} ds. \qquad (1.5.5)\end{aligned}$$

For proving Theorem 1.3.1, we state firstly the following two Lemmas.

Lemma 1.5.1. *With the same notation and hypothesis as in Theorem 1.3.1, it holds*

$$P(C_n) \geq c^n \exp\left(-I_g(C_n) - \tau_n - n \log g(a_n)\right),$$

where

$$\tau_n = ng\left(a_n + \frac{1}{g(a_n)}\right) - ng(a_n). \tag{1.5.6}$$

Proof. By convexity of the function g, and using condition (C_n), applying Jensen's inequality, with $x_1 = ... = x_n = a_n$ it holds

$$I_g(C_n) = ng(a_n).$$

We now consider the largest lower bound for

$$\log Volume\left(C_n \cap S_g(I_g(C_n) + \tau_n)\right).$$

Denote $B = \left\{\mathbf{x} : x_i \in \left[a_n, a_n + \frac{1}{g(a_n)}\right], i = 1, ..., n\right\}$, $S_g(I_g(C_n)+\tau_n) = \{\mathbf{x} : \sum_{i=1}^n g(x_i) \leq ng(a_n) + \tau_n\}$.

For large n and any $\mathbf{x}$ in B, it holds

$$\sum_{i=1}^n g(x_i) \leq \sum_{i=1}^n g\Big(a_n + \frac{1}{g(a_n)}\Big) = ng\Big(a_n + \frac{1}{g(a_n)}\Big) = ng(a_n) + \tau_n,$$

where we used the fact that $g(x)$ is an increasing function for large x. Hence

$$B \subset S_g(I_g(C_n) + \tau_n).$$

It follows that

$$\log Volume\left(C_n \cap S_g(I_g(C_n) + \tau_n)\right) \geq \log Volume(B) = \log\left(\frac{1}{g(a_n)}\right)^n = -n \log g(a_n), \tag{1.5.7}$$

which in turn using (1.5.5) and (1.5.7) implies

$$\begin{aligned}
\log P(C_n) &:= \log c^n \int_{C_n} \exp\left(-\sum_{1\leq i\leq n} g(x_i)\right) dx_1...dx_n \\
&\geq n \log c + \log\left(\exp(-I_g(C_n)) \int_{\tau_n}^{\infty} Volume(C_n \cap S_g(I_g(C_n) + s))e^{-s}ds\right) \\
&\geq n \log c - I_g(C_n) - \tau_n + \log Volume(C_n \cap S_g(I_g(C_n) + \tau_n)) \\
&\geq n \log c - I_g(C_n) - \tau_n - n \log g(a_n),
\end{aligned}$$

This proves the claim. □

Lemma 1.5.2. *With the same notation and hypothesis as in Theorem 1.3.1, it holds*

$$P(I_n^c \cap C_n) \leq c^n \exp\left(-I_g(I_n^c \cap C_n) + n \log I_g(I_n^c \cap C_n) + \log(n+1)\right).$$

Proof. For positive s, let

$$S_g(I_g(C_n)+s) = \left\{ \mathbf{x} : \sum_{1\leq i\leq n} g(x_i) \leq I_g(C_n)+s \right\}$$

and

$$F = \left\{ \mathbf{x} : g(x_i) \leq I_g(C_n)+s, i=1,...,n \right\}.$$

It holds

$$S_g(I_g(C_n)+s) \subset F.$$

Since $\lim_{x\to\infty} g(x)/x = +\infty$

$$F \subset \{\mathbf{x} : x_i \leq (I_g(C_n)+s), i=1,...,n\},$$

which yields

$$S_g(I_g(C_n)+s) \subset \{\mathbf{x} : x_i \leq (I_g(C_n)+s), i=1,...,n\},$$

from which we obtain

$$Volume(C_n \cap S_g(I_g(C_n)+s)) \leq Volume(S_g(I_g(C_n)+s)) \leq (I_g(C_n)+s)^n.$$

With this inequality and (1.5.5) we get as $n\to\infty$

$$\begin{aligned}
\log P(C_n) &= \log c^n \int_{C_n} \exp\left(- \sum_{1\leq i\leq n} g(x_i)\right) dx_1...dx_n \\
&= n\log c - I_g(C_n) + \log \int_0^\infty Volume(C_n \cap S_g(I_g(C_n)+s))e^{-s}ds \\
&\leq n\log c - I_g(C_n) + \log \int_0^\infty \left(I_g(C_n)+s\right)^n e^{-s}ds,
\end{aligned}$$

with integrating repeatedly by parts it holds

$$\begin{aligned}
&\int_0^\infty \left(I_g(C_n)+s\right)^n e^{-s}ds \\
&= I_g(C_n)^n + n\int_0^\infty \left(I_g(C_n)+s\right)^{n-1} e^{-s}ds \\
&= I_g(C_n)^n + nI_g(C_n)^{n-1} + n(n-1)\int_0^\infty \left(I_g(C_n)+s\right)^{n-2} e^{-s}ds \\
&\leq (n+1)I_g(C_n)^n, \qquad (1.5.8)
\end{aligned}$$

where the inequality holds because of $I_g(C_n) = ng(a_n)$ with $n\to\infty$, hence we have

$$\begin{aligned}
\log P(C_n) &\leq n\log c - I_g(C_n) + \log\left((n+1)I_g(C_n)^n\right) \\
&= n\log c - I_g(C_n) + n\log I_g(C_n) + \log(n+1).
\end{aligned}$$

Replace C_n by $I_n^c \cap C_n$. We then obtain

$$P(I_n^c \cap C_n) \leq c^n \exp\left(-I_g(I_n^c\cap C_n) + n\log I_g(I_n^c\cap C_n) + \log(n+1)\right)$$

as sought. □

Proof of Theorem 1.3.1. We will complete the proof, showing that

$$\lim_{n\to\infty} \frac{P(I_n^c \cap C_n)}{P(C_n)} = 0. \tag{1.5.9}$$

By Lemma 1.3.1,

$$I_g(I_n^c \cap C_n) = \min\left(F_{g_1}(a_n, \epsilon_n), F_{g_2}(a_n, \epsilon_n)\right).$$

Using Lemma 1.5.1 and Lemma 1.5.2 it holds

$$\frac{P(I_n^c \cap C_n)}{P(C_n)} \leq \exp\left(-H(a_n, \epsilon_n) + n \log I_g(I_n^c \cap C_n) + \tau_n + n \log g(a_n) + \log(n+1)\right).$$

Under conditions (1.3.3), by (1.5.6) when $n \to \infty$, we have

$$\frac{\tau_n}{H(a_n, \epsilon_n)} = \frac{nG(a_n)}{H(a_n, \epsilon_n)} \longrightarrow 0, \tag{1.5.10}$$

Using condition (1.3.2), when $n \to \infty$,

$$\frac{n \log g(a_n)}{H(a_n, \epsilon_n)} \longrightarrow 0, \qquad \textit{and} \qquad \frac{\log(n+1)}{H(a_n, \epsilon_n)} \longrightarrow 0. \tag{1.5.11}$$

As to the term $n \log I_g(I_n^c \cap C_n)$, we have

$$\begin{aligned} n \log I_g(I_n^c \cap C_n) &= n \log \min\left(F_{g_1}(a_n, \epsilon_n), F_{g_2}(a_n, \epsilon_n)\right) \\ &\leq n \log\left(ng(a_n + \epsilon_n)\right) \\ &= n \log n + n \log g\left(a_n + \epsilon_n\right). \end{aligned}$$

Under condition (1.3.2), $n \log g\left(a_n + \epsilon_n\right)$ is of small order with respect to $H(a_n, \epsilon_n)$ as n tends to infinity. Under condition (1.3.1), for a_n large enough, there exists some positive constant Q such that $\log n \leq Q \log g(a_n)$. Hence we have

$$n \log n \leq Qn \log g(a_n)$$

which under condition (1.3.2), yields that $n \log n$ is negligible with respect to $H(a_n, \epsilon_n)$. Hence when $n \to \infty$, it holds

$$\frac{n \log\left(I_g(I_n^c \cap C_n)\right)}{H(a_n, \epsilon_n)} \longrightarrow 0. \tag{1.5.12}$$

Further, (1.5.10), (1.5.11) and (1.5.12) make (1.5.9) hold. This completes the proof. □

1.5.3 Proof of Theorem 1.4.1

The proof is is in the same vein as that of Theorem 1.3.1; some care has to be taken in order to get similar bounds as developped in the convex case.

Denote $\mathbf{x} = (x_1, ..., x_n)$ in $\mathbb{R}_+^n$ and, for a Borel set $A \in {\mathbb{R}_+}^n$ define

$$I_{g,q}(A) = \inf_{\mathbf{x}\in A} I_{g,q}(\mathbf{x}),$$

where

$$I_{g,q}(\mathbf{x}) := \sum_{1\le i\le n} (g(x_i)+q(x_i)).$$

Also for any positive r define

$$S_{g,q}(r) = \left\{ \mathbf{x} : \sum_{1\le i\le n} (g(x_i)+q(x_i)) \le r \right\}.$$

Then it holds

$$\begin{aligned} P(A) &= c^n \int_A \exp\left(- \sum_{1\le i\le n} (g(x_i)+q(x_i)) \right) dx_1...dx_n \\ &= c^n \exp(-I_{g,q}(A)) \int_A dx_1...dx_n \int 1_{\left[\sum_{1\le i\le n}(g(x_i)+q(x_i))-I_{g,q}(A),\infty\right)}(s)e^{-s}ds \\ &= c^n \exp(-I_{g,q}(A)) \int_0^\infty Volume(A\cap S_{g,q}(I_{g,q}(A)+s))e^{-s}ds. \end{aligned} \tag{1.5.13}$$

The proof of Theorem 1.4.1 relies on the following four Lemmas.

Lemma 1.5.3. *With the same notation and hypothesis as in Theorem 1.4.1, it holds*

$$I_{g,q}(C_n) \ge ng(a_n) - nN\log g(a_n).$$

Proof. For large x it holds

$$g(x) - M(x) \le g(x) + q(x) \le g(x) + M(x). \tag{1.5.14}$$

Set $g_1(x) = g(x) - M(x)$ and $g_2(x) = g(x) + M(x)$, then it follows

$$I_{g_1}(C_n) \le I_{g,q}(C_n) \le I_{g_2}(C_n). \tag{1.5.15}$$

In the same way, it holds

$$I_{g_1}(I_n^c \cap C_n) \le I_{g,q}(I_n^c \cap C_n) \le I_{g_2}(I_n^c \cap C_n). \tag{1.5.16}$$

By condition (1.4.1), there exists some sufficiently large positive y_0 and some positive constant N such that for $x \in [y_0, \infty)$

$$M(x) \le N \log g(x). \tag{1.5.17}$$

Set $r(x) = g(x) - N\log g(x)$, the second order derivative of $r(x)$ is

$$r''(x) = g''(x)\left(1 - \frac{N}{g(x)}\right) + \frac{N\,(g'(x))^2}{g^2(x)},$$

where the second term is positive. The function g is increasing on some interval $[Y,\infty)$ where we also have $g(x) > x$. Hence there exists some $y_1 \in [Y,\infty)$ such that $g(x) > N$ when $x \in [y_1,\infty)$. This implies that $r''(x) > 0$ and $r'(x) > 0$ and therefore $r(x)$ is convex and increasing on $[y_1,\infty)$.

In addition, $M(x)$ is bounded on any finite interval; there exists some $y_2 \in [y_1,\infty)$ such that for all $x \in (0, y_2)$

$$M(x) \le N \log g(y_2). \tag{1.5.18}$$

The function g is convex and increasing on $[y_2, \infty)$. Thus there exists y_3 such that

$$g'(y_3) > 2g'(y_2) \quad \textit{and} \quad g(y_3) > 2N. \tag{1.5.19}$$

We now construct a function h as follows. Let

$$h(x) = r(x)\mathbf{1}_{[y_3,\infty)}(x) + s(x)\mathbf{1}_{(0,y_3)}(x), \tag{1.5.20}$$

where $s(x)$ is defined by

$$s(x) = r(y_3) + r'(y_3)(x - y_3). \tag{1.5.21}$$

We will show that

$$g_1(x) \geq h(x) \tag{1.5.22}$$

for $x \in (0, \infty)$.

If $x \in [y_3, \infty)$, then by (1.5.17), it holds

$$h(x) = r(x) = g(x) - N \log g(x) \leq g(x) - M(x) = g_1(x). \tag{1.5.23}$$

If $x \in (y_2, y_3)$, using (1.5.21), we have

$$s(x) \leq r(x) = g(x) - N \log g(x) \leq g(x) - M(x) = g_1(x), \tag{1.5.24}$$

where the first inequality comes from the convexity of $r(x)$. We now show that (1.5.22) holds when $x \in (0, y_2]$ if y_3 is large enough. For this purpose, set

$$t(x) = g(x) - s(x) - N \log g(y_2).$$

Take the first order derivative of t and use the convexity of g on $(0, y_2]$. We have

$$\begin{aligned} t'(x) &= g'(x) - s'(x) = g'(x) - r'(y_3) = g'(x) - \left(g'(y_3) - \frac{Ng'(y_3)}{g(y_3)}\right) \\ &= g'(x) - \left(1 - \frac{N}{g(y_3)}\right) g'(y_3) \leq g'(y_2) - \left(1 - \frac{N}{g(y_3)}\right) g'(y_3) \\ &< \frac{1}{2} g'(y_3) - \left(1 - \frac{N}{g(y_3)}\right) g'(y_3) < 0, \end{aligned}$$

where the inequalities in the last line hold from (1.5.19). Therefore t is decreasing on $(0, y_2]$. It follows that

$$t(x) \geq t(y_2) = g(y_2) - N \log g(y_2) - s(y_2) \geq g(y_2) - N \log g(y_2) - r(y_2) = 0,$$

which, together with (1.5.18), yields, when $x \in (0, y_2]$

$$g_1(x) = g(x) - M(x) \geq g(x) - N \log g(y_2) \geq s(x).$$

Together with (1.5.23), (1.5.24) and (1.5.20), this last display means that (1.5.22) holds.

We now prove that h is a convex function on $(0, \infty)$; indeed for x such that $0 < x \leq y_3$, $h''(x) = 0$, and if $x > y_3$, $h''(x) = r''(x) > 0$. The left derivative of $h(x)$ at y_3 is $h'(y_3^-) = r'(y_3)$, and it is obvious that the right derivative of $h(x)$

at y_3 is also $h'(y_3^+) = r'(y_3)$; hence h is derivable at y_3 and $h'(y_3) = r'(y_3)$, hence $h''(y_3) = r''(y_3) > 0$. This shows that h is convex on $(0, \infty)$.

Now under condition (C_n), using the convexity of h and (1.5.22), it holds

$$I_{g_1}(\mathbf{x}) = \sum_{i=1}^{n} (g(x_i) - M(x_i)) \geq \sum_{i=1}^{n} h(x_i) \geq nh\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = nh(a_n).$$

Using (1.5.15), we obtain the lower bound of $I_{g,q}(C_n)$ under condition (C_n) for a_n large enough (say, $a_n > y_3$)

$$I_{g,q}(C_n) \geq I_{g_1}(C_n) \geq nh(a_n) = nr(a_n) = ng(a_n) - nN \log g(a_n).$$

□

Lemma 1.5.4. *With the same notation and hypothesis as in Theorem 1.4.1, the following lower bound of $P(C_n)$ holds*

$$P(C_n) \geq c^n \exp\left(-I_{g,q}(C_n) - \tau_n - n \log g(a_n)\right),$$

where τ_n is defined by

$$\begin{aligned}\tau_n &= ng\left(a_n + \frac{1}{g(a_n)}\right) - ng(a_n) + nN \log g\left(a_n + \frac{1}{g(a_n)}\right) + nN \log g(a_n) \\ &= nG(a_n) + nN \log g(a_n) + nN \log g\left(a_n + \frac{1}{g(a_n)}\right). \qquad (1.5.25)\end{aligned}$$

Proof. Denote $B = \left\{\mathbf{x} : x_i \in \left[a_n, a_n + \frac{1}{g(a_n)}\right], i = 1, ..., n\right\}$. If $\mathbf{x} \in B$, by (1.5.17), which holds for large a_n (say, $a_n > y_3$ and g is an increasing function on (y_3, ∞)), we have

$$\begin{aligned}I_{g,q}(\mathbf{x}) &\leq \sum_{i=1}^{n} (g(x_i) + M(x_i)) \leq \sum_{i=1}^{n} (g(x_i) + N \log g(x_i)) \\ &\leq \sum_{i=1}^{n} \left(g\left(a_n + \frac{1}{g(a_n)}\right) + N \log g\left(a_n + \frac{1}{g(a_n)}\right)\right) \\ &= ng\left(a_n + \frac{1}{g(a_n)}\right) + nN \log g\left(a_n + \frac{1}{g(a_n)}\right) \\ &= \tau_n + ng(a_n) - nN \log g(a_n) \leq \tau_n + I_{g,q}(C_n),\end{aligned}$$

where the last inequality holds from Lemma 1.5.3. Since $B \subset C_n$, we have

$$B \subset C_n \cap S_{g,q}(I_{g,q}(C_n) + \tau_n).$$

Now we may obtain the lower bound

$$\log Volume\left(C_n \cap S_{g,q}(I_{g,q}(C_n) + \tau_n)\right) \geq \log Volume(B) = -n \log g(a_n). \qquad (1.5.26)$$

Using (1.5.13) and (1.5.26), it holds

$$\begin{aligned}\log P(C_n) &= \log c^n \int_{C_n} \exp\left(-\sum_{1\le i\le n}(g(x_i)+q(x_i))\right) dx_1...dx_n \\ &= n\log c - I_{g,q}(C_n) + \log\int_0^\infty Volume(C_n\cap S_{g,q}(I_{g,q}(C_n)+s))e^{-s}ds \\ &\ge n\log c - I_{g,q}(C_n) + \log\int_{\tau_n}^\infty Volume(C_n\cap S_{g,q}(I_{g,q}(C_n)+\tau_n))e^{-s}ds \\ &\ge n\log c - I_{g,q}(C_n) - \tau_n - n\log g(a_n),\end{aligned}$$

so the lemma holds. □

Lemma 1.5.5. *With the same notation and hypothesis as in Theorem 1.4.1, the following upper bound holds*

$$P(I_n^c\cap C_n) \le c^n \exp\left(-I_{g,q}(I_n^c\cap C_n) + n\log I_g(I_n^c\cap C_n) + \log(n+1) + n\log 2\right).$$

Proof. For any positive s,

$$S_{g,q}(I_{g,q}(C_n)+s) = \left\{\mathbf{x} : \sum_{1\le i\le n}(g(x_i)+q(x_i)) \le I_{g,q}(C_n)+s\right\}$$

is included in $\{\mathbf{x} : g(x_i)+q(x_i) \le I_{g,q}(C_n)+s, i=1,...,n\}$ which in turn is included in $F = \{\mathbf{x} : g(x_i) - M(x_i) \le (I_{g,q}(C_n)+s), i=1,...,n\}$ by (1.5.14).

Set $H = \{\mathbf{x} := (x_1,...,x_n) : x_i \le 2(I_{g,q}(C_n)+s), i=1,...,n\}$, we will show it holds for a_n large enough

$$F \subset H. \tag{1.5.27}$$

Suppose that for some $\mathbf{x} := (x_1,...,x_n)$ in F, some x_i is larger than $2(I_{g,q}(C_n)+s)$. For a_n large enough, by Lemma 1.5.3, it holds

$$\begin{aligned}x_i &\ge 2(I_{g,q}(C_n)+s) \ge 2\left(ng(a_n) - nN\log g(a_n)\right) \\ &> 2\left(ng(a_n) - \frac{1}{4}ng(a_n)\right) = \frac{3}{2}ng(a_n).\end{aligned}$$

Since $\frac{3}{2}ng(a_n) \ge \frac{3}{2}na_n$ for large n, by (1.5.17) and since $x \to g(x) - N\log g(x)$ is increasing, we have

$$\begin{aligned}g(x_i) - M(x_i) &\ge g(x_i) - N\log g(x_i) \ge g\left(2(I_{g,q}(C_n)+s)\right) - N\log g\left(2(I_{g,q}(C_n)+s)\right) \\ &> g\left(2(I_{g,q}(C_n)+s)\right) - \frac{1}{2}g\left(2(I_{g,q}(C_n)+s)\right) \\ &\ge \frac{1}{2}\left(2(I_{g,q}(C_n)+s)\right) = I_{g,q}(C_n)+s.\end{aligned}$$

Therefore since $\mathbf{x}\in F$, $x_i \le 2(I_{g,q}(C_n)+s)$ for every i, which implicates that (1.5.27) holds. Thus we have

$$S_{g,q}(I_{g,q}(C_n)+s) \subset H,$$

from which we deduce that

$$\begin{aligned} Volume\left(C_n \cap S_{g,q}(I_{g,q}(C_n)+s)\right) &\leq Volume\left(S_{g,q}(I_{g,q}(C_n)+s)\right) \\ &\leq Volume(H) = 2^n (I_{g,q}(C_n)+s)^n. \end{aligned}$$

With this inequality, the upper bound of integration (1.5.13) can be given when $n \to \infty$ through

$$\begin{aligned} \log P(C_n) &= \log c^n \int_{C_n} \exp\left(-\sum_{1\leq i\leq n}(g(x_i)+q(x_i))\right) dx_1...dx_n \\ &= n\log c - I_{g,q}(C_n) + \log \int_0^\infty Volume(C_n \cap S_{g,q}(I_{g,q}(C_n)+s))e^{-s}ds \\ &\leq n\log c - I_{g,q}(C_n) + \log \int_0^\infty \left(I_{g,q}(C_n)+s\right)^n e^{-s}ds + n\log 2. \end{aligned}$$

According to (1.5.8), it holds

$$\int_0^\infty \left(I_{g,q}(C_n)+s\right)^n e^{-s}ds \leq (n+1)I_{g,q}(C_n)^n,$$

Hence we have

$$\begin{aligned} \log P(C_n) &\leq n\log c - I_{g,q}(C_n) + \log\left((n+1)I_{g,q}(C_n)^n\right) + n\log 2 \\ &= n\log c - I_{g,q}(C_n) + n\log I_{g,q}(C_n) + \log(n+1) + n\log 2. \end{aligned}$$

This completes the proof with replacing C_n by $I_n^c \cap C_n$. □

Lemma 1.5.6. *With the same notation and hypothesis as in Theorem 1.4.1, we derive the crude upper bound for* $I_{g_2}(C_n)$

$$I_{g_2}(C_n) \leq ng(a_n) + nN\log g(a_n), \tag{1.5.28}$$

the lower bound for $I_{g_1}(I_n^c \cap C_n)$

$$I_{g_1}(I_n^c \cap C_n) \geq \min\left(F_{g_1}(a_n,\epsilon_n), F_{g_2}(a_n,\epsilon_n)\right) - nN\log g\left(a_n+\epsilon_n\right), \tag{1.5.29}$$

and the upper bound for $\log I_{g_2}(I_n^c \cap C_n)$

$$\log I_{g_2}(I_n^c \cap C_n) \leq \log n + \log(N+1) + \log g\left(a_n + \frac{\epsilon_n}{n-1}\right). \tag{1.5.30}$$

Proof. From (1.5.17) and (1.5.18), there exists some $a_n \in [Y,\infty)$ (say, $a_n > y_2$) such that

$$M(x) \leq \max(N\log g(a_n), N\log g(x)) \tag{1.5.31}$$

holds on $(0,\infty)$. Hence for a_n large enough

$$g_2(x) = g(x) + M(x) \leq g(x) + \max(N\log g(a_n), N\log g(x)),$$

which in turn yields

$$I_{g_2}(C_n) \leq \inf_{\mathbf{x}\in C_n}\left(\sum_{i=1}^n g(x_i) + \sum_{i=1}^n \max(N\log g(a_n), N\log g(x_i))\right). \tag{1.5.32}$$

It holds

$$\inf_{\mathbf{x}\in C_n}\left(\sum_{i=1}^n \max(N\log g(a_n), N\log g(x_i))\right) = nN\log g(a_n) \tag{1.5.33}$$

which implies that

$$\begin{aligned}
&\inf_{\mathbf{x}\in C_n}\left(\sum_{i=1}^n g(x_i) + \sum_{i=1}^n \max(N\log g(a_n), N\log g(x_i))\right)\\
&= \inf_{\mathbf{x}\in C_n}\left(\sum_{i=1}^n g(x_i)\right) + \inf_{\mathbf{x}\in C_n}\left(\sum_{i=1}^n \max(N\log g(a_n), N\log g(x_i))\right)\\
&= \inf_{\mathbf{x}\in C_n}\left(\sum_{i=1}^n g(x_i)\right) + nN\log g(a_n)\\
&= I_g(C_n) + nN\log g(a_n) = ng(a_n) + nN\log g(a_n).
\end{aligned}$$

Thus we obtain the inequality (1.5.28).

We now provide a lower bound of $I_{g_1}(I_n^c\cap C_n)$. Consider the inequality of (1.5.22) in Lemma 1.5.3, where we have showed that h is convex for x large enough; hence, using (1.5.22) when a_n is sufficiently large, it holds

$$I_{g_1}(I_n^c\cap C_n) \geq I_h(I_n^c\cap C_n) = \min\left(F_{h_1}(a_n,\epsilon_n), F_{h_2}(a_n,\epsilon_n)\right),$$

where the second inequality holds from Lemma 1.3.1. By the definition of the function h in (1.5.20), for large x it holds $h(x) = r(x)$ which yields the following lower bound of $I_{g_1}(I_n^c\cap C_n)$

$$I_{g_1}(I_n^c\cap C_n) \geq I_h(I_n^c\cap C_n) = I_r(I_n^c\cap C_n) = \min\left(F_{r_1}(a_n,\epsilon_n), F_{r_2}(a_n,\epsilon_n)\right).$$

By Lemma 1.3.1, it holds

$$\begin{aligned}
F_{r_1}(a_n,\epsilon_n) &= g(a_n+\epsilon_n) + (n-1)g\left(a_n - \frac{1}{n-1}\epsilon_n\right)\\
&\qquad - N\log g(a_n+\epsilon_n) - (n-1)N\log g\left(a_n - \frac{1}{n-1}\epsilon_n\right)\\
&\geq g(a_n+\epsilon_n) + (n-1)g\left(a_n - \frac{1}{n-1}\epsilon_n\right) - nN\log g\left(a_n+\epsilon_n\right),
\end{aligned}$$

by the same way, we have also

$$F_{r_2}(a_n,\epsilon_n) \geq g(a_n-\epsilon_n) + (n-1)g\left(a_n + \frac{1}{n-1}\epsilon_n\right) - nN\log g\left(a_n+\epsilon_n\right),$$

hence (1.5.29) holds.

The method of the estimation of the upper bound of $I_{g_2}(I_n^c\cap C_n)$ is similar to that used for $I_{g_2}(C_n)$ above. In (1.5.32), replace C_n by $I_n^c\cap C_n$; we obtain

$$\begin{aligned}
I_{g_2}(I_n^c\cap C_n) &\leq \inf_{\mathbf{x}\in I_n^c\cap C_n}\left(\sum_{i=1}^n g(x_i) + \sum_{i=1}^n \max(N\log g(a_n), N\log g(x_i))\right)\\
&\leq \inf_{\mathbf{x}\in I_n^c\cap C_n}\left(\sum_{i=1}^n g(x_i) + \sum_{i=1}^n \max\left(N\log g\left(a_n+\frac{\epsilon_n}{n-1}\right), N\log g(x_i)\right)\right).
\end{aligned}$$

Similarly to (1.5.33), it holds

$$\inf_{\mathbf{x}\in I_n^c\cap C_n}\left(\sum_{i=1}^n \max\left(N\log g\left(a_n+\frac{\epsilon_n}{n-1}\right), N\log g(x_i)\right)\right) = nN\log g\left(a_n+\frac{\epsilon_n}{n-1}\right),$$

where equality is attained setting $x_1 = ... = x_{n-1} = a_n + \epsilon_n/(n-1), x_n = a_n - \epsilon_n$. Hence we have, when $n \to \infty$

$$\begin{aligned}
I_{g_2}(I_n^c \cap C_n) &\leq \inf_{\mathbf{x}\in I_n^c\cap C_n}\left(\sum_{i=1}^n g(x_i) + \sum_{i=1}^n \max\left(N\log g\left(a_n+\frac{\epsilon_n}{n-1}\right), N\log g(x_i)\right)\right)\\
&= \inf_{\mathbf{x}\in I_n^c\cap C_n}\sum_{i=1}^n g(x_i) + nN\log g\left(a_n+\frac{\epsilon_n}{n-1}\right)\\
&= I_g(I_n^c\cap C_n) + nN\log g\left(a_n+\frac{\epsilon_n}{n-1}\right)\\
&\leq g(a_n-\epsilon_n) + (n-1)g\left(a_n+\frac{1}{n-1}\epsilon_n\right) + nN\log g\left(a_n+\frac{\epsilon_n}{n-1}\right)\\
&\leq ng\left(a_n+\frac{\epsilon_n}{n-1}\right) + nN\log g\left(a_n+\frac{\epsilon_n}{n-1}\right)\\
&\leq n(N+1)g\left(a_n+\frac{\epsilon_n}{n-1}\right).
\end{aligned}$$

Therefore we obtain (1.5.30). □

Proof of Theorem 1.4.1. We complete the proof of Theorem 1.4.1 by showing that

$$\lim_{n\to\infty}\frac{P(I_n^c\cap C_n)}{P(C_n)} = 0. \tag{1.5.34}$$

Using the upper bound of $P(I_n^c \cap C_n)$ in Lemma 1.5.5, together with the lower bound of $P(C_n)$ in Lemma 1.5.4, we have when a_n is large enough

$$\begin{aligned}
\frac{P(I_n^c\cap C_n)}{P(C_n)} &\leq \exp\Big(-\left(I_{g,q}(I_n^c\cap C_n) - I_{g,q}(C_n)\right) + n\log I_{g,q}(I_n^c\cap C_n)\\
&\qquad + \tau_n + n\log g(a_n) + \log(n+1) + n\log 2\Big)\\
&\leq \exp\left(-\left(I_{g,q}(I_n^c\cap C_n) - I_{g,q}(C_n)\right) + n\log I_{g,q}(I_n^c\cap C_n) + \tau_n + 2n\log g(a_n)\right)\\
&\leq \exp\left(-\left(I_{g_1}(I_n^c\cap C_n) - I_{g_2}(C_n)\right) + n\log I_{g_2}(I_n^c\cap C_n) + \tau_n + 2n\log g(a_n)\right).
\end{aligned}$$

The last inequality holds from (1.5.15) and (1.5.16). Replace $I_{g_1}(I_n^c \cap C_n), I_{g_2}(C_n)$ by the upper bound of (1.5.28) and the lower bound of (1.5.29), respectively, we obtain

$$\begin{aligned}
I_{g_1}(I_n^c\cap C_n) - I_{g_2}(C_n) &\geq \min\left(F_{g_1}(a_n,\epsilon_n), F_{g_2}(a_n,\epsilon_n)\right) - nN\log g\left(a_n+\epsilon_n\right)\\
&\qquad - \left(ng(a_n) + nN\log g(a_n)\right)\\
&= H(a_n,\epsilon_n) - nN\log g\left(a_n+\epsilon_n\right) - nN\log g(a_n)\\
&\geq H(a_n,\epsilon_n) - 2nN\log g\left(a_n+\epsilon_n\right).
\end{aligned} \tag{1.5.35}$$

Under condition (1.4.2), there exists some Q such that $n\log n \leq Qn\log g(a_n)$, which, together with (1.5.30) and (1.5.35), gives

$$\begin{aligned}\frac{P(I_n^c\cap C_n)}{P(C_n)} &\leq \exp\Big(-\left(H(a_n,\epsilon_n)-2nN\log\left(a_n+\epsilon_n\right)\right)+n\log n+n\log(N+1)\\ &\qquad +n\log g\Big(a_n+\frac{\epsilon_n}{n-1}\Big)+\tau_n+2n\log g(a_n)\Big)\\ &\leq \exp\Big(-H(a_n,\epsilon_n)+n(2N+1)\log g\left(a_n+\epsilon_n\right)\\ &\qquad +\tau_n+2n\log g(a_n)+n\log n+n\log(N+1)\Big)\\ &\leq \exp\left(-H(a_n,\epsilon_n)+n(2N+1)\log g\left(a_n+\epsilon_n\right)+\tau_n+2n\log g(a_n)+2n\log n\right)\\ &\leq \exp\left(-H(a_n,\epsilon_n)+n(2N+1)\log g\left(a_n+\epsilon_n\right)+\tau_n+(2Q+2)n\log g(a_n)\right)\\ &\leq \exp\left(-H(a_n,\epsilon_n)+n(2N+2Q+3)\log g\left(a_n+\epsilon_n\right)+\tau_n\right). \qquad (1.5.36)\end{aligned}$$

The second term in the exponent in the last line above and τ_n are both of small order with respect to $H(a_n,\epsilon_n)$. Indeed under condition (1.4.3), when $a_n\to\infty$, it holds

$$\lim_{n\to\infty}\frac{n(2N+2Q+3)\log g\left(a_n+\epsilon_n\right)}{H(a_n,\epsilon_n)}=0. \qquad (1.5.37)$$

For τ_n which is defined in (1.5.25) under conditions (1.4.3), (1.4.4), $nN\log g(a_n)$ and $nG(a_n)$ are both of smaller order than $H(a_n,\epsilon_n)$. As regards to the third term of τ_n, it holds

$$\begin{aligned}nN\log g\left(a_n+\frac{1}{g(a_n)}\right) &= nN\log\left(g\left(a_n+\frac{1}{g(a_n)}\right)-g(a_n)+g(a_n)\right)\\ &\leq nN\log\left(2\max\left(G(a_n),g(a_n)\right)\right)\\ &= nN\log 2+\max\left(nN\log G(a_n),nN\log g(a_n)\right).\end{aligned}$$

Under conditions (1.4.3) and (1.4.4), both $nN\log G(a_n)$ and $nN\log g(a_n)$ are small with respect to $H(a_n,\epsilon_n)$; therefore $nN\log g\left(a_n+1/g(a_n)\right)$ is small with respect to $H(a_n,\epsilon_n)$ when $n\to\infty$. Hence it holds when $n\to\infty$

$$\lim_{n\to\infty}\frac{\tau_n}{H(a_n,\epsilon_n)}=0. \qquad (1.5.38)$$

Finally, (1.5.36), together with (1.5.37) and (1.5.38), implies that (1.5.34) holds. □

Chapter 2

A conditional limit theorem for random walks under extreme deviation

2.1 Introduction

Let $X_1^n := (X_1, ..., X_n)$ denote n independent unbounded real valued random variables and $S_1^n := X_1 + ... + X_n$ denote their sum. The purpose of this paper is to explore the limit distribution of the generic variable X_1 conditioned on extreme deviations (ED) pertaining to S_1^n. By extreme deviation we mean that S_1^n/n is supposed to take values which are going to infinity as n increases. Obviously such events are of infinitesimal probability. Our interest in this question stems from a first result which assesses that under appropriate conditions, when the sequence a_n is such that

$$\lim_{n\to\infty} a_n = \infty$$

then there exists a sequence ε_n which tends to 0 as n tends to infinity such that

$$\lim_{n\to\infty} P\left(\cap_{i=1}^n \left(X_i \in (a_n - \varepsilon_n, a_n + \varepsilon_n)\right)\middle|\, S_1^n/n > a_n\right) = 1$$

which is to say that when the empirical mean takes exceedingly large values, then all the summands share the same behaviour. This result obviously requires a number of hypotheses, which we simply quote as "light tails" type. See Chapter 1 for this result and the connection with earlier related works.

The above result is clearly to be put in relation with the so-called Gibbs conditional Principle which we recall briefly in its simplest form.

Consider the case when the sequence $a_n = a$ is constant with value larger than the expectation of X_1. Hence we consider the behaviour of the summands when $(S_1^n/n > a)$, under a large deviation (LD) condition about the empirical mean. The asymptotic conditional distribution of X_1 given $(S_1^n/n > a)$ is the well known tilted distribution of P_X with parameter t associated to a. Let us introduce some notation to put this in light. The hypotheses to be stated now together with notation are kept throughout the entire paper.

It will be assumed that P_X, which is the distribution of X_1, has a density p with respect to the Lebesgue measure on $\mathbb{R}$. The fact that X_1 has a light tail is captured in the hypothesis that X_1 has a moment generating function

$$\Phi(t) := E \exp tX_1$$

which is finite in a non void neighborhood $\mathcal{N}$ of 0. This fact is usually refered to as a Cramer type condition.

Defined on $\mathcal{N}$ are the following functions. The functions

$$t \to m(t) := \frac{d}{dt} \log \Phi(t)$$

$$t \to s^2(t) := \frac{d}{dt} m(t)$$

$$t \to \mu_j(t) := \frac{d}{dt} s^2(t) \quad , \quad j = 3, 4$$

are the expectation and the three first centered moments of the r.v. $\mathcal{X}_t$ with density

$$\pi_t(x) := \frac{\exp tx}{\Phi(t)} p(x)$$

which is defined on $\mathbb{R}$ and which is the tilted density with parameter t. When Φ is steep, meaning that

$$\lim_{t \to t^+} m(t) = \infty$$

where $t^+ := ess \sup \mathcal{N}$ then m parametrizes the convex hull of the support of P_X. We refer to Barndorff-Nielsen [6] for those properties. As a consequence of this fact, for all a in the support of P_X, it will be convenient to define

$$\pi^a = \pi_t$$

where a is the unique solution of the equation $m(t) = a$.

We now come to some remark on the Gibbs conditional principle in the standard above setting. A phrasing of this principle is:

As n tends to infinity the conditional distribution of X_1 given $(S_1^n/n > a)$ is Π^a, the distribution with density π^a.

Indeed we prefer to state Gibbs principle in a form where the conditioning event is a point condition $(S_1^n/n = a)$. The conditional distribution of X_1 given $(S_1^n/n = a)$ is a well defined distribution and Gibbs conditional principle states that this conditional distribution converges to Π^a as n tends to infinity. In both settings, this convergence holds in total variation norm. We refer to [32] for the local form of the conditioning event; we will mostly be interested in the extension of this form in the present paper.

For all α (depending on n or not) we will denote p_α the density of the random vector X_1^k conditioned upon the local event $(S_1^n = n\alpha)$. The notation $p_\alpha \left(X_1^k = x_1^k\right)$ is sometimes used to denote the value of the density p_α at point x_1^k. The same notation is used when $X_1, ..., X_k$ are sampled under some Π^α, namely $\pi^\alpha(X_1^k = x_1^k)$.

In [16] some extension of the above Gibbs principle has been obtained. When $a_n = a > EX_1$ a second order term provides a sharpening of the conditioned Gibbs principle, stating that

$$\lim_{n \to \infty} \int |p_a(x) - g_a(x)| \, dx = 0 \tag{2.1.1}$$

where

$$g_a(x) := Cp(x)\mathfrak{n}\left(a, s_n^2, x\right). \tag{2.1.2}$$

Hereabove $\mathfrak{n}(a, s_n, x)$ denotes the normal density function at point x with expectation a, with variance s_n^2, and $s_n^2 := s^2(t)(n-1)$. In the above display, C is a normalizing constant. Obviously developing in this display yields

$$g_a(x) = \pi^a(x)\,(1 + o(1))$$

which proves that (2.1.1) is a weak form of Gibbs principle, with some improvement due to the second order term.

The paper is organized as follows. Notation and hypotheses are stated in Section 2.2, along with some necessary facts from asymptotic analysis in the context of light tailed densities. In Section 2.3, the approximations of the expectation and the two first centered moments of the tilted density are given. Section 2.4 states the Edgeworth expansion under extreme normalizing factors. Section 2.5 provides a local Gibbs conditional principle under EDP, namely producing the approximation of the conditional density of $X_1, ..., X_k$ conditionally on $((1/n)\,(X_1 + ... + X_n) = a_n)$ for sequences a_n which tend to infinity, and where k is fixed, independent on n. The approximation is local. This result is further extended to typical paths under the conditional sampling scheme, which in turn provides the approximation in variation norm for the conditional distribution; in this extension, k is equal to 1, although the result clearly also holds for fixed $k > 1$. The method used here follows closely the approach by [16]. The differences between the Gibbs principles in LDP and EDP are discussed. Section 2.6 states similar results in the case when the conditioning event is $((1/n)\,(X_1 + ... + X_n) > a_n)$.

The main tools to be used come from asymptotic analysis and local limit theorems, developed from [36] and [10]; we also have borrowed a number of arguments from [54]. A number of technical lemmas have been postponed to Section 2.7.

2.2 Notation and hypotheses

In this paper, we consider the uniformly bounded density function $p(x)$

$$p(x) = c\exp\Big(-\Big(g(x) - q(x)\Big)\Big) \qquad x \in \mathbb{R}_+, \tag{2.2.1}$$

where c is some positive normalized constant. Define $h(x) := g'(x)$. We assume that there exists some positive constant ϑ , for large x, it holds

$$\sup_{|v-x|<\vartheta x} |q(v)| \leq \frac{1}{x\sqrt{h(x)}}. \tag{2.2.2}$$

The function g is positive and satisfies

$$\frac{g(x)}{x} \longrightarrow \infty, \qquad x \to \infty. \tag{2.2.3}$$

Not all positive g's satisfying (2.2.3) are adapted to our purpose. Regular functions g are defined as follows. We define firstly a subclass R_0 of the family of *slowly varying* function. A function l belongs to R_0 if it can be represented as

$$l(x) = \exp\Big(\int_1^x \frac{\epsilon(u)}{u} du \Big), \qquad x \geq 1, \tag{2.2.4}$$

where $\epsilon(x)$ is twice differentiable and $\epsilon(x) \to 0$ as $x \to \infty$.

We follow the line of Juszczak and Nagaev [54] to describe the assumed regularity conditions of h.

Class R_β : $h(x) \in R_\beta$, if, with $\beta > 0$ and x large enough, $h(x)$ can be represented as

$$h(x) = x^\beta l(x),$$

where $l(x) \in R_0$ and in (2.2.4) $\epsilon(x)$ satisfies

$$\limsup_{x\to\infty} x|\epsilon'(x)| < \infty, \qquad \limsup_{x\to\infty} x^2|\epsilon''(x)| < \infty. \tag{2.2.5}$$

Class R_∞ : Further, $l \in \widetilde{R_0}$, if, in (2.2.4), $l(x) \to \infty$ as $x \to \infty$ and

$$\lim_{x\to\infty} \frac{x\epsilon'(x)}{\epsilon(x)} = 0, \qquad \lim_{x\to\infty} \frac{x^2\epsilon''(x)}{\epsilon(x)} = 0, \tag{2.2.6}$$

and for some $\eta \in (0, 1/8)$

$$\liminf_{x\to\infty} x^\eta \epsilon(x) > 0. \tag{2.2.7}$$

We say that $h \in R_\infty$ if h is increasing and strictly monotone and its inverse function ψ defined through

$$\psi(u) := h^{\leftarrow}(u) := \inf\{x : h(x) \geq u\} \tag{2.2.8}$$

belongs to $\widetilde{R_0}$.

Denote $\mathfrak{R} := R_\beta \cup R_\infty$. In fact, $\mathfrak{R}$ covers one large class of functions, although, R_β and R_∞ are only subsets of *Regularly varying* and *Rapidly varying* functions, respectively.

Remark 2.2.1. *The role of (2.2.4) is to make $h(x)$ smooth enough. Under (2.2.4) the third order derivative of $h(x)$ exists, which is necessary in order to use a Laplace method for the asymptotic evaluation of the moment generating function $\Phi(t)$ as $t \to \infty$, where*

$$\Phi(t) = \int_0^\infty e^{tx} p(x) dx = c \int_0^\infty \exp\Big(K(x,t) + q(x)\Big) dx, \qquad t \in (0, \infty)$$

in which

$$K(x,t) = tx - g(x).$$

If $h \in \mathfrak{R}$, $K(x,t)$ is concave with respect to x and takes its maximum at $\hat{x} = h^{\leftarrow}(t)$. The evaluation of $\Phi(t)$ for large t follows from an expansion of $K(x,t)$ in a neighborhood of $\hat{x}$; this is Laplace's method. This expansion yields

$$K(x,t) = K(\hat{x},t) - \frac{1}{2}h'(\hat{x})\Big(x-\hat{x}\Big)^2 - \frac{1}{6}h''(\hat{x})\Big(x-\hat{x}\Big)^3 + \varepsilon(x,t),$$

where $\varepsilon(x,t)$ is some error term. Conditions (2.2.6) (2.2.7) and (2.2.5) guarantee that $\varepsilon(x,t)$ goes to 0 when t tends to ∞ and x belongs to some neighborhood of $\hat{x}$.

Example 2.2.1. ***Weibull Density.*** *Let p be a Weibull density with shape parameter $k>1$ and scale parameter 1, namely*

$$\begin{aligned} p(x) &= kx^{k-1}\exp(-x^k), \qquad x\geq 0\\ &= k\exp\Big(-\big(x^k-(k-1)\log x\big)\Big). \end{aligned}$$

Take $g(x)=x^k-(k-1)\log x$ and $q(x)=0$. Then it holds

$$h(x)=kx^{k-1}-\frac{k-1}{x}=x^{k-1}\big(k-\frac{k-1}{x^k}\big).$$

Set $l(x)=k-(k-1)/x^k, x\geq 1$, then (2.2.4) holds, namely,

$$l(x)=\exp\Big(\int_1^x\frac{\epsilon(u)}{u}du\Big), \qquad x\geq 1,$$

with

$$\epsilon(x)=\frac{k(k-1)}{kx^k-(k-1)}.$$

The function ϵ is twice differentiable and goes to 0 as $x\to\infty$. Additionally, ϵ satisfies condition (2.2.5). Hence we have shown that $h\in R_{k-1}$.

Example 2.2.2. ***A rapidly varying density.*** *Define p through*

$$p(x)=c\exp(-e^{x-1}), \qquad x\geq 0.$$

Then $g(x)=h(x)=e^{x-1}$ and $q(x)=0$ for all non negative x. We show that $h\in R_\infty$. It holds $\psi(x)=\log x+1$. Since $h(x)$ is increasing and monotone, it remains to show that $\psi(x)\in\widetilde{R_0}$. When $x\geq 1$, $\psi(x)$ admits the representation of (2.2.4) with $\epsilon(x)=1/(\log x+1)$. Also conditions (2.2.6) and (2.2.7) are satisfied. Thus $h\in R_\infty$.

Throughout the paper we use the following notation. When a r.v. X has density p we write $p(X=x)$ instead of $p(x)$. This notation is useful when changing measures. For example $\pi^a(X=x)$ is the density at point x for the variable X generated under π^a, while $p(X=x)$ states for X generated under p. This avoids constant changes of notation.

2.3 Approximations of the expectation and moments of the tilted density

We inherit of the definition of the tilted density π^a defined in Section 2.1, and of the corresponding definitions of the functions m, s^2 and μ_3. Because of (2.2.1) and the various conditions on g those functions are defined as $t\to\infty$. The following Theorem is basic for the proof of the remaining results.

Theorem 2.3.1. *Let $p(x)$ be defined as in (2.2.1) and $h(x)\in\mathfrak{R}$. Denote by*

$$m(t)=\frac{d}{dt}\log\Phi(t), \qquad s^2(t)=\frac{d}{dt}m(t), \qquad \mu_3(t)=\frac{d^3}{dt^3}\log\Phi(t),$$

then with ψ defined as in (2.2.8) it holds as $t \to \infty$

$$m(t) \sim \psi(t), \qquad s^2(t) \sim \psi'(t), \qquad \mu_3(t) \sim \frac{M_6 - 9}{6}\psi''(t),$$

where M_6 is the sixth order moment of standard normal distribution.

Proof. The proof of this result relies on a series of Lemmas. Lemmas (2.7.2), (2.7.3), (2.7.4) and (2.7.5) are used in the proof. Lemma (2.7.1) is instrumental for Lemma (2.7.5). The proof of Theorem 2.3.1 and these Lemmas are postponed to Section 2.7.1. □

Corollary 2.3.1. *Let $p(x)$ be defined as in (2.2.1) and $h(x) \in \mathfrak{R}$. Then it holds as $t \to \infty$*

$$\frac{\mu_3(t)}{s^3(t)} \longrightarrow 0. \tag{2.3.1}$$

Proof. Its proof relies on Theorem 2.3.1 and is also put in Section 2.7.1. □

2.4 Edgeworth expansion under extreme normalizing factors

With π^{a_n} defined through

$$\pi^{a_n}(x) = \frac{e^{tx}p(x)}{\Phi(t)},$$

and t determined by $m(t) = a_n$, define the normalized density of π^{a_n} by

$$\bar{\pi}^{a_n}(x) = s\pi^{a_n}(sx + a_n),$$

where s is defined in Section 2.1 (notice that it depends on a_n here). Denote the n-convolution of $\bar{\pi}^{a_n}(x)$ by $\bar{\pi}_n^{a_n}(x)$, and denote by ρ_n the normalized density of n-convolution $\bar{\pi}_n^{a_n}(x)$,

$$\rho_n(x) := \sqrt{n}\bar{\pi}_n^{a_n}(\sqrt{n}x).$$

The following result extends the local Edgeworth expansion of the distribution of normalized sums of i.i.d. r.v.'s to the present context, where the summands are generated under the density $\bar{\pi}^{a_n}$. Therefore the setting is that of a triangular array of rowwise independent summands; the fact that $a_n \to \infty$ makes the situation unusual. We mainly adapt Feller's proof (Chapter 16, Theorem 2 [36]).

Theorem 2.4.1. *With the above notation, uniformly upon x it holds*

$$\rho_n(x) = \phi(x)\Big(1 + \frac{\mu_3}{6\sqrt{n}s^3}\big(x^3 - 3x\big)\Big) + o\Big(\frac{1}{\sqrt{n}}\Big).$$

where $\phi(x)$ is standard normal density.

proof. The proof of this Theorem relies on Lemmas 2.7.6 and 2.7.7. Its proof and these two Lemmas are postponed to Section 2.7.2. □

2.5 Gibbs' conditional principles under extreme events

We now explore Gibbs conditional principles under extreme events. The first result is a pointwise approximation of the conditional density $p_{a_n}\left(y_1^k\right)$ on $\mathbb{R}^k$ for fixed k. As a by-product we also address the local approximation of p_{A_n} where

$$p_{A_n}\left(y_1^k\right) := p\left(X_1^k = y_1^k \middle| S_1^n > na_n\right).$$

However this local approximation is of poor interest when comparing p_{a_n} to its approximation.

We consider the case $k = 1$. For Y_1 a random variable with density p_{a_n} we first provide a density g_{a_n} on $\mathbb{R}$ such that

$$p_{a_n}(Y_1) = g_{a_n}(Y_1)(1 + R_n)$$

where R_n is a function of the vector Y_1^n which goes to 0 as n tends to infinity. The above statement may also be written as

$$p_{a_n}(y_1) = g_{a_n}(y_1)\left(1 + o_{P_{a_n}}(1)\right) \tag{2.5.1}$$

where P_{a_n} is the joint probability measure of the vector Y_1^n under the condition $(S_1^n = na_n)$. This statement is of a different nature with respect to the above one, since it amounts to prove the approximation on typical realisations under the conditional sampling scheme. We will deduce from (2.5.1) that the L^1 distance between p_{a_n} and g_{a_n} goes to 0 as n tends to infinity. It would be interesting to extend these results to the case when $k = k_n$ is close to n, as done in [16] in all cases from the CLT to the LDP ranges. The extreme deviation case is more envolved, which led us to restrict this study to the case when $k = 1$ (or k fixed, similarly).

2.5.1 A local result in $\mathbb{R}^k$

Fix $y_1^k := (y_1, ..., y_k)$ in $\mathbb{R}^k$ and define $s_i^j := y_i + ... + y_j$ for $1 \le i < j \le k$. Define t through $m(t) = a_n$, similarly, define t_i through

$$m(t_i) := \frac{na_n - s_1^i}{n - i}. \tag{2.5.2}$$

For the sake of brevity, we write m_i instead of $m(t_i)$, and define $s_i^2 := s^2(t_i)$. Consider the following condition

$$\lim_{n\to\infty} \frac{\psi(t)^2}{\sqrt{n}\psi'(t)} = 0, \tag{2.5.3}$$

which can be seen as a growth condition on a_n, avoiding too large increases of this sequence.

For $0 \le i \le k-1$, define z_i through

$$z_i = \frac{m_i - y_{i+1}}{s_i\sqrt{n - i - 1}}.$$

Remark 2.5.1. *Formula (2.5.3) states the precise behaviour of the sequence a_n which defines the present extended Gibbs principle. In the case when the common density $p(x)$ is Weibull with shape parameter k, using Theorem 2.3.1, we obtain $\psi(t) \sim m(t) = a_n$ and $\psi'(t) \sim a_n^{2-k}$. Replace $\psi(t)$ and $\psi'(t)$ in (2.5.3) by these two terms, we have*

$$\lim_{n\to\infty} \frac{a_n^k}{\sqrt{n}} = 0.$$

This rate controls the growth of a_n to infinity.

Lemma 2.5.1. *Assume that $p(x)$ satisfies (2.2.1) and $h(x) \in \mathfrak{R}$. Let t_i be defined in (2.5.2). Assume that $a_n \to \infty$ as $n \to \infty$ and that (2.5.3) holds. Then as $n \to \infty$*

$$\lim_{n\to\infty} \sup_{0\le i\le k-1} z_i = 0, \qquad \text{and} \qquad \lim_{n\to\infty} \sup_{0\le i\le k-1} z_i^2 = o\left(\frac{1}{\sqrt{n}}\right).$$

Proof. When $n \to \infty$, it holds

$$z_i \sim m_i/s_i\sqrt{n-i-1} \sim m_i/(s_i\sqrt{n}).$$

From Theorem 2.3.1, it holds $m(t) \sim \psi(t)$ and $s(t) \sim \sqrt{\psi'(t)}$. Hence we have

$$z_i \sim \frac{\psi(t_i)}{\sqrt{n\psi'(t_i)}}. \tag{2.5.4}$$

By (2.5.2), $m_i \sim m(t)$ as $n \to \infty$. Then

$$m_i \sim \psi(t) = a_n.$$

In addition, $m_i \sim \psi(t_i)$ by Theorem 2.3.1, this implies

$$\psi(t_i) \sim \psi(t). \tag{2.5.5}$$

Case 1: if $h(x) \in R_\beta$. We have $h(x) = x^\beta l_0(x), l_0(x) \in R_0, \beta > 0$. Hence

$$h'(x) = x^{\beta-1} l_0(x)\big(\beta + \epsilon(x)\big),$$

set $x = \psi(u)$, we get

$$h'\Big(\psi(u)\Big) = \Big(\psi(u)\Big)^{\beta-1} l_0\Big(\psi(u)\Big)\Big(\beta + \epsilon\Big(\psi(u)\Big)\Big). \tag{2.5.6}$$

Notice $\psi'(u) = 1/h'\Big(\psi(u)\Big)$, combine (2.5.5) with (2.5.6), we obtain

$$\frac{\psi'(t_i)}{\psi'(t)} = \frac{h'\Big(\psi(t)\Big)}{h'\Big(\psi(t_i)\Big)} = \frac{\Big(\psi(t)\Big)^{\beta-1} l_0\Big(\psi(t)\Big)\Big(\beta + \epsilon\Big(\psi(t)\Big)\Big)}{\Big(\psi(t_i)\Big)^{\beta-1} l_0\Big(\psi(t_i)\Big)\Big(\beta + \epsilon\Big(\psi(t_i)\Big)\Big)} \longrightarrow 1, \tag{2.5.7}$$

where we use the slowly varying propriety of l_0. Thus it holds

$$\psi'(t_i) \sim \psi'(t),$$

which, together with (2.5.5), is put into (2.5.4) to yield

$$z_i \sim \frac{\psi(t)}{\sqrt{n\psi'(t)}}. \tag{2.5.8}$$

Hence we have under condition (2.5.3)

$$z_i^2 \sim \frac{\psi(t)^2}{n\psi'(t)} = \frac{\psi(t)^2}{\sqrt{n}\psi'(t)}\frac{1}{\sqrt{n}} = o\Big(\frac{1}{\sqrt{n}}\Big), \tag{2.5.9}$$

which implies further $z_i \to 0$. Note that the final step is used in order to relax the strength of the growth condition on a_n.

Case 2: if $h(x) \in R_\infty$. By (2.5.2), it holds $m(t_i) \geq m(t)$ as $n \to \infty$. Since the function $t \to m(t)$ is increasing, we have

$$t \leq t_i.$$

Notice the function $x \to \psi(x)$ is also increasing, we get

$$\psi(t_i) \geq \psi(t).$$

The function $x \to \psi^{'}(x)$ is decreasing, since

$$\psi^{''}(x) = -\frac{\psi(x)}{x^2}\epsilon(x)\Big(1+o(1)\Big) < 0 \qquad as \quad x \to \infty. \tag{2.5.10}$$

Therefore as $n \to \infty$

$$\psi'(t) \geq \psi'(t_i) > 0.$$

Perform one Taylor expansion of $\psi(t_i)$ for some $\theta_1 \in (0,1)$

$$\begin{aligned}\psi(t_i) - \psi(t) &= \psi'(t)(t_i - t) + \frac{1}{2}\psi^{''}(t+\theta_1(t_i-t))(t_i-t)^2 \\ &= \frac{\psi(t)\epsilon(t)}{t}(t_i-t) + \frac{1}{2}\psi^{''}(t+\theta_1(t_i-t))(t_i-t)^2. \end{aligned} \tag{2.5.11}$$

By (2.5.5)

$$\frac{\psi(t_i)-\psi(t)}{\psi(t)} \longrightarrow 0,$$

which together with (2.5.10) and (2.5.11) yields

$$\frac{\epsilon(t)}{t}(t_i - t) \longrightarrow 0. \tag{2.5.12}$$

Perform one Taylor expansion of $\psi'(t_i)$ for some $\theta_2 \in (0,1)$

$$\begin{aligned}\psi'(t_i) - \psi'(t) &= \psi^{''}(t)(t_i - t) + \frac{1}{2}\psi^{'''}(t+\theta_2(t_i-t))(t_i-t)^2 \\ &= -\frac{\psi(t)\epsilon(t)}{t^2}(t_i-t)\Big(1+o(1)\Big) + \frac{1}{2}\psi^{'''}(t+\theta_2(t_i-t))(t_i-t)^2,\end{aligned}$$

where the first term goes to 0 as $n \to \infty$ by (2.5.12), and the second term is infinitely small with respect to the first term (see Section 2.7, e.g. (2.7.18)). Hence

$$\psi'(t_i) \sim \psi'(t).$$

The proof is completed by repeating steps (2.5.8) and (2.5.9).

□

Theorem 2.5.1. *With the same notation and hypotheses as in Lemma 2.5.1, it holds*

$$p_{a_n}(y_1^k) = p(X_1^k = y_1^k | S_1^n = na_n) = g_m(y_1^k)\Big(1 + o\Big(\frac{1}{\sqrt{n}}\Big)\Big),$$

with

$$g_m(y_1^k) = \prod_{i=0}^{k-1}\Big(\pi^{m_i}(X_{i+1} = y_{i+1})\Big).$$

Proof. Using Bayes formula,

$$\begin{aligned} p_{a_n}\left(y_1^k\right) &:= p(X_1^k = y_1^k | S_1^n = na_n) \\ &= p(X_1 = y_1 | S_1^n = na_n)\prod_{i=1}^{k-1} p(X_{i+1} = y_{i+1} | X_1^i = y_1^i, S_1^n = na_n) \\ &= \prod_{i=0}^{k-1} p(X_{i+1} = y_{i+1} | S_{i+1}^n = na_n - s_1^i). \end{aligned} \tag{2.5.13}$$

We make use of the following invariance property: for all y_1^k and all $\alpha > 0$

$$p(X_{i+1} = y_{i+1} | X_1^i = y_1^i, S_1^n = na_n) = \pi^\alpha(X_{i+1} = y_{i+1} | X_1^i = y_1^i, S_1^n = na_n)$$

where on the LHS, the r.v's X_1^i are sampled i.i.d. under p and on the RHS, sampled i.i.d. under π^α. It thus holds

$$\begin{aligned} &p(X_{i+1} = y_{i+1} | S_{i+1}^n = na_n - S_1^i) = \pi^{m_i}(X_{i+1} = y_{i+1} | S_{i+1}^n = na_n - s_1^i) \\ &= \pi^{m_i}(X_{i+1} = y_{i+1})\frac{\pi^{m_i}(S_{i+2}^n = na_n - s_1^{i+1})}{\pi^{m_i}(S_{i+1}^n = na_n - s_1^i)} \\ &= \frac{\sqrt{n-i}}{\sqrt{n-i-1}}\pi^{m_i}(X_{i+1} = y_{i+1})\frac{\widetilde{\pi_{n-i-1}}\big(\frac{m_i - y_{i+1}}{s_i\sqrt{n-i-1}}\big)}{\widetilde{\pi_{n-i}}(0)}, \end{aligned} \tag{2.5.14}$$

where $\widetilde{\pi_{n-i-1}}$ is the normalized density of S_{i+2}^n under i.i.d. sampling with the density π^{m_i}; correspondingly, $\widetilde{\pi_{n-i}}$ is the normalized density of S_{i+1}^n under the same sampling. Note that a r.v. with density π^{m_i} has expectation m_i and variance s_i^2.

Write $z_i = \frac{m_i - y_{i+1}}{s_i\sqrt{n-i-1}}$, and perform a third-order Edgeworth expansion of $\widetilde{\pi_{n-i-1}}(z_i)$, using Theorem 2.4.1. It follows

$$\widetilde{\pi_{n-i-1}}(z_i) = \phi(z_i)\Big(1 + \frac{\mu_3^i}{6s_i^3\sqrt{n-1}}(z_i^3 - 3z_i)\Big) + o\Big(\frac{1}{\sqrt{n}}\Big), \tag{2.5.15}$$

The approximation of $\widetilde{\pi_{n-i}}(0)$ is obtained from (2.5.15)

$$\widetilde{\pi_{n-i}}(0) = \phi(0)\Big(1 + o\Big(\frac{1}{\sqrt{n}}\Big)\Big). \tag{2.5.16}$$

Put (2.5.15) and (2.5.16) into (2.5.14) to obtain

$$\begin{aligned} &p(X_{i+1} = y_{i+1} | S_{i+1}^n = na_n - S_1^i) \\ &= \frac{\sqrt{n-i}}{\sqrt{n-i-1}}\pi^{m_i}(X_{i+1} = y_{i+1})\frac{\phi(z_i)}{\phi(0)}\Big[1 + \frac{\mu_3^i}{6s_i^3\sqrt{n-1}}(z_i^3 - 3z_i) + o\Big(\frac{1}{\sqrt{n}}\Big)\Big] \\ &= \frac{\sqrt{2\pi(n-i)}}{\sqrt{n-i-1}}\pi^{m_i}(X_{i+1} = y_{i+1})\phi(z_i)\Big(1 + R_n + o(1/\sqrt{n})\Big), \end{aligned} \tag{2.5.17}$$

where

$$R_n = \frac{\mu_3^i}{6s_i^3\sqrt{n-1}}(z_i^3 - 3z_i).$$

Under condition (2.5.3), using Lemma 2.5.1, it holds $z_i \to 0$ as $a_n \to \infty$, and under Corollary (2.3.1), $\mu_3^i/s_i^3 \to 0$. This yields

$$R_n = o\Big(1/\sqrt{n}\Big),$$

which, combined with (2.5.17), gives

$$p(X_{i+1} = y_{i+1}|s_{i+1}^n = na_n - S_1^i) = \frac{\sqrt{2\pi(n-i)}}{\sqrt{n-i-1}}\pi^{m_i}(X_{i+1} = y_{i+1})\phi(z_i)\Big(1 + o(1/\sqrt{n})\Big)$$
$$= \frac{\sqrt{n-i}}{\sqrt{n-i-1}}\pi^{m_i}(X_{i+1} = y_{i+1})\Big(1 - z_i^2/2 + o(z_i^2)\Big)\Big(1 + o(1/\sqrt{n})\Big),$$

where we use one Taylor expansion in second equality. Using once more Lemma 2.5.1, under conditions (2.5.3), we have as $a_n \to \infty$

$$z_i^2 = o(1/\sqrt{n}),$$

hence we get

$$p(X_{i+1} = y_{i+1}|S_{i+1}^n = na_n - s_1^i) = \frac{\sqrt{n-i}}{\sqrt{n-i-1}}\pi^{m_i}(X_{i+1} = y_{i+1})\Big(1 + o(1/\sqrt{n})\Big),$$

which together with (2.5.13) yields

$$\begin{aligned} p(X_1^k = y_1^k|S_1^n = na_n) &= \prod_{i=0}^{k-1}\Big(\frac{\sqrt{n-i}}{\sqrt{n-i-1}}\pi^{m_i}(X_{i+1} = y_{i+1})\Big(1 + o(1/\sqrt{n})\Big)\Big) \\ &= \prod_{i=0}^{k-1}\Big(\pi^{m_i}(X_{i+1} = y_{i+1})\Big)\prod_{i=0}^{k-1}\Big(\frac{\sqrt{n-i}}{\sqrt{n-i-1}}\Big)\prod_{i=0}^{k-1}\Big(1 + o\Big(\frac{1}{\sqrt{n}}\Big)\Big) \\ &= \Big(1 + o\Big(\frac{1}{\sqrt{n}}\Big)\Big)\prod_{i=0}^{k-1}\Big(\pi^{m_i}(X_{i+1} = y_{i+1})\Big), \end{aligned}$$

The proof is completed. □

In the present case, namely for fixed k, an equivalent statement is

Theorem 2.5.2. *Under the same notation and hypotheses as in the previous Theorem, it holds*

$$p_{a_n}(y_1^k) = p(X_1^k = y_1^k|S_1^n = na_n) = g_{a_n}(y_1^k)\Big(1 + o\Big(\frac{1}{\sqrt{n}}\Big)\Big),$$

with

$$g_{a_n}(y_1^k) = \prod_{i=1}^{k}\Big(\pi^{a_n}(X_i = y_i)\Big).$$

Proof. Using the notations of Theorem 2.5.1, by (2.5.13), we obtain

$$p(X_1^k = y_1^k | S_1^n = na_n) = \prod_{i=0}^{k-1} p(X_{i+1} = y_{i+1} | S_{i+1}^n = na_n - S_1^i). \tag{2.5.18}$$

(2.5.14) is replaced by

$$p(X_{i+1} = y_{i+1} | S_{i+1}^n = na_n - S_1^i) = \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \pi^{a_n}(X_{i+1} = y_{i+1}) \frac{\widetilde{\pi_{n-i-1}^{a_n}}\left(\frac{(i+1)a_n - S_1^{i+1}}{s\sqrt{n-i-1}}\right)}{\widetilde{\pi_{n-i}^{a_n}}\left(\frac{ia_n - S_1^i}{s\sqrt{n-i}}\right)}, \tag{2.5.19}$$

where $\widetilde{\pi_{n-i-1}^{a_n}}$ is the normalized density of S_{i+2}^n under i.i.d. sampling with π^{a_n}, here π^{a_n} has the expectation a_n and variance s. Correspondingly, $\widetilde{\pi_{n-i}^{a_n}}$ is the normalized density of S_{i+1}^n under the same sampling.

Write $z_i = \frac{(i+1)a_n - S_1^{i+1}}{s\sqrt{n-i-1}}$, by Theorem 2.4.1 one three-order Edgeworth expansion yields

$$\widetilde{\pi_{n-i-1}^{a_n}}(z_i) = \phi(z_i)\Big(1 + R_n^i\Big) + o\Big(\frac{1}{\sqrt{n}}\Big), \tag{2.5.20}$$

where

$$R_n^i = \frac{\mu_3}{6s^3\sqrt{n-1}}(z_i^3 - 3z_i).$$

Set $i = i-1$, the approximation of $\widetilde{\pi_{n-i}^{a_n}}$ is obtained from (2.5.20)

$$\widetilde{\pi_{n-i}}(z_{i-1}) = \phi(z_{i-1})\Big(1 + R_n^{i-1}\Big) + o\Big(\frac{1}{\sqrt{n}}\Big). \tag{2.5.21}$$

When $a_n \to \infty$, using Theorem 2.3.1, it holds

$$\begin{aligned} \sup_{0 \le i \le k-1} z_i^2 \sim \frac{(i+1)^2 a_n^2}{s^2 n} \le \frac{2k^2 a_n^2}{s^2 n} = \frac{2k^2 (m(t))^2}{s^2 n} \\ \sim \frac{2k^2(\psi(t))^2}{\psi'(t) n} = \frac{2k^2(\psi(t))^2}{\sqrt{n}\psi'(t)} \frac{1}{\sqrt{n}} = o\Big(\frac{1}{\sqrt{n}}\Big), \end{aligned} \tag{2.5.22}$$

where last step holds under condition (2.5.3). Hence it holds $z_i \to 0$ for $0 \le i \le k-1$ as $a_n \to \infty$, and by Corollary (2.3.1), $\mu_3/s^3 \to 0$, then it follows

$$R_n^i = o\Big(1/\sqrt{n}\Big) \qquad R_n^{i-1} = o\Big(1/\sqrt{n}\Big),$$

then put (2.5.20) and (2.5.21) into (2.5.19), we obtain

$$\begin{aligned} p(X_{i+1} = y_{i+1} | S_{i+1}^n = na_n - S_1^i) &= \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \pi^{a_n}(X_{i+1} = y_{i+1}) \frac{\phi(z_i)}{\phi(z_{i-1})} \Big(1 + o(1/\sqrt{n})\Big) \\ &= \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \pi^{a_n}(X_{i+1} = y_{i+1}) \Big(1 - (z_i^2 - z_{i-1}^2)/2 + o(z_i^2 - z_{i-1}^2)\Big)\Big(1 + o(1/\sqrt{n})\Big), \end{aligned}$$

where we use one Taylor expansion in second equality. Using (2.5.22), we have as $a_n \to \infty$

$$|z_i^2 - z_{i-1}^2| = o(1/\sqrt{n}),$$

hence we get

$$p(X_{i+1}=y_{i+1}|S_{i+1}^n=na_n-S_1^i)=\frac{\sqrt{n-i}}{\sqrt{n-i-1}}\pi^{a_n}(X_{i+1}=y_{i+1})\Big(1+o(1/\sqrt{n})\Big),$$

which together with (2.5.18) yields

$$\begin{aligned} p(X_1^k=y_1^k|S_1^n=na_n) &= \prod_{i=0}^{k-1}\Big(\pi^{a_n}(X_{i+1}=y_{i+1})\sqrt{\frac{n}{n-k}}\Big)\prod_{i=0}^{k-1}\Big(1+o\big(\frac{1}{\sqrt{n}}\big)\Big) \\ &= \Big(1+o\big(\frac{1}{\sqrt{n}}\big)\Big)\prod_{i=0}^{k-1}\Big(\pi^{a_n}(X_{i+1}=y_{i+1})\Big). \qquad (2.5.23) \end{aligned}$$

This completes the proof. □

Remark 2.5.2. *The above result shows that asymptotically the point condition $(S_1^n=na_n)$ leaves blocks of k of the $X_i's$ independent. Obviously this property does not hold for large values of k, close to n. A similar statement holds in the LDP range, conditioning either on $(S_1^n=na)$ (see Diaconis and Friedman 1988)), or on $(S_1^n\geq na)$; see Csiszar 1984 for a general statement on asymptotic conditional independence.*

When $a_n=a$, decide t by $m(t)=a$, use the same proof as Theorem (2.5.2), we obtain the following corollary.

Corollary 2.5.1. *$X_1,...,X_n$ are i.i.d. random variables with density $p(x)$ defined in (2.2.1) and $h(x)\in\mathfrak{R}$. Then it holds*

$$p_a(y_1^k)=p(X_1^k=y_1^k|S_1^n=na)=g_a(y_1^k)\Big(1+o\big(\frac{1}{\sqrt{n}}\big)\Big),$$

with

$$g_a(y_1^k)=\prod_{i=1}^{k}\Big(\pi^a(X_i=y_i)\Big).$$

2.5.2 Strenghtening of the local Gibbs conditional principle

We now turn to a stronger approximation of p_{a_n}. Consider Y_1 with density p_{a_n} and the resulting random variable $p_{a_n}(Y_1)$. We prove the following result

Theorem 2.5.3. *With the same notation and hypotheses as in Theorem 2.5.2, it holds*

$$p_{a_n}(Y_1)=g_{a_n}(Y_1)(1+R_n)$$

where

$$g_{a_n}=\pi^{a_n}$$

the tilted density at point a_n, and where R_n is a function of Y_1^n such that $P_{a_n}(|R_n|>\delta\sqrt{n})\to 0$ as $n\to\infty$ for any positive δ.

This result is of much greater relevance than the previous ones. Indeed under P_{a_n} the r.v. Y_1 may take large values. On the contrary simple approximation of p_{a_n} by g_{a_n} on $\mathbb{R}_+$ only provides some knowledge on p_{a_n} on sets with smaller and smaller probability under p_{a_n}. Also it will be proved that as a consequence of the above result, the L^1 norm between p_{a_n} and g_{a_n} goes to 0 as $n \to \infty$, a result out of reach through the aforementioned results.

In order to adapt the proof of Theorem 2.5.2 to the present setting it is necessary to get some insight on the plausible values of Y_1 under P_{a_n}. It holds

Lemma 2.5.2. *Under P_{a_n} it holds*

$$Y_1 = O_{P_{a_n}}(a_n)$$

Proof. This is a consequence of Markov Inequality:

$$P\left(Y_1 > u \middle| S_1^n = na_n\right) \leq \frac{E\left(Y_1 \middle| S_1^n = na_n\right)}{u} = \frac{a_n}{u}$$

which goes to 0 for all $u = u_n$ such that $\lim_{n\to\infty} u_n/a_n = \infty$. □

We now turn back to the proof of our result, replacing y_{i+1} by Y_1 in (2.5.19). It holds

$$P\left(X_1 = Y_1 \middle| S_1^n = na_n\right) = P(X_1 = Y_1)\frac{P\left(S_2^n = na_n - Y_1\right)}{P\left(S_1^n = na_n\right)}$$

in which the tilting substitution of measures is performed, with tilting density π^{a_n}, followed by normalization. Now if the growth condition (2.5.3) holds, namely

$$\lim_{n\to\infty} \frac{\psi(t)^2}{\sqrt{n}\psi'(t)} = 0$$

with $m(t) = a_n$ it follows that

$$P\left(X_1 = Y_1 \middle| S_1^n = na_n\right) = \pi^{a_n}\left(Y_1\right)\left(1 + R_n\right)$$

as claimed where the order of magnitude of R_n is $o_{P_{a_n}}(1/\sqrt{n})$. We have proved Theorem 2.5.3.

Denote the conditional probabilities by P_{a_n} and G_{a_n} which correspond to the density functions p_{a_n} and g_{a_n}, respectively.

2.5.3 Gibbs principle in variation norm

We now consider the approximation of P_{a_n} by G_{a_n} in variation norm.

The main ingredient is the fact that in the present setting approximation of p_{a_n} by g_{a_n} in probability plus some rate implies approximation of the corresponding measures in variation norm. This approach has been developped in Broniatowski and Caron [16]; we state a first lemma which states that wether two densities are equivalent in probability with small relative error when measured according to the first one, then the same holds under the sampling of the second.

Let $\mathfrak{R}_n$ and $\mathfrak{S}_n$ denote two p.m's on $\mathbb{R}^n$ with respective densities $\mathfrak{r}_n$ and $\mathfrak{s}_n$.

Lemma 2.5.3. *Suppose that for some sequence ϖ_n which tends to 0 as n tends to infinity*

$$\mathfrak{r}_n\left(Y_1^n\right)=\mathfrak{s}_n\left(Y_1^n\right)\left(1+o_{\mathfrak{R}_n}(\varpi_n)\right) \tag{2.5.24}$$

as n tends to ∞. Then

$$\mathfrak{s}_n\left(Y_1^n\right)=\mathfrak{r}_n\left(Y_1^n\right)\left(1+o_{\mathfrak{S}_n}(\varpi_n)\right). \tag{2.5.25}$$

Proof. Denote

$$A_{n,\varpi_n}:=\left\{y_1^n:(1-\varpi_n)\mathfrak{s}_n\left(y_1^n\right)\leq\mathfrak{r}_n\left(y_1^n\right)\leq\mathfrak{s}_n\left(y_1^n\right)(1+\varpi_n)\right\}.$$

It holds for all positive δ

$$\lim_{n\to\infty}\mathfrak{R}_n\left(A_{n,\delta\varpi_n}\right)=1.$$

Write

$$\mathfrak{R}_n\left(A_{n,\delta\varpi_n}\right)=\int\mathbf{1}_{A_{n,\delta\varpi_n}}\left(y_1^n\right)\frac{\mathfrak{r}_n\left(y_1^n\right)}{\mathfrak{s}_n(y_1^n)}\mathfrak{s}_n(y_1^n)dy_1^n.$$

Since

$$\mathfrak{R}_n\left(A_{n,\delta\varpi_n}\right)\leq(1+\delta\varpi_n)\mathfrak{S}_n\left(A_{n,\delta\varpi_n}\right)$$

it follows that

$$\lim_{n\to\infty}\mathfrak{S}_n\left(A_{n,\delta\varpi_n}\right)=1,$$

which proves the claim. □

Applying this Lemma to the present setting yields

$$g_{a_n}\left(Y_1\right)=p_{a_n}\left(Y_1\right)\left(1+o_{G_{a_n}}\left(1/\sqrt{n}\right)\right)$$

as $n\to\infty$, which together with Theorem 2.5.3 implies

$$p_{a_n}\left(Y_1\right)=g_{a_n}\left(Y_1\right)\left(1+o_{G_{a_n}}\left(1/\sqrt{n}\right)\right). \tag{2.5.26}$$

This fact entails, as in Broniatowski and Caron [16]

Theorem 2.5.4. *Under all the notation and hypotheses above the total variation norm between P_{a_n} and G_{a_n} goes to 0 as $n\to\infty$.*

Proof. For all $\delta>0$, let

$$E_\delta:=\left\{y\in\mathbb{R}:\left|\frac{p_{a_n}\left(y\right)-g_{a_n}\left(y\right)}{g_{a_n}\left(y\right)}\right|<\delta\right\}$$

which by Theorem 2.5.3 and (2.5.26) satisfies

$$\lim_{n\to\infty}P_{a_n}\left(E_\delta\right)=\lim_{n\to\infty}G_{a_n}\left(E_\delta\right)=1. \tag{2.5.27}$$

It holds

$$\sup_{C\in\mathcal{B}(\mathbb{R})}\left|P_{a_n}\left(C\cap E_\delta\right)-G_{a_n}\left(C\cap E_\delta\right)\right|\leq\delta\sup_{C\in\mathcal{B}(\mathbb{R})}\int_{C\cap E_\delta}g_{a_n}\left(y\right)dy\leq\delta.$$

By (2.5.27)

$$\sup_{C\in\mathcal{B}(\mathbb{R})}\left|P_{a_n}\left(C\cap E_{k,\delta}\right)-P_{a_n}\left(C\right)\right|<\eta_n$$

and

$$\sup_{C\in\mathcal{B}(\mathbb{R})}\left|G_{a_n}\left(C\cap E_{\delta}\right)-G_{a_n}\left(C\right)\right|<\eta_n$$

for some sequence $\eta_n\to 0$; hence

$$\sup_{C\in\mathcal{B}(\mathbb{R})}\left|P_{a_n}\left(C\right)-G_{a_n}\left(C\right)\right|<\delta+2\eta_n$$

for all positive δ, which proves the claim.

As a consequence, applying Scheffé's Lemma

$$\int\left|p_{a_n}-g_{a_n}\right|dx\to 0\ \text{ as } n\to\infty.$$

□

Remark 2.5.3. *This result is to be paralleled with Theorem 1.6 in Diaconis and Freedman [32] and Proposition 2.15 in Dembo and Zeitouni [30] which provide a rate for this convergence in the LDP range.*

2.5.4 The asymptotic location of X under the conditioned distribution

This section intends to provide some insight on the behaviour of X_1 under the condition $\left(S_1^n=na_n\right)$; this will be extended further on to the case when $\left(S_1^n\geq na_n\right)$ and to be considered in parallel with similar facts developed in [16] for larger values of a_n.

It will be seen that conditionally on $\left(S_1^n=na_n\right)$ the marginal distribution of the sample concentrates around a_n. Let $\mathcal{X}_t$ be a r.v. with density π^{a_n} where $m(t)=a_n$ and a_n satisfies (2.5.3). Recall that $\mathrm{E}\mathcal{X}_t=a_n$ and $\mathrm{Var}\mathcal{X}_t=s^2$. We evaluate the moment generating function of the normalized variable $\left(\mathcal{X}_t-a_n\right)/s$. It holds

$$\log E\exp\lambda\left(\mathcal{X}_t-a_n\right)/s=-\lambda a_n/s+\log\Phi\left(t+\frac{\lambda}{s}\right)-\log\Phi\left(t\right).$$

A second order Taylor expansion in the above display yields

$$\log E\exp\lambda\left(\mathcal{X}_t-a_n\right)/s=\frac{\lambda^2}{2}\frac{s^2\left(t+\frac{\theta\lambda}{s}\right)}{s^2}$$

where $\theta=\theta(t,\lambda)\in(0,1)$. It holds

Lemma 2.5.4. *Under the above hypotheses and notation, for any compact set K,*

$$\lim_{n\to\infty}\sup_{u\in K}\frac{s^2\left(t+\frac{u}{s}\right)}{s^2}=1.$$

Proof. **Case 1:** if $h(t) \in R_\beta$. By Theorem 2.3.1, it holds $s^2 \sim \psi'(t)$ with $\psi(t) \sim t^{1/\beta} l_1(t)$, where l is some slowly varying function. Consider $\psi'(t) = 1/h'\Big(\psi(t)\Big)$, hence by (2.5.6)

$$\begin{aligned}\frac{1}{s^2} &\sim h'\Big(\psi(t)\Big) = \psi(t)^{\beta-1} l_0\Big(\psi(t)\Big)\Big(\beta+\epsilon\Big(\psi(t)\Big)\Big)\\ &\sim \beta t^{1-1/\beta} l_1(t)^{\beta-1} l_0\Big(\psi(t)\Big) = o(t),\end{aligned}$$

where $l_0 \in R_0$. This implies for any $u \in K$

$$\frac{u}{s} = o(\sqrt{t}),$$

which together with (2.5.7) yields

$$\begin{aligned}\frac{s^2\left(t+u/s\right)}{s^2} &\sim \frac{\psi'(t+u/s)}{\psi'(t)} = \frac{\psi(t)^{\beta-1} l_0\Big(\psi(t)\Big)\Big(\beta+\epsilon\Big(\psi(t)\Big)\Big)}{\Big(\psi(t+u/s)\Big)^{\beta-1} l_0\Big(\psi(t+u/s)\Big)\Big(\beta+\epsilon\Big(\psi(t+u/s)\Big)\Big)}\\ &\sim \frac{\psi(t)^{\beta-1}}{\psi(t+u/s)^{\beta-1}} \sim \frac{t^{1-1/\beta} l_1(t)^{\beta-1}}{(t+u/s)^{1-1/\beta} l_1(t+u/s)^{\beta-1}} \longrightarrow 1.\end{aligned}$$

Case 2: if $h(t) \in R_\infty$. Then $\psi(t) \in \widetilde{R_0}$, hence it holds

$$\frac{1}{st} \sim \frac{1}{t\sqrt{\psi'(t)}} = \sqrt{\frac{1}{t\psi(t)\epsilon(t)}} \longrightarrow 0,$$

which last step holds from condition (2.2.7). Hence for any $u \in K$, we get as $n \to \infty$

$$\frac{u}{s} = o(t),$$

thus using the slowly varying propriety of $\psi(t)$ we have

$$\begin{aligned}\frac{s^2\left(t+u/s\right)}{s^2} &\sim \frac{\psi'(t+u/s)}{\psi'(t)} = \frac{\psi(t+u/s)\epsilon(t+u/s)}{t+u/s}\frac{t}{\psi(t)\epsilon(t)}\\ &\sim \frac{\epsilon(t+u/s)}{\epsilon(t)} = \frac{\epsilon(t)+O\Big(\epsilon'(t)u/s\Big)}{\epsilon(t)} \longrightarrow 1,\end{aligned} \tag{2.5.28}$$

where we use one Taylor expansion in the second line, and last step holds from condition (2.2.6). This completes the proof. □

Applying the above Lemma it follows that the normalized r.v's $\left(\mathcal{X}_t - a_n\right)/s$ converge to a standard normal variable $N(0,1)$ in distribution, as $n \to \infty$. This amount to say that

$$\mathcal{X}_t = a_n + sN(0,1) + o_{\Pi^{a_n}}(1).$$

which implies that $\mathcal{X}_t$ concentrates around a_n with rate s. Due to Theorem 2.5.4 the same holds for X_1 under $\left(S_1^n = na_n\right)$.

2.5.5 Differences between Gibbs principle under LDP and under ED

It is of interest to confront the present results with the general form of the Gibbs principle under linear contraints in the LDP range. We recall briefly and somehow unformally the main classical facts in a simple setting similar as the one used in this paper.

Let $X_1, ..., X_n$ denote n i.i.d. real valued r.v's with distribution P and density p and let $f : \mathbb{R} \to \mathbb{R}$ be a measurable function such that $\Phi_f(\lambda) := E \exp \lambda f(X_1)$ is finite for λ in a non void neighborhood of 0 (the so-called Cramer condition). Denote $m_f(\lambda)$ and $s_f^2(\lambda)$ the first and second derivatives of $\log \Phi_f(\lambda)$. Consider the point set condition $E_n := \left(\frac{1}{n}\sum_{i=1}^n f(X_i) = 0\right)$ and let Ω be the set of all probability measures on $\mathbb{R}$ such that $\int f(x)dQ(x) = 0$.

The classical Gibbs conditioning principle writes as follows:

The limiting distribution P^* of X_1 conditioned on the family of events E_n exists and is defined as the unique minimizer of the Kullback-Leibler distance between P and Ω, namely

$$P^* = \arg\min \{K(Q, P), Q \in \Omega\}$$

where

$$K(Q, P) := \int \log \frac{dQ}{dP} dQ$$

whenever Q is absolutely continuous w.r.t. P, and $K(Q, P) = \infty$ otherwise. Also it can be proved that P^* has a density, which is defined through

$$p^*(x) = \frac{\exp \lambda f(x)}{\Phi_f(\lambda)} p(x)$$

with λ the unique solution of the equation $m_f(\lambda) = 0$. Take $f(x) = x - a$ with a fixed to obtain

$$p^*(x) = \pi^a(x)$$

with the current notation of this paper.

Consider now the application of the above result to r.v's $Y_1, ..., Y_n$ with $Y_i := (X_i)^2$ where the $X_i's$ are i.i.d. and are such that the density of the i.i.d. r.v's $Y_i's$ satisfy (2.2.1), where let $h \in R_\beta \cup R_\infty$ with $\beta > 1$. By the Gibbs conditional principle, for *fixed* a, conditionally on $(\sum_{i=1}^n Y_i = na)$ the generic r.v. Y_1 has a non degenerate limit distribution

$$p_Y^*(y) := \frac{\exp ty}{E \exp tY_1} p_Y(y)$$

and the limit density of X_1 under $(\sum_{i=1}^n X_i^2 = na)$ is

$$p_X^*(y) := \frac{\exp tx^2}{E \exp tX_1^2} p_X(y)$$

whereas, when $a_n \to \infty$, Y_1's conditional limit distribution is degenerate and concentrates around a_n. As a consequence the distribution of X_1 under the condition $(\sum_{i=1}^n X_i^2 = na_n)$ concentrates sharply at $-\sqrt{a_n}$ and $+\sqrt{a_n}$.

2.6 EDP under exceedance

The following proposition states the marginally conditional density under condition $A_n = \{S_1^n \geq na_n\}$, we denote this density by p_{A_n} to differentiate it from p_{a_n} which is under condition $\{S_1^n = na_n\}$. For the purpose of proof, we need the following lemma, based on Theorem 6.2.1 of Jensen [49], to provide one asymptotic estimation of tail probability $P(S_1^n \geq na_n)$ and n-convolution density $p(S_1^n/n = u)$ for $u > a_n$.

Define

$$I(x) := xm^{-1}(x) - \log \Phi\Big(m^{-1}(x)\Big). \tag{2.6.1}$$

Lemma 2.6.1. *$X_1, ..., X_n$ are i.i.d. random variables with density $p(x)$ defined in (2.2.1) and $h(x) \in \mathfrak{R}$. Set $m(t) = a_n$. Suppose as $n \to \infty$*

$$\frac{\psi(t)^2}{\sqrt{n}\psi'(t)} \longrightarrow 0, \tag{2.6.2}$$

then it holds

$$P(S_1^n \geq na_n) = \frac{\exp(-nI(a_n))}{\sqrt{2\pi}\sqrt{n}ts(t)}\Big(1 + o\Big(\frac{1}{\sqrt{n}}\Big)\Big). \tag{2.6.3}$$

Let further t_τ be decided by $m(t_\tau) = \tau$ with $\tau \geq a_n$, it then holds

$$p(S_1^n = n\tau) = \frac{\exp(-nI(\tau))}{\sqrt{2\pi}\sqrt{n}s(t_\tau)}\Big(1 + o\Big(\frac{1}{\sqrt{n}}\Big)\Big). \tag{2.6.4}$$

Proof. For the density $p(x)$ defined in (2.2.1), we show $g(x)$ is convex when x is large enough. If $h(x) \in R_\beta$, for x large enough

$$g^{''}(x) = h^{'}(x) = \frac{h(x)}{x}\Big(\beta + \epsilon(x)\Big) > 0. \tag{2.6.5}$$

If $h(x) \in R_\infty$, its reciprocal function $\psi(x) \in \widetilde{R_0}$. Set $x = \psi(v)$, hence

$$g^{''}(x) = h^{'}(x) = \frac{1}{\psi'(v)} = \frac{v}{\psi(v)\epsilon(v)} > 0, \tag{2.6.6}$$

where the inequality holds since $\epsilon(v) > 0$ under condition (2.2.7) when v is large enough. (2.6.5) and (2.6.6) imply that $g(x)$ is convex for x large enough.

Therefore, the density $p(x)$ with $h(x) \in \mathfrak{R}$ satisfies the conditions of Jensen's Theorem 6.2.1 ([49]). Denote by p_n the density of $\bar{X} = (X_1 + ... + X_n)/n$. We obtain with the third order's Edgeworth expansion from formula (2.2.6) of ([49])

$$P(S_1^n \geq na_n) = \frac{\Phi(t)^n \exp(-nta_n)}{\sqrt{n}ts(t)}\Big(B_0(\lambda_n) + O\Big(\frac{\mu_3(t)}{6\sqrt{n}s^3(t)}B_3(\lambda_n)\Big)\Big), \tag{2.6.7}$$

where $\lambda_n = \sqrt{n}ts(t)$, $B_0(\lambda_n)$ and $B_3(\lambda_n)$ are defined by

$$B_0(\lambda_n) = \frac{1}{\sqrt{2\pi}}\Big(1 - \frac{1}{\lambda_n^2} + o(\frac{1}{\lambda_n^2})\Big), \qquad B_3(\lambda_n) \sim -\frac{3}{\sqrt{2\pi}\lambda_n}.$$

We show, under condition (2.6.2), it holds as $a_n \to \infty$

$$\frac{1}{\lambda_n^2} = o\Big(\frac{1}{n}\Big). \tag{2.6.8}$$

Since $n/\lambda_n^2 = 1/(t^2 s^2(t))$, (2.6.8) is equivalent to show

$$t^2 s^2(t) \longrightarrow \infty. \tag{2.6.9}$$

By Theorem 2.3.1, $m(t) \sim \psi(t)$ and $s^2(t) \sim \psi'(t)$, combined with $m(t) = a_n$, it holds $t \sim h(a_n) l_1(a_n)$, where l_1 is some slowly varying function.

If $h \in R_\beta$, notice

$$\psi'(t) = \frac{1}{h'(\psi(t))} = \frac{\psi(t)}{h\Big(\psi(t)\Big)\Big(\beta + \epsilon(\psi(t))\Big)} \sim \frac{a_n}{h(a_n)\Big(\beta + \epsilon(\psi(t))\Big)},$$

hence

$$t^2 s^2(t) \sim h(a_n)^2 l_1(a_n)^2 \frac{a_n}{h(a_n)\Big(\beta + \epsilon(\psi(t))\Big)} = \frac{a_n h(a_n) l_1(a_n)^2}{\beta + \epsilon(\psi(t_n))} \longrightarrow \infty. \tag{2.6.10}$$

If $h \in R_\infty$, then $\psi(t) \in \widetilde{R_0}$, thus

$$t^2 s^2(t) \sim t^2 \frac{\psi(t)\epsilon(t)}{t} = t\psi(t)\epsilon(t) \longrightarrow \infty, \tag{2.6.11}$$

where last step holds from condition (2.2.7). We have showed (2.6.8) , therefore it holds

$$B_0(\lambda_n) = \frac{1}{\sqrt{2\pi}}\Big(1 + o(\frac{1}{n})\Big).$$

By (2.6.9), λ_n goes to ∞ as $a_n \to \infty$, this implies further $B_3(\lambda_n) \to 0$. On the other hand, by (2.3.1) it holds $\mu_3/s^3 \to 0$. Hence we obtain from (2.6.7)

$$P(S_1^n \geq n a_n) = \frac{\Phi(t)^n \exp(-n t a_n)}{\sqrt{2\pi n} t s(t)}\Big(1 + o\Big(\frac{1}{\sqrt{n}}\Big)\Big),$$

which together with (2.6.1) gives (2.6.3).

By Jensen's Theorem 6.2.1 and formula (2.2.4) in [49] it follows uniformly in τ

$$p(S_1^n/n = \tau) = \frac{\sqrt{n}\Phi(t_\tau)^n \exp(-n t_\tau \tau)}{\sqrt{2\pi} s(t_\tau)}\Big(1 + o\Big(\frac{1}{\sqrt{n}}\Big)\Big),$$

which, together with $p(S_1^n = n\tau) = (1/n)p(S_1^n/n = \tau)$, gives (2.6.4). □

Proposition 2.6.1. *$X_1, ..., X_n$ are i.i.d. random variables with density $p(x)$ defined in (2.2.1) and $h(x) \in \mathfrak{R}$. Set $m(t) = a_n$. Suppose as $n \to \infty$*

$$\frac{\psi(t)^2}{\sqrt{n}\psi'(t)} \longrightarrow 0, \tag{2.6.12}$$

and

$$\eta_n \longrightarrow 0 \qquad \text{and} \qquad nm^{-1}(a_n)\eta_n \longrightarrow \infty, \tag{2.6.13}$$

then

$$p_{A_n}(y_1) = p(X_1 = y_1 | S_1^n \geq na_n) = g_{A_n}(y_1)\Big(1 + o\big(\frac{1}{\sqrt{n}}\big)\Big),$$

with

$$g_{A_n}(y_1) = ts(t)e^{nI(a_n)} \int_{a_n}^{a_n+\eta_n} g_\tau(y_1) \exp\Big(- nI(\tau) - \log s(t_\tau)\Big) d\tau,$$

where $g_\tau = \pi^\tau$ *with* t_τ *decided by* $m(t_\tau) = \tau$.

Proof. Denote $p_{A_n}(y_1)$ by the integration of $p_{a_n}(y_1)$

$$\begin{aligned}
p_{A_n}(y_1) &= \int_{a_n}^{\infty} p(X_1 = y_1 | S_1^n = n\tau) p(S_1^n = n\tau | S_1^n \geq na_n) d\tau \\
&= \frac{p(X_1 = y_1)}{P(S_1^n \geq na_n)} \int_{a_n}^{\infty} p(S_2^n = n\tau - y_1) d\tau \\
&= \Big(1 + \frac{P_2}{P_1}\Big) \frac{p(X_1 = y_1)}{P(S_1^n \geq na_n)} \int_{a_n}^{a_n+\eta_n} p(S_2^n = n\tau - y_1) d\tau \\
&= \Big(1 + \frac{P_2}{P_1}\Big) \int_{a_n}^{a_n+\eta_n} p(X_1 = y_1 | S_1^n = n\tau) p(S_1^n = n\tau | S_1^n \geq na_n) d\tau
\end{aligned} \tag{2.6.14}$$

where the second equality is obtained by Bayes formula, and $P_1 = \int_{a_n}^{a_n+\eta_n} p(S_2^n = n\tau - y_1)d\tau$, $P_2 = \int_{a_n+\eta_n}^{\infty} p(S_2^n = n\tau - y_1)d\tau$. In fact P_2 is one infinitely small term with respect to P_1, which is proved below. Further we have

$$\begin{aligned}
P_2 &= \frac{1}{n} P\Big(S_2^n \geq n(a_n + \eta_n) - y_1\Big) = \frac{1}{n} P\Big(S_2^n \geq (n-1)c_n\Big), \\
P_1 + P_2 &= \frac{1}{n} P\Big(S_2^n \geq na_n - y_1\Big) = \frac{1}{n} P\Big(S_2^n \geq (n-1)d_n\Big),
\end{aligned}$$

where $c_n = \Big(n(a_n + \eta_n) - y_1\Big)/(n-1)$ and $d_n = (na_n - y_1)/(n-1)$. Denote $t_{c_n} = m^{-1}(c_n)$ and $t_{d_n} = m^{-1}(d_n)$. Using Lemma (2.6.1), it holds

$$\frac{P_2}{P_1 + P_2} = \Big(1 + o\big(\frac{1}{\sqrt{n}}\big)\Big) \frac{t_{d_n} s(t_{d_n})}{t_{c_n} s(t_{c_n})} \exp\Big(- (n-1)\big(I(c_n) - I(d_n)\big)\Big). \tag{2.6.15}$$

Using the convexity of the function I, it holds

$$\begin{aligned}
\exp\Big(- (n-1)\big(I(c_n) - I(d_n)\big)\Big) &\leq \exp\Big(- (n-1)(c_n - d_n)m^{-1}(d_n)\Big)\Big) \\
&= \exp\Big(- n\eta_n m^{-1}(d_n)\Big)
\end{aligned}$$

Consider $u \to m^{-1}(u)$ is increasing. Since $d_n \geq a_n$ as $a_n \to \infty$, it holds $m^{-1}(d_n) \geq m^{-1}(a_n)$, hence under condition (2.6.13)

$$\exp\Big(- (n-1)\big(I(c_n) - I(d_n)\big)\Big) \leq \exp\Big(- n\eta_n m^{-1}(a_n)\Big) \longrightarrow 0. \tag{2.6.16}$$

Then we show

$$\frac{t_{d_n} s(t_{d_n})}{t_{c_n} s(t_{c_n})} \longrightarrow 1. \tag{2.6.17}$$

By definition, $c_n/d_n \to 1$ as $a_n \to \infty$. If $h \in R_\beta$, by (2.6.10), it holds

$$\Big(\frac{t_{d_n} s(t_{d_n})}{t_{c_n} s(t_{c_n})}\Big)^2 \sim \Big(\frac{d_n h(d_n)}{\beta + \epsilon\big(\psi(d_n)\big)}\Big)^2 \Big(\frac{\beta + \epsilon\big(\psi(c_n)\big)}{c_n h(c_n)}\Big)^2 \sim \Big(\frac{h(d_n)}{h(c_n)}\Big)^2 \longrightarrow 1.$$

If $h \in R_\infty$, by (2.6.11),

$$t^2 s^2(t) \sim t\psi(t)\epsilon(t),$$

hence

$$\Big(\frac{t_{d_n} s(t_{d_n})}{t_{c_n} s(t_{c_n})}\Big)^2 \sim \frac{d_n \psi(d_n)\epsilon(d_n)}{c_n \psi(c_n)\epsilon(c_n)} \sim \frac{\epsilon(d_n)}{\epsilon(c_n)} = \frac{\epsilon\big(c_n - n\eta_n/(n-1)\big)}{\epsilon(c_n)} \longrightarrow 1,$$

where last step holds by using the same argument as in the second line of (2.5.28).

Using (2.6.15), (2.6.16) and (2.6.17), we obtain

$$\frac{P_2}{P_1} = o\big(1\big).$$

Turn back to (2.6.14), $p_{A_n}(y_1)$ can be approximated by

$$p_{A_n}(y_1) = \Big(1 + o\big(1\big)\Big) \int_{a_n}^{a_n+\eta_n} p(X_1 = y_1 | S_1^n = n\tau) p(S_1^n = n\tau | S_1^n \geq n a_n) d\tau. \tag{2.6.18}$$

By Lemma 2.6.1, it follows uniformly when $\tau \in [a_n, a_n + \eta_n]$

$$\begin{aligned} p(S_1^n = n\tau | S_1^n \geq n a_n) &= \frac{p(S_1^n = n\tau)}{P(S_1^n \geq n a_n)} \\ &= \Big(1 + o\big(\frac{1}{\sqrt{n}}\big)\Big) \frac{ts(t)}{s(t_\tau)} \exp\Big(-n(I(\tau) - I(a_n))\Big), \end{aligned} \tag{2.6.19}$$

Inserting (2.6.19) into (2.6.18), we obtain

$$p_{A_n}(y_1) = \Big(1 + o\big(\frac{1}{\sqrt{n}}\big)\Big) ts(t) e^{nI(a_n)} \int_{a_n}^{a_n+\eta_n} g_\tau(y_1) \exp\Big(-nI(\tau) - \log s(t_\tau)\Big) d\tau,$$

this completes the proof. □

2.7 Proofs

2.7.1 Proofs of Theorem 2.3.1 and Corollary 2.3.1

For density functions $p(x)$ defined in (2.2.1) satisfying also $h(x) \in \mathfrak{R}$, denote by $\psi(x)$ the reciprocal function of $h(x)$ and $\sigma^2(v) = \Big(h'(v)\Big)^{-1}$, $v \in \mathbb{R}_+$. For brevity, we write $\hat{x}, \sigma, l$ instead of $\hat{x}(t), \sigma\Big(\psi(t)\Big), l(t)$.

When t is given, $K(x,t)$ attain its maximum at $\hat{x} = \psi(t)$. The fourth order Taylor expansion of $K(x,t)$ on $x \in [\hat{x} - \sigma l, \hat{x} + \sigma l]$ yields

$$K(x,t) = K(\hat{x},t) - \frac{1}{2}h'(\hat{x})\Big(x-\hat{x}\Big)^2 - \frac{1}{6}h''(\hat{x})\Big(x-\hat{x}\Big)^3 + \varepsilon(x,t), \tag{2.7.1}$$

with some $\theta \in (0,1)$

$$\varepsilon(x,t) = -\frac{1}{24}h^{'''}\Big(\hat{x} + \theta(x-\hat{x})\Big)(x-\hat{x})^4. \tag{2.7.2}$$

For proving Theorem 2.3.1 and Corollary 2.3.1, we state firstly the following Lemmas.

Lemma 2.7.1. *For $p(x)$ in (2.2.1), $h(x) \in \mathfrak{R}$, it holds when $t \to \infty$,*

$$\frac{|\log \sigma\Big(\psi(t)\Big)|}{\int_1^t \psi(u)du} \longrightarrow 0. \tag{2.7.3}$$

Proof. If $h(x) \in R_\beta$, by Theorem (1.5.12) of [10], there exists some slowly varying function such that it holds $\psi(x) \sim x^{1/\beta}l_1(x)$. Hence as $t \to \infty$ (see [36], Chapter 8)

$$\int_1^t \psi(u)du \sim t^{1+\frac{1}{\beta}}l_1(t). \tag{2.7.4}$$

On the other hand, $h'(x) = x^{\beta-1}l(x)\Big(\beta + \epsilon(x)\Big)$, thus we have as $x \to \infty$

$$\begin{aligned} |\log \sigma(x)| &= \left|\log \Big(h'(x)\Big)^{-\frac{1}{2}}\right| = \left|\frac{1}{2}\Big((\beta-1)\log x + \log l(x) + \log(\beta + \epsilon(x))\Big)\right| \\ &\leq \frac{1}{2}(\beta+1)\log x, \end{aligned}$$

set $x = \psi(t)$, then when $t \to \infty$, it holds $x < 2t^{1/\beta}l_1(t) < t^{1/\beta+1}$, hence we get

$$|\log \sigma\Big(\psi(t)\Big)| < \frac{(\beta+1)^2}{2\beta}\log t,$$

which, together with (2.7.4), yields (2.7.3).

If $h(x) \in R_\infty$, then by definition $\psi(x) \in \widetilde{R_0}$ is slowly varying as $x \to \infty$, and as $t \to \infty$ (see [36], Chapter 8)

$$\int_1^t \psi(u)du \sim t\psi(t). \tag{2.7.5}$$

Additionally, we have $h'(x)=1/\psi'(t)$ with $x=\psi(t)$, it follows

$$|\log\sigma(x)| = \left|\log\Big(h'(x)\Big)^{-\frac{1}{2}}\right| = \frac{1}{2}|\log\psi'(t)|.$$

Since $\psi(t)\in\widetilde{R_0}$, it holds

$$\begin{aligned}|\log\sigma\Big(\psi(t)\Big)| &= \frac{1}{2}|\log\psi'(t)| = \frac{1}{2}\left|\log\left(\psi(t)\frac{\epsilon(t)}{t}\right)\right| \\ &= \frac{1}{2}\Big|\log\psi(t)+\log\epsilon(t)-\log t\Big| \\ &\leq \log t+\frac{1}{2}|\log\epsilon(t)| \leq 2\log t, \end{aligned} \tag{2.7.6}$$

where last inequality follows from (2.2.7). (2.7.5) and (2.7.6) imply (2.7.3). This completes the proof. □

Lemma 2.7.2. *For $p(x)$ in (2.2.1), $h\in\mathfrak{R}$, then for any varying slowly function $l(t)\to\infty$ as $t\to\infty$, it holds*

$$\sup_{|x|\leq\sigma l} h'''(\hat{x}+x)\sigma^4 l^4 \longrightarrow 0 \qquad as \quad t\to\infty. \tag{2.7.7}$$

Proof. **Case 1:** $h\in R_\beta$. We have $h(x)=x^\beta l_0(x), l_0(x)\in R_0, \beta>0$. Then

$$h^{''}(x)=\beta(\beta-1)x^{\beta-2}l_0(x)+2\beta x^{\beta-1}l_0^{'}(x)+x^\beta l_0^{''}(x). \tag{2.7.8}$$

and

$$h^{'''}(x)=\beta(\beta-1)(\beta-2)x^{\beta-3}l_0(x)+3\beta(\beta-1)x^{\beta-2}l_0^{'}(x)+3\beta x^{\beta-1}l_0^{''}(x)+x^\beta l_0^{'''}(x). \tag{2.7.9}$$

Since $l_0\in R_0$, it is easy to obtain

$$l_0^{'}(x)=\frac{l_0(x)}{x}\epsilon(x), \qquad l_0^{''}(x)=\frac{l_0(x)}{x^2}\Big(\epsilon^2(x)+x\epsilon'(x)-\epsilon(x)\Big), \tag{2.7.10}$$

and

$$l_0^{'''}(x)=\frac{l_0(x)}{x^3}\Big(\epsilon^3(x)+3x\epsilon'(x)\epsilon(x)-3\epsilon^2(x)-2x\epsilon^{'}(x)+2\epsilon(x)+x^2\epsilon^{''}(x)\Big).$$

Under condition (2.2.5), there exists some positive constant Q such that it holds

$$|l_0^{''}(x)|\leq Q\frac{l_0(x)}{x^2}, \qquad |l_0^{'''}(x)|\leq Q\frac{l_0(x)}{x^3},$$

which, together with (2.7.9), yields with some positive constant Q_1

$$|h^{'''}(x)|\leq Q_1\frac{h(x)}{x^3}. \tag{2.7.11}$$

By definition, we have $\sigma^2(x)=1/h^{'}(x)=x/\Big(h(x)(\beta+\epsilon(x))\Big)$, thus it follows

$$\sigma^2=\sigma^2(\hat{x})=\frac{\hat{x}}{h(\hat{x})(\beta+\epsilon(\hat{x}))}=\frac{\psi(t)}{t(\beta+\epsilon(\psi(t)))}=\frac{\psi(t)}{\beta t}\Big(1+o(1)\Big), \tag{2.7.12}$$

this implies $\sigma l = o(\psi(t)) = o(\hat{x})$. Thus we get with (2.7.11)

$$\sup_{|x|\le\sigma l} |h^{'''}(\hat{x}+x)| \le \sup_{|x|\le\sigma l} Q_1 \frac{h(\hat{x}+x)}{(\hat{x}+x)^3} \le Q_2 \frac{t}{\psi^3(t)}, \tag{2.7.13}$$

where Q_2 is some positive constant. Combined with (2.7.12), we obtain

$$\sup_{|x|\le\sigma l} |h^{'''}(\hat{x}+x)|\sigma^4 l^4 \le Q_2 \frac{t}{\psi^3(t)}\sigma^4 l^4 = \frac{Q_2 l^4}{\beta^2 t\psi(t)} \longrightarrow 0,$$

as sought.

Case 2: $h \in R_\infty$. Since $\hat{x} = \psi(t)$, we have $h(\hat{x}) = t$. Thus it holds

$$h'(\hat{x}) = \frac{1}{\psi'(t)} \qquad and \quad h''(\hat{x}) = -\frac{\psi''(t)}{\big(\psi'(t)\big)^3}, \tag{2.7.14}$$

further we get

$$h'''(\hat{x}) = -\frac{\psi^{'''}(t)\psi^{'}(t) - 3\big(\psi^{''}(t)\big)^2}{\big(\psi'(t)\big)^5}. \tag{2.7.15}$$

Notice if $h(\hat{x}) \in R_\infty$, then $\psi(t) \in \widetilde{R_0}$. Therefore we obtain

$$\psi^{'}(t) = \frac{\psi(t)}{t}\epsilon(t), \tag{2.7.16}$$

and

$$\begin{aligned}\psi^{''}(t) &= -\frac{\psi(t)}{t^2}\epsilon(t)\Big(1-\epsilon(t)-\frac{t\epsilon^{'}(t)}{\epsilon(t)}\Big)\\ &= -\frac{\psi(t)}{t^2}\epsilon(t)\Big(1+o(1)\Big) \qquad as \quad t\to\infty,\end{aligned} \tag{2.7.17}$$

where last equality holds from (2.2.6). Using (2.2.6) once again, we have also $\psi^{'''}(t)$

$$\begin{aligned}\psi^{'''}(t) &= \frac{\psi(t)}{t^3}\epsilon(t)\Big(2+\epsilon^2(t)+3t\epsilon^{'}(t)-3\epsilon(t)-\frac{2t\epsilon^{'}(t)}{\epsilon(t)}+\frac{t^2\epsilon^{''}(t)}{\epsilon(t)}\Big)\\ &= \frac{\psi(t)}{t^3}\epsilon(t)\Big(2+o(1)\Big) \qquad as \quad t\to\infty.\end{aligned} \tag{2.7.18}$$

Put (2.7.16) (2.7.17) and (2.7.18) into (2.7.15) we get

$$h^{'''}(\hat{x}) = \frac{t}{\psi^3(t)\epsilon^3(t)}\Big(1+o(1)\Big)$$

Thus by (2.2.7) as $t\to\infty$

$$\begin{aligned}\sup_{|v|\le t/4} h'''\Big(\psi(t+v)\Big) &= \sup_{|v|\le t/4} \frac{t+v}{\psi^3(t+v)\epsilon^3(t+v)}\Big(1+o(1)\Big)\\ &\le \sup_{|v|\le t/4} \frac{2(t+v)^{11/8}}{\psi^3(t+v)} \le \frac{4t^{11/8}}{\psi^3(t)},\end{aligned} \tag{2.7.19}$$

where last inequality holds from the slowly varying propriety: $\psi(t+v) \sim \psi(t)$. Using $\sigma = \Big(h^{'}(\hat{x})\Big)^{-1/2}$, it holds

$$\sup_{|v|\leq t/4} h^{'''}\Big(\psi(t+v)\Big)\sigma^4 \leq \frac{4t^{11/8}}{\psi^3(t)}\frac{1}{(h^{'}(\hat{x}))^2} = \frac{4t^{11/8}}{\psi^3(t)}\frac{\psi^2(t)\epsilon^2(t)}{t^2} = \frac{4\epsilon^2(t)}{\psi(t)t^{5/8}} \longrightarrow 0.$$

Hence for any slowly varying function $l(t) \to \infty$ it holds as $t \to \infty$

$$\sup_{|v|\leq t/4} h^{'''}\Big(\psi(t+v)\Big)\sigma^4 l^4 \longrightarrow 0.$$

Consider $\psi(t) \in \widetilde{R_0}$, thus $\psi(t)$ is increasing, we have the relation

$$\sup_{|v|\leq t/4} h^{'''}\Big(\psi(t+v)\Big) = \sup_{|\zeta|\leq[\zeta_1,\zeta_2]} h^{'''}(\hat{x}+\zeta),$$

where

$$\zeta_1 = \psi(3t/4) - \hat{x}, \qquad \zeta_2 = \psi(5t/4) - \hat{x}.$$

Hence we have showed

$$\sup_{|\zeta|\leq[\zeta_1,\zeta_2]} h^{'''}(\hat{x}+\zeta)\sigma^4 l^4 \longrightarrow 0.$$

For completing the proof, it remains to show

$$\sigma l \leq \min(|\zeta_1|, \zeta_2) \qquad as \quad t \to \infty. \tag{2.7.20}$$

Perform first order Taylor expansion of $\psi(3t/4)$ at t, for some $\alpha \in [0,1]$, it holds

$$\zeta_1 = \psi(3t/4) - \hat{x} = \psi(3t/4) - \psi(t) = -\psi^{'}\Big(t - \alpha t/4\Big)\frac{t}{4} = -\frac{\psi\Big(t-\alpha t/4\Big)}{4-\alpha}\epsilon\Big(t - \alpha t/4\Big),$$

thus using (2.2.7) and slowly varying propriety of $\psi(t)$ we get as $t \to \infty$

$$|\zeta_1| \geq \frac{\psi\Big(t-\alpha t/4\Big)}{4}\epsilon\Big(t-\alpha t/4\Big) \geq \frac{\psi(t)}{5}\epsilon\Big(t-\alpha t/4\Big) \geq \frac{\psi(t)}{5t^{1/8}}. \tag{2.7.21}$$

On the other hand, we have $\sigma = \Big(h^{'}(\hat{x})\Big)^{-1/2} = \Big(\psi(t)\epsilon(t)/t\Big)^{1/2}$, which, together with (2.7.21), yields

$$\frac{\sigma}{|\zeta_1|} \leq 5\sqrt{\frac{\epsilon(t)}{\psi(t)\sqrt{t}}} \longrightarrow 0 \qquad as \quad t \to \infty,$$

which implies for any slowly varying function $l(t)$ it holds $\sigma l = o(|\zeta_1|)$. By the same way, it is easy to show $\sigma l = o(\zeta_2)$. Hence (2.7.20) holds, as sought. □

Lemma 2.7.3. *For $p(x)$ in (2.2.1), $h \in \mathfrak{R}$, then for any varying slowly function $l(t) \to \infty$ as $t \to \infty$, it holds*

$$\sup_{|x|\leq\sigma l} \frac{h'''(\hat{x}+x)}{h''(\hat{x})}\sigma l \longrightarrow 0 \qquad as \quad t \to \infty, \tag{2.7.22}$$

and

$$h^{''}(\hat{x})\sigma^3 l \longrightarrow 0, \qquad h^{''}(\hat{x})\sigma^4 \longrightarrow 0. \tag{2.7.23}$$

Proof. **Case 1:** $h \in R_\beta$. Using (2.7.8) and (2.7.10), we get $h^{''}(x) = \Big(\beta(\beta-1) + o(1)\Big)x^{\beta-2}l_0(x)$ as $x \to \infty$, where $l_0(x) \in R_0$. Hence it holds

$$h^{''}(\hat{x}) = \Big(\beta(\beta-1) + o(1)\Big)\psi(t)^{\beta-2}l_0(\psi(t)), \tag{2.7.24}$$

which, together with (2.7.12) and (2.7.13), yields with some positive constant Q_3

$$\sup_{|x|\le\sigma l}\left|\frac{h'''(\hat{x}+x)}{h''(\hat{x})}\sigma l\right| \le Q_3\frac{t}{\psi^3(t)}\frac{1}{\psi(t)^{\beta-2}l_0(\psi(t))}\sqrt{\frac{\psi(t)}{\beta t}}l = \frac{Q_3}{\sqrt{\beta}}\frac{\sqrt{t}}{\psi(t)^{\beta+1/2}l_0(\psi(t))}l.$$

Notice $\psi(t) \sim t^{1/\beta}l_1(t)$ for some slowly varying function $l_1(t)$, then it holds $\sqrt{t}l = o\Big(\psi(t)^{\beta+1/2}\Big)$. Hence we get (2.7.22).

From (2.7.12) and (2.7.24), we obtain as $t \to \infty$

$$\begin{aligned} h^{''}(\hat{x})\sigma^3 l &= \Big(\beta(\beta-1) + o(1)\Big)\psi(t)^{\beta-2}l_0(\psi(t))\Big(\frac{\psi(t)}{\beta t}\Big)^{3/2} l \\ &= \Big(\beta(\beta-1) + o(1)\Big)\frac{\psi(t)^{\beta-1/2}}{\beta^{3/2}t^{3/2}}l_0(\psi(t))l \\ &\sim \frac{\beta-1}{\beta^{1/2}}\frac{l_1(t)^{\beta-1/2}}{t^{1/2+1/2\beta}}l_0(\psi(t))l \end{aligned} \tag{2.7.25}$$

This implies the first formula of (2.7.23) holds.

Case 2: $h \in R_\infty$. Using (2.7.14) and (2.7.17) we obtain

$$h''(\hat{x}) = -\frac{\psi''(t)}{\Big(\psi'(t)\Big)^3} = \frac{t}{\psi^2(t)\epsilon^2(t)}\Big(1 + o(1)\Big). \tag{2.7.26}$$

Combine (2.7.19) and (2.7.26), using $\sigma = \Big(h^{'}(\hat{x})\Big)^{-1/2}$, we have as $t \to \infty$

$$\sup_{|v|\le t/4}\frac{h^{'''}\Big(\psi(t+v)\Big)}{h^{''}(\hat{x})}\sigma \le \frac{5\epsilon^2(t)t^{3/8}}{\psi(t)}\frac{1}{\sqrt{h^{'}(\hat{x})}} = \frac{5\epsilon(t)^{5/2}}{t^{1/8}\sqrt{\psi(t)}} \to 0,$$

where $\epsilon(t) \to 0$ and $\psi(t)$ varies slowly. Hence for arbitrarily slowly varying function $l(t)$ it holds as $t \to \infty$

$$\sup_{|v|\le t/4}\frac{h^{'''}\Big(\psi(t+v)\Big)}{h^{''}(\hat{x})}\sigma l \longrightarrow 0.$$

Define ζ_1, ζ_2 as in Lemma 2.7.2, we have showed

$$\sup_{|\zeta|\le[\zeta_1,\zeta_2]}\frac{h^{'''}(\hat{x}+\zeta)}{h^{''}(\hat{x})}\sigma l \longrightarrow 0.$$

(2.7.22) is obtained by using (2.7.20). Using (2.7.26), for any slowly varying function, it holds

$$h^{''}(\hat{x})\sigma^3 l \sim \frac{l}{\sqrt{\psi(t)\epsilon(t)t}} \longrightarrow 0.$$

By the same method as proving $h^{''}(\hat{x})\sigma^3 l \to 0$, it is easy to get for **Case 1** and **Case 2**

$$h^{''}(\hat{x})\sigma^4 \longrightarrow 0.$$

Hence the proof. □

Lemma 2.7.4. *For $p(x)$ in (2.2.1), $h \in \mathfrak{R}$, then for any slowly varying function $l(t) \to \infty$ as $t \to \infty$, it holds*

$$\sup_{y\in[-l,l]} \frac{|\xi(\sigma y + \hat{x}, t)|}{h^{''}(\hat{x})\sigma^3} \longrightarrow 0,$$

where $\xi(x,t) = \varepsilon(x,t) + q(x)$.

Proof. A close look to the proof of Lemma 2.7.3, it is straightforward that (2.7.22) can be slightly modified as

$$\sup_{|x|\le\sigma l} \frac{h'''(\hat{x}+x)}{h''(\hat{x})}\sigma l^4 \longrightarrow 0 \qquad as \quad t\to\infty.$$

Hence for $y \in [-l,l]$, by (2.7.2) and Lemma 2.7.3 it holds as $t \to \infty$

$$\left|\frac{\varepsilon(\sigma y + \hat{x}, t)}{h^{''}(\hat{x})\sigma^3}\right| \le \sup_{|x|\le\sigma l}\left|\frac{h'''(\hat{x}+x)}{h''(\hat{x})}\right|\sigma l^4 \longrightarrow 0. \tag{2.7.27}$$

Under condition (2.2.2), set $x = \psi(t)$, we get

$$\sup_{|v-\psi(t)|\le\vartheta\psi(t)} |q(v)| \le \frac{1}{t\sqrt{\psi(t)}}.$$

Then we show

$$\left|\frac{q(\sigma y + \hat{x})}{h''(\hat{x})\sigma^3}\right| \longrightarrow 0. \tag{2.7.28}$$

Case 1: $h \in R_\beta$. We have $h(x) = x^\beta l_0(x), l_0(x) \in R_0, \beta > 0$. Hence

$$h^{'}(x) = x^{\beta-1} l_0(x)\Big(\beta + \epsilon(x)\Big).$$

Notice $\psi^{'}(t) = 1/h^{'}\Big(\psi(t)\Big)$, it holds as $t \to \infty$

$$\begin{aligned}\frac{\sigma l}{\vartheta\psi(t)} &= \frac{\sqrt{\psi^{'}(t)}l}{\vartheta\psi(t)} = \frac{l}{\vartheta\psi(t)\sqrt{h^{'}(\psi(t))}} \\ &= \frac{l}{\vartheta\Big(\psi(t)\Big)^{(\beta+1)/2} l_0\Big(\psi(t)\Big)^{1/2}\Big(\beta+\epsilon\Big(\psi(t)\Big)\Big)^{1/2}} \longrightarrow 0.\end{aligned}$$

It follows

$$\sup_{|v-\psi(t)|\le\sigma l} |q(v)| \le \frac{1}{t\sqrt{\psi(t)}}.$$

Using the above inequality and by the second line of (2.7.25), when $y \in [-l, l]$, it holds as $t \to \infty$

$$\begin{aligned}\left|\frac{q(\sigma y + \hat{x})}{h''(\hat{x})\sigma^3}\right| &\sim \left|q(\sigma y + \hat{x})\left(\frac{\beta - 1}{\sqrt{\beta}}\frac{\psi(t)^{\beta-1/2}}{t^{3/2}}l_0(\psi(t))\right)^{-1}\right| \\ &\leq 2\left|\frac{\sqrt{\beta}}{\beta - 1}\right| \sup_{|v-\psi(t)|\leq \sigma l} |q(v)| \frac{t^{3/2}}{\psi(t)^{\beta-1/2}l_0(\psi(t))} \\ &\leq 2\left|\frac{\sqrt{\beta}}{\beta - 1}\right| \frac{\sqrt{t}}{\psi(t)^{\beta}l_0(\psi(t))} \longrightarrow 0,\end{aligned}$$

where last step holds since $\psi(t) \sim t^{1/\beta}l_1(t)$ for some slowly varying function l_1.

Case 2: $h \in R_\infty$. For any slowly varying function $l(t)$ as $t \to \infty$

$$\frac{\sigma l}{\vartheta\psi(t)} = \frac{\sqrt{\psi'(t)}l}{\vartheta\psi(t)} = \sqrt{\frac{\epsilon(t)}{t\psi(t)}}\frac{l}{\vartheta} \longrightarrow 0,$$

hence

$$\sup_{|v-\psi(t)|\leq \sigma l} |q(v)| \leq \frac{1}{t\sqrt{\psi(t)}}.$$

Using this inequality and (2.7.26), when $y \in [-l, l]$, it holds as $t \to \infty$

$$\left|\frac{q(\sigma y + \hat{x})}{h''(\hat{x})\sigma^3}\right| \leq 2|q(\sigma y + \hat{x})|\sqrt{\psi(t)\epsilon(t)t} \leq 2 \sup_{|v-\psi(t)|\leq \sigma l} |q(v)|\sqrt{\psi(t)\epsilon(t)t} \leq \sqrt{\epsilon(t)/t} \to 0.$$

(2.7.28), together with (2.7.27), completes the proof. □

Lemma 2.7.5. *For $p(x)$ belonging to (2.2.1), $h(x) \in \mathfrak{R}$, $\alpha \in \mathbb{N}$, denote by*

$$\Psi(t, \alpha) := \int_0^\infty (x - \hat{x})^\alpha e^{tx} p(x)dx,$$

then there exists some slowly varying function $l(t)$ such that it holds as $t \to \infty$

$$\Psi(t, \alpha) = c\sigma^{\alpha+1}e^{K(\hat{x},t)}T_1(t, \alpha)\Big(1 + o(1)\Big),$$

where

$$T_1(t, \alpha) = \int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} y^\alpha \exp\Big(-\frac{y^2}{2}\Big)dy - \frac{h''(\hat{x})\sigma^3}{6}\int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} y^{3+\alpha} \exp\Big(-\frac{y^2}{2}\Big)dy.$$

Proof. By (2.7.2) and Lemma 2.7.2, for any slowly varying function $l(t)$ it holds as $t \to \infty$

$$\sup_{|x-\hat{x}|\leq \sigma l} |\varepsilon(x, t)| \to 0.$$

Given a slowly varying function l with $l(t) \to \infty$ and define the interval I_t as follows

$$I_t := \Big(-\frac{l^{1/3}\sigma}{\sqrt{2}}, \frac{l^{1/3}\sigma}{\sqrt{2}}\Big).$$

For large enough τ, when $t \to \infty$ we can partition $\mathbb{R}_+$ as

$$\mathbb{R}_+ = \{x : 0 < x < \tau\} \cup \{x : x \in \hat{x} + I_t\} \cup \{x : x \geq \tau, x \notin \hat{x} + I_t\},$$

where τ large enough such that it holds for $x > \tau$

$$p(x) < 2ce^{-g(x)}. \tag{2.7.29}$$

Obviously, for fixed τ, $\{x : 0 < x < \tau\} \cap \{x : x \in \hat{x} + I_t\} = ‰$ since for large t we have $\min\Big(x : x \in \hat{x} + I_t\Big) \to \infty$ as $t \to \infty$. Hence it holds

$$\begin{aligned}\Psi(t, \alpha) &= \int_0^\tau (x - \hat{x})^\alpha e^{tx} p(x)dx + \int_{x \in \hat{x}+I_t} (x - \hat{x})^\alpha e^{tx} p(x)dx + \int_{x \notin \hat{x}+I_t, x>\tau} (x - \hat{x})^\alpha e^{tx} p(x)dx \\ &:= \Psi_1(t, \alpha) + \Psi_2(t, \alpha) + \Psi_3(t, \alpha). \end{aligned}\tag{2.7.30}$$

We estimate sequentially $\Psi_1(t, \alpha), \Psi_2(t, \alpha), \Psi_3(t, \alpha)$ in **Step 1, Step 2 and Step 3**.

Step 1: Using (2.7.29), for τ large enough, we have

$$\begin{aligned}|\Psi_1(t, \alpha)| &\leq \int_0^\tau |x - \hat{x}|^\alpha e^{tx} p(x)dx \leq 2c \int_0^\tau |x - \hat{x}|^\alpha e^{tx-g(x)} dx \\ &\leq 2c \int_0^\tau \hat{x}^\alpha e^{tx} dx \leq 2ct^{-1}\hat{x}^\alpha e^{t\tau}. \end{aligned}\tag{2.7.31}$$

We show it holds for $h \in \mathfrak{R}$ as $t \to \infty$

$$t^{-1}\hat{x}^\alpha e^{t\tau} = o(\sigma^{\alpha+1} e^{K(\hat{x},t)} h''(\hat{x})\sigma^3). \tag{2.7.32}$$

(2.7.32) is equivalent to

$$\sigma^{-\alpha-4} t^{-1} \hat{x}^\alpha e^{t\tau} \Big(h''(\hat{x})\Big)^{-1} = o(e^{K(\hat{x},t)}),$$

which is implied by

$$\exp\Big(- (\alpha + 4)\log\sigma - \log t + \alpha \log \hat{x} + \tau t - \log h''(\hat{x})\Big) = o(e^{K(\hat{x},t)}).$$

Since $\hat{x} = \psi(t)$, it holds

$$K(\hat{x}, t) = t\psi(t) - g(\psi(t)) = \int_1^t \psi(u)du + \psi(1) - g(1), \tag{2.7.33}$$

where the second equality can be easily verified by the change of variable $u = h(v)$. By Lemma (2.7.1), we know $\log\sigma = o(e^{K(\hat{x},t)})$ as $t \to \infty$. So it remains to show $t = o(e^{K(\hat{x},t)})$, $\log \hat{x} = o(e^{K(\hat{x},t)})$ and $\log h''(\hat{x}) = o(e^{K(\hat{x},t)})$.

If $h(x) \in R_\beta$, by Theorem (1.5.12) of [10], it holds $\psi(x) \sim x^{1/\beta} l_1(x)$ with some slowly varying function $l_1(x)$. (2.7.4) and (2.7.33) yield $t = o(e^{K(\hat{x},t)})$. In addition, $\log \hat{x} = \log \psi(t) \sim \Big((1/\beta)\log t + \log l_1(t)\Big) = o(e^{K(\hat{x},t)})$. By (2.7.24), it holds $\log h''(\hat{x}) = o(t)$. Thus (2.7.32) holds.

If $h(x) \in R_\infty$, $\psi(x) \in \widetilde{R_0}$ is slowly varying as $x \to \infty$. Therefore, by (2.7.5) and (2.7.33), it holds $t = o(e^{K(\hat{x},t)})$ and $\log \hat{x} = \log \psi(t) = o(e^{K(\hat{x},t)})$. Using (2.7.26), we have $\log h^{''}(\hat{x}) \sim \log t - 2\log \hat{x} - 2\log \epsilon(t)$. Under condition (2.2.7), $\log \epsilon(t) = o(t)$, thus it holds $\log h^{''}(\hat{x}) = o(t)$. We get (2.7.32).

(2.7.31) and (2.7.32) yield together

$$|\Psi_1(t,\alpha)| = o(\sigma^{\alpha+1} e^{K(\hat{x},t)} h^{''}(\hat{x})\sigma^3). \tag{2.7.34}$$

Step 2: Notice $\min\Big(x : x \in \hat{x} + I_t\Big) \to \infty$ as $t \to \infty$, which implies both $\varepsilon(x,t)$ and $q(x)$ go to 0 when $x \in \hat{x} + I_t$. By (2.2.1) and (2.7.1), as $t \to \infty$

$$\begin{aligned}\Psi_2(t,\alpha) &= \int_{x\in\hat{x}+I_t} (x-\hat{x})^\alpha c \exp\Big(K(x,t) + q(x)\Big)dx \\ &= \int_{x\in\hat{x}+I_t} (x-\hat{x})^\alpha c \exp\Big(K(\hat{x},t) - \frac{1}{2}h'(\hat{x})\Big(x-\hat{x}\Big)^2 \\ &\qquad - \frac{1}{6}h''(\hat{x})\Big(x-\hat{x}\Big)^3 + \xi(x,t)\Big)dx,\end{aligned}$$

where $\xi(x,t) = \varepsilon(x,t) + q(x)$. Make the change of variable $y = (x-\hat{x})/\sigma$, it holds

$$\Psi_2(t,\alpha) = c\sigma^{\alpha+1}\exp\Big(K(\hat{x},t)\Big)\int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} y^\alpha \exp\Big(-\frac{y^2}{2} - \frac{h^{''}(\hat{x})\sigma^3}{6}y^3 + \xi(\sigma y + \hat{x}, t)\Big)dy. \tag{2.7.35}$$

On $y \in \Big(-l^{1/3}/\sqrt{2}, l^{1/3}/\sqrt{2}\Big)$, by (2.7.23), $|h^{''}(\hat{x})\sigma^3 y^3| \le |h^{''}(\hat{x})\sigma^3 l| \to 0$ as $t \to \infty$. Perform the first order Taylor expansion, as $t \to \infty$

$$\exp\Big(-\frac{h^{''}(\hat{x})\sigma^3}{6}y^3 + \xi(\sigma y + \hat{x}, t)\Big) = 1 - \frac{h^{''}(\hat{x})\sigma^3}{6}y^3 + \xi(\sigma y + \hat{x}, t) + o_1(t,y),$$

where

$$o_1(t,y) = o\Big(-\frac{h^{''}(\hat{x})\sigma^3}{6}y^3 + \xi(\sigma y + \hat{x}, t)\Big).$$

Hence we obtain

$$\begin{aligned}&\int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} y^\alpha \exp\Big(-\frac{y^2}{2} - \frac{h^{''}(\hat{x})\sigma^3}{6}y^3 + \xi(\sigma y + \hat{x}, t)\Big)dy \\ &= \int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} \Big(1 - \frac{h^{''}(\hat{x})\sigma^3}{6}y^3 + \xi(\sigma y + \hat{x}, t) + o_1(t,y)\Big)y^\alpha \exp\Big(-\frac{y^2}{2}\Big)dy \\ &= \int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} y^\alpha \exp\Big(-\frac{y^2}{2}\Big)dy - \frac{h^{''}(\hat{x})\sigma^3}{6}\int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} y^{3+\alpha}\exp\Big(-\frac{y^2}{2}\Big)dy \\ &\qquad + \int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} \Big(\xi(\sigma y + \hat{x}, t) + o_1(t,y)\Big)y^\alpha \exp\Big(-\frac{y^2}{2}\Big)dy.\end{aligned}$$

Define $T_1(t,\alpha)$ and $T_2(t,\alpha)$ as follows

$$\begin{aligned}T_1(t,\alpha) &= \int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} y^\alpha \exp\Big(-\frac{y^2}{2}\Big)dy - \frac{h^{''}(\hat{x})\sigma^3}{6}\int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} y^{3+\alpha}\exp\Big(-\frac{y^2}{2}\Big)dy, \\ T_2(t,\alpha) &= \int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} \Big(\xi(\sigma y + \hat{x}, t) + o_1(t,y)\Big)y^\alpha \exp\Big(-\frac{y^2}{2}\Big)dy.\end{aligned} \tag{2.7.36}$$

For $T_2(t,\alpha)$, it holds

$$\begin{aligned}
|T_2(t,\alpha)| &\leq \int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} \Big(|\xi(\sigma y+\hat{x},t)| + |o_1(t,y)|\Big)|y|^\alpha \exp\Big(-\frac{y^2}{2}\Big)dy \\
&\leq \sup_{y\in[-l,l]} |\xi(\sigma y+\hat{x},t)| \int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} |y|^\alpha \exp\Big(-\frac{y^2}{2}\Big)dy + \int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} |o_1(t,y)||y|^\alpha \exp\Big(-\frac{y^2}{2}\Big)dy \\
&\leq \sup_{y\in[-l,l]} |\xi(\sigma y+\hat{x},t)| \int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} |y|^\alpha \exp\Big(-\frac{y^2}{2}\Big)dy \\
&\qquad + \int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} \Big(\Big|o\Big(\frac{h''(\hat{x})\sigma^3}{6}y^3\Big)\Big| + \Big|o\Big(\xi(\sigma y+\hat{x},t)\Big)\Big|\Big)|y|^\alpha \exp\Big(-\frac{y^2}{2}\Big)dy \\
&\leq 2\sup_{y\in[-l,l]} |\xi(\sigma y+\hat{x},t)| \int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} |y|^\alpha \exp\Big(-\frac{y^2}{2}\Big)dy + |o(h''(\hat{x})\sigma^3)| \int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} |y|^{3+\alpha} \exp\Big(-\frac{y^2}{2}\Big)dy \\
&= |o(h''(\hat{x})\sigma^3)|\Big(\int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} |y|^\alpha \exp\Big(-\frac{y^2}{2}\Big)dy + \int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} |y|^{3+\alpha} \exp\Big(-\frac{y^2}{2}\Big)dy\Big),
\end{aligned}$$

where last equality holds from Lemma 2.7.4. Since the integrals in the last equality are both bounded, it holds as $t\to\infty$

$$T_2(t,\alpha) = o(h''(\hat{x})\sigma^3).$$

When α is even, the second term of $T_1(t,\alpha)$ vanishes. When α is odd, the first term of $T_1(t,\alpha)$ vanishes. $h''(\hat{x})\sigma^3 \to 0$ by (2.7.23), thus $T_1(t,\alpha)$ is at least the same order than $h''(\hat{x})\sigma^3$. It follows as $t\to\infty$

$$T_2(t,\alpha) = o(T_1(t,\alpha)). \tag{2.7.37}$$

Using (2.7.35), (2.7.36) and (2.7.37) we get

$$\Psi_2(t,\alpha) = c\sigma^{\alpha+1}\exp\Big(K(\hat{x},t)\Big)T_1(t,\alpha)\Big(1+o(1)\Big). \tag{2.7.38}$$

Step 3: Given $h\in\mathfrak{R}$, for any t, $K(x,t)$ as a function of x $(x>\tau)$ is concave since

$$K''(x,t) = -h'(x) < 0.$$

Thus we get for $x\notin \hat{x}+I_t$ and $x>\tau$

$$K(x,t)-K(\hat{x},t) \leq \frac{K(\hat{x}+\frac{l^{1/3}\sigma}{\sqrt{2}}sgn(x-\hat{x}),t)-K(\hat{x},t)}{\frac{l^{1/3}\sigma}{\sqrt{2}}sgn(x-\hat{x})}(x-\hat{x}), \tag{2.7.39}$$

where

$$sgn(x-\hat{x}) = \begin{cases} 1 & if \quad x\geq \hat{x}, \\ -1 & if \quad x<\hat{x}. \end{cases}$$

Using (2.7.1), we get

$$K(\hat{x} + \frac{l^{1/3}\sigma}{\sqrt{2}} sgn(x - \hat{x}), t) - K(\hat{x}, t) \le -\frac{1}{8} h'(\hat{x}) l^{2/3} \sigma^2 = -\frac{1}{8} l^{2/3},$$

which, combined with (2.7.39), yields

$$K(x,t) - K(\hat{x},t) \le -\frac{\sqrt{2}}{8} l^{1/3} \sigma^{-1} |x - \hat{x}|.$$

We obtain

$$\begin{aligned}
|\Psi_3(t,\alpha)| &\le 2c \int_{x \notin \hat{x} + I_t, x > \tau} |x - \hat{x}|^\alpha \exp\Big(K(x,t)\Big) dx \\
&\le 2c e^{K(\hat{x},t)} \int_{|x-\hat{x}| > \frac{l^{1/3}\sigma}{\sqrt{2}}} |x - \hat{x}|^\alpha \exp\Big(-\frac{\sqrt{2}}{8} l^{1/3} \sigma^{-1} |x - \hat{x}|\Big) dx \\
&= 2c e^{K(\hat{x},t)} \sigma^{\alpha+1} \int_{|y| > \frac{l^{1/3}}{\sqrt{2}}} |y|^\alpha \exp\Big(-\frac{\sqrt{2}}{8} l^{1/3} |y|\Big) dy \\
&= 2c e^{K(\hat{x},t)} \sigma^{\alpha+1} \int_{|y| > \frac{l^{1/3}}{\sqrt{2}}} \exp\Big(-\frac{\sqrt{2}}{8} l^{1/3} |y| + \alpha \log |y|\Big) dy \\
&= 2c e^{K(\hat{x},t)} \sigma^{\alpha+1} \Big(2 e^{-l^{2/3}/8} \Big(1 + o(1)\Big)\Big),
\end{aligned}$$

where last equality holds when $l \to \infty$ (see e.g. Theorem 4.12.10 of [10]). With (2.7.38), we obtain

$$\Big|\frac{\Psi_3(t,\alpha)}{\Psi_2(t,\alpha)}\Big| \le \frac{8 e^{-l^{2/3}/8}}{|T_1(t,\alpha)|}.$$

In **Step 2**, we know $T_1(t,\alpha)$ has at least the order $h''(\hat{x})\sigma^3$. Hence there exists some positive constant Q and some slowly varying function l_2 with $l_2(t) \to \infty$ such that it holds as $t \to \infty$

$$\Big|\frac{\Psi_3(t,\alpha)}{\Psi_2(t,\alpha)}\Big| \le \frac{Q e^{-l_2^{2/3}/8}}{h''(\hat{x})\sigma^3}.$$

For example, we can take $l_2(t) = (\log t)^3$.

If $h \in R_\beta$, one close look to (2.7.25), it is easy to know $h''(\hat{x})\sigma^3 \ge 1/t^{1+1/(2\beta)}$, with the choice of l_2 as above, we have

$$\Big|\frac{\Psi_3(t,\alpha)}{\Psi_2(t,\alpha)}\Big| \le Q \exp\Big(-l_2^{2/3}/8 + (1 + 1/(2\beta)) \log t\Big) \longrightarrow 0.$$

If $h \in R_\infty$, using (2.7.26), then it holds as $t \to \infty$

$$\begin{aligned}
\Big|\frac{\Psi_3(t,\alpha)}{\Psi_2(t,\alpha)}\Big| &\le 2Q \exp\Big(-l_2^{2/3}/8 + \log \sqrt{t\psi(t)\epsilon(t)}\Big) \\
&= 2Q \exp\Big(-l_2^{2/3}/8 + (1/2)\Big(\log t + \log \psi(t) + \log \epsilon(t)\Big)\Big) \\
&\longrightarrow 0, \qquad (2.7.40)
\end{aligned}$$

where last line holds since $\log \psi(t) = O(\log t)$. The proof is completed by combining (2.7.30), (2.7.34), (2.7.38) and (2.7.40). □

Proof of Theorem 2.3.1. By Lemma 2.7.5, if $\alpha = 0$, it holds $T_1(t,0) \sim \sqrt{2\pi}$ as $t \to \infty$, hence for $p(x)$ defined in (2.2.1), we can approximate X's moment generating function $\Phi(t)$

$$\Phi(t) = \int_0^\infty e^{tx} p(x) dx = c\sqrt{2\pi}\sigma e^{K(\hat{x},t)}\Big(1 + o(1)\Big). \tag{2.7.41}$$

If $\alpha = 1$, it holds as $t \to \infty$,

$$T_1(t,1) = -\frac{h''(\hat{x})\sigma^3}{6} \int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} y^4 \exp\Big(-\frac{y^2}{2}\Big) dy = -\frac{\sqrt{2\pi} h''(\hat{x})\sigma^3}{2}\Big(1 + o(1)\Big),$$

hence we have with $\Psi(t,\alpha)$ defined in Lemma 2.7.5

$$\Psi(t,1) = -c\sqrt{2\pi}\sigma^2 e^{K(\hat{x},t)} \frac{h''(\hat{x})\sigma^3}{2}\Big(1 + o(1)\Big) = -\Phi(t)\frac{h''(\hat{x})\sigma^4}{2}\Big(1 + o(1)\Big), \tag{2.7.42}$$

which, together with the definition of $\Psi(t,\alpha)$, yields

$$\int_0^\infty x e^{tx} p(x) dx = \Psi(t,1) + \hat{x}\Phi(t) = \Big(\hat{x} - \frac{h''(\hat{x})\sigma^4}{2}\Big(1 + o(1)\Big)\Big)\Phi(t). \tag{2.7.43}$$

Hence we get

$$m(t) = \frac{d \log \Phi(t)}{dt} = \frac{\int_0^\infty x e^{tx} p(x) dx}{\Phi(t)} = \hat{x} - \frac{h''(\hat{x})\sigma^4}{2}\Big(1 + o(1)\Big). \tag{2.7.44}$$

By (2.7.23), as $t \to \infty$

$$m(t) \sim \hat{x} = \psi(t). \tag{2.7.45}$$

Set $\alpha = 2$, as $t \to \infty$, it follows

$$\begin{aligned}\Psi(t,2) &= c\sigma^3 e^{K(\hat{x},t)} \int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} y^2 \exp\Big(-\frac{y^2}{2}\Big) dy \Big(1 + o(1)\Big) \\ &= c\sqrt{2\pi}\sigma^3 e^{K(\hat{x},t)}\Big(1 + o(1)\Big) = \sigma^2\Phi(t)\Big(1 + o(1)\Big).\end{aligned} \tag{2.7.46}$$

Using (2.7.42), (2.7.44) and (2.7.46), we have

$$\begin{aligned}&\int_0^\infty \Big(x - m(t)\Big)^2 e^{tx} p(x) dx = \int_0^\infty \Big(x - \hat{x} + \hat{x} - m(t)\Big)^2 e^{tx} p(x) dx \\ &= \int_0^\infty \Big(x - \hat{x}\Big)^2 e^{tx} p(x) dx + 2\Big(\hat{x} - m(t)\Big) \int_0^\infty (x - \hat{x}) e^{tx} p(x) dx + \Big(\hat{x} - m(t)\Big)^2 \Phi(t) \\ &= \Psi(t,2) + 2\Big(\hat{x} - m(t)\Big)\Psi(t,1) + \Big(\hat{x} - m(t)\Big)^2 \Phi(t) \\ &= \sigma^2\Phi(t)\Big(1 + o(1)\Big) - h''(\hat{x})\sigma^4\Big(\Phi(t)\frac{h''(\hat{x})\sigma^4}{2}\Big)\Big(1 + o(1)\Big) + \Big(\frac{h''(\hat{x})\sigma^4}{2}\Big)^2 \Phi(t)\Big(1 + o(1)\Big) \\ &= \sigma^2\Phi(t)\Big(1 + o(1)\Big) - \frac{(h''(\hat{x})\sigma^3)^2}{4}\sigma^2\Phi(t)\Big(1 + o(1)\Big) = \sigma^2\Phi(t)\Big(1 + o(1)\Big),\end{aligned}$$

where last equality holds since $h^{''}(\hat{x})\sigma^3$ goes to 0 by (2.7.23), thus as $t \to \infty$

$$s^2(t) = \frac{d^2 \log \Phi(t)}{dt^2} = \frac{\int_0^\infty \Big(x - m(t)\Big)^2 e^{tx} p(x) dx}{\Phi(t)} \sim \sigma^2 = \psi'(t). \tag{2.7.47}$$

Set $\alpha = 3$, the first term of $T_1(t, 3)$ vanishes, we obtain as $t \to \infty$

$$\begin{aligned} \Psi(t,3) &= -c\sqrt{2\pi}\sigma^4 e^{K(\hat{x},t)} \frac{h^{''}(\hat{x})\sigma^3}{6} \int_{-\frac{l^{1/3}}{\sqrt{2}}}^{\frac{l^{1/3}}{\sqrt{2}}} \frac{1}{\sqrt{2\pi}} y^6 \exp\Big(-\frac{y^2}{2}\Big) dy \\ &= -cM_6\sqrt{2\pi} e^{K(\hat{x},t)} \frac{h^{''}(\hat{x})\sigma^7}{6}\Big(1 + o(1)\Big) = -M_6 \frac{h^{''}(\hat{x})\sigma^6}{6} \Phi(t)\Big(1 + o(1)\Big), \end{aligned} \tag{2.7.48}$$

where M_6 denotes the sixth order moment of standard normal distribution. Using (2.7.42), (2.7.44), (2.7.46) and (2.7.48), we have as $t \to \infty$

$$\begin{aligned} &\int_0^\infty \Big(x - m(t)\Big)^3 e^{tx} p(x) dx = \int_0^\infty \Big(x - \hat{x} + \hat{x} - m(t)\Big)^3 e^{tx} p(x) dx \\ &= \int_0^\infty \Big((x-\hat{x})^3 + 3(x-\hat{x})^2\Big(\hat{x} - m(t)\Big) + 3(x-\hat{x})\Big(\hat{x} - m(t)\Big)^2 + \Big(\hat{x} - m(t)\Big)^3\Big) e^{tx} p(x) dx \\ &= \Psi(t,3) + 3\Big(\hat{x} - m(t)\Big)\Psi(t,2) + 3\Big(\hat{x} - m(t)\Big)^2 \Psi(t,1) + \Big(\hat{x} - m(t)\Big)^3 \Phi(t) \\ &= -M_6 \frac{h^{''}(\hat{x})\sigma^6}{6} \Phi(t)\Big(1 + o(1)\Big) + (3/2) h^{''}(\hat{x})\sigma^4 (\sigma^2 \Phi(t))\Big(1 + o(1)\Big) \\ &\quad - 3\Big(\frac{h^{''}(\hat{x})\sigma^4}{2}\Big)^2 \Phi(t) \frac{h^{''}(\hat{x})\sigma^4}{2}\Big(1 + o(1)\Big) + \Big(\frac{h^{''}(\hat{x})\sigma^4}{2}\Big)^3 \Phi(t)\Big(1 + o(1)\Big) \\ &= \frac{9 - M_6}{6} h^{''}(\hat{x})\sigma^6 \Phi(t)\Big(1 + o(1)\Big) - h^{''}(\hat{x})\sigma^6 \Phi(t) \frac{(h^{''}(\hat{x})\sigma^3)^2}{4}\Big(1 + o(1)\Big) \\ &= \frac{9 - M_6}{6} h^{''}(\hat{x})\sigma^6 \Phi(t)\Big(1 + o(1)\Big), \end{aligned}$$

where last equality holds since $h^{''}(\hat{x})\sigma^3 \to 0$ by (2.7.23). Hence we get as $t \to \infty$

$$\begin{aligned} \mu_3(t) &= \frac{d^3 \log \Phi(t)}{dt^3} = \frac{\int_0^\infty \Big(x - m(t)\Big)^3 e^{tx} p(x) dx}{\Phi(t)} \\ &\sim \frac{9 - M_6}{6} h^{''}(\hat{x})\sigma^6 = -\frac{9 - M_6}{6} \frac{\psi^{''}(t)}{\psi'(t)^3} \psi'(t)^3 = \frac{M_6 - 9}{6} \psi^{''}(t). \end{aligned} \tag{2.7.49}$$

The proof is completed by combining (2.7.45) (2.7.47) with (2.7.49).

□

Proof of Corollary 2.3.1. The proof is immediate by (2.7.23) of Lemma 2.7.3, from which we get $h^{''}(\hat{x})\sigma^3 \to 0$ since $l(t) \to \infty$ as $t \to \infty$. By (2.7.47) and (2.7.49), it holds as $t \to \infty$

$$\frac{\mu_3}{s^3} \sim \frac{9 - M_6}{6} h^{''}(\hat{x})\sigma^3 \longrightarrow 0. \tag{2.7.50}$$

□

2.7.2 Proof of Theorem 2.4.1

Lemma 2.7.6. *With the same notation as in Theorem 2.4.1, it holds*

$$\begin{aligned}&\left|\rho_n(x)-\phi(x)-\frac{\mu_3}{6\sqrt{n}s^3}\Big(x^3-3x\Big)\phi(x)\right|\\ &\leq \frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\Big(\varphi^{a_n}(\tau/\sqrt{n})\Big)^n-e^{-\frac{1}{2}\tau^2}-\frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3e^{-\frac{1}{2}\tau^2}\right|d\tau.\end{aligned}$$

Proof. Let

$$G(x):=\rho_n(x)-\phi(x)-\frac{\mu_3}{6\sqrt{n}s^3}\Big(x^3-3x\Big)\phi(x).$$

From

$$\phi(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\tau x}e^{-\frac{1}{2}\tau^2}d\tau, \tag{2.7.51}$$

it follows that

$$\phi'''(x)=-\frac{1}{2\pi}\int_{-\infty}^{\infty}(i\tau)^3e^{-i\tau x}e^{-\frac{1}{2}\tau^2}d\tau. \tag{2.7.52}$$

On the other hand

$$\phi'''(x)=-(x^3-3x)\phi(x),$$

which, together with (2.7.52), gives

$$(x^3-3x)\phi(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}(i\tau)^3e^{-i\tau x}e^{-\frac{1}{2}\tau^2}d\tau. \tag{2.7.53}$$

Let $\varphi^{a_n}(\tau)$ be the characteristic function (c.f) of $\bar{\pi}^{a_n}$; the c.f of ρ_n is $\Big(\varphi^{a_n}(\tau/\sqrt{n})\Big)^n$. Hence it holds by Fourier inversion theorem

$$\rho_n(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\tau x}\Big(\varphi^{a_n}(\tau/\sqrt{n})\Big)^n d\tau. \tag{2.7.54}$$

Using (2.7.51), (2.7.53) and (2.7.54), we have

$$G(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\tau x}\bigg(\Big(\varphi^{a_n}(\tau/\sqrt{n})\Big)^n-e^{-\frac{1}{2}\tau^2}-\frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3e^{-\frac{1}{2}\tau^2}\bigg)d\tau.$$

Hence the proof is completed. □

Lemma 2.7.7. *With the same notation as in Theorem 2.4.1, the characteristic function φ^{a_n} of $\bar{\pi}^{a_n}(x)$ satisfies*

$$\sup_{a_n\in\mathbb{R}^+}\int|\varphi^{a_n}(\tau)|^2d\tau<\infty \qquad \text{and} \qquad \sup_{a_n\in\mathbb{R}^+,|\tau|\geq\varrho>0}|\varphi^{a_n}(\tau)|<1,$$

for any positive ϱ .

Proof. It is easy to verify that r-order ($r \geq 1$) moment μ^r of $\pi^{a_n}(x)$ satisfies

$$\mu^r(t) = \frac{d^r \log \Phi(t)}{dt^r} \quad \textit{with } t = m^{\leftarrow}(a_n),$$

By Parseval identity

$$\int |\varphi^{a_n}(\tau)|^2 d\tau = 2\pi \int (\bar{\pi}^{a_n}(x))^2 dx \leq 2\pi \sup_{x\in\mathbb{R}} \bar{\pi}^{a_n}(x). \tag{2.7.55}$$

For the density function $p(x)$ in (2.2.1), Theorem 5.4 of Juszczak and Nagaev [54] states that the normalized conjugate density of $p(x)$, namely, $\bar{\pi}^{a_n}(x)$ has the propriety

$$\lim_{a_n\to\infty} \sup_{x\in\mathbb{R}} |\bar{\pi}^{a_n}(x) - \phi(x)| = 0.$$

Thus for arbitrary positive δ, there exists some positive constant M such that it holds

$$\sup_{a_n\geq M} \sup_{x\in\mathbb{R}} |\bar{\pi}^{a_n}(x) - \phi(x)| \leq \delta,$$

which entails that $\sup_{a_n\geq M} \sup_{x\in\mathbb{R}} \bar{\pi}^{a_n}(x) < \infty$. When $a_n < M$, $\sup_{a_n<M} \sup_{x\in\mathbb{R}} \bar{\pi}^{a_n}(x) < \infty$; hence we have

$$\sup_{a_n\in\mathbb{R}^+} \sup_{x\in\mathbb{R}} \bar{\pi}^{a_n}(x) < \infty,$$

which, together with (2.7.55), gives the first inequality of the Lemma. Furthermore, $\varphi^{a_n}(\tau)$ is not periodic, hence the second inequality of the Lemma holds from Lemma 4 (Chapter 15, section 1) of [36]. □

Proof of Theorem 2.4.1. We complete the proof by showing that for n large enough

$$\int_{-\infty}^{\infty} \left| \left(\varphi^{a_n}(\tau/\sqrt{n})\right)^n - e^{-\frac{1}{2}\tau^2} - \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3 e^{-\frac{1}{2}\tau^2} \right| d\tau = o\left(\frac{1}{\sqrt{n}}\right). \tag{2.7.56}$$

For arbitrarily positive sequence a_n we have

$$\sup_{a_n\in\mathbb{R}^+} \left|\varphi^{a_n}(\tau)\right| = \sup_{a_n\in\mathbb{R}^+} \left|\int_{-\infty}^{\infty} e^{i\tau x} \bar{\pi}^{a_n}(x) dx\right| \leq \sup_{a_n\in\mathbb{R}^+} \int_{-\infty}^{\infty} \left|e^{i\tau x} \bar{\pi}^{a_n}(x)\right| dx = 1.$$

In addition, $\pi^{a_n}(x)$ is integrable, by Riemann-Lebesgue theorem, it holds when $|\tau| \to \infty$

$$\sup_{a_n\in\mathbb{R}^+} \left|\varphi^{a_n}(\tau)\right| \longrightarrow 0.$$

Thus for any strictly positive ω, there exists some corresponding N_ω such that if $|\tau| > \omega$, it holds

$$\sup_{a_n\in\mathbb{R}^+} \left|\varphi^{a_n}(\tau)\right| < N_\omega < 1. \tag{2.7.57}$$

We now turn to (2.7.56) which is splitted on $|\tau| > \omega\sqrt{n}$ and on $|\tau| \leq \omega\sqrt{n}$.

It holds

$$\begin{aligned}
&\sqrt{n}\int_{|\tau|>\omega\sqrt{n}}\Big|\Big(\varphi^{a_n}(\tau/\sqrt{n})\Big)^n - e^{-\frac{1}{2}\tau^2} - \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3 e^{-\frac{1}{2}\tau^2}\Big|d\tau \\
&\le \sqrt{n}\int_{|\tau|>\omega\sqrt{n}}\Big|\Big(\varphi^{a_n}(\tau/\sqrt{n})\Big)\Big|^n d\tau + \sqrt{n}\int_{|\tau|>\omega\sqrt{n}}\Big|e^{-\frac{1}{2}\tau^2} + \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3 e^{-\frac{1}{2}\tau^2}\Big|d\tau \\
&\le \sqrt{n}N_\omega^{n-2}\int_{|\tau|>\omega\sqrt{n}}\Big|\Big(\varphi^{a_n}(\tau/\sqrt{n})\Big)\Big|^2 d\tau + \sqrt{n}\int_{|\tau|>\omega\sqrt{n}} e^{-\frac{1}{2}\tau^2}\Big(1+\Big|\frac{\mu_3\tau^3}{6\sqrt{n}s^3}\Big|\Big)d\tau.
\end{aligned} \tag{2.7.58}$$

where the first term of the last line tends to 0 for n large enough, since

$$\begin{aligned}
&\sqrt{n}N_\omega^{n-2}\int_{|\tau|>\omega\sqrt{n}}\Big|\Big(\varphi^{a_n}(\tau/\sqrt{n})\Big)\Big|^2 d\tau \\
&= \exp\Big(\frac{1}{2}\log n + (n-2)\log N_\omega + \log\int_{|\tau|>\omega\sqrt{n}}\Big|\Big(\varphi^{a_n}(\tau/\sqrt{n})\Big)\Big|^2 d\tau\Big) \longrightarrow 0,
\end{aligned} \tag{2.7.59}$$

where the last step holds from Lemma 2.7.7 and (2.7.57). As for the second term of (2.7.58), by Corollary (2.3.1), for n large enough, we have $|\mu_3/s^3| \to 0$. Hence it holds for n large enough

$$\begin{aligned}
&\sqrt{n}\int_{|\tau|>\omega\sqrt{n}} e^{-\frac{1}{2}\tau^2}\Big(1+\Big|\frac{\mu_3\tau^3}{6\sqrt{n}s^3}\Big|\Big)d\tau \\
&\le \sqrt{n}\int_{|\tau|>\omega\sqrt{n}} e^{-\frac{1}{2}\tau^2}|\tau|^3 d\tau = \sqrt{n}\int_{|\tau|>\omega\sqrt{n}} \exp\Big\{-\frac{1}{2}\tau^2 + 3\log|\tau|\Big\}d\tau \\
&= 2\sqrt{n}\exp\Big(-\omega^2 n/2 + o(\omega^2 n/2)\Big) \longrightarrow 0,
\end{aligned} \tag{2.7.60}$$

where the second equality holds from, for example, Chapter 4 of [10]. (2.7.58), (2.7.59) and (2.7.60) implicate that, for n large enough

$$\int_{|\tau|>\omega\sqrt{n}}\Big|\Big(\varphi^{a_n}(\tau/\sqrt{n})\Big)^n - e^{-\frac{1}{2}\tau^2} - \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3 e^{-\frac{1}{2}\tau^2}\Big|d\tau = o\Big(\frac{1}{\sqrt{n}}\Big). \tag{2.7.61}$$

If $|\tau| \le \omega\sqrt{n}$, it holds

$$\begin{aligned}
&\int_{|\tau|\le\omega\sqrt{n}}\Big|\Big(\varphi^{a_n}(\tau/\sqrt{n})\Big)^n - e^{-\frac{1}{2}\tau^2} - \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3 e^{-\frac{1}{2}\tau^2}\Big|d\tau \\
&= \int_{|\tau|\le\omega\sqrt{n}} e^{-\frac{1}{2}\tau^2}\Big|\Big(\varphi^{a_n}(\tau/\sqrt{n})\Big)^n e^{\frac{1}{2}\tau^2} - 1 - \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3\Big|d\tau \\
&= \int_{|\tau|\le\omega\sqrt{n}} e^{-\frac{1}{2}\tau^2}\Big|\exp\Big\{n\log\varphi^{a_n}(\tau/\sqrt{n}) + \frac{1}{2}\tau^2\Big\} - 1 - \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3\Big|d\tau.
\end{aligned} \tag{2.7.62}$$

The integrand in the last display is bounded through

$$|e^\alpha - 1 - \beta| = |(e^\alpha - e^\beta) + (e^\beta - 1 - \beta)| \le (|\alpha-\beta| + \frac{1}{2}\beta^2)e^\gamma, \tag{2.7.63}$$

where $\gamma \ge \max(|\alpha|,|\beta|)$; this inequality follows replacing e^α, e^β by their power series, for real or complex α, β. Denote by

$$\gamma(\tau) = \log\varphi^{a_n}(\tau) + \frac{1}{2}\tau^2.$$

Since $\gamma'(0) = \gamma''(0) = 0$, the third order Taylor expansion of $\gamma(\tau)$ at $\tau = 0$ yields

$$\gamma(\tau) = \gamma(0) + \gamma'(0)\tau + \frac{1}{2}\gamma''(0)\tau^2 + \frac{1}{6}\gamma'''(\xi)\tau^3 = \frac{1}{6}\gamma'''(\xi)\tau^3,$$

where $0 < \xi < \tau$. Hence it holds

$$\left|\gamma(\tau) - \frac{\mu_3}{6s^3}(i\tau)^3\right| = \left|\gamma'''(\xi) - \frac{\mu_3}{s^3}i^3\right|\frac{|\tau|^3}{6}.$$

Here γ''' is continuous; thus we can choose ω small enough such that $|\gamma'''(\xi)| < \rho$ for $|\tau| < \omega$. Meanwhile, for n large enough, according to Corollary (2.3.1) , we have $|\mu_3/s^3| \to 0$. Hence it holds for n large enough

$$\left|\gamma(\tau) - \frac{\mu_3}{6s^3}(i\tau)^3\right| \le \Big(|\gamma'''(\xi)| + \rho\Big)\frac{|\tau|^3}{6} < \rho|\tau|^3. \tag{2.7.64}$$

Choose ω small enough, such that for n large enough it holds for $|\tau| < \omega$

$$\left|\frac{\mu_3}{6s^3}(i\tau)^3\right| \le \frac{1}{4}\tau^2, \qquad |\gamma(\tau)| \le \frac{1}{4}\tau^2.$$

For this choice of ω, when $|\tau| < \omega$ we have

$$\max\left(\left|\frac{\mu_3}{6s^3}(i\tau)^3\right|, |\gamma(\tau)|\right) \le \frac{1}{4}\tau^2. \tag{2.7.65}$$

Replacing τ by $\tau/\sqrt{n}$, it holds for $|\tau| < \omega\sqrt{n}$

$$\begin{aligned}
&\left|n\log\varphi^{a_n}(\tau/\sqrt{n}) + \frac{1}{2}\tau^2 - \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3\right| \\
&= n\left|\log\varphi^{a_n}(\tau/\sqrt{n}) + \frac{1}{2}\left(\frac{\tau}{\sqrt{n}}\right)^2 - \frac{\mu_3}{6s^3}\left(\frac{i\tau}{\sqrt{n}}\right)^3\right| \\
&= n\left|\gamma\left(\frac{\tau}{\sqrt{n}}\right) - \frac{\mu_3}{6s^3}\left(\frac{i\tau}{\sqrt{n}}\right)^3\right| < \frac{\rho|\tau|^3}{\sqrt{n}},
\end{aligned} \tag{2.7.66}$$

where the last inequality holds from (2.7.64). In a similar way, with (2.7.65), it also holds for $|\tau| < \omega\sqrt{n}$

$$\begin{aligned}
&\max\left(\left|n\log\varphi^{a_n}(\tau/\sqrt{n}) + \frac{1}{2}\tau^2\right|, \left|\frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3\right|\right) \\
&= n\max\left(\left|\gamma\left(\frac{\tau}{\sqrt{n}}\right)\right|, \left|\frac{\mu_3}{6s^3}\left(\frac{i\tau}{\sqrt{n}}\right)^3\right|\right) \le \frac{1}{4}\tau^2.
\end{aligned} \tag{2.7.67}$$

Apply (2.7.63) to estimate the integrand of last line of (2.7.62), with the choice of ω in (2.7.64) and (2.7.65), using (2.7.66) and (2.7.67) we have for $|\tau| < \omega\sqrt{n}$

$$\begin{aligned}
&\left|\exp\left\{n\log\varphi^{a_n}(\tau/\sqrt{n}) + \frac{1}{2}\tau^2\right\} - 1 - \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3\right| \\
&\leq \left(\left|n\log\varphi^{a_n}(\tau/\sqrt{n}) + \frac{1}{2}\tau^2 - \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3\right| + \frac{1}{2}\left|\frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3\right|^2\right) \\
&\qquad \times \exp\left[\max\left(\left|n\log\varphi^{a_n}(\tau/\sqrt{n}) + \frac{1}{2}\tau^2\right|, \left|\frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3\right|\right)\right] \\
&\leq \left(\frac{\rho|\tau|^3}{\sqrt{n}} + \frac{1}{2}\left|\frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3\right|^2\right)\exp\left(\frac{\tau^2}{4}\right) \\
&= \left(\frac{\rho|\tau|^3}{\sqrt{n}} + \frac{\mu_3^2\tau^6}{72ns^6}\right)\exp\left(\frac{\tau^2}{4}\right).
\end{aligned}$$

Use this upper bound to (2.7.62), we obtain

$$\begin{aligned}
&\int_{|\tau|\leq\omega\sqrt{n}}\left|\left(\varphi^{a_n}(\tau/\sqrt{n})\right)^n - e^{-\frac{1}{2}\tau^2} - \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3 e^{-\frac{1}{2}\tau^2}\right|d\tau \\
&\leq \int_{|\tau|\leq\omega\sqrt{n}}\exp\left(-\frac{\tau^2}{4}\right)\left(\frac{\rho|\tau|^3}{\sqrt{n}} + \frac{\mu_3^2\tau^6}{72ns^6}\right)d\tau \\
&= \frac{\rho}{\sqrt{n}}\int_{|\tau|\leq\omega\sqrt{n}}\exp\left(-\frac{\tau^2}{4}\right)|\tau|^3 d\tau + \frac{\mu_3^2}{72ns^6}\int_{|\tau|\leq\omega\sqrt{n}}\exp\left(-\frac{\tau^2}{4}\right)\tau^6 d\tau,
\end{aligned}$$

where both the first integral and the second integral are finite, and ρ is arbitrarily small; additionally, by Corollary (2.3.1), $\mu_3^2/s^6 \to 0$ when n large enough, hence it holds for n large enough

$$\int_{|\tau|\leq\omega\sqrt{n}}\left|\left(\varphi^{a_n}(\tau/\sqrt{n})\right)^n - e^{-\frac{1}{2}\tau^2} - \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3 e^{-\frac{1}{2}\tau^2}\right|d\tau = o\left(\frac{1}{\sqrt{n}}\right). \tag{2.7.68}$$

Now (2.7.61) and (2.7.68) give (2.7.56). Further, using (2.7.56) and Lemma 2.7.6, we obtain

$$\left|\rho_n(x) - \phi(x) - \frac{\mu_3}{6\sqrt{n}s^3}\left(x^3 - 3x\right)\phi(x)\right| = o\left(\frac{1}{\sqrt{n}}\right),$$

which concludes the proof. □

Chapter 3

Weighted sampling, MLE and minimum divergence estimator

3.1 Motivation and context

This paper explores Maximum Likelihood (ML) paradigm in the context of sampling. It mainly quotes that inference criterion is strongly connected with the sampling scheme generating the data. Under a given model, when i.i.d. sampling is considered and some standard regularity is assumed, then the Maximum Likelihood principle loosely states that conditionally upon the observed data, resampling under the same i.i.d. scheme should resemble closely to the initial sample only when the resampling distribution is close to the initial unknown one.

Keeping the same definition it appears that under other sampling schemes, the Maximum Likelihood Principle yields a wide range of statistical procedures. Those have in common with the classical simple i.i.d. sampling case that they can be embedded in a natural class of methods based on minimization of $\phi-$divergences between the empirical measure of the data and the model. In the classical i.i.d. case the divergence is the Kullback-Leibler one, which yields the standard form of the Likelihood function. In the case of the weighted bootstrap, the divergence to be optimized is directly related to the distribution of the weights.

This paper discusses the choice of an inference criterion in parametric setting. We consider a wide range of commonly used statistical criterions, namely all those induced by the so-called power divergence, including therefore Maximum Likelihood, Kullback-Leibler, Chi-square, Hellinger distance, etc. The steps of the discussion are as follows.

We first insert Maximum Likelihood paradigm at the center of the scene, putting forwards its strong connection with large deviation probabilities for the empirical measure. The argument can be sketched as follows: for any putative θ in the parameter set, consider n virtual simulated r.v's $X_{i,\theta}$ with corresponding empirical measure $P_{n,\theta}$. Evaluate the probability that $P_{n,\theta}$ is close to P_n , conditionally on P_n, the empirical measure pertaining to the observed data; such statement is refered to as a conditional Sanov theorem, and for any θ this probability is governed by the Kullback-Leibler distance between P_θ and P_{θ_T} where θ_T stands for the true value of the parameter. Estimate this probability for any θ, obviously based on the observed data. Optimize in θ; this provides the MLE, as shown in the two cases of the i.i.d.

sample scheme; our first example is the case when the observations take values in a finite set, and the second case (infinite case), helps to set the arguments to be put forwards. Introducing MLE's through Large deviations for the empirical measure is in the vein of various recent approaches; see Grendar and Judge [43].

We next consider a generalized sampling scheme inherited from the bootstrap, which we call weighted sampling; it amounts to introduce a family of i.i.d. weights $W_1, ..., W_n$ with mean and variance 1. The corresponding empirical measure pertaining to the data set $x_1, ..., x_n$ is just the weighted empirical measure. The MLE is defined through a similar procedure as just evoqued. The conditional Sanov Theorem is governed by a divergence criterion which is defined through the distribution of the weights. Hence MLE results in the optimization of a divergence measure between distributions in the model and the weighted empirical measure pertaining to the dataset. Resulting properties of the estimators are studied.

Optimization of $\phi-$divergences between the empirical measure of the data and the model is problematic when the support of the model is not finite. A number of authors have considered so-called dual representation formulas for divergences or, globally, for convex pseudodistances between distributions. We will make use of the one exposed in [19]; see also [15] for an easy derivation.

3.1.1 Notation

Divergences

The space S is a Polish space endowed with its Borel field $\mathcal{B}(S)$. We consider an identifiable parametric model $\mathcal{P}_\Theta$ on $(S, \mathcal{B}(S))$, hence a class of probability distributions P_θ indexed by a subset Θ included in $\mathbb{R}^d$; Θ needs not be open. The class of all probability measures on $(S, \mathcal{B}(S))$ is denoted $\mathcal{P}$ and $\mathcal{M}(S)$ designates the class of all finite signed measures on $(S, \mathcal{B}(S))$.

A non negative convex function φ with values in $\overline{\mathbb{R}^+}$ belonging to $C^2(\mathbb{R})$ and satisfying $\varphi(1) = \varphi'(1) = 0$ and $\varphi''(1)$ is a *divergence function.* An important class of such functions is defined through the power divergence functions

$$\varphi_\gamma(x) := \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma - 1)} \tag{3.1.1}$$

defined for all real $\gamma \neq 0, 1$ with $\varphi_0(x) := -\log x + x - 1$ (the likelihood divergence function) and $\varphi_1(x) := x \log x - x + 1$ (the Kullback-Leibler divergence function). This class is usually refered to as the Cressie-Read family of divergence functions, a custom we will follow, although its origin takes from [74]. When x is such that $\varphi_\gamma(x)$ is undefined by the above definitions, we set $\varphi_\gamma(x) := +\infty$, by which the definition above is satisfied for all φ_γ. It consists in the simplest power-type class of functions (with the limits in $\gamma \to 0, 1$) which fulfill the definition. The L_1 divergence function $\varphi(x) := |x - 1|$ is not captured by the Cressie-Read family of functions.

Associated with a divergence function φ is the *divergence pseudodistance* between a probability measure and a finite signed measure; see [22].

For P and Q in $\mathcal{M}$ define

$$\begin{aligned}\phi(Q,P) &:= \int \varphi\left(\frac{dQ}{dP}\right) dP \text{ whenever } Q \text{ is a.c. w.r.t. } P\\ &:= +\infty \quad \text{otherwise.}\end{aligned}$$

The divergence $\phi(Q,P)$ is best seen as a mapping $Q \rightarrow \phi(Q,P)$ from $\mathcal{M}$ onto $\overline{\mathbb{R}^+}$ for fixed P in $\mathcal{M}$. Indexing this pseudodistance by γ and using φ_γ as divergence function yields the likelihood divergence $\phi_0(Q,P) := -\int \log\left(\frac{dQ}{dP}\right) dP$, the Kullback-Leibler divergence $\phi_1(Q,P) := \int \log\left(\frac{dQ}{dP}\right) dQ$, the Hellinger divergence $\phi_{1/2}(Q,P) := \frac{1}{2}\int\left(\sqrt{\frac{dQ}{dP}}-1\right)^2 dP$, the modified χ^2 divergence $\phi_{-1}(Q,P) := \frac{1}{2}\int\left(\frac{dQ}{dP}-1\right)^2\left(\frac{dQ}{dP}\right)^{-1} dP$. All these divergences are defined on $\mathcal{P}$. The χ^2 divergence $\phi_2(Q,P) := \frac{1}{2}\int\left(\frac{dQ}{dP}-1\right)^2 dP$ is defined on $\mathcal{M}$. We refer to [19] for the advantage to extend the definition to possibly signed measures in the context of parametric inference for non regular models.

The conjugate divergence function of φ is defined through

$$\widetilde{\varphi}(x) := x\varphi\left(\frac{1}{x}\right) \tag{3.1.2}$$

and the corresponding divergence pseudodistance $\widetilde{\phi}(P,Q)$ is

$$\widetilde{\phi}(P,Q) := \int \widetilde{\varphi}\left(\frac{dP}{dQ}\right) dQ$$

which satisfies

$$\widetilde{\phi}(P,Q) = \phi(Q,P)$$

whenever defined, and equals $+\infty$ otherwise. When $\varphi = \varphi_\gamma$ then $\widetilde{\varphi} = \varphi_{1-\gamma}$ as follows by substitution. Pairs $(\varphi_\gamma, \varphi_{1-\gamma})$ are therefore *conjugate pairs*. Inside the Cressie-Read family, the Hellinger divergence function is self-conjugate.

In parametric models $\varphi-$divergences between two distributions take a simple variational form. It holds, when φ is a differentiable function, and under a commonly met regularity condition, denoted (RC) in [15]

$$\phi(P_\theta, P_{\theta_T}) = \sup_{\alpha\in\mathcal{U}} \int \varphi'\left(\frac{dP_\theta}{dP_\alpha}\right) dP_\theta - \int \varphi^{\#}\left(\frac{dP_\theta}{dP_\alpha}\right) dP_{\theta_T} \tag{3.1.3}$$

where $\varphi^{\#}(x) := x\varphi'(x) - \varphi(x)$. In the above formula, $\mathcal{U}$ designates a subset of Θ containing θ_T such that for any θ, θ' in $\mathcal{U}$, $\phi(P_\theta, P_{\theta'})$ is finite. This formula holds for any divergence in the Cressie Read family, as considered here.

Denote

$$h(\theta,\alpha,x) := \int \varphi'\left(\frac{dP_\theta}{dP_\alpha}\right) dP_\theta - \varphi^{\#}\left(\frac{dP_\theta}{dP_\alpha}(x)\right)$$

from which

$$\phi(P_\theta, P_{\theta_T}) := \sup_{\alpha\in\mathcal{U}} \int h(\theta,\alpha,x) dP_{\theta_T}(x). \tag{3.1.4}$$

For CR divergences

$$h(\theta,\alpha,x)=\frac{1}{\gamma-1}\left[\int\left(\frac{dP_\theta}{dP_\alpha}\right)^{\gamma-1}dP_\theta-1\right]-\frac{1}{\gamma}\left[\left(\frac{dP_\theta}{dP_\alpha}(x)\right)^{\gamma}-1\right].$$

Weights

For a given real valued random variable W denote

$$M(t):=\log E\exp tW \tag{3.1.5}$$

its cumulant generating function which we assume to be finite in a non void interval including 0 (this is the so-called Cramer condition). The Fenchel Legendre transform of M is also called the Chernoff function and is defined through

$$\varphi^W(x)=M^*(x):=\sup_t tx-M(t). \tag{3.1.6}$$

The function $x\rightarrow\varphi^W(x)$ is non negative, is C^2 and convex. We also assume that $EW=1$ together with $VarW=1$ which implies $\varphi^W(1)=\left(\varphi^W\right)'(1)=0$ and $\left(\varphi^W\right)''(1)=1$. Hence $\varphi^W(x)$ is a divergence function with corresponding divergence pseudodistance ϕ^W . Associated with φ^W is the conjugate divergence $\widetilde{\phi^W}$ with divergence function $\widetilde{\varphi^W}$, which therefore satisfies

$$\phi^W(Q,P)=\widetilde{\phi^W}(P,Q).$$

Whether there exists a random variable V satisfying

$$\widetilde{\phi^W}=\phi^V$$

is discussed further on in the paper.

Measure spaces

This paper makes extensive use of Sanov type large deviation results for empirical measures or weighted empirical measures. This requires some definitions and facts.

The vector space $\mathcal{M}(S)$ is endowed with the $\tau-$topology, which is the coarest making all mappings $Q\rightarrow\int fdQ$ continuous for any $Q\in\mathcal{M}(S)$ and any $f\in B(S)$ which denotes the class of all bounded measurable functions on $(S,\mathcal{B}(S))$. A slightly stronger topology will be used in this paper, the τ_0 topology, introduced in [27], which is the natural setting for our sake. This topology can be described through the following basis of neighborhoods. Consider $\mathfrak{P}$ the class of all partitions of S and for $k\geq1$ the class $\mathfrak{P}_k$ of all partitions of S into k disjoint sets, $\mathcal{P}_k:=(A_1,...,A_k)$ where the A_i's belong to $\mathcal{B}(S)$. For fixed P in $\mathcal{M}$, for any k, any such partition $\mathcal{P}_k$ in $\mathfrak{P}_k$ and any positive ε define the open neighborhood $U(P,\varepsilon,\mathcal{P}_k)$ through

$$U(P,\varepsilon,\mathcal{P}_k):=\left\{Q\in\mathcal{M}\text{ such that }\max_{1\leq i\leq k}|P(A_i)-Q(A_i)|<\varepsilon\text{ and }Q(A_i)=0\text{ if }P(A_i)=0\right\}.$$

The additional requirement $Q(A_i) = 0$ if $P(A_i) = 0$ in the above definition with respect to the classical definition of the basis of the $\tau-$topology is essential for the derivation of Sanov type theorems. Endowed with the τ_0-topology, $\mathcal{M}$ is a Hausdorff locally convex vector space.

The following Pinsker type property holds

$$\sup_k \sum_{i=1}^{k} \varphi \left(\frac{Q(A_i)}{P(A_i)} \right) P(A_i) = \phi(Q, P)$$

see [45].

For any P in $\mathcal{M}$ the mapping $Q \to \phi(Q, P)$ is lower semi continuous; see [18], Proposition 2.2. Denoting (a, b) the domain of φ whenever

$$\lim_{\substack{x \to a \\ x > a}} \frac{\varphi(x)}{x} = \lim_{\substack{x \to b \\ x < b}} \frac{\varphi(x)}{x} = +\infty$$

then for any positive C, the level set $\{Q : \phi(Q, P) \leq C\}$ is τ_0-compact, making $Q \to \phi(Q, P)$ a so-called good rate function. Divergence functions φ satisfying this requirement for example are φ_γ with $\gamma > 1$; see [18] for different cases.

Minimum dual divergence estimators

The above formula (3.1.3) defines a whole range of plug in estimators of $\phi(P_\theta, P_{\theta_T})$ and of θ_T. Let $X_1, ..., X_n$ denote n i.i.d. r.v's with common didistribution P_{θ_T}. Denoting

$$P_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

the empirical measure pertaining to this sample. The plug in estimator of $\phi(P_\theta, P_{\theta_T})$ is defined through

$$\phi_n(P_\theta, P_{\theta_T}) := \sup_{\alpha \in \mathcal{U}} \int h(\theta, \alpha, x) dP_n(x)$$

and the family of M-estimators indexed by θ

$$\alpha_n(\theta) := \arg \sup_{\alpha \in \mathcal{U}} \int h(\theta, \alpha, x) dP_n(x)$$

approximates θ_T. In the above formulas $\mathcal{U}$ is defined after (3.1.3). See [19] and [76] for asymptotic properties and robustness results.

Since $\phi(P_{\theta_T}, P_{\theta_T}) = 0$ a natural estimator of θ_T which only depends on the choice of the divergence function φ is defined through

$$\begin{aligned} \theta_n &:= \arg \inf_\theta \phi_n(P_\theta, P_{\theta_T}) \\ &= \arg \inf_{\theta \in \mathcal{U}} \sup_{\alpha \in \mathcal{U}} \int h(\theta, \alpha, x) dP_n(x); \end{aligned}$$

see [19] for limit properties.

3.1.2 Main topics treated in this paper

We now turn to the derivation of the ML estimate through the approximations obtained hereabove. Although much of the development seems unneccessary due to the very form of the Kullback-Leibler divergence. We intend to stress the fact that a large deviation rate in Sanov Large deviation Theorem defined through a divergence function φ induces a Maximum Likelihood estimator which is an M-estimator for a statistical functional defined through the conjugate divergence function of φ. This is developped in details for the classical ML pertaining to the simple sampling scheme and will be extended to the weighted bootstrap sampling scheme in the next section.

3.2 Large deviation and maximum likelihood

3.2.1 ML under finite supported distributions and simple sampling

Suppose that all probability measures P_θ in $\mathcal{P}_\Theta$ share the same finite support $S := \{1, ..., k\}$. Let $X_1, ..., X_n$ be a set of n independent random variables with common probability measure P_{θ_T} and consider the Maximum Likelihood estimator of θ_T . A common way to define the ML paradigm is as follows: For any θ consider independent random variables $(X_{1,\theta}, ..., X_{n,\theta})$ with probability measure P_θ , thus *sampled in the same way as the* X_i*'s*, but under some altermative θ. Define θ_{ML} as the value of the parameter θ for which the probability that, up to a permutation of the order of the $X_{i,\theta}$'s, the probability that $(X_{1,\theta}, ..., X_{n,\theta})$ occupies S as does $X_1, ..., X_n$ is maximal, conditionaly on the observed sample $X_1, ..., X_n$. In formula, let σ denote a random permutation of the indexes $\{1, 2, ..., n\}$ and θ_{ML} is defined through

$$\theta_{ML} := \arg\max_\theta \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}} P_\theta \left(\left(X_{\sigma(1),\theta}, ..., X_{\sigma(n),\theta} \right) = (X_1, ..., X_n) \middle| (X_1, ..., X_n) \right) \tag{3.2.1}$$

where the summation is extended on all equally probable permutations of $\{1, 2, ..., n\}$.

Denote

$$P_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

and

$$P_{n,\theta} := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_{i,\theta}}$$

the empirical measures pertaining respectively to $(X_1, ..., X_n)$ and $(X_{1,\theta}, ..., X_{n,\theta})$

An alternative expression for θ_{ML} is

$$\theta_{ML} := \arg\max_\theta P_\theta \left(P_{n,\theta} = P_n \middle| P_n \right). \tag{3.2.2}$$

An explicit enumeration of the above expression $P_\theta \left(P_{n,\theta} = P_n | P_n \right)$ involves the quantities

$$n_j := card \{i : X_i = j\}$$

for $j = 1, ..., k$ and yields

$$P_\theta\left(P_{n,\theta} = P_n \middle| P_n\right) = \frac{\prod_{j=1}^{k} n_j! P_\theta(j)^{n_j}}{n!} \tag{3.2.3}$$

as follows from the classical multinomial distribution. Optimizing on θ in (3.2.3) yields

$$\begin{aligned}\theta_{ML} &= \arg\max_\theta \sum_{j=1}^{k} \frac{n_j}{n} \log P_\theta(j) \\ &= \arg\max_\theta \frac{1}{n} \sum_{i=1}^{n} \log P_\theta(X_i).\end{aligned}$$

Consider now the Kullback-Leibler distance between P_θ and P_n which is non commutative and defined through

$$\begin{aligned}KL(P_n, P_\theta) &:= \sum_{j=1}^{k} \varphi\left(\frac{n_j/n}{P_\theta(j)}\right) P_\theta(j) \\ &= \sum_{j=1}^{k} (n_j/n) \log \frac{n_j/n}{P_\theta(j)}\end{aligned} \tag{3.2.4}$$

where

$$\varphi(x) := x \log x - x + 1 \tag{3.2.5}$$

which is the Kullback-Leibler divergence function. Minimizing the Kullback-Leibler distance $KL(P_n, P_\theta)$ upon θ yields

$$\begin{aligned}\theta_{KL} &= \arg\min_\theta KL(P_n, P_\theta) \\ &= \arg\min_\theta - \sum_{j=1}^{k} \frac{n_j}{n} \log P_\theta(j) \\ &= \arg\max_\theta \sum_{j=1}^{k} \frac{n_j}{n} \log P_\theta(j) \\ &= \theta_{ML}.\end{aligned}$$

Introduce the *conjugate divergence function* $\widetilde{\varphi}$ of φ, inducing the modified Kullback-Leibler, or so-called Likelihood divergence pseudodistance KL_m which therefore satisfies

$$KL_m(P_\theta, P_n) = KL(P_n, P_\theta).$$

We have proved that minimizing the Kullback-Leibler divergence $KL(P_n, P_\theta)$ amounts to minimizing the Likelihood divergence $KL_m(P_\theta, P_n)$ and produces the ML estimate of θ_T.

Kullback-Leibler divergence as defined above by $KL(P_n, P_\theta)$ is related to the way P_n keeps away from P_θ when θ is not equal to the true value of the parameter θ_T generating the observations X_i's and is closely related with the type of sampling of the X_i's. In the present case i.i.d. sampling of the $X_{i,\theta}$'s under P_θ results in the asymptotic property, named Large Deviation Sanov property

$$\lim_{n\to\infty} \frac{1}{n} \log P_\theta \left(P_{n,\theta} = P_n \right| P_n) = -KL \left(P_{\theta_T}, P_\theta \right). \tag{3.2.6}$$

This result can easily be obtained from (3.2.3) using Stirling formula to handle the factorial terms and the law of large numbers which states that for all j's, n_j/n tends to $P_{\theta_T}(j)$ as n tends to infinity. Comparing with (3.2.4) we note that the ML estimator θ_{ML} estimates the minimizer of the natural estimator of $KL\left(P_{\theta_T}, P_\theta\right)$ in θ, substituting the unknown measure generating the X_i's by its empirical counterpart P_n . Alternatively as will be used in the sequel, θ_{ML} minimizes upon θ the Likelihood divergence $KL_m\left(P_\theta, P_{\theta_T}\right)$ between P_θ and P_{θ_T} substituting the unknown measure P_{θ_T} generating the X_i's by its empirical counterpart P_n . Summarizing we have obtained:

The ML estimate can be obtained from a LDP statement as given in (3.2.6), optimizing in θ in the estimator of the LDP rate where the plug-in method of the empirical measure of the data is used instead of the unknown measure P_{θ_T}. Alternatively it holds

$$\theta_{ML} := \arg\min_\theta \widehat{KL_m}\left(P_\theta, P_{\theta_T}\right) \tag{3.2.7}$$

with

$$\widehat{KL_m}\left(P_\theta, P_{\theta_T}\right) := KL_m\left(P_\theta, P_n\right).$$

In the rest of this section we will develop a similar approach for a model $\mathcal{P}_\Theta$ whose all members P_θ share the same infinite (countable or not) support S.

The statistical properties of θ_{ML} are obtained under the i.i.d. sampling having generated the observed values.

This principle will be kept throughout this paper: the estimator is defined as maximizing the probability that the simulated empirical measure be close to the empirical measure as observed on the sample, conditionally on it, following the same sampling scheme. This yilds a maximum likelihood estimator, and its properties a re then obtained when randomness is introduced as resulting from the sampling scheme.

3.2.2 ML under general distributions and simple sampling

When the support of the generic r.v. X_1 is not finite some of the arguments above are not valid any longer and some discretization scheme is required in order to get occupation probabilities in the spirit of (3.2.3) or (3.2.6). Since all distributions P_θ in $\mathcal{P}_\Theta$ have infinite support, i.i.d. sampling under any P_θ yields $(X_{1,\theta}, ..., X_{n,\theta})$ such that

$$P_\theta \left(P_{n,\theta} = P_n \right| P_n) = 0$$

for all n, so that we are lead to consider the optimization upon θ of probabilities of the type $P_\theta \left(P_{n,\theta} \in V\left(P_n\right) \right| P_n)$ where $V\left(P_n\right)$ is a (small) neighborhood of P_n. Considering the distribution of the outcomes of the simulating scheme P_θ results in the definition of neighborhoods through partitions of S, hence through the τ_0-topology.

When P_n is the empirical measure for some observed r.v's $X_1, ...X_n$, an $\varepsilon-$neighborhood of P_n contains distributions whose support is not necessarily finite, and may indeed be equivalent to the measures in the model $\mathcal{P}_\Theta$ when defined on the Borel $\sigma-$field $\mathcal{B}(S)$.

Let $\mathcal{P}_k := (A_1, ..., A_k)$ be some partition in $\mathfrak{P}_k$. Denote

$$V_{k,\varepsilon}(P_n) := \left\{ Q \in \mathcal{M} \text{ such that } \max_{i=1,\ldots,k} |P_n(A_i) - Q(A_i)| < \varepsilon \text{ and } Q(A_i) = 0 \text{ if } P_n(A_i) = 0 \right\} \tag{3.2.8}$$

an open neighborhood of P_n.

We also would define the Kullback-Leibler divergence between two probability measures Q and P on the partition $\mathcal{P}_k$ through

$$KL_{A_k}(Q, P) := \sum_{A_j \in \mathcal{P}_k} \log \left(\frac{Q(A_j)}{P(A_j)} \right) Q(A_j).$$

Also we define the corresponding Likelihood divergence on $\mathcal{P}_k$ through

$$(KL_m)_{\mathcal{P}_k}(Q, P) := KL_{\mathcal{P}_k}(P, Q). \tag{3.2.9}$$

As in the finite case for any θ in Θ denote $(X_{1,\theta}, ..., X_{n,\theta})$ a set of n i.i.d. random variables with common distribution P_θ. We have

Lemma 3.2.1. *For large n*

$$\begin{aligned}\frac{1}{n} \log P_\theta \left(P_{n,\theta} \in V_{k,\varepsilon}(P_n) \middle| P_n \right) &\geq -KL_{\mathcal{P}_k}\left(V_{k,\varepsilon}(P_n), P_\theta\right) - \frac{k\log(n+1)}{n} \\ &:= -\inf_{Q \in V_{k,\varepsilon}(P_n)} KL_{\mathcal{P}_k}(Q, P_\theta) - \frac{k\log(n+1)}{n}\end{aligned}$$

Proof. The proof uses similar arguments as in [27] Lemma 4.1. For fixed k and large n, P_{θ_T} belongs to $V_{k,\varepsilon}(P_n)$, by the law of large numbers. Indeed for large n , $P_n(A_j)$ is positive and $|P_{\theta_T}(A_j) - P_n(A_j)| < \varepsilon$ for all j in $\{1, ..., k\}$. Assuming that for all θ in Θ

$$KL(P_{\theta_T}, P_\theta) < \infty$$

and taking into account the fact (see [71]) that for any probability measures P and Q, $K(P, Q) = \sup_k \sup_{\mathcal{P}_k \in \mathfrak{P}_k} KL_{\mathcal{P}_k}(P, Q)$ where $\mathfrak{P}_k$ is the class of all partitions of S in k sets in $\mathcal{B}(S)$, it follows that

$$KL_{\mathcal{P}_k}\left(V_{k,\varepsilon}(P_n), P_\theta\right) \text{ is finite}$$

for all fixed k and large n. For positive δ let $P^{(n)}$ in $V_{k,\varepsilon}(P_n)$ with

$$KL_{\mathcal{P}_k}\left(P^{(n)}, P_\theta\right) < KL_{\mathcal{P}_k}\left(V_{k,\varepsilon}(P_n), P_\theta\right) + \delta.$$

Let $0 < \varepsilon' < \varepsilon$ and non negative numbers r_j , $1 \leq j \leq k$ such that

$$\left| r_j - P^{(n)}(A_j) \right| < \varepsilon', \text{ and } r_j = 0 \text{ if } P^{(n)}(A_j) = 0 \text{ and } \sum_{j=1}^{k} r_j = 1.$$

The probability vector $(r_1, ..., r_k)$ defines a probability measure R on $(S, \mathcal{P}_k)$, and R belongs to $V_{k,\varepsilon}(P_n)$. By continuity of the mapping $x \to x \log \frac{x}{P_\theta(A_j)}$ it is possible to fit the r_j's such that for all j between 1 and k

$$\left| r_j \log \frac{r_j}{P_\theta(A_j)} - P^{(n)}(A_j) \log \frac{P^{(n)}(A_j)}{P_\theta(A_j)} \right| < \frac{\delta}{k}. \tag{3.2.10}$$

Indeed since all the P_θ's share the same support, if $P_\theta(A_j) = 0$ then $P_{\theta_T}(A_j) = 0$ which in turn yields $P_n(A_j) = 0$ which through (3.2.8) implies $P^{(n)}(A_j) = 0$. This plus the conventions $0/0 = 0$ and $0 \log 0 = 0$ implies that (3.2.10) holds true for some choice of the r_j's. Choose further the r_j's in such a way that $l_j := nr_j$ is an integer for all j. Let $P_{n,\theta}$ denote the empirical distribution of the $X_{i,\theta}$'s. We now proceed to the evaluation of $P_\theta\left(P_{n,\theta} \in V_{k,\varepsilon}(P_n) \middle| P_n\right)$. It holds

$$\begin{aligned} P_\theta\left(P_{n,\theta} \in V_{k,\varepsilon}(P_n) \middle| P_n\right) &\geq P_\theta\left(P_{n,\theta}(A_j) = r_j, 1 \leq j \leq k \middle| P_n\right) \\ &= \frac{\prod_{j=1}^{k} l_j!}{n!} \prod_{j=1}^{k} P_\theta(A_j)^{l_j} \\ &\geq (n+1)^{-k} \exp -n \sum_{j=1}^{k} r_j \log \frac{r_j}{P_\theta(A_j)} \end{aligned}$$

where we used the same argument as in [27], Lemma 4.1. In turn using (3.2.10)

$$\begin{aligned} \sum_{j=1}^{k} r_j \log \frac{r_j}{P_\theta(A_j)} &\leq \sum_{j=1}^{k} P^{(n)}(A_j) \log \frac{P^{(n)}(A_j)}{P_\theta(A_j)} + \delta \\ &\leq KL_{\mathcal{P}_k}\left(V_{k,\varepsilon}(P_n), P_\theta\right) + 2\delta \end{aligned}$$

and the proof is completed. □

The reverse inequality is as in [27] p 790: The set $V_{k,\varepsilon}(P_n)$ is completely convex, in the terminology of [27], whence it follows

Lemma 3.2.2. *For all n*

$$\frac{1}{n} \log P_\theta\left(P_{n,\theta} \in V_{k,\varepsilon}(P_n) \middle| P_n\right) \leq -KL_{\mathcal{P}_k}\left(V_{k,\varepsilon}(P_n), P_\theta\right)$$

Lemmas 3.2.1 and 3.2.2 link the Maximum Likelihood Principle with the Large deviation statements. Define

$$\theta_{ML} := \arg\max_\theta \frac{1}{n} \log P_\theta\left(P_{n,\theta} \in V_{k,\varepsilon}(P_n) \middle| P_n\right) \tag{3.2.11}$$

and

$$\theta_{LDP} := \arg\min_\theta -KL_{\mathcal{P}_k}\left(V_{k,\varepsilon}(P_n), P_\theta\right)$$

assuming those parameters defined, possibly not in a unique way. Denote

$$L_{k,\varepsilon}(\theta) := \frac{1}{n} \log P_\theta\left(P_{n,\theta} \in V_{k,\varepsilon}(P_n) \middle| P_n\right)$$

and

$$K_{k,\varepsilon}(\theta) := -KL_{\mathcal{P}_k}\left(V_{k,\varepsilon}(P_n), P_\theta\right).$$

We then deduce that

$$-\frac{k}{n} \log(n+1) \leq L_{k,\varepsilon}(\theta_{ML}) - K_{k,\varepsilon}(\theta_{ML}) \leq 0$$

$$0 \leq -L_{k,\varepsilon}(\theta_{LDP}) - K_{k,\varepsilon}(\theta_{LDP}) \leq \frac{k}{n} \log(n+1)$$

whence

$$0 \leq L_{k,\varepsilon}(\theta_{ML}) - L_{k,\varepsilon}(\theta_{LDP}) \leq \frac{k}{n}\log(n+1) \tag{3.2.12}$$

from which θ_{LDP} is a good substitute for θ_{ML} for fixed k and ε in the partitioned based model. Note that the bounds in (3.2.12) do not depend on the peculiar choice of $\mathcal{P}_k$ in $\mathfrak{P}_k$.

Fix $k = k_n$ such that $\lim_{n\to\infty} k_n = \infty$ together with $\lim_{n\to\infty} k_n/n = 0$. Define the partition $\mathcal{P}_k$ such that $P_n(A_j) = k_n/n$ for all $j = 1, ..., k$. Hence A_j contains only k sample points. Let $\varepsilon > 0$ such that $\max_{1\leq j\leq k} |P_{\theta_T}(A_j) - k_n/n| < \varepsilon$. Then clearly P_{θ_T} belongs to $V_{k,\varepsilon}(P_n)$ and $V_{n,\varepsilon}(P_n)$ is included in $V_{k,2\varepsilon}(P_{\theta_T})$. Therefore for any θ it holds

$$KL_{\mathcal{P}_k}(V_{k,2\varepsilon}(P_{\theta_T}), P_\theta) \leq KL_{\mathcal{P}_k}(V_{k,\varepsilon}(P_n), P_\theta) \leq KL_{\mathcal{P}_k}(P_{\theta_T}, P_\theta) \tag{3.2.13}$$

which proves that $\inf_\theta KL_{\mathcal{P}_k}(V_{k,\varepsilon}(P_n), P_\theta) = 0$ with attainment on θ' such that $P_{\theta'}$ and P_{θ_T} coincide on $\mathcal{P}_k$.

We now turn to the study of the RHS term in (3.2.13). Introducing the likelihood divergence $\widetilde{\varphi}$ defined in (3.2.9) leads

$$KL_{\mathcal{P}_k}(P_{\theta_T}, P_\theta) = (KL_m)_{\mathcal{P}_k}(P_\theta, P_{\theta_T})$$

whence minimizing $KL_{\mathcal{P}_k}(P_{\theta_T}, P_\theta)$ over θ in Θ amounts to minimizing the likelihood divergence $\theta \to (KL_m)_{\mathcal{P}_k}(P_\theta, P_{\theta_T})$. Set therefore

$$\theta_{LDP,\mathcal{P}_k} := \arg\min_\theta KL_{\mathcal{P}_k}(P_{\theta_T}, P_\theta) = \arg\min_\theta (KL_m)_{\mathcal{P}_k}(P_\theta, P_{\theta_T}).$$

Based on the $\sigma-$field generated by $\mathcal{P}_k$ on S the dual form (3.1.3) of the Likelihood divergence pseudodistance $(KL_m)_{\mathcal{P}_k}(P_\theta, P_{\theta_T})$ yields

$$\begin{aligned}\arg\min_\theta (KL_m)_{\mathcal{P}_k}(P_\theta, P_{\theta_T}) = \arg\min_\theta \sup_\eta \sum_{B_j\in\mathcal{P}_k} \widetilde{\varphi}\left(\frac{P_\theta}{P_\eta}(A_j)\right) P_\theta(A_j) \\ - \sum_{B_j\in\mathcal{P}_k} (\widetilde{\varphi})^* \left(\frac{P_\theta}{P_\eta}(A_j)\right) P_{\theta_T}(A_j).\end{aligned} \tag{3.2.14}$$

with $\widetilde{\varphi}(x) = -\log x + x - 1$ and $(\widetilde{\varphi})^*(x) = -\log(1-x)$. With the present choice for $\widetilde{\varphi}$ the terms in P_η vanish in the above expression ; however we complete a full development, as required in more envolved sampling schemes. Now an estimate of θ_T is obtained substituting P_{θ_T} by P_n in (3.2.14) leading, denoting n_j the number of X_i's in A_j

$$\widehat{\theta}_{LDP,\mathcal{P}_k} := \arg\min_\theta \sup_\eta \sum_{A_j\in\mathcal{P}_k} \widetilde{\varphi}\left(\frac{P_\theta}{P_\eta}(A_j)\right) P_\theta(A_j) - \sum_{A_j\in\mathcal{P}_k} \frac{n_j}{n} (\widetilde{\varphi})^* \left(\frac{P_\theta}{P_\eta}(A_j)\right).$$

Letting n tend to infinity yields (recall that $k = k_n$)

$$\begin{aligned}\lim_{n\to\infty} \sup_\eta \Bigg| \Bigg[\sum_{A_j\in\mathcal{P}_k} \widetilde{\varphi}\Big(\frac{P_\theta}{P_\eta}(A_j)\Big) - \sum_{A_j\in\mathcal{P}_k} (\widetilde{\varphi})^* \Big(\frac{P_\theta}{P_\eta}(Aj)\Big) P_{\theta_T}(A_j)\Bigg] \\ - \Bigg[\int \widetilde{\varphi}\Big(\frac{p_\theta}{p_\eta}(x)\Big) p_\theta(x)\,dx - \int (\widetilde{\varphi})^* \left(\frac{p_\theta}{p_\eta}(x)\right) dP_n(x)\Bigg]\Bigg| = 0\end{aligned}$$

w.p. 1 which in turn implies

$$\lim_{n\to\infty} \widehat{\theta}_{LDP,\mathcal{P}_k} - \widehat{\theta}_{ML} = 0$$

where $\widehat{\theta}_{ML}$ is readily seen to be the usual ML estimator of θ defined through

$$\widehat{\theta}_{ML} := \arg\sup_{\theta} \prod_{i=1}^{n} p_\theta (X_i)$$

where $X_1, ..., X_n$ have the common density p_θ.

3.3 Weighted sampling

This section extends the previous arguments for weighted sampling schemes. We will show that the Maximum Likelihood paradigm as defined above can be extended for these schemes, leading to operational procedures involving the minimization of specific divergence pseudodistances defined in strong relation with the distribution of the weights.

The sampling scheme which we consider is commonly used in connection with the bootstrap and is refered to as the *weighted* or *generalized bootstra*p, sometimes called *wild bootstrap*, first introduced by Newton and Mason [64]. The main simplification which we consider in the present setting lies in the fact that we assume that the weights W_i are i.i.d. while being exchangeable random variables in the generalized bootstrap setting.

Let $x_1, ..., x_n$ be n independent realizations of n i.i.d. r.v's $X_1, ..., X_n$ with common distribution P_{θ_T}. It will be assumed that

$$\text{For all } \theta \text{ in } \Theta,\ E_\theta X \text{ and } E_\theta X^2 \text{ are finite.} \tag{3.3.1}$$

This entails that both

$$\frac{1}{n}\sum_{i=1}^{n} x_i \text{ and } \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

converge $P_{\theta_T}-$a.e. to $E_{\theta_T}X$ and $E_{\theta_T}X^2$ respectively; also the same holds with θ_T substituted by any θ in Θ when $x_1, ..., x_n$ is sampled under P_θ. This assumption is necessary when studying the properties of the estimates of θ_T and of $\phi(\theta_T, \theta)$ under some alternative θ.

Consider a collection $W_1, ..., W_n$ of independent copies of W, whose distribution satisfies the conditions stated in Section 1. The weighted empirical measure P_n^W is defined through

$$P_n^W := \frac{1}{n}\sum_{i=1}^{n} W_i \delta_{x_i}.$$

This empirical measure need not be a probability measure, since its mass may not equal 1. Also it might not be positive, since the weights may take negative values. The measure P_n^W converges almost surely to P_{θ_T} when the weights W_i's satisfy the hypotheses stated in Section 1. Indeed general results pertaining to this sampling procedure state that under regularity, functionals of the measure P_n^W are asymptotically distributed as are the same functionals of P_n when the X_i's are i.i.d. Therefore the weighted sampling procedure mimicks the i.i.d. sampling fluctuation in a two steps procedure: choose n values of x_i such that they asymptotically fit to P_{θ_T}, which means

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} = P_{\theta_T}$$

deterministically and then play the W_i's on each of the x_i's. Then get P_n^W, a proxy to the random empirical measure P_n .

For any θ in Θ consider a similar sampling procedure under the weights W_i' 's which are i.i.d. copies of the W_i's. Let therefore $x_{1,\theta}, ..., x_{n,\theta}$ denote n i.i.d. realizations of $X_{1,\theta}, ..., X_{n,\theta}$ with distribution P_θ yielding the empirical measure

$$P_{n,\theta}^{W'} := \frac{1}{n}\sum_{i=1}^{n} W_i'\delta_{x_{i,\theta}}$$

the corresponding empirical measure. Note that except for the choice of the generating measure P_θ , $P_{n,\theta}^{W'}$ is obtained in the same way as P_n^W. The ML principle turns out to select the value of θ making $P_{n,\theta}^{W'}$ as close as possible from P_n^W, conditionally upon P_n^W.

The resulting estimates are optimal in many respects, as is the classical ML estimator for regular models in the i.i.d. sampling scheme. The proposal which is presented here also allows to obtain optimal estimators for some non regular models. This approach is in line with [19] who developped a whole range of first order optimal estimation procedures in the case of the i.i.d. sampling, based on divergence minimization.

Using the notations of section 3.1.1, we endow $\mathcal{M}(S)$ with τ_0-topology rather than the weak topology, and define accordingly the σ-field $\mathcal{B}(\mathcal{M})$ on $\mathcal{M}(S)$. Denote by $\mathcal{M}_1(S)$ the space of probability measure on S, endowed with the τ_0-topology.

3.3.1 A Sanov conditional theorem for the weighted empirical measure

The procedure which we are going to develop can be stated as follows.

Similarly as in the simple i.i.d. setting select some (small) neighborhood $V_\epsilon\left(P_n^W\right)$of P_n^W and define the MLE of θ_T as the value of θ which optimizes the probability that the simulated empirical measure $P_{n,\theta}^{W'}$ belongs to $V_\epsilon\left(P_n^W\right)$. This requires a conditional Sanov type result, substituting Lemmas 3.2.1 and 3.2.2. This result is produced in Theorem 3.3.1 in Section 3.3.1. In the same vein as in Lemmas 3.2.1 and 3.2.2, maximizing in θ this probability amounts to minimizing a LDP rate between P_θ and $V_\epsilon\left(P_{\theta_T}\right)$. The rate is in strong relation with the distribution of the W_i's. Call it $\phi^W\left(V_\epsilon\left(P_{\theta_T}\right), P_\theta\right) := \inf\left\{\phi^W\left(Q, P_\theta\right), Q \in V_\epsilon\left(P_{\theta_T}\right)\right\}$.

Since ϵ is small, this rate is of order $\phi^W\left(P_{\theta_T}, P_\theta\right)$; this is Corollary 3.3.1 in Section 3.3.1. Turn to the original data and estimate $\phi^W\left(P_{\theta_T}, P_\theta\right)$ by some plug in method to be stated in Section 3.3.2. Define the ML estimator of θ_T through the minimization of the proxy of $\phi^W\left(P_{\theta_T}, P_\theta\right)$. We will prove that minimum divergence estimators play a key role in this setting.

In order to state our conditional Sanov theorem we put forwards the following lemma, which is in the vein of Theorem 2.2 of Najim [66] which states the Sanov large deviation theorem, where the weights are i.i.d random variables. Trashorras and Wintenberger [77] have investigated the large deviations properties of weighted (bootstrapped) empirical measure with exchangeable weights under appropriate assumptions of the weights. Both papers equip $\mathcal{M}(S)$ with the weak topology.

The lemma's proof is defered to Section 3.7.

Lemma 3.3.1. *Assume that $P_\theta(U) > 0$ for any non-empty open set $U \in S$, and that $\lim_{n\to\infty} P_n = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \delta_{x_i} = P_\theta \in \mathcal{M}_1(S)$, where the convergence holds under τ_0. Then $P_{n,\theta}^W$ satisfies the LDP in $(\mathcal{M}(S), \mathcal{B}(\mathcal{M}))$ equipped with the τ_0-topology with the good convex rate function:*

$$\begin{aligned}\phi^W(\zeta, P_\theta) &= \sup_{f\in B(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} f(x)\zeta(dx) - \int_{\mathbb{R}^d} M(f(x))P_\theta(dx)\right\} \\ &= \begin{cases} \int_{\mathbb{R}^d} M^*(\frac{d\zeta}{dP_\theta})dP_\theta, & \text{if } \zeta \text{ is a.c. w.r.t. } P_\theta \\ \infty, & \text{otherwise}\end{cases}\end{aligned}$$

where $M^*(x) = \sup_t tx - M(t)$ for all real x and $M(t)$ is the moment generating function of W.

Let $\mathcal{P}_k = (A_1, ..., A_k)$ denote an arbitrary partition of S with A_i in $B(S)$ for all $i = 1, ..., k$, and define the pseudometric $d_{\mathcal{P}_k}$ on $\mathcal{M}(S)$ by

$$d_{\mathcal{P}_k}(Q, R) = \max_{1\le j\le k} |Q(B_j) - R(B_j)|, \qquad Q, R \in \mathcal{M}(S).$$

For any positive ϵ, let

$$V_\epsilon(P_n^W) = \{Q \in \mathcal{M}(S) : d_{\mathcal{P}_k}(Q, P_n^W) < \epsilon\}$$

denote an open neighborhood of the weighted empirical measure P_n^W in the τ_0 - topology. Then we have the following conditional LDP theorem.

Theorem 3.3.1. *With the above notation and assuming that P_{θ_T} is absolutely continuous with respect to P_θ, for any positive ϵ, the following conditional LDP result holds*

$$\lim_{n\to\infty} \frac{1}{n} \log P_\theta\Big(P_{n,\theta}^{W'} \in V_\epsilon(P_n^W)|P_n\Big) = -\phi^W(V_\epsilon(P_{\theta_T}), P_\theta).$$

Proof. In the following proof, $\mathcal{P}_k$ is an arbitrary partition on S.

$$\begin{aligned}P_\theta\Big(P_{n,\theta}^{W'} \in V_\epsilon(P_n^W)|P_n\Big) &= P_\theta\Big(d_{\mathcal{P}_k}(P_{n,\theta}^{W'}, P_n^W) < \epsilon|P_n\Big) \\ &\ge P_\theta\Big(d_{\mathcal{P}_k}(P_{n,\theta}^{W'}, P_{\theta_T}) + d_{\mathcal{P}_k}(P_{\theta_T}, P_n^W) < \epsilon|P_n\Big) \\ &= P_\theta\Big(d_{\mathcal{P}_k}(P_{n,\theta}^{W'}, P_{\theta_T}) < \epsilon - d_{\mathcal{P}_k}(P_{\theta_T}, P_n^W)|P_n\Big).\end{aligned}$$

Since $d_{\mathcal{P}_k}(P_{\theta_T}, P_n^W) \to 0$ when $n \to \infty$, for any positive δ and sufficiently large n we have:

$$P_\theta\Big(P_{n,\theta}^{W'} \in V_\epsilon(P_n^W)|P_n\Big) \ge P_\theta\Big(d_{\mathcal{P}_k}(P_{n,\theta}^{W'}, P_{\theta_T}) < \epsilon - \delta\Big) = P_\theta\Big(P_{n,\theta}^{W'} \in V_{\epsilon-\delta}(P_{\theta_T})\Big).$$

By Lemma 3.3.1, we obtain the conditioned LDP lower bound

$$\liminf_{n\to\infty} \frac{1}{n} \log P_\theta\Big(P_{n,\theta}^{W'} \in V_\epsilon(P_n^W)|P_n\Big) \geq -\phi^W(V_{\epsilon-\delta}(P_{\theta_T}), P_\theta),$$

In a similar way, we obtain the large deviation upper bound

$$\begin{aligned} P_\theta\Big(P_{n,\theta}^{W'} \in V_\epsilon(P_n^W)|P_n\Big) &= P_\theta\Big(d_{\mathcal{P}_k}(P_{n,\theta}^{W'}, P_n^W) < \epsilon|P_n\Big) \\ &\leq P_\theta\Big(d_{\mathcal{P}_k}(P_{n,\theta}^{W'}, P_{\theta_T}) - d_{\mathcal{P}_k}(P_{\theta_T}, P_n^W) < \epsilon|P_n\Big) \\ &\leq P_\theta\Big(d_{\mathcal{P}_k}(P_{n,\theta}^{W'}, P_{\theta_T}) < \epsilon + \delta'\Big) = P_\theta\Big(P_{n,\theta}^{W'} \in V_{\epsilon+\delta'}(P_{\theta_T})\Big), \end{aligned}$$

for some positive δ'. We thus obtain

$$\limsup_{n\to\infty} \frac{1}{n} \log P_\theta\Big(P_{n,\theta}^{W'} \in V_\epsilon(P_n^W)|P_n\Big) \leq -\phi^W(V_{\epsilon+\delta'}(P_{\theta_T}), P_\theta).$$

Let $\delta'' = max(\delta, \delta')$, we have

$$\begin{aligned} -\phi^W(V_{\epsilon-\delta''}(P_{\theta_T}), P_\theta) &\leq \liminf_{n\to\infty} \frac{1}{n} \log P_\theta\Big(P_{n,\theta}^{W'} \in V_\epsilon(P_n^W)|P_n\Big) \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \log P_\theta\Big(P_{n,\theta}^{W'} \in V_\epsilon(P_n^W)|P_n\Big) \leq -\phi^W(V_{\epsilon+\delta''}(P_{\theta_T}), P_\theta). \end{aligned}$$

Denote $cl_{\tau_0}(V_\epsilon(P_{\theta_T}))$ the closure of the open set $V_\epsilon(P_{\theta_T})$ in the τ_0-topology, and note δ'' is arbitrarily small, then it holds

$$\begin{aligned} -\phi^W(V_\epsilon(P_{\theta_T}), P_\theta) &\leq \liminf_{n\to\infty} \frac{1}{n} \log P_\theta\Big(P_{n,\theta}^{W'} \in V_\epsilon(P_n^W)|P_n\Big) \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \log P_\theta\Big(P_{n,\theta}^{W'} \in V_\epsilon(P_n^W)|P_n\Big) \leq -\phi^W\Big(cl_{\tau_0}(V_\epsilon(P_{\theta_T})), P_\theta\Big). \end{aligned}$$

It remains to show that

$$\phi^W(V_\epsilon(P_{\theta_T}), P_\theta) = \phi^W\Big(cl_{\tau_0}(V_\epsilon(P_{\theta_T})), P_\theta\Big). \tag{3.3.2}$$

Since P_{θ_T} is absolutely continuous with respect to P_θ, by Lemma 3.3.1 we have

$$\phi^W\Big(cl_{\tau_0}(V_\epsilon(P_{\theta_T})), P_\theta\Big) \leq \phi^W(V_\epsilon(P_{\theta_T}), P_\theta) \leq \phi^W(P_{\theta_T}, P_\theta) < \infty. \tag{3.3.3}$$

Given some small positive constant ω, then there exists $\mu \in cl_{\tau_0}(V_\epsilon(P_{\theta_T}))$ satisfying

$$\phi^W(\mu, P_\theta) < \phi^W\Big(cl_{\tau_0}(V_\epsilon(P_{\theta_T})), P_\theta\Big) + \omega.$$

Set $v \in V_\epsilon(P_{\theta_T})$, and define $z(\alpha) = \alpha\mu + (1-\alpha)v$, where $0 < \alpha < 1$. Obviously, we have $z(\alpha) \in V_\epsilon(P_{\theta_T})$. By Lemma 3.3.1, the map $\zeta \to \phi(\zeta, P_\theta)$ is convex, hence we get

$$\begin{aligned} \phi^W(V_\epsilon(P_{\theta_T}), P_\theta) &\leq \lim_{\alpha\to 1} \phi^W(z(\alpha), P_\theta) \leq \lim_{\alpha\to 1} \Big(\alpha\phi^W(\mu, P_\theta) + (1-\alpha)\phi^W(v, P_\theta)\Big) \\ &= \phi^W(\mu, P_\theta) < \phi^W\Big(cl_{\tau_0}(V_\epsilon(P_{\theta_T})), P_\theta\Big) + \omega, \end{aligned} \tag{3.3.4}$$

where the equality holds since $\phi^W(v, P_\theta)$ is finite by (3.3.3). Combine (3.3.3) with (3.3.4) to get (3.3.2). This proves the conditional large deviation result. □

Using the above theorem, we obtain the following corollary.

Corollary 3.3.1. *Under the assumptions of Theorem 3.3.1, it holds*

$$\lim_{\epsilon\to 0}\phi^W(V_\epsilon(P_{\theta_T}),P_\theta)=\phi(P_{\theta_T},P_\theta).$$

Proof. By Lemma 3.3.1, the rate function $\phi^W(\mu,P_\theta)$ is a good rate function, hence it is lower semi-continuous; this implies

$$\lim_{\epsilon\to 0}\phi^W(V_\epsilon(P_{\theta_T}),P_\theta)\geq\phi(P_{\theta_T},P_\theta). \tag{3.3.5}$$

For any $\epsilon>0$, we have $\phi^W(P_{\theta_T},P_\theta)\geq\phi^W(V_\epsilon(P_{\theta_T}),P_\theta)$; this together with (3.3.5) completes the proof.

□

3.3.2 Divergences associated to the weighted sampling scheme

For any Q in $V_\epsilon(P_{\theta_T})$ rewrite the good rate function using the divergence notation

$$\phi^W(Q,P_\theta)=\int M^*\left(\frac{dQ}{dP_\theta}\right)dP_\theta=\int\varphi^W\left(\frac{dQ}{dP_\theta}\right)dP_\theta \tag{3.3.6}$$

from which $\phi^W(Q,P_\theta)$ is the divergence associated with the divergence function $\varphi^W:=M^*$.

Commuting P_{θ_T} and P_θ in (3.3.6) and introducing the conjugate divergence function $\widetilde{\varphi^W}$ yields

$$\phi^W(Q,P_\theta)=\int\varphi^W\left(\frac{dQ}{dP_\theta}\right)dP_\theta=\int\widetilde{\varphi^W}\left(\frac{dP_\theta}{dQ}\right)dQ=\widetilde{\phi^W}(P_\theta,Q). \tag{3.3.7}$$

By Theorem 3.3.1, maximizing $P_\theta(P_{n,\theta}^{W'}\in V_\epsilon(P_n^W)|P_n)$ amounts to minimize $\phi^W(V_\epsilon(P_{\theta_T}),P_\theta)$. A final approximation now yields the form of the criterion to be estimated in order to define the MLE in the present setting. As $\epsilon\to 0$ the asymptotic order of $\phi^W(V_\epsilon(P_{\theta_T}),P_\theta)$ is equal to $\widetilde{\phi^W}(P_\theta,P_{\theta_T})$ by Corollary 3.3.1 and (3.3.7), which is a proxy of $\phi^W(P_{\theta_T},P_\theta)$ and therefore the theoretical criterion to be optimized in θ.

We now state the dual form of the theoretical criterion $\widetilde{\phi^W}(P_\theta,P_{\theta_T})$ using the dual form (3.1.3) and (3.1.4). It holds

$$\widetilde{\phi^W}(P_\theta,P_{\theta_T})=\sup_{\alpha\in\mathcal{U}}\int\widetilde{h}(\theta,\alpha,x)dP_{\theta_T}(x) \tag{3.3.8}$$

with

$$\widetilde{h}(\theta,\alpha,x)=\int\left(\widetilde{\varphi^W}\right)'\left(\frac{dP_\theta}{dP_\alpha}\right)dP_\theta-\left(\widetilde{\varphi^W}\right)^{\#}\left(\frac{dP_\theta}{dP_\alpha}(x)\right)$$

We now turn to the definition of the MLE in this context, estimating the criterion and deriving the estimate.

3.3.3 MLE under weighted sampling

Using the dual representation of divergences, the natural estimator of $\phi(P_\theta, P_{\theta_T})$ is

$$\widetilde{\phi_n}(P_\theta, P_{\theta_T}) := \sup_{\alpha \in \mathcal{U}} \left\{ \int \widetilde{h}(\theta, \alpha, x)\ dP_n^W(x) \right\}. \tag{3.3.9}$$

From now on, we will use $\phi(\theta, \theta_T)$ to denote $\phi(P_\theta, P_{\theta_T})$; whence the resulting estimator of $\phi(\theta_T, \theta_T)$ is

$$\widetilde{\phi_n}(\theta_T, \theta_T) := \inf_{\theta \in \Theta} \widetilde{\phi_n}(\theta, \theta_T) = \inf_{\theta \in \Theta} \sup_{\alpha \in \mathcal{U}} \left\{ \int \widetilde{h}(\theta, \alpha, x)\ dP_n^W(x) \right\}$$

and the resulting MLE of θ_T is obtained as the minimum dual $\widetilde{\phi^W}$ estimator

$$\widehat{\theta}_{ML,W} := \arg \inf_{\theta \in \Theta} \sup_{\alpha \in \mathcal{U}} \left\{ \int \widetilde{h}(\theta, \alpha, x)\ dP_n^W(x) \right\}. \tag{3.3.10}$$

Formula (3.3.10) indeed defines a Maximum Likelihood estimator, in the vein of (3.2.1) and (3.2.11). This estimator requires no grouping nor smoothing.

3.4 Bahadur slope of minimum divergence tests for weighted data

Consider the test of some null hypothesis H0: $\theta_T = \theta$ versus a simple hypothesis H1 $\theta_T = \theta'$.

We consider two competitive statistics for this problem. The first one is based on the estimate of $\widetilde{\phi}^W(P_\alpha, P_\beta)$ defined for all (α, β) in $\Theta \times \Theta$ through

$$T_n(\alpha) := \sup_{\eta \in \Theta} \int \widetilde{\varphi}^W \left(\frac{p_\alpha}{p_\eta} \right) p_\eta d\mu - \int \left(\widetilde{\varphi}^W \right)^* \left(\frac{p_\alpha}{p_\beta} \right) dP_n^W$$

where the i.i.d. sample $X_1, ..., X_n$ has distribution P_β. The test statistics $T_n(\theta)$ converges to 0 under H0.

A competitive statistics $\widehat{\psi}(\theta)$ writes

$$\widehat{\psi}(\theta) := \psi\left(\theta, P_n^W\right)$$

where $Q \to \psi(\theta, Q)$ is assumed to satisfy $\psi(\theta, P_\theta) = 0$, and is $\tau-$continuous with respect to Q, which implies that under H0 the following Large Deviation Principle holds

$$\begin{aligned} \lim_{n \to \infty} \frac{1}{n} \log P_\theta \left(\widehat{\psi}(\theta) \geq t \right) &= -I(t) \\ &= -\inf \left\{ \phi^W(P_\theta, Q), \psi(\theta, Q) \geq t \right\} \end{aligned} \tag{3.4.1}$$

for any positive t. Also we assume that under H1,$\widehat{\psi}(\theta)$ converges to $\psi(\theta, P_{\theta'})$

$$\lim_{n \to \infty} \widehat{\psi}(\theta) =_{\theta'} \psi(\theta, P_{\theta'}) \tag{3.4.2}$$

where (3.4.2) stands in probability under θ'.

We now state the Bahadur slope of the test $\widehat{\phi^W}(\theta,\theta)$.

Under H0

$$\begin{aligned}\lim_{n\to\infty}\frac{2}{n}\log P_\theta\left(T_n(\theta)\geq t\right) &= -2\inf\left\{\phi^W\left(P_\theta,Q\right),\widetilde{\phi}^W\left(Q,P_\theta\right)\geq t\right\}\\ &= -2\inf\left\{\phi^W\left(P_\theta,Q\right),\phi^W\left(P_\theta,Q\right)\geq t\right\}\\ &= -2t\end{aligned}$$

while, under H1

$$\lim_{n\to\infty}T_n(\theta)=\phi^W\left(P_\theta,P_{\theta'}\right)\text{ in probability}$$

since P_n^W converges weakly to $P_{\theta'}$.

It follows that the Bahadur slope of the minimum divergence test $\widehat{\phi^W}(\theta,\theta)$ is

$$e_{T_n(\theta)}=-2\phi^W\left(P_\theta,P_{\theta'}\right).$$

Let us evaluate the Bahadur slope of the test $\widehat{\psi}(\theta)$.

Following (3.4.1) and (3.4.2) it holds

$$e_{\widehat{\psi}(\theta)}=-2\inf\left\{\phi^W\left(P_\theta,Q\right),\psi\left(\theta,Q\right)\geq\psi\left(\theta,P_{\theta'}\right)\right\}.$$

Since $\inf\left\{\phi^W\left(P_\theta,Q\right),\psi\left(\theta,Q\right)\geq\psi\left(\theta,P_{\theta'}\right)\right\}\leq\phi^W\left(P_\theta,P_{\theta'}\right)$ it follows that $e_{\widehat{\psi}(\theta)}\leq e_{T_n(\theta)}$.

We have proved

Proposition 3.4.1. *Under the weighted sampling the test statistics $\widehat{\psi}(\theta)$ is Bahadur efficient among all tests which are empirical versions of τ_0- continuous functionals.*

3.5 Weighted sampling in exponential families

In this short section we show that MLE's associated with weighted sampling are specific with respect to the weighting; this is in contrast with the unweighted sampling (i.i.d. simple sampling), under which all minimum divergence estimators coincide with the standard MLE; see [15].

Let

$$p_\theta(x)=\exp\left[\theta t(x)-C(\theta)\right]d\mu(x)\tag{3.5.1}$$

be an exponential family with natural parameter θ in an open set Θ in $\mathbb{R}^d$, and where μ denotes a common dominating measure for the model. We assume that this family is full i.e. that the Hessian matrix $(\partial^2/\partial\theta^2)\,C(\theta)$ is definite positive. Recall that under the standard i.i.d. $X_1,...,X_n$ sampling the MLE θ_{ML} of θ satisfies

$$\nabla C(\theta)_{\theta_{ML}}=\frac{1}{n}\sum_{i=1}^{n}t\left(X_i\right).$$

Under the weighted sampling $W_1\ ,...,W_n$ corresponding to the divergence function φ^W, conditionally on the observed data $x_1,...,x_n$ the MLE writes

$$\theta_{ML,W} := \arg \inf_{\theta \in \Theta} \sup_{\alpha \in \mathcal{U}} \int \left(\widetilde{\varphi^W}\right)' \left(\frac{p_\theta}{p_\alpha}\right) p_\theta d\mu - \int \left(\widetilde{\varphi^W}\right)^{\#} \left(\frac{p_\theta}{p_\alpha}\right) dP_n^W.$$

We prove that $\theta_{ML,W}$ satisfies

$$\nabla C(\theta)_{\theta_{ML,W}} = \frac{1}{n} \sum_{i=1}^{n} W_i t\left(x_i\right).$$

Denote

$$M_n\left(\theta, \alpha\right) := \int \left(\widetilde{\varphi^W}\right)' \left(\frac{p_\theta}{p_\alpha}\right) p_\theta d\mu - \int \left(\widetilde{\varphi^W}\right)^{\#} \left(\frac{p_\theta}{p_\alpha}\right) dP_n^W.$$

Clearly, subsituting using (3.5.1) it holds for all θ

$$\inf_{\theta \in \Theta} \sup_{\alpha \in \mathcal{U}} M_n\left(\theta, \alpha\right) \geq M_n\left(\theta, \theta\right) = 0. \tag{3.5.2}$$

We prove that $M_n\left(\theta_{ML,W}, \alpha\right)$ is maximal for $\alpha = \theta_{ML,W}$ which closes the proof.

Let $X_1, ..., X_n$ be n i.i.d. random variables with common distribution P_{θ_T} with θ_T in Θ. Introduce

$$M_n\left(\theta, \alpha\right) := \int \varphi' \left(\frac{dP_\theta}{dP_\alpha}\right) dP_\theta - \frac{1}{n} \sum_{i=1}^{n} \varphi^{\#} \left(\frac{dP_\theta}{dP_\alpha}\left(X_i\right)\right)$$

We prove that

$$\alpha = \theta_{ML,W} \text{ is the unique maximizer of } M_n\left(\theta_{ML,W}, \alpha\right) \tag{3.5.3}$$

which yields

$$\inf_{\theta} \sup_{\alpha} M_n\left(\theta, \alpha\right) \leq \sup_{\alpha} M_n\left(\theta_{ML,W}, \alpha\right) = M_n\left(\theta_{ML,W}, \theta_{ML,W}\right) = 0 \tag{3.5.4}$$

which together with (3.5.2) completes the proof.

Define

$$\begin{aligned} M_{n,1}\left(\theta, \alpha\right) &:= \int \varphi' \left(\exp A(\theta, \alpha, x)\right) \exp B\left(\theta, x\right) d\lambda(x) \\ M_{n,2}\left(\theta, \alpha\right) &:= \frac{1}{n} \sum_{i=1}^{n} W_i \exp\left(A\left(\theta, \alpha, x_i\right)\right) \varphi' \left(\exp A(\theta, \alpha, x_i)\right) \\ M_{n,3}\left(\theta, \alpha\right) &:= \frac{1}{n} \sum_{i=1}^{n} W_i \varphi\left(\exp A(\theta, \alpha, x_i)\right) \end{aligned}$$

with

$$\begin{aligned} A(\theta, \alpha, x) &:= T(x)'\left(\theta - \alpha\right) + C(\alpha) - C(\theta) \\ B(\theta, x) &:= T(x)'\theta - C(\theta). \end{aligned}$$

It holds

$$M_n\left(\theta, \alpha\right) = M_{n,1}\left(\theta, \alpha\right) - M_{n,2}\left(\theta, \alpha\right) + M_{n,3}\left(\theta, \alpha\right)$$

with

$$\frac{\partial}{\partial\alpha}M_{n,1}\left(\theta,\alpha\right)_{\alpha=\theta}=-\varphi^{(2)}\left(1\right)\left[\nabla C\left(\theta\right)-\nabla C\left(\alpha\right)_{\alpha=\theta}\right]=0$$

for all θ,

$$\frac{\partial}{\partial\alpha}M_{n,2}\left(\theta,\alpha\right)_{\alpha=\theta_{ML,W}}=\varphi^{(2)}\left(1\right)\frac{1}{n}\sum_{i=1}^{n}W_i\left[-T(x_i)+\nabla C\left(\alpha\right)_{\alpha=\theta_{ML,W}}\right]=0$$

and

$$\frac{\partial}{\partial\alpha}M_{n,3}\left(\theta_{ML,W},\alpha\right)=\frac{1}{n}\sum_{i=1}^{n}W_i\left[-T(x_i)+\nabla C\left(\alpha\right)_{\alpha=\theta_{ML,W}}\right]=0$$

where the two last displays hold iff $\alpha=\theta_{ML}$. Now

$$\begin{aligned}\frac{\partial^2}{\partial\alpha^2}M_{n,1}\left(\theta,\alpha\right)_{\alpha=\theta_{ML,W}}&=\left(\varphi^{(3)}(1)+2\varphi^{(2)}(1)\right)\left(\partial^2/\partial\theta^2\right)C(\theta_{ML,W})\\ \frac{\partial^2}{\partial\alpha^2}M_{n,2}\left(\theta,\alpha\right)_{\alpha=\theta_{ML,W}}&=\left(\varphi^{(3)}(1)+4\varphi^{(2)}(1)\right)\left(\partial^2/\partial\theta^2\right)C(\theta_{ML,W})\\ \frac{\partial^2}{\partial\alpha^2}M_{n,3}\left(\theta_{ML},\alpha\right)_{\alpha=\theta_{ML,W}}&=\varphi^{(2)}(1)\left(\partial^2/\partial\theta^2\right)C(\theta_{ML,W}),\end{aligned}$$

whence

$$\begin{aligned}\frac{\partial}{\partial\alpha}M_{n}\left(\theta,\alpha\right)_{\alpha=\theta_{ML,W}}&=0\\ \frac{\partial^2}{\partial\alpha^2}M_{n}\left(\theta,\alpha\right)_{\alpha=\theta_{ML,W}}&=-\varphi^{(2)}(1)\left(\partial^2/\partial\theta^2\right)C(\theta_{ML,W})\end{aligned}$$

which proves (3.5.3), and closes the proof.

In contrast with the i.i.d. sampling case minimum divergence estimators in exponential families under appropriate weighted sampling do not coincide independently upon the divergence.

3.6 Weak behavior of the weighted sampling MLE's

The distribution of the estimator is obtained under the sampling scheme which determines its form. Hence under the weighted sampling one. So the observed sample $x_1,...,x_n$ is considered non random, and is assumed to satisfy

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}=P_{\theta_T}$$

and randomness is due to the set of i.i.d. weights $W_1,...,W_n$.

All those estimators can be written as approximate linear functionals of the weighted empirical measure P_n^W. Therefore all the proofs in [19] can be adapted to the present estimators. Even the asymptotic variances of the estimators are the same, and subsequently, Wilk's tests , confidence areas, minimum sample sizes certifying a given asympptotic power, etc, remain unchanged. The only arguments to be noted are the following: All arguments pertaining to laws of large numbers for functionals of the empirical measure carry over to the present setting, conditionally on the observations $x_1, ..., x_n$. Indeed consider a statistics

$$U_n := \frac{1}{n}\sum_{i=1}^{n} W_i f(x_i)$$

where the function f satisfies

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(x_i) = \mu_{1,f} < \infty$$

and

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f^2(x_i) = \mu_{2,f} < \infty.$$

Then clearly

$$\lim_{n\to\infty} EU_n = \mu_{1,f}$$

and

$$\lim_{n\to\infty} VarU_n = \mu_{2,f} - (\mu_{1,f})^2 .$$

Weak behavior of the estimates follow also from similar arguments: Consider for example the statistics

$$T_n := \sqrt{n}\,(U_n - \mu_{1,f})\,/\sqrt{\mu_{2,f} - (\mu_{1,f})^2}.$$

Using Lindeberg Central limit theorem for triangular arrays , we obtain that T_n is asymptotically standard normal conditionally upon $x_1, ..., x_n$. It follows that the limit distributions of $\widetilde{\phi^W}(\theta, \theta_T)$ and of $\widehat{\theta}_{ML,W}$ conditionally on $x_1, ..., x_n$ coincide with those of $\phi_n(\theta, \theta_T)$ and of $\widehat{\theta}_n$ as stated in [19] under the i.i.d. sampling. Also all results pertaining to tests of hypotheses are similar, as is the possibility to handle non regular models.

3.7 Proof of Lemma 3.3.1

Proof. Recall that $B(S)$ denotes the class of all bounded measurable functions on S. Write $B'(S)$ as the algebraic dual of $B(S)$. We equip $B'(S)$ with $B(S)$-topology, it is the weakest topology which makes continuous the following linear functional:

$$\zeta \mapsto < f, \zeta >: B'(S) \to \mathbb{R},\ \text{for all } f \text{ in } B(S),$$

where $< f, \zeta >$ denotes the value of $f(\zeta)$. It follows that $\mathcal{M}(S)$ is included in $B'(S)$ and is endowed with the τ_0-topology induced by $B(S)$. Construct the projection: $p_{f_1,...,f_m} : B'(S) \to \mathbb{R}^m, m \in \mathbb{Z}_+$, namely, $p_{f_1,...,f_m}(\zeta) = (< f_1, \zeta >, ..., < f_m, \zeta >), f_1, ..., f_m \in B(S)$. Then for $p_{f_1,...,f_m}(P^W_{n,\theta}) = (< f_1, P^W_{n,\theta} >, ..., < f_m, P^W_{n,\theta} >)$ we define the corresponding limit logarithm moment generating function as follows

$$h(t) := \lim_{n\to\infty} \frac{1}{n}\log \mathbb{E}(\exp(n < t, Y_m >)) = \lim_{n\to\infty} \frac{1}{n}\log \mathbb{E}(\exp(\sum_{j=1}^{m} < t_j f_j, \sum_{i=1}^{n} W_i \delta_{x_i} >))$$

$$= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \log \mathbb{E}\exp\left(\sum_{j=1}^{m} t_j f_j(x_i) W_i\right) = \int \left(\sum_{j=1}^{m} M(t_j f_j)\right) dP_\theta$$

where $t = (t_1, ..., t_m) \in \mathbb{R}^m$ and $Y_m = (< f_1, P_{n,\theta}^W >, ..., < f_m, P_{n,\theta}^W >)$. The function $h(t)$ is finite since $f \in B(S)$. $M(f)$ is Gateaux-differentiable since the function $s \to M(f + sg)$ is differentiable at $s = 0$ for any $f, g \in B(S)$

$$\frac{d}{ds} M(f + sg)|_{s=0} = \frac{\int g e^f dP_W}{\int e^f dP_W},$$

where P_W is the law of W. Further, the Gateaux-differentiability of $M(f)$ together with the interchange of integration and differentiation justified by dominated convergence theorem show that $h(t)$ is also Gateaux-differentiable in $t = (t_1, ..., t_m)$. Hence by the Gartner-Ellis Theorem (see e.g. Theorem 2.3.6 of [31]), $p_{f_1,...,f_m}(P_{n,\theta}^W)$ satisfies the LDP in $\mathbb{R}^m$ with the good rate function

$$\begin{aligned}\Phi_{f_1,...,f_m}(< f_1, \zeta >, ..., < f_m, \zeta >) &= \sup_{t_1,...,t_m \in \mathbb{R}} \left\{ \sum_{i=1}^m t_i < f_i, \zeta > - \int M\left(\sum_{i=1}^m t_i f_i\right) dP_\theta \right\} \\ &\leq \sup_{f \in B(S)} \Phi_f(< f, \zeta >) := \phi^W(\zeta, P_\theta). \end{aligned} \qquad (3.7.1)$$

Since m is arbitrary positive integer, by Dawson-Gartner's Theorem (see e.g. Theorem 4.6.1 of [31]), $P_{n,\theta}^W$ satisfies the LDP in $B'(S)$ with the good rate function $\phi^W(\zeta, P_\theta)$, which is:

$$\begin{aligned}\phi^W(\zeta, P_\theta) &= \sup_{f \in B(S)} \Phi_f(< f, \zeta >) = \sup_{f \in B(S)} \left\{ \int_S f(x)\zeta(dx) - \int_S M(f) P_\theta(dx) \right\} \\ &= \int_S M^*\left(\frac{d\zeta}{dP_\theta}\right) dP_\theta,\end{aligned}$$

note that $B'(S)$ is endowed with the τ_0-topology, the proof of last equality is given below. Here we always assume ζ is absolutely continuous with respect to P_θ, otherwise $\phi^W(\zeta, P_\theta) = \infty$. Consider $\mathcal{M}(S) \subset B'(S)$, and set $\phi^W(\zeta, P_\theta) = \infty$ when $\zeta \notin \mathcal{M}(S)$. Hence $P_{n,\theta}^W$ satisfies the LDP in $\mathcal{M}(S)$ with the rate function $\phi^W(\zeta, P_\theta)$, for $\zeta \in \mathcal{M}(S)$. As mentioned before, $\mathcal{M}(S)$ is endowed with the topology induced by $B'(S)$, namely the τ_0-topology.

Nown we give another representation of the rate function $\phi^W(\zeta, P_\theta)$. We have:

$$\begin{aligned}&\sup_{\zeta \in \mathcal{M}(S)} \left\{ \int_S f(x)\zeta(dx) - \int_S M^*\left(\frac{d\zeta}{dP_\theta}\right) dP_\theta \right\} \\ &= \sup_{\zeta \in \mathcal{M}(S)} \left\{ \int_S \left(\int_S f d\zeta - M^*\left(\frac{d\zeta}{dP_\theta}\right)\right) dP_\theta \right\} \leq \int_S M(f) dP_\theta,\end{aligned}$$

where the inequality holds from the duality lemma and when $d\zeta = (dP_\theta)M'(f)$ the equality holds. Using once again the duality lemma, we obtain the following identity:

$$\int_S M^*\left(\frac{d\zeta}{dP_\theta}\right) dP_\theta = \sup_{\zeta \in \mathcal{M}(S)} \left\{ \int_S f(x)\zeta(dx) - \int_S M(f) dP_\theta \right\} = \phi^W(\zeta, P_\theta).$$

The convexity of the rate function $\zeta \to \phi^W(\zeta, P_\theta)$ holds from Theorem 7.2.3 of [31] where they show the convexity of $\phi^W(\zeta, P_\theta)$ on $\mathcal{M}(S)$ endowed with $B(S)$-topology. Hence this is also applied to τ_0-topology which is induced by $B(S)$-topology.

This completes the proof of the lemma.

□

Remark 3.7.1. *By the classical Gartner-Ellis Theorem, in (3.7.1), the essential smoothness of* $h(t)$ *is needed for* $\Phi_{f_1,\ldots,f_m}$ *to be a "good rate function". But on a locally convex Hausdorff topological vector space, the essential smoothness of* $h(t)$ *can be reduced to Gateaux differentiability; see Corollary* 4.6.14 *(page* 167*) and the proof Theorem* 6.2.10 *(page* 265*) of* [31].

Remark 3.7.2. *Since* $\Phi_{f_1,\ldots,f_m}(< f_1, \zeta >, \ldots, < f_m, \zeta >)$ *is a good rate function in* $\mathbb{R}^m$*, its level sets* $\Phi^{-1}_{f_1,\ldots,f_m}(\alpha) = \{(y_1, \ldots, y_m) \in \mathbb{R}^m : \Phi_{f_1,\ldots,f_m}(y_1, \ldots, y_m) \leq \alpha\}$ *are compact, for all* α *in* $[0, \infty)$*. Denote the projective limit of* $\Phi^{-1}_{f1,\ldots,f_m}(\alpha)$ *by* $\Phi^{-1}_f(\alpha) = \varprojlim \Phi^{-1}_{f1,\ldots,f_m}(\alpha)$*. According to Tychonoff's theorem, the projective limit* $\Phi^{-1}_f(\alpha)$ *of the compact set* $\Phi^{-1}_{f1,\ldots,f_m}(\alpha)$ *is still compact, so* $\phi^W(\zeta, P_\theta) = \sup_{f\in B(S)} \Phi_f(< f, \zeta >)$ *is also a good rate function in* $(\mathcal{M}(S), \mathcal{B}(\mathcal{M}))$.

Bibliography

[1] Arenal-Gutierrez, E. and Matran, C. (1996). A zero–one law approach to the central limit theorem for the weighted bootstrap mean, *Ann. Probab.* Vol. 24, pp. 532–540.

[2] Armendáriz, I. and Loulakis, M. (2011). Conditional distribution of heavy tailed random variables on large deviations of their sum. *Stochastic Process. Appl.* 121, no. 5, 1138–1147.

[3] Barbe, P. and Bertail, P. (1995). The Weighted Bootstrap, Lecture Notes in Statistics, Springer-Verlag, New York.

[4] Barbe, P. and Broniatowski, M. (2004). Blowing number of a distribution for a statistics and loyal estimators. *Statist. Probab. Lett.* 69, no. 4, 465–475.

[5] Bar-Lev, Shaul K. and Enis, P. (1986). Reproducibility and natural exponential families with power variance functions. *Ann. Statist.* 14 , no. 4, 1507–1522.

[6] Barndorff-Nielsen, O. (1978). Information and exponential families in statistical theory, Wiley Series in Probability and Mathematical Statistics. Chichester: John Wiley & Sons.

[7] Barvinok, Alexander (2005). Lecture note: Math 710: Measure Concentration. http://www.math.lsa.umich.edu/ barvinok/total710.pdf

[8] Beirlant, J., Broniatowski, M., Teugels, Jozef L. and Vynckier, P. (1995). The mean residual life function at great age: applications to tail estimation. Extreme value theory and applications (Villeneuve d'Ascq, 1992). *J. Statist. Plann. Inference* 45, no. 1-2, 21–48.

[9] Bhattacharya, R. N. and Rao, Ranga R. (2010). Normal approximation and asymptotic expansions, Society for Industrial and Applied Mathematics, Philadlphia.

[10] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). Regular Variation Cambridge University Press, Cambridge.

[11] Book, Stephen A. (1978). Probabilities of very large deviations. *J. Austral. Math. Soc. Ser. A* 25, no. 3, 332–347.

[12] Bouzebda, S. and Keziou, A. (2008). A test of independence in some copula models. *Math. Methods Statist.* 17 , no. 2, 123–137.

[13] Broniatowski, M. (1987). Grandes, très grandes et petites déviations pour des suites de variables aléatoires réelles indépendantes. (French) [Large, very large and small deviations for sequences of independent identically

distributed real random variables] *C. R. Acad. Sci. Paris Sér. I Math.* 305, no. 13, 627–630

[14] Broniatowski, M. (2003). Estimation of the Kullback-Leibler divergence. *Math. Methods Statist.* 12, no. 4, 391–409 .

[15] Broniatowski, M. (2011). Minimum divergence estimators, maximum likelihood and exponential families , arXiv:1108.0772.

[16] Broniatowski, M., Caron, V. (2012). Long runs under a conditional limit distribution arXiv:1202.0731

[17] Broniatowski, M. and Fuchs, A. (1995). Tauberian theorems, Chernoff inequality, and the tail behavior of finite convolutions of distribution functions. *Adv. Math.* 116, no. 1, 12–33.

[18] Broniatowski, M. and Keziou, A. (2006). Minimization of φ-divergences on sets of signed measures. *Studia Sci. Math. Hungar.* 43, no. 4, 403–442.

[19] Broniatowski, M. and Keziou, A. (2009). Parametric estimation and tests through divergences and the duality technique. *J. Multivariate Anal.* 100 , no. 1, 16–36.

[20] Broniatowski, M., Leorato, S. (2006). An estimation method for the Neyman chi-square divergence with application to test of hypotheses. *J. Multivariate Anal.* 97, no. 6, 1409–1436.

[21] Broniatowski, M. and Mason, D. M. (1994). Extended large deviations. *J. Theoret. Probab.* 7, no. 3, 647–666.

[22] Broniatowski, M. and Vajda, I. (2012). Several applications of divergence criteria in continuous families, to appear *Kybernetika.*

[23] Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Stat.,* Vol. 23, pp. 493-507.

[24] Chernoff, H. and Scheffe, H. (1952). A generalized of the neyman-pearson fundamental lemma, *The Annals of Math. Stat.*, Vol. 23, No. 2, pp. 213-225.

[25] Cramer, H. (1938). Sur un nouveau theoreme-limite de la theorie des probabilités, *Actualites Sci. Ind.,* No. 736, Paris.

[26] Csiszár, I. (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. (German) *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 8 85–108.

[27] Csiszár (1984). Imre Sanov property, generalized I-projection and a conditional limit theorem. *Ann. Probab.* 12, no. 3, 768–793,

[28] Dantzig, G.B. and Wald, A. (1951). On the fundamental lemmna of Neyman and Pearson, *Annals of Math. Stat.*, Vol. 22, pp. 87-93.

[29] Darling, D. A. (1952). The influence of the maximum term in the addition of independent random variables. Trans. Amer. Math. Soc. 73. 95–107.

[30] Dembo, A. and Zeitouni, O. (1996). Refinements of the Gibbs conditioning principle. *Probab. Theory Related Fields* 104, no. 1, 1–14.

[31] Dembo, A. and Zeitouni, O. (1998). Large Deviations Techniques And Applications. Springer Verlag, New York, second edition.

[32] Diaconis, P. and Freedman, D. A. (1988). Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. *J. Theoret. Probab.* 1, no. 4, 381–410.

[33] Donoho, David and Huber, Peter J. (1983). The notion of breakdown point. *A Festschrift for Erich L. Lehmann, pp. 157-184, Wadsworth Statist./Probab. Ser., Wadsworth, Belmont, CA.*

[34] Efron, B. (1979). Bootstrap methods: another look at the jackknife, *Ann. Statist.*, Vol. 7, No. 1, pp. 1-26.

[35] Erdős, Paul and Rényi, Alfréd (1970). On a new law of large numbers. *J. Analyse Math.* 23 103–111.

[36] Feller, William (1971). An introduction to probability theory and its applications. Vol. II. Second edition John Wiley & Sons, Inc., New York-London-Sydney xxiv+669 pp

[37] Field, C. and Ronchetti, E. (1990). Small sample asymptotics. Institute of Mathematical Statistics Lecture Notes-Monograph Series, 13. Institute of Mathematical Statistics, Hayward, CA.

[38] Frisch, U. and Sornette, D. (1997). Extreme deviations and Applications. *J. Phys.* I, France, 7, 1155-1171

[39] Geffroy, Jean (1958). Stabilité presque complète des valeurs extrêmes d'un échantillon et convergence presque complète du milieu vers une limite certaine. (French) *C. R. Acad. Sci. Paris* 246, 224-226.

[40] Geffroy, Jean (1958). Etude de la stabilité presque certaine des valeurs extrêmes d'un échantillon et de la convergence presque certaine de son milieu. (French) *C. R. Acad. Sci. Paris* 246, 1154-1156.

[41] Geffroy, Jean (1958). Contribution à la théorie des valeurs extrêmes. (French) *Publ. Inst. Statist. Univ. Paris* 7 no. 3/4, 37-121.

[42] Ghosal, S., Ghosh, J. K. and van der Vaart, A. W. (2000). Convergence rates of posterior distributions, *Ann. Statist.,* 28 pp. 500–531.

[43] Grendar, M. and Judge, G. (2009). Asymptotic equivalence of empirical likelihood and Bayesian MAP, *Ann. Statist.,* Vol. 37, pp. 2445-2457.
g

[44] Hampel, Frank R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.,* 42, 1887-1896.

[45] Harremoës, P. and Vajda, I. (2008). On the Bahadur-efficient testing of uniformity by means of the entropy. *IEEE Trans. Inform. Theory* 54 , no. 1, 321–331.

[46] He, X., Jurečková, J., Koenker, R. and Portnoy, S. (1990). Tail Behavior of Regression Estimators and their Breakdown Points, *Ecnometrica*, Vol. 58, No. 5, pp. 1195-1214.

[47] Hodges, J. L., Jr. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. *Proc. Fifth Berkeley*

Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66) Vol. I: Statistics, pp. 163-186 Univ. California Press, Berkeley, Calif. 62.70.

[48] Höglund, Thomas (1979). A unified formulation of the central limit theorem for small and large deviations from the mean. *Z. Wahrsch. Verw. Gebiete* 49, no. 1, 105-117.

[49] Jensen, J.L. (1995). Saddlepoint Approximations. Oxford Statistical Science Series, vol 16. The Clarendon Press Oxford University Press, New York, Oxford Science Publications.

[50] Jørgensen, Bent (1987). Exponential dispersion models. With discussion and a reply by the author. *J. Roy. Statist. Soc. Ser. B* 49, no. 2, 127–162.

[51] Jurečková, J. (1979). Finite-sample comparison of L-estimators of location. *Comment. Math. Univ. Carolin.* 20 , no. 3, 507-518.

[52] Jurečková, J. (1981a). Tail-behavior of location estimators. *Ann. Statist.* 9, no. 3, 578-585.

[53] Jurečková, J. (1981b). Tail-behaviour of location estimators in nonregular cases. *Comment. Math. Univ. Carolin.* 22, no. 2, 365-375.

[54] Juszczak, D. and Nagaev, A. V. (2004). Local large deviation theorem for sums of i.i.d. random vectors when the Cramer condition holds in the whole space, *Probability and Mathematical Statistics*, Vol. 24, pp. 297-320.

[55] Keziou, A. (2003). Dual representation of φ-divergences and applications. *C. R. Math. Acad. Sci. Paris* 336, no. 10, 857–862

[56] Kleijn, B. J. K., van der Vaart, A. W. (2006). Misspecification in infinite-dimensional Bayesian statistics, *Ann. Statist.,* Vol. 34 pp. 837-877.

[57] Kušnier, J. and Mizera, I. (2001). Tail behavior and breakdown properties of equivariant estimators of location. *Ann. Inst. Statist. Math.* 53, no. 2, 244-261.

[58] Lanford, O. E. (1973). Entropy and equilibrium states in classical statistical mechanics, in Statistical Mechanics and Mathematical Problems. Berlin: Springer-Verlag.

[59] Letac, Gérard and Mora, Marianne (1990). Natural real exponential families with cubic variance functions. *Ann. Statist.* 18, no. 1, 1–37.

[60] Liese, F., Vajda, I. (1987). Convex statistical distances. Teubner-Texte zur Mathematik [Teubner Texts in Mathematics], 95. BSB B. G. Teubner Verlagsgesellschaft, Leipzig, ISBN: 3-322-00428-7 .

[61] Liese, F., Vajda, I. (2006). On divergences and informations in statistics and information theory. *IEEE Trans. Inform. Theory* 52, no. 10, 4394–4412

[62] Lynch, J., Sethuraman, J. (1987). Large deviations for processes with independent increments, *Ann. Probab.,* 15(2), pp. 610-627.

[63] Mammen, E. and Nandi, S. (2004). Bootstrap and resampling. Handbook of computational statistics, 467–495, Springer, Berlin.

[64] Mason, D. M. and Newton, M. A. (1992) A rank statistic approach to the consistency of a general bootstrap, *Ann. Statist.* Vol. 20, pp. 1611-1624.

[65] Morris, Carl N. (1982). Natural exponential families with quadratic variance functions. *Ann. Statist.* 10, no. 1, 65–80.

[66] Najim, J. (2002). A Cramer type theorem for weighted random variables, *Electron. J. Probab.,* Vol. 7.

[67] Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses, *Philos. Trans. Roy. Soc. London Ser. A*, Vol. 231, pp. 289-337.

[68] Neyman, J. and Pearson, E. S. (1936). Contributions to the theory of testing statistical hypotheses, *Stat. Res. Memoirs*, Vol.1 , pp. 1-37.

[69] Pelletier, Bruno (2011). Inference in ϕ-families of distributions. *Statistics* 45, no. 3, 223-236.

[70] Petrov, V. V. (1975). Sums of independent random variables. Translated from the Russian by A. A. Brown. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82. Springer-Verlag, New York-Heidelberg.

[71] Pinsker, M. S. (1964). Information and information stability of random variables and processes. Translated and edited by Amiel Feinstein Holden-Day, Inc., San Francisco, Calif.-London-Amsterdam xii+243 pp, 94.20

[72] Praestgaard, J. and Wellner, J.A. (1993). Exchangeably weighted bootstrap of the general empirical process, *Ann. Probab.* Vol. 21, pp. 2053–2086.

[73] Read, T. R. C. and Cressie, N. A. C. (1988). Goodness-of-fit statistics for discrete multivariate data. Springer Series in Statistics. Springer-Verlag, New York. xii+211 pp. ISBN: 0-387-96682-X

[74] Rényi, A. (1961). On measures of entropy and information. Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I pp. 547–561 Univ. California Press, Berkeley, Calif.

[75] Sornette, Didier (2004). Critical phenomena in natural sciences. Chaos, fractals, selforganization and disorder: concepts and tools. Second edition. Springer Series in Synergetics. Springer-Verlag, Berlin.

[76] Toma, A. and Broniatowski, M. (2011). Dual divergence estimators and tests: robustness results. *J. Multivariate Anal.* 102, no. 1, 20–36.

[77] Trashorras J. and Wintenberger O. Large deviations for bootstrapped empirical measures. arXiv:1110.4620v1

[78] Tweedie, M. C. K. (1947). Functions of a statistical variate with given means, with special reference to Laplacian distributions. Proc. Cambridge Philos. Soc. 43. 41–49.

[79] Withers, C. S. and Nadarajah, S. (2011). On the compound Poisson-gamma distribution. *Kybernetika* (Prague) 47, no. 1, 15–37.