Modèles hiérarchiques de Dirichlet à temps continu

Hafedh Faires

PUCVL
Pôle Universités
Centre Val de Loire

UNIVERSITÉ D'ORLÉANS

UNIVERSITE D'ORLEANS

ÉCOLE DOCTORALE SCIENCES ET TECHNOLOGIES

LABORATOIRE DE MATHÉMATIQUES ET APPLICATIONS, PHYSIQUE MATHÉMATIQUES D'ORLÉANS

THÈSE présentée par :

Hafedh Faires

Soutenue le : **3 Octobre 2008**

pour obtenir le grade de **Docteur de l'univérsité d'Orléans**

Discipline : **Mathématiques**

Modèles hiérarchiques de Dirichlet à temps continu

THÈSE dirigée par :

Mr. Richard EMILION	Professeur, Université d'Orléans.

RAPPORTEURS :

M. Fabrice GAMBOA	Professeur, IMT Université Paul Sabatier.
M. Lancelot JAMES	Professeur, Hong Kong university.

JURY :

M. Paul DEHEUVELS	Professeur, Université Paris 6.
Mme. Anne ESTRADE	Professeur, Université de Paris 5.
M. Richard EMILION	Professeur, Université d'Orléans.
M. Fabrice GAMBOA	Professeur, IMT Université Paul Sabatier.
M. Abdelhamid HASSAIRI	Professeur, Université de Sfax, Tunisie.
M. Dominique LEPINGLE	Professeur, Université d'Orléans.

Résumé

Nous étudions les processus de Dirichlet dont le paramètre est une mesure proportionnelle à la loi d'un processus temporel, par exemple un mouvement Brownien ou un processus de saut Markovien. Nous les utilisons pour proposer des modèles hiérarchiques bayésiens basés sur des équations différentielles stochastiques en milieu aléatoire. Nous proposons une méthode pour estimer les paramètres de tels modèles et nous l'illustrons sur l'équation de Black-Scholes en milieu aléatoire.

Abstract

We consider Dirichlet processes whose parameter is a measure proportional to the distribution of a continuous time process, such as a Brownian motion or a Markov jump process. We use them to propose some Bayesian hierarchical models based on stochastic differential equations in random environment. We propose a method for estimating the parameters of such models and illustrate it on the Black-Scholes equation in random environment.

Remerciements

Tout d'abord, je tiens à exprimer mon profond respect et ma grande reconnaissance au professeur Richard EMILION, mon directeur de thèse, qui m'a proposé un sujet de recherche très intéressant, m'a initié à la recherche et a encadré mes travaux. Je le remercie pour les nombreuses connaissances qu'il m'a transmises mais aussi pour sa patience et sa bonne humeur.
Mes remerciements s'adressent également aux Professeurs Fabrice GAMBOA et Paul Deheuvels, les rapporteurs de ce manuscrit de thèse.

Je souhaite aussi remercier le jury Fabrice Gamboa, Anne Estrade, Abdelhamid Hassairi Lancelot James et Dominique Lepingle.

Des remerciements particuliers pour les professeurs Dominique LEPINGLE, Romain ABRAHAM, Gilles ROYER, Pierre ANDREOLETTI qui m'ont apporté des réponses pertinentes à certaines questions.

Mes remerciements vont à l'égard du Directeur, le professeur Jean-Philippe ANKER, et des membres du laboratoire MAPMO de l'Université d'Orléans pour leur acceuil sympathique, les conditions de travail, et pour toutes les discussions pendant ces années de thèse.

Ce travail passionnant a été effectué au côté de camarades de bureau dynamiques que je remercie amicalement : El Safadi Mohammad, Lina Aballah, Pierre Clare, Bassirou Diagne et d'autres avec lesquels j'ai eu l'honneur de partager ces années riches d'expériences et de projets.

Un merci particulier à mon ancien enseignant Kamel MOKNI, pour sa motivation et ses encouragements à mon attention.

Finalement, je voudrais saluer le soutien constant de mes parents et leur confiance dans les choix et les décisions que j'ai prises et je voudrais chaleureusement remercier toute ma famille qui m'a apporté un climat serein indispensable pour un travail de thèse.

Contents

Chapter 1

Introduction

L'objectif de ce travail est de proposer un nouveau modèle hiérarchique comprenant un processus de Dirichlet comme loi a priori, on dira brièvement modèle hiérarchique de Dirichlet, qui soit adapté à l'analyse de trajectoires temporelles, notamment celles qui sont régies par des EDS (équations differentielles stochastiques) en milieu aléatoire.

Le processus de Dirichlet est une loi aléatoire, c'est-à-dire une variable aléatoire à valeurs dans l'ensemble $\mathcal{P}(V)$ des mesures de probabilités sur un ensemble V d'observations. Nous utiliserons l'abréviation anglaise RD (Random Distribution).
Les RDs sont très intéréssants aussi bien du point de vue théorique que du point de vue appliqué.

Nous utiliserons dans ce travail quatre points, considéres comme importants dans l'histoire de ce processus.
• En 1969, dans un article fondamental très célèbre, Thomas S. Ferguson construit le processus de Dirichlet, généralisation des lois de Dirichlet, devenu depuis un outil remarquable et classique en Statistique bayésienne non paramétrique.
• En 1973, J.F.C. Kingman définit des RDs, dits de Poisson-Dirichlet, aux

propriétés intéressantes et liées à la représentation des processus de Dirichlet utilisant le processus Gamma.

• En 1974, motivé par les applications, C.A. Antoniak introduit et étudie les mélanges de processus de Dirichlet.

• En 1994, une méthode constructive des processus de Dirichlet, dite *stick-breaking*, utilisée lors de mises en oeuvre informatique, est élaborée par Jayaram Sethuraman [34].

Les applications concernent pratiquement tous les domaines : biologie, écologie, génétique , informatique, etc...

Récemment ce champ d'application a été étendu en utilisant avec succès des modèles hiérarchiques de Dirichlet en classification par estimation de mélanges de lois à partir de données non temporelles, voir par exemple : Ishwaran et Zarepour (2000), Ishwaran et James (2002) and (2003), Brunner et Lo (2002), Emilion (2001, 2003, 2004), Bigelow and Dunson, (2007), Kacperczyk et al., (2003). Dans ces articles le paramètre du processus de Dirichlet est une mesure proportionnelle à une loi classique sur $\mathbb{R}^n$.

Le présent travail consiste à étudier l'extension de ces modèles hiérarchiques au cas de données temporelles, en utillisant notamment le processus de Dirichlet sur des espaces de trajectoires, le paramètre étant une mesure proportionnelle à une loi de processus temporel (Emilion, 2005).

A partir de l'observation d'une seule trajectoire, il nous est possible de détecter des régimes de durée aléatoire, lorsque le processus temporel suit une EDS en milieu aléatoire. Le milieu est représenté par une chaîne de Markov à temps continu dont les états, qui modélisent les régimes, jouent le rôle que jouent les classes en classification.

Le modèle hiérarchique bayésien que nous introduisons place notamment un processus de Dirichlet comme a priori sur l'espace des trajectoires de cette chaîne. L'estimation des paramètres est bâtie à partir d'un échantillonneur de Gibbs.

Nous traiterons à titre d'exemple l'EDS de Black-Scholes en finance, le drift et la volatilité étant stochastiques. Le modèle hiérarchique utilisé dans ce cas ne suppose donc plus le processus gaussien puisque ses marginales sont des mélanges compliqués de gaussiennes.

La thèse est organisée de la façon suivante :
Les *Chapitres 1, 2* traitent des lois de Dirichlet, des lois de Poisson-Dirichlet et des processus de Dirichlet et leurs mélanges. La fin du *chapitre 2* est consacré à certain nouveau modéle introduit dans des articles trés resent

Au *Chapitre 3*, nous commencons également la partie originale du travail en considérant un processus de Dirichlet ayant pour paramètre une mesure proportionnelle à la mesure de Wiener W. Ce processus, nommé *processus Brownien-Dirichlet*, admet une représentation :

$$X_t(\omega) = \sum_{i=1}^{\infty} p_i(\omega)\delta_{B_t^i(\omega)}$$

où les B^i sont des mouvements Browniens i.i.d. de loi W et $p = (p_i)$ suit une loi de Poisson-Dirichlet de paramètre $c > 0$ indépendant de $(B_t^i)_{i\in\mathbb{N}^*}$. Il sera noté $\mathcal{D}(cW)$.

Nous montrons notamment que l'on a une formule de type Ito et la décomposition classique de Doob-Meyer :

$$< X_t(\omega) - X_0(\omega),\ f >= \sum_{i=1}^{\infty} p_i(\omega)(f(B_t^i) - f(B_0^i)) = M_t + V_t$$

où (M_t) est une martingale, (V_t) est un processus à variation bornée et f une fonction deux fois dérivable vèrifiant $\|f'\|_{[0,T]<+\infty}$.

On montre aussi l'existence d'un temps local et d'une intégrale stochastique par rapport à ce processus.

Dans la dernière partie de ce *Chapitre*, on effectue des calculs de lois *a posteriori* pour des *mélanges* de processus de Brownien-Dirichlet lorsque

• La mesure mélangeante est une loi de Bernoulli $H = p\delta_0 + (1-p)\delta_1$:
Si P est un mélange de processus de Brownian-Dirichlet

$$P \sim \int \mathcal{D}(cW_u)dH(u)$$

et si $f_1,\ f_2, \ldots,\ f_n$ est une échantillon de taille n de P alors la distribution *a posteriori*

$$P\ |_{f_1,\,f_2,\,\ldots,\,f_n} \sim pH_1\ \mathcal{D}\left(cW_1 + \sum_{i=1}^{n}\delta_{f_i}\right) + (1-p)\ F_1\ \mathcal{D}\left(cW_0 + \sum_{i=1}^{n}\delta_{f_i}\right)$$

où F_1 et H_1 sont deux constantes qui dépendent de W', la derivée de Radon-Nikodym de W par rapport à une mesure μ définie dans le lemme d'Antoniak (Section 3.2.6), et òu W_0 et W_1 sont deux mesures de Wiener de moyenne repectivement 0 et 1.
• La mesure mélangeante est une gaussienne $H = \mathcal{N}(m,\ \sigma^2)$:
Si P est un mélange de processus de Brownien-Dirichlet

$$P \sim \int \mathcal{D}(cW_u)dH(u)$$

avec $(W_u)_{u\in\mathbb{R}}$ une famille de mesure de Wiener de moyen u. et si $\theta_1^t,\ \theta_2^t$ est une échantillon de taille 2 de $P_t,\ t \in \mathbb{R}_+$, alors la distribution conditionnelle de P_t sachant $\theta_1^t,\ \theta_2^t$ est un mélange de processus de Dirichlet tel que

$$P_t\ |\ \theta_1^t,\ \theta_2^t \sim \int \mathcal{D}(c\mathcal{N}_u + \sum_{i=1}^{2}\delta_{\theta_i^t})d\hat{H}_t(u)$$

où $\hat{H}_t(u) = H(u\ |_{\theta_1^t,\,\theta_2^t}) \sim \mathcal{N}(\mu_1^t,\ \sigma_{1,\,t}^2)$.

Le *Chapitre 5* est divisé en trois parties.
• Le mouvement Brownien en milieu aléatoire de Dirichlet
Nous l'introduisons comme limite en loi d'une marche aléatoire

$$\frac{1}{n^{1/2}}(U_1 + U_2 + \ldots + U_{[nt]})$$

construite de manière hiérarchique à partir du processus de Dirichlet :

$$\begin{cases} U_i \mid \mathcal{V} = \sigma^2 \overset{iid}{\sim} \mathcal{N}(0,\, \sigma^2) \\ \mathcal{V}^{-1} \mid P \sim P \\ P \mid c \sim \mathcal{D}(c\Gamma(\nu_1,\, \nu_2)) \\ c \sim \Gamma(\eta_1,\, \eta_2). \end{cases}$$

Nous simulons et estimons les paramètres d'un tel processus.

Comme à l'habitude, le système précèdent se lit de bas en haut :

c suit une loi $\Gamma(\eta_1,\, \eta_2)$, conditionellement à c, P suit une loi $\mathcal{D}(c\Gamma(\nu_1,\, \nu_2))$, contionellement à P suit une loi P et conditionellement à $\mathcal{V}$ les U_i sont des gaussiennes i.i.d.

• EDS en milieu aléatoire de Dirichlet.

Nous considérons, pour fixer les idées, l'EDS de type Black-Scholes, avec variance et drift aléatoirement fixés pendant chaque *régime*, toujours suivant un modèle hiérarchique de Dirichlet

$$dX_t = \sum_{j=1}^{L} \mu_{R_j} 1_{[T_{j-1},\, T_j)}(t)dt + \sum_{j=1}^{L} \sigma_{R_j} 1_{[T_{j-1},\, T_j)}(t)dB_t$$

où les R_j sont des entiers choisis aléatoirement dans $\{1,\, \ldots,\, N\}$ et constant sur les intervalles aléatoires de temps $[T_{j-1},\, T_j)$, avec

$$0 = T_0 < T_1 < T_2 < \ldots < T_L = T.$$

Pour estimer les paramètres de ce modèle où le temps est discrétisé, nous utilisons une version de l'échantillonneur de Gibbs utilisant un shéma *stick-breaking* fini (*blocked Gibbs sampling*) Ishwaran - Zarepour (2000) et Ishwaran - James (2002) [44]) shéma présenté au chapitre 2.

• Classification bayésienne de trajectoires d'actions selon leur volatilité.

La volatilité est supposée dépendre du temps :

$$dX_t = b(t,\, X_t)dt + \theta(t)h(X_t)dB_t$$

où $X_t = \log(S_t)$, (S_t) étant le processus du prix de l'action.
Sous certaines conditions l'EDS peut se simplifier en :

$$dX_t = b_t(t,\, X_t)dt + \theta(t)dB_t.$$

On développe alors la volatilité $\theta(t)$ dans une base d'ondelettes (V_i) et on classifie les trajectoires en classifiant les vecteurs des premiers coefficients par estimation d'un modèle hiérarchique de Dirichlet de mélange de lois normales. Ce travail a nécessité l'extension au cas vectoriel des calculs de lois *a posteriori* d'Ishwaran-Zarepour (2000) et Ishwaran-James (2002) [44].

Le *Chapitre 5* contient une partie essentielle de notre travail.
On se place dans le cas de l'observation (à des instants discrétisés) d'une trajectoire d'une EDS, par exemple de type Black-Scholes, en milieu aléatoire : drift et volatilité évoluent selon les états (qui modélisent les *régimes*) d'une chaîne de Markov à temps continu, de loi H à grande variance. Dans la littérature ce principe apparaît en mathématique financière dans les travaux sur les *Regime switching markets*.
Dans notre approche les régimes jouent le rôle que jouent les classes en classification : toute observation temporelle appartient à un régime.
La nouveauté ici est que nous plaçons un processus de Dirichlet de paramètre αH comme loi a priori sur l'espace des trajectoires de cette chaîne. Le nombre α exprime un degré de confiance en la loi H.
Des lois a priori sont mis sur les divers paramètres. L'algorithme consiste à dabord simuler un grand nombre de trajectoires qui sont très différentes à cause de la variance, ce qui permet d'envisager plusieurs scénarios.
On choisit ensuite à chaque itération une des trajectoires selon des poids donnés distribués *a priori* par un schéma stick-breaking. On calcule des lois *a posteriori*, puis selon la vraissemblance de la trajectoire observée, on met à jour poids et paramètres et on utilise un échantillonneur de Gibbs.
L'algorithme a été implémenté en langage C et testé sur des données simulées

puis sur des données réelles.

Le dernier *Chapitre 6* concerne la Conclusion et les Perspectives, notamment le calcul d'option en utilisant le modèle introduit au Chapitre 5.

Introduction

The aim of this work is to propose a new hierarchical model with a Dirichlet process as a prior distribution, shortly a Dirichlet hierarchical model, which is adapted to the analysis of temporal trajectories analysis, particularly those which are governed by an SDE (stochastic differential equation) in random environment.

The Dirichlet process is a random distribution (RD), i.e. a random variable taking its values in the set $\mathcal{P}(V)$ of all probability measures defined on a set V of observations.

The RDs are very interesting both for their theoretical aspects and their applied ones.

In our work, we will use four points, considered as very important in the history of this process.

- In 1969, in a fundamental and celebrated paper, Thomas S. Ferguson built the Dirichlet process as a generalization of a Dirichlet distribution. From this time the Dirichlet process is a remarkable and classical tool in nonparametric Bayesian statistics.
- In 1973, J.F.C. Kingman introduced a new RD, called Poisson-Dirichlet distribution, having some interesting properties and related to the representation of a Dirichlet process through the Gamma process.
- In 1974, motivated by applications, C.A. Antoniak introduced and studied mixtures of Dirichlet processes.
- In 1994, J. Sethuraman introduced a constructive method of a Dirichlet process, which is crucial for implementations.

The applications of Dirichlet processes deal with quite all fields: biology, ecology, computer science and so on.

This field was extended by using successfully Dirichlet hierarchical models in classification, more precisely in estimating mixtures of distributions from non temporal data, see e.g. Ishwaran and Zarepour (2000), Ishwaran and

James (2002), Kacperczyk et al., (2003), Bigelow and Dunson, (2007). Recently, Rodregez et al. introduce finite mixture versions of the nPD which is inspired from the work of Ishwaran and James (2002).

In all these papers, the Dirichlet process parameter is a measure proportional to a standard probability distribution in $\mathbb{R}^n$.

The present work consists in studying the extension of these hierarchical models to the case of temporal data, more precisely in introducing the Dirichlet process on a path space, the parameter being a measure proportional to the distribution of a continuous time process (Emilion 2005).

By observing just one path, we are able to detect some *regimes* of random durations, when the stochastic process is generated by an SDE in random environment. The random environment is represented by a continuous time Markov chain whose states modellize the regimes (for example the states of the financial market). These ones play the same role as the clusters in classification.

The Bayesian hierarchical model that we introduce, places a Dirichlet process as a prior on the path space of this chain. We show that the parameters can be estimated by using Gibbs sampling.

As an illustration of our work, we will consider a Black-Scholes SDE in finance, in random environment, the drift and the volatility being stochastic. This hierarchical model does not assume that the process is Gaussian since its finite marginal distributions are complicated mixtures Gaussian.

The thesis is organized as follows:

Chapters 1, 2 deal with Dirichlet distributions, Poisson-Dirichlet distribution, Dirichlet processes and their mixtures. The end of *Chapter 2* is devoted to new models introduced in some very recent papers.

After that, from Chapter 3 we start the original part of this work, firstly considering a Dirichlet process with parameter proportional to a Wiener measure W, shortly a *Brownian-Dirichlet* process, which has the following representation:

$$X_t(\omega) = \sum_{i=1}^{\infty} p_i(\omega)\delta_{B_t^i(\omega)}$$

where the B^i's are i.i.d. Brownian motions having for distribution W, and $p = (p_i)$ is Poisson-Dirichlet with parameter $c > 0$ and is independent of $(B_i)_{i=1,\,2,\ldots}$. This processes will be denoted $\mathcal{D}(cW)$

We show an Ito type formula and a classical Doob-Meyer decomposition

$$< X_t - X_0,\, f >= M_t + V_t$$

where M_t is a martingale and V_t is a process with bounded variation.

We also observe the existence of a local time and a stochastic integral with respect to a Brownian-Dirichlet process.

In the last part of *Chapter 3* we calculate the *posterior* distribution for *mixtures* of Brownian-Dirichlet when

• The mixing measure is a Bernoulli distribution $H = p\delta_0 + (1-p)\delta_1$:

If P is a mixture of Brownian-Dirichlet processes

$$P \sim \int \mathcal{D}(cW_u)dH(u)$$

and if $f_1,\, f_2, \ldots,\, f_n$ is a sample of size n of P, then the posterior distribution satisfies the following formula

$$P\,|_{f_1,\, f_2,\ldots,\, f_n} \sim pH_1\mathcal{D}\left(cW_1 + \sum_{i=1}^{n}\delta_{f_i}\right) + (1-p)\,F_1\mathcal{D}\left(cW_0 + \sum_{i=1}^{n}\delta_{f_i}\right)$$

where F_1 and H_1 are two constants depending of W', the Radon-Nikodym derivative of W w.r.t. a probability measure μ which will be defined later in Antoniak lemma (section 3.2.6) and where W_0 and W_1 are two Wiener measures with mean 0 and 1, respectively.

• The mixing measure is Gaussian distribution $H = \mathcal{N}(m,\, \sigma^2)$:

If P is a continuous time Dirichlet process

$$P \sim \int \mathcal{D}(cW_u)dH(u)$$

and if θ_1^t, θ_2^t is a sample of size 2 of P_t, $t \in \mathbb{R}_+$, then the conditional distribution of P_t given θ_1^t, θ_2^t is a mixture of Dirichlet processes such that

$$P_t \mid \theta_1^t,\, \theta_2^t \sim \int \mathcal{D}(c\mathcal{N}_u + \sum_{i=1}^{2} \delta_{\theta_i^t}) d\hat{H}_t(u)$$

where $\hat{H}_t(u) = H(u \mid_{\theta_1^t, \theta_2^t}) \sim \mathcal{N}(\mu_1^t,\, \sigma_{1,\,t}^2)$.

The *Chapter 4* is divided in three parts.

• The Brownian motion in Dirichlet random environment. We introduced as the limit in distribution of a random walk

$$\frac{1}{n^{1/2}}(U_1 + U_2 + \ldots + U_{[nt]})$$

based on the following a hierarchical Dirichlet model:

$$\begin{cases} U_i \mid \mathcal{V} = \sigma \overset{iid}{\sim} \mathcal{N}(0,\, \sigma^2) \\ \mathcal{V}^{-1} \mid P \sim P \\ P \mid c \sim \mathcal{D}(c\Gamma(\nu_1,\, \nu_2)) \\ c \sim \Gamma(\eta_1,\, \eta_2). \end{cases}$$

We proceed to the simulation and the estimation of the parameters of such a motion.

As usual, the above system has to be read from bottom to top: c has a $\Gamma(\eta_1,\, \eta_2)$ distribution, given c, P has $\mathcal{D}(c\Gamma(\nu_1,\, \nu_2))$ distribution, given P, $\mathcal{V}^{-1}$ has for distribution P and given $\mathcal{V}$ the U_i's are i.i.d. Gaussians.

• SDE in Dirichlet random environment.

As an illustration, we consider Black-Scholes SDE type, with variance and drift randomly fixed during each *regime* and derived from a Dirichlet hierarchical model

$$dX_t = \sum_{j=1}^{L} \mu_{R_j} 1_{[T_{j-1},\, T_j)}(t)dt + \sum_{j=1}^{L} \sigma_{R_j} 1_{[T_{j-1},\, T_j)}(t)dB_t$$

where the R_j are integers randomly chosen in $\{1, \ldots, N\}$ and constant on the random time intervals $[T_{j-1},\, T_j)$, where

$$0 = T_0 < T_1 < T_2 < \ldots < T_L = T.$$

To estimate the parameters of this model, where time is discretized, we use a blocked Gibbs sampling method (Ishwaran - Zarepour (2000)et Ishwaran - James (2002) [24]) which hinges on stick-breaking scheme.

• Bayesian classification of shares according to their volatility

The volatility is assumed to be depending on time and varies according to the share:

$$dX_t = b(t,\, X_t)dt + \theta(t)h(X_t)dB_t$$

where $X_t = \log(S_t)$ and (S_t) is the process describing the share.

Under some conditions this SDE reduces to:

$$dX_t = b_t(t,\, X_t)dt + \theta(t)dB_t.$$

Expanding the volatility $\theta(t)$ in a (wavelet) basis (V_i) we classify the paths by classifying the vectors of the first coefficients, estimating a hierarchical Dirichlet model of Normal distributions mixture: to this end, it is necessary to extend the calculus of posterior distributions (Ishwaran - Zarepour (2000), Ishwaran - James (2002)) to the vector case.

Chapter 5 contains an essential part of our work.

We observe an SDE path at discrete times, for example the Black-Scholes SDE in random environment: drift and volatility evolve according to the state *regime* of the market which is modellized by a continuous time Markov chain, having a distribution H with large variance. This appears in mathematical finance literature as *regime switching markets*.

In our approach, regimes play the role that play clusters in classification: each temporal observation belong to a regime.

The novelty here is that we place a prior, a Dirichlet process with parameter αH, on the path space of the Markov chain. The number α is a confidence degree on H, the distribution of the Markov chain.

We also place a prior distribution on each parameter.
The algorithm consists in first simulating a large number of paths which are very different, due to the variance. This gives us a large variety of scenarios. Next, in each iteration we choose a path according to random weights, initially given by a stick-breaking scheme. A calculation of posterior distributions is performed. Then according to the likelyhood w.r.t. the observed path, we perform a Gibbs sampling procedure, by first updating the weights and the parameters.
The program is implemented in C language and tested on a set of simulated data and real data.

The last *Chapter 6* concerns Conclusion and Perspectives, in particular, the calculation of option prices when using the model introduced in chapter 5.

Chapter 2

Dirichlet distribution

The Dirichlet distribution is intensively used in various fields: biology EMILION, R. (2005), astronomy ISHWARAN, H. and JAMES, L.F. (2002), text mining DAHL, D. B. (2003), ...

It can be seen as a random distribution on a finite set. Dirichlet distribution is a very popular prior in Bayesian statistics because the posterior distribution is also a Dirichlet distribution. In this chapter we give a complete presentation of this interesting law: representation by Gamma's distribution, limit distribution in a contamination model. (The Polya urn scheme), ...

2.1 Random probability vectors

Consider a partition of a nonvoid finite set E with cardinality $\sharp E = n \in \mathbb{N}^*$ into d nonvoid disjoint subsets. To such a partition corresponds a partition of the integer n, say $c_1, \ldots, c_d$, that is a finite family of positive integers, such that $c_1 + \ldots + c_d = n$. Thus, if $p_j = \frac{c_j}{n}$, we have $p_1 + \ldots + p_d = 1$.

In biology for example, p_j can represent the percentage of the j^{th} specy in a population.

So we are lead to introduce the following d-dimentional simplex:

$$\triangle_{d-1} = \{(p_1, \ldots, p_d) : p_j \geq 0, \sum_{j=1}^{d} p_j = 1\}.$$

When n tends to infinity, this yields to the following notion:

Definition 2.1.1 *One calls mass-partition any infinite numerical sequence*

$$p = (p_1, p_2, \ldots)$$

such that $p_1 \geq p_2 \geq \ldots$ *and* $\sum_1^\infty p_j = 1$.
The space of mass-partitions is denoted by

$$\nabla_\infty = \{(p_1, p_2, \ldots) : p_1 \geq p_2 \geq \ldots; p_j \geq 0, j \geq 1, \sum_{j=1}^{\infty} p_j = 1\}.$$

Lemma 2.1.1 *(Bertoin [28] page 63) Let* $x_1, \ldots, x_{d-1}$ *be* $d-1$ *i.i.d. random variables uniformly distributed on* $[0, 1]$ *and let* $x_{(1)} < \ldots < x_{(d-1)}$ *denote its order statistic, then the random vector*

$$(x_{(1)}, \ldots, x_{(d-1)} - x_{(d-2)}, 1 - x_{(d-1)})$$

is uniformly distributed on $\triangle_{d-1}$.

2.2 Polya urn (Blackwell and MacQueen) [3]

We consider an urn that contains d colored balls numbered from 1 to d. Initially, there is only one ball of each color in the urn. We draw a ball, we observe its color and we put it back in the urn with another ball having the same color. Thus at the instant n we have $n+d$ balls in the urn and we have added $n = N_1 + \ldots + N_d$ balls with N_j balls of color j.
We are going to show that the distribution of $(\frac{N_1}{n}, \frac{N_2}{n}, \ldots, \frac{N_d}{n})$ converges to a limit distribution.

2.2.1 Markov chain

Proposition 2.2.1

$$\lim_{n\longrightarrow\infty}(\frac{N_1}{n},\ldots,\,\frac{N_d}{n})\stackrel{d}{=}(Z_1,\,Z_2,\ldots,\,Z_d)$$

where $(Z_1,\,Z_2,\,\ldots,\,Z_d)$ *have a uniform distribution on the simplex* $\triangle_{d-1}$.

Proof

Denote the projection operation

$$\pi_i : \mathbb{R}^d \to \mathbb{R}$$

$$x = (x_1,\,\ldots,\,x_d) \mapsto x_i$$

and

$$\theta_i(x) = (x_1\,,\ldots,\,x_{i-1},\,x_i+1,\,x_{i+1},\,\ldots,\,x_d).$$

Let

$$S(x) = \sum_{i=1}^{d} x_i$$

and

$$f_i(x) = \frac{\pi_i(x)+1}{S(x)+d}.$$

Define a transition kernel as follows

$$P(x,\,\theta_i(x)) = \frac{\pi_i(x)+1}{S(x)+d}$$

$$P(x,\,y) = 0, \quad \text{if} \quad y \notin \{\theta_1(x),\ldots,\,\theta_d(x)\}.$$

Recall that for any non-negative (resp. bounded) measurable function g defined on $\mathbb{R}^d$, the function Pg is defined as

$$Pg(x) = \int_{\mathbb{R}^d} g(y)P(x,\,dy).$$

Here we see that

$$Pg(x) = \sum_{i=1}^{d} g(\theta_j(x)) \frac{\pi_i(x)+1}{S(x)+d}.$$

First step :

Consider $Y_n = (Y_n^1, \ldots, Y_n^d)$ where $(Y_n^i)_{0\le i\le d}$ is the number of balls of color i added to the urn at n^{th} step. We clearly see that (Y_{n+1}) only depends on the n^{th} step so that $(Y_n)_n$ is Markov chain with transition kernel

$$P(Y_n, \theta_i(Y_n)) = \frac{\pi_i(Y_n)+1}{S(Y_n)+d}$$

and

$$Y_0 = (0, \ldots, 0).$$

On the other hand,

$$Pf_i(Y_n) \;=\; \textstyle\sum_{j=1}^{d} \frac{\pi_i(Y_n)+1}{S(Y_n)+d} \frac{\pi_i(\theta_j(Y_n))+1}{S(\theta_j(Y_n))+d}$$

since

$$\begin{cases} \pi_i(\theta_j(Y_n)) = \pi_i(Y_n) \quad \text{if} \quad i \neq j, \\ \pi_i(\theta_i(Y_n)) = \pi_i(Y_n)+1 \quad \text{if} \quad i = j, \\ S(\theta_i(Y_n)) = S(Y_n)+1. \end{cases} \tag{2.1}$$

Then

$$\begin{aligned} Pf_i(Y_n) &= \textstyle\sum_{i\neq j} \frac{\pi_i(Y_n)+1}{S(Y_n)+d} \frac{\pi_j(Y_n)+1}{S(Y_n)+d+1} + \frac{\pi_i(Y_n)+1}{S(Y_n)+d} \frac{\pi_i(Y_n)+2}{S(Y_n)+d+1} \\ &= \textstyle\frac{\pi_i(Y_n)+1}{(S(Y_n)+d)(S(Y_n)+d+1)} [\pi_i(Y_n) + 2 + \sum_{i\neq j} \pi_j(Y_n) + 1] \\ &= \textstyle\frac{\pi_i(Y_n)+1}{(S(Y_n)+d)(S(Y_n)+d+1)} [\pi_i(Y_n) + 2 + (S(Y_n) + d - 1 - \pi_i(Y_n))] \\ &= f_i(Y_n). \end{aligned}$$

implies that $f_i(Y_n)$ is a positive martingale which converges almost sure towards a random variable Z_i. Since $f_i(Y_n)$ is bounded by 1, it is also convergent in the L^p spaces, according to the bounded convergence theorem. We then see that :

$$\frac{\pi_i(Y_n)}{n} = \frac{n+d}{n} f_i(Y_n) - \frac{1}{n}$$

converges to the same limit Z_i almost surely and in L^p.
By the martingale properties we have moreover that

$$\mathbb{E}(f_i(Y_n)) = \mathbb{E}(f_i(Y_0)).$$

Consequently

$$\begin{aligned} \mathbb{E}(\lim_{n\to\infty} f(Y_n)) &= \lim_{n\to\infty} \mathbb{E}(f(Y_n)) \\ &= \mathbb{E}(f(Y_0)), \end{aligned}$$

so

$$\mathbb{E}(Z_i) = \mathbb{E}(f_i(Y_0)) = \frac{1}{d}.$$

Second step:
Let

$$\wedge_{d-1} = \{(p_1, \ldots, p_{d-1}) : \, p_i \geq 0 \sum_{i=1}^{d-1} p_i \leq 1\},$$

and

$$h_u(Y_n) = \frac{(S(Y_n)+d-1)!}{\prod_{i=1}^{d} \pi_i(Y_n)!} u_1^{\pi_1(Y_n)} \ldots u_d^{\pi_d(Y_n)}$$

The uniform measure λ_d on $\triangle_{d-1}$ is defined as follows: for any borelian bounded function $F(u_1, \ldots, u_d)$ we have:

$$\int_{\triangle_{d-1}} F(u)\lambda_d(du) = \int_{\wedge_{d-1}} F(u_1, \ldots, u_{d-1}, 1-u_1-u_2-\ldots-u_{d-1})du_1.\ldots.du_{d-1}$$

Now, let us compare the moments of $(Z_1, Z_2, \ldots, Z_d)$ with the ones of λ_d.
Using formula (1.1)

$$h_u(\theta_i(Y_n)) \quad = \quad \tfrac{S(Y_n)+d}{\pi_i(Y_n)+1} u_i h_u(Y_n).$$

hence

$$Ph_u(Y_n) = h_u(Y_n)(\sum_i^d u_i) = h_u(Y_n).$$

implies that $(h_u(Y_n))$ is a martingale and similarly

$$g_k(Y_n) = \int_{\triangle_{d-1}} h_u(Y_n) u_1^{k_1} \dots u_d^{k_d} \lambda_d(du)$$

is a martingale because

$$\begin{aligned} Pg_k(Y_n) &= \sum_i^d P(Y_n, \theta_i(Y_n)) \int_{\triangle_{d-1}} h_u(Y_n) u_1^{k_1} \dots u_d^{k_d} \lambda_d(du) \\ &= \int_{\triangle_{d-1}} Ph_u(Y_n) u_1^{k_1} \dots u_d^{k_d} \lambda_d(du) \\ &= g_k(Y_n). \end{aligned}$$

This gives

$$\mathbb{E}(g_k(Y_n)) = \mathbb{E}(g_k(Y_0)).$$

On the other hand

$$\begin{aligned} g_k(Y_n) &= \frac{\prod_{i=1}^{i=d} [\pi_i(Y_n) + 1] \dots [\pi_i(Y_n) + k_i]}{(n+d) \dots (n + s(k) + d - 1)} \\ &= \frac{\prod_{i=1}^{i=d} \frac{[\pi_i(Y_n)+1]}{n} \dots \frac{[\pi_i(Y_n)+k_i]}{n}}{\frac{(n+d)}{n} \dots \frac{(n+s(k)+d-1)}{n}} \end{aligned}$$

so that

$$0 \leq g_k(Y_n) \leq \prod_{i=1}^{d} 2^{k_i} = 2^{S(k)}.$$

Therefore by the bounded convergence theorem

$$\begin{aligned} \lim_{n\to\infty} \mathbb{E}(g_k(Y_n)) &= \mathbb{E}(\lim_{n\to\infty} g_k(Y_n)) \\ &= \mathbb{E}(Z_1^{k_1} \dots Z_d^{k_d}) \\ &= \frac{(d-1)! \prod_{i=1}^d k_i!}{(S(k)+d-1)!} \\ &= c_d \int_{\triangle_{d-1}} u_1^{k_1} \dots u_d^{k_d} \lambda_d(du), \end{aligned}$$

where $c_d = (d-1)!$

Indeed if

$$m_k = \int_{\triangle_{d-1}} u_1^{k_1} \dots u_d^{k_d} c_d \lambda_d(du)$$

integrations and recurrences yield,

$$m_k = \frac{\prod_{i=1}^d k_i!}{(S(k)+d-1)!}.$$

Taken $(k_1, \dots, k_d) = (0, \dots, 0)$, we see that $c_d = (d-1)!$.

Further, if μ is the distribution of $(Z_1, \dots, Z_d)$, then $c_d\lambda_d$ and μ have the same moments and since $\triangle_{d-1}$ is compact, the theorem of monotone class yields, $\mu = c_d\lambda_d$.

2.2.2 Gamma, Beta and Dirichlet densities

Let $\alpha > 0$, the gamma distribution with parameter α, denoted $\Gamma(\alpha, 1)$, is defined by the probability density function:

$$f(y) = y^{\alpha-1}\frac{e^{-y}}{\Gamma(\alpha)}1\!\!1_{\{y>0\}}.$$

Let $Z_1, \dots, Z_d$ be d independent real random variables with gamma distributions $\Gamma(\alpha_1, 1), \dots, \Gamma(\alpha_d, 1)$, respectively, then it is well-known that $Z = Z_1 + \dots + Z_d$ has distribution $\Gamma(\alpha_1 + \dots + \alpha_d, 1)$.

Let $a, b > 0$, a beta distribution with parameter (a, b), denoted $\beta(a, b)$, is defined by the probability density function:

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}1\!\!1_{\{0<x<1\}}.$$

From these densities it is easily seen that the following function is a density function:

Definition 2.2.1 *For any* $\underline{\alpha} = (\alpha_1, \ldots, \alpha_d)$ *where* $\alpha_i > 0$ *for any* $i = 1, \ldots, d$*, the density function* $d(y_1, y_2, \ldots, y_{d-1} \mid \underline{\alpha})$ *defined as*

$$\frac{\Gamma(\alpha_1 + \ldots + \alpha_k)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_k)} y_1^{\alpha_1 - 1} \ldots y_{d-1}^{\alpha_{d-1}-1} (1 - \sum_{h=1}^{d-1} y_h)^{\alpha_d - 1} \mathbb{1}_{\wedge_{d-1}}(y) \qquad (2.2)$$

is called the **Dirichlet density** *with parameter* $(\alpha_1, \ldots, \alpha_d)$.

Proposition 2.2.2 *Let* $(Z_1, Z_2, \ldots, Z_d)$ *be uniformly distributed on* $\triangle_{d-1}$*. Then the random vector* $(Z_1, Z_2, \ldots, Z_{d-1})$ *has the the Dirichlet density (1.2) with parameters (1, 1,...,1).*

Proof

Let $\lambda_i \in \mathbb{N}$ for any $i \in \{1, \ldots, d\}$.

Let $(Y_1, Y_2, \ldots, Y_{d-1})$ be a random vector with Dirichlet density defined in (1.2).

Let $Y_d = 1 - \sum_{i=1}^{d-1} y_i$. Then

$$\begin{aligned}
\mathbb{E}(Y_1^{\lambda_1} \ldots Y_d^{\lambda_d}) &= \mathbb{E}(Y_1^{\lambda_1} \ldots Y_d^{\lambda_{d-1}} [1 - \textstyle\sum_{i=1}^{d-1} Y_i]^{\lambda_d}) \\
&= \tfrac{\Gamma(\alpha_1 + \ldots + \alpha_k)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_d)} \int_{\wedge_{d-1}} y_1^{\alpha_1 + \lambda_1 - 1} \ldots y_{d-1}^{\alpha_{d-1} + \lambda_{d-1} - 1} \\
&\quad [1 - \textstyle\sum_{i=1}^{d-1} y_i]^{\alpha_d + \lambda_d - 1} dy_1 \ldots dy_{d-1} \\
&= \tfrac{\Gamma(\alpha_1 + \ldots + \alpha_d)\Gamma(\alpha_1 + \lambda_1) \ldots \Gamma(\alpha_d + \lambda_d)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_d) \Gamma((\alpha_1 + \ldots + \alpha_d) + \sum_{i=1}^{d} \lambda_i)}.
\end{aligned}$$

Consequently, if λ_i, $i \in \{1, \ldots, d\}$ are non-negative integers and $\alpha_1 = \ldots = \alpha_d = 1$, then

$$\mathbb{E}(Y_1^{\lambda_1} \ldots Y_d^{\lambda_d}) = \frac{(d-1)! \prod_{i=1}^{d} \lambda_i!}{((d-1) + S(\lambda))!}.$$

Now the proof of the preceding proposition 1.2.1 shows that $(Z_1, Z_2, \ldots, Z_d)$ and $(Y_1, \ldots, Y_d)$ have the same moments, and thus the same distribution. Consequently $(Z_1, Z_2, \ldots, Z_{d-1})$ has the same distribution as $(Y_1, \ldots, Y_{d-1})$ which is by construction $d(y_1, y_2, \ldots, y_{d-1} \mid \underline{\alpha})$.

2.3 Dirichlet distribution

The Dirichlet density is not easy to be handled and the following theorem gives an interesting construction where appears this density.

Theorem 2.3.1 *Let $Z_1, \ldots, Z_d$ be d independent real random variables with gamma distributions $\Gamma(\alpha_1, 1), \ldots, \Gamma(\alpha_d, 1)$ respectively and let $Z = Z_1 + \ldots + Z_d$. Then the random vector $(\frac{Z_1}{Z}, \ldots, \frac{Z_{d-1}}{Z})$ has a Dirichlet density with parameters $(\alpha_1, \ldots, \alpha_d)$.*

Proof

The mapping

$$(y_1, \ldots, y_d) \mapsto (\frac{y_1}{y_1 + \ldots + y_d}, \ldots, \frac{y_{d-1}}{y_1 + \ldots + y_d}, y_1 + \ldots + y_d)$$

is a diffeomorphism from $[0, \infty)^d$, to $\wedge_{d-1} \times]0, \infty)$ with Jacobian y_d^{d-1} and reciprocal function:

$$(y_1, \ldots, y_d) \mapsto (y_1 y_d, \ldots, y_{d-1} y_d, y_d[1 - \sum_{i=1}^{d-1} y_i]).$$

The density of $(Z_1, \ldots, Z_{d-1}, Z)$ at point $(y_1, \ldots, y_d)$ is therefore equal to:

$$e^{-y_d} y_1^{\alpha_1 - 1} \ldots y_{d-1}^{\alpha_{d-1} - 1} (1 - \sum_{i=1}^{d-1} y_i)^{\alpha_d - 1} \frac{y_d^{\alpha_1 + \ldots + \alpha_d - d}}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_d)} y_d^{d-1}.$$

Integrating w.r.t. y_d and using the equality $\int_0^\infty e^{-y_d} y_d^{\alpha - 1} dy_d = \Gamma(\alpha_1 + \ldots + \alpha_d)$, we see that the density of $(\frac{Z_1}{Z}, \ldots, \frac{Z_{d-1}}{Z})$ is a Dirichlet density with parameters $(\alpha_1, \ldots, \alpha_d)$. $\square$

Definition 2.3.1 *Let $Z_1, \ldots, Z_d$ be d independent real random variables with gamma distributions $\Gamma(\alpha_1, 1), \ldots, \Gamma(\alpha_d, 1)$, respectively, and let $Z = Z_1 + \ldots + Z_d$. The Dirichlet distribution with parameters $(\alpha_1, \ldots, \alpha_d)$ is the distribution of the random vector $(\frac{Z_1}{Z}, \ldots, \frac{Z_d}{Z})$.*

Not that the Dirichlet distribution is singular w.r.t Lebesgue measure in $\mathbb{R}^d$ since it is supported by Δ_{d-1} which has Lebesgue measure 0.
The following proposition can be easily proved

Proposition 2.3.1 *With the same notation as in Theorem 1.3.1 let* $Y_i = \frac{Z_i}{Z}$, $i = 1, \ldots, d$ *then* Y_i *has a beta distribution* $\beta(\alpha_i, \, \alpha_1 + \ldots + \alpha_{i-1} + \alpha_{i+1} + \ldots + \alpha_d)$ *and*

$$\mathbb{E}(y_i) = \frac{\alpha_i}{\alpha_1 + \ldots + \alpha_d}, \, \mathbb{E}(y_i y_j) = \frac{\alpha_i \alpha j}{(\alpha_1 + \ldots + \alpha_k)(\alpha_1 + \ldots + \alpha_d + 1)}.$$

Lemma 2.3.1 *Let* $\gamma = (\gamma_1, \, \gamma_2, \ldots, \gamma_k)$ *and* $\rho = (\rho_1, \, \rho_2, \ldots, \rho_k)$ *be* k*-dimensional vectors. Let* U, V *be independent* k*-dimensional random vectors with Dirichlet distributions* $\mathcal{D}(\gamma)$ *and* $\mathcal{D}(\rho)$*, respectively. Let* W *be independent of* $(U, \, V)$ *and have a Beta distribution* $\beta(\sum_{i=1}^k \gamma_i, \, \sum_{i=1}^k \rho_i)$*. Then the distribution of*

$$WU + (1 - W)V$$

is the Dirichlet distribution $\mathcal{D}(\gamma + \rho)$*.*

Lemma 2.3.2 *Let* e_j *denote the* k*-dimensional vector consisting of* 0*'s, except of the* j^{th} *co-ordinate, with equal to 1. Let* $\gamma = (\gamma_1, \, \gamma_2, \ldots, \gamma_k)$ *and let* $\beta_j = \frac{\gamma_j}{\sum_{i=1}^k \gamma_i}$, $j = 1, \, 2, \ldots, \, k$*.*
Then

$$\sum \beta_j \mathcal{D}(\gamma + e_j) = \mathcal{D}(\gamma).$$

This conclusion can also be written as $\mathbb{E}(\mathcal{D}(\rho + \gamma) = \mathcal{D}(\gamma)$*.*

The proofs of these two Lemma are found in Wilks ((1962), section 7),

2.4 Posterior distribution and Bayesian estimation

Consider the Dirichlet distribution $\mathcal{D}(\alpha_1, \ldots, \alpha_d)$ as a prior on $p = (p_1, \, p_2, \ldots, \, p_d) \in \Delta_{d-1}$.

Let X be a random variable assuming values in $\{1, \ldots, d\}$, such that $P(X = i \mid p) = p_i$. Then the posterior distribution $p \mid X = i$ is Dirichlet $\mathcal{D}(\alpha_1, \ldots, \alpha_{i-1}, \alpha_i + 1, \ldots, \alpha_d)$.

Indeed let $N_i = \sum_{j=1}^{n} 1\!\!1_{X_j=i}$, $1 \leq i \leq d$. The likelihood of the sample is

$$\prod_{i=1}^{d-1} p^{N_i}(1 - \sum_{i=1}^{d-1} p_i)^{N_d}.$$

If the prior distribution of p is $\mathcal{D}(\alpha_1, \ldots, \alpha_d)$, the posterior density will be proportional to

$$\prod_{i=1}^{d-1} p_i^{\alpha_i+N_i}(1 - \sum_{i=1}^{d-1} p_i)^{N_d+\alpha_d}.$$

Thus the posterior distribution of p is $\mathcal{D}(\alpha_1 + N_1, \alpha_2 + N_2, \ldots, \alpha_k + N_d)$.

If $(X_1, \ldots, X_n)$ is a sample of law $p = (p_1, \ldots, p_d)$ on $\{1, \ldots, d\}$ then the average Bayesian estimation of p is:

$$p' = (\frac{\alpha_1 + N_1}{\sum_{i=1}^{d} \alpha_i + 1}, \frac{\alpha_2 + N_2}{\sum_{i=1}^{d} \alpha_i + 1}, \ldots, \frac{\alpha_k + N_d}{\sum_{i=1}^{d} \alpha_i + 1}).$$

Proposition 2.4.1 ([19]) *Let $r_1, \ldots, r_l$ be l integers such that $0 < r_1 < \ldots < r_l = d$.*

1. *If $(Y_1, \ldots, Y_d) \sim \mathcal{D}(\alpha_1, \ldots, \alpha_d)$, then*

$$\Big(\sum_{1}^{r_1} Y_i, \sum_{r_1+1}^{r_2} Y_i, \ldots, \sum_{r_{l-1}}^{r_l} Y_i\Big) \sim \mathcal{D}\Big(\sum_{1}^{r_1} \alpha_i, \sum_{r_1+1}^{r_2} \alpha_i, \ldots, \sum_{r_{l-1}}^{r_l} \alpha_i\Big).$$

2. *If the prior distribution of $(Y_1, \ldots, Y_d)$ is $\mathcal{D}(\alpha_1, \ldots, \alpha_d)$ and if*

$$P(X = j \mid Y_1, \ldots, Y_d) = Y_j$$

a.s for $j = 1, \ldots, d$, then the posterior distribution of $(Y_1, \ldots, Y_d)$ given $X = j$ is $\mathcal{D}(\alpha_1^{(j)}, \ldots, \alpha_k^{(j)})$ where

$$\alpha_i^{(j)} = \begin{cases} \alpha_i & \text{if } i \neq j \\ \alpha_j + 1 & \text{if } i = j \end{cases}$$

3. *Let $D(y_1, \ldots, y_d \mid \alpha_1, \ldots, \alpha_d)$ denote the distribution function of the Dirichlet distribution $\mathcal{D}(\alpha_1, \ldots, \alpha_d)$, that is*

$$D(y_1, \ldots, y_d \mid \alpha_1, \ldots, \alpha_d) = P(Y_1 \leq y_1, \ldots, Y_d \leq y_d).$$

Then,

$$\int_0^{z_1} \cdots \int_0^{z_d} y_j dD(y_1, \ldots, y_d \mid \alpha_1, \ldots, \alpha_d) = \frac{\alpha_j}{\alpha} D(z_1, \ldots, z_d \mid \alpha_1^{(j)}, \ldots, \alpha_d^{(j)}).$$

Proof

1. Recall that: if $Z_1 \sim \Gamma(\alpha_1)$, $Z_2 \sim \Gamma(\alpha_2)$, and if Z_1 and Z_2 are independent then $Z_1+Z_2 \sim \Gamma(\alpha_1+\alpha_2)$. Hence 1 may be obtained by recurrence.

2. Is obtained then by induction.

3. Using 2

$$\begin{aligned} P(X = j, Y_1 \leq z_1, \ldots, Y_d \leq z_d) &= P(X = j) P(Y_1 \leq z_1, \ldots, Y_d \leq z_d \mid X = j) \\ &= \mathbb{E}(\mathbb{E}(\mathbb{1}_{X=j} \mid Y_1, \ldots, Y_d)) \\ &\times D(z_1, \ldots, z_d \mid \alpha_1^{(j)}, \ldots, \alpha_{(d)}^{j}) \\ &= \mathbb{E}(Y_j) D(z_1, \ldots, z_d \mid \alpha_1^{(j)}, \ldots, \alpha_d^{(j)}) \\ &= \tfrac{\alpha_j}{\alpha} D(z_1, \ldots, z_d \mid \alpha_1^{(j)}, \ldots, \alpha_d^{(j)}). \end{aligned}$$

On the other hand

$$\begin{aligned} P(X=j,\, Y_1 \leq z_1, \ldots, Y_d \leq z_d) &= \mathbb{E}(\mathbb{1}_{\{X=j,\, Y_1 \leq z_1, \ldots, Y_d \leq z_d\}}) \\ &= \mathbb{E}(\mathbb{E}(\mathbb{1}_{\{X=j,\, Y_1 \leq z_1, \ldots, Y_d \leq z_d\}} \mid Y_1, \ldots, Y_K) \\ &= E(\mathbb{1}_{\{Y_1 \leq z_1, \ldots, Y_d \leq z_d\}} E(\mathbb{1}_{\{X=j\}} \mid Y_1, \ldots, Y_d)) \\ &= \mathbb{E}(\mathbb{1}_{\{Y_1 \leq Z_1, \ldots, Y_d \leq z_d\}} Y_j)) \\ &= \int_0^{z_1} \ldots \int_0^{z_d} Y_j dD(Y_1, \ldots, Y_d \mid \alpha_{(1)}, \ldots, \alpha_{(d)}). \end{aligned}$$

2.5 Definition and proprieties on Poisson-Dirichlet distribution

The Poisson-Dirichlet distribution is a probability measure introduced by J.F.C Kingman [31] on the set

$$\bigtriangledown_\infty = \{(p_1, p_2, \ldots);\, p_1 \geq p_2 \geq \ldots,\, p_i \geq 0,\, \sum_{j=1}^{\infty} p_j = 1\}.$$

It can be considered as a limit of some specific Dirichlet distributions and is also, as shown below, the distribution of the sequence of the jumps of a Gamma process arranged by decreasing order and normalized .
We will also see how Poisson-Dirichlet distribution is related to Poisson processes.

2.5.1 Gamma process and Dirichlet distribution

Definition 2.5.1 *We say that $X = (X_t)_{t \in \mathbb{R}^+}$ is a Levy process if for every $s, t \geq 0$, the increment $X_{t+s} - X_t$ is independent of the process $(X_v,\, 0 \leq v \leq t)$ and has the same law as X_s, in particular, $\mathcal{P}(X_0 = 0) = 1$.*

Definition 2.5.2 *A subordinator is a Levy process taking values in* $[0, \infty)$, *which implies that its sample paths are increasing.*

Definition 2.5.3 *The law of a random variable* X *is infinitely divisible, if for all* $n \in \mathbb{N}$ *there exist i.i.d. random variables* $X_1^{(1/n)}, \ldots, X_n^{(1/n)}$ *such that*

$$X \stackrel{d}{=} X_1^1 + \ldots + X_n^1$$

Equivalently, the law of X *is infinitely divisible, if for all* $n \in \mathbb{N}$ *there exists a random variable* $X^{(1/n)}$, *such that the characteristic function of* X,

$$\varphi_X(u) = (\varphi_{X^{(1/n)}}(u))^n.$$

Definition 2.5.4 *The law of a random variable* X *is infinitely divisible if and only if there exists a triplet* (b, c, ν), *with* $b \in \mathbb{R}$, $c \in \mathbb{R}_+$ *and a measure satisfying* $\nu(\{0\}) = 0$ *and* $\int_{\mathbb{R}}(1 \Lambda |x|^2)\nu(dx) < \infty$, *such that*

$$\mathbb{E}[\exp(uX)] = \exp[ibu - \frac{u^2c}{2} + \int_{\mathbb{R}}(e^{iux} - 1 - iux\mathbb{1}_{\{|x|<1\}})\nu(dx)]. \qquad (2.3)$$

The triplet $(b,\ c,\ \nu)$ *is called the Lévy triplet and the exponent in (1.3)*

$$\psi(u) = ibu - \frac{u^2c}{2} + \int_{\mathbb{R}}(e^{iux} - 1 - iux\mathbb{1}_{\{|x|<1\}})\nu(dx)$$

is called the Lévy exponent. Moreover, $b \in \mathbb{R}$ *is called the drift term,* $c \in \mathbb{R}_+$ *the Gaussian coefficient and* ν *the Lévy measure.*

Definition 2.5.5 *A Gamma process is a subordinator such that its Lévy measure is* $\gamma(dx) = x^{-1}e^{-x}dx$.

Remark 2.5.1 *Let* ξ *be a gamma process. Let* $\alpha_1, \ldots, \alpha_n > 0$, $t_0 = 0$, $t_j = \alpha_1 + \ldots + \alpha_j$, *for* $1 \leq j \leq n$ *and* $Y_j = \xi(t_j) - \xi(t_{j-1})$ *then*

$$Y_j \sim \Gamma(\alpha_j).$$

Moreover, $Y_1, Y_1 \ldots, Y_n$ *are independent.*
Let $Y = Y_1 + \ldots + Y_n = \xi(t_n)$ *and* $p = (p_1, \ldots, p_n)$ *with* $p_j = \frac{Y_j}{Y}$ *then* p *is a random vector on* $\wedge_{n-1}$ *having* $\mathcal{D}(\alpha_1, \ldots, \alpha_n)$ *distribution. Therefore we get a random vector having Dirichlet distribution.*

2.5.2 The limiting order statistic

Let $\mathcal{D}(\alpha_1, \ldots, \alpha_n)$ be a Dirichlet distribution defined as in chapter 1 and let:

$$f_{\alpha_1,\ldots,\alpha_d}(p_1, p_2, \ldots, p_d) = \frac{\Gamma(\alpha_1 + \ldots + \alpha_d)}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_d)} p_1^{\alpha_1 - 1} \ldots p_d^{\alpha_d - 1} \mathbb{1}_{\triangle_{d-1}}. \tag{2.4}$$

Assume that the α_i are equal, then $f_{\alpha_1,\ldots,\alpha_d}(p_1, p_2, \ldots, p_d)$ reduces to

$$d(p_1, p_2, \ldots, p_d \mid \underline{\alpha}) = \frac{\Gamma(N\alpha)}{\Gamma(\alpha)^d}(p_1 \ldots p_d)^{\alpha-1}. \tag{2.5}$$

In this section we prove the following theorem which exhibits the limiting joint distribution of the order statistics $p_{(1)} \geq p_{(2)} \geq \ldots$ an element of the subset $\bigtriangledown_\infty$ of the set

$$\triangle_\infty = \{(p_1, p_2, \ldots); p_i \geq 0, \sum_{j=1}^{\infty} p_j = 1\}.$$

Consider the following mapping

$$\psi : \triangle_\infty \longrightarrow \bigtriangledown_\infty$$

$$(p_1, p_2, \ldots) \longmapsto (p_{(1)}, p_{(2)}, \ldots).$$

If P is any probability measure on $\bigtriangledown_\infty$, and n is any positive integer, then the random n-vector $(p_{(1)}, p_{(2)}, \ldots, p_{(n)})$ has a distribution depending on P,which might be called the n^{th} marginal distribution of P. The measure P is uniquely determined by its marginal distributions.

Theorem 2.5.1 (Kingman) *(1974) For each $\lambda \in]0, \infty[$, there exists a probability measure P_λ on $\bigtriangledown_\infty$ with the following property. If for each N the random vector p is distributed over $\triangle_N$ according to the distribution* (2.1) *with $\alpha = \alpha_N$, and if $N\alpha_N \to \lambda$ as $N \to \infty$, then for any n the distribution of the random vector $p = (p_{(1)}, p_{(2)}, \ldots, p_{(n)})$ converges to the n^{th} marginal distribution of P_λ as $N \to \infty$.*

Proof

Let y_1, $y_2, \ldots, y_N$ be independent random variables, each having a gamma distribution $\Gamma(\lambda,\, 1)$. We know that if $S = y_1 + y_2 + \ldots + y_N$, then $(y_1/S,\, y_2/S, \ldots,\, y_N/S)$ has a Dirichlet distribution $\mathcal{D}(\lambda, \ldots,\, \lambda)$.

To exploit this fact, consider as above a gamma process ξ, that is a stationary random process $(\xi(t),\, t \geq 0)$ with $\xi(0) = 0$. The process ξ increases only in jumps. The positions of these jump forms a random countable dense subset $J(\xi)$ of $(0,\, \infty)$, with

$$P\{t \in J(\xi)\} = 0 \tag{2.6}$$

for all $t > 0$. For each value of N, write

$$q_j(N) = \frac{\xi(j\alpha_N) - \xi\Big((j-1)\alpha_N\Big)}{\xi(N\alpha_N)} \tag{2.7}$$

by the result cited above, the vector $q = (q_{(1)},\, q_{(2)}, \ldots,\, q_{(N)})$ has the same distribution as p and it therefore suffices to prove the theorem with p replaced by q. We shall in fact prove that

$$\lim_{N\to\infty} q_{(j)}(N) = \delta\xi_{(j)}/\xi(\lambda) \tag{2.8}$$

where the $(\delta\xi_{(j)})_{j\in\mathbb{N}}$'s are the magnitudes of the jumps in $(0,\, \lambda)$ arranged in descending order. This will suffice to prove the theorem, with P_λ the distribution of the sequence

$$(\delta\xi_{(j)}/\xi(\lambda);\, j = 1, 2 \ldots) \tag{2.9}$$

since this sequence lies in $\bigtriangledown_\infty$ as a consequence of the equality

$$\xi(\lambda) = \sum_{j=1}^{\infty} \delta\xi_{(j)}. \tag{2.10}$$

For any integer n, choose N_0 so large that, for any $N \geq N_0$, the discontinuities of height $\delta\xi_{(j)}\,(j = 1,\, 2, \ldots,\, n)$ are contained in distinct intervals

$((i-1)\alpha_N,\, i\alpha_N)$. Then

$$\xi(N\alpha_N)q_{(j)} \geq \delta\xi_{(j)} \quad (1 \leq j \leq n, \quad N \geq N_0),$$

so that

$$\underline{\lim}\, q_{(j)} \geq \delta\xi_{(j)}/\xi(\lambda). \tag{2.11}$$

For $j = 1, 2, \ldots, n$. Since n is arbitrary, (2.8) holds for all j, and moreover, Fatou's lemma and (2.7) give

$$\overline{\lim}\, q_{(j)} = \overline{\lim}\{1-\sum_{i\neq j} q_{(i)}\} \leq 1-\sum_{i\neq j} \underline{\lim}\, q_{(j)} \leq 1-\sum_{i\neq j}\{\delta\xi_{(i)}/\xi(\lambda)\} = \delta\xi_{(j)}/\xi(\lambda).$$

Hence,

$$\delta\xi_{(j)}/\xi(\lambda) \leq \underline{\lim}\, q_{(j)} \leq \overline{\lim}\, q_{(j)} \leq \delta\xi_{(j)}/\xi(\lambda).$$

Thus,

$$\lim q_{(j)} = \delta\xi_{(j)}/\xi(\lambda).$$

□

By definition of $\delta\xi_{(j)}/\xi(\lambda)$, we have

$$\delta\xi_{(1)}/\xi(\lambda) \geq \delta\xi_{(2)}/\xi(\lambda) \geq \ldots,$$

and

$$\sum_{k=0}^{\infty} \delta\xi_{(k)}/\xi(\lambda) = 1.$$

We will write

$$(\delta\xi_{(1)}/\xi(\lambda),\, \delta\xi_{(2)}/\xi(\lambda), \ldots) \sim \mathcal{PD}(0,\, \lambda)$$

where $\mathcal{PD}(0,\, \lambda)$ is the Poisson-Dirichlet distribution define as follows:

Definition 2.5.6 *Let $0 < \lambda < \infty$. Let $(\xi(t),\, t \in [0, \lambda])$ be a gamma subordinator and let $J_1 \geq J_2 \geq \ldots \geq 0$ be the ordered sequence of its jumps. The distribution on $\wedge_\infty$ of the random variable $(\frac{J_1}{\xi(\lambda)},\, \frac{J_2}{\xi(\lambda)},\, \ldots)$ is called the* **Poisson-Dirichlet** *distribution with parameter λ and is denoted by $\mathcal{PD}(0,\, \lambda)$.*

Theorem 2.1.1 shows that if

$$(p_1, \ldots, p_N) \sim \mathcal{D}(\alpha_N, \ldots, \alpha_N)$$

then the distribution of $(p_{(1)}, \ldots, p_{(N)})$ approximates $\mathcal{PD}(0, \lambda)$, if N is fairly large, the α_N being uniformly small and $N\alpha_N$ closed to λ.

Chapter 3

Introduction on Dirichlet Processes

Nonparametric methods try to avoid assumptions about the probability distributions in order to generate methods that can be used in settings where regular parametric assumptions do not work. Although applicable in more general circumstances, nonparametric models can lead to very complex mathematics in all but the simplest models. Also, there is an implicit tradeoff between the generality of nonparametric tests and the power to detect differences between populations. From a frequentist perspective, a parametric t-test has a higher power if the normality assumption is indeed true, but might badly under perform the sign test if it is false, given the same type I error. From a Bayesian perspective, posterior distributions obtained from nonparametric models tend to have larger variances than their parametric counterparts. Nonparametric methods have a long history in modern frequentist statistics, starting with Fisher's exact test (Fisher, 1922). In Bayesian statistics, nonparametric models are constructed through priors on rich families of distributions. Therefore, the term Bayesian nonparametrics is really a misnomer. Bayesian nonparametric models are not parameter free, but have an infinite number of parameters. Raiffa and Schlaifer (1961) and Ferguson (1973) in their seminal work on Bayesian nonparametrics mention some characteristics that should be kept in mind when constructing priors on spaces of distributions:

1. The class should be analytically tractable. Therefore, the posterior distribution should be easily computed, either analytically or through simulation.

2. The class should be rich, in the sense of having a large enough support.

3. The hyperparameters defining the prior should be easily interpreted.

The Dirichlet process can also be regarded as a type of stick-breaking prior (Sethuraman, 1994; Pitman, 1996; Ishwaran and James, 2001; Ongaro and Cattaneo, 2004).
This chapter makes a quick review of Bayesian nonparametric models and defintions, making special emphasis on the Dirichlet process.

3.1 Dirichlet processes

In a celebrated paper [19], Thomas S. Ferguson introduced a random distribution, called a Dirichlet process DP, such that its marginal w.r.t. any finite partition has a Dirichlet Distribution as defined in Chapter 1. A Dirichlet process is a random discrete distribution which is a very useful tool in nonparametric Bayesian statistics. The work of (Ferguson, 1973, 1974; Blackwell and MacQueen, 1973; Sethuraman, 1994) is the base for the most widely used nonparametric models for random distributions in Bayesian statistics, mainly due to the availability of efficient computational techniques. Some recent applications of the Dirichlet Process include finance (Kacperczyk et al., 2003), econometrics (Chib and Hamilton, 2002; Hirano, 2002), epidemiology (Dunson, 2005), genetics (Medvedovic and Sivaganesan, 2002; Dunson et al., 2007a), astronomic (Ishwaran et James (2002)) and auditing (Laws and OH́agan, 2002).

3.1.1 Definition and proprieties of the Dirichlet process

Let $\mathcal{H}$ be a set and let $\mathcal{A}$ be a $\sigma-$field on $\mathcal{H}$. We define below a random probability, on ($\mathcal{H}$, $\mathcal{A}$) by defining the joint distribution of the random variables ($P(A_1)$, ..., $P(A_m)$) for every m and every finite sequence of measurable sets ($A_i \in \mathcal{A}$ for all i). We then verify the Kolmogorov consistency conditions to show there exists a probability, $\mathcal{P}$, on ($[0,1]^{\mathcal{A}}$, $B\mathcal{F}^{\mathcal{A}}$) yielding these distributions. Here $[0, 1]^{\mathcal{A}}$ represents the space of all functions from $\mathcal{A}$ into $[0, 1]$, and $B\mathcal{F}^{\mathcal{A}}$ represents the σ-field generated by the field of cylinder sets .

It is more convenient to define the random probability P, by defining the joint distribution of ($P(B_1)$, ..., $P(B_m)$) for all k and all finite measurable partitions (B_1, ..., B_m) of $\mathcal{H}$.

If $B_i \in \mathcal{A}$ for all i, $B_i \cap B_j = \emptyset$ for $i \neq j$, and $\cup_{j=1}^{k} B_j = \mathcal{H}$. From these distributions, the joint distribution of ($P(A_1)$, ..., $P(A_m)$) for arbitrary measurable sets A_1, ..., A_m may be defined as follows.

Given arbitrary measurable sets A_1, ..., A_m, we define $B_{x_1,\ldots,x_m}$ where $x_j =$ 0 or 1, as

$$B_{x_1,\ldots,x_m} = \cap_{j=1}^{m} A_j^{x_j}$$

where $A_j^1 = A_j$, and $A_j^0 = A_j^c$. Thus $\{B_{x_1,\ldots,x_m}\}$ form a partition of $\mathcal{H}$. If we are given the joint distribution of

$$\{P(B_{x_1,\ldots,x_m});\, x_j = 0, \quad \text{or}\, 1 \quad j = 1, \ldots, m\} \tag{3.1}$$

then we may compute the joint distribution of ($P(A_1)$, ..., $P(A_m)$) by

$$P(A_i) = \sum_{\{(x_1,\ldots,x_m);\, x_i=1\}} P(B_{x_1,\ldots,x_i=1,\ldots,x_m}). \tag{3.2}$$

We note that if $A_1, \ldots, A_m$ is a measurable partition to start with, then this does not lead to contradictory definitions provided $P(\emptyset)$ is degenerate at 0. If we are given a system of distribution of $(P(B_1), \ldots, P(B_k))$ for all k and all measurable partitions $B_1, \ldots, B_k$, is one consistency criterion that is needed; namely,

CONDITION C :

If $(B'_1, \ldots, B'_k)$, and $(B_1, \ldots, B_k)$ are measurable partitions, and if $(B'_1, \ldots, B'_k)$ is a refinement of $(B_1, \ldots, B_k)$ with

$$B_1 = \cup_1^{r_1} B'_i, \quad B_2 = \cup_{r_1+1}^{r_2} B'_i, \ldots, B_k = \cup_{r_{k-1}+1}^{k'} B'_i,$$

then the distribution of

$$(\sum_1^{r_1} P(B'_i), \sum_{r_1+1}^{r_2} P(B'_i), \ldots, \sum_{r_{k-1}}^{k'} P(B'_i)),$$

as determined from the joint distribution of $(P(B'_1), \ldots, P(B'_{k'}))$, is identical to the distribution of $(P(B_1), \ldots, P(B_m))$

Lemma 3.1.1 *If a system of joint distributions of $(P(B_1), \ldots, P(B_m))$ for all k and measurable partition $(B_1, \ldots, B_k)$ satisfies condition C, and if for arbitrary measurable sets $A_1, \ldots, A_m$, the distribution of $(P(A_1), \ldots, P(A_m))$ is defined using (3.2), then there exists a probability $\mathcal{P}$ on, $([0,1]^{\mathcal{A}}, B\mathcal{F}^{\mathcal{A}})$ yielding these distribution.*

Proof

See [21] page 214.

Definition 3.1.1 *Let α be a non-null finite measure on $(\mathcal{H}, \mathcal{A})$.*
We say P is a Dirichlet process on $(\mathcal{H}, \mathcal{A})$ with parameter α if for every $k = 1, 2, \ldots,$ and measurable partition $(B_1, \ldots, B_k)$ of $\mathcal{H}$, the distribution of $(P(B_1), \ldots, P(B_k))$ is Dirichlet $\mathcal{D}(\alpha(B_1), \ldots, \alpha(B_k))$.

The measure α can be represented by cH, where $c = \alpha(\mathcal{H})$, the parameter of precision and $H(.) = \frac{(.)}{\alpha(\mathcal{H})}$.

Proposition 3.1.1 *Let P be a Dirichlet process on $(\mathcal{H}, \mathcal{A})$ with parameter α and let $A \in \mathcal{A}$. If $\alpha(A) = 0$, then $P(A) = 0$ with probability one. If $\alpha(A) > 0$, then $P(A) > 0$ with probability one. Furthermore,* $\mathbb{E}(P(A)) = \frac{\alpha(A)}{\alpha(\mathcal{H})}$.

Proof

By considering the partition (A, A^c), it is seen that $P(A)$ has a beta distribution, $\beta(\alpha(A), \alpha(A^c))$. Therefore

$$\mathbb{E}(P(A)) = \frac{\alpha(A)}{\alpha(\mathcal{H})}.$$

The Dirichlet process can be alternatively characterized in terms of its predictive rule (Blackwell and MacQueen, 1973). If $(\theta_1, \ldots, \theta_n)$ is an iid sample from $P \sim \mathcal{D}(cH)$, we can integrate out the unknown P and obtain the conditional predictive distribution of a new observation,

$$\theta_n | \theta_n, \ldots, \theta_1 \sim \frac{c}{c+n-1} H + \sum_{l=1}^{n-1} \frac{1}{c+n-1} \delta_{\theta_l}$$

whereδ_{θ_l} is the Dirac probability measure concentrated at θ_l. Exchangeability of the draws ensures that the full conditional distribution of any θ_l has this same form. This result, which relates the Dirichlet process to a Pólya urn, is the basis for the usual computational tools used to fit models based on the Dirichlet process.

The Dirichlet process can also be regarded as a type of stick-breaking prior (Sethuraman, 1994; Pitman, 1996; Ishwaran and James, 2001; Ongaro and Cattaneo, 2004). A stick-breaking prior has the form

$$P^N(.) = \sum_{i=1}^{N} p_k \delta_{\theta_k}(.) \qquad \theta_k \sim H$$

$$p_k = v_k \prod_{i=1}^{k-1} (1 - v_k) \qquad v_k \sim \beta(a_k, b_k) k = 1, \ldots, N \quad and \quad v_N = 1$$

where the number of atoms N can be finite (either known or unknown) or infinite. For example, taking $N = 1, a_k = 1 - a$ and $b_k = b + ka$ for $0 \leq a < 1$

and $b > -a$ yields the two-parameter Poisson-Dirichlet Process, also known as Pitman- Yor Process (Pitman, 1996), with the choice $a = 0$ and $b = c$ resulting in the Dirichlet Process (Sethuraman, 1994).

The stick-breaking representation is probably the most versatile definition of the Dirichlet Process. It has been exploited to generate efficient alternative MCMC algorithms and as the starting point for the definition of many generalizations that allow dependence across a collection of distributions, including the DDP (MacEachern, 2000), the π DDP (Griffin and Steel, 2006b) and the GSDP (Duan et al., 2007).

Finally, the Dirichlet Process can be obtained as the asymptotic limit of certain finite mixture models (Green and Richardson, 2001; Ishwaran and Zarepour, 2002). In particular consider the finite-dimensional Dirichlet-Multinomial prior

$$P^N(.) = \sum_{i=1}^{N} p_k \delta_{\theta_k}(.) \qquad p \sim \mathcal{D}(\frac{c}{N}, \ldots, \frac{c}{N}) \qquad \theta_k \sim H$$

which differs from a truncated stick-breaking representation of the Dirichlet Process in the way the weights have been defined. Ishwaran and Zarepour (2002) prove that for each measurable function g which is integrable with respect to H,

$$\int g(\theta) P^N(d\theta) \xrightarrow{P} \int g(\theta) P(d\theta)$$

where $P \sim \mathcal{D}(cH)$, i.e., the finite-dimensional Dirichlet-Multinomial prior converges in distribution to the Dirichlet process. This result not only provides another useful approximation, but also justifies frequently used finite mixture models as approximating a Dirichlet Process.

Conjugacy is another appealing property of the Dirichlet process. If $\theta_1, \ldots, \theta_n \sim P$ and $P \sim \mathcal{D}(cH)$, then

$$P|\theta_1, \ldots, \theta_n \sim \mathcal{D}\Big(cH + \sum_{i=1}^{n} \delta_i\Big)$$

Therefore, the optimal estimator under squared error loss for P is

$$\hat{P}(.) = \frac{c}{c+n}H(.) + \frac{1}{c+n}\sum_{i=1}^{n}\delta_{\theta_i}(.)$$

which converges to the empirical distribution as $n \to \infty$.
Antoniak (1974) studies the properties of draws from a distribution that follow a Dirichlet process. In particular, he proves that, if H is nonatomic, the probability of k distinct values on a sample $\theta_1, \ldots, \theta_n$ of size n is

$$\mathbb{P}(k) = c_n(k)n!c^k\frac{\Gamma(c)}{c+n}$$

for $k = 1, \ldots, n$, where $c_n(k)$ is a constant that can be obtained using recurrence formulas for Stirling numbers. The expected number of distinct values can be calculated as

$$\mathbb{E}(k|c,n) = \sum_{i=1}^{n}\frac{c}{c+n-1} \approx c\log(\frac{c+n}{c})$$

These results will be used later to construct computational algorithms that treat α as an unknown parameter and to elicit prior distributions for this parameter.

3.1.2 Mixtures of Dirichlet processes (MDP)

The following definitions are due to C. Antoniak [1].
Let $(U, \mathcal{B}, H)$ be a probability space called the index space. Let $(\Theta, \mathcal{A})$ be a measurable space of parameters.

Definition 3.1.2 *A transition measure on $U \times \mathcal{A}$ is a mapping α from $U \times \mathcal{A}$ into $[0, \infty)$ such that*

1. *for any $u \in U$, $\alpha(u, .)$ is a finite, nonnegative non-null measure on $(\Theta, \mathcal{A})$*

2. *for every $A \in \mathcal{A}$, $\alpha(., A)$ is measurable on $(U, \mathcal{B})$.*

Note that this differs from the definition of a transition probability in that $\alpha(u, \Theta)$ need not be identically one as we want $\alpha(u, .)$ to be a parameter for a Dirichlet process.

Definition 3.1.3 *A random distribution P is a mixture of Dirichlet processes on $(\Theta, \mathcal{A})$ with mixing distribution H and transition measure α, if for all $k = 1, 2, ...$ and any measurable partition $A_1, A_2, \ldots, A_k$ of Θ we have*

$$\mathcal{P}\{P(A_1) \leq y_1, \ldots, P(A_k) \leq y_k\} = \int_U D(y_1, \ldots, y_k | \alpha(u, A_1), \ldots, \alpha(u, A_k)) dH(u),$$

where $\mathcal{D}(y_1, \ldots, y_k | \alpha_1, \ldots, \alpha_k)$ denotes the distribution function of Dirichlet distribution with parameters $(\alpha_1, \ldots, \alpha_k)$.
In concise symbols we will use the heuristic notation:

$$P \sim \int_U \mathcal{D}(\alpha(u, .)) dH(u).$$

Roughly, we may consider the index u as a random variable with distribution H and given u, P is a Dirichlet process with parameter $\alpha(u, .)$. In fact U can be defined as the identity mapping random variable and we will use the notation $|_u \; for$ "$U = u$". In alternative notation

$$\begin{cases} u \sim H \\ P \,|_u \sim \mathcal{D}(\alpha_u) \end{cases} \tag{3.3}$$

where $\alpha_u = \alpha(u, .)$.

3.1.3 Dirichlet processes Mixtures

Since the DP and MDP models put probability one on the space of discrete measures, they are typically not good choices for modelling continuous data. Instead, they are more naturally employed as priors on the random mixing distribution over the parameters of a continuous distribution K with density k,

$$z \sim g(.) \qquad g(.) = \int k(.|\theta)H(d\theta) \qquad H \sim \mathcal{D}(cH_0); \tag{3.4}$$

resulting in a DP mixture (DPM) model (Lo, 1984; Escobar, 1994; Escobar and West, 1995). The DPM induces a prior on g indirectly through a prior on the mixing distribution H. A popular choice is the DPM of Gaussian distributions, where $\theta = (\mu, \Sigma)$ and $k(.|\theta) = \phi_p(.|\mu, \Sigma)$ is a p-variate normal kernel with mean μ and covariance matrix Σ.

Given an i.i.d sample $z_n = (z_1, \ldots, z_n)$, the posterior of the mixing distribution, $H_n(z^n)$, is distributed as a mixture of Dirichlet processes (MDP), i.e,

$$H_n(.|z_n) \sim \int \mathcal{D}\Big(cH + \sum_{i=1}^{n} \delta_{\theta_i}\Big) p(d\theta_1, \ldots, d\theta_n|z^n)$$

and the optimal density estimator under squared error loss, $g^n(z)$, is the posterior predictive distribution

$$\begin{aligned} g^n(z) = \mathbb{E}\Big[k(z|\theta)H^n(d\theta|z^n)\Big] &= \int k(z|\theta)\mathbb{E}[H^n(d\theta|z^n)] \\ &= \int k(z|\theta)\frac{cH_0(n)+\sum_{i=1}^{n}\delta_{\theta_i(\eta)}}{c+n} p(d\theta_1, \ldots, \theta_n|z^n). \end{aligned}$$

Density estimates arising from location-and-scale DP mixtures can be interpreted as Bayesian kernel density estimates with adaptive bandwidth selection. This interpretation is extremely appealing because it provides a direct link with well-known frequentist techniques and demonstrates the versatility of the model. Due to the discrete nature of the DP prior, the DPM model divides the observations into independent groups, each one of them assumed to follow a distribution implied by the kernel k. Therefore, DPM models can be used for clustering as well as for density estimation. In this setting, the model automatically allows for an unknown number of clusters.

3.2 Some properties and computions for DPMs

Computation for DPM models is typically carried out using one of the three different approaches: Pólya urn schemes that marginalize out the unknown distribution H (MacEachern, 1994; Escobar and West, 1995; MacEachern and Méuller, 1998; Neal, 2000, Ishwaran, H. and James, L. F. (2003)), truncation methods that use finite mixture models to approximate the DP (Ishwaran and James, 2001; Green and Richardson, 2001), and Reversible Jump algorithms (Green and Richardson, 2001; Jain and Neal, 2000; Dahl, 2003).
For computational purposes, it is convenient to rewrite model 1.2 using latent variables $\theta_1, \dots, \theta_n$ corresponding to observations $z_1, \dots, z_n$. In turn, these latent variables can be rewritten in terms of a set of $k \leq n$ unique values $\theta_1^*, \dots, \theta_k^*$ and a set of indicators $\zeta_1, \dots, \zeta_n$, such that $\theta_i = \theta^*_{\zeta_i}$.
Pólya urn samplers, also called marginal samplers, are popular in practice because they are relatively easy to implement and produce exact samples from the posterior distribution of θ. However, they are more useful when the baseline measure H_0 is conjugate to the kernel k. Escobar and West (1995) original algorithm uses the Pólya urn directly to simultaneously sample group indicators and group parameters. They note that

$$p(\theta_i|\theta_{-i}, z) = q_{i0} p(\theta_i|z_i, H_0) + \sum_{l=1,\, l\neq i}^{n} q_{i,l} \delta_{\theta_l}(\theta_i)$$

where $q_{i0} = c \int k(z_i|\theta) H_0(d\theta)$, $q_{il} = k(z_i|\theta_l)$ for $l \geq 1$ and $p(\theta_i|z_i,\, H_0)$ is the posterior distribution for θ_i based on the prior H_0 and a single observation z_i. MacEachern (1994) points out that mixing can be slow in this setting, and proposes to add an additional step to the Gibbs sampler that resamples the group parameters conditional on the indicators. Taking this idea one step forward, Bush and MacEachern (1996) note that, in the conjugate case, the group parameters can be easily integrated out, yielding a more efficient sampler. Finally, MacEachern and Méuller (1998) propose an algorithm that can be used in the nonconjugate case. Neal (2000) provides an excellent review of marginal methods.

Blocked samplers are a more recent idea and are based on approximations to the Dirichlet process by finite mixture models. They are straightforward to code, tend to have better mixing properties than marginal samplers and, unlike them, directly produce (approximate) draws from the posterior distribution $H^n(d\theta|Z^n)$. Their main drawback is that the samples only approximately follow the desired distribution. As an example, consider the truncation sampler of Ishwaran and James (2001), which starts with the finite stick breaking prior

$$P^N = \sum_{k=1}^{K} p_k \delta_{\theta_k}(.) \qquad \theta_k \sim H$$

$$p_k = v_k \prod_{i=1}^{k-1}(1 - v_k), \qquad v_k \sim \beta(a_k, b_k), \quad k = 1, \ldots, N-1 \quad \text{and} \quad V_N = 1$$

After proving that P^N converges in distribution to a Dirichlet process when $N \longrightarrow \infty$, the authors are able to construct a simple Gibbs sampler that exploits conjugacy between the generalized Dirichlet distribution and the multinomial distribution. A related approach is the retrospective sampler (Roberts and Papaspiliopoulos, 2007), who also use the stick breaking representation of the Dirichlet process to generate a sampler that avoids truncations but shares some of the advantages of the blocked sampler.

3.2.1 Dependent Dirichlet Process

The dependent Dirichlet process (DDP) (MacEachern, 1999, 2000) induces dependence in a collection of distributions by replacing the elements of the stick-breaking representation (Sethuraman, 1994) with stochastic processes. It has been employed by DeIorio et al. (2004) to create ANOVA-like models for densities, and by Gelfand et al. (2005) to generate spatial processes that allow for non-normality and nonstationarity. This last class of models is extended in Duan et al. (2007) to create generalized spatial Dirichlet processes (GSDP) that allow different surface selection at different locations, among others.

Along similar lines, the hierarchical Dirichlet process (HDP) (Teh et al., 2006) is another approach to introduce dependence. In this setting, multiple group-specific distributions are assumed to be drawn from a common Dirichlet Process whose base- 12 line measure is in turn a draw from another Dirichlet process. This allows the different distributions to share the same set of atoms but have distinct sets of weights. More recently, Griffin and Steel (2006b) proposed an order-dependent Dirichlet Process (πDDP), where the correspondence between atoms and weights is allowed to vary with the covariates. Also, Dunson and Park (2007) propose a kernel stick breaking that allows covariate dependent weights and fixed atoms.

An alternative approach to the DDP is to introduce dependence through linear combinations of realizations of independent Dirichlet processes. For example, Méuller et al. (2004), motivated by a similar problem to Teh et al. (2006), define the distribution of each group as the mixture of two independent samples from a DP process: one component that is shared by all groups and one that is idiosyncratic. Dunson (2006) extended this idea to a time setting, and Dunson et al. (2007b) propose a model for density regression using a kernel-weighted mixture of Dirichlet Processes defined at each value of the covariate.

Definition 3.2.1 *(**MacEachern** [2000]) Let I be an index set, let $\{\theta(t) : t \in I\}$ and $\{v(t) : t \in I\}$ be stochastic processes over I such that $z(t) \sim \beta(1, \alpha(t))$ for any $t \in I$ and define*

$$H_t = \sum_{i=1}^{\infty} p_i^*(t) \delta_{\theta_i^*(t)}(.), \tag{3.5}$$

where $\{\theta_i^*(t)\}_{i=1}^{\infty}$ are mutually independent collections of independent real-

izations of the stochastic processes $\{\theta(t) : t \in I\}$ *and* $\{v(t) : t \in I\}$*, and* $p_i^*(t) = v_i^*(t)\prod_{s=1}^{i-1}(1 - v_s^*(t))$*. The colloction of the probability measures* $\mathcal{H}_I = \{H_t : t \in I\}$ *is to follow a dependent Dirichlet process (DDP).*

DDP models are dense on a large class of distributions. Indeed, under mild conditions, the DDP assigns positive probability to every ϵ-ball centered on a finite collection of distributions that are absolutely continuous to the baseline measures corresponding to the same locations of the index space D (MacEachern, 2000). One of the most popular variates of the DDP is the "single-p" model, where the weights are assumed to be constant over I while the atoms are allowed to vary. Models of this form can be rewritten as regular DP models with atoms arising from a stochastic process. Therefore, standard Gibbs sampling algorithms can be used to perform inferences for the "single-p" DDP models. The main drawback of this approach is its inability to produce a collection of independent distributions. The hierarchical Dirichlet process (HDP) (Teh et al., 2006) can also be recast as a DDP model. The HDP places a prior on a collection of exchangeable distributions $\{G_1, \ldots, G_J\}$. Conditional on a probability measure G_0, the distributions in the collection are assumed to be iid samples from a regular Dirichlet process centered around G_0. In order to induce dependence, G_0 is in turn given another Dirichlet process prior. In summary,

$$G_i | G_0 \sim \mathcal{D}(cG_0)$$

$$G_0 \sim \mathcal{D}(\beta H)$$

Since G_0 is, by construction, almost surely discrete, the distributions G_i share the same set of random atoms (corresponding to those of G_0), but assign strictly different (although dependent) weights to each one of them. As is to be expected, H corresponds to the common expected value for each of the distributions in the collection, and β and c control the variance around

H and the dependence between distributions. Computation for the HDP is performed using a generalized Pólya urn scheme.

3.2.2 Nested Dirichelt process

Motivated by the multicenter studies, Abel Rodriguez et.al (2006) introduce nested Dirichlet process. In fact, subjects in different centers have different outcome distributions. The problem of nonparametric modeling of these distributions, borrowing information across centers while also allowing centers to be clustered. Starting with a stick-breaking representation of the Dirichlet process (DP), he replaces the random atoms with random probability measures drawn from a DP. This results in a nested Dirichlet process (nDP) prior, which can be placed on the collection of distributions for the different centers, with centers drawn from the same DP component automatically clustered together.

3.3 Some recent advances in Dirichlet models

Popular approaches for nonparametric functional estimation can be broadly divided in three main groups. One simple yet powerful alternative is kernel regression methods. These methods represent the unknown function as a linear combination of the observed values of the outcome variables, using covariate-based weights (Altman, 1992; Chu and Marron, 1991; Fan et al., 1995). Another class of methods assumes that the functions of interest can be represented as a linear combination of basis functions. The problem of estimating the function reduces to estimation of the basis coefficients. Splines, wavelets and reproducing kernel methods fall in this broad category (Vidakovic, 1999; Truong et al., 2005). A third alternative is to assume that the functions in question are realizations of stochastic processes, with the Gaussian process (GP) being a common choice (Rasmussen and Williams, 2006). Different approaches have been used to extend these methodologies to collections of functions. For example, when the function of interest is modelled as a linear combination of basis functions, hierarchical models on the basis coefficients can be used to accommodate different types of dependence. This approach has been successfully exploited by authors such as Rice and Silverman (1991); Wang (1998); Guo (2002); Wu and Zhang (2002) and Morris and Carroll (2006) to construct ANOVA and random effect models for curves. Along similar lines, Bigelow and Dunson (2007) and Ray and Mallick (2006) have used Dirichlet process priors as part of the hierarchical specification of the model in order to induce clustering across curves. Behseta et al. (2005) develop a hierarchical Gaussian process (GP) model, which treats individual curves as realizations of a GP centered on a GP mean function.

Recently, Abel Rodregez et al. propose a hierarchical model that allows us to simultaneously estimate multiple curves nonparametrically by using dependent Dirichlet Process mixtures of Gaussians to characterize the joint distribution of predictors and outcomes. About stick-breaking, recently, YeeWhye Teh et al.(2007) introduce The Indian buffet process (IBP) is a Bayesian nonparametric distribution where by objects are modelled using an unbounded number of latent features. He derives a stick-breaking representation for the IBP. Based on this new representation, he develops slice samplers for the IBP.

Chapter 4

Mixtures of continuous time Dirichlet processes

In this chapter, we first define, in section 1, continuous time Dirichlet processes. In section 2 we examine the case of the Brownian-Dirichlet process (BDP) whose parameter is proportional to a standard Wiener measure.
Next we show that some stochastic calculus formulas (Ito's formula, local time occupation formula) hold for BDP's.
Next, in section 3, we define mixtures of continuous time Dirichlet processes and we extend some, rather nontrivial computations of Antoniak (1974) [2].

4.1 Continuous time Dirichlet processes

From now, we take for $\mathcal{H}$ any standard Polish space of real functions defined on an interval $I \subset [0, \infty)$, for example the space $\mathcal{C}(I)$ (resp. $\mathcal{D}(I)$) of continuous (resp. cadlag) functions. For any $t \in I$, let $\pi_t : x \longrightarrow x(t)$ denote the usual projection at time t from the space $\mathcal{H}$ to $\mathbb{R}$. Recall that π_t maps any measure μ on $\mathcal{H}$ into a measure $\pi_t\mu$ on $\mathbb{R}$ defined by $\pi_t\mu(A) = \mu(\pi_t^{-1}(A))$ for any Borel subset A of $\mathbb{R}$.

The following proposition defines a continuous time process (X_t) such that for any $t \in \mathbb{R}$ X_t is a Ferguson-Dirichlet random distribution.

Proposition 4.1.1 *(Emilion, 2005) Let α be any finite measure on $\mathcal{H}$, let X be a Ferguson-Dirichlet random distribution $\mathcal{D}(\alpha)$ on $\mathcal{H}$ and let $X_t = \pi_t X$. Then the time continuous process $(X_t)_{t \in I}$ is such that for each $t \in I$, X_t is a Ferguson-Dirichlet random distribution on $\mathbb{R}$ $\mathcal{D}(\alpha_t)$ where $\alpha_t = \pi_t \alpha$. Moreover if $V^{(i)}$ is any iid sequence on $\mathcal{H}$ such that $V^{(i)} \sim \frac{\alpha}{\alpha(\mathcal{H})}$ and*

$$X(\omega) \stackrel{d}{=} \sum_{i=1}^{\infty} p_i(\omega) \delta_{V^{(i)}(\omega)}$$

where the sequence (p_i) is independent of the $V^{(i)}$'s and has a Poisson-Dirichlet distribution $\mathcal{PD}(\alpha(\mathcal{H}))$, then

$$X_t(\omega) \stackrel{d}{=} \sum_{i=1}^{\infty} p_i(\omega) \delta_{V^{(i)}(\omega)(t)}.$$

For sake of simplicity we deal with just one parameter α, but it can be noticed that two-parameter $X_{t,\alpha,\beta}$ continuous time Dirichlet process can be defined similarly by using two-parameter Poisson-Dirichlet distributions introduced in Pitman Yor (1997) [44].

Proof

Let $k \in \{1, 2, 3, ...\}$ and $A_1, ..., A_k$ a measurable partition of $\mathbb{R}$.
Then for any $t \in \mathbb{R}$, $\pi_t^{-1}(A_1), ..., \pi_t^{-1}(A_k)$ is a measurable partition of $\mathcal{H}$ so that, by definition of X, the joint distribution of the random vector

$$(X(\pi_t^{-1}(A_1)), ..., X(\pi_t^{-1}(A_k)))$$

is Dirichlet with parameters $(\alpha(\pi_t^{-1}(A_1)), ..., \alpha(\pi_t^{-1}(A_k))$. In other words $(X_t(A_1)), ..., X_t(A_k))$ is Dirichlet with parameters $(\alpha_t(A_1), ..., \alpha_t(A_k))$ and $X_t \sim \mathcal{D}(\alpha_t)$.

A consequence of the definition of π_t is that

$$\pi_t(\sum_{i=1}^{\infty} \mu_i) = \sum_{i=1}^{\infty} \pi_t \mu_i$$

for any sequence of positive measures on $\mathcal{H}$ and $\pi_t(\lambda\mu) = \lambda\pi_t(\mu)$ for any positive real number λ. Hence if $V^{(i)}$ is any i.i.d. sequence on $\mathcal{H}$ such that $V^{(i)} \sim \frac{\alpha}{\alpha(\mathcal{H})}$ and

$$X(\omega) \overset{d}{=} \sum_{i=1}^{\infty} p_i(\omega)\delta_{V^{(i)}(\omega)}$$

where (p_i) has a Poisson-Dirichlet distribution $\mathcal{PD}(\alpha(\mathcal{H}))$, then

$$X_t(\omega) = \pi_t(X(\omega)) \overset{d}{=} \sum_{i=1}^{\infty} p_i(\omega)\pi_t(\delta_{V^{(i)}(\omega)}) \ = \sum_{i=1}^{\infty} p_i(\omega)\delta_{V^{(i)}(\omega)(t)}$$

the last equality being due to the fact that $\pi_t(\delta_f) = \delta_{f(t)}$ for any $f \in \mathcal{H}$, as it can be easily seen. In addition the $V^{(i)}(t)$'s are iid with $V^{(i)}(t) \sim \pi_t(\frac{\alpha}{\alpha(\mathcal{H})}) = \frac{1}{\alpha(\mathcal{H})}\pi_t(\alpha) = \frac{1}{\alpha_t(R)}\alpha_t$. Moreover (p_i) has a Poisson-Dirichlet distribution $\mathcal{PD}(\alpha(\mathcal{H})) = \mathcal{PD}(\alpha_t(R))$ so that the preceding expression of $X_t(\omega)$ is exactly the expression of a Ferguson-Dirichlet random distribution $\mathcal{D}(\alpha_t)$ as a random mixture of random Dirac masses. □

As a corollary of the above proof and of Sethuraman stick-breaking construction (1994), we have the following result which is of interest for simulating continuous time Dirichlet processes. It shows that such processes of random distributions can be used to generate stochastic paths and to classify random curves.

Corollary 4.1.1 *(Continuous time stick-breaking construction) Let α be any finite measure on $\mathcal{H}$ and $\alpha_t = \pi_t\alpha$. Let $c = \alpha(\mathcal{H})$ and $H = \alpha/c$. For any integer N, let $V_1, \cdots, V_{N-1}$ be iid Beta$(1, c)$ and $V_N = 1$. Let $p_1 = V_1$, $p_k = (1 - V_1)\dots(1 - V_{k-1})V_k$, $k = 2, \cdots, N$. Let Z_k be iid H. Then, $P_{N,t} = \sum_{k=1}^{N} p_k\delta_{Z_{k,t}}$ converges a.e. to a continuous time Dirichlet process $\mathcal{D}(\alpha_t)$.*

Corollary 4.1.2 *Let X_t be as in the preceding proposition, then for any Borel subset A of $\mathbb{R}$, $(X_t(A))_{t\geq 0}$ is a Beta process, ie for any $t \geq 0$*

$$X_t(A) \sim Beta(\alpha_t(A), \alpha_t(A^c)).$$

4.2 Brownian-Dirichlet process

We suppose here that the parameter α is proportional to a standard Wiener measure W so that the $V^{(i)}$'s above are i.i.d. standard Brownian motions that we denote by B^i. The sequence (p_i) is assumed to be Poisson-Dirichlet(c) independent of $(B^i)_{i=0,1,\ldots}$

Definition 4.2.1 *Let X be a Dirichlet process such that $X \sim \mathcal{D}(cW)$, then the continous-time process (X_t) defined by $X_t = \pi_t X$, for any $t > 0$, is called a Brownian-Dirichlet process (BDP).*

As observed in the previous proposition, X_t is a random probability measure such that $X_t \sim \mathcal{D}(c\mathcal{N}(0,t))$ and if we have a representation

$$X(\omega) = \sum_{i=1}^{\infty} p_i(\omega)\delta_{B^i(\omega)},$$

then we also have

$$X_t(\omega) = \sum_{i=1}^{\infty} p_i(\omega)\delta_{B^i_t(\omega)}.$$

We show that stochastic calculus can be extended to such processes (X_t). Consider the filtration defined by

$$\mathcal{F}_0 = \sigma(p_i,\, i \in \mathbb{N}^*),$$

and for any $s > 0$,

$$\mathcal{F}_s = \mathcal{F}_0 \cup (\cup_i \sigma(B^i_u,\, u < s)).$$

4.2.1 Ito's formula

Proposition 4.2.1 *Let $f \in C^2$ be such that there exist a constant $c \in \mathbb{R}$ such that $\int_0^s (f'(B^i_u)^2 du < c$ for any i and any $s > 0$. Then,*

1. *$M_t = \sum_{i=1}^{\infty} p_i(\omega) \int_0^t f'(B^i_u) dB^i_u$ is a well-defined $(\mathcal{F}_s) -$ martingale,*

2. $V_t = \frac{1}{2}\sum_{i=1}^{+\infty} p_i(\omega)\int_0^t f"(B_u^i)du$ *is a well-defined process with bounded variation, and*

3. $< X_t - X_0,\ f >= M_t + V_t.$

Proof. Let

$$M_t^n(\omega) = \sum_{i=1}^{n} p_i(\omega)\int_0^t f'(B_u^i)dB_u^i,$$

and let $s < t$. Let $0 = t_1^{(k)} < t_2^{(k)} < \ldots < t_{r_k}^{(k)} = t$ be a sequence of subdivisions of $[0,\ t]$ such that

$$\int_0^t f'(B_u^i)dB_u^i = \lim_{k\longrightarrow+\infty}\sum_{l=1}^{r_k} f'(B^i_{t_l^{(k)}})(B^i_{t_{l+1}^{(k)}} - B^i_{t_l^{(k)}}),$$

the limit being taken in L_2-norm. We now show that M_t^n is a martingale. Note that we don't use below the fact that the sequence p_i has a Poisson-Dirichlet distribution. For sake of simplicity, in what follows, we omit the superscript (k) in $t_l^{(k)}$. We have

$$\begin{aligned}
\mathbb{E}(M_t^n \mid \mathcal{F}_s) &= \textstyle\sum_{i=1}^n \mathbb{E}\Big(p_i \int_0^t f'(B_i^i)dB_u^i \mid \mathcal{F}_s\Big) \\
&= \lim_{k\to\infty}\{\textstyle\sum_{i=1}^n \mathbb{E}\Big(p_i \sum_{\{l\,:\,t_l<s\}} f'(B_{t_l}^i)(B_{t_{l+1}}^i - B_{t_l}^i) \mid \mathcal{F}_s\Big) \\
&+ \textstyle\sum_{i=1}^n \mathbb{E}\Big(p_i \sum_{\{l:t_l>s\}} f'(B_{t_l}^i)(B_{t_{l+1}}^i - B_{t_l}^i) \mid \mathcal{F}_s\Big)\}.
\end{aligned}$$

In the case $t_l < s$, if we have in addition $t_{l+1} < s$ then

$$\mathbb{E}\Big(f'(B_{t_l}^i)(B_{t_{l+1}}^i - B_{t_l}^i) \mid \mathcal{F}_s\Big) = f'(B_{t_l}^i)(B_{t_{l+1}}^i - B_{t_l}^i)$$

while if $t_{l+1} > s$, writing $B_{t_{l+1}}^i - B_{t_l}^i = B_{t_{l+1}}^i - B_s^i + B_s^i - B_{t_l}^i$, we see that

$$\mathbb{E}\Big(f'(B_{t_l}^i)(B_{t_{l+1}}^i - B_{t_l}^i) \mid \mathcal{F}_s\Big) = f'(B_{t_l}^i)(B_s^i - B_{t_l}^i).$$

On the other hand in the case $t_l > s$ we have

$$\begin{aligned}
\mathbb{E}\Big(f'(B^i_{t_l})(B^i_{t_{l+1}} - B^i_{t_l}) \mid \mathcal{F}_s\Big) &= \mathbb{E}\Big(E(f'(B^i_{t_l})(B^i_{t_{l+1}} - B^i_{t_l}) \mid \mathcal{F}_{t_l}) \mid \mathcal{F}_s\Big) \\
&= \mathbb{E}\Big(f'(B^i_{t_l})\mathbb{E}(B^i_{t_{l+1}} - B^i_{t_l} \mid \mathcal{F}_{t_l}) \mid \mathcal{F}_s\Big) \\
&= \mathbb{E}\Big(f'(B^i_{t_l})\mathbb{E}(B^i_{t_{l+1}} - B^i_{t_l}) \mid \mathcal{F}_s\Big) = 0.
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbb{E}(M^n_t \mid \mathcal{F}_s) &= \textstyle\sum_{i=1}^n p_i \lim_{k\longrightarrow\infty} \Big(\sum_{\{l:t_{l+1}<s\}} f'(B^i_{t_l})(B^i_{t_{l+1}} - B^i_{t_l})\Big) \\
&+ f'(B^i_{t_s})(B^i_s - B^i_{t_s})
\end{aligned}$$

where t_s denotes the unique $t^{(k)}_l$ such that $t^{(k)}_l < s$ and $t^{(k)}_{l+1} > s$. Therefore

$$\mathbb{E}(M^n_t \mid \mathcal{F}_s) = \sum_{i=1}^n p_i(\omega) \int_0^s f'(B^i_u)dB^i_u = M^n_s$$

proving that M^n_t is a martingale. Moreover, since

$$\begin{aligned}
\mathbb{E}\Big((M^{(n)}_s)^2\Big) &= 2\textstyle\sum_{\{1\le i<j\le n\}} \mathbb{E}\Big(p_i p_j \int_0^s f'(B^i_u)dB^i_u \int_0^s f'(B^j_u)dB^j_u\Big) \\
&+ \textstyle\sum_{i=1}^n \mathbb{E}\Big[p_i^2(\int_0^s f'(B^i_u)dB^i_u)^2\Big] \\
&= \textstyle\sum_{i=1}^n \mathbb{E}(p_i^2)\mathbb{E}(\int_0^s f'(B^i_u)dB^i_u)^2) \\
&= \textstyle\sum_{i=1}^n \mathbb{E}(p_i^2)\mathbb{E}\Big(\int_0^s (f'(B^i_u))^2 du\Big) \le c\sum_{i=1}^\infty \mathbb{E}(p_i) = c
\end{aligned}$$

the martingale convergence theorem implies that M^n_t converges to a martingale

$$M_t = \sum_{i=1}^\infty p_i(\omega) \int_0^t f'(B^i_u)dB^i_u.$$

Finally, applying Ito's formula to each B^i, we get

$$\begin{aligned} < X_t(\omega) - X_0(\omega),\, f > &= \textstyle\sum_{i=1}^{\infty} p_i(\omega)(f(B_t^i) - f(B_0^i)) \\ &= \textstyle\sum_{i=1}^{\infty} p_i(\omega) \int_0^t f'(B_u^i) dB_u^i \\ &+ \tfrac{1}{2} \textstyle\sum_{i=1}^{\infty} p_i(\omega) \int_0^t f''(B_u^i) du \\ &= M_t + V_t \end{aligned}$$

where V_t is obviously a bounded variation process.

Corollary 4.2.1 *(***Stochastic integral***) Let X_t be a BDP given by*

$$X_t(\omega) = \sum_{i=1}^{\infty} p_i(\omega) \delta_{B_t^i(\omega)}.$$

Let (Y_t) be a real valued stochastic process and ϕ a bounded function defined on $\mathbb{R}$. Then the stochastic integral $\int \phi(Y_t) dX_t$ is defined as the measure such that

$$< \int \phi(Y_t) dX_t,\, f >= \sum_{i=1}^{\infty} \int \phi(Y_t) p_i(\omega) f'(B_t^i) dB_t^i + \frac{1}{2} \sum_{i=1}^{\infty} \int \phi(Y_t) p_i(\omega) f''(B_t^i) dt,$$

for any function f verifying the conditions of the preceding proposition.

4.2.2 Local time

The following result exhibits the local time of a Brownian-Dirichlet process as a density of occupation time.

Proposition 4.2.2 *Let (X_t) be a BDP*

$$X_t(\omega) = \sum_{i=1}^{\infty} p_i(\omega) \delta_{B_t^i(\omega)}.$$

Then for each $(T, x) \in \mathbb{R}_+ \times \mathbb{R}$*, there exist a random distribution* $L(T,\, x)$ *such that*

$$\int_{\mathbb{R}} L(T,\, x) f(x) dx = \int_0^T < X_s,\, f > ds,$$

for any f *Borel measurable and locally integrable on* $\mathbb{R}$*.*

Proof. Let $L_i(T, x)$ be the local time w.r.t. to $B^{(i)}$ so that for any $i \in \mathbb{N}$ we have

$$\int_{\mathbb{R}} L_i(T,\, x) f(x) dx = \int_0^T f(B_s^i) ds$$

and

$$\int_{\mathbb{R}} \sum_{i=1}^{n} p_i L_i(T, x) f(x) dx = \int_0^T \sum_{i=1}^{n} p_i f(B_s^i) ds.$$

Then, if $f \in L_\infty^+$, set of positif bounded functions, the monotone convergence theorem yields

$$\int_{\mathbb{R}} \sum_{i=1}^{\infty} p_i L_i(T,\, x) f(x) dx = \int_0^T \sum_{i=1}^{\infty} p_i f(B_s^i) ds$$

and the same holds if $f \in L_\infty$ by using $f = f_+ - f_-$. Letting $L(T,\, x) = \sum_{i=1}^{\infty} p_i L_i(T,\, x)$ we get the desired result. □

4.2.3 Diffusions

Definition 4.2.2 *A stochastic process* (ψ_t) *is called a diffusion w.r.t. to the BDP* (X_t) *if it has a.s. continuous paths and can be represented as*

$$\psi_t = \psi_0 + \int_0^t a(s) ds + \sum_{i=0}^{\infty} p_i(\omega) \int_0^t b_{i,\,s} dB_s^i$$

where $a \in L_1(\mathbb{R}_+)$ *and* $b_i \in L_2(\mathbb{R}_+)$ *for any integer* i*.*

The following result can be proved using the Banach fixed point theorem, similar to the classical case of a single Brownian motion.

Proposition 4.2.3 *Suppose that f and g_i, $i = 0, 1, \ldots$ are Lipshcitz functions from $\mathbb{R}$ to $\mathbb{R}$. Let u_0 be an $\mathcal{F}_0$-measurable square integrable r.v. Then there exist a diffusion (ψ_t) w.r.t. to the BDP (X_t) such that*

$$\begin{aligned} d\psi_t &= f(\psi_t)dt + \textstyle\sum_{i=0}^{\infty} p_i g_i(\psi_t) dB_t^i, \\ \psi_0 &= u_0. \end{aligned} \tag{4.1}$$

4.2.4 Mixtures of continuous time Dirichlet processes

We now consider the case where α_u is a finite measure on a function space like $\mathcal{C}(I)$ and $\mathcal{D}(I)$ (spaces defined in section 1).
The following proposition defines a continuous time process $(P_t)_t$ such that each P_t is a mixture of Dirichlet processes.

Proposition 4.2.4 *Let P be a mixture of Dirichlet distributions*

$$P \sim \int_U \mathcal{D}(\alpha_u) dH(u).$$

Let $P_t = \pi_t P$. Then, for each $t \geq 0$, P_t is a mixture of Dirichlet processes:

$$P_t \sim \int_U \mathcal{D}(\alpha_{u,t}) dH(u)$$

where $\alpha_{u,t} = \alpha_u(\pi_t^{-1}(.))$.

Proof

Let $A_1, A_2, \ldots, A_k$ be a partition of $\mathbb{R}$.

$$\begin{aligned} \mathcal{P}[P_t(A_1) \leq y_1, \ldots, P_t(A_k) \leq y_k] &= \mathcal{P}[P\pi_t^{-1}(A_1) \leq y_1, \ldots, P\pi_t^{-1}(A_k) \leq y_k] \\ &= \textstyle\int_U D(y_1, y_2, \ldots, y_k \mid (\alpha_u(\pi_t^{-1} A_i))_{1 \leq i \leq k}) dH(u), \end{aligned}$$

since $\pi_t^{-1}(A_1), \pi_t^{-1}(A_2), \ldots, \pi_t^{-1}(A_k)$ is a partition of Θ.

Therefore

$$P_t \sim \int_U \mathcal{D}(\alpha_{u,t}) dH(u).$$

□

4.2.5 Posterior distributions

We suppose now that the sample space of observations is $\mathcal{X} = \mathcal{C}(\mathbb{R}^+)$, where $\mathcal{C}(\mathbb{R}^+)$ denote the space of continuous functions from $\mathbb{R}^+$ to $\mathbb{R}$.
Let F be a transition probability from $\Theta \times \zeta$ into $[0, 1]$.
Let θ_t be a sample from P_t, i.e. $\theta_t \mid_{P_t, u} \sim P_t$ and $X(t) \mid_{P_t, \theta_t, u} \sim F(\theta_t, .)$.
Let H_x denote the conditional distribution of (θ_t, u) given $X(t) = x$.
Let H_{θ_t} denote the conditional distribution of u given θ_t.
The following proposition shows that if (P_t) is a mixture of Dirichlet processes then for each $t \in \mathbb{R}^+$ the posterior probability of P_t is also a mixture of Dirichlet processes.

Proposition 4.2.5 *If for any* $t \in \mathbb{R}^+$

$$\begin{cases} P_t \mid_u \sim \mathcal{D}(\alpha_{u,t}) \\ u \sim H \\ P_t \sim \int_U \mathcal{D}(\alpha_{u,t}) dH(u) \\ \theta_t \mid_{P_t, u} \sim P_t \\ X(t) \mid_{P_t, \theta_t, u} \sim F(\theta_t, .) \end{cases} \tag{4.2}$$

then

$$P_t \mid_{X(t)=x} \sim \int_{\Theta \times U} \mathcal{D}(\alpha_{u,t} + \delta_{\theta_t}) dH_x(\theta_t, u).$$

Proof

Let $A_1, A_2, \ldots, A_k$ be a partition of $\mathbb{R}$

$$\begin{aligned} \mathcal{P}[P_t(A_i) \le y_i, 1 \le i \le k \mid_{X(t)=x}] &= E[\mathcal{P}[P_t(A_i) \le y_i, i = 1, \ldots, k \mid_{X(t)=x, \theta_t, u}] \mid_{X(t)=x}] \\ &= E[D(y_1, y_2, \ldots, y_k \mid_{\beta_{u,t}(A_1), \ldots, \beta_{u,t}(A_k)}) \mid_{X(t)=x}] \\ &= \int_{U \times \Theta} D(y_1, \ldots, y_k \mid_{\beta_{u,t}(A_1), \ldots, \beta_{u,t}(A_k)}) dH_x(u, \theta). \end{aligned}$$

where $\beta_{u,t}(A_i) = \alpha_{t,u}(A_i) + \delta_{\theta_t}(A_i)$, for any $i = 1, \ldots, k$.
Therefore

$$P_t \mid_{X(t)=x} \sim \int_U \mathcal{D}(\alpha_{u,t} + \delta_{\theta_t}) dH_x(\theta_t, u).$$

□

As a corollary, let us show that the same result holds, if (P_t) is simply a continuous time Dirichlet process: the posterior distribution of P_t given $X(t) = x$ is still a mixture of continuous time Dirichlet processes.

Corollary 4.2.2 *If*

$$\begin{cases} P_t \sim \mathcal{D}(\alpha_t) \\ \theta_t \sim P_t \\ X(t) \mid_{P_t, \theta_t} \sim F(\theta_t, \, .) \end{cases} \tag{4.3}$$

then

$$P_t \mid_{X(t)=x} \sim \int_U \mathcal{D}(\alpha_{u,\,t} + \delta_{\theta_t}) dH_x(\theta_t).$$

Proof

Let $A_1, A_2, \ldots, A_k$ be a partition of $\mathbb{R}$

$$\begin{aligned} \mathcal{P}[P_t(A_i) \leq y_i, \, 1 \leq i \leq k \mid_{X(t)=x}] &= E[\mathcal{P}[P_t(A_i) \leq y_i, \, 1 \leq i \leq k \mid_{X(t)=x,\, \theta_t,\, u}] \mid_{X(t)=x}] \\ &= E[D(y_1, \, y_2, \ldots, \, y_k \mid_{\beta_{A_1,\,t}, \beta_{A_2,\,t}, \ldots, \beta_{A_k,\,t}}) \mid_{X(t)=x}] \\ &= \textstyle\int_\Theta D(y_1, \, y_2, \ldots, \, y_k \mid_{\beta_{A_1,\,t}, \beta_{A_2,\,t}, \ldots, \beta_{A_k,\,t}}) dH_x(\theta_t), \end{aligned}$$

where $\beta_{A_i,\,t} = \alpha_{t,\,u}(A_i) + \delta_{\theta_t}(A_i), \; i \in \{1, \, 2, \ldots, \, k\}$. Therefore

$$P_t \mid_{X(t)=x} \sim \int_\Theta \mathcal{D}(\alpha_t + \delta_{\theta_t}) dH_x(\theta_t).$$

Corollary 4.2.3 *If for any $t \in \mathbb{R}^+$*

$$P_t \sim \int_U \mathcal{D}(\alpha_{u,\,t}) dH(u)$$

and

$$\theta_t \sim P_t$$

then for any $t \in \mathbb{R}^+$

$$P_t \mid_{\theta_t} \sim \int_U \mathcal{D}(\alpha_{u,\,t} + \delta_{\theta_t}) dH_{\theta_t}(u).$$

Proof

Let $A_1,\, A_2, \ldots,\, A_k$ be a partition of $\mathbb{R}$

$$\begin{aligned}\mathcal{P}[P_t(A_i) \leq y_i,\, i = 1, \ldots,\, k\,|_{\theta_t}] &= E[\mathcal{P}[P_t(A_i) \leq y_i,\, i = 1, \ldots,\, k\,|_{\theta_t, u}]\,|_{\theta_t}] \\ &= E[D(y_1,\, y_2, \ldots,\, y_k \mid \beta_{u,t}(A_1), \ldots,\, \beta_{u,t}(A_k)) \mid \theta_t] \\ &= \int_U D(y_1,\, y_2, \ldots,\, y_k \mid \beta_{u,t}(A_1), \ldots,\, \beta_{u,t}(A_k)) dH_{\theta_t}(u).\end{aligned}$$

Therefore

$$P_t\,|_{\theta_t}\!\sim \int_U \mathcal{D}(\alpha_{u,t} + \delta_{\theta_t}) dH_{\theta_t}(u).$$

4.2.6 A Lemma of Antoniak

The following result will yield explicit expressions of conditional distributions. It is just an application of a Lemma of C. Antoniak to each P_t but we prefer to give its proof for completeness.

Consider the following notations and hypothesis.

Let $P \sim \int_U \mathcal{D}(\alpha_u) dH(u)$ as in theorem 3.

Let $\theta^\star = (\theta_1,\, \theta_2, ...,\, \theta_n)$ be a sample of size n from P.

Suppose that there exists a $\sigma - finite$, $\sigma - additive$ measure μ on $(\Theta,\, \mathcal{A})$ such that for each $u \in U$:

$i)$ α_u is $\sigma - additive$ and absolutely continuous with respect to μ

$ii)$ the measure μ has mass one at each atom of α_u.

Let $\alpha'_u\,(.)$ denote the Radon-Nikodym derivative of $\alpha_u\,(.)$ with respect to μ.

Let θ'_i denote the i^{th}-distinct value of θ_t in $\theta^\star$.

Let $n(\theta'_i)$ denote the number of times the value θ'_i occurs in $\theta^\star$.

Let $M_u = \alpha_u(\Theta)$ and let $m_u(\theta) = \alpha'_u(\theta'_i)$ if θ'_i is an atom of α_u, zero otherwise.

Last, let $x^{(n)} = x(x+1)(x+2)...(x+n-1)$, $n \in \mathbb{N} - \{0\}$.

Lemma 4.2.1 *Under the preceding hypotheses and notations, the condi-*

tional distribution $u|\theta^*$

$$dH_{\theta^\star}(u) = \frac{\frac{1}{M_u^{(n)}}\prod_{i=1}^{r}\alpha'_u(\theta'_i)(\alpha'_u(\theta'_i)+1)^{(n(\theta'_i)-1)}dH(u)}{\int_U \frac{1}{M_u^{(n)}}\alpha'_u(\theta'_i)(\alpha_u(\theta'_i)+1)^{(n(\theta'_i)-1)}dH(u)}$$

Proof

Referring to the proof of Proposition 3 in [1], we see that the likelihood of θ^t_{k+1}, given u, θ_1, $\theta_2, \ldots,$ θ_k is equal to $\frac{\alpha'_u(\theta_{k+1})d\mu}{M_u+k}$ for a value of θ_{k+1} which has not occurred previously in θ_1, $\theta_2, \ldots,$ θ_k, and is equal to $[\frac{\alpha_u(\theta_{k+1})+jd\mu}{\alpha_u(\Theta)+k)}]$ for a value of θ_{k+1} which has occurred previously j times in θ_1, $\theta_2, \ldots,$ θ_k. Hence the likelihood of $(u, \theta_1, \theta_2, \ldots, \theta_k)$ is

$$\begin{aligned} L(u,\, \theta_1,\, \theta_2, \ldots,\, \theta_n) &= L(\theta_i\, |_{u,\theta_1,\theta_2,\ldots,\theta_{n-1}})L(u,\theta_1,\theta_2,\ldots,\,\theta_{n-1}) \\ &= \prod_{i=1}^{k} L(\theta_i\, |_{u,\theta_1,\theta_2,\ldots,\theta_{i-1}}) \\ &= \frac{1}{M_u^{(n)}}\prod_{i=1}^{r}\alpha'_u(\theta'_i)(\alpha'_u(\theta'_i)+1)^{(n(\theta'_i)-1)}dH(u). \end{aligned}$$

Therefore,

$$L(\theta_1,\theta_2,\ldots,\theta_n) = \int_U \frac{1}{M_u^{(n)}}\prod_{i=1}^{r}\alpha'_u(\theta'_i)(\alpha'_u(\theta'_i)+1)^{(n(\theta'_i)-1)}dH(u).$$

where r is the number of distinct components of the random vector $(\theta_1,\, \theta_2, \ldots,\, \theta_k)$. We obtain $dH_{\theta^\star}$ by multiplying the above by $dH(u)$ and dividing by the unconditional distribution of $\theta^\star_t$. So,

$$dH_{\theta^\star_t}(u) = \frac{\frac{1}{M_u^{(n)}}\prod_{i=1}^{r}\alpha'_u(\theta'_i)(\alpha'_u(\theta'_i)+1)^{(n(\theta'_i)-1)}dH(u)}{\int_U \frac{1}{M_u^{(n)}}\alpha'_u(\theta'_i)(\alpha_u(\theta'_i)+1)^{(n(\theta'_i)-1)}dH(u)}. \quad \square$$

4.3 Explicit posteriors

4.3.1 Example 1 : α Wiener measure and H Bernoulli

Let W denote the standard Wiener measure on $\Theta = \mathcal{C}(\mathbb{R}_+)$, where $\mathcal{C}(\mathbb{R}_+)$ denote the space of continuous functions from $\mathbb{R}_+$ to $\mathbb{R}$. Let the space $U = \{0,\, 1\}$

Theorem 4.3.1 *Let P be a finite mixture of Dirichlet processes on $\mathbb{R}$ with transition measure $\alpha_u = cW_u$, where W_u is a Wiener measure, and mixing distribution $H \sim$ Bernoulli(p) with parameter $p \in]0, 1[$ and let $f_1, f_2, \ldots, f_n$ be a sample of size n of P. Then*

$$P \mid_{f_1, f_2, \ldots, f_n} \sim pH_1\ \mathcal{D}\left(cW_1 + \sum_{i=1}^{n} \delta_{f_i}\right) + (1-p)\, F_1\ \mathcal{D}\left(cW_0 + \sum_{i=1}^{n} \delta_{f_i}\right)$$

where F_1 and H_1 are two constants depending on W'_0 and W'_1, the Radon-Nikodym derivative of W_0 and W_1, respectively, w.r.t. $\mu = W_0 + W_1 + \sum_{i=1}^{n} \delta_{f_i}$.

Proof

According to Lemma 4.3.1

$$\begin{aligned} dH(u \mid_{f_1}) &= \frac{dW'_{f_1|u} dH(u)}{\int_{\{0,1\}} dW'_{f_1|u} dH(u)} \\ &= \frac{W'_u(f_1) dH(u)}{pW'_1(f_1) + (1-p)W'_0(f_1)}. \end{aligned}$$

Therefore the conditional distribution of $P \mid f_1$ is a mixture of Dirichlet processes given by :

$$\frac{pW'_1(f_1)}{pW'_1(f_1) + (1-p)W'_1(f_1)} \mathcal{D}\left(cW_1 + \delta_{f_1}\right) + (1-p) \frac{cW'_0(f_1)}{pcW'_1(f_1) + (1-p)W'_0(f_1)} \mathcal{D}(cW_0 + \delta_{f_1}).$$

Let us first examine the case of a sample of size 2. Again by Lemma 3.2.6, we have

$$\begin{aligned} dH\left(u \mid_{f_1, f_2}\right) &= \frac{\frac{cW'_u(f_1)W'_u(f_2)dH(u)}{(cW_u(\Theta)+1)W_u(\Theta)}}{\int_{\{0,1\}} \frac{cW'_u(f_1)W'_u(f_2)dH(u)}{(cW_u(\Theta)+1)W_u(\Theta)}} \\ &= \frac{\frac{cW'_u(f_1)W'_u(f_2)dH(u)}{(cW_u(\Theta)+1)W_u(\Theta)}}{p\frac{cW'_1(f_1)W'_1(f_2)}{(cW_1(\Theta)+1)W_1(\Theta)} + (1-p)\frac{cW'_0(f_1)W'_0(f_2)}{(cW_u(\Theta)+1)W_0(\Theta)}}. \end{aligned}$$

Therefore

$$P \mid_{f_1, f_2} \sim pH\mathcal{D}(cW_1 + \sum_{i=1}^{2} \delta_{f_i}) + (1-p)F\mathcal{D}(cW_0 + \sum_{i=1}^{2} \delta_{f_i})$$

where $H = H(0)$ and $F = H(1)$ are such that

$$H = \frac{\frac{cW_1'(f_1)W_1'(f_2)dH(u)}{(cW_1(\Theta)+1)W_1(\Theta)}}{p\frac{cW_1'(f_1)W_1'(f_2)}{(cW_1(\Theta)+1)W_1(\Theta)} + (1-p)\frac{cW_0'(f_1)W_0'(f_2)}{(cW_0(\Theta)+1)W_0(\Theta)}}$$

and

$$F = \frac{\frac{cW_0'(f_1)W_0'(f_2)}{(cW_0(\Theta)+1)W_0(\Theta)}}{p\frac{cW_1'(f_1)W_1'(f_2)}{(cW_1(\Theta)+1)W_1(\Theta)} + (1-p)\frac{cW_0'(f_1)W_0'(f_2)}{(cW_0(\Theta)+1)W_0(\Theta)}}$$

In the general case of a sample of size n, Lemma 1 yields

$$dH(u\,|_{f_1, f_2, \ldots, f_n}) \;=\; \frac{\frac{1}{M^{(n)}}\prod_{i=1}^{r}\frac{cW_u'(f_i)(cW_u'(f_i)+1)^{(n(f_i)-1)}dH(u)}{(cW_u(\Theta))^{(n)}}}{\int_{\{0,\,1\}}\frac{1}{M^{(n)}}\frac{\prod_{i=1}^{r}cW_u'(f_i)(cW_u'(f_i)+1)^{(n(f_i)-1)}dH(u)}{(cW_u(\Theta))^{(n)}}}$$

and

$$P\,|_{f_1, f_2, \ldots, f_n} \sim (1-p)H_1\mathcal{D}(cW_0 + \sum_{i=1}^{n}\delta_{f_i}) + pF_1\mathcal{D}(cW_1 + \sum_{i=1}^{n}\delta_{f_i})$$

where

$$H_1 = \frac{\frac{1}{M^{(n)}}\prod_{i=1}^{r}\frac{cW_0'(f_i)(cW_0'(f_i)+1)^{(n(f_i)-1)}dH(u)}{(cW_0(\Theta))^{(n)}}}{(p-1)\frac{1}{M^{(n)}}\prod_{i=1}^{r}\frac{cW_0'(f_i)(cW_0'(f_i)+1)^{(n(f_i)-1)}}{(cW_0(\Theta))^{(n)}} + p\frac{1}{M^{(n)}}\prod_{i=1}^{r}\frac{cW_1'(f_i)(cW_1'(f_i)+1)^{(n(f_i)-1)}}{(cW_1(\Theta))^{(n)}}},$$

$$F_1 = \frac{\frac{1}{M^{(n)}}\prod_{i=1}^{r}\frac{cW_1'(f_i)(cW_1'(f_i)+1)^{(n(f_i)-1)}dH(u)}{(cW_1(\Theta))^{(n)}}}{(p-1)\frac{1}{M^{(n)}}\prod_{i=1}^{r}\frac{cW_0'(f_i)(cW_0'(f_i)+1)^{(n(f_i)-1)}}{(cW_0(\Theta))^{(n)}} + p\frac{1}{M^{(n)}}\prod_{i=1}^{r}\frac{cW_1'(f_i)(cW_1'^{\prod}(f_i)+1)^{(n(f_i)-1)}}{(cW_1(\Theta))^{(n)}}}.$$

and where r is the number of distinct components of the random vector $(f_1,\, f_2, \ldots,\, f_n)$. $\square$

Remark 4.3.1 *We can generalize this theorem to the case of a finite mixture where H is distributed on $\{1,\, 2, \ldots,\, k\}$.*

4.3.2 Example 2 : α Wiener measure and H Gaussian

Let W denote the standard Wiener measure on $\mathcal{C}(\mathbb{R}_+)$. For any $u \in \mathbb{R}_+$ let W_u denote a Wiener measure with marginal distributions $\mathcal{N}(u,\, t\sigma^2)$, $t \in \mathbb{R}_+$.

Theorem 4.3.2 *Let P be a mixture of continuous time Dirichlet processes,*

$$P \sim \int \mathcal{D}(cW_u)dH(u)$$

with $u \sim H = \mathcal{N}(0,\, \rho^2)$, then for any $t \in \mathbb{R}$

$$P_t \sim \int \mathcal{D}(c\mathcal{N}(u,\, t\sigma^2))dH(u).$$

Let $\theta_1^t,\, \theta_2^t$ be a sample of size 2 from P_t. Then the conditional distribution of P_t given $\theta_1^t,\, \theta_2^t$ is a mixture of continuous time Dirichlet processes such that

$$P_t \mid \theta_1^t,\, \theta_2^t \sim \int \mathcal{D}(c\mathcal{N}_u + \sum_{i=1}^{2} \delta_{\theta_i^t})d\hat{H}_t(u)$$

where $\hat{H}_t(u) = H(u\,|_{\theta_1^t,\theta_2^t}) \sim \mathcal{N}(\mu_1^t,\, \sigma_{1,\,t}^2)$ is given in the proof below

Proof

According to corollary 4.3.2, the conditional distribution of a mixture of Dirichlet distributions $P_t \mid \theta_1^t,\, \theta_2^t$, is also a mixture of Dirichlet distributions, with parameter $c\mathcal{N}(u(t),\, t\sigma^2) + \sum_{i=1}^2 \delta_{\theta_i^t}$.

According to Lemma 4.3.1 the mixing distribution $\hat{H}(u)$ of u given θ_1^t, θ_2^t can be computed as follows.

Case $\theta_1^t \neq \theta_2^t$:

$$\begin{aligned} dH(u\,|_{\theta_1^t,\theta_2^t}) &= \frac{\frac{1}{(\alpha_{u,\,t}(\Theta))^{(2)}}\alpha'_{u,\,t}(\theta_1^t)\alpha'_{u,t}(\theta_2^t)dH(u)}{\int_{-\infty}^{+\infty}\frac{1}{(\alpha_{u,\,t}(\Theta))^{(2)}}\alpha'_{u,\,t}(\theta_1^t)\alpha'_{u,\,t}(\theta_2^t)dH(u)} \\ &= \frac{\frac{M}{\sqrt{t2\pi}\sigma}e^{\frac{-1}{2t\sigma^2}(\theta_1^t-u)^2}\frac{M}{\sqrt{t2\pi}\sigma}e^{\frac{-1}{2t\sigma^2}(\theta_2^t-u)^2}\frac{1}{\sqrt{t2\pi}\rho}e^{\frac{-1}{2t\rho^2}u^2}du}{\int_{-\infty}^{+\infty}\frac{M}{\sqrt{2t\pi}\sigma}e^{\frac{-1}{2t\sigma^2}(\theta_1^t-u)^2}\frac{M}{\sqrt{2t\pi}\sigma}e^{\frac{-1}{2t\sigma^2}(\theta_2^t-u)^2}\frac{1}{\sqrt{t2\pi}\rho}e^{\frac{-1}{2t\rho^2}u^2}du}. \end{aligned}$$

After simplification we get,

$$dH(u\,|_{\theta_1^t,\theta_2^t}) = \frac{\frac{1}{t\rho^2\sigma^2(2\pi)^{3/2}} e^{\frac{-1}{2}(u^2(\frac{2}{t\sigma^2}+\frac{1}{t\rho^2})-\frac{4(\theta_1^t+\theta_2^t)u}{t\sigma^2}} e^{\frac{-1}{2}(\frac{(\theta_1^t)^2+(\theta_2^t)^2}{t\sigma^2})} du}{\int_{\mathbb{R}} \frac{1}{t\rho^2\sigma^2(2\pi)^{3/2}} e^{\frac{-1}{2}(u^2(\frac{2}{t\sigma^2}+\frac{1}{t\rho^2})-\frac{4(\theta_1^t+\theta_2^t)u}{t\sigma^2}} e^{\frac{-1}{2}(\frac{(\theta_1^t)^2+(\theta_2^t)^2}{t\sigma^2})} du}$$

$$= \frac{1}{\sqrt{2\pi}} \frac{\sqrt{2t\rho^2+t\sigma^2}}{t\sigma\rho} e^{\frac{-1}{2}\frac{2t\rho^2+t\sigma^2}{t\sigma^2 t\rho^2}(u-\frac{2(\theta_1^t+\theta_2^t)}{2t\rho^2+t\sigma^2})^2}.$$

Hence,

$$H(u\,|_{\theta_1^t,\theta_2^t}) = \mathcal{N}(\mu_1^t, \sigma_{1,t}^2) \tag{4.4}$$

where $\mu_1^t = \frac{(\theta_1^t+\theta_2^t)\rho^2}{2\rho^2+\sigma^2}$, and $\sigma_{1,t}^2 = t\frac{\sigma^2\rho^2}{2\rho^2+\sigma^2}$.

Case $\theta_1^t = \theta_2^t$:

$$dH(u\,|_{\theta_1^t,\theta_2^t}) = \frac{\frac{1}{(\alpha_u(\Theta))^{(2)}}\alpha'_{u,t}(\theta_1^t)\alpha'_{u,t}(\theta_2^t)dH(u)}{\int_{-\infty}^{+\infty}\frac{1}{(\alpha_u(\Theta))^{(2)}}\alpha'_u(\theta_1^t)\alpha'_u(\theta_2^t)dH(u)}$$

$$= \frac{\frac{M}{\sqrt{t2\pi}\sigma} e^{\frac{-1}{2t\sigma^2}(\theta_1^t-u)^2} \frac{1}{\sqrt{t2\pi}\rho} e^{\frac{-1}{2t\sigma^2}u^2} du}{\int_{-\infty}^{+\infty}\frac{M}{\sqrt{t2\pi}\sigma} e^{\frac{-1}{2t\sigma^2}(\theta_1^t-u)^2} \frac{1}{\sqrt{t2\pi}\rho} e^{\frac{-1}{2t\sigma^2}u^2} du}$$

$$= \frac{\frac{1}{2t\pi\sigma\rho} e^{\frac{-1}{2}(u^2(\frac{1}{t\sigma^2}+\frac{1}{t\rho^2})-\frac{(\theta_1^t)u}{t\sigma^2}+\frac{(\theta_1^t)u}{t\sigma^2}} du}{\int_{-\infty}^{+\infty}\frac{1}{2\pi\sigma\rho} e^{\frac{-1}{2}(u^2(\frac{1}{t\sigma^2}+\frac{1}{t\rho^2})-\frac{(\theta_1^t)u}{t\sigma^2}+\frac{(\theta_1^t)u}{t\sigma^2})} du}.$$

As above, we get

$$dH(u\,|_{\theta_1^t,\theta_2^t}) = \frac{\frac{1}{\sqrt{2\pi}}\frac{\sqrt{t\rho^2+t\sigma^2}}{t\sigma\rho} e^{\frac{-1}{2}\frac{t\rho^2+t\sigma^2}{t\sigma^2 t\rho^2}(u-t\frac{2\theta_1^t\rho^2}{t\rho^2+t\sigma^2})^2 du}}{\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\frac{\sqrt{t\rho^2+t\sigma^2}}{t\sigma\rho} e^{\frac{-1}{2}\frac{t\rho^2+t\sigma^2}{t\sigma^2 t\rho^2}(u-\frac{t2\theta_1^t\rho^2}{t\rho^2+t\sigma^2})^2 du}}$$

$$= \frac{1}{\sqrt{2\pi}}\frac{\sqrt{t\rho^2+t\sigma^2}}{t\sigma\rho} e^{\frac{-1}{2}\frac{t\rho^2+t\sigma^2}{t\sigma^2 t\rho^2}(u-\frac{\theta_1^t t\rho^2}{t\rho^2+t\sigma^2})^2 du}.$$

Therefore

$$H(u\,|_{\theta_1^t,\theta_2^t}) \sim \mathcal{N}(\mu_1^t, \sigma_{1,t}^2) \tag{4.5}$$

where $\mu_1^t = \frac{\rho^2\theta_1^t}{\rho^2+\sigma^2}$, and $\sigma_{1,t}^2 = t\frac{\sigma^2\rho^2}{\rho^2+\sigma^2}$.

Remark 4.3.2 *Note that the mixing distribution H is gaussian depending on the parameter t.*

4.4 Parameter estimation problems

In this section we incorporate the time parameter in a sampling model of C.Antoniak ([1] page 1165) which leads to estimates different from standard Bayesian analysis.

Let

$$G : \Omega \longrightarrow \mathbf{P}(\mathcal{C}(\mathbb{R}_+))$$
$$G \sim \int \mathcal{D}(\alpha_u) dH(u)$$

Let $\theta_1,\, \theta_2, ..., \,\theta_n \in \mathbf{P}(\mathcal{C}(\mathbb{R}))$ be a sample from G and $\theta_i^t = \pi_t(\theta_i)$.

Let

$$G_t : \Omega \longrightarrow \mathbf{P}(\mathbb{R})$$

where $G_t(.) = G(\pi_t(.))$

$$G_t \sim \int \mathcal{D}(\alpha_{u,\,t}) dH(u).$$

If α_u is the Wiener W_u measure then

$$\alpha_{u,t} = \mathcal{N}(u,\, t\sigma^2).$$

Hence $G_{m,\,t} = G_t] - \infty,\, m]$ is a distribution function from a mixture of Beta distributions with parameter $\alpha_{u,\,t}$ and mixture distribution H.

Let $g^t_{1,\,m},\, g^t_{2,\,m}, ...,\, g^t_{n,\,m}$ be a sample of size n from $G_{m,\,t}$ and let $X^t_{i1}, ...,\, X^t_{im_i}$ be a sample of size m_i from $F_{\theta^t_{i,\,m}}(x)$.

As in [1], consider the two following problems

(a) Estimating the index of the parameter

(b) Estimating the mixing distribution function.

In problem (a), if we wish to estimate u with square error loss, then the Bayes estimate is simply

$$U'_t = E(u\mid_{\theta^t_1, ...,\, \theta^t_n})$$

if the θ^t_i are observed directly, and

$$U'_t = E(u\mid_{X^t_{i1}, ...,\, X^t_{nm_n}})$$

if we only observe X_{ij}^t.
In problem (b) $\hat{G}_t = E(G_t \mid_{\theta_1^t, \theta_2^t, \ldots, \theta_n^t})$ is the Bayes estimate when the θ_i, are observed and $\hat{G}_t = E(G_t \mid_{X_{i1}^t, \ldots, X_{nmn}^t})$ when only the X_{ij}^t are observed.
Using ([1] page 1166) we get

$$G_t \mid_{\theta_1^t, \theta_2^t} \sim \int_{-\infty}^{+\infty} \mathcal{D}(\alpha_{u,\,t} + \delta_{\theta_1^t} + \delta_{\theta_2^t}) dH(u \mid_{\theta_1^t, \theta_2^t}),$$

where $H(u \mid_{\theta_1^t, \theta_2^t}) = \mathcal{N}(\mu_1^t, \sigma_{1,\,t}^2)$ (see theorem 2). Further

$$G \mid_{X_1^t} \sim \int_{-\infty}^{+\infty} \mathcal{D}(\alpha_u + \delta_{\theta_1^t}) dH_{X_1}(\theta_2^t,\, u),$$

where $H_{X_1}$is a bivariate Normal with parameters

$$\begin{cases} \mu_{1,\,t} = X_1^t (t\rho^2 + t\sigma^2 + t\tau^2)^{-1} (t\rho^2 + t\sigma^2) \\ \mu_{2,\,t} = t X_1^t (t\rho^2 + t\sigma^2 + t\tau^2)^{-1} \rho^2 \\ \sigma_{1,\,t}^2 = \alpha^t t\tau (t\rho^2 + t\sigma^2) \\ \sigma_{2,\,t}^2 = (t\rho^2 + t\sigma^2 + t\tau^2)^{-1} t\rho^2 (t\rho^2 + t\sigma^2) \\ \sigma_{21,\,t} = t\alpha^t \tau^2 \rho^2. \end{cases} \tag{4.6}$$

For (a) we get $U'_{\theta^t} = \frac{2(\theta_1^t + \theta_2^t)\rho^2}{2\rho^2 + \sigma^2}$ when $\theta_1^t \neq \theta_2^t$ and $U'_{\theta^t} = \frac{2(\theta_1^t + \theta_2^t)\rho^2}{\rho^2 + \sigma^2}$ when $\theta_1^t = \theta_2^t$. Since we do note observe whether $\theta_1^t = \theta_2^t$ or not, we must weight these two estimates according to the posterior probability, given X_1^t and X_2^t and we get an estimate

$$U_t'' = p_{d,\,t} U_t' + p_{s,\,t} U_t'^{*}$$

where $p_{s,t} = P(\theta_1^t = \theta_2^t \mid X_1^t,\, X_2^t)$, and $p_{d,\,t} = 1 - p_{s,\,t}$.
Concerning problem (b), the computation of $E(G_{\theta,\,t} \mid u,\, \theta_1^t,\, \theta_2^t)$ is slightly different from those in ([1] page 1167) because the time parameter also appears in H:

$$E(G_{\theta,\,t} \mid_{u,\, \theta_1^t,\, \theta_2^t}) = \frac{\alpha_{u,\,t}(]-\infty,\, \theta]) + \delta_{\theta_1^t}(]-\infty,\, \theta]) + \delta_{\theta_2^t}(]-\infty,\, \theta])}{\alpha_{u,\,t}(\mathbb{R}) + \delta_{\theta_1^t}(\mathbb{R}) + \delta_{\theta_2^t}(\mathbb{R})},$$

hence for $\theta_1^t \neq \theta_2^t$ we have

$$\begin{aligned}
E(G_{\theta,\,t}\mid_{\theta_1^t,\,\theta_2^t}) &= \int_{-\infty}^{+\infty} E(G_{\theta,\,t}\mid_{u,\theta_1^t,\,\theta_2^t}) dH(u\mid_{\theta_1^t,\theta_2^t}) \\
&= \int_{-\infty}^{+\infty} \tfrac{\alpha_{u,\,t}(]-\infty,\,\theta])+\delta_{\theta_1^t}(]-\infty,\theta])+\delta_{\theta_2^t}(]-\infty,\,\theta])}{\alpha_u(\mathbb{R})+\delta_{\theta_1}(\mathbb{R})+\delta_{\theta_2}(\mathbb{R})} dH(u\mid_{\theta_1^t,\theta_2^t}) \\
&= \int_{-\infty}^{+\infty} \tfrac{M}{M+2}\mathcal{N}(u,\,t\sigma^2)(]-\infty,\,\theta])dH(u\mid_{\theta_1^t,\theta_2^t}) + \tfrac{2}{M+2}F_2(]-\infty,\,\theta]).
\end{aligned}$$

$$\begin{aligned}
\hat{G}_t(\theta) &= E(G_t(]-\infty,\,\theta])\mid_{\theta_1^t,\theta_2^t}) \\
&= \tfrac{M}{M+2}\int_{\mathbb{R}} \Phi(\tfrac{\theta-u}{\sigma})dH(u\mid_{\theta_1^t,\theta_2^t}) + \tfrac{2}{M+2}F_2(]-\infty,\,\theta]) \\
&= \tfrac{M}{M+2}\int_{-\infty}^{+\infty} \tfrac{1}{\sqrt{2\pi t}\sigma\sqrt{2\pi t}\sigma_1}\left(\int_{-\infty}^{\theta} e^{\frac{-1}{2}\frac{(x-u)^2}{t\sigma^2}}dx\right)e^{\frac{-1}{2}\frac{(u-\mu_{1,t})^2}{t\sigma_1^2}}du + \tfrac{2}{M+2}F_2(]-\infty,\,\theta]).
\end{aligned}$$

Using Fubini formula, we get

$$\begin{aligned}
\hat{G}_t(\theta) &= \tfrac{M}{M+2}\int_{-\infty}^{\theta}\left(\int_{-\infty}^{+\infty} \tfrac{1}{\sqrt{2\pi t}\sigma\sqrt{2\pi t}\sigma_1} e^{\frac{-1}{2}\frac{(x-u)^2}{t\sigma^2}} e^{\frac{-1}{2}\frac{(u-\mu_{1,\,t})^2}{t\sigma_1^2}}du\right)dx + \tfrac{2}{M+2}F_2(]-\infty,\,\theta]) \\
&= \tfrac{M}{M+2}\int_{-\infty}^{\theta} \tfrac{\sqrt{t\sigma_1^2+t\sigma^2}}{\sqrt{2\pi}} e^{\frac{-1}{2}\frac{1}{t\sigma_1^2+t\sigma^2}\frac{(x-\mu_1)^2}{t\sigma^2}}dx \\
&= \mathcal{N}(\mu_{1,\,t},\,t\sigma_1^2+t\sigma^2)(]-\infty,\,\theta]) + \tfrac{2}{M+2}F_2(]-\infty,\,\theta]).
\end{aligned}$$

Therefore for $\theta_1^t \neq \theta_2^t$ we get

$$\begin{aligned}
\hat{G}_t &= \tfrac{M}{M+2}\mathcal{N}(\mu_{1,t},\,t\sigma_1^2+t\sigma^2) + \tfrac{\delta_{\theta_1}+\delta_{\theta_2}}{M+2} \\
&= \tfrac{M}{M+2}\mathcal{N}(\mu_{1,\,t},\,\tfrac{t^2(\sigma_1)^2+3t\rho^2t\sigma^2}{2t\rho^2+t\sigma^2}) + \tfrac{\delta_{\theta_1^2}+\delta_{\theta_2^2}}{M+2}.
\end{aligned}$$

If $\theta_1^t = \theta_2^t$, then for reasons given above, we get

$$\hat{G}_t = \frac{M}{M+2}\mathcal{N}\left(\frac{\rho^2\theta_1^t}{\rho^2+\sigma^2},\,\frac{(t\sigma_1)^2+2t\rho^2t\sigma^2}{t\rho^2+t\sigma^2}\right) + \frac{2\delta_{\theta_1^t}}{M+2}.$$

Chapter 5

Continuous time Dirichlet hierarchical models

In some recent and interesting papers, hierarchical models with a Dirichlet prior, shortly Dirichlet hierarchical models, were used in probabilistic classification applied to various fields such as biology ANTONIAK, C.E. (1974)., astronomy ISHWARAN, H. and JAMES, L.F. (2002). or text mining BLEI, D. and JORDAN., I. J. (2005). Actually, these models can be seen as complex mixtures of real Gaussian distributions fitted to non-temporal data.
The aim of this chapter is to extend these models and estimate their parameters in order to deal with temporal data following a stochastic differential equation (SDE).
The chapter is organized as follows. In section 2 we briefly recall Dirichlet hierarchical models. In section 3 we consider the case of a Brownian motion with a Dirichlet prior on its variance which is shown to be a limit of a random walk in Dirichlet random environment. As an application, we estimate, in section 4, regime switching models with stochastic drift and volatility.
In section 5, we consider the case of functional data such as signals or solutions of SDE's. Computing some posterior distributions in the multivariate

case, the preceding method is extended in order to classify such functional data.

5.1 Dirichlet hierarchical models

Let $P \sim \mathcal{D}(cH)$ denote a Dirichlet process with precision parameter $c > 0$ and mean parameter H, where H is a probability measure on a Polish space $\mathcal{X}$. It is well-Known that P can be approximated by

$$P = \sum_{k=1}^{N} p_k \delta_{X_k(.)}$$

where

$$\begin{cases} X_i \overset{iid}{\sim} H \\ (p_i) \sim SB(c,\, N) \quad (5.1) \\ (p_i) \perp (X_i), \end{cases}$$

$SB(c,\, N)$ denoting the stick-breaking scheme of Sethuraman. We will say that $(X_i)_{1,\,2,\ldots,}$ follows a Dirichet hierarchical model if

$$\begin{cases} X_i \mid P \overset{iid}{\sim} P \\ P \sim \mathcal{D}(c,\, H). \end{cases} \quad (5.2)$$

5.2 Brownian motion in Dirichlet random environment

5.2.1 Random walks in random Dirichlet environment

Let $\mathcal{D}(c\alpha)$ denote a Dirichlet process with parameters $c > 0$ and α, a finite measure on a polish space $\mathcal{X}$.

Consider a random variable $\mathcal{H}$ and a sequence (U_i) of random variables defined by the following hierarchical model

$$\begin{cases} U_i \mid \mathcal{V} = \sigma \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \\ \mathcal{V}^{-1} \mid P \sim P \\ P \mid c \sim \mathcal{D}(c\Gamma(\nu_1, \nu_2)) \\ c \sim \Gamma(\eta_1,\, \eta_2). \end{cases} \tag{5.3}$$

Since $\mathcal{V}$ is sampled from a Dirichlet process, we have $\sigma < \infty$ a.e. because

$$\mathcal{P}(\mathcal{V} < \infty) = \mathbb{E}(\mathbb{E}(\mathcal{V} \in \mathbb{R} \mid P,\, P(\mathbb{R}))) = \mathbb{E}(P(\mathbb{R})) = 1$$

Hence, we are allowed to consider the following random walk $(S_n)_{n\in\mathbb{N}}$ in Dirichlet random environment, starting from 0:

$$S_n = U_1 + U_2 + \ldots + U_n.$$

For any real number $t \geq 0$ let

$$S_t^n = \frac{1}{n^{1/2}} S_{[nt]} \tag{5.4}$$

where $[x]$ denotes the integer part of x.

Let $B^\sigma = \sigma B$ denote a zero mean Brownian motion with variance σ^2, B denoting a standard Brownian motion independent from $\mathcal{V}$.

Proposition 5.2.1

$$(S_t^n)_{t\geq 0} \xrightarrow{d} \mathcal{V}B.$$

Proof

Let $E = \mathcal{C}(\mathbb{R}_+)$ be the space of real-valued continuous functions defined on $\mathbb{R}_+$. For any bounded continuous function f defined on E we have

$$\int f((S_t^n)) d\mathcal{P} = \int_{\mathbb{R}} (\int_E f(x) d\mathcal{P}_{S_t^n|\sigma'=\sigma}) d\mathcal{P}(\sigma).$$

But, a standard result on the convergence of Gaussian random walks is that

$$\int_E f(x) d\mathcal{P}_{S_t^n|\mathcal{V}=\sigma} \longrightarrow \int_E f(x) d\mathcal{P}_{B^\sigma}$$

and this integral is dominated by $\| f \|$.

Hence by the dominated convergence theorem we have

$$\begin{aligned}\int (f(S_t^n)_{t\geq 0}) d\mathcal{P} \quad &\longrightarrow \quad \int_{\mathbb{R}} \Big(\int_E f(x) d\mathcal{P}_{B^\sigma}(x) \Big) d\mathcal{P}_\sigma(\sigma) \\ &= \quad \int_{\mathbb{R}} \Big(\int_E f(\sigma x) d\mathcal{P}_B \Big) d\mathcal{P}_\sigma(\sigma) \\ &= \quad \int f(\sigma B) d\mathcal{P}\end{aligned}$$

the last equality being due to the fact that B and σ' are independent.

Definition 5.2.1 *A Brownian motion in Dirichlet random environment (BMDE) is a process Z such that*

$$\begin{cases} Z \mid \mathcal{V} = \sigma = \mathcal{L}(B^\sigma) \\ \mathcal{V}^{-1} \mid P \sim P \\ P \mid c \sim \mathcal{D}(c\Gamma(\nu_1,\, \nu_2)) \\ c \sim \Gamma(\eta_1,\, \eta_2). \end{cases}$$

So, the above random walks in Dirichlet environment converge to a BMDE.

5.2.2 Simulation algorithm

An order to simulate a M paths $Z^1, \ldots, Z^M$ of $BMDE$, proceed as follows:
A path of a BMDE process $(Z_0 = 0, Z_{t_1}, \ldots, Z_{t_n})$ can be simulated as follows:
Let $dt = t_{i+1} - t_i > 0$ be small enough and let K be the stick-breaking precision.
Draw c from $\Gamma(\eta_1, \eta_2)$ and draw $q = (q_1, q_2, \ldots, q_K)$ from $SB(c, N)$.
Draw $x = (x_1, x_2, \ldots, x_K)$ with x_i's $\overset{iid}{\sim} \Gamma(\nu_1,\, \nu_2)$.
Repeat M times:
Draw σ^{-1} from $\sum_{i=1}^K q_i \delta_{x_i}$, draw $Z_0 = 0$ and n points Z_{t_i} such that $Z_{t_{i+1}} - Z_{t_i} \overset{iid}{\sim} \mathcal{N}(0,\, \sigma^2 dt)$.

Simulations

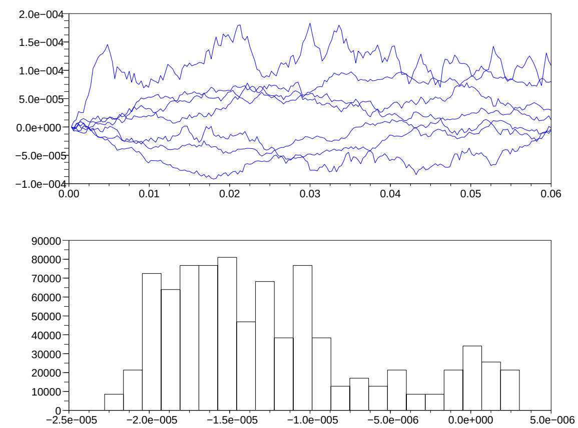

Figure 5.1: M Paths of BMDE and non Gaussian density of $(Z^1_{t_i}, \ldots, Z^M_{t_i})$.

5.2.3 Estimation

Due to proposition 1, given an observed path $(z_{t_i}$ of a BMDE, an estimation of its parameters can be obtained by performing Ishwaran and James blocked Gibbs algorithm with 0 means and equal variances on the data $z_{t_{i+1}} - z_{t_i}$ (see Ishwaran - James paper, Section 3).

5.3 Description of the model

let $(\omega,\, \mathcal{F},\, \mathcal{F}_t,\, P)$ be a stochastic basis and (W_t) a one dimensional Wiener process adapted to $(\omega,\, \mathcal{F},\, \mathcal{F}_t, P)$. We consider a stochastic process satisfying the following SDE:

$$dX_t = b(t,\, X_t)dt + \theta(t)h(X_t)dW_t$$

where the function $h(.)$ is assumed to be unknown, the volatility coefficient $\theta(.)$ is a known function of time and has to be correctly estimated, the drift coefficient $b(t,\, x)$ may be unknown. We observe one sampling path of the process $(X_t,\, t \in [0,\, T])$ at the discrete times $t_i = i\triangle$ for $i = 1, \ldots, N$. The sampling interval $\triangle$ is small in comparison of T. Let assume that $N := T\triangle^{-1}$

is an integer.

We will use the following assumptions:

• (A0): $\theta(t)$ is adapted to the filtration $\mathcal{F}_t$, $b(t,.)$ is non-anticipative map, $b \in C^{-1}(\mathbb{R}^+, \mathbb{R})$ and the exist $L_T > 0$ such that $\forall L_T > 0$ such that $\forall t \in [0, T]$, $\mathbb{E}(\theta^4) \leq L_T$ and $\mathbb{E}(\theta^8) \leq L_T$.

• (A1): $\theta(.) = \sum_{\rho=0}^{f} \theta_\rho \mathbb{1}_{[t_\rho,\, t_{\rho+1})}(.)$ where t_ρ is the volatility jump times.

• (A2): $\exists > 0$ such that $\theta^2(.)$ is almost surely Hölder continuous of order m with a constant $K(\omega)$ and $\mathbb{E}(K(\omega)^2) < +\infty$.

If we assume that the volatility jump times correspond to the sampling times $t_i = i\triangle$, we have

• (A1'): $\theta(.) = \sum_{i=0}^{N} \theta_i \mathbb{1}_{[t_i,\, t_{i+1})}(.)$ we denote $\delta\theta^2 = \theta_{i+1}^2 - \theta_i^2$.

and if moreover there is at most one change time in each window we get $(A3)$.

• (A3): (A1) and (A1') are satisfied and $\inf_{\rho=0,\dots,f} |t_{\rho+1} - t_\rho| \geq A\triangle$.

Remark 5.3.1 *If $\theta(t)$ satisfies a S.D.E. then* (A2) *is fulfilled, see e.g [A. Revuz and M.Yor, (1991)].*

We need to control $\int_{t_i}^{t_{i+1}} b^4(s,\, X_s)ds$, so we will use:

$(B1)$ $\exists K_T > 0, \quad \forall t \in [0,\, T], \quad \mathbb{E}(b(t,\, X_t)^4) \leq K_T$ *In all the sequel we work on the simplified model:*

$$dX_t = b_t(t,\, X_t)dt + \theta(t)dW_t.$$

Under some natural assumptions, the model (2) *becomes* (3) *after the following change of variable:*

Proposition 5.3.1 *(Pierre Bertrand) Assume that there exists a domain $D \subseteq \mathbb{R}$ such that $h \in \mathcal{C}(D, \mathbb{R}_+ - \{0\})$ the space of continuous function from D to $\mathbb{R}_+ - \{0\}$, $h^{-1} \in L^1_{loc}(D)$ and for (X_t) solution of* (2) *satisfying* $\mathbb{P}(X_t \in D, \forall\, t \in [0,T]) = 1$.

Let $H(x) =\in h^{-1}(\xi)d\xi$. *Then* $Y_t = H(X_t)$ *satisfies the S.D.E* (3) *with* $b_1(t, x) = h^{-1}(x)a(t, x) - \frac{1}{2}h^{'}(x)\theta^2(t)$.

5.4 Estimation of the Volatility using Haar wavelets basis

Since the size of the window appears in numerical applications as a free parameter to be arbitrarily chosen, we give a description of the Estimator introduced by Pierre Bertrand

$$H_{A,\triangle}(t) = \sum_{k=1}^{N/A-1} \left\{ A^{-1} \sum_{k=1}^{A-1} (X_{t_{kA+i+1}} - X_{t_{kA+i}})^2 \right\} 1\!\!1_{[t_{kA};t_{(k+1)A})}(t). \tag{5.5}$$

5.5 SDE in Dirichlet random environment

More generally, consider the following model. During the observation time interval $[0, T]$ the process X_t, evolves according to various regimes. Regime R_j holds during a random time interval $[T_{j-1}, T_j)$ where

$$0 = T_0 < T_1 < T_2 < \ldots < T_L = T.$$

The drift and the variance are randomly chosen in each regime but they do not change during this regime, so

$$dX_t = \sum_{j=1}^{L} \mu_{R_j} 1_{[T_{j-1}, T_j)}(t)dt + \sum_{j=1}^{L} \sigma_{R_j} 1_{[T_{j-1}, T_j)}(t)dB_t$$

where the R_j's $\in \{1, \ldots, N\}$ are random positive integers such that

$$\begin{cases} R_j \mid_p \overset{iid}{\sim} \sum_{k=1}^{N} p_k \delta_k(.) \\ (\mu_k,\, \sigma_k) \mid \theta \sim \mathcal{N}(\theta,\, \sigma_\mu) \otimes \Gamma(\eta_1,\, \eta_2), \quad k = 1, \ldots,\, L \\ p \mid \alpha \sim SB(\alpha,\, N) \\ \alpha \sim \Gamma(\nu_1,\, \nu_2) \\ \theta \sim \mathcal{N}(0,\, A). \end{cases}$$

5.5.1 Estimation and empirical results

The above process (X_t) is observed at discrete times, say idt, $i = 0, 1, 2, \ldots, n$. It is also assumed that the regime changes occur at these times. The estimation of the above parameters can be done through Ishwaran and James Blocked Gibbs algorithm where their class label variable K is our regime R.

$$\begin{cases} \Delta X_i \mid_{R,\mu,\sigma} \overset{ind}{\sim} \mathcal{N}(\mu_{R_i}, \sigma_{R_i}) \\ R_i \mid_p \overset{iid}{\sim} \sum_{k=1}^{N} p_k \delta_k(.) \\ \mu_i \mid_\theta \sim \mathcal{N}(\theta, \sigma_\mu) \\ \sigma_i \sim \Gamma(\eta_1, \eta_2) \\ p \mid_\alpha \sim SB(\alpha, N) \\ \alpha \sim \Gamma(\nu_1, \nu_2) \\ \theta \sim \mathcal{N}(0, A). \end{cases}$$

Our method was tested on the index of the Indian stock exchange market (www.nseindia.com), where the number of data is n=300. We have found 3 regimes:

	Regime 1	Regime 2	Regime 3
μ	4635.765	4924.502	5348.373
σ^2	59579	12879.15	19773.46
Probability	0.38	0.44	0.17

The analysis based on 25000 iterations following an initial 2000 iteration burn-in.

5.5.2 Option pricing in a regime switching market

The above setting can be used in the option pricing problem with $X_t = \log(S_t)$ where $(S_t)_{t\geq 0}$ is the stock price process governed by a geometric Brownian motion, and σ_{R_i} is a stochastic volatility during regime R_i. Observe that the estimations are done here without using any sliding windows technique and without assuming that $T_j - T_{j-1}$ is exponentially distributed, as it is done with Markov chains in regime switching markets.

Definition 5.5.1 *Suppose X is an $n \times p$ matrix, each row of which is independently drawn from p-variate normal distribution with zero mean:*

$$X_{(i)} = (x_i^1, \dots, x_i^p)^T \sim \mathcal{N}_p(0, V).$$

Then the Wishart distribution is the probability distribution of the $p \times p$ random matrix

$$W = XX^T = \sum_{i=1}^{n} X_{(i)} X_{(i)}^T.$$

One indicates that W has that probability distribution by writing

$$W \sim \mathcal{W}(n, V).$$

The positive integer n is the number of degrees of freedom.

5.6 Classification of trajectories

We consider the problem of classifying a set of n functions representing signals, stock prices and so on. Each function is known through a finite dimensional vector of observed points. In order to classify these functions, we now extend the blocked Gibbs algorithm to vector data. First let us precise our model.

5.6.1 Hierarchical Dirichlet Model for vector data

In the finite d-dimensional normal mixture problem, we observe data $f = (f_1,\, f_2, \ldots,\, f_n)$, where f_i are iid random curves with finite Wiener mixture density, the curves f_i can be represented and approximated by the vector $\tilde{f}_i = (\triangle_1 f_i,\, \triangle_2 f_i, \ldots,\, \triangle_L f_i)$

$$\psi_P(f) = \int_{\mathbb{R}\times\mathbb{R}^+} \phi(f\mid_{\sigma(y)})dP(y) = \Sigma_{k=1}^{d} p_{k,0}\phi(f\mid_{\sigma_k}) \tag{5.6}$$

where $\phi(f\mid_\sigma)$ represents a d-dimensional normal distribution with mean 0 and variance matrix σ.

Based on the data , we would like to estimate the unknown mixture distribution P. We can devise a Gibbs sampling scheme for exploring the posterior $\mathcal{P}_N \mid f$.

Notice that the model derived from (5) also contains hidden variables $K = \{K_1, \ldots,\, K_m\}$ since it can also be expressed as

$$\begin{cases} \tilde{f}_i \mid K,\, W,\, \mu \overset{iid}{\sim} \mathcal{N}_L(\mu_{K_i},\, \triangle t_i W_{K_i}) \\ K_i \mid p \sim \sum_{k=1}^{N} p_k \delta_k(.) \\ \mu_k \mid \theta \sim \mathcal{N}_L(\theta,\, \sigma_\mu) \\ W_k \sim \mathcal{W}(s,\, V) \\ \theta \sim \mathcal{N}_k(0,\, A) \end{cases} \tag{5.7}$$

where $\mathcal{W}(s,\, V)$ and $\mathcal{N}_L(\mu,\, \sigma)$ denote a Wishart and a multivariate Gaussian distribution respectively, and $p \sim SB(c,\, N)$.

Note that a similar model for vector data appear in Caron F. *et al.* (2006) but in our case the parameters of the Whishart prior are updated at each iteration. In addition, we have a problem of clustering which justifies the use of the hidden variables K_i's. In particular we will need to compute the posterior distribution of the class variable K and of the weight variable p. To implement the blocked Gibbs sampler we iteratively draw values from the following conditional distributions:

$$\mu \mid K,\, W,\, \theta,\, f$$

$$W \mid K,\, \mu,\, K,\, f$$

$$K \mid p,\, \sigma,\, Z,\, f$$

$$p \mid K,\, \alpha$$

$$\alpha \mid p$$

$$\theta \mid \mu.$$

5.6.2 Posterior computations

Blocked Gibbs Algorithm for vector data .

Let $\{K_1^\star, \ldots, K_m^\star\}$ denote the current m unique values of K. In each iteration of the Gibbs sampler we simulate:

(a) Conditional for μ: For each $j \in \{K_1^\star, \ldots, K_m^\star\}$, draw

$$\mu_j \mid W,\, K,\, \theta,\, f \overset{ind}{\sim} \mathcal{N}_l(\mu_j^\star,\, W_j^\star)$$

where $\mu_j^\star = \sum_{\{i:K_i=j\}} \tilde{f}_i + \theta$ and $W_j^\star = \sigma_\mu$, also for each $j \in K - K^\star$, independently simulate $\mu_j \sim \mathcal{N}_l(\theta,\, \sigma_\mu)$.

(b) Conditional for W: For each $j \in \{K_1^\star, \ldots, K_m^\star\}$, draw

$$W_j \mid \mu,\, K,\, f \overset{ind}{\sim} \mathcal{W}(s, \sum_{\{i:K_i=j\}} (\tilde{f}_i - \mu_j)(\tilde{f}_i - \mu_j)^T + V)$$

where $\mathcal{W}(V,\, p)$ denote the Wishart distribution with parameters V and p.

(c) Conditional for K:

$$K_i \mid p,\, \mu,\, W,\, f \overset{iid}{\sim} \sum_{h=1}^{N} p_{h,\,i}\delta_h(.), \qquad i = 1, \ldots, l$$

where for each $h = 1, 2, \dots, N$

$$p_{h,i} \propto p_h\Big(\frac{1}{(2\pi)^{l/2}\Big(det(W_h)\Big)^{1/2}}\Big)^{n_h} \exp\Big\langle \sum_{\{d,\, K_d^\star = h\}} (\tilde{f}_d - \mu_h)(\tilde{f}_d - \mu_h)^T,\, W_h\Big\rangle,$$

and $< A, B >$ is the trace of AB.

(d) Conditional for p:

For any integer N, let $V_1, \dots, V_{N-1}$ be iid $\beta(1, c)$ and $V_N = 1$. Let $p_1 = V_1^\star,\ p_k = (1 - V_1^\star)\dots(1 - V_{k-1}^\star)V_k^\star,\ k = 2, \dots, N$
where

$$V_k^\star = \beta\Big(1 + r_k,\, \alpha + \sum_{l=k+1}^{N} r_l\Big), \quad for \quad k = 1, \dots, N-1$$

and (as before) r_k records the number of K_i values which equal k.

(e) Conditional for α:

$$\alpha \mid p \sim \Gamma\Big(N + \eta_1 - 1,\, \eta_2 - \sum_{k=1}^{N-1} log(1 - V_k^\star)\Big),$$

for the same values of $V_k^\star$ used in the simulation for p.

(f) Conditional for θ:

$$\theta \mid \mu \sim \mathcal{N}_L(\theta^\star,\, \sigma^\star),$$

where

$$\theta^\star = \sum_{k=1}^{N} \mu_k \quad and \quad \sigma^\star = A.$$

Proof

Let ϕ denote the distribution function, for every $j \in \{K_1^\star, \dots, K_m^\star\}$

(a) Conditional for μ:

$$\begin{aligned}
\phi_{\mu_j|W,K,\theta,f}(y) &= \phi_{f|\mu_j=y,W,K,\theta}(y)\phi_{\mu_j|W,K,\theta}(y)\phi_{W,K,\theta} \\
&= \textstyle\prod_{\{d,\,K_d^\star=j\}} \phi_{\tilde{f}_d|\mu_j=y,\,W,\,K,\,\theta}(y)\phi_{\mu_j|W,\,K,\,\theta}(y)\phi_{W,\,K,\,\theta} \\
&= \left(\textstyle\prod_{\{d,\,K_d^\star=j\}} e^{iy^T\tilde{f}_d} e^{-\frac{1}{2}\tilde{f}_d{}^T W_j \tilde{f}_d}\right) e^{iy^T\theta-\frac{1}{2}y^T\sigma_\mu y} \\
&= e^{iy^T\sum_{\{d,\,K_d^\star=j\}}\tilde{f}_d} e^{-\frac{1}{2}\sum_{\{d,\,K_d^\star=s\}}(\tilde{f}_d{}^T W_j \tilde{f}_d)} e^{iy^T\theta-\frac{1}{2}y^T\sigma_\mu y} \\
&= \left(e^{-\frac{1}{2}\sum_{\{d,\,K_d^\star=s\}}(\tilde{f}_d{}^T W_j \tilde{f}_d)}\right) e^{iy^T(\theta+\sum_{\{d,\,K_d^\star=j\}}\tilde{f}_d)-\frac{1}{2}y^T\sigma_\mu y}
\end{aligned}$$

hence

$$\mu_j \mid W,\, K,\, \theta,\, f \overset{ind}{\propto} \mathcal{N}_l(\theta + \sum_{\{d,\,K_d^\star=j\}} \tilde{f}_d,\, \sigma_\mu)$$

(b) Conditional for W: For each $j \in \{K_1^\star, \ldots, K_m^\star\}$

$$\begin{aligned}
\phi_{W_j^{-1}|\mu, K, f}(M) &= \phi_{X|W_j=M, K}(M)\phi_{W_j^{-1}|K, \mu}(M)\phi_{\mu, K}(z, t) \\
&= \Big(\prod_{\{d, K_d^\star=j\}} e^{-\frac{1}{2}(\tilde{f}_d-\mu_j)^T M(\tilde{f}_d-\mu_j)}\Big) \\
&\times \frac{det(M^{\frac{n-l-1}{2}})^{\frac{n-l-1}{2}}}{2^{\frac{nl}{2}} det(V)^{\frac{n}{2}} \Gamma_p(\frac{n}{2})} e^{-\frac{1}{2}Tr(V^{-1}M)} \phi_{\mu, K}(z, t) \\
&= e^{-\frac{1}{2}Tr\Big(\sum_{\{d, K_d^\star=j\}}(\tilde{f}_d-\mu_j)(\tilde{f}_d-\mu_j)^T M\Big)} \\
&\times \frac{det(M^{\frac{n-l-1}{2}})^{\frac{n-l-1}{2}}}{2^{\frac{nl}{2}} det(V)^{\frac{n}{2}} \Gamma_p(\frac{n}{2})} e^{-\frac{1}{2}Tr(V^{-1}M)} \phi_{\mu,K}(z, t) \\
&= \frac{det(M^{\frac{n-l-1}{2}})^{\frac{n-l-1}{2}}}{2^{\frac{nl}{2}} det(V)^{\frac{n}{2}} \Gamma_p(\frac{n}{2})} e^{-\frac{1}{2}Tr\Big((\sum_{\{d, K_d^\star=j\}}(\tilde{f}_d-\mu_j)(\tilde{f}_d-\mu_j)^T+V^{-1})M\Big)} \\
&\times \phi_{\mu, K}(z, t)
\end{aligned}$$

therefore,

$$W_j \mid \mu, K, f \overset{ind}{\propto} W\Big(n, (\sum_{\{i: K_i=j\}} (\tilde{f}_i - \mu_j)(\tilde{f}_i - \mu_j)^T + V)^{-1}\Big).$$

(c) Conditional for K:

$$\begin{aligned}
P\{K_i = j \mid p, \mu, W, f\} &= P\{f \mid p, W, K_i = j, \mu\}P\{K_i = s \mid W, \mu\}P\{\mu\}P\{W\} \\
&\propto P\{f \mid p, W, K_i = j, \mu\}P\{K_i = s \mid W, \mu\} \\
&= \Big(\prod_{\{d, K_d^\star=s\}} \frac{p_s}{(2\pi)^{l/2}\Big(det(W_s)\Big)^{1/2}} e^{-\frac{1}{2}(\tilde{f}_d-\mu_s)^T W_s(\tilde{f}_d-\mu_s)}\Big).
\end{aligned}$$

Hence,

$$p_{s, i} \propto p_s\Big(\frac{1}{(2\pi)^{l/2}\Big(det(W_s)\Big)^{1/2}}\Big)^{n_s} \exp\Big\langle \sum_{\{d, K_d^\star=s\}} (\tilde{f}_d - \mu_s)(\tilde{f}_d - \mu_s)^T, W_s\Big\rangle$$

where n_s is the number of time $K_s^\star$ occurs in K.

(d) Conditional for θ:

$$\begin{aligned}
\phi_{\theta|\mu=\mu'}(\theta) &\propto \phi_{\mu|\theta}(\mu')\phi_\theta(\theta) \\
&= \prod_{j=1}^N \phi_{\mu|\theta}(\mu'_j)\phi_\theta(\theta) \\
&= \left(\prod_{j=1}^N e^{i\theta^T\mu'_j} e^{-\frac{1}{2}\mu_j'^T\sigma_\mu\mu'_j}\right) e - \frac{1}{2}\theta^T A\theta \\
&= \left(e^{i\sum_{j=1}^N \theta^T\mu'_j} e^{-\frac{1}{2}\theta^T A\theta}\right) e^{\sum_{j=1}^N -\frac{1}{2}\mu_j'^T\sigma_\mu\mu'_j}.
\end{aligned}$$

Hence the distribution of $\theta \mid \mu \propto \mathcal{N}_L(\sum_{j=1}^N \mu_j,\, A)$.

5.6.3 Classes of volatility

Let (S_t) be the stock price process and suppose that $X_t = \log(S_t)$, satisfies:

$$dX_t = b(t,\, X_t)dt + \theta(t)h(X_t)dB_t \tag{5.8}$$

where the function $h(.)$ is assumed to be known, the volatility coefficient $\theta(.)$ is a random function of time and has to be estimated and the drift coefficient $b(t,\, x)$ is unknown. We observe a path of the process $(X_t,\, t \in [0,\, T])$ sampled at discrete times $t_i = i\triangle$, for $i = 1, \ldots, N$.

Under some conditions and after a change of variable (see e.g. [5]), equation (5.8) reduces to

$$dX_t = b_t(t,\, X_t)dt + \theta(t)dB_t.$$

A refined method to estimate $\theta(t)$ consists in using wavelets. Consider $(V_j,\, j \in \mathbb{Z})$ an r-regular Multi Resolution Analysis of $L^2(\mathbb{R})$ such that the associated scale function Φ and the wavelet function ψ are compactly supported. For all j, the family $\{\Phi_{j,\,k}(t) = 2^{j/2}\Phi(2^j t - k), k \in \mathbb{Z}\}$ is an orthogonal basis of V_j. Time being sampled with $\triangle = 2^{-n}$, S_t, the estimator is then:

$$\theta^2(t) = \sum_k \mu_{j(n),\,k}\Phi_{j(n),\,k}(t) \tag{5.9}$$

for $j(n) < n$, where

$$\mu_{j(n),k} = \sum_{i=1}^{N-1} \Phi_{j(n),k}(t_i)(X_{t_{i+1}} - X_{t_i})^2. \tag{5.10}$$

Suppose that we have observed n trajectories $X_1, \ldots, X_l, \ldots, X_n$ sampled as above, and that we want to classify them according to their volatility component, that is, we want to classify the θ_l's estimated by (5.9).
We then see that we have just to apply the preceding algorithm to the vectors $\mu^l_{j(n),k}$ which are finite dimensional representations of the θ_l's.

5.7 Conclusion

We have extended Dirichlet hierarchical models in order to deal with temporal data such as solutions of SDE with stochastic drift and volatility. It can be thought that the process on which are based these parameters belongs to a certain well-known class of processes, such as continuous time Markov chains. Then, we think that a Dirichlet prior can be put on the path space, that is a functional space. The estimation procedure in such a context is the topic the next chapter.

Chapter 6

Markov regime switching with Dirichlet Prior. Application to Modelling Stock Prices

We have seen in Chapter 3, some examples of continuous time Dirichlet processes with parameters proportional to the distribution of continuous time processes, such as the Wiener measure one.
In the present Chapter, motivated by some mathematical models in finance dealing with 'Regime switching markets', we consider the case where the continuous time process is a continuous time Markov chain whose state at time t modellizes the state of the market at time t.
Indeed, while in preceding Chapter 5, volatility was constant during some time interval of random length without any hypothesis on the switching process, here the switching depends on a Markov chain which states represent the different regimes. Also, the various values of the trend and the volatility depend on the state of this chain which 'chooses' these values among some i.i.d. ones. Clearly, we deal with *stochastic volatility*
In our approach, the regimes play the same role as the classes play in classi-

fication: each temporal observation therefore belongs to a class that is to a regime.

Our contribution consists in placing a Dirichlet process prior on the path space of the Markov chain, which is a cadlag function space. This idea is new as it has never been used in the literature.

In the first Section, we present our model. Section 2 deals with the estimation procedure, the computations of the posteriors follow from those done in Chapter 5. In the last Section 3, we give some indications on the implementation of the algorithm in C language and some numerical results are presented.

6.1 Markov regime switching with Dirichlet prior

In this section , we take $\bar{\alpha} = H$, the distribution of a continuous time Markov chain on a finite set of states and we propose a new hierarchical model that is specified, as an example, in the setting of mathematical finance. Of course, this can be similarly used in many other cases. We consider the Black-Scholes SDE in random environment with a Dirichlet prior on the path space of the chain, the states of the chain representing the environment due to the market. We model the stock price using a geometric Brownian motion with drift and variance depending on the state of the market. The state of the market is modeled as a continuous time Markov chain with a Dirichlet prior. In what follows, the notation σ will be used to denote the variance rather than the standard deviations.

The following notations will be adopted:

1. n will denote the number of observed data and also the length of an observed path.

2. M will denote the number of states of the Markov chain.

3. The state space of the chain will be denoted by $S = \{i : 1 \le i \le M\}$.

4. N will denote the number of simulated paths.

5. m will denote the number of distinct states of a path.

- The stock price follows the following SDE:

$$\frac{dS_t}{S_t} = \beta(X_t)dt + \sqrt{\sigma(X_t)}dB_t, \qquad t \ge 0,$$

  where B_t is a standard Brownian motion. By the Ito's formula, the process $Z_t = \log(S_t)$ satisfies the SDE,

$$dZ_t = \mu(X_t)dt + \sqrt{\sigma(X_t)}dB_t, \qquad t \ge 0,$$

  where $\mu(X_t) = \beta(X_t) - \frac{1}{2}\sigma(X_t)$. The observed data is of the form $Z_0, Z_1, \ldots, Z_n$.

- The process (X_t) is assumed to be a continuous time Markov process taking values in the set $S = \{i : 1 \le i \le M\}$. The transition probabilities of this chain are denoted by p_{ij}, i, $j \in S$ and the transition rate matrix is $Q_0 = (q_{ij})_{i,j \in S}$ with

$$\lambda_i > 0, \qquad q_{ij} = \lambda_i p_{ij} \quad \text{if} \quad i \ne j, \qquad \text{and} \qquad q_{ii} = -\sum_{j \ne i} q_{ij}, \qquad i, j \in S.$$

  Define the log-returns, $Y_t = Z_t - Z_{t-1} = \log(S_t/S_{t-1})$, $t = 1, 2, \ldots, n$. Suppose we know the path $X = \{X_s, 0 \le s \le n\}$. Let $T_j(t)$ be the time spent by the path X in state j in the time interval $[t-1, t]$. Define

$$\mu(t) := \sum_{j=1}^{M} \mu(j) T_j(t); \qquad \sigma(t) := \sum_{j=1}^{M} \sigma(j) T_j(t). \tag{6.1}$$

  Then, conditional on the path X, Y_t are i.i.d. $\mathcal{N}(\mu_t, \sigma_t)$, $t = 1, 2, \ldots, n$.

- For each $i = 1, 2, \ldots, M$, the priors on $\mu_i = \mu(i)$ and $\sigma_i = \sigma(i)$ are specified by

$$\mu_i \overset{ind}{\sim} \mathcal{N}(\theta, \tau^{\mu}), \qquad \text{with} \qquad \theta \sim \mathcal{N}(0, A), \qquad A > 0, \tag{6.2}$$

$$\sigma_i \overset{ind}{\sim} \Gamma(\nu_1, \nu_2). \tag{6.3}$$

- The Markov chain $\{X_t, t \geq 0\}$ has prior $\mathcal{D}(\alpha\ H)$, where H is a probability measure on the path space of cadlag functions $D([0, \infty), S)$. The initial distribution according to H is the uniform distribution $\pi_0 = (1/M, \ldots, 1/M)$, and the transition rate matrix is Q with $p_{ij} = 1/(M-1)$ and $\lambda_i = \lambda > 0$. Thus the Markov chain under Q will spend an exponential time with mean $1/\lambda$ in any state i and then jump to state $j \neq i$ with probability $1/(M-1)$.

  A realization of the Markov chain from the above prior is generated as follows: Generate a large number of paths $X_i = \{x_s^i : 0 \leq s \leq n\}$, $i = 1, 2, \ldots, N$, from H. Generate the vector of probabilities $(p_i, i = 1, \ldots, N)$ from a Poisson Dirichlet distribution with parameter α, using stick breaking. Then draw a realization of the Markov chain from

$$p = \sum_{i=1}^{N} p_i \delta_{X_i}, \tag{6.4}$$

  which is a probability measure on the path space $D([0, n), S)$. The parameter λ is chosen to be small so that the variance is large and hence we obtain a large variety of paths to sample from at a later stage. The prior for α is given by,

$$\alpha \sim \Gamma(\eta_1, \eta_2). \tag{6.5}$$

6.2 Estimation

Estimation is done using the simulation of a large number of paths of the Markov chain which will be selected according to a probability vector (generated by stick-breaking) and then using the blocked Gibbs sampling technique. This technique uses the posterior distribution of the various parameters.

We denote by μ, and σ, the current values of the vectors $(\mu_1, \mu_2, \ldots, \mu_n)$, $(\sigma_1, \sigma_2, \ldots, \sigma_n)$, respectively. Let Y be the vector of observed data $(Y_1, \ldots, Y_n)$. Let $X = (x_1, x_2, \ldots, x_n)$ be the vector of current values of the states of the Markov chain at times $t = 1, 2, \ldots, n$, respectively. Let $X^* = (x_1^*, \ldots, x_m^*)$ be the distinct values in X.

6.2.1 Modifying the observed data set

In order to obtain the conditional distribution of the parameters, we first need to extract the change in the log-returns between the jump times of the Markov chain. Let $0 = t_0 < t_1 < t_2 < \ldots t_J$ be the times at which the path X changes state. Define the log-returns between the jump times, $W_k = \log(S_{t_k}/S_{t_{k-1}})$, $k = 1, 2, \ldots, J$. To obtain realizations of the W_k from the observed Y process, we need to simulate Gaussian random variables conditioned on their sums.

Consider any $t \in \{0, 1, \ldots, n\}$ for which the chain changes state atleast once in the time interval $[t-1, t]$. Let $t_{k-1} < t-1 \leq t_k < \ldots < t_{k+p} < t < t_{k+p+1}$, be the jump times that lie in $[t-1, t]$, for some $p \geq 1$. Let $V_t^1 = \log(S_{t_k}/S_{t-1})$ and $V_t^2 = \log(S_t/S_{t_{k+p}})$. Then,

$$Y_t = V_t^1 + \sum_{i=1}^{p} W_{k+i} + V_t^2. \tag{6.6}$$

Suppose for some the chain X is in state j_i in the time interval $[t_{k+i-1}, t_{k+i})$, $i = 0, 1, \ldots, p+1$. Set $s_0 = t_k - t - 1$, $s_i = t_{k+i} - t_{k+i-1}$, $i = 1, 2, \ldots, p$, and $s_{p+1} = t - t_{k+p}$. Let $m_j = \mu(j_i)s_i$ and $v_j = \sigma(j_i)s_i$, $i = 0, 1, \ldots, p+1$. Recall that $Y_t \sim \mathcal{N}(\mu_t, \sigma_t)$, where $\mu(t), \sigma(t)$ are as defined in (6.1). It is easy to see that the joint conditional density of $(V_t^1, W_{k+1}, \ldots, W_{k+p})$ given $Y_t = y$

$$f(u_0, u_1, \ldots, u_p) = C \prod_{i=0}^{p} \exp\left(-\frac{1}{2} \frac{v_i + v_{p+1}}{v_i \; v_{p+1}} \left(u_i - \frac{v_{p+1} m_i + v_i (y - m_{p+1})}{v_i + v_{p+1}} \right)^2 \right), \tag{6.7}$$

where C is a constant that depends on y and the parameters. Thus, one can simulate the variables $V_t^1, W_k, W_{k+1}, \ldots, W_{k+p}$ from independent Gaussians and then obtain V_t^2 using (6.6).

Using the above procedure, we can obtain a realization for all W_k for which $[t_{k-1}, t_k] \subseteq [t-1, t]$, for some $t \in \{0, 1, \ldots, n\}$. Now for any k for which there is a $q \geq 0$, such that $t-1 \leq t_{k-1} < t < t+1 < \ldots < t+q \leq t_k < t+q+1$, we can obtain W_k using the relation

$$W_k = V_t^2 + \sum_{i=1}^{q} Y_{t+i} + V_{t+q+1}^1. \tag{6.8}$$

Note that the W values depend on the path X and need to be computed in each iteration.

6.2.2 The Gibbs sampling procedure

We are now ready to estimate the posterior distributions of the parameters using Gibbs sampling. Each iteration produces one realization of the parameters from their approximate posterior distribution. Each iteration consists of a large number of samples obtained recursively for each parameter conditioned on the current values of the other parameters and the data.

- **Conditional for** μ. For each $j \in X^*$ draw

$$(\mu_j | \theta, \tau^\mu, \sigma, X, W) \stackrel{ind}{\sim} \mathcal{N}(\mu_j^*, \sigma_j^*), \tag{6.9}$$

where

$$\mu_j^* = \sigma_j^* \left(\sum_{k: X_{t_{k-1}} = j} \frac{W_k}{\sigma_j (t_k - t_{k-1})} + \frac{\theta}{\tau^\mu} \right),$$

$$\sigma_j^* = \left(\frac{n_j}{\sigma_j} + \frac{1}{\tau^\mu}\right)^{-1},$$

and n_j being the number of times j occurs in X. For each $j \in X \setminus X^*$, independently simulate $\mu_j \sim \mathcal{N}(\theta, \tau^\mu)$.

- **Conditional for** σ. For each $j \in X^*$ draw

$$(\sigma_j|\mu, \nu, X, W) \overset{ind}{\sim} \Gamma(\nu_1 + \frac{n_j}{2}, \nu_{2,j}^*), \tag{6.10}$$

where

$$\nu_{2,j}^* = \nu_{2,j} + \sum_{k:X_{t_{k-1}}=j} \frac{(W_k - \mu_j(t_k - t_{k-1}))^2}{2(t_k - t_{k-1})}.$$

Also for each $j \in X \setminus X^*$, independently simulate $\sigma_j \sim \Gamma(\nu_1, \nu_2)$.

- **Conditional for** X.

$$(X|p) \sim \sum_{i=1}^N p_i^* \delta_{X_i}, \tag{6.11}$$

where

$$p_i^* \propto \prod_{j=1}^m \Big(\prod_{\{k: x^{i,*}_{t^i_{k-1}}=j\}} \frac{1}{(2\pi\sigma_j(t_k - t_{k-1}))^{1/2}} e^{-\frac{1}{2\sigma_j}(W_k^i - \mu_j(t_k - t_{k-1}))^2} \Big) \, p_i, \tag{6.12}$$

where $(x_1^{i,*}, \ldots, x_m^{i,*})$ denote the current $m = m(i)$ unique values of the states and t_k^i, W_k^i are as defined in subsection 6.2.1 for the path X_i, $i = 1, \ldots, N$.

- **Conditional for** p.

$$p_1 = V_1^*, \text{ and } p_k = (1 - V_1^*) \cdots (1 - V_{k-1}^*) V_k^*, \quad k = 2, 3, \ldots, N-1, \tag{6.13}$$

where

$$V_k^* \overset{ind}{\sim} \beta\Big(1 + r_k, \alpha\Big),$$

r_k equal 1 if $i = k$ and 0 else.

- **Conditional for α.**

$$(\alpha|p) \sim \Gamma\left(N + \eta_1 - 1,\, \eta_2 - \sum_{i=1}^{N-1} \log(1 - V_i^*)\right),$$

where the V^* values are those obtained in the simulation of p in the above step.

- **Conditional for θ.**

$$(\theta|\mu) \sim \mathcal{N}(\theta^*,\, \tau^*), \tag{6.14}$$

where

$$\theta^* = \frac{\tau^*}{\tau^\mu} \sum_{j=1}^{M} \mu_j,$$

and

$$\tau^* = \left(\frac{M}{\tau^\mu} + \frac{1}{A}\right)^{-1}.$$

Proof.

(a) The computation of the posterior distributions for μ, σ and θ follow in the same manner as in Ishwaran and James (2002) and Ishwaran and Zarepour (2000). Here, $X_t = s$ means that the class variable is equal to s.

(b) Conditional for X:

$$P\{X = X_i \mid p,\, \mu,\, \sigma,\, W\} = P\{W \mid p,\, \sigma,\, X = X_i,\, \mu\}P\{X = X_i \mid \sigma,\, \mu,\, p\}P\{\mu,\, \sigma\}$$

$$\propto \prod_{j=1}^{m}\Bigg(\prod_{\{k:\, x^{i,*}_{t^i_{k-1}} = j\}} \frac{1}{(2\pi\sigma_j(t_k - t_{k-1}))^{1/2}} e^{-\frac{1}{2\sigma_j}(W_k^i - \mu_j(t_k - t_{k-1}))^2}\Bigg)\, p_i$$

where $X_i = (x_1^i, \ldots, x_n^i)$ and $(x_1^{i,*}, \ldots, x_m^{i,*})$ denote the current m unique values in the path X_i.

(c) Conditional for p : The Sethuraman stick-breaking scheme can be extended to the two-parameter Beta distributions, see Ishwaran James (2001) and Walker Muliere (1997, 1998):

Let $V_k \stackrel{ind}{\sim} \beta(a_k, b_k)$, for each $k = 1, \ldots, N$. Let

$$p_1 = V_1, \text{ and } p_k = (1 - V_1) \cdots (1 - V_{k-1}) V_k, \quad k = 2, 3, \ldots, N - 1.$$

We will write the above random vector, in short as

$$p \sim SB(a_1, b_2, \ldots, a_{N-1}, b_{N-1}).$$

By Connor and Mosimann (1969), the density of p is

$$\Big(\prod_{k=1}^{N-1} \frac{\Gamma(a_k - b_k)}{\Gamma(a_k)\Gamma(b_k)} \Big) p_1^{a_1 - 1} \ldots p_{N-1}^{a_{N-1}-1} p_N^{b_{N-1}-1} \times$$

$$\times (1 - P_1)^{b_1 - (a_2 - b_2)} \ldots (1 - P_{N-2})^{b_{N-2} - (a_{N-1} - b_{N-1})},$$

where $P_k = p_1 + \ldots + p_k$.

From this, it easily follows that the distribution is conjugate for multinomial sampling, and consequently the posterior distribution of p given X, when $a_k = 1$ and $b_k = \alpha$ for each k, is

$$SB(a_1^*, b_2^*, \ldots, a_{N-1}^*, b_{N-1}^*),$$

where

$$\begin{aligned} b_k^* &= \alpha \\ a_k^* &= 1 + r_k, \end{aligned}$$

and r_k equal 1 if $i = k$ and 0 else, k=1,..., N-1. □

6.3 Implementation

The algorithm presented in the previous section was implemented in C language. The implementation includes:
- functions that simulate standard probability distributions: Uniform, Normal, Gamma, Beta, Exponential.
- a function that returns an index $\in \{1, \ldots, n\}$ according to a vector of probability $p_1, \ldots, p_n$.
- a function that simulates a probability vector according to stick-breaking scheme.
- a function that simulates n paths of a Markov chain.
- a function that records the number of times a state appears in a path.
- a function that chooses one of the paths according to a vector of probability.
- a function that modifies the parameters of prior distributions according to the formulas of the posteriori distributions.

After having simulated a number of paths, we perform the iterations. At each iteration a path is randomly selected and the parameters are updated according to posteriori formulas. At the end of each iteration of the Gibbs sampling, we obtain a path X of the Markov chain. From this, the parameters π and Q_0 can be re-estimated. From Q_0 the parameters λ_i and p_{ij} can be derived.

6.3.1 Simulated data

We fit the model, using the algorithm developed above, to a simulated series of lenght $n = 480$, with a number of states (regimes) $M = 4$, mean and variance in each state being chosen as follows:

$$\begin{aligned}(\mu_1, \sigma_1) &= (-1.15, 0.450)\\ (\mu_2, \sigma_2) &= (-0.93, 0.450)\\ (\mu_3, \sigma_3) &= (-0.60, 0.440)\\ (\mu_4, \sigma_4) &= (1.40, 0.500).\end{aligned}$$

We have performed our algorithm on that series with number of states $M = 10$, number of paths $N = 100$ and number of iterations = 25,000. Then, we have observed that the algorithm is able to put most of the mass (in terms of the stationary distribution of the MC) on 4 regimes, which are close to the ones chosen above. At the end of the iterations we compute a confidence interval for the mean and for the variance w.r.t. each regime. We can conclude that the algorithm is able to identify the parameters of the simulated data set.

The confidence intervals for the mean and the variance are given below.

Regime 1:

$$I_m = [-1.208, -1.12423] \quad \text{and} \quad I_v = [0.431, 0.4738].$$

Regime 2:

$$I_m = [-0.9351, -0.9296] \quad \text{and} \quad I_v = [0.442, 0.4538].$$

Regime 3:

$$I_m = [-0.63446, -0.5140] \quad \text{and} \quad I_v = [0.4319, 0.4491].$$

Regime 4:

$$I_m = [1.30114, 1.43446] \quad \text{and} \quad I_v = [0.4949, 0.5081].$$

6.3.2 Real data

We have also applied our algorithm to the Bsemidcap index data of the Indian National Stock Exchange (NSE) from 21/12/2006 to 15/11/2007 (www.nseindia.com).

For this dataset we have, $n = 250$, $\Delta t = 1$, and we deal with $N = 100$ of paths while Gamma(2, 4) is the prior for α.
With the above choice, we obtain 6 regimes for which the estimates for the mean, variance and stationary probabilities are as follows:

	R 1	R 2	R 3	R 4	R 5	R 6
μ	0.001124	-0.009479	0.000629	-0.004579	0.000829	0.001109
σ	2.9132 e-05	7.2166 e-05	2.3023 e-05	7.3800 e-05	1.186 e-05	3.3372 e-05
π	20 %	3 %	29%	5 %	10 %	33 %

The most frequent Markov chain path, its parameters λ_is and the matrix of transition probability $(p_{i,j})_{1\leq i\neq j\leq 6}$ are respectively equal to:

3 5 3 6 3 6 3 6 1 6 5 1 3 6 3 5 3 3 6 6 5 6 3 6 1 1 4 1 6 1 3 3 6 6 6 3 1 3 3 3 6 3 3 3 4 5 6 6 6 6 4 6 1 1 1 6 6 6 6 6 1 3 3 3 1 6 1 3 3 5 6 3 3 1 6 5 4 1 3 6 4 6 3 3 5 6 3 6 2 3 6 1 3 3 6 1 6 6 5 5 1 1 5 3 5 3 3 6 1 6 5 6 1 6 6 3 1 6 3 1 1 6 2 3 6 6 6 3 3 2 6 6 6 1 3 3 6 6 3 1 3 6 6 1 6 6 1 1 6 1 5 3 5 1 3 5 3 4 1 3 3 5 3 1 3 6 6 6 1 3 5 6 5 3 3 6 3 6 1 3 5 6 6 6 5 1 6 3 3 1 1 6 6 6 3 6 1 3 6 3 6 6 6 6 6 3 6 3 6 6 4 6 3 6 1 1 6 4 6 1 3 4 3 6.

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
0.8	1	0.7	1	0.95	0.75

$$\begin{pmatrix} & 0 & 0.48 & 0.03 & 0.06 & 0.42 \\ 0 & & 0.66 & 0 & 0 & 0.33 \\ 0.16 & 0.02 & & 0.062 & 0.2 & 0.54 \\ 0.375 & 0 & 0 & & 0.125 & 0.5 \\ 0.157 & 0 & 0.42 & 0.052 & & 0.36 \\ 0.36 & 0.038 & 0.384 & 0.077 & 0.134 & \end{pmatrix}$$

It is interesting to note that in the high volatility states, the index has a negative drift as is usually observed in analysis of empirical data. A by-product of our algorithm is the distribution of the current state of the volatility, which is required to compute the price of an option (see [?] and references therein).

6.4 Validation.

Consider the stock price data for duration $1 \leq t \leq T_1$. Estimate the model based on this data. The carry out a 1-step forecast on the time interval $T_1 \leq t \leq T_2$ using the estimated model. Compare the MSE with other models like GBM with fixed variance, GARCH (Rene Carmona), simple Markov switched model etc.

6.5 Option Pricing

The model we follow is as in Ghosh and Deshpande (G-D), except that we now have a prior on the variables. So, essentially we have to take take several realizations of our parameters and for each of them compute the option price and then average over these values.

Suppose we have stock price for time $0 \leq t \leq T_1$, then use formula (4.1) in G-D to compute the option price with $s = S_{T_1}$ which is the current price and take $t = T$, to be say 15 (the day the option matures). This will give us the values $(\phi(T, S_{T_1}, (k,i), k = 1, \dots N, i = 1, \dots, M)$.

Note that the vector ϕ is written as

$$(\phi(t,s,(1,1)), \phi(t,s,(1,2)), \dots, \phi(t,s,(1,M)), \phi(t,s,(2,1)), \dots, \phi(t,s,(N,M)))$$

and the transition matrix and the other matrices accordingly. For example, the matrix Σ in (4.1) will be a block diagonal matrix with N blocks each of which is $(diag(\sigma_1, \dots, \sigma_M)$. Thus Σ will be a $NM \times NM$ matrix.

Once we solve (4.1), then, given the history of the price upto time T_1, we have to estimate the probability that the Markov chain is in state (k, i). Then we have to average the option price over these probabilities.

This option price should be compared with the usual Black-Scholes formula for GBM with fixed σ.

In this numerical work, we can keep the interest rate fixed. See some literature on option pricing for choice of the interest rate.

Chapter 7

Conclusion and Perspectives

Our main subject of interest was to investigate Dirichlet processes when the parameter is proportional to the distribution of a stochastic process (Brownian motion, jump processes, ...) and to propose continuous time hierarchical models involving continous-time Dirichlet processes.

Although this area requires some rather nontrivial techniques, we have shown that such a setting can be of interest in modelling SDEs in random environment and that the proposed estimation procedure works.

Let us finally mention some perspectives.

It is clear that it would be interesting to extend the method to other SDEs and to other kind of processes, we think of replacing, in the last chapter, the markov chain by a diffusion, a spatio-temporal process or a multivariate process.

It would be also of interest to use the estimated model for prediction and to compare this prediction with other models.

Concerning the algorithm in the last chapter it can be observed that for each iteration, an option price w.r.t. the selected path can be computed by using for example the formula in Ghosh and Deshpande. After performing all the iterations, we will have a distribution of option prices that can be used for decision-making on the final option price. This should be compared to other decision procedures.

Bibliography

[1] Altman, N. (1992).An introduction to kernel and nearest-neighbor nonparametric regression. *American Statistician* **46**, 175-185.

[2] Antoniak, C.E. (1974). Mixtures of Dirichlet processes. *Ann. Statist.* **2**, 6, 1152-1174.

[3] Behseta, S., Kass, R. E., and Wallstrom, G. L. (2005). Hierarchical models for assessing variability among functions. *Biometrika* **92**, 419-434.

[4] Bertoin, J (2006). Random fragmantation and coagulation processes.

[5] Bigelow, J. L. and Dunson, D. B. (1992). Posterior simulation across nonparametric models for functional clustering. *AJournal of the Royal Statistical Society, Series B.* , **under revision**.

[6] Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Polya urn schemes. *Ann. Statist.* **2**, 1, 353-355.

[7] Blei, D. and Jordan., I. J. (2005). Variational inference for Dirichlet process mixtures. *Bayesian Analysis, 1* 121-144.

[8] Blei, D., Ng, A. and Jordan, M. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research,* 3: 993-1022.

[9] Bertrand, P (1996). Estimation of the Stochastic Volatility of a Diffusion Process I. Comparison of Haar basis Estimator and Kernel Estimators. *INRIA*.

[10] Brunner, L.J. and Lo, A. Y. (2002). Bayesian classification To appear.

[11] , C. A. and MacEachern, S. N. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika* **83** 275-285.

[12] Chib, S. and Hamilton, B. H. (2002). Semiparametric Bayes analysis of longitudinal data treatment models. *Journal of Econometrics* **110** , 67-89.

[13] Chu, C.-K. and Marron, J. S. (1991). Choosing a kernel regression estimator. Statistical Science, *Statistical Science* **6** , 404-419.

[14] Cifarelli, D. M. and Melilli, E. (2000). Some new results for Dirichlet priors, *Journal of the American Statistical Association. Ann. Statist.* **28** 1390-1413.

[15] Cifarelli, D. and Regazzini, E. (1978). Problemi Statistici Non Parametrici in Condizioni di Scambiabilit‘ a Parziale e Impiego di Medie Associative. *Tech. rep., Quaderni Istituto Matematica Finanziaria dell-Universitàdi Torino.*

[16] Cifarelli, D. M. and Regazzini, E. (1990). Distribution functions of means of a Dirichlet process. *Ann. Statist.* **18** 429-442.

[17] Dahl, D. B. (2003). Modeling differential gene expression using a Dirichlet Process mixture model. *in Proceedings of the American Statistical Association, Bayesian Statistical Sciences Section. American Statistical Association, Alexandria, VA.*

[18] Deshpande, A. and Ghosh, M. K. (2007). Risk Minimizing Option Pricing in a Regime Switching Market. *Stochastic Analysis and Applications.* Vol. 28, 2008. To appear.

[19] DeIorio, M., Méuller, P., Rosner, G. L., and MacEachern, S. N. (2004). An anova model for dependent random measures. *Journal of the American Statistical Asso- ciation* **99** 205-215.

[20] DONNET, S., and SAMSON, A. (2005). Parametric Estimation for Diffusion Processes from Discrete-time and Noisy Observations. *Journal of the American Statistical Association*, 92, 894-902.

[21] Doss, H. and Sellke, T (1982). The tails of probabilities chosen from a Dirichlet prior. *Ann. Statist.* 10 1302-1305.

[22] Duan, J. A. (2006).Bayesian dynamic modeling of latent trait distributions. *Bio- statistics* **7** 551-568.

[23] Duan, J. A., Guindani, M., and Gelfand, A. E. (2007). Generalized spatial Dirichlet process models. *Biometrika* **in press.**

[24] Dunson, D. B., Herring, A. H., and Mulheri-Engel, S. A. (2007a). Bayesian selection and clustering of polymorphisms in functionally-related genes. *Journal of the Royal Statistical Society* , **in press**.

[25] Dunson, D. B. and Park, J.-H. (2007). Kernel stick-breaking processes. *Tech. rep., Institute of Statistics and Decision Sciences - Duke University.*

[26] Dunson, D. B., Pillai, N., and Park, J.-H. (2007). Bayesian density regression. *Journal of the Royal Statistical Society Series B.* **69** 163-183.

[27] Emilion, R. (2001). Classification and mixtures of processes. *SFC 2001 and C.R. Acad. Sci. Paris, série I* **335**, 189-193.

[28] Emilion, R. and Pasquignon, D. (2005). Random distributions in image analysis. *Preprint*.

[29] Emilion, R. (2005). Process of random distributions. *Afrika Stat*, vol 1, 1, pp. 27-46, http://www.ufrsat.org/jas (contenus).

[30] Escobar, M. D. (1994). Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association* **89** 268-277.

[31] Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90** 577-588.

[32] Fan, J. Q., Hickman, N. E., and Wand, M. P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *Journal of the American Statistical Association* **99** 141-150.

[33] Ferguson, T.S (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209-230.

[34] Ferguson, T.S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2** 615-629.

[35] Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. *Recent Advances in Statistics (H. Rizvi and J. Rustagi, eds.), New York: Academic Press.*

[36] Fisher, R. (1974). On the interpretation of χ_2 from contingency tables, and the calculation of p. *Journal of the Royal Statistical Society* **85** 87-94.

[37] Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association* **100** 1021-1035.

[38] Griffin, J. E. and Steel, M. F. J. (2006b). Order-based dependent Dirichlet processes. *Journal of the American Statistical Association* **101** 179-194.

[39] Green, P and Richardson, S.. Modelling heterogeneity with and without the Dirichlet process *Scandinavian Journal of Statistics* **28** 355-375.

[40] Hal Daume III and Daniel Marcu (2005). A Bayesian Model for Supervised Clustering with the Dirichlet Process Prior. *Journal of Machine Learning Research 6* **1**, 1551-1577.

[41] Huillet, T and Christian Paroissin (2005). A Bayesian Model for Supervised Clustering with the Dirichlet Process Prior. *preprint*.

[42] In Dey, D., Muller, P. and Sinha, D. eds. (1969). Computational methods for mixture of Dirichlet process models. *Practical Nonparametric and Semiparametric Bayesian Statistics,* 23-44. Springer.

[43] Ishwaran, H. and James, L.F. (2002). Approximate Dirichlet processes computing in finite normal mixtures: smoothing and prior information. *J. Comp. Graph. Stat.* **11** 209-230.

[44] Ishwaran H., James, L.F and Sun, J. (2000). Bayesian Model Selection in Finite Mixtures by Marginal Density Decompositions. *conditionally accepted by Journal of the American Statistical Association.*

[45] Ishwaran, H. and Zarepour, M. (2000). Markov Chain Monte Carlo in Approximate Dirichlet and Beta Two-Parameter Process Hierarchical Models,. *EJP* **1**, 1-28.

[46] Ishwaran, H. and James, L.F. (2003). Generalized weighted Chinese restaurant processes for species sampling mixture models. *Statistica Sinica,* **13** 1211-1235.

[47] Jason A. DUAN, Michele GUINDANI, and Alan E. GELFAND (2007). Generalized spatial Dirichlet process models. *Biometrika,* 94, 4, pp. 809-825.

[48] Jain, S. and Neal, R. M. (2000). A split-merge Markov chain Monte Carlo procedure for the Dirichlet process mixture model. University of Toronto.

[49] Kacperczyk, M., Damien, P., and Walker, S. G.(2003). A new class of Bayesian semiparametric models with applications to option pricing. *Tech. rep., University of Michigan Bussiness School.*

[50] Kingmann, J. F. C. (1974). Approximate Dirichlet processes computing in finite normal. *Royal statistical society section B*

[51] Kingmann, J. F. C. and James, L.F. (2002). Random discrete Distributions.

[52] Kingmann, J. F. C. (1993). Poisson Processes. Clarendon, Oxford University Press. *Royal statistical society section B*

[53] Kottas, T. and Gelfand, A. E. (2002). A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* .

[54] Kottas, T., A. Duan, J. and Gelfand, A. E. (2007). Modeling Disease Incidence Data with Spatial and Spatio-Temporal Dirichlet Process Mixtures. *Biometrical Journal,* 5, 1-14 DOI: 10.1002/bimj.200610375.

[55] Laws, D. J. and O'Hagan, A. (1975). A hierarchical Bayes model for multilocation auditing. *Journal of the Royal Statistical Society, Series D* **37,** 431-450.

[56] Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. density estimates. *Annals of Statistics* **12** 351-357.

[57] Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4** 639-650.

[58] Sethuraman, J. and West, M. (1994). Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association, 90, pp*, 577-588.

[59] Sethuraman, J. and Tiwari, R. C. (1982). Convergence of Dirichlet Measures and the Interpretation of Their Parameters. *Statistical Decision Theory and Related Topics* III, 2, 305-315.

[60] Soule, A., Salamatian, K. and Emilion, R. (2004). Classification of histograms. *Sigmetrics 2004, New York*. http://www-rp.lip6.fr/site_npa/site_rp/publications.php.

[61] Soule, A., Salamatian, K., Taft, N., Emilion, R. and Papagiannaki, K. (2004). Classification of Internet flows. *Sigmetrics'04*, New-York.

[62] Soule, A., Salamatian, K., Taft, N., Emilion, R. and Papagiannaki, K. (2004). Classification of Internet flows. *Sigmetrics'04*, New-York.

[63] MacEachern, S. N. (1994). Estimating normal means with a conjugate style Dirichlet process prior. *Commnunications in Statistics, Part B - Simulation and Com- putation* **23** 727-741.

[64] MacEachern, S. N. (1999). Dependent nonparametric processes. *In ASA Proceed- ings of the Section on Bayesian Statistical Science,* 50-55.

[65] MacEachern, S. N. (1999). Dependent Dirichlet processes. *Tech. rep., Ohio State University, Department of Statistics.*

[66] MacCloskey, J.W. (1965). Model for the Distribution of Individuals by Species in an Environment. *Unpublished Ph.D. thesis, Michigan State University.*

[67] Meuller, P., Quintana, F., and Rosner, G. (1998). Hierarchical meta-analysis over related non-parametric Bayesian models. *Journal of Royal Statistical Society, Se- ries B* **66**, 735-749.

[68] Medvedovic, M. and Sivaganesan, S. (2002). Bayesian infinite mixture model-based clustering of gene expression profiles. *Bioinformatics* **18**, 1194-1206.

[69] Morris, J. S. and Carroll, R. J. (2002). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B* **68**, 179-199.

[70] N. Johnson and S. Kotz. (1978). Urn Models and Their Applications: an approach to Modern Discrete Probability Theory. *Technometrics, Vol. 20, No. 4, Part 1 (Nov., 1978), p. 501.*

[71] Radford M. NEAL (1996). Density Modeling and Clustering Using Dirichlet Diffusion Trees. *BAYESIAN STATISTICS 7, pp.* 619-629.

[72] Roeder, K. and Wasserman, L. (1997). Practical Bayesian Density Estimation Using Mixtures of Normals. *Journal of the American Statistical Association* , 92, 894-902.

[73] P. Damien, J. C. Wakefield, and S. G.Walker. (1999). S. G.Walker. Gibbs sampling for Bayesian nonconjugate and hierarchical models using auxiliary variables. *Journal of the Royal Statistical Society Series B,* **61:** 331-344.

[74] Pitman, J. and Yor, M. (1996). Random discrete distributions derived from self-similar random sets. *EJP* **1**, 1-28.

[75] Pitman, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from stable subordiantors. *Ann. Proba* **25**, 2, 855-900.

[76] Pitman, J. (2003). Poisson-Kingman partitions. In: Science and Statistics: A Festschrift for Terry Speed. D. R. Goldstein editor. Lecture Notes - Monograph Series 30 1-34. *Institute of Mathematical Statistics Hayward, California.* **2** Lecture Notes - Monograph Series 30 1-34.

[77] Wu, H. and Zhang, J. T. (2002). Local polynomial mixed-effects models for longitudinal data. *Tech. rep., Duke University - Institute of Statistics and Decision Sciences.*

[78] West, M., Aguilar, O., and Lourdes, V. (2003).Va hospital quality monitors: 1988- 1997. *Tech. rep., Duke University - Institute of Statistics and Decision Sciences.*

[79] YeeWhye Teh, Dilan Gorur and Zoubin Ghahramani (2007). Stick-breaking Construction for the Indian Buffet Process.

Hafedh Faires

Modèles hiérarchiques de Dirichlet à temps continu

Résumé :

Nous étudions les processus de Dirichlet dont le paramètre est une mesure proportionnelle à la loi d'un processus temporel, par exemple un mouvement Brownien ou un processus de saut Markovien. Nous les utilisons pour proposer des modèles hiérarchiques bayésiens basés sur des équations différentielles stochastiques en milieu aléatoire. Nous proposons une méthode pour estimer les paramètres de tels modèles et nous l'illustrons sur l'équation de Black-Scholes en milieu aléatoire.

Mots-clés : Statistiques Bayesien, Mouvement Brownien, Échantillonneur de Gibbs, Chaîne de Markov, Mélanges, Milieu aléatoire, Regime-switching, Calculs stochastiques, Équations différentielle stochastiques, volatilités stochastiques, mesure de Wiener.

CONTINUOUS TIME DIRICHLET HIERARCHICAL MODELS

Abstract :

We consider Dirichlet processes whose parameter is a measure proportional to the distribution of a continuous time process, such as a Brownian motion or a Markov jump process. We use them to propose some Bayesian hierarchical models based on stochastic differential equations in random environment. We propose a method for estimating the parameters of such models and illustrate it on the Black-Scholes equation in random environment.

Key words : Bayesian statistics, Brownian motion, Classification, Dirichlet process, Gibbs sampling, Markov chain, Mixtures, Random environment, Regime-switching, Stochastic calculus, Stochastic differential equations, stochastic volatility, Wiener measure.

MAPMO, Université d'Orléans
UFR Sciences Bâtiment de mathématiques

Route de Chartres B.P. 6759, 45067 Orléans cedex 2 FRANCE