



**HAL**  
open science

# Extrema de processus stochastiques. Propriétés asymptotiques de tests d'hypothèses

Cécile Mercadier

► **To cite this version:**

Cécile Mercadier. Extrema de processus stochastiques. Propriétés asymptotiques de tests d'hypothèses. Mathématiques [math]. Université Paul Sabatier - Toulouse III, 2005. Français. NNT : . tel-00010070

**HAL Id: tel-00010070**

**<https://theses.hal.science/tel-00010070>**

Submitted on 8 Sep 2005

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE

pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ TOULOUSE III**

*Discipline* : Mathématiques Appliquées

*Spécialité* : Statistique

présentée par

**Cécile MERCADIER**

---

**Extrema de processus stochastiques.  
Propriétés asymptotiques de tests d'hypothèses.**

---

Soutenue publiquement le 1er juillet 2005 devant le jury composé de :

M. Jean-Marc AZAÏS	Professeur à l'Université Toulouse III	Directeur de Thèse
M. Bernard BERCU	Professeur à l'Université Toulouse III	Examinateur
M. Éric MOULINES	Professeur à l'ENST	Rapporteur
M. Holger ROOTZÉN	Professeur à l'Université de Göteborg	Rapporteur
M. Igor RYCHLIK	Professeur à l'Université de Lund	Examinateur
M. Mario WSCHEBOR	Professeur à l'Université de Montevideo	Examinateur







# THÈSE

pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ TOULOUSE III**

*Discipline* : Mathématiques Appliquées

*Spécialité* : Statistique

présentée par

**Cécile MERCADIER**

---

**Extrema de processus stochastiques.  
Propriétés asymptotiques de tests d'hypothèses.**

---

Soutenue publiquement le 1er juillet 2005 devant le jury composé de :

M. Jean-Marc AZAÏS	Professeur à l'Université Toulouse III	Directeur de Thèse
M. Bernard BERCU	Professeur à l'Université Toulouse III	Examineur
M. Éric MOULINES	Professeur à l'ENST	Rapporteur
M. Holger ROOTZÉN	Professeur à l'Université de Göteborg	Rapporteur
M. Igor RYCHLIK	Professeur à l'Université de Lund	Examineur
M. Mario WSCHEBOR	Professeur à l'Université de Montevideo	Examineur



## *Remerciements*

Je tiens à adresser mes plus vifs remerciements à Jean-Marc Azaïs. Son enthousiasme contagieux et sa gentillesse m'ont permis d'intégrer dans les meilleures conditions le monde de la recherche. À partir d'un sujet passionnant, il m'a aidé à développer mon raisonnement par des conseils judicieux et des discussions stimulantes. Je salue également sa constante disponibilité et conserve précieusement les enseignements que j'ai reçu de lui à propos du travail d'un chercheur. Pour toutes ces raisons, je souhaite que cette thèse soit le point de départ d'une longue collaboration.

À l'occasion de deux conférences j'ai constaté, avec une grande joie, l'intérêt que portait Holger Rootzén à mes communications. Je lui exprime toute ma gratitude pour avoir accepté le rôle de rapporteur. Ses précieux commentaires ont enrichi le texte qui suit et lui offrent de nouvelles perspectives.

Je remercie également Eric Moulines pour avoir accepté d'être rapporteur de ce travail. Je lui suis reconnaissante pour le soin de sa lecture me permettant de bénéficier de ses points de vue très instructifs.

Je décerne officiellement à Mario Wschebor le titre de co-directeur de thèse. J'ai pu lui exposer mes problèmes techniques comme lui confier mon humeur du moment. Grâce à lui, je me suis familiarisée avec les mathématiques de l'hémisphère sud et même d'autres thèmes tant sa culture est immense. Je le remercie chaleureusement car ce séjour s'est révélé être bien plus qu'un enrichissement de ma thèse.

Je noue aussi une relation amicale avec Igor Rychlik depuis sa généreuse invitation à l'Université de Lund. Nos discussions ont stimulé ma curiosité sur ses intérêts de recherche et amélioré le contenu de mon travail. Je suis honorée par sa présence dans mon jury et plus généralement par la constante attention qu'il témoigne à mes projets.

Enfin, je suis très heureuse que Bernard Bercu fasse partie de mon jury. Il a toujours trouvé les mots justes pour apaiser mon inquiétude lors de ces trois années. Sa profonde gentillesse est une leçon de vie.

Elisabeth Gassiat est le visage féminin de mes collaborateurs. Elle a partagé sans retenue ses compétences sur les modèles de mélange. Je la remercie pour sa patience et particulièrement pour ses messages de soutien lorsque pesait dans nos bureaux respectifs cette redoutable canicule.

Je souhaite saluer toutes les personnes du laboratoire de Statistique et Probabilités de Toulouse. Je pense en particulier aux membres d'une amicale mafia composée de Serge Cohen, Fabrice Gamboa, Jean-Claude Fort, Jean-Michel Loubes et Elie Maza. Je pense aussi aux doctorants avec lesquels les avant-séminaires ou les pauses-anniversaires ont été des moments de détente. De même, un grand merci aux maîtres de conférences Anne-Laure, Clémentine, Sébastien, Jérémie, Thierry et Jean-François de n'être pas avares d'écoute et de conseils.



Au cours de ma thèse, j'ai passé trois mois en Uruguay. J'en suis revenue avec une famille de coeur. Susana m'a ouvert les portes de sa maison et je veux dire ici, pour l'honorer, que je n'ai manqué de rien tant elle a de l'amour à donner. Je pense aussi Al Vota, à ma petite soeur Laura, à Claudia et Melina qui m'entraînaient la nuit y hasta siempre !

Côté famille toujours, j'adresse une pensée particulière à Jackie & Elie, Pierre, Anne et Isabelle qui me motivent depuis longtemps.

Le bilan de trois années de thèse et plus généralement de ces huit années à Paul Sabatier ne serait pas si positif sans la présence de mes potes. J'ai agréablement cohabité avec Delphine, Marielle, Christophe, Guillaume, Lionel et les hispanos. Je les remercie pour m'avoir fait évacuer mon stress à coups de fous rire aussi bien dans le bureau qu'à "l'upsi". J'ai plaisir à évoquer ici les escapades à la "Tireuse", les diners chez Clem, les sorties ciné commentées par Jp et Seb, les visites de "Montpel" animées par Céline & Philippe. Une place particulière revient à Nico pour l'ensemble des spectacles d'un comique sans pareil qu'ils nous a donné en soirées. J'adresse enfin mon plus beau sourire à Yan dans le but de dissiper les craintes qu'il a émises dans ses propres remerciements !

Place aux filles maintenant, à mes amies de longue ou plus courte date, que je souhaite remercier pour leur fidélité. En respectant la chronologie de nos rencontres, je pense à Emeline, Céline (avec qui j'ai appris la rigueur imposée par la danse), Véro, Audrey (victimes elles-aussi du meilleur comme du pire offert par le lycée Fermat), Caroline, Marilyne, Hélène (qui m'ont démontré que ma folie n'avait rien de rare) et Karine (avec qui je partage le goût de l'espagnol et des routes arpentées avec le camion de Pascal, sans Pascal !).

Évidemment les plus fidèles sont mes parents Michèle & Jean-Pierre et mon grand frère Pierre. Grâce à leurs efforts j'ai naïvement cru, pendant de longues années, que la vie n'était que douce et simple. Même si la réalité m'a rattrapée depuis, leur soutien reste présent et précieux.

*Cécile.*





# Table des matières

<b>I PRÉSENTATION GÉNÉRALE</b>	<b>11</b>
<b>Introduction</b>	<b>13</b>
<b>1 Présentation du modèle</b>	<b>17</b>
1.1 Processus stochastiques . . . . .	17
1.2 Fonctionnelles de processus . . . . .	18
1.3 Exemples d'applications . . . . .	19
Bibliographie du chapitre 1 . . . . .	21
<b>2 Résumé de la partie II</b>	<b>23</b>
2.1 Comportement asymptotique du maximum . . . . .	23
2.1.1 Introduction à la théorie des valeurs extrêmes . . . . .	23
2.1.2 Modèles gaussiens non-stationnaires en temps continu . . . . .	27
Bibliographie de la section 2.1 . . . . .	32
2.2 Estimation non asymptotique du maximum . . . . .	34
2.2.1 Sous propriétés Markoviennes . . . . .	34
2.2.2 Bornes et équivalents classiques . . . . .	36
2.2.3 Conséquences d'une formule implicite . . . . .	40
Bibliographie de la section 2.2 . . . . .	42
<b>3 Résumé de la partie III</b>	<b>45</b>
3.1 Introduction . . . . .	45
3.2 Classes de Donsker . . . . .	47
3.3 Lien avec le maximum d'un processus stochastique . . . . .	48
3.4 Un test non uniformément le plus puissant . . . . .	51
Bibliographie du chapitre 3 . . . . .	53
<b>II DISTRIBUTION DU MAXIMUM DE PROCESSUS</b>	<b>55</b>
<b>4 Caractère asymptotiquement Poissonien des extrêmes dans des modèles gaussiens non-stationnaires</b>	<b>57</b>
4.1 Introduction . . . . .	57
4.1.1 Notations and hypotheses . . . . .	58
4.2 Main results . . . . .	60
4.2.1 Example of application . . . . .	61
4.3 Main tools . . . . .	62
4.4 Proofs . . . . .	62
4.4.1 Proof of Theorem 4.2.2 . . . . .	63

4.4.2	Proof of Lemma 4.4.1 . . . . .	65
4.4.3	Proof of Lemma 4.4.2 . . . . .	68
4.4.4	Proof of Lemma 4.4.3 . . . . .	70
4.4.5	Proof of Lemma 4.4.4 . . . . .	72
4.4.6	Proof of Corollary 4.2.1 . . . . .	73
	Bibliography of Chapter 4 . . . . .	74
<b>5</b>	<b>Bornes numériques pour la loi du maximum d'un processus à un ou deux paramètres</b>	<b>77</b>
5.1	Introduction . . . . .	78
5.2	Smooth stochastic processes . . . . .	82
5.3	Numerical applications for smooth Gaussian processes . . . . .	84
5.3.1	MATLAB Tool-box . . . . .	85
5.3.2	Checking the program and evaluating the error . . . . .	86
5.3.3	Elements to answer to : "how far is Rice bound numerically si- gnificant?" . . . . .	87
5.3.4	Estimation of the persistence exponent . . . . .	89
5.4	Numerical applications for non-smooth Gaussian processes . . . . .	91
5.5	Two-parameter Gaussian processes . . . . .	94
5.6	Numerical applications for two-parameter Gaussian processes . . . . .	100
5.7	WAFO . . . . .	103
5.8	MAGP . . . . .	105
	Bibliography of Chapter 5 . . . . .	106
<b>III</b>	<b>APPLICATIONS STATISTIQUES</b>	<b>109</b>
<b>6</b>	<b>Distribution asymptotique et puissance du test de rapport de vrai- semblance pour les mélanges : cas borné et non borné</b>	<b>111</b>
6.1	Introduction . . . . .	112
6.2	Asymptotic distribution of the LRT for the number of populations in a mixture under null and contiguous hypotheses . . . . .	114
6.2.1	Contamination mixture . . . . .	117
6.2.2	Two populations against a single one . . . . .	123
6.2.3	Contamination with unknown nuisance parameter . . . . .	126
6.2.4	General mixtures with unknown nuisance parameter . . . . .	129
6.3	The LRT for contamination mixtures with large set of parameters . . .	134
6.4	Asymptotic distribution of the LRT for Gaussian contamination mix- tures with unbounded mean under contiguous alternatives . . . . .	137
6.5	Proofs . . . . .	140
6.5.1	Proof of Theorem 6.2.5 . . . . .	140
6.5.2	Proof of Theorem 6.3.1 . . . . .	140
6.5.3	Proof of Corollary 6.3.1 . . . . .	141
6.5.4	Proof of formula (6.25) . . . . .	145
6.5.5	Proof of Lemma 6.4.1 . . . . .	146
	Bibliography of Chapter 6 . . . . .	148

---

<b>7</b>	<b>Test d'une hypothèse sur la matrice de covariance</b>	<b>151</b>
7.1	Introduction . . . . .	151
7.1.1	Analyse en composantes principales . . . . .	151
7.1.2	Matrice de Wishart . . . . .	152
7.2	Deux tests pour la matrice de covariance . . . . .	154
7.2.1	Test du "Spiked model" . . . . .	154
7.2.2	Test de Sphéricité . . . . .	156
7.3	Distribution du nombre de condition . . . . .	156
7.4	Preuves des lemmes . . . . .	161
	Bibliographie du chapitre 7 . . . . .	172
<b>IV</b>	<b>RÉSULTATS ANNEXES</b>	<b>175</b>
<b>8</b>	<b>Processus de Poisson</b>	<b>177</b>
8.1	Définition . . . . .	177
8.2	Convergence faible vers un processus de Poisson . . . . .	178
	Bibliographie du chapitre 8 . . . . .	181
<b>9</b>	<b>Formules de Rice</b>	<b>183</b>
9.1	Introduction . . . . .	183
9.2	Résultats analytiques . . . . .	185
9.2.1	Pour un niveau fixé . . . . .	185
9.2.2	Pour presque tout niveau . . . . .	186
9.3	Formule de Rice avec des poids . . . . .	188
	Bibliographie du chapitre 9 . . . . .	191
	<b>Index des auteurs cités</b>	<b>194</b>
	<b>Summary of the thesis</b>	<b>197</b>



---

---

PREMIÈRE PARTIE

PRÉSENTATION GÉNÉRALE

---

---





Les connexions interdisciplinaires de la Statistique sont de plus en plus nombreuses et de plus en plus variées. Elles jouent d'une part le rôle de motivation pour introduire des modèles et sont d'autre part un terrain d'application. La liste de ces domaines contient la finance, l'assurance, la génétique, l'informatique, la recherche pétrolière, ... Mais la Statistique est aussi un instrument de formalisme où l'ensemble des modèles est très étendu. Cette spécialité mathématique suit le chemin de toute autre en s'enrichissant aussi d'elle-même.

Mon travail de thèse, décrit à travers ce manuscrit, en est une courte illustration. En effet, je présente des résultats théoriques concernant la *statistique des processus* ou la *théorie des tests* tout comme des méthodes numériques répondant à des problèmes venus de la physique ou de la génétique.

Beaucoup de phénomènes physiques sont continus par nature et conduisent à une modélisation mathématique particulière : le processus stochastique à temps continu. Une grande variété de caractéristiques des processus sont sources d'intérêt statistique.

PARTIE II : DISTRIBUTION DU MAXIMUM DE PROCESSUS (page 57)

La seconde partie de ce document rapporte l'analyse de la distribution du maximum global (et de façon analogique, celle du minimum). Elle contient précisément deux travaux distincts.

#### - Chapitre 4, résumé dans la section 2.1.2 -

En collaboration avec J.-M. Azaïs, nous avons étendu un résultat de convergence en distribution du maximum sur l'intervalle  $[0, t]$  d'un processus, noté  $M_t$ . L'apport de ce travail est le caractère non-stationnaire du processus gaussien étudié. Sous un ensemble d'hypothèses raisonnables, nous démontrons l'existence et explicitons l'expression de constantes  $a_t$  et  $b_t$  telles que  $a_t(M_t - b_t)$  admette la loi de Gumbel pour distribution asymptotique.

En somme, par simple changement de variable, nous disposons d'une approximation de la loi du maximum  $M_t$  qui est fine pour  $t$  suffisamment grand. Ce travail s'inscrit pleinement dans la *théorie des extrêmes* (introduite dans la section 2.1.1).

Ce résultat est obtenu comme conséquence directe d'une autre convergence en distribution : celle du processus ponctuel des dépassements de niveau (vers le haut) convenablement normalisé vers un processus de Poisson standard.

**- Chapitre 5, résumé dans la section 2.2.3 -**

Le second travail donne la distribution du maximum d'un processus défini sur un ensemble compact, au moyen d'une formule implicite. A la différence des études traditionnelles qui dénombrent les dépassements de niveau, seule l'existence d'un dépassement de niveau est retenue ici. Cette méthode est développée pour des processus à un ou deux paramètres. Dans ce deuxième cas, j'obtiens la première borne supérieure pour la fonction de survie du maximum d'un processus gaussien et unit-speed. Son expression est explicite et son calcul ne nécessite pas l'utilisation d'outils numériques.

La mise en oeuvre pratique de la méthode générale nécessite à plusieurs reprises la discrétisation de l'espace de définition. Elle a donné naissance à une boîte à outils Matlab dont les fonctions permettent le calcul d'une borne supérieure ou inférieure de la loi du maximum.

Enfin, une étude comparative est menée entre ces nouvelles estimations et les éléments connus comme la borne de Rice, la série de Rice ou l'utilisation de la caractéristique d'Euler (présentés dans la section 2.2.2).

Ces études appartenant au domaine de la statistique théorique s'appliquent à différents problèmes environnementaux. Le résultat présenté dans le chapitre 4 permet en particulier de calculer la probabilité pour qu'une concentration excède un seuil critique au moins une fois sur un intervalle de temps considéré comme grand. De même, les estimations déduites de la méthode donnée au chapitre 5 constituent une aide précieuse pour la détermination de régions critiques dans différents exemples de la théorie des tests.

**PARTIE III : APPLICATIONS STATISTIQUES (page 111)**

Ceci est illustré à travers l'étude des *modèles de mélange* qui sont par exemple utilisés en génétique pour conclure à l'existence d'un gène majeur dans l'expression d'un trait qualitatif.

**- Chapitre 6, résumé dans les sections 3.3 et 3.4 -**

En collaboration avec É. Gassiat et J.-M. Azaïs, nous résumons et complétons les propriétés asymptotiques du test basé sur la statistique du maximum de vraisemblance dans les modèles de mélange (introduits dans la section 3.1). Ce test permet en outre de déterminer la valeur du nombre de composantes du mélange. Dans une première partie, nous revenons sur la distribution asymptotique de cette statistique. Dans le cadre considéré, elle repose sur une fonctionnelle du maximum d'un ou plusieurs processus gaussiens. Pour différentes lois usuelles,

nous minimisons l'ensemble d'hypothèses nécessaires et réduisons, lorsque c'est possible, l'expression asymptotique à une fonctionnelle ne dépendant que d'un seul processus gaussien.

Dans un second temps, nous nous intéressons à des modèles à horizon infini, autrement dit, à des modèles dont les paramètres ne sont pas bornés a priori. Nous caractérisons de nouveau la loi asymptotique mais aussi la puissance asymptotique de cette statistique.

## - Chapitre 7 -

J'expose dans ce chapitre un état de l'art, restreint à mes motivations, des résultats sur les matrices de Wishart. Seules les valeurs propres extrêmes sont retenues dans le but de construire un *test de sphéricité* basé sur le rapport de la plus grande et de la plus petite valeur propre.

Il s'agit aussi d'une étude préliminaire sur la possibilité d'extension de résultats relatifs à ce rapport. À ce titre, nous donnons quelques majorations des densités marginales ou de la densité jointe de ces variables.

Malheureusement certains résultats théoriques ne sont disponibles que dans le cas complexe, ce qui freine l'aboutissement ou l'espoir d'aboutissement d'une telle construction.

## PARTIE IV : RÉSULTATS ANNEXES (page 177)

Les études traitées au cours de cette thèse reposent, pour une part certaine, sur la maîtrise de problèmes connexes à ceux étudiés. Dans la quatrième partie de ce document, nous rappelons au chapitre 8 quelques caractérisations du processus de Poisson et donnons une description bibliographique de la méthode de Rice dans le chapitre 9. Ce dernier contient une extension de la formule de Rice pour un champ gaussien dans laquelle nous attribuons un poids à chacun des dépassements.

Chaque chapitre de ce document possède sa propre bibliographie. Nous donnons en fin de manuscrit un index des auteurs cités (page 194) et ce résumé de thèse en version anglaise (page 197).

Nous poursuivons avec la première partie, dite de présentation générale, qui décrit de façon détaillée le cadre et les idées liés aux résultats obtenus.



# Chapitre 1

---

## PRÉSENTATION DU MODÈLE

Ce premier chapitre est destiné à présenter, de façon simple et peu formelle, l'objet aléatoire considéré tout au long de ce manuscrit. En particulier, nous décrivons les caractéristiques que nous souhaitons étudier et des exemples d'applications concrètes ou abstraites qui motivent leurs analyses.

### 1.1 Processus stochastiques

Le modèle mathématique commun à toutes les études présentées est le processus stochastique. Rappelons rapidement qu'il s'agit d'un ensemble  $\{X(t), t \in \mathcal{T}\}$  de variables aléatoires définies sur  $\mathcal{T}$  et ici, à valeurs dans  $\mathbb{R}^k$ . L'espace des paramètres  $\mathcal{T}$  est usuellement un sous-ensemble de

- $\mathbb{N}$  ou  $\mathbb{Z}$ ; on parle de suite aléatoire ou de série chronologique,
- $\mathbb{R}^+$  ou  $\mathbb{R}$ ; on parle plutôt de processus à temps continu,
- $\mathbb{R}^n$ ; on parle alors de champ aléatoire (si  $k > 1$ ) et de processus stochastique à plusieurs paramètres (si  $k$  vaut 1).

Parmi les processus à temps continu, on compte d'une part les processus ponctuels, dont les changements d'état sont dirigés par une suite de points  $t_i$  finie ou infinie, et d'autre part les processus continus. Il est d'ailleurs intéressant de voir, en particulier dans ce travail, que la résolution d'un problème donné peut nécessiter l'introduction de ces deux types de processus.

Beaucoup de systèmes réels peuvent être modélisés ainsi, citons en exemple :

- l'évolution de la taille d'une population subissant des contraintes extérieures de territoire ou de réserve alimentaire,
- la fluctuation d'un courant dans un circuit électrique en présence de prétendu bruit thermique,
- les changements du niveau des signaux-radio reçus en présence de perturbations météorologiques,
- l'écoulement turbulent d'un liquide ou d'un gaz,
- la hauteur d'une partie de la surface de mer.

A ces derniers s'ajoutent certains processus industriels accompagnés de fluctuations aléatoires, et également certains processus produits de la bio-physique, de la géophysique et des sciences économiques. Prévoir les conséquences de certaines situations grâce à

de tels modèles est une aide précieuse pour établir des stratégies.

Trois classes principales de processus stochastiques se dessinent.

1. Les processus aux accroissements indépendants pour lesquels lorsque les paramètres vérifient  $t_1 < t_2 \leq t_3 < t_4$ , les variables aléatoires  $X(t_2) - X(t_1)$  et  $X(t_4) - X(t_3)$  sont indépendantes.
2. Les modèles de Markov jouent un rôle très important. Historiquement ils introduisent la notion de dépendance alors que, jusqu'en 1913, les aléas étaient perçus comme indépendants les uns des autres. Un processus de Markov est en effet caractérisé par le fait que la probabilité conditionnelle des états possibles aux instants  $t_{n+1}, t_{n+2}, \dots$ , connaissant les états aux instants  $t_1, \dots, t_n$  ne dépend que du dernier instant passé, à savoir  $t_n$ .
3. Les processus gaussiens vérifient pour tout entier  $n$  et tout vecteur  $(t_1, \dots, t_n)$  dans  $\mathbb{R}^n$  : la distribution jointe des variables aléatoires  $X(t_1), \dots, X(t_n)$  est une loi gaussienne (ou normale) multivariée. La loi normale étant déterminée de façon unique par ses premier et deuxième moments, un tel processus stochastique est caractérisé par sa fonction moyenne  $m(t) = \mathbb{E}(X(t))$  et sa fonction de covariance  $r(s, t) = \mathbb{E}(X(s)X(t)) - m(s)m(t)$ . La connaissance des lois normales et plus généralement la place qui leur est faite dans la littérature permet de justifier l'intérêt intrinsèque d'une étude basée sur les processus gaussiens.

Cependant la théorie mathématique des processus stochastiques ou des champs aléatoires comprend aussi de nombreux résultats sur une série de sous-classes, ou au contraire d'extensions des classes de processus présentées ci-dessus.

## 1.2 Fonctionnelles de processus

Nous nous sommes intéressés à certaines caractéristiques d'un processus stochastique  $\{X(t), t \in \mathcal{T}\}$  à valeurs réelles où  $\mathcal{T}$  est un sous-ensemble de  $\mathbb{R}^n$ .

Les principaux résultats présentés dans ce manuscrit concernent les extrema globaux de cette courbe ou surface aléatoire donnés par

$$M_{\mathcal{T}} = \sup_{t \in \mathcal{T}} X(t) \quad \text{et} \quad m_{\mathcal{T}} = \inf_{t \in \mathcal{T}} X(t).$$

Ce problème a un sens lorsque  $\mathcal{T}$  est dénombrable ou bien sous des hypothèses de séparabilité. C'est d'ailleurs ce que nous supposons implicitement tout au long du document,  $\{X(t), t \in \mathcal{T}\}$  est séparable. *Un processus à valeurs réelles défini sur un espace topologique  $\mathcal{T}$  est dit séparable s'il existe un sous-ensemble dense  $\mathcal{D}$  de  $\mathcal{T}$  tel que, presque sûrement sur tout ouvert  $V$  de  $\mathcal{T}$*

$$M_V = \sup_{t \in V \cap \mathcal{D}} X(t) \quad \text{et} \quad m_V = \inf_{t \in V \cap \mathcal{D}} X(t).$$

L'ordre de grandeur des extrema globaux dépend d'une part de la régularité du processus et d'autre part de la taille de l'espace des paramètres  $\mathcal{T}$ . Bien que les travaux présentés dans cette thèse soient de nature variée, deux cadres d'étude se distinguent tout de même.

- Les chapitres 5, 6 et 7 se rejoignent sur le fait que  $\mathcal{T}$  est compact (ou à horizon fini) avec  $\mathcal{T} = [0, T]$ ,  $\mathcal{T} = [-T, T]$  et  $\mathcal{T} = S^{n-1}$  la sphère unité de  $\mathbb{R}^n$ .
- Les chapitres 4 et 6 contiennent le cas où  $\mathcal{T}$  est mobile (ou à horizon infini). Nous observons alors, quand  $t$  tend vers l'infini, le comportement d'une normalisation de

$$M_t = \sup_{s \in [0, t]} X(s) \quad \text{ou} \quad \widetilde{M}_t = \sup_{s \in [-t, t]} X(s).$$

Dans le cadre de l'analyse de ces fonctionnelles du processus  $X$ , une approche classique consiste à introduire les processus ponctuels de franchissement, et de dépassement de niveau vers le haut ou vers le bas. Présentons-les dans le cas simple où le processus considéré  $\{X(t), t \geq 0\}$  est à trajectoires presque sûrement différentiables. Pour tout intervalle  $B$ , on définit

$$U_u^X(B) = \#\{t \in B : X(t) = u, X'(t) > 0\},$$

$$D_u^X(B) = \#\{t \in B : X(t) = u, X'(t) < 0\},$$

$$N_u^X(B) = U_u(B) + D_u(B).$$

$U_u^X(B)$  est appelé le nombre de upcrossings (ou dépassements vers le haut) du niveau  $u$  par  $X$  sur  $B$ . De même,  $D_u^X(B)$  et  $N_u^X(B)$  sont appelés respectivement le nombre de downcrossings (ou dépassements vers le bas) et le nombre de crossings (ou franchissements) du niveau  $u$  par  $X$  sur  $B$ . Se référer à Cramér et Leadbetter (1967) pour une définition dans un cadre plus général. Lorsqu'il n'y a pas d'ambiguïté, nous ne faisons pas apparaître le nom  $X$  du processus observé.

Le lien entre le maximum et le nombre de upcrossings est simple puisque l'événement  $\{M_t > u\}$  est presque sûrement l'union disjointe de

- $\{X(0) > u\}$ , le processus commence sa trajectoire au-dessus du niveau observé,
- et  $\{X(0) \leq u, U_u(0, t) \geq 1\}$ , la trajectoire part en dessous du niveau  $u$  mais le dépasse avant le temps  $t$ .

Nous faisons appel au processus des upcrossings dans les chapitres 4, 5 et 6. Cependant, nous ne nous attachons principalement qu'au calcul de ses premier et second moments faisant référence à la méthode de Rice (décrite dans le chapitre 9).

Un processus semblable est considéré au chapitre 7 pour compter le nombre de dépassements vers le bas du niveau  $u = 0$  par la dérivée du processus initialement étudié.

### 1.3 Exemples d'applications

La distribution des extrema globaux d'un processus stochastique apparaît dans de nombreux problèmes statistiques, à la fois théoriques et appliqués. Nous commençons par présenter des applications associées à des niveaux finis et nous terminons par le point de vue asymptotique.



L'EXPOSANT DE PERSISTANCE. (cf. chapitre 5)

Pour donner une définition simple de la persistance, considérons un processus stochastique qui ne peut prendre que les valeurs -1 ou 1. La persistance jusqu'au temps  $t$  correspond à la situation la plus extrême où la variable n'a pas changé d'état. En d'autres termes, la "particule" a passé tout son temps dans une seule des deux phases possibles. La probabilité de cet événement, probabilité de persistance, décroît avec le temps avec un exposant non trivial. La recherche de ces exposants est un domaine d'investigation très présent dans la littérature des physiciens décrite ici à travers Deloubrière et Hilhorst (2000), Ehrhardt et al. (2002) et Molchan et Khokhlov (2004).

Si l'on s'intéresse maintenant à caractériser le signe d'un processus stochastique centré  $\{X(t), t \geq 0\}$ , le problème de persistance revient à étudier la probabilité de persistance  $p_0$  pour que le maximum reste en dessous du niveau 0. Dans ce cas, sous condition d'existence, la limite

$$q(0) = \lim_{t \rightarrow +\infty} -\frac{1}{t} \log \mathbb{P}(M_t \leq 0)$$

est communément appelée exposant de persistance car elle donne une estimation de la probabilité de persistance  $p_0$  sous la forme  $\exp(-tq(0))$ .

LE TEST DE SPHÉRICITE. (cf. chapitre 7)

Une méthode pour analyser la structure d'un tableau de données  $A$  consiste à estimer la matrice de covariance théorique  $\Sigma$  de la population. Un test de sphéricité est un moyen pour répondre au test de l'hypothèse

$$(H_0) : \Sigma = \sigma I_m \quad \text{contre} \quad (H_1) : \Sigma \neq \sigma I_m.$$

Il peut être construit en regardant des fonctionnelles des valeurs propres extrêmes de la matrice  $A^T A$ , c'est-à-dire

$$\lambda_{\min} = \min_{s \in S^{m-1}} X(s) \quad \text{et} \quad \lambda_{\max} = \max_{s \in S^{m-1}} X(s)$$

avec  $X(s) = s^T A^T A s$ .

LE TEST DU MAXIMUM DE RAPPORT DE VRAISEMBLANCE. (cf. chapitre 6)

Une méthode fondamentale en Statistique pour construire un estimateur de paramètres inconnus est la méthode du maximum de vraisemblance. Supposons que nous avons, pour un échantillon  $Y_1, \dots, Y_n$  de distribution  $\mathbb{P}_\theta$  dépendant d'un paramètre inconnu  $\theta \in \Theta \subseteq \mathbb{R}^k$ , le but d'estimer  $\theta$ . Lorsque toutes les mesures  $\mathbb{P}_\theta$  sont absolument continues relativement à une mesure de référence  $\mu$ , la fonction de vraisemblance est définie par

$$L(\theta) = \prod_{i=1}^n \frac{d\mathbb{P}_\theta}{d\mu}(Y_i).$$

Cette méthode permet de tester l'hypothèse nulle  $(H_0) : \theta = \theta_0$  contre  $(H_1) : \theta \in \Theta_1$ . La statistique de test basée sur le maximum de vraisemblance est

$$\lambda_n = \sup_{\theta \in \Theta_1} \log L(\theta) / L(\theta_0).$$

Dans certains contextes, cette variable aléatoire converge en distribution vers une fonction du maximum d'un processus gaussien dont les caractéristiques dépendent de  $\theta_0$  et  $\Theta_1$  (Dacunha-Castelle et Gassiat, 1997). La construction d'un test basé sur la valeur empirique de  $\lambda_n$  nécessite donc la connaissance de la loi du maximum pour des niveaux plutôt grands.

EN STATISTIQUE DES EXTRÊMES. (cf. chapitre 4)

La théorie des valeurs extrêmes étudie en particulier les questions liées aux très grandes valeurs (ou au contraire, aux très petites) prises par des processus stochastiques. L'ouvrage de Finkenstädt et Rootzén (2004) et les références qu'il contient décrivent des applications dans les domaines aussi variés que la finance, les télécommunications et l'environnement. A la différence du processus gaussien considéré dans ce travail, l'analyse d'un taux d'intérêt ou autre objet financier conduit à une modélisation plus complexe. Les mouvements d'un prix sont, par exemple, représentés sous la forme d'un processus de diffusion donné par une équation différentielle stochastique

$$dX(t) = \mu(X(t)) dt + \sigma(X(t)) dW(t), \quad t > 0, \quad X(0) = x,$$

avec  $W$  le mouvement Brownien standard,  $\mu$  un terme de drift réel et  $\sigma$  à valeurs strictement positives décrivant la volatilité.

Le comportement asymptotique (lorsque  $t$  tend vers l'infini) du maximum joue un rôle considérable dans la modélisation stochastique des événements rares.

## Bibliographie du chapitre 1

- Cramér, H. et Leadbetter, M. R.** (1967). *Stationary and related stochastic processes. Sample function properties and their applications*. John Wiley & Sons Inc., New York.
- Dacunha-Castelle, D. et Gassiat, É.** (1997). Testing in locally conic models, and application to mixture models. *ESAIM Probab. Statist.*, 1 :285–317.
- Deloubrière, O. et Hilhorst, H. J.** (2000). Persistence exponents of non-Gaussian processes in statistical mechanics. *J. Phys. A*, 33(10) :1993–2013.
- Ehrhardt, G. C., Bray, A. J., et Majumdar, S. N.** (2002). Persistence of a continuous stochastic process with discrete-time sampling : non-Markov processes. *Phys. Rev. E (3)*, 65(4) :041102, 13.
- Finkenstädt, B. et Rootzén, H.** (2004). *Extreme Values in Finance, Telecommunications, and the Environment*. Chapman & Hall.
- Molchan, G. et Khokhlov, A.** (2004). Small values of the maximum for the integral of fractional Brownian motion. *J. Statist. Phys.*, 114(3-4) :923–946.



# Chapitre 2

---

## RÉSUMÉ DE LA PARTIE II

Nous décrivons dans ce chapitre les résultats présentés de façon détaillée dans la partie II de ce document. Outre la précision du contenu bibliographique, nous expliquons ici les idées maîtresses ainsi que le bénéfice associés à ces études.

### 2.1 Comportement asymptotique du maximum

Ce travail s'inscrit dans la lignée des résultats composant la théorie des valeurs extrêmes. Ces analyses, à la fois probabilistes et statistiques, s'intéressent avant tout au comportement stochastique du maximum et du minimum de phénomènes donnés. En particulier, nous cherchons à estimer la probabilité d'un événement rare. L'objectif du travail, décrit dans le chapitre 4, est de trouver l'ordre du maximum d'un processus gaussien non-stationnaire au temps  $t$ , lorsque  $t$  tend vers l'infini.

#### 2.1.1 Introduction à la théorie des valeurs extrêmes

Elle est un mélange d'applications très variées et de résultats mathématiques avancés sur les processus ponctuels et les fonctions à variations régulières.

En effet, les applications traitées par ces méthodes sont nombreuses. Nous trouvons d'une part les phénomènes naturels comme par exemple

- les rafales de vent,
- la pollution de l'air,
- la corrosion,
- l'analyse sismique,
- les précipitations et les crues.

Et d'autre part, la fiabilité des structures mécaniques, les émissions radioactives, le management des risques en économie ou en finance.

Les problèmes étudiés sont par exemple la détermination d'un niveau limite pour l'ozone, le calcul de la taille d'un barrage ou la minimisation par les assureurs des risques de réclamations très importantes.

Parallèlement la connaissance théorique acquise, depuis disons un siècle, est vaste elle aussi. Les premiers articles, introduisant de façon nominative le concept de distribution de la plus grande valeur, voient le jour au cours des années 1920. En particulier, Fréchet (1927) identifie une des distributions asymptotiques possibles de la plus grande

statistique d'ordre alors que Fisher et Tippett (1928) montrent qu'elles sont en fait au nombre de trois. En 1943, Gnedenko présente une preuve rigoureuse de ce résultat et en fournit les conditions nécessaires et suffisantes dans le cas indépendant et identiquement distribué (noté par la suite i.i.d.). Citons en référence les livres de Resnick (1987), Leadbetter, Lindgren et Rootzén (1983) pour la théorie et les livres de Castillo (1988), Finkenstädt et Rootzén (2004) pour les applications des valeurs extrêmes.

Considérons  $\{X_n, n \in \mathbb{N}^*\}$  une suite de variables aléatoires i.i.d., de fonction de répartition commune notée  $F$ . Notons le maximum associé

$$M_n = \max_{k=1, \dots, n} X_k$$

qui a  $F^n$  pour fonction de répartition.

Par analogie avec le théorème de la limite centrale, qui décrit la distribution limite d'une somme de variables aléatoires i.i.d., l'objectif de la théorie des valeurs extrêmes est de trouver un résultat analogue pour  $M_n$ . Nous recherchons donc des conditions sous lesquelles il existe des constantes de normalisation  $a_n > 0$ ,  $b_n \in \mathbb{R}$  et une distribution non-dégénérée  $G$  telles que

$$(2.1) \quad a_n(M_n - b_n) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} G$$

où  $\xrightarrow{\mathcal{L}}$  désigne la convergence en loi, ou de façon équivalente telles que

$$\lim_{n \rightarrow +\infty} F(x/a_n + b_n)^n = G(x).$$

Deux fonctions de répartitions  $G$  et  $G_1$  sont dites du même type si l'on peut trouver deux constantes  $c_1 > 0$  et  $c_2 \in \mathbb{R}$  telles que pour tout  $x$  dans  $\mathbb{R}$ ,

$$G(x) = G_1(c_1x + c_2).$$

### Proposition 2.1.1 (Gnedenko)

Soit  $G$  une fonction de répartition non dégénérée. Supposons qu'il existe  $a_n > 0$  et  $b_n \in \mathbb{R}$  pour tout  $n \in \mathbb{N}^*$  tels que

$$\mathbb{P}(a_n(M_n - b_n) \leq x) = F^n(x/a_n + b_n) \xrightarrow[n \rightarrow +\infty]{} G(x)$$

Alors  $G$  est du type de l'une des trois classes suivantes

$$\star G_1(x) = \begin{cases} 0 & \text{si } x < 0 \\ \exp(-x^{-\alpha}) & \text{sinon} \end{cases} \quad \text{pour un certain } \alpha > 0,$$

et la distribution associée à  $G_1$  est appelée loi de Fréchet,

$$\star G_2(x) = \begin{cases} \exp(-(-x)^\alpha) & \text{si } x < 0 \\ 1 & \text{sinon} \end{cases} \quad \text{pour un certain } \alpha > 0,$$

et la distribution associée à  $G_2$  est appelée loi de Weibull,

$$\star G_3(x) = \exp(-e^{-x}) \quad \text{pour } x \in \mathbb{R},$$

et la distribution associée à  $G_3$  est appelée loi de Gumbel.

Cette caractérisation est une classification des comportements des extrêmes dans le cadre i.i.d et c'est pour cette raison que nous allons en donner la preuve.

*Preuve de la Proposition 2.1.1.* Avec  $[x]$  désignant la partie entière de  $x$  et la convergence supposée du théorème, on a pour tout  $t$  positif

$$F^{[nt]}(x/a_{[nt]} + b_{[nt]}) \xrightarrow[n \rightarrow +\infty]{} G(x)$$

et aussi

$$F^{[nt]}(x/a_n + b_n) = (F^n(x/a_n + b_n))^{\frac{[nt]}{n}} \xrightarrow[n \rightarrow +\infty]{} G^t(x).$$

Soit  $y$  un élément de  $\mathbb{R}$ ; en étudiant les ensembles  $\inf\{x, H(x) \geq y\}$  pour  $H = G$  et  $H = G^t$  on peut montrer qu'il existe deux fonctions réelles mesurables  $\alpha$  et  $\beta$  définies sur  $\mathbb{R}$ , avec  $\alpha(t) > 0$  pour tout  $t$ , telles que

$$\lim_{n \rightarrow +\infty} a_{[nt]}/a_n = \alpha(t) \quad \text{et} \quad \lim_{n \rightarrow +\infty} b_{[nt]}/b_n = \beta(t).$$

De plus, il est possible de montrer, via des constructions s'appuyant sur le théorème de Skorohod, que

$$(2.2) \quad G^t(x) = G(\alpha(t)x + \beta(t)).$$

En particulier, on a

$$G^{ts}(x) = G(\alpha(ts)x + \beta(ts))$$

et d'autre part

$$G^{ts}(x) = G(\alpha(s)\alpha(t)x + \alpha(t)\beta(s) + \beta(t))$$

en utilisant (2.2) à deux reprises. La non-dégénérescence de  $G$  permet d'affirmer que pour tout  $t > 0$  et  $s > 0$

$$\alpha(ts) = \alpha(s)\alpha(t) \quad \text{et} \quad \beta(ts) = \alpha(t)\beta(s) + \beta(t) = \alpha(s)\alpha(t) + \beta(s).$$

L'équation en  $\alpha$  n'admet qu'une solution mesurable positive donnée par

$$\alpha(t) = t^{-\theta} \text{ avec } \theta \in \mathbb{R}.$$

Si  $\theta = 0$ , on montre tout d'abord que  $\beta(t) = -c \log(t)$  pour  $t > 0$  avec  $c > 0$  puis on conclut en utilisant le fait que  $G(0)$  s'écrit sous la forme  $\exp(-e^{-p})$ . Cette situation conduit au type  $G_3$ .

Si  $\theta > 0$ , la fonction  $\beta/(1 - \alpha)$  est identiquement égale à une constante notée  $d$ . Ce qui implique facilement l'égalité  $G^t(x + d) = G(t^{-\theta}x + d)$ . On obtient la caractérisation de  $G$  en montrant que  $G(1 + d) \in ]0, 1[$  donc peut s'écrire comme un élément de la forme  $\exp(-p^{-\theta-1})$ . Nous concluons au type  $G_1$ .

Dans le dernier cas, on obtient le type  $G_2$ . □

Pour un type de distribution limite donné, parmi ceux obtenus dans la proposition 2.1.1, nous connaissons des conditions nécessaires et suffisantes sur  $F$  pour que la convergence (2.1) se réalise. Elles sont clairement établies et rappelées dans les ouvrages de référence précédemment cités.

En particulier, le domaine d'attraction de la loi de Gumbel contient la distribution normale. Autrement dit, si  $F$  représente la fonction de répartition d'une loi gaussienne standard (centrée et de variance un), ces conditions impliquent l'existence de constantes  $a_n$  et  $b_n$  telles que

$$(2.3) \quad a_n(M_n - b_n) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} G_3.$$

Parmi les modèles utilisés dans les applications citées au début de la section 2.1.1, beaucoup décrivent l'évolution d'un système en temps continu. De plus, dans la plupart des cas d'étude, les états observés sont fortement dépendants. Ces constats motivent les travaux visant à affaiblir l'indépendance et à prolonger la théorie pour caractériser le comportement extrême des suites de variables aléatoires dépendantes et des processus continus.

Une première démarche pour prendre en compte la dépendance est de considérer un phénomène stationnaire. Si  $\{X_n, n \in \mathbb{N}^*\}$  est une suite stationnaire gaussienne standard telle que

$$\rho(n) \log(n) \rightarrow 0,$$

où  $\rho(n) = \mathbb{E}(X_0 X_n)$ , alors nous savons que la convergence en loi (2.3) se réalise avec les constantes

$$\begin{aligned} a_n &= \sqrt{2 \log(n)}, \\ b_n &= a_n - \frac{1 \log(\log(n)) + \log(4\pi)}{2 a_n}. \end{aligned}$$

Ce type de condition sur la covariance s'appelle hypothèse de Berman. Il existe des versions plus faibles permettant d'obtenir la même convergence. Cependant, nous ne

les détaillons pas mais conseillons de consulter l'état de l'art présenté par Leadbetter, Lindgren et Rootzén (1983).

Pour un processus  $\{X(t), t \geq 0\}$  gaussien standard nous cherchons la loi asymptotique d'une normalisation de

$$M_t = \max_{s \in [0, t]} X(s).$$

La version en temps continu du problème auquel nous nous intéressons consiste à trouver des conditions d'existence de constantes  $a_t > 0$  et  $b_t \in \mathbb{R}$  telles que

$$(2.4) \quad a_t(M_t - b_t) \xrightarrow[t \rightarrow +\infty]{\mathcal{L}} G_3.$$

Lorsque le processus est stationnaire, le maximum admet le comportement décrit par (2.4) sous une condition de régularité et dès que par exemple

$$(2.5) \quad \rho(t) \log(t) \rightarrow 0$$

où  $\rho(t) = \mathbb{E}((X(0)X(t)))$  avec les constantes

$$(2.6) \quad \begin{aligned} a_t &= \sqrt{2 \log(t)}, \\ b_t &= a_t - \frac{\log(2\pi)}{a_t}. \end{aligned}$$

Nous pouvons remarquer que ces normalisations diffèrent d'un cas à l'autre, mais de façon très faible. Chacune d'elles est une fonction qui tend très lentement vers l'infini avec le temps.

Intéressons-nous maintenant à une étape supplémentaire dans le procédé d'extension et regardons le cas non-stationnaire. Pour une suite gaussienne standard telle que si  $i$  et  $j$  sont distincts

$$r(i, j) < \rho_{|i-j|}$$

où  $r(i, j) = \mathbb{E}(X_i X_j)$ ,  $r(i, i) = 1$ ,  $\rho_k < 1$  pour  $k > 1$  et  $\rho_k \log(k) \rightarrow 0$  quand  $k$  tend vers l'infini, le résultat (2.3) reste vrai avec les constantes  $a_n$  et  $b_n$  données précédemment. Les améliorations successives que nous venons de rappeler laissent beaucoup d'espoir quant à la possibilité de traiter le cas d'un processus continu non-stationnaire.

### 2.1.2 Modèles gaussiens non-stationnaires en temps continu

Le chapitre 4 présente l'extension de la distribution asymptotique du maximum explicitée dans (2.4), lorsque  $\{X(t), t \in \mathbb{R}\}$  est un processus gaussien non nécessairement stationnaire mais à variance constante, égale à un. Sans argument pour ramener le processus à un processus standard, nous devons considérer sa fonction moyenne  $m(t) = \mathbb{E}(X(t))$  et sa fonction de covariance  $r(s, t) = \mathbb{E}[(X(t) - m(t))(X(s) - m(s))]$ . Lorsque les trajectoires du processus sont des fonctions suffisamment régulières, nous définissons les dérivées partielles de  $r$

$$r_{ij}(s, t) = \frac{\partial^{i+j}}{\partial s^i \partial t^j} r(s, t)$$



et la fonction

$$(2.7) \quad S(t) = \int_0^t \sqrt{r_{11}(s, s)} ds.$$

Elle permet de construire le processus  $\{Y(t) = X(S^{-1}(t)), t \geq 0\}$  qui vérifie

$$\text{Var}(Y(t)) = \text{Var}(Y'(t)) = 1.$$

Le processus  $Y$  est appelé la transformation "unit-speed" de  $X$ . On note  $r^Y$  sa fonction de covariance.

Voici l'énoncé de notre résultat.

### Corollaire 4.2.1

Soit  $\{X(t), t \in \mathbb{R}\}$  un processus gaussien de variance 1 satisfaisant les hypothèses (G) suivantes.

$$(G1) \quad \forall t \in \mathbb{R}, r_{11}(t, t) > 0 \text{ et } \lim_{t \rightarrow +\infty} S(t) = +\infty,$$

$$(G2) \quad \sup_{|S(s) - S(t)| \geq A} r(s, t) \log |S(s) - S(t)| \xrightarrow{A \rightarrow +\infty} 0,$$

$$(G3) \quad \forall \varepsilon > 0, \sup_{|S(s) - S(t)| > \varepsilon} |r(s, t)| < 1,$$

$$(G4) \quad r \text{ est de classe } \mathcal{C}^4 \text{ avec } s \rightarrow r_{11}(s, s) \text{ de classe } \mathcal{C}^3, \\ \forall \gamma > 0, r_{01}^Y \text{ et } r_{04}^Y \text{ sont bornées sur } \{(s, t) \in \mathbb{R}^2, s > \gamma \text{ et } t > \gamma\},$$

$$(G5) \quad \sqrt{\log \circ S(t)} m(t) \xrightarrow{t \rightarrow +\infty} 0.$$

Alors, nous avons

$$a_t(M_t - b_t) \xrightarrow[t \rightarrow +\infty]{\mathcal{L}} G_3$$

avec  $a_t = \sqrt{2 \log \circ S(t)}$  et  $b_t = a_t - \log(2\pi) a_t^{-1}$ .

#### • COMMENTAIRES SUR LES HYPOTHÈSES DU COROLLAIRE 4.2.1.

(G1) décrit des conditions sur les variations de  $\text{Var}(X'(t))$ . Nous demandons qu'à chaque instant le processus dérivé ne soit pas dégénéré. Les conditions (G2) et (G3) généralisent la condition de Berman (2.5). Les hypothèses de régularité (G4) sont nécessaires à plusieurs étapes : justifier l'approximation polygonale pour évaluer le nombre de upcrossings (défini dans la section 1.2), minorer la loi du maximum grâce au moment factoriel d'ordre deux des upcrossings. Enfin, il n'est pas nécessaire de supposer le processus centré mais seulement asymptotiquement centré avec une vitesse donnée par (G5).

Nous rencontrons un processus vérifiant les hypothèses (G) par exemple dans la théorie des tests de mélange. Il s'agit de tester, pour  $\mathbb{T}$  un sous-ensemble compact de  $\mathbb{R}$  et  $g$  la densité d'une population

$$(H_0) : "g = f_0" \quad \text{contre} \quad (H_1) : "g = (1 - \pi)f_0 + \pi f_t, 0 \leq \pi \leq 1, t \in \mathbb{T}."$$

D'une part, lorsque l'on suppose le mélange gaussien  $f_t(x) = \exp(-(x-t)^2/2)/\sqrt{2\pi}$ , la distribution asymptotique de la statistique de test définie par maximum de vraisemblance est donnée, sous hypothèse nulle, par l'intermédiaire du maximum d'un processus gaussien centré, de fonction de covariance

$$r(s, t) = \frac{\exp(st) - 1}{\sqrt{\exp(s^2) - 1}\sqrt{\exp(t^2) - 1}}$$

vérifiant l'ensemble d'hypothèses (G1–G4). Ce résultat asymptotique est détaillé dans la section 6.2.1.

D'autre part, pour un modèle de mélange dit de translation où

$$f_t(x) = f_0(x - t)$$

le processus gaussien intervenant dans cette distribution asymptotique vérifie les hypothèses du corollaire 4.2.1 dès que la densité  $f_0$ , définissant le mélange, satisfait l'ensemble d'hypothèse (H) associé au corollaire 6.3.1. C'est d'ailleurs un point clé dans la démonstration du corollaire 6.3.1.

• IDÉES DE LA DÉMONSTRATION DU COROLLAIRE 4.2.1.

Supposons que pour toute valeur positive réelle  $s$ , il existe une fonction  $C_u$  de  $u$  telle que

$$(2.8) \quad \lim_{u \rightarrow +\infty} \mathbb{P}(U_u(0, C_u^{-1}s) = 0) = e^{-s}$$

lorsque  $U_u$  représente le processus des upcrossings de niveau  $u$  associé au processus  $X$ . En utilisant le lien entre  $M_t$  et  $U_u$  (décrit à la fin de la section 1.2), on obtient pour  $t = C_u^{-1}s$  et  $s = e^{-x}$

$$\lim_{u \rightarrow +\infty} \mathbb{P}(M_t \leq u) = \exp(-e^{-x}).$$

De la relation précédente liant  $u$  à  $t$  et  $x$ , on obtient la convergence (2.4) avec  $a_t$  et  $b_t$  connues de façon explicite. Il s'en suit que le résultat sur le corollaire 4.2.1 est une conséquence directe de la convergence (2.8). Cette méthode pour obtenir la distribution asymptotique du maximum normalisé est classique. Sous certaines de ces conditions, on peut prouver que  $U_u$  satisfait un résultat plus large que (2.8) : le processus  $U_u(C_u^{-1}\cdot)$  converge vers un processus de Poisson standard (Théorème 4.2.2).

Mais cette méthode de démonstration, pour un résultat comme le corollaire 4.2.1, n'est pas la seule possible. Par exemple, Hüsler (1990, 1995) traite le cas des processus localement stationnaire d'indice  $\alpha \in ]0, 2]$ , gaussiens, à variance 1. Ces derniers ne sont pas suffisamment réguliers pour pouvoir définir le caractère poissonnien des dépassements. La conclusion de notre corollaire 4.2.1 est très proche de celle de Hüsler dans le cas  $\alpha = 2$ , mais il ne dispose d'aucun résultat sur la famille de processus  $(U_u)_u$ .

Nous présentons maintenant l'énoncé de notre second résultat pour un processus "unit-speed".

**Théorème 4.2.2** *Soit  $\{X(t), t \in \mathbb{R}\}$  un processus gaussien de variance 1 satisfaisant les hypothèses (U) suivantes.*

$$(U1) \quad \forall t \in \mathbb{R}, r_{11}(t, t) = 1,$$

$$(U2) \quad \sup_{|s-t| \geq A} r(s, t) \log |s - t| \xrightarrow{A \rightarrow +\infty} 0,$$

$$(U3) \quad \forall \varepsilon > 0, \sup_{|s-t| > \varepsilon} |r(s, t)| < 1,$$

$$(U4) \quad r \text{ est de classe } \mathcal{C}^4 \\ \forall \gamma > 0, r_{01} \text{ et } r_{04} \text{ sont bornées sur } \{(s, t) \in \mathbb{R}^2, s > \gamma \text{ et } t > \gamma\},$$

$$(U5) \quad \sqrt{\log(t)} m(t) \xrightarrow{t \rightarrow +\infty} 0,$$

$$(U6) \quad m'(t) \xrightarrow{t \rightarrow +\infty} 0.$$

Pour  $C_u = \frac{1}{2\pi} e^{-u^2/2}$ , la famille de processus  $U_u(C_u^{-1} \cdot)$  converge faiblement sur  $]0, +\infty[$  vers un processus de Poisson standard.

• COMMENTAIRES BIBLIOGRAPHIQUES.

Revenons un instant sur le cas où  $\{X(t), t \geq 0\}$  est un processus gaussien stationnaire, centré et de variance un. Nous notons  $\rho$  sa fonction de covariance. Nous commençons par lister un certain nombre de caractéristiques possibles pour cette fonction. Les constantes  $\lambda_i$  sont les moments spectraux ; nous renvoyons à la partie 9.1 pour plus de détails.

$$(2.9) \quad \rho(t) = 1 - \frac{\lambda_2}{2!} t^2 + \frac{\lambda_4}{4!} t^4 + o(t^4),$$

$$(2.10) \quad \rho(t) = O(t^{-\alpha}) \text{ pour } \alpha > 0 \text{ et } t \rightarrow +\infty$$

$$(2.11) \quad \rho(t) \log(t) \rightarrow 0, \text{ quand } t \text{ tend vers } +\infty,$$

$$(2.12) \quad \int_0^{+\infty} \rho^2(s) ds < +\infty$$

$$(2.13) \quad \rho(t) = 1 - \frac{\lambda_2}{2!} t^2 + o(t^2),$$

$$(2.14) \quad \exists \eta > 0, \quad \rho(t) = 0, \forall t \geq \eta$$

L'équivalent stationnaire du théorème 4.2.2 est connu depuis 1961 et n'a cessé d'être démontré sous des hypothèses de plus en plus faibles. Nous présentons maintenant un extrait de sa bibliographie.

- Volkonskiï et Rozanov (1961) en supposant (2.9) et des conditions de mélange,
- Cramér et Leadbetter (1967) sous les conditions (2.9) et (2.10),
- Berman (1971) avec les hypothèses (2.13) et l'une des hypothèses (2.11) ou (2.12),
- Leadbetter, Lindgren et Rootzén (1983) sous les conditions (2.13) et (2.11).
- Wschebor (1985) montre le même résultat sous les conditions (2.9) et (2.14), plus restrictives que celles de Leadbetter, Lindgren et Rootzén par exemple, lui permettant d'éviter certaines complications techniques.

Dans le cas stationnaire, nous avons

$$C_u = \frac{\mathbb{E}(U_u(0, t))}{t} = \frac{\sqrt{\lambda_2}}{2\pi} e^{-u^2/2}.$$

Cette valeur est appelée intensité du processus ponctuel des upcrossings.

Le processus  $U_u(C_u^{-1}\cdot)$  est stationnaire (la probabilité pour avoir  $n$  upcrossings sur un intervalle de longueur  $t$  ne dépend pas de la position de l'intervalle). De plus, en utilisant le fait que deux upcrossings ne peuvent pas arriver au même instant, on en déduit qu'il est aussi régulier (la probabilité d'avoir au moins deux upcrossings sur un intervalle de longueur  $t$  est un  $o(t)$  lorsque  $t$  tend vers 0).

La proposition 8.2 du chapitre 8 donne une caractérisation pour conclure que ce processus est (asymptotiquement en  $u$ ) un processus de Poisson.

• COMMENTAIRES SUR LES HYPOTHÈSES DU THÉORÈME 4.2.2.

Les hypothèses ( $U1 - U5$ ) ont en fait donné naissance aux conditions ( $G1 - G5$ ) du corollaire 4.2.1. Les commentaires que nous pourrions en faire sont par conséquent semblables à ceux qui suivent le corollaire 4.2.1. Cependant, le résultat précédent nécessite une sixième hypothèse portant sur la moyenne asymptotique du processus. On peut facilement observer que localement la valeur de  $m'$  peut avoir une influence sur le nombre de upcrossings mais n'en a pas sur le niveau du maximum.

• IDÉES DE LA DÉMONSTRATION DU THÉORÈME 4.2.2.

En choisissant

$$C_u = \frac{1}{2\pi} e^{-u^2/2}$$

nous avons pour sous-ensemble  $B$  de  $\mathbb{R}$ , qui peut s'écrire comme une réunion finie d'intervalles disjoints,

$$\mathbb{E}(U_u(C_u^{-1}B)) = \lambda(B) (1 + o(1))$$

où  $\lambda$  désigne la mesure de Lebesgue sur  $\mathbb{R}$ .

Un résultat de Kallenberg, rappelé dans la proposition 8.5 du chapitre 8, assure qu'il suffit de prouver alors que

$$\mathbb{P}\left(U_u(C_u^{-1}B) = 0\right) \xrightarrow{u \rightarrow \infty} \exp(-\lambda(B))$$

pour en déduire que le processus est asymptotiquement un processus de Poisson standard. Cette propriété s'obtient en introduisant l'approximation polygonale du processus  $X$  et un découpage fin des intervalles considérés.

Le théorème 4.2.2 s'étend au cas d'un processus non-stationnaire général (dans le sens non unit-speed mais à variance constante). Il s'agit du théorème 4.2.3 obtenu en modifiant de façon adéquate les hypothèses et la normalisation précédentes. La convergence du processus de upcrossings reste alors vraie pour

$$U_u(S^{-1}(C_u^{-1}\cdot))$$

où  $S$  est définie par (2.7).

Pour conclure, nous pouvons remarquer deux éléments importants dans ce travail. D'une part, l'hypothèse de variance constante du processus initial  $X$  est cruciale avec cette méthode. Cela pose ouvertement la question du comportement du maximum lorsque cette hypothèse n'est pas vérifiée. Nous pouvons d'ailleurs poser la même question concernant l'hypothèse gaussienne. D'autre part la souplesse sur la stationnarité, entraînée par notre résultat, confirme finalement la robustesse de l'ensemble des résultats énoncés dans cette partie.

## Bibliographie de la section 2.1

- Azaïs, J.-M. et Mercadier, C.** (2003). Asymptotic poisson character of extremes in non-stationary gaussian models. *Extremes*, 6 :301–318.
- Berman, S. M.** (1971). Excursions above high levels for stationary Gaussian processes. *Pacific J. Math.*, 36 :63–79.
- Castillo, E.** (1988). *Extreme value theory in engineering*. Statistical Modeling and Decision Science. Academic Press Inc., Boston, MA.
- Cramér, H. et Leadbetter, M. R.** (1967). *Stationary and related stochastic processes. Sample function properties and their applications*. John Wiley & Sons Inc., New York.
- Finkenstädt, B. et Rootzén, H.** (2004). *Extreme Values in Finance, Telecommunications, and the Environment*. Chapman & Hall.
- Fisher, R. et Tippett, L.** (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Proc. Camb. Phil. Soc.*, 24 :180–190.
- Fréchet, M.** (1927). Sur la loi de probabilité de l'écart maximum. *Ann. Soc. polonaise Math. Cracovie*, t. VI :93–116.
- Gnedenko, B.** (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math. (2)*, 44 :423–453.
- Hüsler, J.** (1990). Extreme values and high boundary crossings of locally stationary Gaussian processes. *Ann. Probab. (18)*, 3 :1141–1158.
- Hüsler, J.** (1995). A note on extreme values of locally stationary Gaussian processes. *J. Stat. Plann. Inference*, 45 :203–213.

- 
- Leadbetter, M. R., Lindgren, G., et Rootzén, H.** (1983). *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer-Verlag, New York.
- Resnick, S. I.** (1987). *Extreme values, regular variation, and point processes*, volume 4 of *Applied Probability. A Series of the Applied Probability Trust*. Springer-Verlag, New York.
- Volkonskiï, V. A. et Rozanov, J. A.** (1961). Some limit theorems for random functions. II. *Teor. Veroyatnost. i Primenen.*, 6 :202–215.
- Wschebor, M.** (1985). *Surfaces aléatoires*, volume 1147 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin. Mesure géométrique des ensembles de niveau. [The geometric measure of level sets].

## 2.2 Estimation non asymptotique du maximum

Nous venons d'exposer les idées associées à l'étude de la distribution du maximum  $M_t$  pour des niveaux de la forme  $u = x/a_t + b_t$  où les constantes de normalisation  $a_t$  et  $b_t$  sont donnés par (2.6). Cependant, il est aussi très intéressant de caractériser la loi du maximum  $M_t$  en temps  $t$  fini, et donc pour des niveaux de tous ordres.

Nous considérons maintenant un processus stochastique  $\{X(t), t \in S\}$  à valeurs réelles défini sur un sous-ensemble compact  $S$  de  $\mathbb{R}^n$  et son maximum sur  $S$  noté  $M_S$ . Lorsque  $n$  vaut 1, nous prenons  $S$  sous la forme d'un intervalle  $[0, T]$ , et notons  $M_T$  le maximum associé. Cette partie rapporte le travail détaillé dans le chapitre 5 qui considère les cas  $n = 1$  et  $n = 2$ . Nous faisons une étude comparative entre les résultats obtenus et les outils déjà connus. Soulignons que nous disposons d'une borne dans le cas bi-dimensionnel, résultat qui n'apparaît pas dans la littérature.

Les fonctions  $\varphi$ ,  $\Phi$  et  $\Psi$  représentent la densité, la fonction de répartition et la fonction de survie respectivement d'une gaussienne standard.

### 2.2.1 Sous propriétés Markoviennes

Il existe un petit nombre de processus, définis sur une partie compacte de  $\mathbb{R}$ , pour lesquels la distribution du maximum est connue, parfois de façon exacte. Pour les présenter, nous les distinguons suivant deux classes.

LA PREMIÈRE CLASSE EST COMPOSÉE DE FONCTIONNELLES DU MVT BROWNIEN.

- Mouvement Brownien noté  $W$ .

*Nous obtenons la distribution du maximum par le principe de réflexion. Lorsque  $u \geq 0$ ,*

$$\mathbb{P}(M_T \geq u) = \mathbb{P}(|W_T| \geq u) = 2 \left(1 - \Phi(u/\sqrt{T})\right).$$

- Pont Brownien noté  $B$  défini sur  $[0, 1]$ .

*Avec des arguments semblables à ceux du cas précédent.*

$$\mathbb{P}(M_1 \geq u) = \exp(-2u^2).$$

- Processus de Slepian  $X(t) = W(t+1) - W(t)$ .

$$\mathbb{P}(M_n < u) = \int_{D'} \dots \int \det[\varphi(y_i - y_{j+1} + u)]_{0 \leq i, j \leq n} dy_1 \dots dy_{n+1}$$

*où  $n$  est un entier,  $D' = \{0 < y_1 < y_2 < \dots < y_{n+1}\}$  et  $y_0 = 0$ .*

*Ce processus est stationnaire de fonction de covariance  $\rho(t) = (1 - |t|)^+$ .*

*Ce résultat de Slepian (1961) a été généralisé au cas où  $n$  n'est plus entier par Shepp (1971).*

- Le mouvement Brownien normalisé  $X(t) = W(t)/\sqrt{t}$  et sa valeur absolue. *DeLong (1981)* donne des tables de leurs distributions. Par simple transformation, on peut en déduire celle du processus stationnaire de covariance  $\rho(t) = \exp(-|t|)$ , appelé *Ornstein-Uhlenbeck*.
- Le mouvement Brownien intégré  $X(t) = a + \int_0^t Y(s)ds$  et  $Y(t) = b + W(t)$ . *McKean (1963)* donne la densité de transition pour

$$\mathbb{P}(\sup_{0 \leq s \leq t} X(s) = X(t) \in dx, Y(t) \in dy).$$

*Goldman (1971)* détermine la densité du premier temps de passage  $\tau_u$  de  $X$  au point  $u > 0$  quand  $a = 0$  et  $b \leq 0$  en fonction des densités de  $(\tau_0, |Y(\tau_0)|)$  et de celle de *McKean*.

- Un processus semblable défini par  $X(t) = \sqrt{3} \exp(-\sqrt{3}t) \int_0^{te^{2/\sqrt{3}}} W(s)ds$ . *Wong (1970)* étudie la densité de la taille de l'intervalle entre deux zéros consécutifs. Il s'agit en fait d'un processus stationnaire de fonction de covariance  $\rho(t) = 3/2 \exp(-|t|/3) (1 - \exp(-2|t|/\sqrt{3})/3)$ .
- Le mouvement Brownien avec un drift  $X_c(t) = W(t) + ct$ . *Shepp (1979)* obtient la densité jointe de  $M_T$ , sa position et la valeur  $X_c(T)$ .

LA SECONDE CLASSE EST COMPOSÉE DE PROCESSUS GAUSSIENS STATIONNAIRES DONT LA COVARIANCE EST NOTÉE  $\rho$ .

$$\text{- pour } \alpha \in [0, 2], \rho(t) = \begin{cases} 1 - \alpha t, & t \in [0, 1] \\ 1 - \alpha(t - 2), & t \in [1, 2] \end{cases} \quad \text{et } \rho(t) = \rho(-t) = \rho(t + 2).$$

*Shepp et Slepian (1976)* calculent la probabilité conditionnelle que le processus reste au-dessous d'un niveau  $u$  sur  $[0, T]$  sachant sa valeur initiale. Son expression devient facilement calculable dans les cas suivants

- i/  $0 \leq T \leq 1$ ,
- ii/  $T = 2$  et  $\alpha < 2$ ,
- iii/  $u = 0$  et  $0 \leq T \leq 1$ ,
- iv/  $u = 0$  et  $T \geq 2$ .

- pour  $\beta \in (0, 1/2]$  et  $T_\beta = 1 - \beta$ ,  $\rho(t) = \max_{t \in [-T_\beta, T_\beta]} (1 - |t|/(1 - \beta), -\beta/(1 - \beta))$ . *Cressie (1980)* donne la probabilité conditionnelle que le processus reste en-dessous d'un niveau  $u$  strictement positif sur un temps au moins égal à 1, sachant la valeur initiale.



- $\rho(t) = \cos(t)$  définit un processus appelé "cosine process".  
Notons  $g(u, t) = \exp(-u^2(1 - \cos(t)) / \sin^2(t))$ .

$$\mathbb{P}(M_T \geq u) = \begin{cases} \Psi(u) + \varphi(u) \frac{T}{\sqrt{2\pi}} & \forall 0 < T < \pi, \\ \Psi(u) + \varphi(u) \frac{T}{\sqrt{2\pi}} - \int_{\pi}^T g(u, t) \frac{dt}{2\pi} & \forall \pi \leq T < 2\pi, \\ \exp\left(-\frac{u^2}{2}\right) & \forall T \geq 2\pi. \end{cases}$$

Ce résultat de Berman (1971) a été généralisé par Delmas (2003).

Chacun de ces travaux s'appuie fortement soit sur les propriétés markoviennes ou "quasi-markovienne" du processus, soit sur sa périodicité. Il est donc difficile de reprendre leurs arguments pour étendre ces résultats à une catégorie plus large de processus.

## 2.2.2 Bornes et équivalents classiques

La littérature dispose tout de même de moyens d'approximation de la distribution du maximum de processus stochastiques n'appartenant pas à une des deux catégories répertoriées dans la partie 2.2.1. Nous reprenons les notations définies au début de la section 2.2.

- CAS  $n = 1$ .

La fonction de survie du maximum est majorée par la borne dite de Rice.

$$\begin{aligned} (2.15) \quad \mathbb{P}(M_T \geq u) &= \mathbb{P}(X(0) \geq u) + \mathbb{P}(X(0) < u, U_u(0, T) \geq 1), \\ &\leq \mathbb{P}(X(0) \geq u) + \mathbb{E}(U_u(0, T)), \\ &\leq \mathbb{P}(X(0) \geq u) + \int_0^T \mathbb{E}(X'(t)^+ / X(t) = u) p_{X(t)}(u) dt. \end{aligned}$$

Cette borne est un très bon équivalent lorsque  $u$  est grand devant  $T$ . Dans ce cas, le nombre de upcrossings est au plus égal à un, et l'inégalité de Markov est heuristiquement proche d'une égalité.

Remarquons que le dernier événement de (2.15) s'écrit  $\{U_u(0, T) \mathbb{1}_{X(0) < u} \geq 1\}$ . Le théorème suivant est une extension de la formule de Rice dans le sens où elle conserve cette modification du nombre de upcrossings. Il en existe une version plus large, traitant des processus non gaussiens mais extrêmement réguliers.

**Théorème 2.2.1 (Azaïs et Wschebor)**

Soit  $X$  un processus gaussien centré stationnaire de fonction de covariance  $\rho$ . Supposons que  $\rho(0) = 1$  et que la série de Taylor de  $\rho$  en  $2T$  soit absolument convergente. Alors

$$\mathbb{P}(M_T \geq u) = \mathbb{P}(X(0) \geq u) + \sum_{m=1}^{+\infty} (-1)^{m+1} \frac{\nu_m}{m!}$$

avec  $\nu_m = \mathbb{E}(N_u^{[m]} \mathbb{1}_{X(0) < u})$  et  $x^{[m]} = x(x-1)\dots(x-m+1)$ .

Du résultat énoncé, nous obtenons des bornes inférieures et supérieures (notées dans le chapitre 5,  $\text{LBAW}(T, u)$  et  $\text{UBAW}(T, u)$  respectivement). En effet, en notant  $S_k$  la somme partielle associée d'indice  $k$ , on a

$$\mathbb{P}(X(0) \geq u) + S_{2k} \leq \mathbb{P}(M_T \geq u) \leq \mathbb{P}(X(0) \geq u) + S_{2k+1}.$$

D'un point de vue théorique cette formule semble résoudre le problème, d'autant qu'une majoration de l'erreur est connue. Malheureusement son application numérique nécessite le calcul de plusieurs termes complexes, ajoutant à chaque étape une erreur d'estimation.

Pour palier à cette surenchère de termes, nous remarquons que seule l'existence d'un premier dépassement est nécessaire. En caractérisant le maximum à l'aide du premier temps de passage, on obtient une formule implicite.

**Théorème 2.2.2 (Rychlik)**

Soit  $\tau_0$  le temps du premier upcrossing du niveau 0 par un processus  $Z$  défini par

$$Z(s) = F(X(s)) - a(s)$$

avec  $F$  réelle continûment différentiable bijective,  $X$  un processus gaussien centré et  $a$  une fonction réelle continue.

Si l'on peut régresser de façon non dégénérée  $X$  par rapport à  $X(0)$ , alors

$$p_{\tau_0}(t) = \mathbb{E}(Z'(t)^+ \mathbb{1}_{\{Z(s) \leq 0, \forall s < t\}} / Z(t) = 0) p_{Z(t)}(0).$$

Il suffit alors d'écrire, pour  $F$  l'identité et  $a$  identiquement égale à  $u$ ,

$$(2.16) \quad \mathbb{P}(M_T \geq u) = \mathbb{P}(X(0) \geq u) + \int_0^T p_{\tau_0}(t) dt.$$

Cette représentation est très implicite car le maximum présent dans le terme de gauche,  $\{M_T \geq u\}$ , se retrouve aussi dans le terme de droite, à travers la condition  $\{M_t \leq u\}$  de la fonction indicatrice. Nous montrons dans ce travail qu'elle conduit cependant à une très bonne précision numérique.

- CAS  $n > 1$ .

Aucune formule exacte n'est connue dans ce cas, seuls des équivalents sont obtenus par l'une des méthodes suivantes.

(A) ADLER (1981) GÉNÉRALISE LE NOMBRE DE UPCROSSINGS PAR "THE INTEGRAL GEOMETRY CHARACTERISTIC". DANS LE CAS OÙ L'ENSEMBLE DE NIVEAU  $A(u) = \{t \in S, X(t) \geq u\}$  NE RENCONTRE PAS LE BORD DE  $S$ , CETTE GÉNÉRALISATION CORRESPOND À LA CARACTÉRISTIQUE D'EULER DE  $A(u)$ .

Cette méthode estime les composantes connexes de  $A(u)$  motivée par l'égalité

$$\mathbb{P}(M_S \geq u) = \mathbb{P}(\text{Nombre de composantes connexes de } A(u) \geq 1).$$

Comme ce problème n'est pas facile à résoudre, le nombre de composantes connexes est estimé grâce à la caractéristique d'Euler notée  $\phi$ . En dimension 1 et 2 le lien entre ces deux objets est relativement simple. En dimensions supérieures, il est difficile de se faire une idée de la caractéristique d'Euler de  $A(u)$ . Cependant, lorsque le niveau  $u$  étudié est grand, le comportement ellipsoïdal des composantes connexes de  $A(u)$  justifie l'équivalence entre ces deux mesures.

Worsley (1995a,b) fournit une expression exacte de  $\mathbb{E}[\phi(A(u))]$  pour un champ aléatoire gaussien centré stationnaire isotrope et  $S$  un sous-ensemble compact de  $\mathbb{R}^n$  dont la frontière est une variété deux fois continûment différentiable. Adler et Taylor (2005) synthétisent toutes les extensions, reposant principalement sur l'ensemble  $S$ , de ce résultat.

(B) L'UTILISATION DU VOLUME DE TUBES POUR DES CHAMPS GAUSSIENS PARTICULIERS DE VARIANCE UN PAR SUN (1993) ET TAKEMURA ET KURIKI (2002).

Soit  $X$  un champ aléatoire gaussien centré de variance un admettant un développement orthogonal d'ordre  $K$

$$X(t) = \sum_{l=1}^K \xi_l \gamma_l(t) =: \langle \gamma_t, \xi \rangle .$$

Ce travail utilise la formule de Weyl pour calculer le volume d'un tube sans bord sur la sphère unité de  $\mathbb{R}^K$  afin de calculer

$$\mathbb{P}(M_S \geq u) = \int_u^\infty \mathbb{P}(\sup_{t \in S} \langle \gamma_t, \xi/r \rangle \geq ur^{-1}/\|\xi\| = r) \mathbb{P}_{\|\xi\|}(dr).$$

Sous certaines conditions de régularité du bord de  $S$  et de la covariance de  $X$ , cette formule permet d'écrire un équivalent lorsque  $u$  devient grand.

(C) LA MÉTHODE DE LA DOUBLE SOMME DE PITERBARG (1996) QUI CONSISTE À DIVISER L'ESPACE DES PARAMÈTRES EN UNE UNION FINIE D'ENSEMBLES PLUS PETITS.

Pour un ensemble  $S$ , on définit une partition finie  $(S_k)$  de  $S$  telle que

- la taille des  $S_k$  est dépendante du niveau  $u$ ,
- les sous-ensembles  $S_k$  ne sont pas forcément disjoints mais ont une intersection de taille petite en comparaison avec leurs propres tailles.

L'idée initiale est l'encadrement de  $\mathbb{P}(M_S \geq u)$  par

$$\sum_k \mathbb{P}(M_{S_k} \geq u)$$

et

$$\sum_k \mathbb{P}(M_{S_k} \geq u) - \sum_{k \neq l} \mathbb{P}(M_{S_k} \geq u, M_{S_l} \geq u).$$

Les probabilités apparaissant sous une somme simple sont traitées grâce à des outils classiques, comme l'inégalité de Borell. Pour les autres probabilités, il faut au préalable s'assurer d'une asymptotique indépendance des événements.

(D) LA GÉNÉRALISATION DU NOMBRE DE UPCROSSINGS AVEC LE NOMBRE DE MAXIMA LOCAUX PAR PITERBARG (1996) ET AZAÏS ET DELMAS (2002).

Notons  $(\partial S)_{i=1, \dots, p}$  les  $p$  bords de  $S$  supposés suffisamment réguliers et définissons

$$\begin{aligned} M_u^X(S) &= \#\{t \in S, X(t) \geq u, X'(t) = 0, X''(t) < 0\}, \\ R_u^X[(\partial S)_i] &= \{\exists t \in (\partial S)_i, X(t) \geq u, X''(t)|_{(\partial S)_i} < 0, X'_{\mathcal{T}}(t) = 0, X'_{\mathcal{N}}(t) > 0\}, \end{aligned}$$

où  $X''(t)|_{(\partial S)_i}$  est la matrice des dérivées secondes de  $X$  restreintes au bord  $(\partial S)_i$ ,  $X'_{\mathcal{T}}(t)$  est le vecteur des dérivées tangentielles au bord  $(\partial S)_i$  et  $X'_{\mathcal{N}}(t)$  est le vecteur des dérivées normales rentrantes au bord  $(\partial S)_i$ . Le maximum est caractérisé par l'égalité

$$\mathbb{P}(M_S \geq u) = \mathbb{P}[\{M_u^X(S) \geq 1\} \cup \cup_{i=1}^n R_u^X((\partial S)_i)],$$

puis encadré grâce à l'inégalité de Bonferroni.

Cette méthode est largement développée dans la thèse de Delmas (2001).

#### CONCLUSION.

L'approximation de la fonction de survie du maximum a ainsi été obtenue grâce à différentes méthodes et sous des hypothèses de plus en plus faibles. Le cadre le plus général est assurément celui d'Adler et Taylor (chapitre 12 et 13, 2005). Afin d'éviter les notations de leur modèle Riemannien, nous présentons l'équivalent obtenu pour un champ aléatoire gaussien centré isotropique défini sur le rectangle

$$S = [0, T]^n.$$

Sous certaines conditions de non-dégénérescence des lois jointes de  $X'$  et  $X''$  et d'asymptotique indépendance des éléments de  $X''$ , lorsque le niveau  $u$  est grand, nous avons

$$\mathbb{P}(M_S \geq u) = \Psi(u/\sigma) + e^{-u^2/(2\sigma^2)} \sum_{k=1}^n \frac{C_n^k T^k \lambda_2^{k/2}}{(2\pi)^{(k+1)/2} \sigma^k} H_{k-1}(u/\sigma) + o[\exp(-u^2/2(1+\delta))]$$

avec  $\text{Var } X = \sigma$ ,  $\text{Var } X' = \lambda_2 I_n$  et  $H_k(x) = (-1)^k e^{x^2/2} \frac{d^k}{dx^k} (e^{-x^2/2})$ .

L'erreur commise par cette approximation devient donc super-exponentiellement faible avec le niveau. Dans le cas  $n = 2$ , un résultat d'Adler montre que pour un champ aléatoire gaussien centré seulement stationnaire, l'erreur est un  $o[\exp(-u^2/2)/u]$ .

### 2.2.3 Conséquences d'une formule implicite

Le but de l'étude exposée dans le chapitre 5 est d'améliorer les approximations présentées dans les sections précédentes. Plus précisément, nous avons constaté que la littérature contient des outils d'approximations de la distribution du maximum pour tout niveau  $u$  lorsque  $n = 1$  et seulement pour de hauts niveaux  $u$  lorsque  $n > 1$ . Cependant les besoins, en particulier de la physique (problème de la persistance décrit dans la section 1.3), poussent à la considération des bas niveaux.

• PLAN DE L'ÉTUDE.

Lorsque  $n = 1$ , la formule d'Azaïs et Wschebor et celle de Rychlik répondent de façon implicite à ce problème. Par conséquent nous nous sommes intéressés à plusieurs choses.

- (P1) Étendre la formule de Rychlik (2.16) à des processus non-nécessairement gaussiens ayant ps des trajectoires absolument continues, théorème 5.2.1.
- (P2) Simplifier la démonstration de la formule (2.16) dans le cas de processus gaussiens à trajectoires presque sûrement continûment différentiables, proposition 5.2.1.
- (P3) Comparer numériquement les méthodes de Rice, Azaïs et Wschebor et Rychlik pour tous types de niveaux, tables 5.1, 5.2, 5.3 et figures 5.1, 5.2.
- (P4) Appliquer ces estimations au problème de la persistance, section 5.3.4.
- (P5) Construire une méthode numérique d'encadrement de la loi du maximum pour des processus à trajectoires seulement continues presque sûrement, section 6.9.

Lorsque  $n = 2$ , les objectifs sont les suivants.

- (C1) Écrire la généralisation de la formule (2.16) pour un champ gaussien à trajectoires "un peu moins" que deux fois continûment différentiables, théorème 5.5.1.
- (C2) Dédire une borne explicite lorsque le processus est "unit-speed", corollaires 5.5.1 et 5.5.2.
- (C3) Comparer numériquement cette méthode et l'approximation d'Adler et Taylor (2005) donnée à la fin de la section 2.2.2, figure 5.5.

• IDÉES DES DÉMONSTRATIONS.

L'idée principale de (P1) est d'introduire la notion de points records. Dans le cas unidimensionnel, il s'agit de l'ensemble des points

$$E = \{t \in [0, T], M_t \leq X(t)\},$$

tels que la valeur du processus en un de ces points  $t$  est supérieure à toutes les valeurs du processus sur le passé de  $t$ . L'avantage de cet ensemble est que la probabilité

$$\mathbb{P}(X(0) < u, U_u(0, T) \geq 1)$$

s'obtient comme l'espérance du nombre de crossings de niveau  $u$  parmi les points records puisque ce cardinal ne peut prendre que la valeur 0 ou 1. Ainsi, sa valeur est

calculée grâce à la méthode de Rice reposant sur la formule de Banach, théorème 9.2 présenté dans le chapitre 9.

Pour (C1) le concept de points records dans le cas multi-dimensionnel n'est pas défini de façon aussi évidente. En effet, il faut au préalable choisir une direction privilégiée de l'espace puis définir l'ensemble  $\Gamma_t$  représentant le passé du point  $t$ . Le déroulement se fait alors en deux étapes, l'analyse du bord puis de l'intérieur de l'ensemble  $S$ , mais reste, dans chacun des cas, semblable à celui présenté au-dessus. L'espérance du nombre de crossings parmi les points records s'obtient maintenant par application de la méthode de Rice basée sur la formule de Federer, théorème 9.4 rappelé dans le chapitre 9. A ce titre nous avons dû étendre le calcul du premier moment du nombre de crossings au calcul du premier moment d'une somme de poids affectés à des crossings, établi dans le théorème 9.6 démontré dans le chapitre 9.

Le résultat le plus novateur est la dérivation d'une borne facilement calculable dans le cas d'un champ "unit-speed" (C2), dont voici l'énoncé.

**Corollaire 5.5.2.** *Soit  $\{X(t), t \in S = [0, T]^2\}$  un champ aléatoire gaussien satisfaisant :*

- (A<sub>1</sub>) *les trajectoires de  $Z = (X, X_{10})$  sont ps continûment différentiables,*
- (A<sub>2</sub>) *pour  $W \in \mathbb{R}^2$  fixé, ps il n'y a pas de point  $t$  tel que  $Z(t) = W$  et  $\det(Z'(t)) = 0$ ,*
- (A<sub>3</sub>) *pour  $t \in S$ , la loi de  $Z(t)$  est non-dégénérée,*
- (A<sub>4</sub>) *pour  $(s, t) \in S^2$ ,  $s_2 = t_2$  et  $s \neq t$ , la loi jointe de  $Z(s), Z(t)$  est non-dégénérée.*
- (A<sub>5</sub>)  $\text{Var}(X(t)) = 1$ ,
- (A<sub>6</sub>)  $\text{Var}(X'(t)) = Id$ ,
- (A<sub>7</sub>)  $\text{Var}(X_{20}(t))$  *est constante.*

*Alors, pour tout  $u$  dans  $\mathbb{R}$*

$$\mathbb{P}(M_{[0,T]^2} \geq u) \leq \Psi(u) + \frac{T}{\pi} \exp(-u^2/2) + \frac{T^2}{(2\pi)^{\frac{3}{2}}} [c \varphi(\frac{u}{c}) + u \Phi(\frac{u}{c})] \exp(-u^2/2)$$

*avec  $c = (\text{Var}(X_{20}) - 1)^{\frac{1}{2}}$ .*

En utilisant le comportement asymptotique de  $\Phi$

$$\varphi(a)/a - \varphi(a)/a^3 \leq 1 - \Phi(a) \leq \varphi(a)/a,$$

nous déduisons un équivalent de la borne du corollaire 5.5.2 à savoir

$$\Psi(u) + \frac{T}{\pi} \exp(-u^2/2) + \frac{T^2 u}{(2\pi)^{\frac{3}{2}}} \exp(-u^2/2),$$

qui est exactement le développement d'Adler et Taylor. Nous obtenons de plus que la différence entre leur équivalent et notre borne est majorée en valeur absolue par

$$\frac{T^2}{\sqrt{2\pi^2} u^2} \exp(-3u^2/4).$$

• OUTILS NUMÉRIQUES.

Nous avons écrit pour ce travail une boîte à outils appelée MAGP, pour 'Maximum Analysis of Gaussian Processes', contenant l'ensemble de nos fonctions MATLAB et disponible librement à l'adresse

<http://www.lsp.ups-tlse.fr/Fp/Mercadier/MAGP.html>

Ces fonctions sont dépendantes de la routine rind.m de la boîte à outils WAFO (Brodtkorb et al., 2000). Cette page web contient une description de WAFO, rind.m et une aide d'utilisation de MAGP données dans ce manuscrit en sections 5.7 et 5.8.

La fonction rind.m permet le calcul d'intégrales gaussiennes avec l'algorithme de Genz (1992). Ce dernier transforme par exemple l'intégrale gaussienne

$$\frac{1}{\sqrt{(2\pi)^m \det \Sigma}} \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} e^{-\frac{1}{2} \theta^t \Sigma^{-1} \theta} d\theta$$

en une intégrale de la forme

$$(e_1 - d_1) \int_0^1 (e_2 - d_2) \dots \int_0^1 (e_m - d_m) \int_0^1 dw_1 \dots dw_m,$$

où les éléments  $e$  et  $d$  sont issus d'une décomposition de Cholesky suivie d'un changement de variable gaussien ( $y_i = \Phi^{-1}(z_i)$ ) et d'une transformation linéaire.

• CONCLUSIONS NUMÉRIQUES.

Cette étude comparative mène au classement suivant :

- $n = 1$  et  $u \ll T$ , les meilleures estimations numériques s'obtiennent avec la méthode du chapitre 5, c'est-à-dire les bornes (5.7) et (5.8) ;
- $n = 1$  et  $u \gg T$ , la borne de Rice est la plus économique ;
- $n = 1$  avec  $u \gg 1$  et  $T \gg 1$ , les résultats sur les extrêmes doivent être testés, section 2.1 et chapitre 4 ;
- $n = 2$  avec  $u \ll T$ , nous conseillons la méthode décrite dans le chapitre 5 à travers les bornes (5.22) et (5.23) ;
- $n \geq 2$  avec  $u \gg T$ , les approximations issues des travaux présentés dans la section 2.2.2 ont le meilleur rapport "temps/qualité" ;
- $n \geq 2$  avec  $u \gg 1$  et  $T \gg 1$ , il faut tester la convergence, obtenue par Piterbarg et Konakov (1995) seulement dans le cas stationnaire, vers la loi de Gumbel.

## Bibliographie de la section 2.2

**Adler, R. J.** (1981). *The geometry of random fields*. John Wiley & Sons Ltd., Chichester. Wiley Series in Probability and Mathematical Statistics.

**Adler, R. J. et Taylor, J.** (2005). *Random fields and geometry*. <http://iew3.technion.ac.il/radler/grf.pdf>. To appear.

**Azaïs, J.-M. et Delmas, C.** (2002). Asymptotic expansions for the distribution of the maximum of Gaussian random fields. *Extremes*, 5(2) :181-212 (2003).

- Azaïs, J.-M. et Wschebor, M.** (2002). The distribution of the maximum of a Gaussian process : Rice method revisited. In *In and out of equilibrium (Mambucaba, 2000)*, volume 51 of *Progr. Probab.*, pages 321–348. Birkhäuser Boston, Boston, MA.
- Berman, S. M.** (1971). Excursions above high levels for stationary Gaussian processes. *Pacific J. Math.*, 36 :63–79.
- Blake, I. F., et Lindsey, W. C.** (1973). Level-crossing problems for random processes. *IEEE Trans. Information Theory*, IT-19 :295–315.
- Brodtkorb, P., Johannesson, P., Lindgren, G., Rychlik, I., Rydén, J., et Sjö, E.** (2000). Wafo - a matlab toolbox for analysis of random waves and loads. In *Proceedings of the 10th International Offshore and Polar Engineering conference*, volume Vol III, 343–350, Seattle. <http://www.maths.lth.se/matstat/wafo/>.
- Cressie, N.** (1980). The asymptotic distribution of the scan statistic under uniformity. *Ann. Probab.*, 8(4) :828–840.
- Delmas, C.** (2001). *Distribution du maximum d'un champ aléatoire et applications statistiques*. PhD thesis, Université Paul Sabatier, Toulouse III.
- Delmas, C.** (2003). Projections on spherical cones, maximum of Gaussian fields and Rice's method. *Statist. Probab. Lett.*, 64(3) :263–270.
- DeLong, D. M.** (1981). Crossing probabilities for a square root boundary by a Bessel process. *Comm. Statist. A—Theory Methods*, 10(21) :2197–2213.
- Genz, A.** (1992). Numerical computation of multivariate normal probabilities. *J. Comp. Graph Stat.*, 1 :141–149.
- Goldman, M.** (1971). On the first passage of the integrated wiener process. *Ann. Math. Statist.*, 42(6) :2150–2155.
- Mercadier, C.** (2005). Numerical bounds for the distribution of one- and two-parameter processes. *Submitted*. Software <http://www.lsp.ups-tlse.fr/Fp/Mercadier/MAGP.html>.
- McKean, Jr., H. P.** (1963). A winding problem for a resonator driven by a white noise. *J. Math. Kyoto Univ.*, 2 :227–235.
- Piterbarg, V. I.** (1996). *Asymptotic methods in the theory of Gaussian processes and fields*, volume 148 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI. Translated from the Russian by V. V. Piterbarg, Revised by the author.
- Piterbarg, V. I. et Konakov, V. D.** (1995). High level excursions of Gaussian fields and the weakly optimal choice of the smoothing parameter. I. *Math. Methods Statist.*, 4(4) :421–434.
- Rychlik, I.** (1987). A note on Durbin's formula for the first-passage density. *Statist. Probab. Lett.*, 5(6) :425–428.



- Shepp, L. A.** (1971). First passage time for a particular gaussian process. *Ann. Math. Statist.*, 42 :946–951.
- Shepp, L. A.** (1979). The joint density of the maximum and its location for a Wiener process with drift. *J. Appl. Probab.*, 16(2) :423–427.
- Shepp, L. A. et Slepian, D.** (1976). First-passage time for a particular stationary periodic Gaussian process. *J. Appl. Probability*, 13(1) :27–38.
- Slepian, D.** (1961). First passage time for a particular Gaussian process. *Ann. Math. Statist.*, 32 :610–612.
- Sun, J.** (1993). Tail probabilities of the maxima of Gaussian random fields. *Ann. Probab.*, 21(1) :34–71.
- Takemura, A. et Kuriki, S.** (2002). On the equivalence of the tube and Euler characteristic methods for the distribution of the maximum of Gaussian fields over piecewise smooth domains. *Ann. Appl. Probab.*, 12(2) :768–796.
- Wong, E.**, (1970). The distribution of intervals between zeros for a stationary Gaussian process. *SIAM J. Appl. Math.*, 18 :67–73.
- Worsley, K. J.** (1995a). Boundary corrections for the expected Euler characteristic of excursion sets of random fields, with an application to astrophysics. *Adv. in Appl. Probab.*, 27(4) :943–959.
- Worsley, K. J.** (1995b). Estimating the number of peaks in a random field using the Hadwiger characteristic of excursion sets, with applications to medical images. *Ann. Statist.*, 23(2) :640–669.

# Chapitre 3

---

## RÉSUMÉ DE LA PARTIE III

Nous exposons les résultats décrits dans le premier chapitre de la partie III mettant en évidence l'intérêt statistique des études menées au cours de la partie II.

Le second chapitre de la partie III, autour du test d'une hypothèse sur la matrice de covariance, n'est pas résumé ici car il ne contient pas de résultat nouveau majeur ; il est rédigé en français.

### Le test du maximum de vraisemblance

Depuis un travail de Ghosh et Sen (1985) nous connaissons mieux les types de lois que peut rencontrer la distribution asymptotique de la statistique de test du maximum du rapport de la log-vraisemblance. Sous certaines conditions, elle est reliée à celle du maximum d'un ou plusieurs processus stochastiques, définis sur un ensemble plus ou moins complexe suivant le modèle considéré. Ce travail s'inscrit donc dans la continuité des résultats présentés dans les sections précédentes.

Cette statistique permet en outre de traiter le problème du nombre de composantes dans un modèle de mélange de distributions appartenant à une famille paramétrique. Dans le chapitre 6, nous nous sommes restreints à ce cadre avec deux objectifs clairement définis. Le premier est d'utiliser les résultats généraux de Gassiat (2002) afin de donner, explicitement et pour les mélanges les plus courants, les caractéristiques des processus gaussiens sous-jacents. Le second est le calcul de la puissance de ce test pour différents modèles lorsque l'ensemble de leurs paramètres est de taille très grande. Remarquons que les arguments utilisés reposent fortement sur l'ordre asymptotique du maximum d'un processus gaussien non-stationnaire, décrit dans la section 2.1 et développé dans le chapitre 4.

### 3.1 Introduction

Les modèles de mélange de distributions sont couramment utilisés pour analyser des données provenant de plusieurs populations homogènes. Par exemple en génétique où l'existence d'un gène majeur associé à un caractère quantitatif est une question importante. Lorsque les caractères biologiques étudiés sont continus, un gène est dit majeur lorsqu'il y a une différence d'un écart-type phénotypique entre les valeurs moyennes associées aux génotypes extrêmes. S'il existe, deux sous-populations sont alors induites par la présence ou l'absence, chez un individu, de ce gène. Plus généralement, ils sont

utilisés dans de nombreuses applications dont les principales sont présentées dans les livres de Titterington, Smith et Makov (1985) et Lindsay (1995).

Par souci de simplicité, on associe à chaque population un simple modèle paramétrique (par exemple la densité d'une loi usuelle) et par conséquent, chaque observation est regardée comme la réalisation de leur mélange.

Formellement soit  $\mathcal{F} = \{f_{\mathbf{t},\alpha}, \mathbf{t} \in \mathbb{T}, \alpha \in \mathbb{A}\}$  un ensemble paramétrique de densités par rapport à une mesure de référence  $\nu$ , avec

$\mathbb{T}$  un sous-ensemble compact de  $\mathbb{R}^k$ ,  
et  $\mathbb{A}$  un sous-ensemble compact de  $\mathbb{R}^h$ .

Définissons alors un mélange de  $p$  densités de la façon suivante :

$$g_{p,\pi,\mathbf{T},\alpha} = \sum_{i=1}^p \pi_i f_{\mathbf{t}^i,\alpha},$$

où  $\pi = (\pi_1, \dots, \pi_p)$  est un vecteur normalisé dont les composantes sont positives ou nulles,  $\mathbf{T} = (\mathbf{t}^1, \dots, \mathbf{t}^p) \in \mathbb{T}^p$  et  $\alpha \in \mathbb{A}$  est commun à chacune des  $p$  densités.

Ce type de modèle constitue un compromis entre un modèle paramétrique basique qui impose une trop lourde structure sur les données et un modèle non-paramétrique. Ici nous pouvons penser que la signification propre de chaque composante facilite l'estimation des paramètres. Ceci est d'autant plus vrai si le modèle est identifiable, c'est-à-dire si chacun de ses éléments n'admet qu'une seule représentation. Le modèle  $G_p$  contenant l'ensemble des mélanges  $g_{p,\pi,\mathbf{T},\alpha}$  n'est donc pas identifiable pour les paramètres  $\pi = (\pi_1, \dots, \pi_p)$  et  $\mathbf{T} = (\mathbf{t}^1, \dots, \mathbf{t}^p)$ . Cependant, nous allons travailler sous hypothèse d'identifiabilité de  $G_p$  au sens faible. Nous supposons que l'égalité presque partout

$$g_{p,\pi^0,\mathbf{T}_0,\alpha} = g_{p,\pi^1,\mathbf{T}_1,\alpha}$$

implique l'égalité des distributions

$$\sum_{i=1}^p \pi_i^0 \delta_{\mathbf{t}_0^i} = \sum_{i=1}^p \pi_i^1 \delta_{\mathbf{t}_1^i}.$$

Le principal problème lié aux modèles de mélange est le choix du nombre de composantes. Il est donc important de savoir tester si l'introduction d'une classe supplémentaire est justifiée. Autrement dit, en revenant à l'exemple génétique de la section 3.1, nous voulons déterminer la valeur de  $p$  :

- $p = 1$ , la population est homogène, le caractère étudié obéit seulement à un déterminisme polygénique, il n'y a pas de gène à effet majeur,
- $p = 2$ , la population résulte d'un croisement de deux sous-populations homogènes, dont l'une possède le gène majeur. Le caractère étudié est régi par un déterminisme mixte, où l'effet majeur s'ajoute à l'effet polygénique.

Différents tests ont été introduits pour analyser ces modèles. Citons la méthode des moments où la statistique de test est par exemple la moyenne ou bien la variance empirique. Citons aussi l'approche bayésienne, où les paramètres du modèle sont appréhendés comme des variables aléatoires. D'autre part, plusieurs travaux ont proposé une méthode basée sur une fonction contraste, le maximum de vraisemblance en fait

partie. Ainsi pour un échantillon  $X_1, \dots, X_n$  de densité inconnue  $g$ , nous souhaitons utiliser la statistique du test du rapport de maximum de log-vraisemblance (noté TMRV et LRT en anglais) pour tester l'hypothèse nulle

$$(H_0) : "g \text{ est un mélange de } p_0 \text{ populations}"$$

contre l'hypothèse alternative

$$(H_1) : "g \text{ est un mélange de } p \text{ populations}"$$

avec  $p_0 < p$ . Introduisons les ensembles contenant les distributions associées :

$$\begin{aligned} \mathcal{M}_0 &= \left\{ g_{p_0, \pi_0, \mathbf{T}_0, \alpha}, \pi_0 \in [0, 1]^{p_0}, \mathbf{T}_0 \in \mathbb{T}^{p_0}, \alpha \in \mathbb{A}, \sum_{i=1}^{p_0} \pi_i = 1, i = 1, \dots, p_0 \right\}, \\ \mathcal{M} &= \left\{ g_{p, \pi, \mathbf{T}, \alpha}, \pi \in [0, 1]^p, \mathbf{T} \in \mathbb{T}^p, \alpha \in \mathbb{A}, \sum_{i=1}^p \pi_i = 1, i = 1, \dots, p \right\}. \end{aligned}$$

En notant  $\ell_n(g) = \sum_{i=1}^n \log g(X_i)$  la log-vraisemblance, la statistique du TRMV est définie de la manière suivante :

$$\lambda_n = \sup_{g \in \mathcal{M}} \ell_n(g) - \sup_{g \in \mathcal{M}_0} \ell_n(g).$$

Rappelons que pour des modèles réguliers (ce n'est pas le cas des modèles de mélange que nous considérons) et sous certaines conditions,  $\lambda_n$  converge en distribution sous l'hypothèse nulle vers

$$\frac{1}{2} \chi_{N-N_0}^2$$

où  $N$  et  $N_0$  sont les dimensions respectives de  $\mathcal{M}$  et  $\mathcal{M}_0$ .

## 3.2 Classes de Donsker

Cette section s'inspire de la partie 19.2 du livre de van der Vaart (1998) sur les distributions empiriques. Considérons  $X_1, \dots, X_n$  un échantillon de distribution  $\mathbb{P}$ . Pour une fonction mesurable  $f$ , nous notons  $\mathbb{P}_n f$  l'espérance de  $f$  sous la mesure empirique  $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$  et  $\mathbb{P} f$  son espérance sous la loi  $\mathbb{P}$ ,

$$\mathbb{P}_n f = n^{-1} \sum_{i=1}^n f(X_i) \quad \mathbb{P} f = \int f d\mathbb{P}.$$

Le processus empirique au point  $f$  est défini par

$$\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n f - \mathbb{P} f).$$

En appliquant le théorème de la limite centrale à un ensemble fini de fonctions mesurables  $f_i$  telles que  $\mathbb{P} f_i^2 < +\infty$  nous avons

$$(\mathbb{G}_n f_1, \dots, \mathbb{G}_n f_k) \rightarrow (\mathbb{G}_{\mathbb{P}} f_1, \dots, \mathbb{G}_{\mathbb{P}} f_k),$$

où le vecteur limite admet une loi normale multivariée centrée de covariance

$$\mathbb{E}(\mathbb{G}_{\mathbb{P}} f \mathbb{G}_{\mathbb{P}} g) = \mathbb{P} f g - \mathbb{P} f \mathbb{P} g.$$

L'intérêt d'une classe  $\mathcal{F}$  (de fonctions  $f$ ) dite  $\mathbb{P}$ -Donsker est d'avoir ce résultat de façon uniforme. La suite de processus  $\{\mathbb{G}_n f, f \in \mathcal{F}\}$  converge en distribution vers un processus gaussien  $\mathbb{G}_{\mathbb{P}}$  centré de fonction de covariance déjà donnée.

Regardons maintenant une caractérisation des classes de Donsker. Pour deux fonctions  $l$  et  $u$ , le crochet  $[l, u]$  est l'ensemble des fonctions  $f \in \mathcal{F}$  tel que  $l \leq f \leq u$ . Un  $\varepsilon$ -crochet dans  $L^r(\mathbb{P})$  est un crochet  $[l, u]$  avec  $\mathbb{P}(l - u)^r < \varepsilon^r$ . Nous notons  $N_{\mathcal{F}, \square, r}(\varepsilon)$  le nombre de  $\varepsilon$ -crochets nécessaires pour recouvrir  $\mathcal{F}$ . L'entropie entre crochet  $H_{\mathcal{F}, \square, r}(\varepsilon)$  est alors le logarithme de ce nombre.

### Théorème 3.2.1 (van der Vaart)

Une classe  $\mathcal{F}$  de fonctions mesurables est  $\mathbb{P}$ -Donsker dès que  $\int_0^1 \sqrt{H_{\mathcal{F}, \square, r}(\varepsilon)} d\varepsilon < +\infty$ .

EXEMPLE POUR UNE CLASSE PARAMÉTRIQUE.

Soit  $\mathcal{F} = \{f_{\theta}, \theta \in \Theta\}$  un ensemble de fonctions mesurables indexé par un sous-ensemble borné  $\Theta$  de  $\mathbb{R}^d$ . Supposons qu'il existe une fonction mesurable  $m$  telle que pour tout  $\theta_1$  et  $\theta_2$

$$|f_{\theta_1}(x) - f_{\theta_2}(x)| \leq m(x) \|\theta_1 - \theta_2\|.$$

Si  $\mathbb{P}|m|^r < +\infty$  alors il existe une constante  $K$  dépendante de  $\Theta$  et  $d$  seulement, telle que

$$N_{\mathcal{F}, \square, r}(\varepsilon (\mathbb{P}|m|^r)^{1/r}) \leq K \left( \frac{\text{diam} \Theta}{\varepsilon} \right)^d$$

pour tout  $0 < \varepsilon < \text{diam} \Theta$ .

Par suite, l'entropie est d'ordre plus petit que  $\log(1/\varepsilon)$ .

Pour établir la borne supérieure, nous utilisons des crochets du type  $[f_{\theta} - \varepsilon m, f_{\theta} + \varepsilon m]$  avec  $\theta$  dans un sous-ensemble de  $\Theta$ . Le diamètre de ce crochet est  $2\varepsilon (\mathbb{P}|m|^r)^{1/r}$ . Si  $\theta$  parcourt un  $\varepsilon$ -réseau de  $\Theta$ , alors les crochets recouvrent  $\mathcal{F}$  grâce à la condition de Lipschitz, si  $\|\theta_1 - \theta_2\| \leq \varepsilon$ ,  $f_{\theta_1} - \varepsilon m \leq f_{\theta_2} \leq f_{\theta_1} + \varepsilon m$ . Nous avons alors besoin de boules de rayon  $\varepsilon/2$  pour recouvrir  $\Theta$ .

## 3.3 Lien avec le maximum d'un processus stochastique

La théorie asymptotique classique rappelée à la fin de la section 3.1 ne s'applique pas sous nos hypothèses. En particulier la non-identifiabilité implique une dégénérescence de la matrice d'information de Fisher et par suite l'impossibilité de borner de façon uniforme le reste du développement de Taylor de la log-vraisemblance.

Ghosh et Sen (1985) sont les premiers à déterminer un ensemble d'hypothèses (comprenant une forte hypothèse de séparation des populations) sous lequel une autre distribution limite de la statistique du TRMV est obtenue.

Dacunha-Castelle et Gassiat (1997) ont affaibli leurs hypothèses et donné des généralisations. Leur idée est de paramétrer le modèle avec deux paramètres de sorte que

le premier, mesurant la distance à la vraie densité, soit identifiable sous l'hypothèse nulle. Ils l'appellent la paramétrisation localement conique. Naturellement elle perd, par rapport aux résultats obtenus sous hypothèses de régularité, l'asymptotique consistante du deuxième paramètre. L'hypothèse centrale de leur travail est de supposer que la fermeture des dérivées de la log-vraisemblance constitue une classe de Donsker de variables aléatoires. La distribution asymptotique est alors une fonction du supremum d'un processus gaussien. Remarquons que le cadre qu'ils utilisent est plus large que les modèles de mélange. Dacunha-Castelle et Gassiat montrent, à travers l'application à des modèles particuliers, que ce résultat permet de traiter des tests aussi variés que celui de l'hypothèse  $(H_0)$  contre  $(H_1)$  présenté au milieu de la section 3.1.

Cependant, nous avons choisi d'appliquer les résultats de Gassiat (2002). Il s'agit d'une généralisation du travail de Dacunha-Castelle et Gassiat dans laquelle la construction d'une paramétrisation n'est que sous-jacente.

Soit un échantillon  $(X_1, \dots, X_n)$  de densité inconnue  $f_0$  appartenant à un ensemble  $\mathcal{G}$  de densités par rapport à une mesure positive  $\nu$ . Définissons, pour  $\|\cdot\|_2$  la norme dans  $L^2(f_0\nu)$ , l'ensemble des scores

$$\mathcal{S} = \left\{ s_g = \frac{(g - f_0)/f_0}{\|(g - f_0)/f_0\|_2}, g \in \mathcal{G} \right\}.$$

Introduisons aussi l'ensemble des scores  $\mathcal{D}$  comme l'ensemble des fonctions  $d \in L^2(f_0\nu)$  telles que nous puissions trouver une suite  $g_n$  dans  $\mathcal{G}$  satisfaisant :  $\|(g_n - f_0)/f_0\|_2$  et  $\|d - s_{g_n}\|_2$  tendent vers 0. Notons  $\mathbb{P}_n$  l'ensemble des probabilités  $(g_{c/\sqrt{n}}\nu)^{\otimes n}$  où  $g_{c/\sqrt{n}}\nu$  est dans un voisinage d'ordre  $1/\sqrt{n}$  de  $f_0$ . Notons  $d_0$  son score associé. Enfin soit  $\{Z(d), d \in \mathcal{D}\}$  le processus gaussien centré de fonction de covariance le produit scalaire dans  $L^2(f_0\nu)$ .

**Théorème 3.3.1 (Gassiat)**

Supposons que  $\int_0^1 \sqrt{H_{\mathcal{S},[\cdot],2}(\varepsilon)} d\varepsilon < +\infty$ . Alors nous avons

- $\sup_{g \in \mathcal{G}} (\ell_n(g) - \ell_n(f_0)) = \frac{1}{2} \sup_{d \in \mathcal{D}} (\max\{\frac{1}{\sqrt{n}} \sum_{i=1}^n d(X_i), 0\})^2 + o_{\mathbb{P}_n}(1)$ ,
- $(f_0\nu)^{\otimes n}$  et  $\mathbb{P}_n$  sont mutuellement contiguës,
- sous  $\mathbb{P}_n$ ,  $\sup_{g \in \mathcal{G}} (\ell_n(g) - \ell_n(f_0))$  converge vers

$$\frac{1}{2} \sup_{d \in \mathcal{D}} (\max\{Z(d) + c \int dd_0 f_0 d\nu, 0\})^2.$$

Ce théorème s'applique directement au cas où l'hypothèse nulle consiste en un point unique, la vraie densité  $f_0$ . Pour des hypothèses nulles plus complexes, nous écrivons

$$\lambda_n = \sup_{g \in \mathcal{M}} \ell_n(g) - \sup_{g \in \mathcal{M}_0} \ell_n(g) = \sup_{g \in \mathcal{M}} (\ell_n(g) - \ell_n(f_0)) - \sup_{g \in \mathcal{M}_0} (\ell_n(g) - \ell_n(f_0)).$$

En utilisant à deux reprises le théorème 3.3.1 nous déduisons que la distribution asymptotique de  $2\lambda_n$  est donnée via la différence de carrés de suprema de processus gaussiens.

Dans la section 6.2 du chapitre 6 nous avons explicité, voire simplifié, l'expression de la distribution asymptotique de  $\lambda_n$  pour différents modèles de mélange.

• MÉLANGE SIMPLE (section 6.2.1) :  $\mathcal{M}_0 = \{f_{\mathbf{0}}\}$  et pour  $\mathbb{T}$  un sous-ensemble compact de  $\mathbb{R}^k$  dont l'intérieur contient  $\mathbf{0}$ ,

$$\mathcal{M} = \{g_{\pi, \mathbf{t}} = (1 - \pi)f_{\mathbf{0}} + \pi f_{\mathbf{t}}, 0 \leq \pi \leq 1, \mathbf{t} \in \mathbb{T}\}.$$

Plusieurs exemples précis sont donnés, pour lesquels les hypothèses sont réduites,

- le mélange de translation,  $f_{\mathbf{t}}(\cdot) = f_{\mathbf{0}}(\cdot - \mathbf{t})$ ,
- le mélange gaussien où  $f_{\mathbf{t}}$  est la densité d'une loi normale multivariée  $\mathcal{N}(\mathbf{t}, I)$ ,
- le mélange binomial avec  $f_{\mathbf{t}} \sim \text{Bi}(k, \theta_0 + t)$
- le mélange dans une famille exponentielle ; pour une statistique exhaustive  $T(x)$  prenons

$$f_{\mathbf{t}}(x) = f_{\mathbf{0}}(x) \exp\left(\sum_{i=1}^k t_i T_i(x) - \phi(\mathbf{t})\right).$$

• TEST D'UNE POPULATION CONTRE DEUX (section 6.2.2) :  $p_0 = 1$  et  $p = 2$ . Nous déduisons la loi asymptotique de  $\lambda_n$  d'une expression dépendant du supremum d'un seul processus gaussien. Nous appliquons de nouveau les résultats aux lois particulières considérées dans le cas du mélange simple.

• MÉLANGE SIMPLE AVEC UN PARAMÈTRE DE NUISANCE COMMUN AUX DEUX ALTERNATIVES (section 6.2.3). Soit  $\mathbb{A}$  un sous-ensemble compact de  $\mathbb{R}^h$ ,  $\mathcal{M}_0 = \{f_{\mathbf{0}, \alpha}, \alpha \in \mathbb{A}\}$ , et

$$\mathcal{M} = \{g_{\pi, \mathbf{t}, \alpha} = (1 - \pi)f_{\mathbf{0}, \alpha} + \pi f_{\mathbf{t}, \alpha}, 0 \leq \pi \leq 1, \mathbf{t} \in \mathbb{T}, \alpha \in \mathbb{A}\}.$$

La réduction de la distribution asymptotique de  $\lambda_n$  à une fonctionnelle d'un seul processus gaussien est de nouveau établie. Nous reprenons l'étude de la translation et de la loi gaussienne de variance inconnue.

• TEST DE  $p_0$  POPULATIONS CONTRE  $p = p_0 + 1$  POPULATIONS (section 6.2.4). La loi asymptotique de  $\lambda_n$  est exprimée sous l'hypothèse nulle grâce au supremum d'un processus gaussien. Nous commentons le cas d'un test général de  $p_0$  populations contre  $p$  populations sous hypothèse nulle et hypothèses contiguës.

Cette étude, basée sur le cadre abstrait obtenu par Gassiat, analyse des modèles assez différents par une même méthode. De plus, l'étude de lois ou de nuisance particulières apporte un intérêt synthétique à ce travail.

### 3.4 Un test non uniformément le plus puissant

Il n'y a aucune raison définitive pour qualifier la méthode de maximum de vraisemblance d'optimale. Cette opinion est cependant très répandue. L'idée de son efficacité s'est construite sur le grand succès avec lequel elle a été appliquée à de nombreux problèmes concrets, et partiellement sur des propriétés asymptotiques d'optimalité rigoureusement établies.

Revenons par exemple aux applications dans le domaine de la génétique, déjà évoquées. La connaissance expérimentale permet d'affiner la représentation des modèles de mélanges testés.

- Thode, Finch et Mendell (1988) montrent numériquement que l'approximation provenant de la théorie classique est satisfaisante dans leur problème dès que la taille de l'échantillon dépasse 1000 individus.
- Dans le papier de Goffinet, Loisel et Laurent (1992) le poids  $\pi$  est connu. Ils étudient la puissance de la statistique du TRMV et montrent que celle-ci reste bonne dans le cas où la proportion n'est estimée qu'à  $10^{-1}$  près. Des hypothèses semblables sont supposées dans l'article de Loisel, Goffinet, Monod et Montes de Oca (1994) pour le test d'une population contre trois.

Il faut noter que les distributions asymptotiques de la log-vraisemblance sont alors "entre la théorie classique et la théorie donnée dans la section précédente".

Il est clair que ces modèles sont justifiés par de simples considérations génétiques en amont.

Malgré la souplesse d'adaptation de ce test, il est important de calculer sa puissance. Pour cela, nous regardons l'hypothèse nulle et une hypothèse alternative qui sont mutuellement contiguës. Plus précisément, pour le cas d'un mélange simple avec  $\mathbb{T} = [-T, T]$ , considérons  $\pi_n$  et  $t_n$  des suites vérifiant :

$$\lim_{n \rightarrow +\infty} \sqrt{n\pi_n} \|(f_{t_n} - f_0)/f_0\|_2 = c, \text{ pour une constante positive } c.$$

Par conséquent,

soit  $t_n$  tend vers  $t_0 \neq 0$  et  $\sqrt{n\pi_n}$  tend vers une constante positive,

soit  $t_n$  tend vers 0, et  $t_n/\|t_n\|$  converge vers une limite notée  $\tau$ .

Notons alors

$$\mathbb{P}_{\pi_n, t_n} = \left( g_{\pi_n, t_n} \cdot \nu \right)^{\otimes n} \text{ et } \mathbb{P}_0 = \left( f_0 \cdot \nu \right)^{\otimes n}.$$

Dans le cas où l'ensemble des paramètres  $\mathbb{T}$  est un compact de  $\mathbb{R}$ , nous connaissons les lois asymptotiques de  $\lambda_n$  sous ces hypothèses. La puissance n'étant pas triviale nous nous demandons ce qu'il en est lorsque l'ensemble des paramètres  $\mathbb{T}$  a une taille qui devient de plus en plus grande. Nous recherchons l'effet provoqué, sur la puissance asymptotique de ce test, par  $T \rightarrow +\infty$ .

Dans le théorème 6.3.1 nous supposons que les hypothèses associées au mélange simple sont vérifiées. Nous savons alors que la statistique  $\lambda_n$  admet une loi asymptotique sous l'hypothèse nulle et sous alternative contiguë données par l'intermédiaire d'un processus gaussien. En supposant de plus que ces processus satisfont les hypothèses du corollaire 4.2.1, nous parvenons à calculer la puissance asymptotique du test.



Lorsque  $T$  diverge vers  $+\infty$ , une normalisation du maximum du processus limite converge en distribution, sous les deux types d'hypothèses, vers la même loi (la distribution de Gumbel). Autrement dit, lorsque  $c_{T,\alpha}$  est le seuil du test défini par

$$\lim_{n \rightarrow +\infty} \mathbb{P}_0(\lambda_n > c_{T,\alpha}) = \alpha,$$

alors pour toute alternative contiguë, la puissance limite du TRMV est égal à son niveau :

$$\lim_{T \rightarrow +\infty} \lim_{n \rightarrow +\infty} \mathbb{P}_{\pi_n, t_n}(\lambda_n > c_{T,\alpha}) = \alpha.$$

Nous avons explicité l'ensemble d'hypothèses  $(H)$  pour le problème de mélange de translation dans le corollaire 6.3.1.

$$(H1) \quad \forall x \in \mathbb{R}, f_0(x) > 0, f_0 \text{ est quatre fois continûment différentiable,} \\ \text{et } \forall i = 1, \dots, 4, \exists K_i > 0, \forall x \in \mathbb{R}, \left| \frac{f_0^{(i)}}{f_0}(x) \right| \leq K_i,$$

$$(H2) \quad \forall x \in \mathbb{R}, \lim_{t \rightarrow +\infty} \frac{f_0(x+t)}{f_0(t)} = \lim_{t \rightarrow -\infty} \frac{f_0(x+t)}{f_0(t)} = 1,$$

$$(H3) \quad \exists M > 0, \forall x, t \in \mathbb{R}, \frac{f_0(x)f_0(t)}{f_0(x+t)} \leq M,$$

$$(H4) \quad \exists F \in L^2(\lambda) : \sup_{|t| \geq 1} \log |t| \sqrt{f_0(x+t)} \leq F(x),$$

$$(H5) \quad \lim_{t \rightarrow +\infty} \log(t) \sqrt{f_0(t)} = 0.$$

Ainsi, lorsqu'un modèle de translation est construit à partir d'une densité  $f_0$  vérifiant les conditions  $(H)$ , nous pouvons affirmer que la puissance asymptotique du TRMV devient, avec la valeur de  $T$ , aussi proche que nous le souhaitons du niveau asymptotique. Autrement dit, le TRMV ne distingue plus l'hypothèse nulle de l'hypothèse alternative.

Nous avons ensuite considéré un modèle de mélange différent de ceux présentés jusqu'ici. Il s'agit du mélange simple gaussien dans lequel aucune hypothèse n'est faite sur le domaine  $\mathbb{T}$ . Plus précisément, nous prenons  $\mathbb{T} = \mathbb{R}$  et des suites  $\pi_n, t_n$  telles que  $\lim_{n \rightarrow +\infty} \sqrt{n} \pi_n t_n = \gamma \in \mathbb{R}$  et  $\lim_{n \rightarrow +\infty} t_n = t_0 \in \mathbb{R}$ .

Sans aucun a priori sur l'ensemble des paramètres il est clair que nous ne pouvons pas appliquer le résultat de Dacunha-Castelle et Gassiat (1997) ou même celui de Gassiat (2002). L'idée est de normaliser directement la statistique  $\lambda_n$  pour avoir une loi limite non triviale. Cette normalisation est donnée sous l'hypothèse nulle par le résultat suivant.

### **Théorème 3.4.1 (Liu et Shao)**

Lorsque  $n$  tend vers  $+\infty$ ,  $2 \lambda_n - \log \circ \log n + \log(2\pi^2)$  converge en distribution vers la distribution de Gumbel sous l'hypothèse nulle.

Dans le Théorème 6.4.1, nous affirmons que cette convergence reste vraie sous alternatives contiguës. *En définissant la région de rejet comme :  $(\lambda_n > c_{\alpha,n})$  avec*

$$c_{\alpha,n} = \frac{1}{2}(G^{1-\alpha} + \log \circ \log n - \log(2\pi^2)),$$

où  $G^{1-\alpha}$  est le fractile d'ordre  $1 - \alpha$  de la loi de Gumbel, nous avons par définition

$$\lim_{n \rightarrow +\infty} \mathbb{P}_0(\lambda_n > c_{\alpha,n}) = \alpha.$$

Et pour tout  $\gamma$  et  $t_0$ , la puissance limite du TRMV vaut

$$\lim_{n \rightarrow +\infty} \mathbb{P}_{\pi_n, t_n}(\lambda_n > c_{\alpha,n}) = \alpha.$$

L'idée de la démonstration de ce résultat est d'appliquer la généralisation du troisième lemme de Le Cam établie par van der Vaart (1998) associée à une approximation forte donnée par Bickel et Chernoff (1993).

Ces derniers résultats entraînent que le TRMV est asymptotiquement moins puissant que les tests de moments par exemple. Cependant la vitesse des convergences utilisées implique que cela n'est vrai que pour des échantillons de taille très grande. Nous nous remettons donc pas en cause l'utilisation pratique du TRMV. Ce travail doit être vu comme un plaidoyer en faveur de son étude théorique dans un cadre compact.

## Bibliographie du chapitre 3

- Azaïs, J.-M., Gassiat, E., et Mercadier, C.** (2004). Asymptotic distribution and power of the likelihood ratio test for mixtures : bounded and unbounded cases. *Submitted*.
- Bickel, P. et Chernoff, H.** (1993). Asymptotic distribution of the likelihood ratio statistic in a prototypical non regular problem. *Statistics and Probability : A Raghu Raj Bahadur Festschrift. Ghosh et al. Eds. Wiley Eastern Limited, Publishers. 83-96*.
- Dacunha-Castelle, D. et Gassiat, É.** (1997). Testing in locally conic models, and application to mixture models. *ESAIM Probab. Statist.*, 1 :285–317.
- Gassiat, É.** (2002). Likelihood ratio inequalities with applications to various mixtures. *Ann. Inst. H. Poincaré Probab. Statist.*, 6 :897–906.
- Ghosh, J. et Sen, P.** (1985). On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983)*, Wadsworth Statist./Probab. Ser., pages 789–806, Belmont, CA. Wadsworth.
- Goffinet, B., Loisel, P., et Laurent, B.** (1992). Testing in normal mixture models when the proportions are known. *Biometrika*, 79(4) :842–846.

- Lindsay, B. G.** (1995). *Mixture Models : Theory, Geometry and Applications*. NSF-CBMS Series. Institute of Mathematical Statistics, Hayward.
- Liu, X. et Shao, Y.** (2000). Asymptotic distributions of the likelihood ratio test in two-component normal mixture models. *Ann. Statist.*
- Loisel, P., Goffinet, B., Monod, H., et Montes de Oca, G.** (1994). Detecting a major gene in a f2 population. *Biometrics*, 50 :512–516.
- Thode, H.C., Finch, S.J. et Mendell** (1988). Simulated percentage points for the null distribution of the likelihood ratio test for a mixture of two normals. *Biometrics*, 44 :1195–1201.
- Titterington, D. M., Smith, A. F. M., et Makov, U. E.** (1985). *Statistical analysis of finite mixture distributions*. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics. John Wiley & Sons Ltd., Chichester.
- van der Vaart, A. W.** (1998). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

---

---

DEUXIÈME PARTIE

DISTRIBUTION DU MAXIMUM  
DE PROCESSUS

---

---



# Chapitre 4

---

## CARACTÈRE ASYMPTOTIQUEMENT POISSONNIEN DES EXTRÊMES DANS DES MODÈLES GAUSSIENS NON-STATIONNAIRES

Ce chapitre présente dans sa version originale l'article "Asymptotic Poisson character of extremes in non-stationary Gaussian models" d'Azaïs et Mercadier (2003) dont les références sont données dans la section bibliographique.

Considérons  $X$  un processus gaussien non-stationnaire, asymptotiquement centré avec une variance constante. Nous étudions  $U_u(t)$  le nombre de upcrossings de niveau  $u$  par le processus  $X$  sur l'intervalle  $(0, t]$ . Sous certaines conditions nous montrons que la suite de processus ponctuels  $(U_u)_{u>0}$  converge faiblement, après normalisation, vers un processus de Poisson standard, quand  $u$  tend vers l'infini. Comme conséquence à cette étude, nous obtenons la faible convergence du supremum renormalisé vers une distribution de Gumbel.

### Abstract

Let  $X$  be a non-stationary Gaussian process, asymptotically centered with constant variance. We study  $U_u(t)$  the number of upcrossings of level  $u$  by the process  $X$  on the interval  $(0, t]$ . Under some conditions we prove that the sequence of point processes  $(U_u)_{u>0}$  converges weakly, after normalization, to a standard Poisson process as  $u$  tends to infinity. In consequence of this study we obtain the weak convergence of the normalized supremum to a Gumbel distribution.

### 4.1 Introduction

Since the heuristic consideration of S. Rice and the development of V. A. Volkonskiï and J. A. Rozanov, one knows that high excursions of a regular stationary Gaussian process have a limit distribution (after renormalization of time). More precisely if the excursions beyond the level  $u$  are identified by their starting point (called upcrossings of level  $u$ ) this point process converges weakly, as  $u$  tends to infinity, to a Poisson process. Several proofs of such kind of results have been given with increasingly more

general hypotheses or original arguments by Volkonskiĭ and Rozanov (1961), Cramér and Leadbetter (1967), Berman (1971a), Leadbetter et al. (1983) and Wschebor (1985). However all these works rely on the stationary assumption which sometimes is not very well adapted to some modelling situations.

The aim of the present paper is to show that in fact the assumption of constant variance, instead of stationarity, is sufficient to obtain the same type of convergence (Theorem 4.2.3). In addition we show that the process does not need to be exactly centered : the first result remains true for a process with an expectation that tends to zero at a certain speed. Classically we obtain, as a corollary of our result, that a normalization of the supremum converges weakly to a Gumbel distribution (Corollary 4.2.1). Related studies were conducted by Piterbarg and Konakov (1995) for homogeneous and isotropic Gaussian fields under the assumption of compact supported covariance.

The key of our development is to begin the study in a simple case when the variance of the derived process is also supposed to be constant (Theorem 4.2.2). Next we define the "unit-speed" transformation of time to come down to such context. Furthermore, we use some improvements with respect to known tools :

- ★ a sharp evaluation of the distribution of the supremum on intervals which sizes tend to  $+\infty$  at a certain speed (Lemma 4.4.1),
- ★ the normal comparison Lemma in a version due to Li and Shao (2002) (Proposition 4.3.2).

The organization of the paper is as following. In Section 2 we state our main results. Principal tools for their demonstrations are reminded in Section 3. Section 4 is devoted to the proofs.

### 4.1.1 Notations and hypotheses

We consider  $X = \{X_t, t \in \mathbb{R}\}$  a Gaussian process. Denote by  $m$  its expectation function and  $r$  its covariance function. Assume that the variance of  $X$  is constant, without loss of generality we can take it equal to one, for all  $t$  in  $\mathbb{R}$ ,  $r(t, t) = \text{Var}(X_t) = 1$ . The partial derivatives of  $r$  are given for all  $(i, j)$  in  $\mathbb{N}^2$  by  $r_{ij}(s, t) = \frac{\partial^{i+j}}{\partial s^i \partial t^j} r(s, t)$ . Throughout the paper  $X$  will satisfy one of the two sets of hypotheses expressed below.

- The "unit-speed" set :

$$(U1) \quad \forall t \in \mathbb{R}, \quad r_{11}(t, t) = 1,$$

$$(U2) \quad r(s, t) \log |s - t| \xrightarrow{|s-t| \rightarrow +\infty} 0,$$

$$(U3) \quad \forall \varepsilon > 0 \quad \sup_{|s-t| > \varepsilon} |r(s, t)| < 1,$$

$$(U4) \quad r \text{ is of class } \mathcal{C}^4 \\ \forall \gamma > 0, r_{01} \text{ and } r_{04} \text{ are bounded on } \{(s, t) \in \mathbb{R}^2, s > \gamma \text{ and } t > \gamma\},$$

$$(U5) \quad \sqrt{\log(t)} m(t) \xrightarrow[t \rightarrow +\infty]{} 0,$$

$$(U6) \quad m'(t) \xrightarrow[t \rightarrow +\infty]{} 0.$$

• In general case we define :

$$\star \forall t \in \mathbb{R}, \quad S(t) = \int_0^t \sqrt{r_{11}(s, s)} ds,$$

$\star Y = \{Y_t = X_{S^{-1}(t)}, t \in \mathbb{R}\}$  the “unit-speed” transformation of  $X$  and  $r^Y$  its covariance function.

Then the general group of hypotheses on  $X$  consists in assuming that the transformed process  $Y$  satisfies Conditions (U2, ..., U6) since by construction  $Y$  verifies (U1). It consists of :

$$(G1) \quad \forall t \in \mathbb{R}, \quad r_{11}(t, t) > 0 \quad \text{and} \quad \lim_{t \rightarrow +\infty} S(t) = +\infty,$$

$$(G2) \quad r(s, t) \log |S(s) - S(t)| \xrightarrow[|S(s)-S(t)| \rightarrow +\infty]{} 0,$$

$$(G3) \quad \forall \varepsilon > 0 \quad \sup_{|S(s)-S(t)| > \varepsilon} |r(s, t)| < 1,$$

$$(G4) \quad r \text{ is of class } \mathcal{C}^4 \text{ with } s \rightarrow r_{11}(s, s) \text{ of class } \mathcal{C}^3, \\ \forall \gamma > 0, r_{01}^Y \text{ and } r_{04}^Y \text{ are bounded on } \{(s, t) \in \mathbb{R}^2, s > \gamma \text{ and } t > \gamma\},$$

$$(G5) \quad \sqrt{\log \circ S(t)} m(t) \xrightarrow[t \rightarrow +\infty]{} 0,$$

$$(G6) \quad \frac{m'(t)}{\sqrt{r_{11}(t, t)}} \xrightarrow[t \rightarrow +\infty]{} 0.$$

$\star \varphi$  is the standard Gaussian density and  $\Phi$  its distribution function. Set  $\bar{\Phi} = 1 - \Phi$ .

$\star$  For a positive real number  $u$  and a process  $X$  with  $\mathcal{C}^1$ -sample paths, we define the point process of upcrossings by  $X$  of level  $u$  :

$$\forall B \subset \mathbb{R}, \quad U_u(B) = \# \{t \in B, X_t = u, X'_t > 0\}.$$

$\star$  For any value of  $u$ , we define  $C_u$  as the asymptotic mean number of upcrossings on any interval of size one under “unit-speed” parameterization :

$$C_u = \frac{1}{2\pi} \exp\left(-\frac{u^2}{2}\right) = \frac{1}{\sqrt{2\pi}} \varphi(u).$$

$\star$  For  $t > 0$  and  $B \subset \mathbb{R}$ , we define  $M_t = \sup_{s \in (0, t)} X_s$  and  $M_B = \sup_{s \in B} X_s$ .



★ We write  $|B|$  the length of  $B \subset \mathbb{R}$ .

★ We define on  $(0, +\infty)$  the two normalizing functions

$$a_t = \sqrt{2 \log \circ S(t)} \text{ and } b_t = \sqrt{2 \log \circ S(t)} - \frac{\log(2\pi)}{\sqrt{2 \log \circ S(t)}}.$$

★ The standard Poisson process is, by definition, the Poisson process associated to Lebesgue measure.

★ Throughout this paper ( $Const$ ) symbolizes a generic positive constant. Its value may change from one occurrence to another.

## 4.2 Main results

We start by reminding the result within the stationary framework, as it is stated in Leadbetter et al. (1983).

### Theorem 4.2.1

Let  $X = \{X_t, t \in \mathbb{R}\}$  be a zero-mean stationary Gaussian process with unit variance. Assume that the covariance function  $\Pi$  satisfies  $\Pi(t) \log(t) \xrightarrow[t \rightarrow +\infty]{} 0$ , and that the second spectral moment  $\lambda_2 = \Pi''(0)$  is finite. Define the family of point processes  $R_u$  by

$$\forall B \subset \mathbb{R}, \quad R_u(B) = U_u(C_u^{-1} \lambda_2^{-\frac{1}{2}} B).$$

Then as  $u \rightarrow +\infty$ ,  $R_u$  converges weakly to a standard Poisson process.

By comparison our results are :

### Theorem 4.2.2

Let  $X = \{X_t, t \in \mathbb{R}\}$  be a Gaussian process with unit variance and suppose that Assumptions (U1-U6) are satisfied. Define the family of point processes  $R_u$  by

$$\forall B \subset \mathbb{R}, \quad R_u(B) = U_u(C_u^{-1} B).$$

Then as  $u \rightarrow +\infty$ ,  $R_u$  converges weakly, on  $(0, +\infty)$ , to a standard Poisson process.

### Theorem 4.2.3

Let  $X = \{X_t, t \in \mathbb{R}\}$  be a Gaussian process with unit variance and suppose that Assumptions (G1-G6) are satisfied. Define the family of point processes  $R_u$  by

$$\forall B \subset \mathbb{R}, \quad R_u(B) = U_u(S^{-1}(C_u^{-1} B)).$$

Then as  $u \rightarrow +\infty$ ,  $R_u$  converges weakly, on  $(0, +\infty)$ , to a standard Poisson process.

Comments :

★ If  $\sqrt{\log(t)} m(t) \rightarrow \tau$  instead of (U5) and  $\sqrt{\log \circ S(t)} m(t) \rightarrow \tau$  instead of (G5),  $R_u$  converges weakly on  $(0, +\infty)$  as  $u \rightarrow +\infty$  to an homogeneous Poisson process with intensity  $e^\tau dx$ . Arguments of the demonstration are the same as those used for the case  $\tau = 0$ .

★ For a stationary process, Conditions (U1), second statement of (U4), (U5) and (U6) are obviously satisfied. Moreover in this case (U3) is a direct consequence of (U2). Thus our hypotheses are comparable to those of Leadbetter, Lindgren and Rootzén.

★ Remark that under (U1) and stationary hypotheses we get  $\lambda_2 = r_{11}(t, t) = 1$ . Hence we still look at the same family  $(R_u)_{u>0}$ .

★ The restriction to the positive half line is due to the bound (U5) or (G5) on  $m$  that makes sense only for  $t$  positive. For zero-mean processes the convergence of the family of point processes  $R_u$  is valid on  $\mathbb{R}$ .

★ Of course by space-reversibility the same kind of theorem holds true for  $u$  that tends to  $-\infty$ .

★ Note that hypotheses (U4) and (G4) imply that the process  $X$  admits a modification with  $\mathcal{C}^1$ -sample paths, so the number of upcrossings is well defined.

As it is well known the asymptotic Poisson character implies that the supremum tends weakly to a Gumbel distribution.

### Corollary 4.2.1

Let  $X = \{X_t, t \in \mathbb{R}\}$  be a Gaussian process with unit variance satisfying (G1-G5). Then

$$\mathbb{P}(a_t(M_t - b_t) \leq x) \xrightarrow{t \rightarrow +\infty} \exp(-e^{-x}).$$

Comments :

★ If (G5) is replaced by  $\sqrt{\log \circ S(t)} m(t) \rightarrow \tau \in \mathbb{R}$  we can prove that the preceding probability converges to  $\exp(-e^{\tau-x})$ .

★ Note that the Condition (G6) is not needed. This comes from a stochastic argument of comparison as detailed at the end of the proof.

### 4.2.1 Example of application

An example of process satisfying the assumptions of Corollary 4.2.1 is given by the process that describe the behavior of the likelihood test statistic for the simple Gaussian mixture. We consider the following testing situation :

$$\begin{cases} (H_0) : N(0, 1) \\ (H_1) : (1 - p_0) N(0, 1) + p_0 N(t_0, 1) \end{cases}$$

where  $p_0 \in [0, 1]$  and  $t_0 \in \Gamma = [-T, T], t_0 \neq 0$ . It is proved in Dacunha-Castelle and Gassiat (1997, 1999) that under a local alternative assumption ( $t_0$  fixed and  $p_0 = \frac{\delta}{\sqrt{n}}$  for instance) the asymptotic distribution (as the number  $n$  of independent observations tends to infinity) of the likelihood test statistic is related to that of the supremum of a Gaussian process with mean and covariance

$$m(t) = \frac{\delta^2(e^{tt_0} - 1)}{\sqrt{e^{t^2} - 1}} \quad \text{and} \quad r(s, t) = \frac{e^{st} - 1}{\sqrt{e^{s^2} - 1}\sqrt{e^{t^2} - 1}}.$$

It is deduced in Azais et al. (2004) that, for such kind of mixture and other general mixtures, the log-likelihood ratio test is worthless when the parameter  $T$  tends to infinity.

### 4.3 Main tools

In this part we express classical results about Poisson convergence and normal comparison. Let  $\mathcal{J}$  denote the family that consists of finite unions of bounded and disjoint intervals of  $\mathbb{R}$ .

**Proposition 4.3.1 (Kallenberg (1973))**

Let  $(N_n)_{n \in \mathbb{N}}$  a sequence of point processes on  $\mathbb{R}$  satisfying for any  $B \in \mathcal{J}$ ,

$$a) \quad \mathbb{E}\left(N_n(B)\right) \xrightarrow[n \rightarrow \infty]{} |B|,$$

$$b) \quad \mathbb{P}\left(N_n(B) = 0\right) \xrightarrow[n \rightarrow \infty]{} \exp(-|B|).$$

Then as  $n \rightarrow +\infty$ ,  $N_n$  converges weakly to a standard Poisson process.

Comment :

It is enough to take  $\mathcal{J}$  as the family consisting in elements of the form of  $B = \bigcup_{i=1}^p (c_i, d_i]$ , called  $\tilde{\mathcal{J}}$ .

A central tool in our study is the discretization of the initial Gaussian process  $X$ . Actually we need to show the independence of the maximum of the discretized process on disjoint and bounded sets of  $(0, +\infty)$  (Lemma 4.4.3 proved in Section 4.4). It will be useful to know how to compare in distribution two Gaussian vectors. This result, based on early works of Plackett (1954), Slepian (1962) and Berman (1964, 1971b), is presented here under its version improved by Li and Shao (2002).

**Proposition 4.3.2 (Normal comparison Lemma)**

Let us consider two Gaussian vectors  $U \sim N_k(0, \Sigma)$  and  $V \sim N_k(0, \Lambda)$ .

Assume that  $\Sigma_{ii} = \Lambda_{ii} = 1$  for all  $i \in \{1, \dots, k\}$ .

Denote by  $\rho_{ij} = \max(|\Sigma_{ij}|, |\Lambda_{ij}|)$ . Then

$$\mathbb{P}(U_i \leq w_i, i = 1, \dots, k) - \mathbb{P}(V_i \leq w_i, i = 1, \dots, k)$$

$$\leq \frac{1}{2\pi} \sum_{1 \leq i < j \leq k} (\arcsin(\Sigma_{ij}) - \arcsin(\Lambda_{ij}))^+ \exp\left(-\frac{w_i^2 + w_j^2}{2(1 + \rho_{ij})}\right).$$

### 4.4 Proofs

Once Theorem 4.2.2 has been proved, the proof of Theorem 4.2.3 is direct using the “unit-speed” transformation. Thus we will only prove Theorem 4.2.2 and work under the set of Hypotheses (U1 – U6). The domain of convergence of  $R_u$  being  $(0, +\infty)$ , it suffices to check the characterization of Kallenberg over  $\tilde{\mathcal{J}}$  which contains the family

of sets  $B = \cup_{i=1}^p (c_i, d_i]$  where  $0 < c_1 < d_1 < c_2 < d_2 < \dots < d_p < +\infty$ .  
Throughout the different proofs we will consider

★ "semi-dilated" intervals

$$I_u = (t_u a, t_u b) \text{ with } 0 < a < b \text{ and } t_u = O(C_u^{-\frac{1}{2}}),$$

★ "dilated" intervals

$$(4.1) \quad D_{u,i} = (C_u^{-1}c_i, C_u^{-1}d_i] \text{ for } i \in \{1, \dots, p\} \text{ and } B = \cup_{i=1}^p (c_i, d_i] \in \tilde{\mathcal{J}}$$

$$(4.2) \quad D_u = (C_u^{-1}c, C_u^{-1}d] \text{ with } 0 < c < d,$$

★  $(I_{u,j})_{j=1, \dots, n}$  a partition of  $D_u$  into  $n = n_u$  "semi-dilated" intervals. For the sequel, it is important to keep in mind that  $n = O(C_u^{-\frac{1}{2}})$ .

We will use discretization arguments in order in order to apply normal comparison Lemma. Consequently, let  $q = q_u$  a discretization step that depends on  $u$  such that (U7)  $qu \rightarrow 0$  as  $u \rightarrow +\infty$ .

Define for any subset  $B$  of  $\mathbb{R}$

$$U_u^q(B) = \#\{l \in \mathbb{Z}, q(l-1) \in B, ql \in B, X(q(l-1)) < u < X(ql)\}.$$

#### 4.4.1 Proof of Theorem 4.2.2

Before turning to the proof of our main result, we express intermediate results.

##### Lemma 4.4.1

Under Assumptions (U2-U5),

$$\mathbb{P}(M_{I_u} > u) = \mathbb{E}(U_u(I_u)) + o(C_u^{\frac{1}{2}}),$$

with  $o(C_u^{\frac{1}{2}})$  uniform for all intervals  $I_u$  such that  $a \geq \gamma > 0$ .

##### Lemma 4.4.2

Under Assumptions (U1, U4-U6, U7),

$$\mathbb{E}(U_u(I_u) - U_u^q(I_u)) = o(C_u^{\frac{1}{2}}),$$

with  $o(C_u^{\frac{1}{2}})$  uniform for all  $I_u$  such that  $0 < \gamma \leq a < b \leq \Gamma$ .

##### Lemma 4.4.3

Under Assumptions (U2-U7),

$$\mathbb{P}(X(ql) \leq u, \forall ql \in \bigcup_{j=1}^n I_{u,j}) = \prod_{j=1}^n \mathbb{P}(X(ql) \leq u, \forall ql \in I_{u,j}) + o(1),$$

$$\mathbb{P}(X(ql) \leq u, \forall ql \in \bigcup_{i=1}^p D_{u,i}) = \prod_{i=1}^p \mathbb{P}(X(ql) \leq u, \forall ql \in D_{u,i}) + o(1),$$

$$\mathbb{P}\left(\bigcap_{i=1}^p \{M_{D_{u,i}} \leq u\}\right) = \prod_{i=1}^p \mathbb{P}(M_{D_{u,i}} \leq u) + o(1).$$

**Lemma 4.4.4**

Under Assumptions (U1 – U6),

$$\mathbb{P}(M_{D_u} \leq u) = \exp(- (d - c)) + o(1).$$

The proofs of these lemmas are postponed up to next sections.

Now let  $B = \bigcup_{i=1}^p (c_i, d_i]$  be an element of the family  $\tilde{\mathcal{J}}$ . Define the function  $\mu(t) = \sup_{s>t} |m(s)|$ . Assumption (U5) implies that for any positive  $c$

$$\lim_{u \rightarrow +\infty} u \mu(C_u^{-1}c) = 0.$$

We start by checking Condition a) of Proposition 4.3.1.

Using the fact that the variance is constant, it may be seen that for any positive  $t$  the variables  $X_t$  and  $X'_t$  are independent. Consequently by Rice formula the first moment of  $R_u(B)$  is

$$\mathbb{E}(R_u(B)) = \mathbb{E}\left(U_u(C_u^{-1}B)\right) = \int_{C_u^{-1}B} \mathbb{E}(X'_t^+) \varphi(u - m(t)) dt.$$

Since  $\mathbb{E}(X'_t^+) = \mathbb{E}((G - m'_t)^+)$  where  $G$  is a standard Gaussian variable and

$$\sup_{t \in C_u^{-1}B} |m'_t| \rightarrow 0,$$

it is obvious that  $\mathbb{E}(X'_t^+) = \frac{1}{\sqrt{2\pi}}(1 + o(1))$  with  $o(1)$  uniform with respect to  $t \in C_u^{-1}B$ . After a change of variables we get

$$\mathbb{E}(R_u(B)) = \int_B \exp\left(um(C_u^{-1}v) - \frac{1}{2}m^2(C_u^{-1}v)\right) dv (1 + o(1)).$$

★ Uniformly with respect to  $v \in B$  and according to (U5)

$$\exp\left(um(C_u^{-1}v) - \frac{1}{2}m^2(C_u^{-1}v)\right) \leq \exp\left(u\mu(C_u^{-1}c_1)\right) \leq (Const).$$

★ For every  $v \in B$ ,  $\lim_{u \rightarrow \infty} \exp\left(um(C_u^{-1}v) - \frac{1}{2}m^2(C_u^{-1}v)\right) = 1$ .

Thus by application of Lebesgue Theorem, one obtains

$$(4.3) \quad \mathbb{E}(R_u(B)) = |B| (1 + o(1)).$$

Now we check Assumption b) of Proposition 4.3.1.

$$\mathbb{P}(R_u(B) = 0) = \mathbb{P}\left(\bigcap_{i=1}^p \{R_u((c_i, d_i]) = 0\}\right) = \mathbb{P}\left(\bigcap_{i=1}^p \{U_u(D_{u,i}) = 0\}\right).$$

This last probability is related to the supremum of the process since

$$0 \leq \mathbb{P}\left(\bigcap_{i=1}^p \{U_u(D_{u,i}) = 0\}\right) - \mathbb{P}\left(\bigcap_{i=1}^p \{M_{D_{u,i}} \leq u\}\right) \leq p \bar{\Phi}(u - m(C_u^{-1}c_1)) = o(1).$$

Thus we deduce from Lemma 4.4.3 that

$$\mathbb{P}(R_u(B) = 0) = \mathbb{P}\left(\bigcap_{i=1}^p \{M_{D_{u,i}} \leq u\}\right) + o(1) = \prod_{i=1}^p \mathbb{P}(M_{D_{u,i}} \leq u) + o(1).$$

The desired result is obtained since Lemma 4.4.4 leads to

$$\mathbb{P}(R_u(B) = 0) = \prod_{i=1}^p \exp(- (d_i - c_i)) + o(1) = \exp(-|B|) + o(1). \quad \square$$

#### 4.4.2 Proof of Lemma 4.4.1

The upper bound comes from the equality

$$(4.4) \quad \{M_{I_u} > u\} = \{X(t_u a) \geq u\} \cup \{X(t_u a) < u, U_u(I_u) \geq 1\} \quad a.s.$$

which leads, by Markov inequality, to

$$\mathbb{P}(M_{I_u} > u) \leq \bar{\Phi}(u - m(t_u a)) + \mathbb{E}(U_u(I_u)).$$

The first stage follows from the equivalence  $\bar{\Phi}(u - m(t_u a)) = o(C_u^{\frac{1}{2}})$ .

Set  $\nu_2 = \mathbb{E}(U_u(I_u)(U_u(I_u) - 1))$ . Using (4.4) we deduce, as it is done in Azaïs et al. (1999), that

$$\begin{aligned} \mathbb{P}(M_{I_u} > u) &= \mathbb{P}(X(t_u a) \geq u) + \mathbb{P}(U_u(I_u) \geq 1) - \mathbb{P}(X(t_u a) \geq u, U_u(I_u) \geq 1) \\ &\geq \mathbb{P}(X(t_u a) \geq u) + \mathbb{E}(U_u(I_u)) - \frac{\nu_2}{2} - \mathbb{P}(X(t_u a) > u, U_u(I_u) \geq 1) \\ &\geq \mathbb{E}(U_u(I_u)) - \frac{\nu_2}{2} + o(C_u^{\frac{1}{2}}). \end{aligned}$$

It is then sufficient to analyze the factorial moment of order two of the random variables  $U_u(I_u)$ ,

$$\nu_2 = \iint_{I_u^2} \mathbb{E}(X_s'^+ X_t'^+ / X_s = X_t = u) p_{X_s, X_t}(u, u) ds dt.$$

Set  $A_u(s, t) = \mathbb{E}(X_s'^+ X_t'^+ / X_s = X_t = u) p_{X_s, X_t}(u, u)$ . Observe that by (U2), there exists  $T_0 > 0$  such that

$$R_{T_0} = \sup_{|s-t| \geq T_0} |r(s, t)| < \frac{1}{3}.$$

For a given small enough  $\delta > 0$ , we partition  $I_u^2$  into the three following subsets

$$\begin{aligned} E_u &= \{(s, t) \in I_u^2, |s - t| < \delta\} \\ F_u &= \{(s, t) \in I_u^2, \delta \leq |s - t| < T_0\} \\ G_u &= \{(s, t) \in I_u^2, |s - t| \geq T_0\}. \end{aligned}$$

★ We begin by studying the integral on  $F_u$ . We know by (U4) that  $r_{01}$  is uniformly bounded on  $F_u$  and by (U3) that  $\sup_{(s,t) \in F_u} |r(s,t)| < 1$ . We find that

$$\begin{aligned} \mathbb{E}(X_s' + X_t' / X_s = X_t = u) &\leq \frac{1}{4} \mathbb{E}((X_s' + X_t')^2 / X_s = X_t = u) \\ &\leq \frac{1}{4} \left\{ \mathbb{E}(X_s' + X_t' / X_s = X_t = u)^2 + \text{Var}(X_s' + X_t' / X_s = X_t = u) \right\} \\ &\leq \frac{1}{4} \left\{ \left( \frac{r_{01}(s,t) + r_{10}(s,t)}{1 + r(s,t)} \right)^2 u^2 + \text{Var}(X_s' + X_t') \right\} \\ &\leq \left( \frac{r_{01}(s,t) + r_{10}(s,t)}{1 + r(s,t)} \right)^2 \frac{u^2}{4} + 1 \\ &\leq (\text{Const}) u^2 \end{aligned}$$

where the constant depends only on  $\delta$ . It allows us to upper bound the integral by

$$\iint_{F_u} A_u(s,t) ds dt \leq \iint_{F_u} (\text{Const}) u^2 p_{X_s, X_t}(u, u) ds dt.$$

Since  $R_\delta = \sup_{|t-s| \geq \delta} |r(s,t)| < 1$  because of (U3), we get

$$p_{X_s, X_t}(u, u) \leq (\text{Const}) \exp\left(-\frac{u^2}{1 + R_\delta}\right),$$

and we can conclude that

$$\iint_{F_u} A_u(s,t) ds dt \leq (\text{Const}) u^2 \exp\left(-\frac{u^2}{1 + R_\delta} + \frac{u^2}{4}\right) = o(C_u^{\frac{1}{2}}).$$

★ With the same arguments as for  $F_u$ , the conditional expectation on  $G_u$  is upper bounded by

$$\mathbb{E}(X_s' + X_t' / X_s = X_t = u) \leq (\text{Const}) u^2$$

where now the constant depends only on the process  $X$ . Hence

$$\iint_{G_u} A_u(s,t) ds dt \leq (\text{Const}) u^2 \exp\left(-\frac{u^2}{1 + R_{T_0}} + \frac{u^2}{2}\right) = o(C_u^{\frac{1}{2}}).$$

★ On  $E_u$  we set

$$\mu_s = \mathbb{E}(X_s' / X_s = X_t = u) = \frac{r_{10}(s,t)}{1 + r(s,t)} u$$

$$\mu_t = \mathbb{E}(X_t' / X_s = X_t = u) = \frac{r_{01}(s,t)}{1 + r(s,t)} u$$

$$\sigma_s^2 = \text{Var}(X_s' / X_s = X_t = u) = 1 - \frac{r_{10}^2(s,t)}{1 - r^2(s,t)}$$

$$\sigma_t^2 = \text{Var}(X_t' / X_s = X_t = u) = 1 - \frac{r_{01}^2(s,t)}{1 - r^2(s,t)}.$$

$A_u(s, t)$  being symmetric in  $(s, t)$  we may assume that  $s < t$ . For some  $\xi_1$  and  $\xi_2 \in (s, t)$ , Taylor expansions of  $r$  and  $r_{01}$  about the point  $(s, s)$  give

$$r(s, t) = 1 - \frac{(t-s)^2}{2} + \frac{(t-s)^4}{4!} r_{04}(s, \xi_1)$$

and

$$r_{01}(s, t) = -(t-s) + \frac{(t-s)^3}{3!} r_{04}(s, \xi_2).$$

Using (U4),  $\mu_t = u \times \left( -\frac{t-s}{2} + O((t-s)^3) \right)$  with a uniform remainder. Thus for  $\delta$  sufficiently small we have

$$\mu_t \leq -\frac{t-s}{4}u \quad \text{and} \quad \mu_s \geq \frac{t-s}{4}u$$

for example. It follows that  $\mu_t \leq 0$  and  $\mu_s \geq 0$ . After some calculations we obtain from (U4)

$$\sigma_t^2 = (t-s)^2 \left\{ \frac{r_{04}(s, \gamma)}{3} + \frac{1}{4} + \frac{r_{04}(s, \xi)}{4!} \right\} + O((t-s)^4)$$

with also a uniform remainder. Thus we can find a constant  $C_1$  depending only on  $\delta$  such that :

$$\forall (s, t) \in E_u, \quad \sigma_t^2, \sigma_s^2 \leq C_1 (t-s)^2.$$

Furthermore *under the conditional distribution* we have from Cauchy-Schwarz inequality :

$$\mathbb{E}(X_s'^+ X_t'^+) \leq \sqrt{\mathbb{E}(X_s'^{+2})} \sqrt{\mathbb{E}(X_t'^{+2})}.$$

In addition, if  $Z \sim N(\mu, \sigma^2)$  then

$$\begin{aligned} \mathbb{E}(Z^{+2}) &\leq \mathbb{E}(Z^2) \leq \mu^2 + \sigma^2 \quad \text{for } \mu > 0 \\ \mathbb{E}(Z^{+2}) &\leq \frac{\sigma^2}{2} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \quad \text{for } \mu < 0. \end{aligned}$$

Combining the latter results, we get in our case

$$\begin{aligned} \mathbb{E}(X_s'^+ X_t'^+ / X_s = X_t = u) &\leq (\mu_s^2 + \sigma_s^2)^{\frac{1}{2}} \frac{\sigma_t}{\sqrt{2}} \exp\left(-\frac{\mu_t^2}{4\sigma_t^2}\right) \\ &\leq (Const) (t-s)^2 u \exp\left(-C_2 u^2\right). \end{aligned}$$

where  $C_2$  is a positive constant. Moreover, for  $\delta$  sufficiently small

$$\frac{1}{1-r^2(s, t)} \leq \frac{(Const)}{(t-s)^2},$$

whence

$$p_{X_s, X_t}(u, u) \leq \frac{(Const)}{t-s} \exp\left(-\frac{u^2}{1+r(s, t)}\right),$$



$$A_u(s, t) \leq (\text{Const}) (t - s) u \exp \left( -u^2 \left( \frac{1}{2} + (\text{Const}) \right) \right)$$

and

$$\iint_{E_u} A_u(s, t) ds dt \leq (\text{Const}) u \exp \left( -u^2 \left( \frac{1}{4} + (\text{Const}) \right) \right) = o(C_u^{\frac{1}{2}}).$$

The proof is now complete by fixing  $\delta$  in order to have all relations concerning the upper bound on  $E_u$ , and then applying the bounds for the integrals over the two other domains. Note that our result is uniform for all intervals  $I_u$  such that  $a \geq \gamma > 0$ .  $\square$

### 4.4.3 Proof of Lemma 4.4.2

The proof is similar to that of Leadbetter et al. (1983, Lemmas 7.3.1 and 8.2.2). With the same arguments as those used in (4.3), we can prove that

$$(4.5) \quad \mathbb{E}(U_u(I_u)) = |I_u| C_u (1 + o(1))$$

with  $o(1)$  uniform for all intervals  $I_u$  such that  $0 < \gamma \leq a < b \leq \Gamma$ . It is thus enough to study  $\mathbb{E}(U_u^q(I_u))$ . Now it is sufficient to prove that for all family of intervals  $J_u$  with end point  $a_u, b_u$  being a multiple of  $q(u)$  and such that  $a_u \sim C_u^{-\frac{1}{2}} a$  and  $b_u \sim C_u^{-\frac{1}{2}} b$  we have

$$\mathbb{E}(U_u^q(J_u)) = |J_u| C_u + o(C_u^{\frac{1}{2}}).$$

This is a direct consequence of the fact that

$$P_{k,q}(u) = \mathbb{P} \left( X(kq(u)) < u < X((k+1)q(u)) \right) = q C_u (1 + o(1))$$

uniformly for  $k$  such that  $C_u^{-\frac{1}{2}} \gamma < kq(u) < C_u^{-\frac{1}{2}} \Gamma$ . This is proved in the next paragraph.

Set  $d_k = kq(u)$ , the event  $\left\{ X(kq(u)) < u < X((k+1)q(u)) \right\}$  can be written as

$$\left\{ X(d_k) < u < X(d_{k+1}) \right\} = \left\{ |Y_{1,k} - u| < \frac{q}{2} Y_{2,k} \right\}$$

where  $Y_{1,k} = \frac{X(d_k) + X(d_{k+1})}{2}$  (resp.  $Y_{2,k} = \frac{X(d_{k+1}) - X(d_k)}{q}$ ) is a Gaussian variable with mean  $m_{1,k} = \frac{m(d_k) + m(d_{k+1})}{2}$  and variance  $\sigma_{1,k}^2 = \frac{1+r(d_k, d_{k+1})}{2}$  (resp. with mean  $m_{2,k} = \frac{m(d_{k+1}) - m(d_k)}{q}$  and variance  $\sigma_{2,k}^2 = \frac{2(1-r(d_k, d_{k+1}))}{q^2}$ ).

$$\begin{aligned} q^{-1} C_u^{-1} P_{q,k}(u) &= C_u^{-1} \frac{\mathbb{P} \left( |Y_{1,k} - u| < \frac{q}{2} Y_{2,k} \right)}{q} \\ &= \psi(u, k) \int_0^\infty dy \int_{-1}^1 dx y \exp(f(y, x, u, k)) \end{aligned}$$

with

$$f(y, x, u, k) = - \left( y - \frac{m_{2,k}}{\sqrt{2}\sigma_{2,k}} \right)^2 - \frac{q^2 x^2 y^2 \sigma_{2,k}^2}{4\sigma_{1,k}^2} - \frac{(u - m_{1,k}) \sigma_{2,k} q x y}{\sqrt{2}\sigma_{1,k}^2}$$

$$\text{and} \quad \psi(u, k) = \frac{\sigma_{2,k}}{\sigma_{1,k}} \exp \left( - \frac{u^2}{2\sigma_{1,k}^2} (1 - \sigma_{1,k}^2) - \frac{m_{1,k}^2}{2\sigma_{1,k}^2} + \frac{u m_{1,k}}{\sigma_{1,k}^2} \right).$$

★  $u m_{1,k} \leq u \mu(C_u^{-\frac{1}{2}}\gamma)$ . Consequently  $u m_{1,k}$  and  $m_{1,k}$  converge to 0.

★  $|m_{2,k}| \leq \sup_{x \in (C_u^{-1/2}\gamma, C_u^{-1/2}\Gamma)} |m'(x)|$ . Hence  $m_{2,k}$  also converges to 0.

★  $\sigma_{1,k}$  and  $\sigma_{2,k}$  converge to 1. Indeed

$$\begin{aligned} |\sigma_{1,k}^2 - 1| &= \left| \frac{r(d_k, d_{k+1}) - 1}{2} \right| = \left| \frac{r(d_k, d_{k+1}) - r(d_k, d_k)}{2} \right| \\ &\leq q \sup_{x \in (d_k, d_{k+1})} |r_{01}(d_k, x)|. \end{aligned}$$

Since  $r_{01}$  is bounded on  $(C_u^{-1/2}\gamma, C_u^{-1/2}\Gamma)$ , we can conclude. Now, let us look at what occurs for  $\sigma_{2,k}^2$ . The Taylor expansion of order four of  $r$  is

$$r(d_k, d_{k+1}) = 1 - \frac{q^2}{2} + \frac{r_{04}(d_k, \tau) q^4}{4}$$

where  $\tau \in (d_k, d_{k+1})$ . Assumption (U4) leads to

$$|\sigma_{2,k}^2 - 1| \leq \frac{r_{04}(d_k, \tau) q^2}{4} \leq (Const) q^2 \xrightarrow{u \rightarrow +\infty} 0.$$

First of all, one can deduce from preceding convergences that  $\psi(u, k)$  converges to 1 uniformly in  $k$ . Moreover, we have the upper bound

$$\begin{aligned} f(y, x, u, k) &\leq -\left(y - \frac{m_{2,k}}{\sqrt{2}\sigma_{2,k}}\right)^2 - \frac{(u - m_{1,k})\sigma_{2,k}qxy}{\sqrt{2}\sigma_{1,k}^2} \\ &\leq -\left(y - \frac{m_{2,k}}{\sqrt{2}\sigma_{2,k}}\right)^2 + \frac{(u - m_{1,k})\sigma_{2,k}qy}{\sqrt{2}\sigma_{1,k}^2} \quad \text{obtained for } x = -1 \\ &\leq -y^2 + \left\{ \frac{\sqrt{2}m_{2,k}}{\sigma_{2,k}} + \frac{(u - m_{1,k})\sigma_{2,k}q}{\sqrt{2}\sigma_{1,k}^2} \right\} y. \end{aligned}$$

Consequently, when  $y > \frac{2\sqrt{2}m_{2,k}}{\sigma_{2,k}} + \frac{\sqrt{2}(u - m_{1,k})\sigma_{2,k}q}{\sigma_{1,k}^2}$ , one has

$$f(y, x, u, k) \leq -\frac{y^2}{2}.$$

Thus the integral for  $(y, x)$  on

$$\left[ \frac{2\sqrt{2}m_{2,k}}{\sigma_{2,k}} + \frac{\sqrt{2}(u - m_{1,k})\sigma_{2,k}q}{\sigma_{1,k}^2}, \infty \right) \times [-1, 1]$$

tends by dominated convergence to

$$2 \int_0^\infty y \exp(-y^2) dy = 1.$$

The remaining integral is upper bounded by

$$\left( \frac{2\sqrt{2}m_{2,k}}{\sigma_{2,k}} + \frac{\sqrt{2}(u - m_{1,k})\sigma_{2,k}q}{\sigma_{1,k}^2} \right)^2 \exp\left( 2 \left\{ \frac{\sqrt{2}m_{2,k}}{\sigma_{2,k}} + \frac{(u - m_{1,k})\sigma_{2,k}q}{\sqrt{2}\sigma_{1,k}^2} \right\}^2 \right)$$

which converges to 0. In short since  $q^{-1} C_u^{-1} P_{q,k}(u) \rightarrow 1$  the proof is complete.  $\square$

#### 4.4.4 Proof of Lemma 4.4.3

The first two statements come from same arguments. To prove the first part we denote by  $\Lambda$  the variance matrix of the vector  $(X(q_l), \forall q_l \in \bigcup_{j=1}^n I_{u,j})$ . It can be written using  $\Lambda_{j_1 j_2}$ , the variance matrix of  $X(q_l)$  for all  $q_l \in I_{u,j_1}$  and  $X(q_l)$  for all  $q_l \in I_{u,j_2}$ . Let  $\Sigma$  be the diagonal matrix exit of  $\Lambda$ .

$$\Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & \dots & \Lambda_{1n} \\ \Lambda_{21} & \Lambda_{22} & \dots & \Lambda_{2n} \\ \vdots & \vdots & \ddots & \\ \Lambda_{n1} & \Lambda_{n2} & & \Lambda_{nn} \end{bmatrix} \quad \Sigma = \begin{bmatrix} \Lambda_{11} & 0 & \dots & 0 \\ 0 & \Lambda_{22} & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & \Lambda_{nn} \end{bmatrix}$$

We write  $\{d_1, \dots, d_N\}$  the set of discretized points present in  $D_u = \bigcup_{j=1}^n I_{u,j}$ . One can describe these two matrices in the following way :

- elements of  $\Lambda$  are the  $\lambda_{l,k} = r(d_l, d_k)$  for  $(l, k) \in \{1, \dots, N\}^2$
- those of  $\Sigma$  are denoted by  $\sigma_{l,k}$  for  $(l, k) \in \{1, \dots, N\}^2$ .

By applying twice Normal comparison Lemma, we have :

$$\begin{aligned} & \left| \mathbb{P}(X(q_l) \leq u, \forall q_l \in \bigcup_{j=1}^n I_{u,j}) - \prod_{j=1}^n \mathbb{P}(X(q_l) \leq u, \forall q_l \in I_{u,j}) \right| \\ & \leq \frac{1}{2\pi} \sum_{1 \leq l < k \leq N} \left| \arcsin(\lambda_{l,k}) - \arcsin(\sigma_{l,k}) \right| \exp \left( - \frac{(u - m(d_l))^2 + (u - m(d_k))^2}{2(1 + |\lambda_{l,k}|)} \right). \end{aligned}$$

On the diagonal of  $\Lambda$ , in other words when  $l$  and  $k$  satisfy  $d_l, d_k \in I_{u,j}$ , we know that  $\lambda_{l,k} = \sigma_{l,k}$ . Thus, it is sufficient to carry out this sum on the couples

$$K = \{(l, k), 1 \leq l < k \leq N, \exists j_1 \neq j_2, d_l \in I_{u,j_1}, d_k \in I_{u,j_2}\}.$$

The preceding inequality becomes :

$$\begin{aligned} (4.6) \quad & \left| \mathbb{P}(X(q_l) \leq u, \forall q_l \in \bigcup_{j=1}^n I_{u,j}) - \prod_{j=1}^n \mathbb{P}(X(q_l) \leq u, \forall q_l \in I_{u,j}) \right| \\ & \leq (Const) \sum_{(l,k) \in K} |r(d_l, d_k)| \exp \left( - \frac{(u - m(d_l))^2 + (u - m(d_k))^2}{2(1 + |r(d_l, d_k)|)} \right) \\ & \leq (Const) \sum_{(l,k) \in K} |r(d_l, d_k)| \exp \left( - \frac{u^2}{1 + |r(d_l, d_k)|} + \frac{u(m(d_l) + m(d_k))}{1 + |r(d_l, d_k)|} \right). \end{aligned}$$

Using (U2), we can find  $T_0$  such that  $R_{T_0} = \sup_{|s-t| > T_0} |r(s, t)| < \frac{1}{3}$ . Let us introduce the following sets :

$$\begin{aligned} K_1 &= \{(l, k) \in K, |d_l - d_k| < T_0\} \\ K_2 &= \{(l, k) \in K, T_0 < |d_l - d_k| < \frac{C_u^{-1}}{n}(d - c)\} \\ K_3 &= \{(l, k) \in K, \frac{C_u^{-1}}{n}(d - c) < |d_l - d_k| < C_u^{-1}(d - c)\}. \end{aligned}$$

Set  $Z(a)$  the number of occurrences of  $(l, k) \in K$  such that  $|d_l - d_k| = qa$ . It is easy to see that :

$$\begin{aligned} Z(a) &= a(n-1) && \text{for } qa < \frac{C_u^{-1}}{n}(d-c) \\ Z(a) &\leq \frac{C_u^{-1}}{q}(d-c) && \text{for every } a. \end{aligned}$$

★ Using  $\#K_1 \leq (n-1)(T_0 + T_0 - 1 + \dots + 2 + 1)$  and  $0 \leq |r(d_l, d_k)| \leq 1$ , we have :

$$\begin{aligned} \sum_{(l,k) \in K_1} |r(d_l, d_k)| \exp\left(-\frac{u^2}{1+|r(d_l, d_k)|} + \frac{u(m(d_l)+m(d_k))}{1+|r(d_l, d_k)|}\right) \\ \leq (Const) C_u^{-\frac{1}{2}} \exp\left(-\frac{u^2}{2} + (Const) u\right) \\ \leq (Const) \exp\left(-\frac{u^2}{4} + (Const) u\right) \xrightarrow{u \rightarrow +\infty} 0. \end{aligned}$$

★ The condition on  $R_{T_0}$  ensures that :

$$\begin{aligned} \sum_{(l,k) \in K_2} |r(d_l, d_k)| \exp\left(-\frac{u^2}{1+|r(d_l, d_k)|} + \frac{u(m(d_l)+m(d_k))}{1+|r(d_l, d_k)|}\right) \\ = \sum_{qa=[T_0]+1}^{\frac{C_u^{-1}}{n}(d-c)} Z(a) \exp\left(-\frac{u^2}{1+R_{T_0}} + (Const) u\right) \\ \leq \frac{C_u^{-1}}{n} \frac{C_u^{-1}}{n} (n-1) \exp\left(-\frac{u^2}{1+R_{T_0}} + (Const) u\right) \\ \leq (Const) \frac{C_u^{-2}}{n} \exp\left(-\frac{u^2}{1+R_{T_0}} + (Const) u\right) \xrightarrow{u \rightarrow +\infty} 0. \end{aligned}$$

★ We write  $\rho(\tau) = \sup_{|s-t|>\tau} |r(s, t)|$ , using changes of variables we get successively :

$$\begin{aligned} \sum_{(l,k) \in K_3} |r(d_l, d_k)| \exp\left(-\frac{u^2}{1+|r(d_l, d_k)|} + \frac{u(m(d_l)+m(d_k))}{1+|r(d_l, d_k)|}\right) \\ \leq \frac{C_u^{-1}}{q^2} \int_{\frac{C_u^{-1}(d-c)}{n}}^{C_u^{-1}(d-c)} \rho(t) \exp\left(-\frac{u^2}{1+\rho(t)} + 2u\mu(C_u^{-1}c)\right) dt \\ \leq \frac{C_u^{-2}}{q^2 u^2} \int_{\frac{(d-c)}{n}}^{d-c} u^2 \rho(t C_u^{-1}) \exp\left(-\frac{u^2}{1+\rho(t C_u^{-1})} + 2u\mu(C_u^{-1}c)\right) dt \\ \leq \frac{1}{q^2 u^2} \int_{\frac{(d-c)}{n}}^{d-c} u^2 \rho(t C_u^{-1}) \exp\left(u^2 \rho(t C_u^{-1}) + 2u\mu(C_u^{-1}c)\right) dt. \end{aligned}$$

Since  $u^2 \rho(t C_u^{-1}) \rightarrow 0$  by (U2) and using (U5), Lebesgue Theorem implies that

$$\int_{\frac{(d-c)}{n}}^{d-c} u^2 \rho(t C_u^{-1}) \exp\left(u^2 \rho(t C_u^{-1}) + 2u\mu(C_u^{-1}c)\right) dt \xrightarrow{u \rightarrow +\infty} 0.$$

Thus we get the convergence to 0 of the sum on  $K_3$  for a convenient choice of  $q$  and finally the convergence to 0 of (4.6).

Now we prove the last statement of Lemma 4.4.3. As in the relation (4.9) we can show that

$$(4.7) \quad 0 \leq \mathbb{P}\left(\bigcap_{i=1}^p \{X(ql) \leq u, \forall ql \in D_{u,i}\}\right) - \mathbb{P}\left(\bigcap_{i=1}^p \{M_{D_{u,i}} \leq u\}\right) \leq o(C_u^{\frac{1}{2}}).$$

Applying what precedes, we have

$$(4.8) \quad \mathbb{P}\left(\bigcap_{i=1}^p \{X(ql) \leq u, \forall ql \in D_{u,i}\}\right) = \prod_{i=1}^p \mathbb{P}(X(ql) \leq u, \forall ql \in D_{u,i}) + o(1)$$

and then using (4.7) applied to  $D_{u,1}, \dots, D_{u,p}$  and to each  $D_{u,i}$  separately we get the result.  $\square$

#### 4.4.5 Proof of Lemma 4.4.4

The demonstration is derived from a partition of  $D_u$  into "semi-dilated" intervals. According to Lemma 4.4.2, we know that  $\mathbb{E}(U_u(I_{u,j}) - U_u^q(I_{u,j}))$  is an  $o(C_u^{\frac{1}{2}})$  uniform for  $(I_{u,j})_{j=1, \dots, n}$ , which allows us to write

$$\begin{aligned} 0 &\leq \mathbb{P}(X(ql) \leq u, \forall ql \in \bigcup_{j=1}^n I_{u,j}) - \mathbb{P}(M_{D_u} \leq u) \\ &\leq \sum_{j=1}^n \left\{ \mathbb{P}(X(\alpha_j) > u) + \mathbb{P}(X(\alpha_j) \leq u, U_u(I_{u,j}) \geq 1, U_u^q(I_{u,j}) = 0) \right\} \\ &\leq \sum_{j=1}^n \left\{ \bar{\Phi}(u - m(\alpha_j)) + \mathbb{E}(U_u(I_{u,j}) - U_u^q(I_{u,j})) \right\} \end{aligned}$$

and to obtain

$$(4.9) \quad 0 \leq \mathbb{P}(X(ql) \leq u, \forall ql \in \bigcup_{j=1}^n I_{u,j}) - \mathbb{P}(M_{D_u} \leq u) \leq n \times o(C_u^{\frac{1}{2}}) = o(1).$$

By using what precedes and the first part of Lemma 4.4.3, we can write

$$\mathbb{P}(M_{D_u} \leq u) = \prod_{j=1}^n \mathbb{P}(X(ql) \leq u, \forall ql \in I_{u,j}) + o(1).$$

As in the proof of the relation (4.9), we have

$$(4.10) \quad \mathbb{P}(X(ql) \leq u, \forall ql \in I_{u,j}) = \mathbb{P}(M_{I_{u,j}} \leq u) + o(C_u^{\frac{1}{2}}).$$

Furthermore, according to Lemma 4.4.1

$$\mathbb{P}(M_{I_{u,j}} \leq u) = 1 - |I_{u,j}| C_u + o(C_u^{\frac{1}{2}}) = 1 - \frac{d-c}{n} + \varepsilon_n$$

with  $(\varepsilon_n)_n$  satisfying  $n \sup_{n \in \mathbb{N}} \varepsilon_n \rightarrow 0$  as  $u \rightarrow +\infty$ . Using (4.10), the same kind of relation is true after discretization

$$\mathbb{P}(X(ql) \leq u, \forall ql \in I_{u,j}) = 1 - \frac{d-c}{n} + \tilde{\varepsilon}_n$$

with again  $n \sup_{n \in \mathbb{N}} \tilde{\varepsilon}_n \rightarrow 0$  as  $u \rightarrow +\infty$ . We obtain

$$\prod_{j=1}^n \mathbb{P}(X(ql) \leq u, \forall ql \in I_{u,j}) \xrightarrow{u \rightarrow +\infty} \exp(-(d-c))$$

by upper and lower bounds. □

#### 4.4.6 Proof of Corollary 4.2.1

##### Step 1

Let  $X$  satisfying assumptions of Corollary 4.2.1 and also (G6), in other words

$$\frac{m'_X(t)}{\sqrt{r_{11}(t,t)}} \xrightarrow{t \rightarrow +\infty} 0.$$

Denote  $Y = \{Y_t = X_{S^{-1}(t)}, t \in \mathbb{R}\}$  the process constructed by “unit-speed” transformation. It satisfies assumptions of Theorem 4.2.2. Set  $M^Y$ ,  $R_u^Y$ ,  $R_u^Y$  and  $m_Y$  the obvious notations related to the process  $Y$ .

Let  $x \in \mathbb{R}$ ,  $\tau = \exp(-x)$  and  $u^2 = 2(\log \circ S(t) + x + \log(\sqrt{2\pi}))$ .

$$\mathbb{P}(M_t \leq u) = \mathbb{P}(M_{S(t)}^Y \leq u) \sim \mathbb{P}(U_u^Y(S(t)) = 0) = \mathbb{P}(R_u^Y(C_u S(t)) = 0)$$

A simple calculation shows that  $C_u S(t) = \tau$ . Moreover, from Theorem 4.2.2, we have the convergence  $\mathbb{P}(R_u^Y(\tau) = 0) \xrightarrow{t \rightarrow \infty} \exp(-\tau)$  from which we get

$$\mathbb{P}(M_t \leq u) \xrightarrow{t \rightarrow \infty} \exp(\exp(-x)).$$

The result follows at once since  $u$  is written in the form  $\frac{x}{a_t} - b_t + o(a_t^{-1})$ .

##### Step 2

We now assume that  $X$  only satisfies Assumptions (G1 – G5).

Let us consider  $K$  a  $\mathcal{C}^\infty$  function with support included in  $[0, 1]$  and satisfying

$$\int K(t) dt = 1.$$

We define

$$m_1(t) = \sup_{s>t} m_Y(s), \quad m_2(t) = m_1 * K(t) \quad \text{and} \quad v(t) = m_2(S(t)).$$

One can check the following assertions :

- ★ since  $m_2$  is  $\mathcal{C}^\infty$  it follows that  $v$  is at least  $\mathcal{C}^1$ ,
- ★  $\lim_{t \rightarrow +\infty} m_2'(t) = 0$  which yields  $\frac{v'(t)}{\sqrt{r_{11}(t,t)}} = m_2'(S(t)) \xrightarrow{t \rightarrow +\infty} 0$ ,
- ★ from  $m_2(t) \geq m_1(t) \geq m_Y(t)$  we deduce that  $v \geq m_X$ ,
- ★  $0 \leq m_2(t) \leq m_1(t-1)$  and so  $\lim_{t \rightarrow +\infty} \sqrt{\log \circ S(t)} v(t) = 0$ .

The Gaussian process  $V$  with mean  $v$  and covariance function  $r$  is stochastically superior to  $X$ . Furthermore, it verifies assumptions of the first part of this proof. Thus

$$\limsup_{t \rightarrow +\infty} \mathbb{P}(a_t(M_t^X - b_t) \leq x) \geq \lim_{t \rightarrow +\infty} \mathbb{P}(a_t(M_t^V - b_t) \leq x) = \exp(-e^{-x}).$$

With a similar proof (using  $\inf_{s>t} m_Y(s)$  instead of  $m_1(t)$ ) we can construct a Gaussian process  $W$  stochastically inferior to  $X$  checking again assumptions of Step 1 and get the inverse inequality.  $\square$

## Bibliography of Chapter 4

- Azaïs, J.-M., Cierco-Ayrolles, C., and Croquette, A.** (1999). Bounds and asymptotic expansions for the distribution of the maximum of a smooth stationary Gaussian process. *ESAIM Probab. Statist.*, 3 :107–129.
- Azaïs, J.-M., Gassiat, É., and Mercadier, C.** (2004). Asymptotic distribution and power of the likelihood ratio test for mixtures : bounded and unbounded cases. *Submitted*.
- Azaïs, J.-M. and Mercadier, C.** (2003). Asymptotic poisson character of extremes in non-stationary gaussian models. *Extremes*, 6 :301–318.
- Berman, S. M.** (1964). Limit theorems for the maximum term in stationary sequences. *Ann. Math. Statist.*, 35 :502–516.
- Berman, S. M.** (1971a). Asymptotic independence of the numbers of high and low level crossings of stationary Gaussian processes. *Ann. Math. Statist.*, 42 :927–945.
- Berman, S. M.** (1971b). Excursions above high levels for stationary Gaussian processes. *Pacific J. Math.*, 36 :63–79.
- Cramér, H. and Leadbetter, M. R.** (1967). *Stationary and related stochastic processes. Sample function properties and their applications*. John Wiley & Sons Inc., New York.
- Dacunha-Castelle, D. and Gassiat, É.** (1997). Testing in locally conic models, and application to mixture models. *ESAIM Probab. Statist.*, 1 :285–317.

- Dacunha-Castelle, D. and Gassiat, É.** (1999). Testing the order of a model using locally conic parametrization : population mixtures and stationary ARMA processes. *Ann. Statist.*, 27(4) :1178–1209.
- Kallenberg, O.** (1973). Characterization and convergence of random measures and point processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 27 :9–21.
- Leadbetter, M. R., Lindgren, G., and Rootzén, H.** (1983). *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer-Verlag, New York.
- Li, W. V. and Shao, Q.-M.** (2002). A normal comparison inequality and its applications. *Probab. Theory Related Fields*, 122(4) :494–508.
- Piterbarg, V. I. and Konakov, V. D.** (1995). High level excursions of Gaussian fields and the weakly optimal choice of the smoothing parameter. I. *Math. Methods Statist.*, 4(4) :421–434.
- Plackett, R. L.** (1954). A reduction formula for normal multivariate integrals. *Biometrika*, 41 :351–360.
- Slepian, D.** (1962). The one-sided barrier problem for Gaussian noise. *Bell System Tech. J.*, 41 :463–501.
- Volkonskiĭ, V. A. and Rozanov, J. A.** (1961). Some limit theorems for random functions. II. *Teor. Veroyatnost. i Primenen.*, 6 :202–215.
- Wschebor, M.** (1985). *Surfaces aléatoires*, volume 1147 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin. Mesure géométrique des ensembles de niveau. [The geometric measure of level sets].





# Chapitre 5

---

## BORNES NUMÉRIQUES POUR LA LOI DU MAXIMUM D'UN PROCESSUS À UN OU DEUX PARAMÈTRES

Nous présentons dans ce chapitre l'intégralité d'un article soumis à publication "Numerical bounds for the distribution of the maximum of a one- or two-parameter process", Mercadier (2005).

Nous considérons la classe de processus stochastiques à valeurs réelles indexés sur un sous-ensemble compact de  $\mathbb{R}$  ou  $\mathbb{R}^2$  avec des trajectoires presque sûrement absolument continues. Nous obtenons une formule implicite pour la distribution de leur maximum. Le résultat principal est la détermination de bornes numériques très précises même pour des niveaux qui ne sont pas hauts.

Nous présentons la première borne explicite pour la fonction de survie du maximum dans le cadre bi-dimensionnel.

Des comparaisons numériques sont effectuées avec des outils connus comme la borne supérieure de Rice et les équivalents basés sur la caractéristique d'Euler.

Nous traitons numériquement la détermination de l'exposant de persistance.

### Abstract

We consider the class of real valued stochastic processes indexed on a compact subset of  $\mathbb{R}$  or  $\mathbb{R}^2$  with almost surely absolutely continuous sample paths. We obtain an implicit formula for the distribution of their maximum. The main result is the derivation of numerical bounds that turn to be very accurate for levels which are not large.

We also present the first explicit upper bound for the survival function of the maximum in the two-dimensional framework.

Numerical comparisons are performed with known tools as Rice upper bound and expansions based on Euler characteristic.

We deal numerically with the determination of the persistence exponent.

## 5.1 Introduction

Let  $\{X(t), t \in S\}$  be a real valued random process defined on a compact subset  $S$  of  $\mathbb{R}$  or  $\mathbb{R}^2$ . The paper examines the question : what is the distribution of its maximum  $M_S$ ?

In the one-parameter framework we will define  $S$  as the interval  $[0, T]$  where  $T$  is a positive real. The literature gives the exact distribution for some Gaussian processes constructed as functions of the Brownian motion and also for five stationary processes (see Azaïs and Wschebor (2002) for an exhaustive list). Most of these results are inferred from Markovian properties. When the process does not have such characteristics, the studies are usually dealt with the survival function. Indeed the event  $\{M_T \geq u\}$  is the disjoint union of  $\{X(0) \geq u\}$  and  $\mathcal{E}(T, u) = \{X(0) < u, M_T \geq u\}$ . Then all the works consist in including the event  $\mathcal{E}(T, u)$  in a bigger one for which the computation of the probability is simpler and still instructive. The most classical bound includes  $\mathcal{E}(T, u)$  in  $\{U_u([0, T]) \geq 1\}$ , which means the existence of an upcrossing by  $X$  of the level  $u$  in  $[0, T]$ . The associated probability is investigated since the early works of S.O. Rice and M. Kac. In all generality, we know by this method that when the sample paths of  $X$  are almost surely absolutely continuous

$$(5.1) \quad \mathbb{P}(M_T \geq u) \leq \mathbb{P}(X(0) \geq u) + \int_0^T \mathbb{E}(X'(t)^+ / X(t) = u) p_{X(t)}(u) dt$$

where  $p_V(v)$  denotes the probability density function of a random vector  $V$  at point  $v$  throughout the paper. It is well understood that for large level  $u$ , the so-called Rice upper bound (5.1) is an accurate equivalent of  $\mathbb{P}(M_T \geq u)$ .

Azaïs and Wschebor (2002) give a general formula for  $\mathbb{P}(M_T \geq u)$  writing that  $\mathcal{E}(T, u)$  is equal to the event  $\{U_u([0, T]) \mathbb{1}_{\{X(0) < u\}} \geq 1\}$ . The survival function of  $M_T$  is thus expressed by means of an infinite series which terms contain the factorial moments of this slight modification of the number of upcrossings. Unfortunately the associated estimating method is not satisfactory when, for a fixed level  $u$ , the parameter  $T$  becomes large.

In a multidimensional framework, we restrict ourselves to a real valued Gaussian field defined on a compact  $S$  of the plane (also called here two-parameter process). As first observation, there is no an exact formula nor bounds for the distribution of its maximum  $M_S$ . However there exists different ways to approximate it that may be organized along the following methods.

- (A) The generalization by Adler (1981) of the number of upcrossings with the Integral Geometry characteristic which is equal to the Euler characteristic of  $A(u) = \{t \in S, X(t) \geq u\}$  if this set does not encounter the boundary of  $S$ .
- (B) The use of the volume of tubes for particular unit-variance Gaussian fields by Sun (1993) and Takemura and Kuriki (2002).
- (C) The double sum method of Piterbarg (1996) that consists in dividing the parameter space into a finite union of smaller sets.
- (D) The generalization of the number of upcrossings with the number of local maxima by Piterbarg (1996) and Azaïs and Delmas (2002).

The books of Adler (1981) and Adler and Taylor (2005) give complete references for the results linked with these approaches. Each of them lead to an expansion of the survival function of  $M_S$ . For instance, when  $X$  is a centered stationary Gaussian field of unit-variance on  $S = [0, T]^2$ , which derivatives joint distribution are non-degenerate, we know since Adler (1981) that  $\mathbb{P}(M_{[0, T]^2} \geq u) = EQA(T, u) + o(u \exp(-u^2/2))$  where

$$(5.2) \quad EQA(T, u) = \frac{T^2 \det(\Sigma)^{\frac{1}{2}}}{(2\pi)^{\frac{3}{2}}} u e^{-\frac{u^2}{2}}$$

and  $\Sigma = \text{Var}(X'(t))$ .

This has been improved by Piterbarg (1996), Azaïs and Delmas (2002) and finally by Taylor et al. (2004). For all positive  $\delta$ , we have  $\mathbb{P}(M_{[0, T]^2} \geq u) = EQT(T, u) + o(\exp(-u^2(1 + \delta)/2))$  where

$$(5.3) \quad EQT(T, u) = \Psi(u) + \frac{T^2 \det(\Sigma)^{\frac{1}{2}}}{(2\pi)^{\frac{3}{2}}} u e^{-\frac{u^2}{2}} + T \frac{\Sigma_{11} + \Sigma_{22}}{2\pi} e^{-\frac{u^2}{2}}$$

and  $\Psi$  is the tail of a standard Gaussian distribution.

These different studies or improvements have a more general version than the present one and we remark that the model considered by Taylor et al. prevails.

In numerous statistical problems involving random processes and fields, the maximum plays potentially an important role. In non-identifiable mixture models the asymptotic distribution of the log-likelihood ratio test statistics relies on that of the maximum of stochastic processes. Their knowledge enables to construct well-defined tests for which one can evaluate the power; we refer to Davies (1977) for pioneer work, Gassiat (2002) and Azaïs et al. (2004). This theory has found practical applications in many areas including image processing and genetics.

In a two-dimensional definition space, a random field is used for instance to model spatial data as the sea level, or more generally every fluctuating surfaces. Results on Euler characteristic have been used by Worsley (1995, a,b) to obtain informations on images arising in astrophysics and medicine.

In physics we meet a concept to control the stability or the memory over a long time of a fluctuating structure (see for instance Molchan and Khokhlov, 2004, and Ehrhardt et al., 2002). This persistence exponent characterizes the probability that a random process remains below a specific level within a given interval of time, in other words the maximum is bounded above. It is important to know that such works consider a level really close or equal to the mean value of the process and consequently we need to have an expression of the value  $\mathbb{P}(M_S \geq u)$  even for not large level  $u$ .

The goal of the present paper is to improve approximations (5.1), (5.2) and (5.3) of the distribution of the maximum  $M_S$ . In both dimensional contexts, the idea consists in introducing the first passage time or location (in a sense that we have to define) at the level  $u$  by  $X$  and compute the associated probability by application of results of Banach or Federer.

In Section 5.2 we give the formula for the distribution of the maximum of absolutely continuous one-parameter processes (Theorem 5.2.1). We prove the bounded and

continuity of the density function of the first passage time (Proposition 5.2.1). Hence the survival function of the maximum takes the simple form (5.6) in the case of continuously differentiable Gaussian processes. With such regularity, this implicit formula can be also obtained from Rychlik (1987).

As a consequence we deduce explicit bounds for the survival function. So Section 5.3 reports an estimating study which enable us to address some problems from the numerical point of view. To this end, we create a "software" based on the WAFO tool-box (Brodtkorb et al., 2000) that estimates Gaussian integrals. Our approximations are compared with the most classical tool that is Rice upper bound and also with estimations of Azaïs and Wschebor (2002). Section 5.4 provides a numerical study of two non-smooth processes : the Brownian motion and the Ornstein-Uhlenbeck process.

We present in Section 5.5 an implicit formula for the distribution of a real valued Gaussian field defined on a part of the plane (Theorem 5.5.1). In addition, we give a simple and explicit upper bound for the survival function in the "unit-speed" case (Corollary 5.5.1) that seems to be the first published bound for the Gaussian fields. In a particular context, we see that this upper bound has EQT for equivalent (Corollary 5.5.2 and remark that follows). We propose in Section 5.6 to compare our estimations to expansions based on the Integral Geometry characteristic.

This paper contains several bounds summarized in the following table.

Bound Name	Description	In the text	Comment
RICE	Rice upper bound	(5.1)	Explicit
UBAW	Upper bound of Azais and Wschebor	5.3.3	Azais and Wschebor (2002)
LBAW	Lower bound of Azais and Wschebor	5.3.3	Azais and Wschebor (2002)
UBPn	Upper bound for smooth Processes	(5.7)	free value of $n$
UBP	Upper bound for smooth Processes	(5.7)	for $n = 30$
LBPn	Lower bound for smooth Processes	(5.8)	free value of $n$
LBP	Lower bound for smooth Processes	(5.8)	mean on several value of $n$
UBNS	Upper bound for non-smooth Processes	(5.15)	for $n = 100$
LBNS	Lower bound for non-smooth Processes	(5.15)	for $n = 100$
EQA	Equivalent of Adler	(5.2)	Adler (1981)
EQT	Equivalent of Taylor et al.	(5.3)	Taylor et al. (2004)
UBUS	Upper bound for "unit-speed" Fields	(5.20)	Explicit
LBF	Lower bound for smooth Fields	(5.22)	for $n = 10$
UBF	Upper bound for smooth Fields	(5.23)	for $n_p = 100, n_q = 15$ and $n_r = 10$

## 5.2 Smooth stochastic processes

Let  $T$  denotes a positive real and consider  $\{X(t), t \in [0, T]\}$  a real valued stochastic process with sample paths supposed to be almost surely absolutely continuous.

### Theorem 5.2.1

Set  $\{X(t), t \in [0, T]\}$  a process with almost surely absolutely continuous sample paths. We suppose that for almost every  $t$  in  $[0, T]$ ,  $X'(t)$  is integrable and that  $X(t)$  admits a density  $p_{X(t)}$ . Then for every  $u$  in  $\mathbb{R}$ ,

$$(5.4) \quad \mathbb{P}(M_T \geq u) = \mathbb{P}(X(0) \geq u) + \lim_{\delta \rightarrow 0} \frac{1}{\delta} \int_u^{u+\delta} dx \int_0^T dt f(t, x)$$

where  $f$  is defined by  $f(t, x) = \mathbb{E}(X'(t)^+ \mathbb{1}_{\{X(s) < x, \forall s < t\}} / X(t) = x) p_{X(t)}(x)$ .

This formula is not explicit, in particular the supremum of the process appears in the indicator function present in the right-hand member. However it is interesting to obtain an equality in this large framework which does not assume  $X$  being Gaussian.

*Proof of Theorem 5.2.1.* We write  $\mathcal{E}(T, u)$  as  $\{\tau_u \in (0, T)\}$  where  $\tau_u$  is defined, with the convention that  $\inf \emptyset = -\infty$ , by

$$(5.5) \quad \tau_u = \inf\{t \in (0, T), X(t) = u, X(s) < u, \forall s < t\}.$$

Our method is based on the indicator variable

$$I_{T,u} = \mathbb{1}_{\{\tau_u \in (0, T)\}}$$

which allows us to write the survival function as  $\mathbb{P}(M_T \geq u) = \mathbb{P}(X(0) \geq u) + \mathbb{E}(I_{T,u})$ . To compute the first moment of  $I_{T,u}$ , let us consider  $B$  and  $E$  some Borelians of  $\mathbb{R}$ ,  $\mathbf{x} : \mathbb{R} \rightarrow \mathbb{R}$  an absolutely continuous function and set  $N(E, \mathbf{x}, u) = \#\{t \in E, \mathbf{x}(t) = u\}$ . Denote by  $\mathbf{x}'$  the derivative of  $\mathbf{x}$  in the sense of absolute continuity. If  $N(E, \mathbf{x}, u)$  is finite for almost every  $u$  we can prove using a formula of Banach that

$$\int_B N(E, \mathbf{x}, u) du = \int_E |\mathbf{x}'(t)| \mathbb{1}_B(\mathbf{x}(t)) dt.$$

Let us define the record points set of the process

$$E = \{t \in [0, T], X(s) < X(t), \forall s < t\}.$$

We have  $N(E, X, u) = I_{T,u}$  and almost surely

$$\int_B I_{T,u} du = \int_0^T |X'(t)| \mathbb{1}_{\{X(s) < X(t), \forall s < t\}} \mathbb{1}_B(X(t)) dt.$$

Taking the expectation, using Fubini Theorem, integrability of  $X'(t)$  and absolute continuity of  $X(t)$ , we get for every Borelian  $B$  of  $\mathbb{R}$

$$\int_B \mathbb{E}(I_{T,u}) du = \int_B du \int_0^T dt \mathbb{E}(|X'(t)| \mathbb{1}_{\{X(s) < u, \forall s < t\}} / X(t) = u) p_{X(t)}(u).$$

Taking into account the indicator function and the conditioning, we remark that the absolute value in the last expression is a positive part and we deduce that for almost every  $u$  in  $\mathbb{R}$ ,  $\mathbb{E}(I_{T,u}) = \int_0^T f(t, u) dt$ . The right continuity of the function  $\mathbb{P}(M_T \geq u) - \mathbb{P}(X(0) \geq u)$  leads to Formula (5.4).  $\square$

Actually we find again the Rice upper bound for processes satisfying hypotheses of Theorem 5.2.1. Defining the conditional expectation  $\mathbb{E}(X'(t)^+/X(t) = u)$  as the limit of  $\frac{1}{\delta} \int_u^{u+\delta} \mathbb{E}(X'(t)^+/X(t) = x) dx$  as  $\delta$  tends to zero we get

$$\mathbb{P}(M_T \geq u) \leq \mathbb{P}(X(0) \geq u) + \int_0^T \mathbb{E}(X'(t)^+/X(t) = u) p_{X(t)}(u) dt.$$

We turn now to the case of Gaussian processes. The value  $f(t, u)$  is the density function at point  $t$  of the passage time  $\tau_u$  obtained by Rychlik (1987) in the case of continuously differentiable sample paths. Consequently for such processes we can directly write that

$$\begin{aligned} \mathbb{P}(M_T \geq u) &= \mathbb{P}(X(0) \geq u) + \mathbb{P}(\tau_u \in (0, T)) \\ &= \mathbb{P}(X(0) \geq u) + \int_0^T f(t, u) dt. \end{aligned}$$

However we propose an original proof of this formula in this framework, which seems simpler.

**Proposition 5.2.1**

Set  $\{X(t), t \in [0, T]\}$  a Gaussian process with almost surely continuously differentiable sample paths. We suppose that  $\sigma_T^2 = \min_{t \in [0, T]} \text{Var}(X(t)) > 0$ . Then the function  $\int_0^T f(t, \cdot) dt$  defined in Theorem 5.2.1 is continuous and for every  $u$  in  $\mathbb{R}$ ,

$$(5.6) \quad \mathbb{P}(M_T \geq u) = \mathbb{P}(X(0) \geq u) + \int_0^T \mathbb{E}(X'(t)^+ \mathbb{1}_{\{X(s) < u, \forall s < t\}}/X(t) = u) p_{X(t)}(u) dt.$$

Consequently the distribution of the maximum is known from only two terms. At first sight the interest of (5.6) could appear limited but it lies in the upper bound (5.7) that is deduced in Section 5.3.

*Proof of Proposition 5.2.1.* Regularity of the sample paths implies that  $\mathbb{E}(X'(t))$  and  $\text{Var}(X'(t))$  are continuous so that  $e_T = \sup_{t \in [0, T]} \mathbb{E}(|X'(t)|)$  and  $s_T^2 = \sup_{t \in [0, T]} \text{Var}(X'(t))$  are finite. We point out that if  $\mathcal{N}$  is a Gaussian random variable then  $\mathbb{E}(\mathcal{N}^+) \leq |\mathbb{E}(\mathcal{N})| + \sqrt{\text{Var}(\mathcal{N})}$ . Applying to  $\mathcal{N} = X'(t)$  under conditional distribution and using the Gaussian regression formula we obtain after simplifications that  $f(t, u) \leq (e_T + 2s_T)/\sigma_T$ .

Clearly it is now sufficient to prove that  $u \mapsto \mathbb{E}(X'(t)^+ \mathbb{1}_{\{X(s) < u, \forall s < t\}}/X(t) = u)$  is continuous. Let  $r(\cdot, \cdot)$  denote the covariance function of the process. Perform a Gaussian regression of  $\{(X(s), X'(s)), \forall s < t\}$ , with respect to the variable  $X(t)$ , by writing

$$\begin{aligned} X(s) &= a^t(s) X(t) + (t-s) U^t(s) \\ X'(s) &= b^t(s) X(t) - U^t(s) + (t-s) \tilde{U}^t(s) \end{aligned}$$

where for all  $s < t$ ,  $a^t(s) = \frac{r(s, t)}{r(t, t)}$  and  $b^t(s) = \frac{\partial}{\partial s} a^t(s)$ . The processes  $U^t$  and  $\tilde{U}^t$  are Gaussian, non-degenerated, independent to  $X(t)$ .



Let us introduce  $D_{t,u} = b^t(t)u - U^t(t)$  and  $B_{t,u} = \mathbb{1}_{\{\mathcal{M}_{t,u} < 0\}}$  where

$$\mathcal{M}_{t,u} = \sup_{s < t} \left\{ U^t(s) - u \frac{1 - a^t(s)}{t - s} \right\}.$$

We get  $\mathbb{E} \left( X'(t)^+ \mathbb{1}_{\{X(s) < u, \forall s < t\}} / X(t) = u \right) = \mathbb{E} \left( D_{t,u}^+ B_{t,u} \right)$ . For  $u_n$  a sequence that converges to  $u$  as  $n$  tends to infinity, we consider

$$|\mathbb{E} \left( D_{t,u_n}^+ B_{t,u_n} \right) - \mathbb{E} \left( D_{t,u}^+ B_{t,u} \right)| \leq \mathbb{E} \left( |D_{t,u_n}^+ - D_{t,u}^+| \right) + \mathbb{E} \left( D_{t,u}^+ |B_{t,u_n} - B_{t,u}| \right).$$

By a dominated convergence argument, the first term of the right hand member tends to zero. For the second term remark that

$$\mathbb{P} \left( B_{t,u_n} \neq B_{t,u} \right) \leq \mathbb{P} \left( \mathcal{M}_{t,u_n} < 0, \mathcal{M}_{t,u} \geq 0 \right) + \mathbb{P} \left( \mathcal{M}_{t,u_n} \geq 0, \mathcal{M}_{t,u} < 0 \right)$$

tends to zero :  $\mathcal{M}_{t,u_n} \rightarrow \mathcal{M}_{t,u}$  as  $n$  tends to infinity and according to Ylvisaker (1968) we have that almost surely  $\mathcal{M}_{t,u} \neq 0$ . Then the integrability of  $D_{t,u}^+$  leads to the conclusion.

The hypothesis  $\sigma_T^2 > 0$  implies using the theorem of Ylvisaker (1968) that  $\mathbb{P} \left( M_T \geq \cdot \right)$  is continuous. The proof is now obvious combining last assertion, Theorem 5.2.1 and last continuity.  $\square$

### 5.3 Numerical applications for smooth Gaussian processes

Let  $\{X(t), t \in [0, T]\}$  be a Gaussian process satisfying conditions of Proposition 5.2.1 and consequently (5.6). The purpose is to write upper and lower bounds for the survival function of the maximum of  $X$ . Through this paper we use  $\varphi$ ,  $\Phi$  and  $\Psi$  to denote the standard normal probability density function, distribution function and tail respectively.

For every  $n$  in  $\mathbb{N}^*$ , we obtain an upper bound replacing the interval  $[0, t)$  by the subdivision  $\Delta_{t,n} = \{kt/n, k = 0, \dots, n-1\}$

$$(5.7) \quad \mathbb{P} \left( M_T \geq u \right) \leq \text{UBPn} \left( T, u, n \right)$$

$$\text{UBPn} \left( T, u, n \right) = \mathbb{P} \left( X(0) \geq u \right) + \int_0^T \mathbb{E} \left( X'(t)^+ \mathbb{1}_{\{X(s) < u, \forall s \in \Delta_{t,n}\}} / X(t) = u \right) p_{X(t)}(u) dt.$$

It is easy to see that (5.7) is valid under assumptions of Theorem 5.2.1 as soon as the components of this inequality are continuous.

On the other hand the probability  $\mathbb{P} \left( M_T \geq u \right)$  has obvious lower bounds obtained by discretization. For every  $n$  in  $\mathbb{N}^*$ , one has

$$(5.8) \quad \mathbb{P} \left( M_T \geq u \right) \geq \text{LBPn} \left( T, u, n \right)$$

with  $\text{LBPn} \left( T, u, n \right) = 1 - \mathbb{P} \left( X(s) \leq u, \forall s \in \tilde{\Delta}_{T,n-1} \right)$  and  $\tilde{\Delta}_{T,n} = \{kT/n, k = 0, \dots, n\}$ .

After the description of our numerical framework we compare our method to the known tools presented in the introduction.

### 5.3.1 MATLAB Tool-box

For the sake of simplicity we now restrict ourselves to centered processes with unit-variance and we denote by  $r$  the covariance function. To illustrate our results we consider the following covariance functions :

- $r_1(t) = \exp(-t^2/2)$  the Gaussian covariance,
- $r_2(t) = (\cosh(t))^{-1}$  the inverse of the hyperbolic cosine,
- $r_3(t) = (\sqrt{3}t)^{-1} \sin(\sqrt{3}t)$  the covariance of a low-pass process that is a white noise passed through a low-pass filter,
- $r_4(t) = e^{-\sqrt{5}|t|} \left( \frac{\sqrt{5}|t|^3}{3} + 2t^2 + \sqrt{5}|t| + 1 \right)$  the normalization of the convolution  $r_0 * r_0 * r_0 * r_0$  of the Ornstein-Uhlenbeck covariance  $r_0(t) = \exp(-|t|)$ ,
- $r_5(t) = \cos(t)$  the covariance of the cosine process,
- $r_6(s, t) = \frac{\exp(st)-1}{\sqrt{\exp(s^2)-1}\sqrt{\exp(t^2)-1}}$  the covariance of the Gaussian process giving the asymptotic distribution of the log-likelihood ratio test in the simple Gaussian mixture, see Ghosh and Sen (1985), Gassiat (2002) and Azaïs et al. (2004).

We insist on the fact that  $r_6$  is non-stationary whereas the others are stationary. For the stationary covariances the time has been normalized such that  $\text{Var}(X'(t)) \equiv 1$ . This normalization, sometimes called "unit-speed" transformation, leads to a very simple form of the right-hand member of (5.1) :  $\Psi(u) + \varphi(u) \frac{T}{\sqrt{2\pi}}$ .

Numerical applications have been run under MATLAB. Our procedure is entirely based on `rind.m` from the version 2.0.5 of WAFO, written by Brodtkorb et al. (2000). This routine computes Gaussian integrals that may be singular and offers several methods of integration. We made all computations with the method associated to Genz algorithm (1992), taking the option `Nit = -1`. The integration in the upper bound (5.7) is carried out by the MATLAB function "quad" which approximates the integral of a function, within a given error, using recursive adaptive Simpson quadrature.

In theory larger is the parameter  $n$ , the better are the bounds (5.7) and (5.8). But the authors of the routine advise to integrate in  $\mathbb{R}^{100}$ . Unfortunately, the time of estimation of the function `UBPn` may become too long for large values of  $n$ . We think that the value  $n = 30$  is a satisfactory compromise.

Moreover for small  $n$ , the survival function is clearly under estimated by (5.8). Whereas when  $n$  is too large in comparison with  $T$ , the covariance matrix associated to elements of the indicator function of (5.8) is almost singular. Since the true value could be overestimated, for middle value of  $T$ , we decide to make a slight randomization for its estimation.

- ★ `UBPn(T, u, n, r)` estimates the upper bound (5.7).
- ★ `UBP(T, u, r) = UBPn(T, u, 30, r)`.
- ★ `LBPn(T, u, n, r)` estimates the lower bound (5.8).
- ★ `LBP(T, u, r)` gives the mean of `LBPn(T, u, n, r)` computed for  $n = 10, 15, 20, \dots, 100$ .
- ★ `RICE(T, u)` estimates the upper bound (5.1) which is trivial except for the covariance  $r_6$ .

These functions are available on the web site Mercadier (2005).

All results are graphically presented with interpolated values by the MATLAB function “spline” (Cubic spline data interpolation).

### 5.3.2 Checking the program and evaluating the error

On a theoretical side the cosine process, corresponding to the covariance  $r_5$ , is the only smooth process from our list for which the distribution of the maximum  $M_T$  is known. Indeed, since Berman (1971) and Delmas (2003), one has for any positive  $u$

$$\mathbb{P}(M_T \geq u) = \begin{cases} \Psi(u) + \varphi(u) \frac{T}{\sqrt{2\pi}} & \forall 0 < T < \pi, \\ \Psi(u) + \varphi(u) \frac{T}{\sqrt{2\pi}} - \int_{\pi}^T \exp\left(-\frac{u^2(1-\cos(t))}{\sin(t)^2}\right) \frac{dt}{2\pi} & \forall \pi \leq T < 2\pi, \\ \exp\left(-\frac{u^2}{2}\right) & \forall T \geq 2\pi. \end{cases}$$

Due to the particular form of  $\mathbb{P}(M_T \geq u)$ , let us fix a value of  $u$  and study the function of the parameter  $T$ . We present our numerical results in Table 5.1.

	$T = 0.5$	$T = 1.5$	$T = 3.1$	$T = 4.5$	$T = 10$	$T = 15$
UBP	0.3788	0.5192	0.7439	0.8709	0.8866	0.8840
True	0.3788	0.5192	0.7439	0.8709	0.8825	0.8825
LBP	0.3786	0.5191	0.7438	0.8699	0.8797	0.8793

TAB. 5.1: Estimation of  $\mathbb{P}(M_T \geq 0.5)$  for  $r_5$  (cosine process). For a given  $T$ , we write UBP, the true value and LBP.

We observe that the upper bounds obtained by the function UBP are exact for  $T < \pi$ : it is not surprising since Rice upper bound is exact on this particular domain. In general we see that the program gives in each case a good interval. But this process is very special since it is periodic. In Table 5.2 we compare our values with another work concerning our first covariance function that is less particular.

	$u = -2$	$u = -1$	$u = 0$	$u = 1$	$u = 2$	$u = 3$
UBP	0.9944	0.9278	0.6526	0.2540	0.0442	0.0031
Estimates	0.9944	0.9279	0.6527	0.2541	0.0442	0.0031
LBP	0.9945	0.9281	0.6527	0.2539	0.0430	0.0025

TAB. 5.2: Estimation of  $\mathbb{P}(M_1 \geq u)$  for  $r_1$ . For a given  $u$ , we write UBP, estimation of Azaïs et al. (1999, page 112) and LBP.

Clearly the estimation of (5.7) and (5.8) by `rind.m` is not exact (see the case Table 5.2,  $u = -1$ ). To present most performing results we restrict our experiences on  $T \leq 25$  and  $u \geq 1$ . Generally speaking we think that on this domain we can trust the result up to  $10^{-3}$ .

### 5.3.3 Elements to answer to : “how far is Rice bound numerically significant ?”

The preceding study cases allow us to assert that we dispose, in some cases, of a new method to estimate the value  $\mathbb{P}(M_T \geq u)$ . But one can wonder whether it makes a significant improvement compared to the classical Rice upper bound (5.1). In Figure 5.1 we present both estimations of the function  $T \mapsto \mathbb{P}(M_T \geq 1)$  for two different covariances.

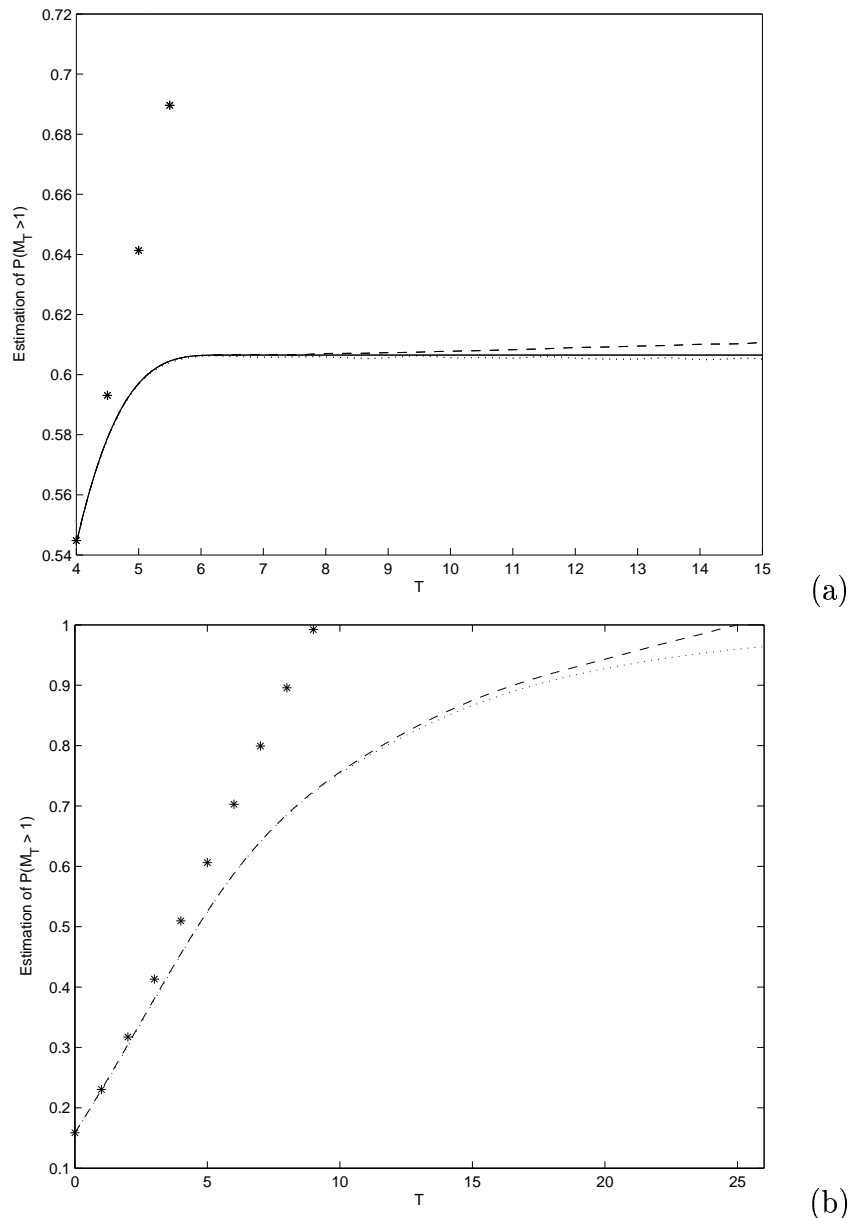


FIG. 5.1: (a) Estimation of  $\mathbb{P}(M_T \geq 1)$  for  $r_5$ . From top to bottom, RICE (star), UBP (broad dotted line), true values (solid line), LBP (fine dotted line). (b) Estimation of  $\mathbb{P}(M_T \geq 1)$  for  $r_6$ . From top to bottom, RICE (star), UBP (broad dotted line) and LBP (fine dotted line).

We will say that an upper bound is numerically significant if its absolute and relative errors are respectively upper bounded by  $10^{-2}$  and  $10^{-1}$ . When the true value is unknown we compute these errors with respect to the mean of (LBP,UBP).

In Table 5.3 we estimate, for a given value of  $u$  and  $r$ , the maximal size of  $T$  leading to a significant Rice upper bound and UBP bound.

	$u = 1$		$u = 1.5$		$u = 2$		$u = 2.5$		$u = 3$	
	RICE	UBP	RICE	UBP	RICE	UBP	RICE	UBP	RICE	UBP
$r_1$	2	20	3	20	8	20	19	>25	>25	>25
$r_5$	4	20	4	20	5	>25	6	>25	6	>25
$r_6$	1	20	3	20	7	20	17	>25	>25	>25

TAB. 5.3: For a given  $(r, u)$  the first (resp. second) column gives the maximal size of  $T$  for which Rice upper bound (resp. UBP bound) remains numerically significant.

The preceding table proves the efficiency of UBP in comparison with Rice upper bound, particularly for not large values of the level  $u$ .

Actually, this paper does not only propose an upper bound but a complete estimating method : the couple (LBP,UBP). Hence let us now compare our method to that of Azaïs and Wschebor (2002) which complexity is closed to ours. Indeed, their upper bound (denoted by UBAW) is constructed with the first three terms of the Rice series, their lower bound (denoted by LBAW) with the first two terms. For instance, take  $u = 1$  and the covariance  $r_1$ , our method becomes clearly more efficient when  $T$  is greater than 10.

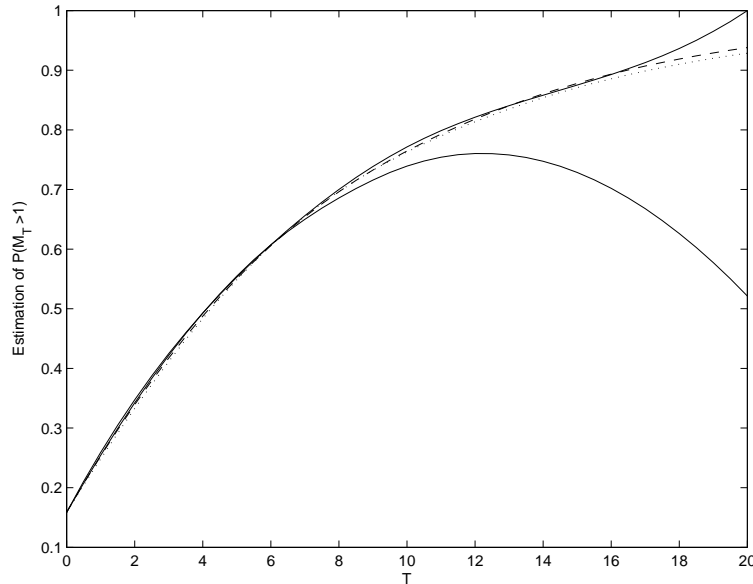


FIG. 5.2: Estimation of  $\mathbb{P}(M_T \geq 1)$  for the covariance  $r_1$ . From top to bottom we plot UBAW, UBP (broad dotted line), LBP (fine dotted line) and LBAW from (Azaïs and Wschebor, 2002, page 342).

### 5.3.4 Estimation of the persistence exponent

We restrict here our attention on the behavior of  $-1/T \ln \mathbb{P}(M_T < u)$  as  $T$  tends to infinity and  $u$  is fixed, a problem mainly dealt with in the literature of the physicists. According to Slepian (1962) and Li and Shao (2004) we know that if  $X$  is an almost surely continuous centered stationary Gaussian process with positive covariance  $r$  then for every  $u$  in  $\mathbb{R}$  this limit exists with

$$q(u) := \lim_{T \rightarrow +\infty} -\frac{1}{T} \ln \mathbb{P}(M_T < u) = \inf_{T > 0} -\frac{1}{T} \ln \mathbb{P}(M_T < u).$$

So let us consider a process  $\{X(t), t \in [0, T]\}$  satisfying such hypotheses. Since the determination of  $q(u)$  remains an open problem, our goal is to describe a method leading to a lower and an upper bound of this exponent  $q_{\text{LB}}(u) \leq q(u) \leq q_{\text{UB}}(u)$ . Taking into account the value for  $T$  equals to zero, a general but reasonable approximation of the probability distribution function is given by

$$(5.9) \quad \mathbb{P}(M_T < u) \sim \Phi(u) \exp(-q(u)T).$$

This formula explains the term of persistence exponent used for  $q(u)$ . So we will also deal with the following definition

$$q(u) = \lim_{T \rightarrow +\infty} -\frac{1}{T} \ln \frac{\mathbb{P}(M_T < u)}{\Phi(u)}$$

which loss the inf representation. Very user-friendly and efficient, the function LBPn gives a lower bound

$$(5.10) \quad q_{\text{LB}}(u) = \lim_{T \rightarrow +\infty} -\frac{1}{T} \ln \frac{1 - \text{LBPn}(T, u, n_T, r)}{\Phi(u)}.$$

Let us start with the most difficult case that is when  $u$  is equal to zero. For the covariance  $r_2$ , Li and Shao (2002) established that  $b = 2q(0)$  lies in  $(0.5, 1]$ . Simulations of Dembo et al. (2002) suggest  $b = 0.76 \pm 0.03$ .

Let us consider our first four covariances. In practice we noted that the limit value (5.10) is achieved from  $T$  equals to 10. Since  $T$  is quit large, we restrict ourselves to values of  $n_T$  greater than 40. So this limit is computed as the mean value on  $T$  in  $\{10, 15, 20\}$  and all the integers  $n_T$  included in  $[40, 100]$ . Figure 5.3 presents the details of our results on the covariance  $r_2$ .

Let us summarize in Table 5.4 our lower bound estimates for each covariance (including  $r_3$  which does not satisfy the positivity hypothesis of Li and Shao).

Remark that our conclusions for the covariance  $r_2$  are in accordance with theoretical and numerical results previously cited. We observe that the persistence exponent  $q(0)$  depends clearly on other characteristics than the local properties since for all considered processes we have at zero

$$r(t) = 1 - t^2 + O(t^4).$$

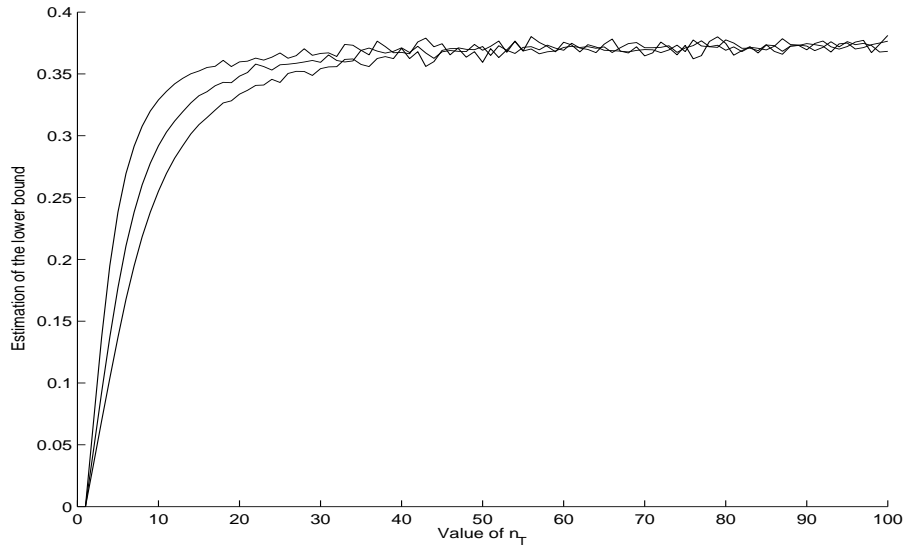


FIG. 5.3: Convergence study to obtain an estimate of (5.10) for the covariance  $r_2$ . Representation of the function, with respect to  $n_T$ ,  $-1/T \ln(1 - \text{LBPn}(T, 0, n_T))$  when  $T = 10, 15$  and  $20$ .

	$r_1$	$r_2$	$r_3$	$r_4$
$q_{\text{LB}}(0)$	0.4143	0.3704	0.4521	0.3879

TAB. 5.4: Lower bound of the persistence exponent  $q(0)$  for our first four covariances.

Thanks to the inf representation,  $q(u)$  is upper bounded by

$$(5.11) \quad q_{\text{UB}}(u) = \inf_{T>0} -\frac{1}{T} \ln(1 - \text{UBPn}(T, u, n, r))$$

and consequently also by

$$(5.12) \quad q_{\text{RICE}}(u) = \inf_{T>0} -\frac{1}{T} \ln(1 - \text{RICE}(T, u, r)).$$

Unfortunately our procedure  $\text{UBPn}$  is unstable to derive an upper bound of  $q(0)$ . On the other hand  $q_{\text{RICE}}(0) = 0.8525$  is less instructive than the theoretical bound 0.5 only known in the case of  $r_2$ .

Now we deal with the case when  $u$  differs from zero, say equals to one or two. It seems that the persistence exponent  $q(u)$  becomes practically the same for every process as we can see in the table 5.5.

Here is an heuristic property of this behavior. For small values of  $T$ , it is known that

$$(5.13) \quad \mathbb{P}(M_T < u) \sim \Phi(u) + T \frac{\exp(-u^2/2)}{2\pi}.$$

	$r_1$	$r_2$	$r_3$	$r_4$
$q_{\text{LB}}(1)$	0.1265	0.1173	0.1308	0.1211
$q_{\text{UB}}(1)$	0.1445	0.1367	0.1504	0.1416
$q_{\text{RICE}}(1)$	0.1960	0.1960	0.1960	0.1960
$q_{\text{LB}}(2)$	0.0222	0.0213	0.0223	0.0217
$q_{\text{UB}}(2)$	0.0244	0.0237	0.0244	0.0240
$q_{\text{RICE}}(2)$	0.0271	0.0271	0.0271	0.0271

TAB. 5.5: Estimation of persistence exponents  $q(1)$  and  $q(2)$  for our first four covariances.

The only function of the type given by the left-hand member of (5.9) which has the Taylor expansion given by (5.13) is obtained after substitution of  $q(u)$  by  $h(u) = \exp(-u^2/2) / (2\pi\Phi(u))$ . Whereas estimations of Table 5.4 were calculated for large values of  $T$ , we remark that they are, in the case of  $u = 2$ , close to this very heuristic equivalent :  $h(2) = 0.0220$ .

## 5.4 Numerical applications for non-smooth Gaussian processes

Let us consider in this section a centered Gaussian process  $X$  whose sample paths are continuous but not differentiable. Again we wish to know the distribution of its maximum. Obviously we cannot apply Formula (5.6) in order to derive upper bounds for the survival function. Two preliminary approaches have been examined.

- (M1) Use of the upper bound (5.7) to smoothed equivalents of  $X$ .
- (M2) Use of the upper bound (5.7) where  $X'(t)$  is approximated by some  $(X(t) - X(s))/(t - s)$ .

The method (M2) is theoretically justified by the result of Durbin (1985). However they both lead to numerical procedures that are very unstable. Consequently, since LBPn is very accurate, we propose to set up in this case an estimating method based on the discretization.

We begin by the description of the general ideas. Let  $\{X(t), t \in [0, T]\}$  be a centered Gaussian process and let  $\{t_1, \dots, t_n\}$  be a subdivision of  $(0, T)$ . We assume that the interval  $[0, T]$  is the disjoint union of the intervals  $[t_k - d, t_k + d]$  for a positive real  $d$ . So we have the lower bound

$$(5.14) \quad \mathbb{P}(M_T \geq u) \geq 1 - \mathbb{P}(X(t_k) < u, \forall k = 1, \dots, n).$$

Now suppose that we know the local behavior of the process  $X$  in the neighborhood on a fixed point. More precisely, assume that there exists a deterministic function  $\mathcal{L}$ , increasing on  $[0, d]$ , such that for all  $t^*$  we have almost surely

$$\limsup_{t \rightarrow t^*} \frac{|X(t) - X(t^*)|}{\mathcal{L}(|t - t^*|)} = 1.$$



Then we can construct an heuristic upper bound for the survival function of  $M_T$  applying the last equivalent to the discretization points. Define  $M_{[a,b]}$  as the random variable  $\sup_{t \in [a,b]} X(t)$ . By compactness each maximum is achieved at a point  $t_k^*$  of  $[t_k - d, t_k + d]$  and by an abusive use of the superior limit, we get for each value of  $k$

$$M_{[t_k-d, t_k+d]} - X(t_k) \lesssim \mathcal{L}(d).$$

In consequence we propose the following ‘‘upper bound’’

$$\begin{aligned} \mathbb{P}(M_T \geq u) &= 1 - \mathbb{P}(M_T < u) \\ &= 1 - \mathbb{P}(M_{[t_k-d, t_k+d]} < u, \forall k = 1, \dots, n) \\ (5.15) \quad &\lesssim 1 - \mathbb{P}(X(t_k) < u - \mathcal{L}(d), \forall k = 1, \dots, n). \end{aligned}$$

For numerical implementations within the non-smooth framework we choose the advised limit value of `rind.m` which computes Gaussian integrals in spaces included in  $\mathbb{R}^{100}$ . Consequently we define the MATLAB functions `LBNS(T, u, r)` and `UBNS(T, u, r)` that estimate (5.14) and (5.15) by calling `LBPn(T, u, 100, r)` and `LBPn(T, u - \mathcal{L}(d), 100, r)` respectively. They are also available on the web site Mercadier (2005).

In the stationary case, we deal with the Ornstein-Uhlenbeck process whose covariance function has been already introduced as  $r_0(t) = \exp(-|t|)$ . Locally the behavior of such diffusion process is the same to that of the Brownian motion. They satisfy the law of the iterated logarithm and we have for every positive  $\tau$

$$\mathcal{L}(\tau) = \sqrt{2\tau \log |\log(1/\tau)|}.$$

This function is increasing on  $[0, d]$  if  $d \leq 0.17$ . We take for every value of  $k$ ,  $t_k = kT/n$  and  $d = T/2n$ . In that way we cover the domain  $[T/2n, T+T/2n]$  which by stationarity is the same as  $[0, T]$ . DeLong (1981, page 2210) gives the values of

$$\mathbb{P}\left(\sup_{1 < t < T} \frac{B(t)}{\sqrt{t}} \geq u\right)$$

where  $B(t)$  is the Brownian motion. We derive that of  $\mathbb{P}(M_T \geq u)$  for the covariance  $r_0$  from the transformation

$$X(t) = B(\exp(2t))/\sqrt{\exp(2t)}.$$

Figure 5.4 (a) presents our results for three different values of  $u$ .

For the non-stationary case, we make the study on the Brownian motion of covariance  $r_7(s, t) = \min(s, t)$ . As said before the function  $\mathcal{L}$  is given by the law of the iterated logarithm. We make a slight modification of the subdivision in comparison with the stationary case. We cover the interval  $[T/2n, T]$  by

$$\cup_{k=1}^{n-1} [kT/n \pm T/2n] \cup [T - T/4n \pm T/4n].$$

Since the values of the process are negligible around zero we estimate  $M_T$  by  $M_{[T/2n, T]}$ . We represent our estimations for  $T = 1$  in Figure 5.4 (b).

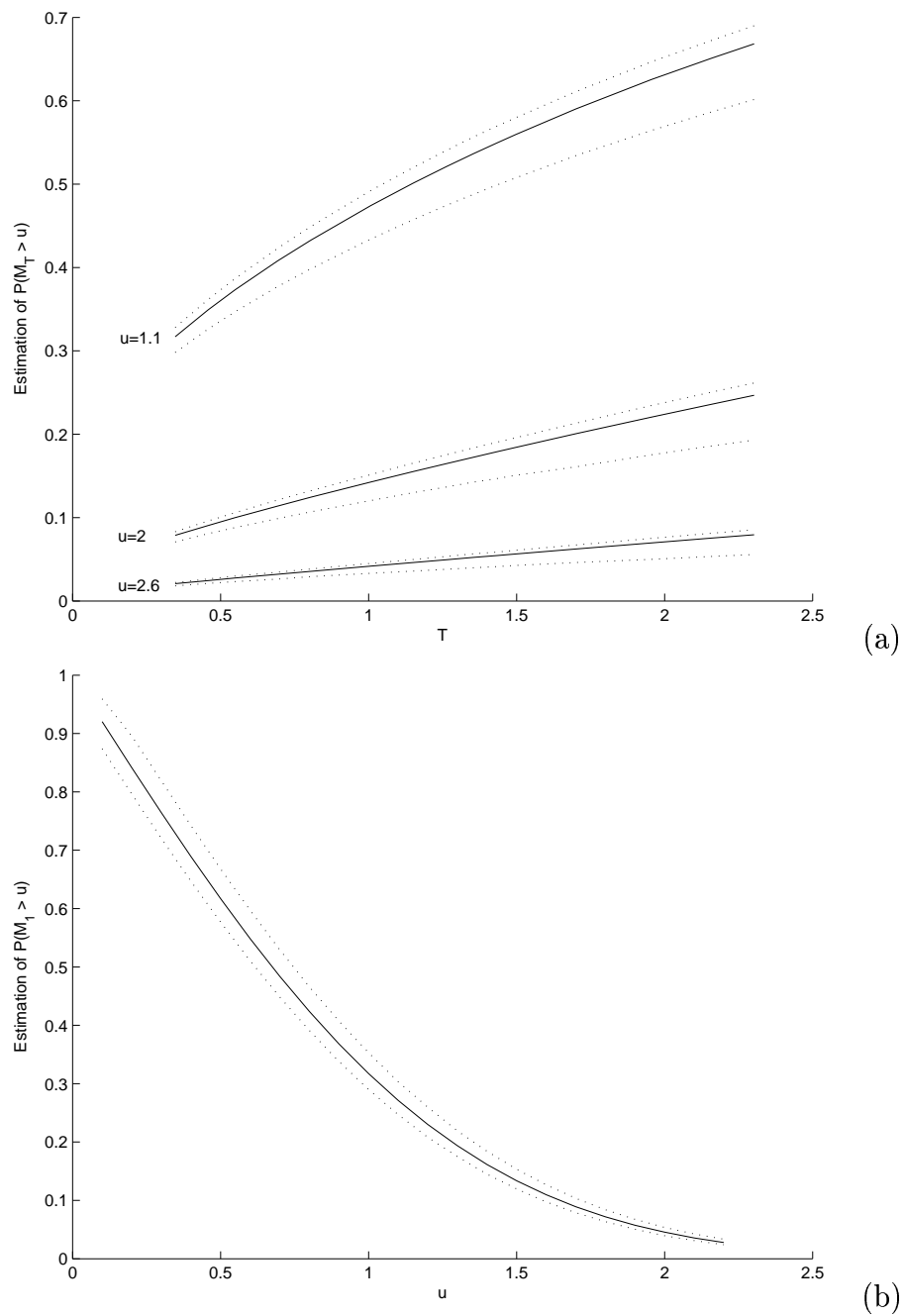


FIG. 5.4: (a) Estimation of  $\mathbb{P}(M_T \geq u)$  for  $r_0$  and three values of  $u$ . From top to bottom, we find UBNS, true values (DeLong, 1981, page 2210) and LBNS. (b) Estimation of  $\mathbb{P}(M_1 \geq u)$  for  $r_7$ . From top to bottom, we find UBNS, true values from reflection principle and LBNS.

We therefore conclude that the local behavior of a process is an interesting tool to construct approximations of the distribution of its global maximum, in particular for the stationary case. Moreover we remark that this method leads to correct bounds even in the non-stationary case.

## 5.5 Two-parameter Gaussian processes

The aim of this part is to obtain the distribution of the maximum of a two-parameter Gaussian process. We present a formula made up of few terms that enable us to compute some numerical bounds.

We restrict the study to particular subsets  $S$  of the plane that satisfy Hypotheses (H) :

(H1)  $S$  is compact,

(H2)  $S$  is simply connected,

(H3) There exists a positive real  $L$  and a function  $\rho : [0, L] \rightarrow \partial S$  such that  $\rho$  is a  $\mathcal{C}^1$ -diffeomorphism except at a finite number of points for whose  $\rho$  is just continuous.

In all generality, we could assume that  $L = 1$ . However the correspondence between our next result and its application to  $S = [0, T]^2$  in Section 5.6 seems clearer if we do not. Remark that extremal points of  $[0, L]$  are identified.

Moreover, we assume without loss of generality that the parameterization  $\rho$  satisfies  $\|\rho'(\theta)\|_2 = 1$  for every  $\theta$  in  $[0, L]$ .

We take the following notations. Denote by  $\pi_1$  and  $\pi_2$  the orthogonal projectors on the x-axis and the y-axis respectively. When  $t$  is a point of  $S$  we write also  $t_i$  for  $\pi_i(t)$ . For  $s$  and  $t$  in  $S$  and for all positive integers  $i, j, k$  and  $l$ , we set

$$X_{ij}(t) = \frac{\partial^{i+j}}{\partial t_1^i \partial t_2^j} X(t).$$

The Gaussian process  $Y$  is the restriction of  $X$  on the boundary  $\partial S$  defined by

$$Y(\theta) = X \circ \rho(\theta)$$

and we call  $Z$  the Gaussian field  $(X, X_{10})$ . Recall that  $M_S = \sup_{t \in S} X(t)$ .

The approach is the generalization of the method described in Section 5.2. Let us privilege a direction of the plane : we arbitrarily choose the y-axis. Along the one-dimensional study, the point  $t = 0$  is viewed as the starting point of the interval  $[0, T]$ . We do the same here assuming, without loss of generality, that  $\pi_2(\rho(0)) = \inf \pi_2(S)$  and taking  $\rho(0)$  as the starting point of the subset  $S$ .

For the one-parameter process, we dealt with the event  $\{X(0) < u, M_T \geq u\}$  by researching the smallest point  $t$  associated to an upcrossing of the level  $u$ . By analogy we will research points of

$$\{t \in \partial S, X(t) = u\} \text{ or } \{t \in \overset{\circ}{S}, X(t) = u, X_{10}(t) = 0\}$$

with minimal y-coordinate. We will see that it can be expressed by means of the subset

$$\Gamma_t = \{s \in S, s_2 \leq t_2\}.$$

**Theorem 5.5.1**

Set  $\{X(t), t \in S\}$  a real valued Gaussian field defined on a subset  $S$  of  $\mathbb{R}^2$  that satisfies Hypotheses (H). Assume that

- (A<sub>1</sub>) the sample paths of  $Z = (X, X_{10})$  are almost surely continuously differentiable,
- (A<sub>2</sub>) for  $W$  fixed in  $\mathbb{R}^2$ , there is almost surely no point  $t$  such that  $Z(t) = W$  and  $\det(Z'(t)) = 0$ ,
- (A<sub>3</sub>) for  $t$  in  $S$ , the distribution of  $Z(t)$  is non-degenerate,
- (A<sub>4</sub>) for  $(s, t)$  in  $S^2$  with  $s_2 = t_2$  and  $s \neq t$ , the joint distribution of  $Z(s), Z(t)$  is non-degenerate.

Then for every  $u$  in  $\mathbb{R}$ ,

$$(5.16) \quad \mathbb{P}(M_S \geq u) = \mathbb{P}(Y(0) \geq u) + \int_0^L \mathbb{E} \left( |Y'(\theta)| \mathbb{1}_{\{X(s) \leq u, \forall s \in \Gamma_{\rho(\theta)}\}} / Y(\theta) = u \right) p_{Y(\theta)}(u) d\theta + \int_S \mathbb{E} (X_{20}(t)^- X_{01}(t)^+ \mathbb{1}_{\{X(s) \leq u, \forall s \in \Gamma_t\}} / X(t) = u, X_{10}(t) = 0) p_{X(t), X_{10}(t)}(u, 0) dt.$$

We refer to Azaïs and Wschebor (2005) for sufficient conditions for (A<sub>2</sub>) to hold. The distribution of the maximum of a two-parameter Gaussian field is implicitly known by means of two integrals. The interest of this formula may appear limited since the maximum appears in both members of the equality. However we will see in Section 5.6 that some accurate upper bounds can be deduced.

*Proof of Theorem 5.5.1.* Let us begin by studying the event

$$\{\exists t \in S, X(t) = u, X(s) \leq u, \forall s \in \Gamma_t\}.$$

Remark that we have almost surely the equality

$$\{\exists t \in \overset{\circ}{S}, X(t) = u, X(s) \leq u, \forall s \in \Gamma_t\} = \{\exists t \in \overset{\circ}{S}, Z(t) = (u, 0), X(s) \leq u, \forall s \in \Gamma_t\}.$$

The following lemma is easy using Bulinskaya method.

**Lemma 5.5.1**

Let  $I$  be a subset of  $\mathbb{R}^d$  of dimension  $m$ . Let  $U$  be a random field defined on an open neighborhood  $O$  of  $I$ , with values in  $\mathbb{R}^{m+k}$  for some positive  $k$ . We assume that the sample paths of  $U$  are almost surely continuously differentiable and that for every  $t$  in  $O$ , the density  $p_{U(t)}(W)$  is bounded in a neighborhood of  $W \in \mathbb{R}^{m+k}$ . Then

$$\mathbb{P}(\exists t \in I, U(t) = W) = 0.$$

According to (A<sub>1</sub>, A<sub>4</sub>), we apply this lemma with  $I_n = \{(s, t) \in \overset{\circ}{S}, s_2 = t_2, \|s - t\| > 1/n\}$  and  $U(s, t) = (Z(s), Z(t))$ . We deduce that almost surely there exists not a pair  $(s, t)$  in

$$I = \{(s, t) \in \overset{\circ}{S}, s_2 = t_2, s \neq t\} \text{ such that } Z(s) = (u, 0) \text{ and } Z(t) = (u, 0).$$

With the same method almost surely we can not find a pair  $(s, t) \in \partial S^2$  with  $s_2 = t_2$  and  $s \neq t$  such that  $X(s) = u$  and  $X(t) = u$ .

Moreover, the Ylvisaker's theorem implies that there is no local maximum with value

$u$ , so there exists not a pair  $(s, t) \in S^2$  with  $s_2 < t_2$  such that  $X(s) = u$ ,  $X(t) = u$  and  $X(z) \leq u, \forall z \in \Gamma_t$ .

As a consequence, almost surely

$$\{M_S \geq u\} = \{Y(0) \geq u\} \cup \{\exists t \in S, X(t) = u, X(s) \leq u, \forall s \in \Gamma_t\}$$

where the union is almost surely disjunctive and where the existence of such point  $t$  implies its uniqueness. So we get  $\mathbb{P}(M_S \geq u) = \mathbb{P}(Y(0) \geq u) + P_1(u) + P_2(u)$  where

$$\begin{aligned} P_1(u) &= \mathbb{E}(\#\{t \in \partial S, X(t) = u, X(s) \leq u, \forall s \in \Gamma_t\}), \\ P_2(u) &= \mathbb{E}(\#\{t \in \overset{\circ}{S}, Z(t) = (u, 0), X(s) \leq u, \forall s \in \Gamma_t\}). \end{aligned}$$

We start by computing the term  $P_1(u)$ . The associated number can be expressed by means of the function  $\rho$  as

$$\#\{\theta \in (0, L), Y(\theta) = u, X(s) \leq u, \forall s \in \Gamma_{\rho(\theta)}\}.$$

With the same arguments as those used in the proof of Theorem 5.2.1 and with the record points set  $E = \{\theta \in (0, L), X(s) \leq Y(\theta), \forall s \in \Gamma_{\rho(\theta)}\}$  we get for almost every  $u$  in  $\mathbb{R}$

$$(5.17) \quad P_1(u) = \int_0^L \mathbb{E} \left( |Y'(\theta)| \mathbb{1}_{\{X(s) \leq u, \forall s \in \Gamma_{\rho(\theta)}\}} / Y(\theta) = u \right) p_{Y(\theta)}(u) d\theta.$$

The main difference between the treatment of  $\int_0^T f(t, \cdot) dt$  in the proof of Proposition 5.2.1 and the right-hand member of (5.17) is that the process  $Y$  may have a finite number of points  $\theta$  for whose some regularity is lost. Consequently here we use the regularity assumption  $(A_1)$  to dominate this integrand independently of  $u$  by an integrable function. Then we also perform a Gaussian regression of  $X(s)$  with respect to  $Y(\theta)$  to deduce that this integral is a continuous function of  $u$ .

We turn to the study of the term  $P_2(u)$  that we compute thanks to next lemma (proved at the end) that is a generalization of the Rice Formula of Azaïs and Wschebor (2005).

**Lemma 5.5.2**

Set  $Z : S \subset \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be a Gaussian field satisfying  $(A_1)$ ,  $(A_2)$  and  $(A_3)$ . Assume that  $g : \mathcal{C}(S, \mathbb{R}) \times S \rightarrow \mathbb{R}$  is continuous and bounded, putting on  $\mathcal{C}(S, \mathbb{R})$  the sup norm. Then for every  $W$  in  $\mathbb{R}^2$

$$\mathbb{E} \left( \sum_{t \in \overset{\circ}{S} \cap Z^{-1}(\{W\})} g(Z, t) \right) = \int_{\overset{\circ}{S}} \mathbb{E} (|\det(Z'(t))| g(Z, t) / Z(t) = W) p_{Z(t)}(W) dt$$

where both members of the equation are continuous functions with respect to  $W$ .

Let  $(k_n)_n$  be a sequence of continuous and monotone functions such that  $k_n(x) = 1$  if  $x \leq 0$  and  $k_n(x) = 0$  if  $x > 1/n$ . We can verify that the limit with respect to  $n$  of  $k_n(x)$  is  $\mathbb{1}_{(-\infty, 0]}(x)$ . We apply Lemma 5.5.2 with  $W = (u, 0)$  and

$$g_n(Z, t) = k_n(\sup_{s \in \Gamma_t} Z(s) - u).$$

We get by monotone and dominated convergence arguments that

$$\begin{aligned} P_2(u) &= \int_{\overset{\circ}{S}} \mathbb{E} (|\det(Z'(t))| \mathbb{1}_{\{X(s) \leq u, \forall s \in \Gamma_t\}} / Z(t) = (u, 0)) p_{Z(t)}(u, 0) dt \\ &= \int_S \mathbb{E} (X_{20}(t)^- X_{01}(t)^+ \mathbb{1}_{\{X(s) \leq u, \forall s \in \Gamma_t\}} / X(t) = u, X_{10}(t) = 0) p_{X(t), X_{10}(t)}(u, 0) dt \end{aligned}$$

where both members are continuous functions of  $u$ .

To summarize, we know that the equality (5.16) is true for almost every  $u$  of  $\mathbb{R}$  and that its right-hand member is a continuous function of  $u$ . We deduce the whole equality by application of the theorem of Ylvisaker (1968).  $\square$

*Proof of Lemma 5.5.2.* Let  $\beta : \mathbb{R}^2 \rightarrow \mathbb{R}$  be a continuous function with compact support and let  $N_W^Z(I)$  denotes the number of solutions  $t$  in a compact subset  $I$  of  $\overset{\circ}{S}$  of the equation  $Z(t) = W$ . Applying the co-area formula of Federer (1969), we can prove that

$$(5.18) \quad \int_{\mathbb{R}^2} \beta(W) \sum_{t \in \overset{\circ}{S} \cap Z^{-1}(\{W\})} g(Z, t) dW = \int_{\overset{\circ}{S}} \beta(Z(t)) |\det(Z'(t))| g(Z, t) dt.$$

The main difficulty here is to dominate the number of elements in  $Z^{-1}(\{W\})$ . Define for each positive  $\eta$ , the subset  $I_\eta = \{t \in S, \|t - s\| \geq \eta, \forall s \notin S\}$  and consider a continuous non-decreasing function  $F$  with  $F(x) = 0$  for  $x \leq 1/2$  and  $F(x) = 1$  for  $x \geq 1$ . Let us take  $\beta(W)$  as the product  $\alpha_{\Delta, \eta}(W) f(W)$  where  $f$  is a continuous function with compact support and

$$\alpha_{\Delta, \eta}(u) = F\left(\frac{1}{2\Delta} \inf_{s \in I_\eta} [\lambda_{\min}(Z'(s)) + \|Z(s) - W\|]\right) \times \left(1 - F\left(\frac{2}{\Delta} \omega_{Z'}(\eta)\right)\right)$$

where  $\lambda_{\min}$  and  $\omega$  are respectively the smallest eigenvalue and the continuity modulus. The function  $\alpha_{\Delta, \eta}(W)$  is a tool to control the value of the number of solutions of  $Z(t) = W$  in  $I_\eta$ .

Taking the expectation of both members of (5.18) with above context we deduce an equality of the following two expressions for almost every  $W \in \mathbb{R}^2$

$$\begin{aligned} &\mathbb{E} \left( \alpha_{\Delta, \eta}(W) \sum_{t \in I_\eta \cap Z^{-1}(\{W\})} g(Z, t) \right) \\ &= \int_{I_\eta} \mathbb{E} (|\det(Z'(t))| \alpha_{\Delta, \eta}(W) g(Z, t) / Z(t) = W) p_{Z(t)}(W) dt. \end{aligned}$$

It is then sufficient to prove that both members are continuous function of  $W$  and to take the limit in  $\eta$  and  $\Delta$  to obtain the conclusion of Lemma 5.5.2. Each step can be written with the arguments of Azaïs and Wschebor (2005, Theorem 2.1).  $\square$

From the last theorem we obtain an upper bound, easily computable, in the case of an "unit-speed" Gaussian field. Since the literature only propose equivalents of such survival functions, this result is new and it is one of our main results.

### Corollary 5.5.1

Set  $\{X(t), t \in S\}$  a centered Gaussian field satisfying conditions of Theorem 5.5.1.

Assume that

$$(A_5) \text{Var}(X(t)) = 1,$$

$$(A_6) \text{Var}(X'(t)) = Id,$$

$$(A_7) \text{Var}(X_{20}(t)) \text{ is constant.}$$

Then for every  $u$  in  $\mathbb{R}$

$$(5.19) (M_S \geq u) \leq \Psi(u) + \frac{\lambda_1(\partial S)}{\pi} \exp(-u^2/2) + \frac{\lambda_2(S)}{(2\pi)^{\frac{3}{2}}} [c \varphi(\frac{u}{c}) + u \Phi(\frac{u}{c})] \exp(-u^2/2)$$

with  $c = (\text{Var}(X_{20}) - 1)^{\frac{1}{2}}$  and where  $\lambda_1$  and  $\lambda_2$  denote the Lebesgue measure on  $\mathbb{R}$  and  $\mathbb{R}^2$  respectively.

*Proof of Corollary 5.5.1.* In Formula (5.16) the integrand associated to the boundary of  $S$  is reduced to

$$\mathbb{E}(|Y'|) p_Y(u) = \frac{2}{\sqrt{2\pi}} \frac{\exp(-\frac{u^2}{2})}{\sqrt{2\pi}}$$

since for any  $\theta \in [0, L]$ ,  $\mathbb{E}(Y(\theta)) = \mathbb{E}(Y'(\theta)) = 0$ ,  $\text{Var}(Y(\theta)) = 1$  and  $\text{Var}(Y'(\theta)) = \|\rho'(\theta)\|_2^2 = 1$ .

Remark that for any  $t$  in  $S$ ,

- $X_{20}(t)$  and  $X_{01}(t)$  are both independent of  $X_{10}(t)$ ,
- $\text{cov}(X_{20}(t)X(t)) = -1$ ,
- $X_{20}(t)$  and  $X_{01}(t)$  are independent conditioning or not by  $X(t)$ ,
- $X_{01}(t)$  and  $X(t)$  are independent.

It follows the following upper bound of the integrand associated to the interior of  $S$

$$\mathbb{E}(X_{01}^+) \mathbb{E}(X_{20}^- / X = u) p_{X, X_{10}}(u, 0) = \frac{1}{\sqrt{2\pi}} [c \varphi(\frac{u}{c}) + u \Phi(\frac{u}{c})] \frac{\exp(-\frac{u^2}{2})}{2\pi}. \quad \square$$

We propose now some improvements of upper bound (5.19) in the particular case where  $S$  is some square  $[0, T]^2$ .

Actually similar improvements hold for every subset  $S$  whose parameterization has a derivative  $\rho'$  that is constant on a sub-interval of  $[0, L]$ .

**Corollary 5.5.2**

Set  $\{X(t), t \in [0, T]^2\}$  a Gaussian field satisfying conditions of Corollary 5.5.1. Then for every  $u$  in  $\mathbb{R}$

$$(5.20) \quad \mathbb{P}(M_{[0,T]^2} \geq u) \leq \text{UBUS}(T, u)$$

$$\text{with } \text{UBUS}(T, u) = \Psi(u) + \frac{T}{\pi} \exp(-u^2/2) + \frac{T^2}{(2\pi)^{\frac{3}{2}}} [c\varphi(\frac{u}{c}) + u\Phi(\frac{u}{c})] \exp(-u^2/2)$$

and  $c = (\text{Var}(X_{20}) - 1)^{\frac{1}{2}}$ .

Using the asymptotics of  $\Phi$  we see that this upper bound satisfies as  $u$  becomes large

$$(5.21) \quad \text{UBUS}(T, u) \sim \Psi(u) + \frac{T}{\pi} \exp(-u^2/2) + \frac{T^2 u}{(2\pi)^{\frac{3}{2}}} \exp(-u^2/2).$$

which is exactly the expansion EQT. This can be also expressed by saying that

$$0 \leq \text{UBUS}(T, u) - \text{EQT}(T, u) \leq \frac{T^2}{\sqrt{2\pi^2}u^2} \exp(-3/4u^2).$$

*Proof of Corollary 5.5.2.* The change between (5.19) and (5.20) comes from the second term, corresponding to the boundary of  $S$ . We begin by an application of Ylvisaker (1968) saying that almost surely the maximum does not take the fix value  $u$  as local extremum. Consequently in (5.16) the second integral is reduced to an integration on  $[0, T] \times \{0\} \cup \{0\} \times [0, T] \cup \{T\} \times [0, T]$ . The integral on  $[0, T] \times \{0\}$  is upper bounded by

$$\int_0^T \mathbb{E}(X_{10}(\theta, 0)^+ / X(\theta, 0) = u) p_{X(\theta, 0)}(u) d\theta = \frac{T}{\sqrt{2\pi}} \varphi(u).$$

Taking into account that  $\{X(s) \leq u, \forall s \in \Gamma_{(0, \theta)}\} \subset \{X_{10}(0, \theta) \leq 0\}$  under the condition  $\{X(\theta, 0) = u\}$ , the integral on  $\{0\} \times [0, T]$  is upper bounded by

$$\int_0^T \mathbb{E}(X_{01}(0, \theta)^+ \mathbb{1}_{\{X_{10}(0, \theta) \leq 0\}} / X(0, \theta) = u) p_{X(0, \theta)}(u) d\theta = \frac{T}{2\sqrt{2\pi}} \varphi(u).$$

Similar arguments imply that the integral on  $\{T\} \times [0, T]$  is upper bounded by

$$\int_0^T \mathbb{E}(X_{01}(T, \theta)^+ \mathbb{1}_{\{X_{10}(T, \theta) \geq 0\}} / X(T, \theta) = u) p_{X(T, \theta)}(u) d\theta = \frac{T}{2\sqrt{2\pi}} \varphi(u).$$

So the second term of (5.19) can be replaced by  $T \exp(-u^2/2)/\pi$ . □



## 5.6 Numerical applications for two-parameter Gaussian processes

This section develops the numerical interest of Formula (5.16) giving the survival function of the maximum of a two-parameter Gaussian field defined on a compact domain. We restrict here our attention on isotropic centered Gaussian fields of unit variance defined on  $[0, T]^2$ . A lower bound is obtained as in Section 5.3 by single discretization of  $S$  taking the domain  $\nabla_{T,n} = \{(kT/n, lT/n), k = 0, \dots, n, l = 0, \dots, n\}$ . We have

$$(5.22) \quad \mathbb{P} (M_{[0,T]^2} \geq u) \geq \text{LBF} (T, u, n)$$

where  $\text{LBF} (T, u, n) = 1 - \mathbb{P} (X (s) < u, \forall s \in \nabla_{T,n})$ .

Now we deduce an upper bound by discretization of  $\Gamma_t$  that appears in the indicator functions of the conditional expectations of (5.16) taking  $\Gamma_{t,n} = \Gamma_t \cap \nabla_{T,n}$ . In the next bound, we apply the same argument as in the proof of Corollary 5.5.2 to reduce in Formula (5.16) the integration on the boundary of  $[0, T]^2$  to that on  $[0, T] \times \{0\} \cup \{0\} \times [0, T] \cup \{T\} \times [0, T]$ . Using the isotropic assumption, the field on the last two segments has the same properties. Thus we get

$$(5.23) \quad \mathbb{P} (M_{[0,T]^2} \geq u) \leq \text{UBF} (T, u, n_p, n_q, n_r)$$

with  $\text{UBF} (T, u, n_p, n_q, n_r) = \Psi (u) + \varphi (u) [P_{n_p} (T, u) + 2Q_{n_q} (T, u) + R_{n_r} (T, u)]$  and

$$P_n (T, u) = \int_0^T \mathbb{E} \left( X_{10} (\theta, 0)^+ \mathbb{1}_{\{X(s) \leq u, \forall s \in \Gamma_{(\theta,0),n}\}} / X (\theta, 0) = u \right) d\theta,$$

$$Q_n (T, u) = \int_0^T \mathbb{E} \left( X_{01} (0, \theta)^+ \mathbb{1}_{\{X(s) \leq u, \forall s \in \Gamma_{(0,\theta),n}\}} / X (0, \theta) = u \right) d\theta,$$

$$R_n (T, u) = \frac{1}{\sqrt{2\pi}\Sigma_{11}} \iint_{[0,T]^2} \mathbb{E} (X_{20} (t)^- X_{01} (t)^+ \mathbb{1}_{\{X(s) \leq u, \forall s \in \Gamma_{t,n}\}} / X (t) = u, X_{10} (t) = 0) dt.$$

Our goal is to compare this estimating method to the known expansions (5.2) and (5.3) presented in the introduction. To this end, we consider the Gaussian covariance function

$$r_8 (s, t) = \exp \left( -\frac{\|s - t\|^2}{2} \right).$$

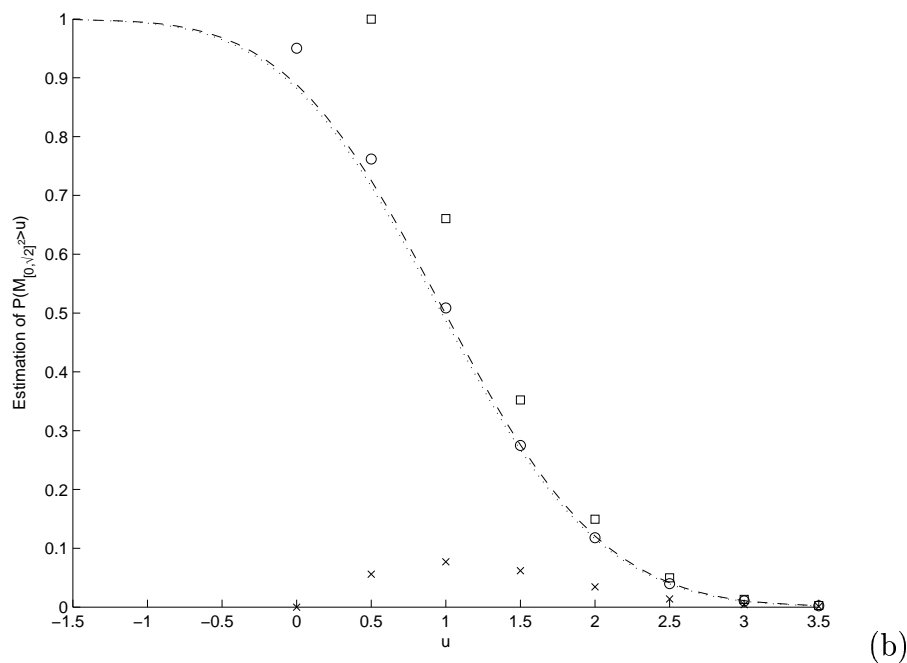
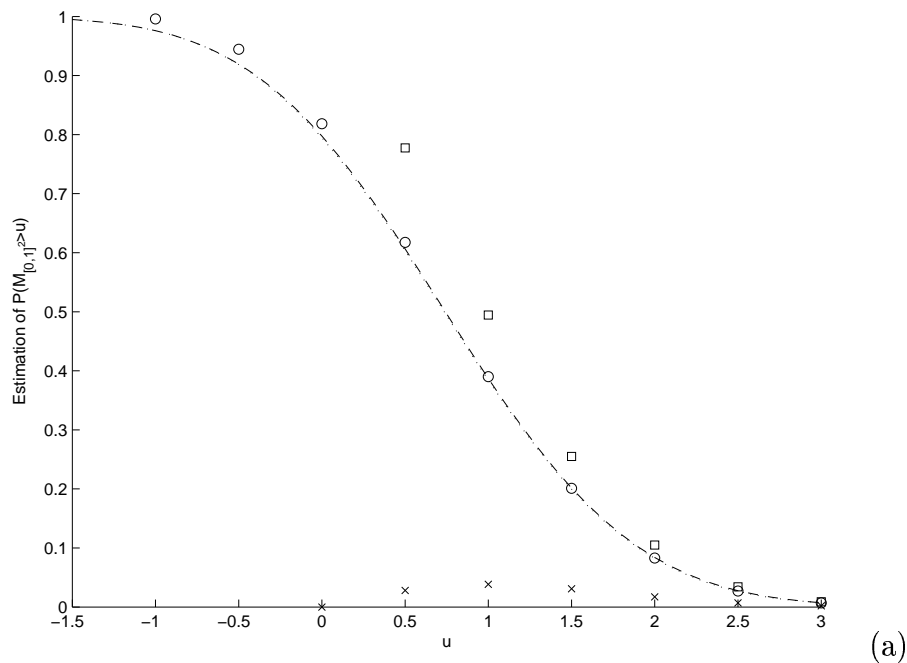
In order to compute UBF and LBF we call the routine `rind.m` of WAFO which computes Gaussian integrals (see Section 5.3 for details on this tool-box).

★  $\text{UBF}(T, u)$  estimates the upper bound (5.23) with  $n_p = 100, n_q = 15$  and  $n_r = 10$ .

★  $\text{LBF}(T, u)$  gives the estimation of (5.22) for  $n = 10$ .

These function are available on the web site Mercadier (2005).

The difference between the order of  $n_p$  and the others parameters comes from the fact that for all  $\theta$  in  $[0, T]$  we have that  $\Gamma_{(\theta,0),n}$  is included in  $[0, T]$ . Considering the definition of  $\nabla_{T,n}$ , it is important to remark that when  $n_q = 15$  we consider actually  $15^2$  points  $s$  in the indicator function associated to  $Q$ .



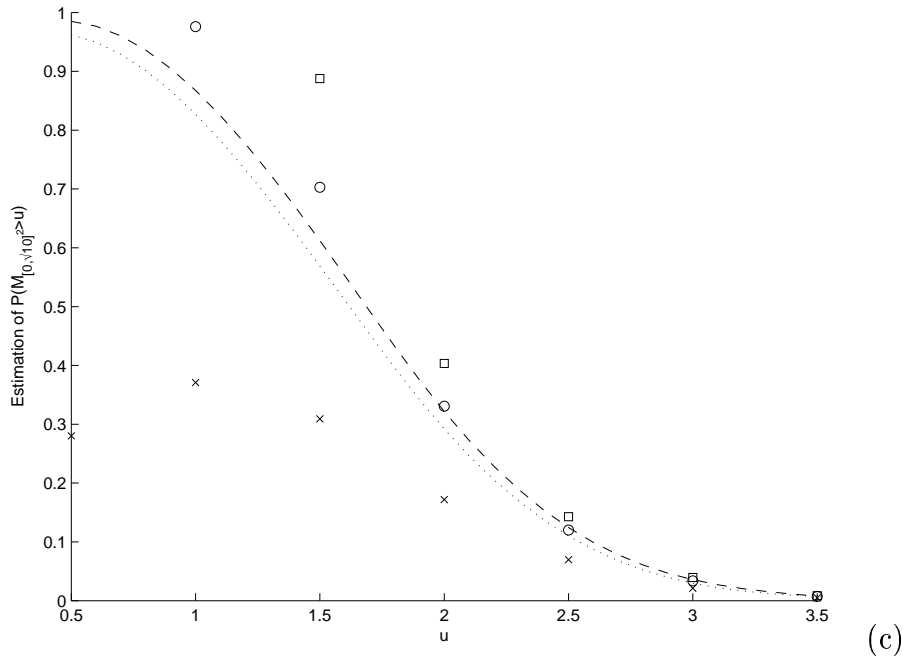


FIG. 5.5: (a) Estimation of  $\mathbb{P}(M_{[0,T]^2} \geq u)$  for  $T^2 = 1$ . From top to bottom, UBUS (square), EQT (circle), UBUS (broad dotted line), LBF (fine dotted line) and EQA (cross). (b) Estimation of  $\mathbb{P}(M_{[0,T]^2} \geq u)$  for  $T^2 = 2$ . (c) Estimation of  $\mathbb{P}(M_{[0,T]^2} \geq u)$  for  $T^2 = 10$ .

The first comment concerning these numerical results is the goodness of the expansion EQT when  $T$  is small with respect to  $u$ . It is also interesting to remark that the upper bound UBUS converges relatively quickly to the true value for middle levels. Consequently such bound could be of interest to justify properly the use of EQT in statistical tests for instance. Finally for small surfaces we can assure that our method gives the distribution of the maximum. For larger surfaces, the method remains accurate taking into account the complexity of this problem.

## Acknowledgments

The author wishes to thank Professor I. Rychlik for his suggestions concerning the field case, and his assistance for the use of WAFO. The author also thanks Professors J.-M. Azais and M. Wschebor for authorize the use of their data in Fig. 5.2 and for their helpful comments.

# ANNEX

## MAGP tutorial

<http://www.lsp.ups-tlse.fr/Fp/Mercadier/MAGP.html>

This text helps for the use of MAGP (Maximum Analysis for Gaussian Processes) which supposes the installation of MATLAB version 5.x/6.x. It is instructive since you can check the program on examples.

This toolbox has been developed for a precise need. It contains estimating procedures associated to the paper "Numerical bounds for the distribution of the maximum of one- and two-parameter processes" (2005).

The functions MATLAB of MAGP are based on the routine `rind.m` of the toolbox WAFO

<http://www.maths.lth.se/matstat/wafo/>

written by Brodtkorb, P.A., Johannesson, P., Lindgren, G., Rychlik, I., Rydén, J., and Sjö, E. Consequently, you need to download WAFO before MAGP.

## 5.7 WAFO

### Introduction

WAFO is a toolbox for MATLAB containing functions for analysis and simulations of waves and random fatigue. It has several sub-directories, associated to particular applications. Indeed, we can compute the theoretical distributions of waves characteristics parameters from observations or from theoretical spectral density of the sea. The other main domain is the statistical analysis of fatigue, there is various modeling functions. Finally, we end this description saying that WAFO contains general statistical tools, for instance linked with the classical extreme theory.

### Downloading

WAFO is free distributed toolbox. Some programs are written with Fortran. Versions for PC/Windows and environments Unix are available. However, for MAGP we used WAFO 2.1.0 with Windows.

1. [http://www.maths.lth.se/matstat/wafo/download/wafo-2.1.0\\_install.txt](http://www.maths.lth.se/matstat/wafo/download/wafo-2.1.0_install.txt)
2. <http://www.maths.lth.se/matstat/wafo/download/wafo-2.1.0.zip>
3. [http://www.maths.lth.se/matstat/wafo/download/wafo-2.1.0\\_pcwin.zip](http://www.maths.lth.se/matstat/wafo/download/wafo-2.1.0_pcwin.zip)

You need to add the path for WAFO under MATLAB. Then

```
>> initwafo
WAFO toolbox paths set: normal initiation
```

Some helps

```
>> wafomenu          >> help wafo
>> help wstats       >> help fatigue
```

## The function rind.m

We present now the function from WAFO that we use in MAGP. It is rind.m, from the sub-directory "trgauss" of WAFO.

```
>> cd wafo/trgauss
>> ls
.          hermitetr.m   private          spec2skew.m     tranproc.m
..         initdata.m   readme.m         spec2tccpdf.m  trmak.m
contents.m iter.m             rind.m          spec2thpdf.m   ttraylpdf.m
createpdf.m iter_mc.m         spec2acat.m     spec2tpdf.m    trunmak.m
dat2gaus.m ochifun.m        spec2acdf.m     spec2tpdf2.m   gaus2dat.m
ochitr.m   spec2cmat.m      spec2vhpdf.m    hermitefun.m   pdfplot.m
spec2mmtpdf.m  trangood.m
```

In the version for PC/windows you need to change "load" in "loaddata". So the last three lines of rind.m must be

```
fxind=loaddata('rind.out');
return
end
```

We write the local help of rind.m.

```
>> help rind
RIND computes  $E[\text{Jacobian} \cdot \text{Indicator} \mid \text{Condition}] \cdot f_{\{X_c\}}(x_c(:,ix))$ 
where "Indicator" =  $I\{H_{lo}(i) < X(i) < H_{up}(i), i=1:N_t+N_d\}$ 
"Jacobian" =  $J(X(N_t+1), \dots, X(N_t+N_d+N_c))$ , special case is
"Jacobian" =  $|X(N_t+1) \dots X(N_t+N_d)| = |X_d(1) \cdot X_d(2) \dots X_d(N_d)|$ 
"condition" =  $X_c = x_c(:,ix), ix=1, \dots, N_x$ .
X = [Xt; Xd; Xc], a stochastic vector of Multivariate Gaussian
variables where Xt, Xd and Xc have the length Nt, Nd and Nc,
respectively. (Recommended limitations  $N_x, N_t \leq 100, N_d \leq 6$  and  $N_c \leq 10$ )
CALL [E, tid] = rind(S,m,xc,[Nt Nj],NIT,speed,indI,B_lo,B_up);
tid = time taken for the fortran executable to integrate
E = expectation/density as explained above size 1 x Nx
S = Covariance matrix of X=[Xt;Xd;Xc] size Ntdc x Ntdc (Ntdc=Nt+Nd+Nc)
m = the expectation of X=[Xt;Xd;Xc] size N x 1
xc = values to condition on size Nc x Nx
Nt = size of Xt
Nj = # of variables of Xt integrated directly (default Nj=0)
NIT = 0,1,2,..., dimension of numerical integration (default NIT=1)
-1 Integrate all by SADAPT for Ndim<9 and by KRBVRC otherwise
-2 Integrate all by SADAPT by Genz (1992) (Fast)
-3 Integrate all by KRBVRC by Genz (1993) (Fast)
-4 Integrate all by KROBOV by Genz (1992) (Fast)
-5 Integrate all by RCRUDE by Genz (1992)
-6 Integrate all by RLHCRUDE using MLHD and center of cell
```

```

-7 Integrate all by RLHCRUDE using LHD and center of cell
-8 Integrate all by RLHCRUDE using MLHD and random point within the cell
-9 Integrate all by RLHCRUDE using LHD and random point within the cell
speed = defines accuracy of calculations by choosing different
        parameters, possible values: 1,2,...,9 (9 fastest, default 4).
indI = vector of indices to the different barriers in the
        indicator function, length NI, where NI = Nb+1 (NB! restriction
        indI(1)=0, indI(NI)=Nt+Nd )
B_lo,B_up = Lower and upper barriers used to compute the integration
            limits, H_lo and H_up, respectively. size Mb x Nb

```

## 5.8 MAGP

MAGP is a toolbox for the estimation of the distribution of the maximum of a Gaussian process  $X$  with one or two parameters. So let  $S$  be the compact set  $[0, T]$  or the compact square  $[0, T]^2$  on which  $X$  is defined. Each function of MAGP gives an estimation of a lower or an upper bound of the survival function

$$\mathbb{P}(\max_{t \in S} X(t) \geq u).$$

```

>> cd wafo/trgauss
>> ls
.          EQA.m      LBF.m      LBPn.m      UBNS.m      UBPN.m
EQT.m     LBNS.m     side.m     UBP.m      UBUS.m      bottom.m
interior.m LBP.m      UBF.m      ubpi.m

```

These functions are well described in the preceding sections, including a summary table in Section 5.1. These functions may have different number of arguments. However, four arguments will correspond to  $(T, u, n, r)$ , three to  $(T, u, r)$  and two to  $(T, u)$ . Some examples of results,

```

>>LBPn(1,2,20,1)= 0.0432          >>LBP(1,2,1)= 0.0430
>>UBPn(1,2,20,1)= 0.0442        >>LBNS(1,2,7)= 0.1199
>>UBNS(1,2,7)= 0.1512           >>LBF(1,0)= 0.7947
>>UBF(1,0)= 0.7976

```

## Bibliography of Chapter 5

- Adler, R. J.** (1981). *The geometry of random fields*. John Wiley & Sons Ltd., Chichester. Wiley Series in Probability and Mathematical Statistics.
- Adler, R. J. and Taylor, J.** (2005). *Random fields and geometry*. <http://iew3.technion.ac.il/radler/grf.pdf>. To appear.
- Azaïs, J.-M., Cierco-Ayrolles, C., and Croquette, A.** (1999). Bounds and asymptotic expansions for the distribution of the maximum of a smooth stationary Gaussian process. *ESAIM Probab. Statist.*, 3 :107–129.
- Azaïs, J.-M. and Delmas, C.** (2002). Asymptotic expansions for the distribution of the maximum of Gaussian random fields. *Extremes*, 5(2) :181–212 (2003).
- Azaïs, J.-M., Gassiat, É., and Mercadier, C.** (2004). Asymptotic distribution and power of the likelihood ratio test for mixtures : bounded and unbounded cases. *Submitted*.
- Azaïs, J.-M. and Wschebor, M.** (2005). On the distribution of the maximum of a gaussian field with  $d$  parameters. *Ann. Appl. Probab.*, 15(1) :254–278.
- Azaïs, J.-M. and Wschebor, M.** (2002). The distribution of the maximum of a Gaussian process : Rice method revisited. In *In and out of equilibrium (Mambucaba, 2000)*, volume 51 of *Progr. Probab.*, pages 321–348. Birkhäuser Boston, Boston, MA.
- Berman, S. M.** (1971). Excursions above high levels for stationary Gaussian processes. *Pacific J. Math.*, 36 :63–79.
- Brodtkorb, P., Johannesson, P., Lindgren, G., Rychlik, I., Rydén, J., and Sjö, E.** (2000). Wafo - a matlab toolbox for analysis of random waves and loads. In *Proceedings of the 10th International Offshore and Polar Engineering conference*, volume Vol III, 343–350, Seattle. <http://www.maths.lth.se/matstat/wafo/>.
- Davies, R. B.** (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64(2) :247–254.
- Delmas, C.** (2003). Projections on spherical cones, maximum of Gaussian fields and Rice's method. *Statist. Probab. Lett.*, 64(3) :263–270.
- DeLong, D. M.** (1981). Crossing probabilities for a square root boundary by a Bessel process. *Comm. Statist. A—Theory Methods*, 10(21) :2197–2213.
- Deloubrière, O. and Hilhorst, H. J.** (2000). Persistence exponents of non-Gaussian processes in statistical mechanics. *J. Phys. A*, 33(10) :1993–2013.
- Dembo, A., Poonen, B., Shao, Q.-M., and Zeitouni, O.** (2002). Random polynomials having few or no real zeros. *J. Amer. Math. Soc.*, 15(4) :857–892.
- Durbin, J.** (1985). The first-passage density of a continuous Gaussian process to a general boundary. *J. Appl. Probab.*, 22(1) :99–122.

- Ehrhardt, G. C., Bray, A. J., et Majumdar, S. N.** (2002). Persistence of a continuous stochastic process with discrete-time sampling : non-Markov processes. *Phys. Rev. E* (3), 65(4) :041102, 13.
- Federer, H.** (1969). *Geometric measure theory*. Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York.
- Gassiat, É.** (2002). Likelihood ratio inequalities with applications to various mixtures. *Ann. Inst. H. Poincaré Probab. Statist.*, 6 :897–906.
- Genz, A.** (1992). Numerical computation of multivariate normal probabilities. *J. Comp. Graph Stat.*, 1 :141–149.
- Ghosh, J. and Sen, P.** (1985). On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983)*, Wadsworth Statist./Probab. Ser., pages 789–806, Belmont, CA. Wadsworth.
- Li, W. V. and Shao, Q.-M.** (2002). A normal comparison inequality and its applications. *Probab. Theory Related Fields*, 122(4) :494–508.
- Li, W. V. and Shao, Q.-M.** (2004). Lower tail probabilities for Gaussian processes. *Ann. Probab.*, 32(1A) :216–242.
- Mercadier, C.** (2005). Numerical bounds for the distribution of one- and two-parameter processes. *Submitted*. Software <http://www.lsp.ups-tlse.fr/Fp/Mercadier/MAGP.html>.
- Molchan, G. and Khokhlov, A.** (2004). Small values of the maximum for the integral of fractional Brownian motion. *J. Statist. Phys.*, 114(3-4) :923–946.
- Piterbarg, V. I.** (1996). *Asymptotic methods in the theory of Gaussian processes and fields*, volume 148 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI. Translated from the Russian by V. V. Piterbarg, Revised by the author.
- Rychlik, I.** (1987). A note on Durbin’s formula for the first-passage density. *Statist. Probab. Lett.*, 5(6) :425–428.
- Slepian, D.** (1962). The one-sided barrier problem for Gaussian noise. *Bell System Tech. J.*, 41 :463–501.
- Sun, J.** (1993). Tail probabilities of the maxima of Gaussian random fields. *Ann. Probab.*, 21(1) :34–71.
- Takemura, A. and Kuriki, S.** (2002). On the equivalence of the tube and Euler characteristic methods for the distribution of the maximum of Gaussian fields over piecewise smooth domains. *Ann. Appl. Probab.*, 12(2) :768–796.



- Taylor, J. E., Adler, R. J., and Takemura, A.** (2004). The Expected Euler Characteristic Heuristic. To appear in *Ann. Probab.*
- Worsley, K. J.** (1995). Estimating the number of peaks in a random field using the Hadwiger characteristic of excursion sets, with applications to medical images. *Ann. Statist.*, 23(2) :640–669.
- Ylvisaker, D.** (1968). A note on the absence of tangencies in Gaussian sample paths. *Ann. Math. Statist.*, 39 :261–262.

---

---

**TROISIÈME PARTIE**

**APPLICATIONS STATISTIQUES**

---

---



# Chapitre 6

---

## DISTRIBUTION ASYMPTOTIQUE ET PUISSANCE DU TEST DE RAPPORT DE VRAISEMBLANCE POUR LES MÉLANGES : CAS BORNÉ ET NON BORNÉ

Ce chapitre présente de façon fidèle l'article soumis à publication "Asymptotic distribution and power of the likelihood ratio test for mixtures : bounded and unbounded cases", Azaïs, Gassiat et Mercadier (2004).

Nous considérons le test du rapport de log-vraisemblance (TRMV) pour tester le nombre de composantes dans un mélange de populations dans une famille paramétrique. Nous donnons la distribution asymptotique de la statistique du TRMV sous l'hypothèse nulle et sous alternatives contiguës quand l'espace des paramètres est borné. De plus, pour le modèle de simple contamination nous montrons que, sous des hypothèses générales, la puissance asymptotique sous alternatives contiguës devient arbitrairement proche du niveau asymptotique quand l'espace des paramètres est suffisamment grand. Dans le problème particulier de distributions normales, nous montrons que, quand la moyenne inconnue n'est pas borné a priori, la puissance asymptotique sous hypothèses contiguës est égale au niveau asymptotique.

### Abstract

In this paper, we consider the log-likelihood ratio test (LRT) for testing the number of components in a mixture of populations in a parametric family. We provide the asymptotic distribution of the LRT statistic under the null hypothesis as well as under contiguous alternatives when the parameter set is bounded. Moreover, for the simple contamination model we prove that, under general assumptions, the asymptotic power under contiguous hypotheses may be arbitrarily close to the asymptotic level when the set of parameters is large enough. In the particular problem of normal distributions, we prove that, when the unknown mean is not a priori bounded, the asymptotic power under contiguous hypotheses is equal to the asymptotic level.

## 6.1 Introduction

Mixtures of populations is a modelling tool widely used in applications and the literature on the subject is vast. For finite mixtures, the first task is the choice of the number of components in the mixture. Some estimation or testing procedures have been proposed for this purpose, see for instance the books of Titterton et al. (1985), Lindsay (1995) and McLachlan and Peel (2000) or the papers of James et al. (2001), Gassiat (2002) and references therein. Asymptotic optimality of the likelihood ratio test (LRT) in several parametric contexts is well known. Using the LRT for testing the number of components in a mixture appears quite natural. In one way, simulation studies show that the LRT performs well in various situations (see Goffinet et al., 1992). In another way, the asymptotic distribution and power of the test have to be evaluated to compare with other known tests.

In this paper, we focus on the asymptotic properties of the LRT for testing that i.i.d. observations  $X_1, \dots, X_n$  come from a mixture of  $p_0$  populations in a parametric set of densities  $\mathcal{F}$  (null hypothesis  $H_0$ ) against a mixture of  $p$  populations (alternative  $H_1$ ), where the integers  $p_0$  and  $p$  satisfy  $p_0 < p$ .

In Section 6.2 we apply results of Gassiat (2002) to obtain the asymptotic distribution of the LRT statistic for testing ( $H_0$ ) against ( $H_1$ ) under the null hypothesis as well as under contiguous alternatives. Indeed, Gassiat (2002) gives a quite weak assumption under which the derivation of the asymptotic distribution of the LRT statistic is made in the general situation when one has to test a small model in a larger one, under the null hypothesis as well as under contiguous hypotheses. This applies to the number of components in a mixture of populations in a parametric set with possibly an unknown nuisance parameter. For instance, we apply the method to multidimensional Gaussian distributions with unknown common variance. By this way, we recover known results for mixtures of one or two populations but under weaker assumptions, or known results concerning particular parametric families such as Gaussian or Binomial distributions; see Ghosh and Sen (1985), Dacunha-Castelle and Gassiat (1997, 1999), Garel (2001), Lemdani and Pons (1997, 1999), Chen and Chen (2001), Mosler and Seidel (2001) Chernoff and Lander (1995). We also obtain more general results than previous ones :

- They apply to general sets of parametric families with unknown nuisance parameter.
- Asymptotic distribution under contiguous alternatives is considered.

However, except for smoothness assumptions, the main point is that these asymptotic results require that the parameter set is bounded.

In Sections 6.3 and 6.4 we study what happens when the set of parameters becomes larger and larger. For simplicity we restrict our attention to the simplest model : the contamination model for family of distributions indexed by a single real parameter. Indeed, roughly speaking, the LRT statistic converges in distribution to half the square of the supremum of some Gaussian process indexed by a compact set of scores. But when this set of scores is enlarged, the covariance of the Gaussian process is close to 0 for sufficiently distant scores, so that the supremum of the Gaussian process may become arbitrarily large. Thus one also knows that for unbounded sets of parameters,

the LRT statistic tends to infinity in probability, as Hartigan first noted for normal mixtures (see Hartigan, 1985). Here, we prove that under some extreme circumstances the LRT can have less power than moment tests or goodness-of-fit tests. At the end of the introduction we draw carefully practical conclusions from this result.

More precisely, let  $\mathbb{T}$  be  $[-T, T]$  and  $\mathcal{F} = \{f_t, t \in \mathbb{T}\}$  be a parametric set of probability densities on  $\mathbb{R}$  with respect to the Lebesgue measure. Using i.i.d. observations  $X_1, \dots, X_n$ , we consider the testing problem for the density  $g$  of the observations.

$$(6.1) \quad (H_0) : "g = f_0" \quad \text{against} \quad (H_1) : "g = (1 - \pi)f_0 + \pi f_t, 0 \leq \pi \leq 1, t \in \mathbb{T}."$$

We prove that :

- For general parametric sets  $\mathcal{F}$ ,  $\mathbb{T} = [-T, T]$  and  $T$  large enough, under contiguous alternatives, the LRT for (6.1) has asymptotic power close to the asymptotic level, under some smoothness assumptions, see Theorem 6.3.1.

A set of assumptions is given for which Theorem 6.3.1 applies in the case of translation mixtures, that is when  $f_t(\cdot) = f_0(\cdot - t)$ , see Corollary 6.3.1. This is done in Section 6.3.

- When  $f_t$  is the standard Gaussian with mean  $t$  we get the normal mixture problem. When the set of means is not a priori bounded, that is  $\mathbb{T} = \mathbb{R}$ , Liu and Shao (2000) obtained the asymptotic distribution of the LRT under the null hypothesis by using the strong approximation proved in Bickel and Chernoff (1993). We prove in Theorem 6.4.1 of Section 6.4 that the asymptotic power under contiguous alternatives is equal to the asymptotic level.

The way to obtain these results is to gather together : expansion of the LRT obtained in Gassiat (2002) to identify contiguity and apply Le Cam's third Lemma (see van der Vaart, 1998), behaviour of the supremum of a Gaussian process on an interval with bounds tending to infinity as obtained in Azaïs and Mercadier (2003), and the normal comparison inequality as refined in Li and Shao (2002). Proof of most results of Sections 6.3 and 6.4 are detailed in Section 6.5.

Independently of our work, for the Gaussian model with unbounded means, Hall and Stewart (2004) obtained the speed of separation of alternatives that ensures asymptotic power to be bigger than asymptotic level. Their result indicates that the speed of separation should be  $\sqrt{\log \circ \log n} / \sqrt{n}$  contrary to the classical parametric situation, where it is  $1/\sqrt{n}$ .

### **Practical application**

Such tests that have power less or equal to level are sometimes called "worthless" (see for example van der Vaart, 1998) but this word would be dangerous because practical interpretation of our result must take into account the following points :

- It is well known that for mixtures of population in general the convergence to the asymptotic distribution is very slow. For example for a very simple test, as the skewness test, Boistard (2003) showed that  $n = 10^3$  observations are needed to meet the asymptotic distribution.

- For maximum likelihood estimates (MLE) and tests, the problem of the speed of convergence to the asymptotic distribution is very difficult to address since in practice MLE are computed through iterative algorithms and are only approximate. The most famous one is the EM algorithm and its variants. All these algorithms depend on tuning constants, in particular concerning the stopping rule. It is shown for example in table 6.3 of McLachlan and Peel (2000) (based on results by Seidel et al., 2000) that the distribution of the LRT depends heavily on these tuning constants. Simulation results by Liu and Shao (2000) suggest that their asymptotic distribution is not met for  $n = 5.10^3$  observations.
- Nowadays some results and softwares are available to compute the distribution of the maximum of Gaussian processes. See for example Garel (2001), Delmas (2003) and Mercadier (2005). In particular these results show that, as soon as the means are contained in some “not huge” set, the asymptotic power under contiguous alternatives of the LRT is generally better than that of moment tests or of goodness-of-fit tests. Nevertheless, the LRT is not uniformly most powerful.

Our result that shows that the LRT is asymptotically less powerful than moment tests is valid in practice only for very large data sets. For all the reasons above it will be very difficult to say precisely when. Simulations have proved that in practice LRT based on Monte-Carlo calculation of threshold or bootstrapping behave well (see Goffinet et al., 1992) for unbounded parameter.

Our opinion is that the main consequence of our result for large or unbounded parameter sets is that the study of the LRT for mixtures in the compact case seems to be the more relevant case.

## 6.2 Asymptotic distribution of the LRT for the number of populations in a mixture under null and contiguous hypotheses

A general theorem in Gassiat (2002) allows to find the asymptotic distribution of the LRT for testing a particular model in a larger one, under the null hypothesis as well as under contiguous alternatives. Roughly speaking, the asymptotic distribution is some function of the supremum of the isonormal process on a set of score functions. The theorem holds under a simple assumption on the bracket entropy of an enlarged set. In many applications, those sets are parameterized by a finite dimensional parameter. In such cases,

- Lipschitz properties allow to compute easily bracket entropies, as in van der Vaart (1998, p. 271). We give some examples in the text.
- The covariance structure of the isonormal process may be computed in an explicit way and identified with the covariance function of a Gaussian field with real parameters.

We shall describe in this section how it applies to mixture models. We first recall the general result of Gassiat (2002) and its application to a very simple contamination mixture model : one has to test between a particular known population with some density  $f_0(\cdot)$  and a mixture of this known one and another with density  $f_t(\cdot)$ ,  $t$  a

multidimensional parameter. Then we detail the case of two populations with possibly unknown nuisance parameter. A typical example will be that of translation mixtures with possibly unknown scale parameter. We end the section by giving general considerations on how to deal with parametric mixture models, allowing an unknown nuisance parameter and setting a general result in such situations.

Assume one would like to use the LRT for testing  $(H_0) : "g \in \mathcal{M}_0"$  against  $(H_1) : "g \in \mathcal{M}"$ , where  $g$  is the generic density of i.i.d. observations  $X_1, \dots, X_n$ ,  $\mathcal{M}_0 \subset \mathcal{M}$  are sets of densities with respect to some measure  $\nu$  on  $\mathbb{R}^k$  (or more generally on some Polish space). Let  $\ell_n(g) = \sum_{i=1}^n \log g(X_i)$  be the log-likelihood, and

$$\lambda_n = \sup_{g \in \mathcal{M}} \ell_n(g) - \sup_{g \in \mathcal{M}_0} \ell_n(g)$$

be the LRT statistic. Let also  $g_0$  be a density in  $\mathcal{M}_0$  that will denote the true (unknown) density of the observations. In the first considered examples, and without loss of generality, we will assume that  $g_0$  coincides with  $f_0$ .

Throughout the paper we use  $\|\cdot\|_2$  to denote the norm in  $L^2(g_0 \cdot \nu)$ .

When studying  $\ell_n(g) - \ell_n(g_0)$ , functions  $\frac{g-g_0}{g_0}$  appear naturally. Define the set  $\mathcal{S}$  as the subset of the unit sphere in  $L^2(g_0 \cdot \nu)$  of such functions when normalized :

$$(6.2) \quad \mathcal{S} = \left\{ \frac{g - g_0}{g_0} / \left\| \frac{g - g_0}{g_0} \right\|_2, g \in \mathcal{M} \setminus \{g_0\} \right\},$$

and  $\mathcal{S}_0$  its subset when  $g \in \mathcal{M}_0$  :

$$(6.3) \quad \mathcal{S}_0 = \left\{ \frac{g - g_0}{g_0} / \left\| \frac{g - g_0}{g_0} \right\|_2, g \in \mathcal{M}_0 \setminus \{g_0\} \right\}.$$

A bracket  $[L, U]$  of length  $\epsilon$  is the set of functions  $b$  such that  $L \leq b \leq U$ , where  $L$  and  $U$  are functions in  $L^2(g_0 \cdot \nu)$  such that  $\|U - L\|_2 \leq \epsilon$ . Define  $H_{\square, 2}(\mathcal{S}, \epsilon)$  the entropy with bracketing of  $\mathcal{S}$  with respect to the norm  $\|\cdot\|_2$ , as the logarithm of the number of brackets of length  $\epsilon$  needed to cover  $\mathcal{S}$ . To apply the theorem in Gassiat (2002), the only needed assumption is :

$$(6.4) \quad \int_0^1 \sqrt{H_{\square, 2}(\mathcal{S}, \epsilon)} d\epsilon < +\infty.$$

This assumption implies in particular that  $\mathcal{S}$  is Donsker and that its closure is compact. As said before, when  $\mathcal{M}$  is parameterized,  $\mathcal{S}$  is also parameterized and smoothness properties will allow to verify (6.4). But in general the parameterization will not be continuous throughout  $\mathcal{S}$ . The delicate point may be that one has to find all possible limit points, in  $L^2(g_0 \cdot \nu)$ , of sequences  $\frac{g_n - g_0}{g_0} / \left\| \frac{g_n - g_0}{g_0} \right\|_2$  when  $\left\| \frac{g_n - g_0}{g_0} \right\|_2$  tends to 0. The set  $\mathcal{D}$  (resp.  $\mathcal{D}_0$ ) of limit points of sequences  $\frac{g_n - g_0}{g_0} / \left\| \frac{g_n - g_0}{g_0} \right\|_2$  where  $\left\| \frac{g_n - g_0}{g_0} \right\|_2$  tends to 0,  $g_n \in \mathcal{M} \setminus \{g_0\}$  (resp.  $g_n \in \mathcal{M}_0 \setminus \{g_0\}$ ) will be parameterized in such a way that Lipschitz properties can be used on subsets.

Let us for example see how it applies to the simple contamination mixture model (6.1). In this case,

$$\mathcal{M}_0 = \{f_0\}, \mathcal{M} = \{g_{\pi, t} = (1 - \pi)f_0 + \pi f_t, 0 \leq \pi \leq 1, t \in [-T, T]\}$$



for a given positive real number  $T$ . Since  $\mathcal{M}_0$  is a singleton, we do not need to define  $\mathcal{S}_0$  and  $\mathcal{D}_0$ . One has  $\frac{g_{\pi_n, t_n} - g_0}{g_0} = \pi \frac{f_t - f_0}{f_0}$ , so that

$$\mathcal{S} = \left\{ d_t = \frac{\frac{f_t - f_0}{f_0}}{\left\| \frac{f_t - f_0}{f_0} \right\|_2}, \quad t \in [-T, 0) \cup (0, T] \right\}.$$

If  $\left\| \frac{f_t - f_0}{f_0} \right\|_2 = 0$ , which occurs if and only if  $t = 0$ , then under smoothness assumptions  $\left\| \frac{g_{\pi_n, t_n} - g_0}{g_0} \right\|_2$  tends to 0 if and only if  $\pi_n$  or  $t_n$  tends to 0. Then  $d_{t_n}$  has two possible limit points (depending on the sign of  $t_n$ ), and

$$\mathcal{D} = \left\{ d_t, t \in [-T, 0) \cup (0, T], d_{0^-} = \frac{-\frac{f'_0}{f_0}}{\left\| \frac{f'_0}{f_0} \right\|_2}, d_{0^+} = \frac{\frac{f'_0}{f_0}}{\left\| \frac{f'_0}{f_0} \right\|_2} \right\}.$$

Here derivatives are taken with respect to parameter  $t$ . Again under smoothness assumptions, it will be possible to prove, considering  $\{d_t, t \in [-T, 0), d_{0^-}\}$  and  $\{d_t, t \in (0, T], d_{0^+}\}$  that the number of brackets of length  $\epsilon$  needed to cover  $\mathcal{S}$  is of order at most  $O(1/\epsilon)$ , so that Assumption (6.4) holds. (A complete proof is given below for contamination models with multidimensional parameterization).

In general when  $\mathcal{M}_0$  contains more than one density,  $\mathcal{D}_0 \subset \mathcal{D}$ , and if the parameterization is smooth enough, it will be possible to define a set  $\mathbb{U}$  in  $\mathbb{R}^{k_0} \times \mathbb{R}^{k_1}$  and a set  $\mathbb{U}_0$  in  $\mathbb{R}^{k_0}$  such that

$$\mathcal{D} = \{d_{\mathbf{u}}, \mathbf{u} \in \mathbb{U}\} \text{ and } \mathcal{D}_0 = \{d_{(\mathbf{v}, \mathbf{0})}, \mathbf{v} \in \mathbb{U}_0\}.$$

Define the covariance function  $r(\cdot, \cdot)$  on  $\mathbb{U} \times \mathbb{U}$  by

$$r(\mathbf{u}_1, \mathbf{u}_2) = \int d_{\mathbf{u}_1} d_{\mathbf{u}_2} g_0 d\nu.$$

Then, under (6.4), applying Theorem 3.1 in Gassiat (2002),

$$2\lambda_n = \sup_{\mathbf{u} \in \mathbb{U}} \left( \max \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n d_{\mathbf{u}}(X_i), 0 \right\} \right)^2 - \sup_{\mathbf{v} \in \mathbb{U}_0} \left( \max \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n d_{(\mathbf{v}, \mathbf{0})}(X_i), 0 \right\} \right)^2 + o_{\mathbb{P}_0}(1),$$

so that  $2\lambda_n$  converges in distribution to

$$(6.5) \quad \sup_{\mathbf{u} \in \mathbb{U}} (\max \{Z(\mathbf{u}), 0\})^2 - \sup_{\mathbf{v} \in \mathbb{U}_0} (\max \{Z(\mathbf{v}, \mathbf{0}), 0\})^2$$

where  $Z(\cdot)$  is the Gaussian process on  $\mathbb{U}$  with covariance  $r(\cdot, \cdot)$  and  $\mathbb{P}_0$  is the joint distribution of the observations  $X_1, \dots, X_n$  under the null hypothesis. In the particular case when  $\mathcal{M}_0$  is reduced to a single element, a direct application of Corollary 3.1 of Gassiat (2002) gives that  $2\lambda_n$  converges in distribution to

$$(6.6) \quad \sup_{\mathbf{u} \in \mathbb{U}} (\max \{Z(\mathbf{u}), 0\})^2.$$

It will be seen in the examples below that  $r(\cdot, \cdot)$  is in general not continuous everywhere on the closure of  $\mathbb{U} \times \mathbb{U}$ .  $Z(\cdot)$  is not a continuous Gaussian field, though the isonormal process on  $\mathcal{D}$  is continuous, so that the suprema involved in (6.5) are a.s. finite. In general,  $r(\cdot, \cdot)$  is continuous almost everywhere. In the simple contamination mixture model (6.1), for non null  $s$  and  $t$ ,

$$(6.7) \quad r(s, t) = \int \left( \frac{\frac{f_t - f_0}{f_0}}{\left\| \frac{f_t - f_0}{f_0} \right\|_2} \right) \left( \frac{\frac{f_s - f_0}{f_0}}{\left\| \frac{f_s - f_0}{f_0} \right\|_2} \right) f_0 d\nu;$$

$r$  is continuous for non zero  $s$  and  $t$  and admits the following limits

$$r(0^+, 0^+) = r(0^-, 0^-) = 1, r(0^+, 0^-) = -1.$$

It is also proved in Gassiat (2002) that if the densities  $g_n$  in  $\mathcal{M} \setminus \mathcal{M}_0$  are such that  $\frac{g_n - g_0}{g_0} / \left\| \frac{g_n - g_0}{g_0} \right\|_2$  converges to some  $d_{\mathbf{u}_0}$  with  $\sqrt{n} \left\| \frac{g_n - g_0}{g_0} \right\|_2$  tending to a positive constant  $c$ , then the distributions  $(g_0 \cdot \nu)^{\otimes n}$  and  $(g_n \cdot \nu)^{\otimes n}$  are mutually contiguous, and  $2 \lambda_n$  converges in distribution under this contiguous alternative to

$$(6.8) \quad \sup_{\mathbf{u} \in \mathbb{U}} (\max \{Z(\mathbf{u}) + c \cdot r(\mathbf{u}, \mathbf{u}_0), 0\})^2 - \sup_{\mathbf{v} \in \mathbb{U}_0} (\max \{Z(\mathbf{v}, \mathbf{0}) + c \cdot r((\mathbf{v}, \mathbf{0}), \mathbf{u}_0), 0\})^2.$$

In general (6.5) and (6.8) reduce to the square of only one supremum, due to the particular structure of the Gaussian process.

We will see, in the subsequent subsections, examples such as : translation mixtures, exponential families, in particular Bernoulli or Gaussian mixtures.

### 6.2.1 Contamination mixture

We consider here the contamination mixture model where parameter  $\mathbf{t}$  may be multidimensional :  $\mathbf{t} \in \mathbb{T}$ ,  $\mathbb{T}$  being a compact subset of  $\mathbb{R}^k$  such that  $\mathbf{0}$  belongs to the interior of  $\mathbb{T}$ . Let  $\|\cdot\|$  and  $\langle \cdot, \cdot \rangle$  denote the Euclidean norm and scalar product in  $\mathbb{R}^k$ . Again,

$$\mathcal{M}_0 = \{f_0\}, \mathcal{M} = \{g_{\pi, \mathbf{t}} = (1 - \pi)f_0 + \pi f_{\mathbf{t}}, 0 \leq \pi \leq 1, \mathbf{t} \in \mathbb{T}\},$$

$$\mathcal{S} = \left\{ d_{\mathbf{t}} = \frac{\frac{f_{\mathbf{t}} - f_0}{f_0}}{\left\| \frac{f_{\mathbf{t}} - f_0}{f_0} \right\|_2}, \mathbf{t} \in \mathbb{T} \right\}.$$

We shall use the following Assumptions (CM), insuring smoothness and some non degeneracy :

(CM)

- $f_{\mathbf{t}} = f_0 \nu$  a.e. if and only if  $\mathbf{t} = \mathbf{0}$ .
- $\mathbf{t} \rightarrow f_{\mathbf{t}}$  is twice continuously differentiable  $\nu$  a.e. at any  $\mathbf{t} \in \mathbb{T}$ .
- $\exists \eta > 0, \forall \mathbf{t} \in \mathbb{T}$  such that  $\|\mathbf{t}\| \leq \eta, \sum_{i=1}^k \tau_i \frac{\partial f_{\mathbf{t}}}{\partial t_i} = 0 \nu$  a.e. if and only if  $\tau = \mathbf{0}$ .
- There exists a positive real  $\eta$  and a function  $B \in L^2(f_0 \cdot \nu)$  that upper bounds all following functions :

$$\frac{f_{\mathbf{t}}}{f_0}, \frac{1}{f_0} \left| \frac{\partial f_{\mathbf{t}}}{\partial t_i} \right|, i = 1, \dots, k, \mathbf{t} \in \mathbb{T},$$

$$\frac{1}{f_0} \left| \frac{\partial^2 f_{\mathbf{t}}}{\partial t_i \partial t_j} \right|, i, j = 1, \dots, k, \mathbf{t} \in \mathbb{T}, \|\mathbf{t}\| \leq \eta.$$

Notice that in this assumption the real number  $\eta$  is fixed. We shall prove that the condition (6.4) holds true for  $\mathcal{S}$  by splitting it into two sets

$$\mathcal{S}_1 = \{d_{\mathbf{t}}, \mathbf{t} \in \mathbb{T}, \|\mathbf{t}\| \geq \eta\} \text{ and } \mathcal{S}_2 = \{d_{\mathbf{t}}, \mathbf{t} \in \mathbb{T}, \|\mathbf{t}\| < \eta\}.$$

Since  $\|(g_{\pi, \mathbf{t}} - f_0)/f_0\|_2 = \pi \|(f_{\mathbf{t}} - f_0)/f_0\|_2$  tends to 0 as soon as  $\pi$  or  $\mathbf{t}$  tends to 0, it is easy to see that a limit point exists only if either  $\mathbf{t}$  converges to a limit different of  $\mathbf{0}$  or  $\frac{\mathbf{t}}{\|\mathbf{t}\|}$  converges to some  $\tau$ . One obtains easily

$$\mathcal{D} = \left\{ d_{\mathbf{t}} = \frac{\frac{f_{\mathbf{t}} - f_0}{f_0}}{\left\| \frac{f_{\mathbf{t}} - f_0}{f_0} \right\|_2}, \mathbf{t} \in \mathbb{T} \right\} \cup \left\{ \bar{d}_{\tau} = \frac{1}{f_0} \sum_{i=1}^k \tau_i \frac{\partial f_0}{\partial t_i} / \left\| \frac{1}{f_0} \sum_{i=1}^k \tau_i \frac{\partial f_0}{\partial t_i} \right\|_2, \|\tau\| = 1 \right\}.$$

Set  $h_{\mathbf{t}} = \frac{f_{\mathbf{t}} - f_0}{f_0}$ . Then, for  $i = 1, \dots, k$ , if (CM) holds,

$$\frac{\partial d_{\mathbf{t}}}{\partial t_i} = \frac{\frac{\partial h_{\mathbf{t}}}{\partial t_i}}{\|h_{\mathbf{t}}\|_2} - \int \left( \frac{\frac{\partial h_{\mathbf{t}}}{\partial t_i}}{\|h_{\mathbf{t}}\|_2} \right) \left( \frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2} \right) f_0 d\nu \cdot \frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2}.$$

This proves that, there exists a constant  $C$  such that, for all  $\mathbf{t}$  and  $\mathbf{s}$  such that  $\|\mathbf{t}\| \geq \eta$  and  $\|\mathbf{s}\| \geq \eta$ ,  $|d_{\mathbf{t}} - d_{\mathbf{s}}| \leq C \cdot B \cdot \|\mathbf{t} - \mathbf{s}\|$ , so that the number of brackets of length  $\epsilon$  needed to cover  $\mathcal{S}_1$  is of order at most  $O(1/\epsilon^k)$  and that Condition (6.4) holds true for the set  $\mathcal{S}_1$ .

Now, for any  $\tau \in \mathbb{T}$  such that  $\|\tau\| = 1$ , one has letting  $\mathbf{t} = \lambda\tau$ ,  $\lambda \in \mathbb{R}$ ,

$$\frac{\partial}{\partial \lambda} (d_{\lambda\tau}) = \frac{\sum_{i=1}^k \tau_i \frac{\partial h_{\mathbf{t}}}{\partial t_i}}{\|h_{\mathbf{t}}\|_2} - \int \left( \frac{\sum_{i=1}^k \tau_i \frac{\partial h_{\mathbf{t}}}{\partial t_i}}{\|h_{\mathbf{t}}\|_2} \right) \left( \frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2} \right) f_0 d\nu \cdot \frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2}.$$

But using Taylor expansions, there exists  $\hat{\lambda}$ ,  $\lambda^*$  and  $\tilde{\lambda}$  in  $[0, \lambda]$ , such that

$$\begin{aligned} h_{\mathbf{t}} &= \lambda \sum_{i=1}^k \tau_i \frac{\partial h_{\hat{\lambda}\tau}}{\partial t_i} = \lambda \sum_{i=1}^k \tau_i \frac{\partial h_0}{\partial t_i} + \frac{\lambda^2}{2} \sum_{i,j=1}^k \tau_i \tau_j \frac{\partial^2 h_{\lambda^*\tau}}{\partial t_i \partial t_j}, \\ \sum_{i=1}^k \tau_i \frac{\partial h_{\mathbf{t}}}{\partial t_i} &= \sum_{i=1}^k \tau_i \frac{\partial h_0}{\partial t_i} + \lambda \sum_{i,j=1}^k \tau_i \tau_j \frac{\partial^2 h_{\tilde{\lambda}\tau}}{\partial t_i \partial t_j}. \end{aligned}$$

All this leads to

$$\frac{\partial}{\partial \lambda} (d_{\lambda\tau}) = \frac{H_{\mathbf{t}}}{\left\| \sum_{i=1}^d \tau_i \frac{\partial h_{\hat{\lambda}\tau}}{\partial t_i} \right\|_2} - \left[ \int \frac{H_{\mathbf{t}}}{\left\| \sum_{i=1}^d \tau_i \frac{\partial h_{\hat{\lambda}\tau}}{\partial t_i} \right\|_2} \left( \frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2} \right) f_0 d\nu \right] \cdot \frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2}.$$

with

$$H_{\mathbf{t}} = \frac{1}{\lambda} \sum_{i=1}^k \tau_i \frac{\partial h_{\mathbf{t}}}{\partial t_i} - \frac{1}{\lambda^2} h_{\mathbf{t}} = \sum_{i,j=1}^k \tau_i \tau_j \frac{\partial^2 h_{\tilde{\lambda}\tau}}{\partial t_i \partial t_j} - \frac{1}{2} \sum_{i,j=1}^k \tau_i \tau_j \frac{\partial^2 h_{\lambda^*\tau}}{\partial t_i \partial t_j}.$$

But using (CM), this implies that for some constant  $C$ ,  $\forall \lambda \in (0, \eta]$ ,

$$\left| \frac{\partial}{\partial \lambda} (d_{\lambda\tau}) \right| \leq C \cdot B,$$

and that

$$\begin{aligned} \lim_{\lambda \rightarrow 0} \frac{\partial}{\partial \lambda} (d_{\lambda\tau}) &= \partial \bar{d}_\tau \\ &= \frac{1}{2} \frac{\sum_{i,j=1}^k \tau_i \tau_j \frac{\partial^2 h_{\mathbf{0}}}{\partial t_i \partial t_j}}{\left\| \sum_{i=1}^k \tau_i \frac{\partial h_{\mathbf{0}}}{\partial t_i} \right\|_2} - \frac{1}{2} \frac{\sum_{i=1}^k \tau_i \frac{\partial h_{\mathbf{0}}}{\partial t_i}}{\left\| \sum_{i=1}^k \tau_i \frac{\partial h_{\mathbf{0}}}{\partial t_i} \right\|_2} \int \left( \frac{\sum_{i,j=1}^k \tau_i \tau_j \frac{\partial^2 h_{\mathbf{0}}}{\partial t_i \partial t_j}}{\left\| \sum_{i=1}^k \tau_i \frac{\partial h_{\mathbf{0}}}{\partial t_i} \right\|_2} \right) \left( \frac{\sum_{i=1}^k \tau_i \frac{\partial h_{\mathbf{0}}}{\partial t_i}}{\left\| \sum_{i=1}^k \tau_i \frac{\partial h_{\mathbf{0}}}{\partial t_i} \right\|_2} \right) f_{\mathbf{0}} d\nu, \end{aligned}$$

so that  $\lambda \rightarrow d_{\lambda\tau}$  is continuously differentiable on  $[0, \eta]$ . Using the fact that

$$\begin{aligned} |d_{\lambda\tau} - d_{\lambda'\tau'}| &\leq |d_{\lambda\tau} - \bar{d}_\tau| + |\bar{d}_\tau - \bar{d}_{\tau'}| + |d_{\lambda'\tau'} - \bar{d}_{\tau'}| \\ &\leq C \cdot B \cdot (\lambda + \lambda') + |\bar{d}_\tau - \bar{d}_{\tau'}| \end{aligned}$$

and that, using (CM), there exists a positive constant  $\tilde{C}$  such that

$$\inf_{\|\tau\|=1} \left\| \sum_{i=1}^k \tau_i \frac{\partial h_{\mathbf{0}}}{\partial t_i} \right\|_2 \geq \tilde{C},$$

we obtain that for some constant  $C^*$ , and any  $\tau, \tau'$  such that  $\|\tau\| = 1, \|\tau'\| = 1$ ,

$$|\bar{d}_\tau - \bar{d}_{\tau'}| \leq C^* \cdot B \cdot \|\tau - \tau'\|.$$

It is straightforward to see that the number of brackets of length  $\epsilon$  needed to cover  $\mathbb{S}_2$  is of order at most  $O(1/\epsilon^k)$ . Thus Assumptions (CM) imply Condition (6.4).

Define now for all non null  $\mathbf{s}$  and  $\mathbf{t}$  in  $\mathbb{T}$ ,

$$(6.9) \quad r(\mathbf{s}, \mathbf{t}) = \int d_{\mathbf{s}} d_{\mathbf{t}} f_{\mathbf{0}} d\nu,$$

and let  $Z(\cdot)$  be the Gaussian field on  $\mathbb{T} \setminus \{\mathbf{0}\}$  with covariance  $r$ . Notice that, on each direction  $\tau$  such that  $\|\mathbf{t}\| \rightarrow 0$  with  $\mathbf{t}/\|\mathbf{t}\| \rightarrow \tau$ , one may extend  $r(\cdot, \cdot)$  by continuity, setting

$$(6.10) \quad \bar{r}(\tau, \mathbf{t}) = \bar{r}(\mathbf{t}, \tau) = \int \bar{d}_\tau d_{\mathbf{t}} f_{\mathbf{0}} d\nu \quad ; \quad \tilde{r}(\tau, \tau') = \int \bar{d}_\tau \bar{d}_{\tau'} f_{\mathbf{0}} d\nu.$$

Let  $\pi_n$  and  $\mathbf{t}_n$  be sequences such that

- $\lim_{n \rightarrow +\infty} \sqrt{n} \pi_n \|(f_{\mathbf{t}_n} - f_{\mathbf{0}})/f_{\mathbf{0}}\|_2 = c$  for some positive  $c$ ,
- either  $\mathbf{t}_n$  tends to some  $\mathbf{t}_0 \neq \mathbf{0}$  and  $\sqrt{n} \pi_n$  tends to some positive constant, or  $\mathbf{t}_n$  tends to  $\mathbf{0}$ , and  $\mathbf{t}_n/\|\mathbf{t}_n\|$  converges to some limit  $\tau$ .

Then :

**Theorem 6.2.1**

Assume (CM). Then  $(f_0 \cdot \nu)^{\otimes n}$  and  $[(1 - \pi_n)f_0 + \pi_n f_{\mathbf{t}_n}] \cdot \nu^{\otimes n}$  are mutually contiguous,  $2\lambda_n$  converges under  $(f_0 \cdot \nu)^{\otimes n}$  in distribution to

$$\sup_{\mathbf{t} \in \mathbb{T}} (\max\{Z(\mathbf{t}), 0\})^2 = \left( \sup_{\mathbf{t} \in \mathbb{T}} Z(\mathbf{t}) \right)^2,$$

and under  $[(1 - \pi_n)f_0 + \pi_n f_{\mathbf{t}_n}] \cdot \nu^{\otimes n}$  to

$$\sup_{\mathbf{t} \in \mathbb{T}} (\max\{Z(\mathbf{t}) + \mu(\mathbf{t}), 0\})^2 = \left( \sup_{\mathbf{t} \in \mathbb{T}} (Z(\mathbf{t}) + \mu(\mathbf{t})) \right)^2$$

with

(6.11)

$\mu(\mathbf{t}) = c \cdot r(\mathbf{t}, \mathbf{t}_0)$  if  $\mathbf{t}_n \rightarrow \mathbf{t}_0 \neq 0$ , and  $\mu(\mathbf{t}) = c \cdot \bar{r}(\mathbf{t}, \tau)$  if  $\|\mathbf{t}_n\| \rightarrow 0$  and  $\mathbf{t}_n/\|\mathbf{t}_n\| \rightarrow \tau$ .

**Remark :** Set  $m \equiv 0$  under  $(f_0 \cdot \nu)^{\otimes n}$  and  $m \equiv \mu$  under  $[(1 - \pi_n)f_0 + \pi_n f_{\mathbf{t}_n}] \cdot \nu^{\otimes n}$ . Letting  $\mathbf{t}$  go to  $\mathbf{0}$  radially in two opposite directions and using covariance properties in the neighbourhood of  $\mathbf{0}$  we see that almost surely  $\sup_{\mathbf{t} \in \mathbb{T}} (Z(\mathbf{t}) + m(\mathbf{t})) > 0$  what justifies equalities in preceding theorem.

Let us give applications of this theorem to particular models :

**Translation mixtures**

We consider the translation mixture model, where  $\nu$  is the Lebesgue measure and

$$f_{\mathbf{t}}(\cdot) = f_0(\cdot - \mathbf{t}).$$

Then, it is easy to see that Theorem 6.2.1 applies as soon as the following Assumptions (CTM) hold :

(CTM)

- $f_0$  is positive on  $\mathbb{R}^k$ ,
- $x \rightarrow f_0(x)$  is twice continuously differentiable  $\nu$  a.e.
- There exists a function  $B \in L^2(f_0 \cdot \nu)$  that upper bounds all following functions :

$$\frac{f_0(x - \mathbf{t})}{f_0(x)}, \frac{1}{f_0(x)} \left| \frac{\partial f_0}{\partial x_i}(x - \mathbf{t}) \right|, i = 1, \dots, k, \mathbf{t} \in \mathbb{T},$$

$$\frac{1}{f_0(x)} \left| \frac{\partial^2 f_0}{\partial x_i \partial x_j}(x - \mathbf{t}) \right|, i, j = 1, \dots, k, \mathbf{t} \in \mathbb{T}, \|\mathbf{t}\| \leq \eta.$$

Indeed, since  $\frac{\partial f_{\mathbf{t}}}{\partial t_i}(x) = -\frac{\partial f_0}{\partial x_i}(x - \mathbf{t})$ , if  $\tau$  is such that  $\sum_{i=1}^k \tau_i \frac{\partial f_{\mathbf{t}}}{\partial t_i} = 0$   $\nu$  a.e. for all  $\|\mathbf{t}\| \leq \eta$ , then  $\sum_{i=1}^k \tau_i \frac{\partial f_0}{\partial x_i} = 0$   $\nu$  a.e., so that  $f_0(x + \lambda\tau) = f_0(x)$  for all  $\lambda \in \mathbb{R}$ , which is impossible unless  $\tau = \mathbf{0}$ .

Here are some examples of situations in which these assumptions are met :  $f_0$  being the inverse of a polynomial with degree at least 2, among which the Cauchy density,

the Gaussian densities and the normalization of  $\text{ch}(x)^{-1}$ .

The covariance function  $r$  is given for non null  $\mathbf{s}$  and  $\mathbf{t}$  by

$$r(\mathbf{s}, \mathbf{t}) = \frac{\int \frac{f_{\mathbf{0}}(x - \mathbf{s})f_{\mathbf{0}}(x - \mathbf{t})}{f_{\mathbf{0}}(x)} d\nu(x) - 1}{\sqrt{\int \frac{f_{\mathbf{0}}(x - \mathbf{s})^2}{f_{\mathbf{0}}(x)} d\nu(x) - 1} \sqrt{\int \frac{f_{\mathbf{0}}(x - \mathbf{t})^2}{f_{\mathbf{0}}(x)} d\nu(x) - 1}},$$

and if the dimension  $k = 1$ , one may define  $r(0^+, 0^-) = -1$ , and for non null  $t$

$$r(0^+, t) = -r(0^-, t) = \frac{\int \frac{-f'_{\mathbf{0}}(x)f_{\mathbf{0}}(x - t)}{f_{\mathbf{0}}(x)} d\nu(x)}{\sqrt{\int \frac{f_{\mathbf{0}}'^2(x)}{f_{\mathbf{0}}(x)} d\nu(x)} \sqrt{\int \frac{f_{\mathbf{0}}(x - t)^2}{f_{\mathbf{0}}(x)} d\nu(x) - 1}},$$

where the derivation is with respect to  $x$ .

### Gaussian mixtures

Without loss of generality we may assume that  $f_{\mathbf{0}}$  is standard normal. Let  $K$  be a bound for  $\|\mathbf{t}\|$ ,  $\mathbf{t} \in \mathbb{T}$ . Then the following bounds show that the function  $B$  exists for any  $\eta$ . For every  $i$  and every  $i \neq j$ ,

$$\begin{aligned} \frac{f_{\mathbf{0}}(x - \mathbf{t})}{f_{\mathbf{0}}(x)} &= \exp(\langle x, \mathbf{t} \rangle - \|\mathbf{t}\|^2/2) \leq \exp(K\|x\|), \\ \frac{1}{f_{\mathbf{0}}(x)} \left| \frac{\partial f_{\mathbf{0}}}{\partial x_i}(x - \mathbf{t}) \right| &= |x_i - t_i| \frac{f_{\mathbf{0}}(x - \mathbf{t})}{f_{\mathbf{0}}(x)} \leq (\|x\| + K) \exp(K\|x\|), \\ \frac{1}{f_{\mathbf{0}}(x)} \left| \frac{\partial^2 f_{\mathbf{0}}}{\partial x_i \partial x_j}(x - \mathbf{t}) \right| &= |x_i - t_i| |x_j - t_j| \frac{f_{\mathbf{0}}(x - \mathbf{t})}{f_{\mathbf{0}}(x)} \leq (\|x\| + K)^2 \exp(K\|x\|), \\ \frac{1}{f_{\mathbf{0}}(x)} \left| \frac{\partial^2 f_{\mathbf{0}}}{\partial x_i^2}(x - \mathbf{t}) \right| &= |(x_i - t_i)^2 - 1| \frac{f_{\mathbf{0}}(x - \mathbf{t})}{f_{\mathbf{0}}(x)} \leq [1 + (\|x\| + K)^2] \exp(K\|x\|). \end{aligned}$$

So (CTM) holds, and Theorem 6.2.1 applies, as soon as  $f_{\mathbf{0}}$  is some Gaussian density on  $\mathbb{R}^k$  and  $\mathbb{T}$  is compact. The covariance of the process  $Z$  is :

$$r(\mathbf{s}, \mathbf{t}) = \frac{\exp(\langle \mathbf{t}, \mathbf{s} \rangle) - 1}{\sqrt{\exp(\|\mathbf{t}\|^2) - 1} \sqrt{\exp(\|\mathbf{s}\|^2) - 1}}.$$

### Binomial mixtures

Here  $\nu$  is the measure with density  $\frac{k!}{x!(k-x)!}$  with respect to the counting measure on the set  $\{0, 1, \dots, k\}$ . We consider the binomial family  $Bi(k, \theta)$  with density  $\theta^x(1 - \theta)^{k-x}$  ;  $x = 0, 1, \dots, k$ . Let  $\theta_0 \in (0, 1)$  and  $f_t$  be the density of  $Bi(k, \theta_0 + t)$ . The most relevant case for genetic applications is the case  $\theta_0 = 1/2$ , see Problem 1 in Chernoff and Lander (1995). We have

$$f_t(x) = (t + \theta_0)^x (1 - t - \theta_0)^{k-x},$$

$$\frac{\partial f_t}{\partial t}(x) = \left( \frac{x}{t + \theta_0} - \frac{k - x}{1 - t - \theta_0} \right) f_t(x),$$

$$\frac{\partial^2 f_t}{\partial t^2}(x) = \left[ \left( \frac{x}{t + \theta_0} - \frac{k - x}{1 - t - \theta_0} \right)^2 - \frac{x}{(t + \theta_0)^2} + \frac{k - x}{(1 - t - \theta_0)^2} \right] f_t(x).$$

It is clear that  $f_t(x)$  and  $\frac{\partial f_t}{\partial t}(x)$  are uniformly upper bounded and that  $\frac{\partial^2 f_t}{\partial t^2}(x)$  is upper bounded for  $t$  small enough, proving Assumptions (CM). Direct calculations lead to

$$\frac{f_t - f_0}{f_0}(x) = \left(1 + \frac{t}{\theta_0}\right)^x \left(1 - \frac{t}{1 - \theta_0}\right)^{k-x} - 1,$$

$$r(s, t) = \frac{\Gamma(s, t)}{\sqrt{\Gamma(s, s)}\sqrt{\Gamma(t, t)}},$$

with

$$\Gamma(s, t) = \sum_{x=0}^k \left[ \left(1 + \frac{s}{\theta_0}\right)^x \left(1 - \frac{s}{1 - \theta_0}\right)^{k-x} - 1 \right] \left[ \left(1 + \frac{t}{\theta_0}\right)^x \left(1 - \frac{t}{1 - \theta_0}\right)^{k-x} - 1 \right] \theta_0^x (1 - \theta_0)^{k-x}$$

which is equivalent to the result of Chernoff and Lander (1995).

### Mixtures in exponential families

This case generalizes the preceding. Let  $f_{\mathbf{t}}$  be a regular exponential family with exhaustive statistic  $T(x) = (T_1(x), \dots, T_k(x))$  :

$$f_{\mathbf{t}}(x) = f_0(x) \exp \left( \sum_{i=1}^k t_i T_i(x) - \phi(\mathbf{t}) \right),$$

and assume  $\mathbb{T}$  is a compact subset in the interior of the definition set of the exponential family. Then  $\mathbf{t} \rightarrow f_{\mathbf{t}}$  is infinitely differentiable on  $\mathbb{T}$ . Let  $F(x) = \sup_{\mathbf{t} \in \mathbb{T}} \exp \left( \sum_{i=1}^k t_i T_i(x) \right)$ , Assumption (CEM) will be :

(CEM)

- There exists  $B$  in  $L^2(f_0 \cdot \nu)$  that upper bounds all following functions :  $F$ ,  $|T_i|F$ ,  $|T_i T_j|F$ ,  $i, j = 1, \dots, k$ .

One can see easily that (CEM) implies (CM), so that Theorem 6.2.1 applies to exponential families as soon as (CEM) holds. Direct calculations again lead to

$$\frac{f_{\mathbf{t}} - f_0}{f_0}(x) = \exp \left( \sum_{i=1}^k t_i T_i(x) - \phi(\mathbf{t}) \right) - 1,$$

$$r(\mathbf{s}, \mathbf{t}) = \frac{\exp(\phi(\mathbf{s} + \mathbf{t}) - \phi(\mathbf{s}) - \phi(\mathbf{t})) - 1}{\sqrt{\exp(\phi(2\mathbf{s}) - 2\phi(\mathbf{s})) - 1} \sqrt{\exp(\phi(2\mathbf{t}) - 2\phi(\mathbf{t})) - 1}}.$$

### 6.2.2 Two populations against a single one

We consider here the case where one wants to test a single population in the family of densities  $f_{\mathbf{t}}, \mathbf{t} \in \mathbb{T}$ ,  $\mathbb{T}$  compact subset of  $\mathbb{R}^k$  against a mixture of two such populations. That is :

$$\mathcal{M}_0 = \{f_{\mathbf{t}}, \mathbf{t} \in \mathbb{T}\},$$

and

$$\mathcal{M} = \{g_{\pi, \mathbf{t}_1, \mathbf{t}_2} = (1 - \pi)f_{\mathbf{t}_1} + \pi f_{\mathbf{t}_2}, 0 \leq \pi \leq 1, \mathbf{t}_1 \in \mathbb{T}, \mathbf{t}_2 \in \mathbb{T}\}.$$

We suppose moreover that  $\mathbf{0}$  is an interior point of  $\mathbb{T}$  and that  $f_{\mathbf{0}}$  is the unknown distribution of the observations (with no loss of generality). We shall use Assumptions (TP), insuring smoothness and some non degeneracy :

(TP)

- $(1-\pi)f_{\mathbf{t}_1} + \pi f_{\mathbf{t}_2} = f_{\mathbf{0}} \nu$  a.e. if and only if  $(\pi = 0$  and  $\mathbf{t}_1 = \mathbf{0})$  or  $(\pi = 1$  and  $\mathbf{t}_2 = \mathbf{0})$  or  $(\mathbf{t}_1 = \mathbf{0}$  and  $\mathbf{t}_2 = \mathbf{0})$ ,
- $\mathbf{t} \rightarrow f_{\mathbf{t}}$  is three times continuously differentiable  $\nu$  a.e. at any  $\mathbf{t} \in \mathbb{T}$ ,
- $\forall \tau \in \mathbb{R}^k, \forall \mathbf{t} \in \mathbb{T}, \forall \mathbf{s} \in \mathbb{T}, \forall \rho \geq 0, \rho(f_{\mathbf{s}} - f_{\mathbf{0}}) + \sum_{i=1}^k \tau_i \frac{\partial f_{\mathbf{t}}}{\partial t_i} = 0 \nu$  a.e. if and only if  $\rho \mathbf{s} = \mathbf{0}$  and  $\tau = \mathbf{0}$ ,  
and  $\exists \eta > 0$ , such that  $\forall \tau \in \mathbb{R}^k, \forall \mathbf{t} \in \mathbb{T}$  with  $\|\mathbf{t}\| \leq \eta \sum_{i,j=1}^k \tau_i \tau_j \frac{\partial^2 f_{\mathbf{t}}}{\partial t_i \partial t_j} = 0 \nu$  a.e. if and only if  $\tau = \mathbf{0}$ ,
- there exists a function  $B \in L^2(f_{\mathbf{0}} \cdot \nu)$  that upper bounds all following functions :

$$\frac{f_{\mathbf{t}}}{f_{\mathbf{0}}}, \frac{1}{f_{\mathbf{0}}} \left| \frac{\partial f_{\mathbf{t}}}{\partial t_i} \right|, \frac{1}{f_{\mathbf{0}}} \left| \frac{\partial^2 f_{\mathbf{t}}}{\partial t_i \partial t_j} \right|, i, j = 1, \dots, k, \mathbf{t} \in \mathbb{T},$$

$$\frac{1}{f_{\mathbf{0}}} \left| \frac{\partial^3 f_{\mathbf{t}}}{\partial t_i \partial t_j \partial t_l} \right|, i, j, l = 1, \dots, k, \mathbf{t} \in \mathbb{T}, \|\mathbf{t}\| \leq \eta.$$

Then  $\mathcal{S} \subset \mathcal{D}, \mathcal{S}_0 \subset \mathcal{D}_0$ , and  $\mathcal{D}$  can be parameterized as follows :

$$\mathcal{D} = \left\{ d_{\mathbf{t}, a, \tau} = \frac{a \frac{f_{\mathbf{t}} - f_{\mathbf{0}}}{f_{\mathbf{0}}} + \sum_{i=1}^k \tau_i \frac{1}{f_{\mathbf{0}}} \frac{\partial f_{\mathbf{0}}}{\partial t_i}}{\|a \frac{f_{\mathbf{t}} - f_{\mathbf{0}}}{f_{\mathbf{0}}} + \sum_{i=1}^k \tau_i \frac{1}{f_{\mathbf{0}}} \frac{\partial f_{\mathbf{0}}}{\partial t_i}\|_2}, \mathbf{t} \in \mathbb{T} \setminus \{\mathbf{0}\}, \tau \in \mathbb{R}^k, a \geq 0, a + \|\tau\| = 1 \right\},$$

$$\mathcal{D}_0 = \{d_{\mathbf{0}, 0, \tau}, \|\tau\| = 1\}.$$

Let  $r(\cdot, \cdot)$  be as in Section 6.2.1 :

$$r(\mathbf{s}, \mathbf{t}) = \int \left( \frac{h_{\mathbf{s}}}{\|h_{\mathbf{s}}\|_2} \right) \left( \frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2} \right) f_{\mathbf{0}} d\nu$$

with  $h_{\mathbf{t}} = (f_{\mathbf{t}} - f_{\mathbf{0}})/f_{\mathbf{0}}$ , and  $Z(\cdot)$  the associated Gaussian field.

Let  $W$  be the  $k$ -dimensional centered Gaussian variable with variance  $\Sigma$  with entries :

$$\Sigma_{i,j} = \int \left( \frac{\frac{1}{f_{\mathbf{0}}} \frac{\partial f_{\mathbf{0}}}{\partial t_i}}{\|\frac{1}{f_{\mathbf{0}}} \frac{\partial f_{\mathbf{0}}}{\partial t_i}\|_2} \right) \left( \frac{\frac{1}{f_{\mathbf{0}}} \frac{\partial f_{\mathbf{0}}}{\partial t_j}}{\|\frac{1}{f_{\mathbf{0}}} \frac{\partial f_{\mathbf{0}}}{\partial t_j}\|_2} \right) f_{\mathbf{0}} d\nu, i, j = 1, \dots, k,$$



and for any  $\mathbf{t}$ , let  $C(\mathbf{t})$  be the  $k$ -dimensional vector of covariances of  $Z(\mathbf{t})$  and  $W$  :

$$C(\mathbf{t})_i = \int \left( \frac{\frac{1}{f_0} \frac{\partial f_0}{\partial t_i}}{\left\| \frac{1}{f_0} \frac{\partial f_0}{\partial \mathbf{t}} \right\|_2} \right) \left( \frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2} \right) f_0 d\nu, \quad i = 1, \dots, k.$$

Then  $\mathcal{D}$  can be reparametrized as follows

$$\mathcal{D} = \{d_{\mathbf{t},a,\tau}; \mathbf{t} \in \mathbb{T} \setminus \{\mathbf{0}\}, \tau \in \mathbb{R}^k, a \geq 0, a^2 + \tau^T \Sigma \tau + 2a\tau^T C(\mathbf{t}) = 1\}.$$

Using the same tricks as for proving Theorem 6.2.1,  $2\lambda_n$  converges under  $(f_0 \cdot \nu)^{\otimes n}$  in distribution to

$$(6.12) \quad \left( \sup_{\substack{a \geq 0, \mathbf{t} \in \mathbb{T}, \tau \in \mathbb{R}^k \\ a^2 + \tau^T \Sigma \tau + 2a\tau^T C(\mathbf{t}) = 1}} (aZ(\mathbf{t}) + \langle \tau, W \rangle) \right)^2 - \left( \sup_{\tau^T \Sigma \tau = 1} \langle \tau, W \rangle \right)^2.$$

Remark that :

$$(6.13) \quad \left( \sup_{\tau^T \Sigma \tau = 1} \langle \tau, W \rangle \right)^2 = W^T \Sigma^{-1} W,$$

and that the supremum is attained for  $\tau$  colinear to  $\Sigma^{-1}W$ . Then consider the matrix :

$$\tilde{\Sigma} = \begin{bmatrix} 1 & C(\mathbf{t})^T \\ C(\mathbf{t}) & \Sigma \end{bmatrix}$$

with inverse

$$(\tilde{\Sigma})^{-1} = \begin{bmatrix} \alpha & u^T \\ u & M \end{bmatrix}$$

where  $M = M(\mathbf{t}) = (\Sigma - C(\mathbf{t})C(\mathbf{t})^T)^{-1}$ ,  $u = u(\mathbf{t}) = -M(\mathbf{t})C(\mathbf{t})$ ,  $\alpha = \alpha(\mathbf{t}) = 1 + C(\mathbf{t})^T M(\mathbf{t})C(\mathbf{t})$ . Now consider the maximization problem in  $a$  and  $\tau$  :

$$(6.14) \quad \left( \sup_{a \geq 0, a^2 + \tau^T \Sigma \tau + 2a\tau^T C(\mathbf{t}) = 1} (aZ(\mathbf{t}) + \langle \tau, W \rangle) \right)^2.$$

If the maximum is attained for  $a > 0$ , then by (6.13) its value is

$$\begin{pmatrix} Z(\mathbf{t}) \\ W \end{pmatrix}^T (\tilde{\Sigma})^{-1} \begin{pmatrix} Z(\mathbf{t}) \\ W \end{pmatrix},$$

which is equal to

$$\alpha \left( Z(\mathbf{t}) + \frac{u^T W}{\alpha} \right)^2 + W^T \left( M - \frac{uu^T}{\alpha} \right) W = \alpha \left( Z(\mathbf{t}) + \frac{u^T W}{\alpha} \right)^2 + W^T \Sigma^{-1} W.$$

This is the case when the first coordinate of

$$(\tilde{\Sigma})^{-1} \begin{pmatrix} Z(\mathbf{t}) \\ W \end{pmatrix}$$

is non-negative that is  $\alpha Z(\mathbf{t}) + \langle u, W \rangle \geq 0$ . In the other case ( $a = 0$ ) the maximum is  $W^T \Sigma^{-1} W$  by (6.13). Finally we have proved that the supremum in (6.14) is equal to

$$(6.15) \quad \left\{ \max \left\{ Z(\mathbf{t}) - \frac{C(\mathbf{t})^T M(\mathbf{t}) W}{1 + C(\mathbf{t})^T M(\mathbf{t}) C(\mathbf{t})}, 0 \right\} \right\}^2 (1 + C(\mathbf{t})^T M(\mathbf{t}) C(\mathbf{t})) + W^T \Sigma^{-1} W.$$

This implies that the limit of  $2 \lambda_n$  under  $(f_0 \cdot \nu)^{\otimes n}$  is equal in distribution to

$$\sup_{\mathbf{t} \in \mathbb{T}} \left\{ \left( Z(\mathbf{t}) - \frac{C(\mathbf{t})^T M(\mathbf{t}) W}{1 + C(\mathbf{t})^T M(\mathbf{t}) C(\mathbf{t})} \right)^2 (1 + C(\mathbf{t})^T M(\mathbf{t}) C(\mathbf{t})) \right\}.$$

Indeed one may see, letting  $\mathbf{t}$  go to  $\mathbf{0}$  radially in two opposite directions, that the supremum of the Gaussian process involved in formula (6.15) is non negative. Let now  $\pi_n$ ,  $\mathbf{t}_1^n$  and  $\mathbf{t}_2^n$  be sequences such that  $\frac{(1-\pi_n)f_{\mathbf{t}_1^n} + \pi_n f_{\mathbf{t}_2^n} - f_0}{f_0} / \left\| \frac{(1-\pi_n)f_{\mathbf{t}_1^n} + \pi_n f_{\mathbf{t}_2^n} - f_0}{f_0} \right\|_2$  tends to some  $d_{\mathbf{t}_0, a_0, \tau_0}$  in  $\mathcal{D}$ , with  $\lim_{n \rightarrow +\infty} \sqrt{n} \left\| \frac{(1-\pi_n)f_{\mathbf{t}_1^n} + \pi_n f_{\mathbf{t}_2^n} - f_0}{f_0} \right\|_2 = c$  for some positive constant  $c$ . Then, using the same tricks again :

**Theorem 6.2.2**

Assume (TP). Then  $(f_0 \cdot \nu)^{\otimes n}$  and  $[((1 - \pi_n)f_{\mathbf{t}_1^n} + \pi_n f_{\mathbf{t}_2^n}) \cdot \nu]^{\otimes n}$  are mutually contiguous,  $2 \lambda_n$  converges under  $(f_0 \cdot \nu)^{\otimes n}$  in distribution to

$$\sup_{\mathbf{t} \in \mathbb{T}} \left\{ \left( Z(\mathbf{t}) - \frac{C(\mathbf{t})^T M(\mathbf{t}) W}{1 + C(\mathbf{t})^T M(\mathbf{t}) C(\mathbf{t})} \right)^2 (1 + C(\mathbf{t})^T M(\mathbf{t}) C(\mathbf{t})) \right\},$$

and under  $[((1 - \pi_n)f_{\mathbf{t}_1^n} + \pi_n f_{\mathbf{t}_2^n}) \cdot \nu]^{\otimes n}$  to

$$\sup_{\mathbf{t} \in \mathbb{T}} \left\{ \left( aZ(\mathbf{t}) + a_0 c r(\mathbf{t}, \mathbf{t}_0) + c C(\mathbf{t})^T \tau_0 - \frac{C(\mathbf{t})^T M(\mathbf{t}) (W + c \Sigma \tau_0 + c a_0 C(\mathbf{t}_0))}{1 + C(\mathbf{t})^T M(\mathbf{t}) C(\mathbf{t})} \right)^2 \times (1 + C(\mathbf{t})^T M(\mathbf{t}) C(\mathbf{t})) \right\},$$

where if  $\mathbf{t}_0 = 0$  then  $a_0 = 0$ .

Notice that, when  $\mathbf{t}_0 = 0$ ,  $d_{0, a_0, \tau_0} = d_{0, 0, \tau_0}$ , and  $\langle d_{0, 0, \tau_0}, d_{\mathbf{t}, a, \tau} \rangle = c C(\mathbf{t})^T \tau_0 + c \Sigma \tau_0$ . This is why one has to take  $a_0 = 0$  when  $\mathbf{t}_0 = 0$  in the last formula of Theorem 6.2.2.

**Examples.**

Results of Section 6.2.2 apply to the same previous examples.

- **Translation mixtures**, under (CTM) with moreover  $x \rightarrow f_0(x)$  is three times continuously differentiable  $\nu$  a.e., and the function  $B \in L^2(f_0 \cdot \nu)$  is also an upper bound for

$$\frac{1}{f_0(x)} \left| \frac{\partial^3 f_0}{\partial x_i \partial x_j \partial x_l} (x - \mathbf{t}) \right|, i, j, l = 1, \dots, k, \mathbf{t} \in \mathbb{T}, \|\mathbf{t}\| \leq \eta.$$

- **Gaussian mixtures**, in this case  $W$  is a standard normal vector and for all  $\mathbf{t} \in \mathbb{T}$   $C(\mathbf{t}) = \frac{\mathbf{t}}{\sqrt{e^{\|\mathbf{t}\|^2} - 1}}$ .

- **Bernoulli mixtures**,
- **Mixtures in exponential families**, under (CEM) with moreover : the function  $B \in L^2(f_0 \cdot \nu)$  is also an upper bound for  $|T_i T_j T_l| F$ ,  $i, j, l = 1, \dots, k$ . In this case,  $W$  is the Gaussian vector with covariance  $\Sigma$  the correlation matrix of the vector  $(T_1(X), \dots, T_k(X))$ , when  $X$  has density  $f_0$ . Recall that the variance matrix of the vector  $(T_1(X), \dots, T_k(X))$  when  $X$  has density  $f_0$  is the matrix  $D^2\phi$  of second derivatives of the function  $\phi$  at point  $\mathbf{0}$ , and the vector  $C(\mathbf{t})$  is given by

$$C(\mathbf{t})_i = \frac{\frac{\partial\phi}{\partial t_i}(\mathbf{t}) - \frac{\partial\phi}{\partial t_i}(\mathbf{0})}{\sqrt{\exp(\phi(2\mathbf{t}) - 2\phi(\mathbf{t})) - 1} \sqrt{(D^2\phi(\mathbf{0}))_{i,i}}}, \quad i = 1, \dots, k.$$

### 6.2.3 Contamination with unknown nuisance parameter

We consider here the contamination mixture model with some unknown parameter, which is the same for all populations. A typical example may be that of mixtures of Gaussian distributions with the same unknown variance, or translation mixtures with the same unknown scale parameter. We shall assume that the nuisance parameter is identifiable, so that its maximum likelihood estimator is consistent. This will allow to reduce the possible nuisance parameters in the definition of the set  $\mathcal{S}$  to be in a neighbourhood of the true unknown one (recall that  $\mathcal{S}$  is only a theoretical tool to verify that some theorem apply, and compute the set of normalized scores, so that this does not restrict the model  $\mathcal{M}$ , for which the nuisance parameter is not restricted to be in a neighbourhood of the true one).

Let  $\mathcal{F} = \{f_{\mathbf{t},\alpha}, \mathbf{t} \in \mathbb{T}, \alpha \in \mathbb{A}\}$  be a set of densities with respect to some dominating measure  $\nu$ , where  $\mathbb{T}$  is a compact subset of  $\mathbb{R}^k$  and  $\mathbb{A}$  is a compact subset of  $\mathbb{R}^h$ . We consider here the case where

$$\mathcal{M}_0 = \{f_{\mathbf{0},\alpha}, \alpha \in \mathbb{A}\},$$

and

$$\mathcal{M} = \{g_{\pi,\mathbf{t},\alpha} = (1 - \pi)f_{\mathbf{0},\alpha} + \pi f_{\mathbf{t},\alpha}, 0 \leq \pi \leq 1, \mathbf{t} \in \mathbb{T}, \alpha \in \mathbb{A}\}.$$

The unknown true distribution of the observations will be  $f_{\mathbf{0},\alpha_0}$ . We suppose that  $(\mathbf{0}, \alpha_0)$  is an interior point of  $\mathbb{T} \times \mathbb{A}$ . We shall use Assumptions (CMN), insuring smoothness and some non degeneracy :

(CMN)

- $(1 - \pi)f_{\mathbf{0},\alpha} + \pi f_{\mathbf{t},\alpha} = f_{\mathbf{0},\alpha_0}$   $\nu$  a.e. if and only if  $\alpha = \alpha_0$  and  $[\pi = 0 \text{ or } \mathbf{t} = \mathbf{0}]$ ,
- $(\mathbf{t}, \alpha) \rightarrow f_{\mathbf{t},\alpha}$  is twice continuously differentiable  $\nu$  a.e. at any  $(\mathbf{t}, \alpha) \in \mathbb{T} \times \mathbb{A}$ ,
- $\exists \eta > 0$ , such that  $\forall \delta \in \mathbb{R}^h, \forall \mathbf{t} \in \mathbb{T}, \forall \alpha \in \mathbb{A}$  with  $\|\alpha - \alpha_0\| \leq \eta, \forall \rho \geq 0$  :  
 $\rho(f_{\mathbf{t},\alpha_0} - f_{\mathbf{0},\alpha_0}) + \sum_{i=1}^h \delta_i \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_i} = 0$   $\nu$  a.e. if and only if  $\rho \mathbf{t} = \mathbf{0}$  and  $\delta = \mathbf{0}$ ,  
and  $\forall \tau \in \mathbb{R}^k, \|\mathbf{t}\| \leq \eta, \|\alpha - \alpha_0\| \leq \eta : \sum_{i=1}^k \tau_i \frac{\partial f_{\mathbf{t},\alpha_0}}{\partial t_i} + \sum_{i=1}^h \delta_i \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_i} = 0$   $\nu$  a.e. if and only if  $\tau = \mathbf{0}$  and  $\delta = \mathbf{0}$ .
- There exists a function  $B \in L^2(f_{\mathbf{0},\alpha_0} \cdot \nu)$  that upper bounds all following functions :

$$\frac{f_{\mathbf{t},\alpha}}{f_{\mathbf{0},\alpha_0}}, \frac{1}{f_{\mathbf{0},\alpha_0}} \left| \frac{\partial f_{\mathbf{t},\alpha}}{\partial t_i} \right|, i = 1, \dots, k, \frac{1}{f_{\mathbf{0},\alpha_0}} \left| \frac{\partial f_{\mathbf{t},\alpha}}{\partial \alpha_i} \right|, i = 1, \dots, h,$$

$$\frac{1}{f_{\mathbf{0},\alpha_0}} \left| \frac{\partial^2 f_{\mathbf{t},\alpha}}{\partial t_i \partial t_j} \right|, i, j = 1, \dots, k, \quad \frac{1}{f_{\mathbf{0},\alpha_0}} \left| \frac{\partial^2 f_{\mathbf{t},\alpha}}{\partial t_i \partial \alpha_j} \right|, i = 1, \dots, k, j = 1, \dots, h,$$

$$\frac{1}{f_{\mathbf{0},\alpha_0}} \left| \frac{\partial^2 f_{\mathbf{t},\alpha}}{\partial \alpha_i \partial \alpha_j} \right|, i, j = 1, \dots, h, \quad (\mathbf{t}, \alpha) \in \mathbb{T} \times \mathbb{A}, \quad \|\alpha - \alpha_0\| \leq \eta, \quad \|\mathbf{t}\| \leq \eta.$$

Then, since the maximum likelihood estimator of parameter  $\alpha$  is consistent, one only needs to verify Assumption (6.4) for

$$\mathcal{S} = \left\{ \frac{(1-\pi)f_{\mathbf{0},\alpha} + \pi f_{\mathbf{t},\alpha} - f_{\mathbf{0},\alpha_0}}{f_{\mathbf{0},\alpha_0}} \middle/ \left\| \frac{(1-\pi)f_{\mathbf{0},\alpha} + \pi f_{\mathbf{t},\alpha} - f_{\mathbf{0},\alpha_0}}{f_{\mathbf{0},\alpha_0}} \right\|_2, \right. \\ \left. 0 \leq \pi \leq 1, \mathbf{t} \in \mathbb{T}, \alpha \in \mathbb{A}, \|\alpha - \alpha_0\| \leq \eta \right\},$$

where we restrict our definition to  $\pi, \mathbf{t}$  and  $\alpha$  such that  $(1-\pi)f_{\mathbf{0},\alpha} + \pi f_{\mathbf{t},\alpha}$  differs from  $f_{\mathbf{0},\alpha_0}$ . One has also

$$\mathcal{S}_0 = \left\{ \frac{f_{\mathbf{0},\alpha} - f_{\mathbf{0},\alpha_0}}{f_{\mathbf{0},\alpha_0}} \middle/ \left\| \frac{f_{\mathbf{0},\alpha} - f_{\mathbf{0},\alpha_0}}{f_{\mathbf{0},\alpha_0}} \right\|_2, 0 \leq \pi \leq 1, \alpha \in \mathbb{A}, \|\alpha - \alpha_0\| \leq \eta \right\}.$$

Define, for  $(\mathbf{t}, \rho, \delta, \tau) \in \mathbb{T} \times \mathbb{R}_+ \times \mathbb{R}^h \times \mathbb{R}^k$ ,

$$H_{\mathbf{t},\rho,\delta,\tau} = \rho(f_{\mathbf{t},\alpha_0} - f_{\mathbf{0},\alpha_0}) + \sum_{i=1}^h \delta_i \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_i} + \sum_{i=1}^k \tau_i \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial t_i},$$

and

$$d_{\mathbf{t},\rho,\delta,\tau} = \frac{H_{\mathbf{t},\rho,\delta,\tau}/f_{\mathbf{0},\alpha_0}}{\|H_{\mathbf{t},\rho,\delta,\tau}/f_{\mathbf{0},\alpha_0}\|_2}.$$

The sets  $\mathcal{D}$  and  $\mathcal{D}_0$  can be parameterized as follows :

$$\mathcal{D} = \left\{ d_{\mathbf{t},\rho,\delta,\tau}, \mathbf{t} \in \mathbb{T}, \rho \geq 0, \delta \in \mathbb{R}^h, \tau \in \mathbb{R}^k, \rho^2 + \|\delta\|^2 + \|\tau\|^2 = 1 \right\},$$

$$\mathcal{D}_0 = \left\{ d_{\mathbf{0},0,\delta,\mathbf{0}}, \delta \in \mathbb{R}^h, \|\delta\| = 1 \right\}.$$

Note that due to the existence of the nuisance parameter which is fixed to  $\alpha_0$ , now  $\mathcal{D}$  does not contain  $\mathcal{S}$ .

It will be possible to obtain the asymptotic distributions in the same way as in Section 6.2.2. Let again

$$r(\mathbf{s}, \mathbf{t}) = \int \left( \frac{h_{\mathbf{s}}}{\|h_{\mathbf{s}}\|_2} \right) \left( \frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2} \right) f_{\mathbf{0},\alpha_0} d\nu$$

with  $h_{\mathbf{t}} = (f_{\mathbf{t},\alpha_0} - f_{\mathbf{0},\alpha_0})/f_{\mathbf{0},\alpha_0}$ , and  $Z(\cdot)$  the associated Gaussian field.

Note that this process is the same as the one of Section 6.2.1 if we set  $f_{\mathbf{0}} = f_{\mathbf{0},\alpha_0}$ . Let also  $W, \Sigma$  and  $C(\mathbf{t})$  be the same as in Section 6.2.1 replacing  $\frac{\partial f_{\mathbf{0}}}{\partial t_i}$  by  $\frac{\partial f_{\mathbf{0},\alpha_0}}{\partial t_i}$ .

Let  $V$  be the  $h$ -dimensional centered Gaussian variable with variance  $\Gamma$  :

$$\Gamma_{i,j} = \int \left( \frac{\frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_i}}{\left\| \frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_i} \right\|_2} \right) \left( \frac{\frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_j}}{\left\| \frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_j} \right\|_2} \right) f_{\mathbf{0},\alpha_0} d\nu, \quad i, j = 1, \dots, h,$$

and for any  $\mathbf{t}$ , let  $G(\mathbf{t})$  be the  $h$ -dimensional vector of covariances of  $Z(\mathbf{t})$  and  $V$  :

$$G(\mathbf{t})_i = \int \left( \frac{\frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_i}}{\left\| \frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_i} \right\|_2} \right) \left( \frac{h_{\mathbf{t}}}{\|h_{\mathbf{t}}\|_2} \right) f_{\mathbf{0},\alpha_0} d\nu, \quad i = 1, \dots, h.$$

Let also  $S$  be the covariance matrix of  $W$  and  $V$ , with entries :

$$S_{i,j} = \int \left( \frac{\frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_i}}{\left\| \frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial \alpha_i} \right\|_2} \right) \left( \frac{\frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial t_j}}{\left\| \frac{1}{f_{\mathbf{0},\alpha_0}} \frac{\partial f_{\mathbf{0},\alpha_0}}{\partial t_j} \right\|_2} \right) f_{\mathbf{0},\alpha_0} d\nu, \quad i = 1, \dots, h, \quad j = 1, \dots, k.$$

Define the matrices  $U(\mathbf{t})$  and  $N(\mathbf{t})$  by

$$U(\mathbf{t}) = \begin{pmatrix} C(\mathbf{t})^T \\ G(\mathbf{t}) \end{pmatrix}, \quad N(\mathbf{t}) = \left( \begin{pmatrix} \Sigma & S^T \\ S & \Gamma \end{pmatrix} - U(\mathbf{t})U(\mathbf{t})^T \right)^{-1}.$$

Let  $\pi_n$ ,  $\mathbf{t}_n$  and  $\alpha_n$  be sequences such that

$$\frac{(1 - \pi_n)f_{\mathbf{0},\alpha_n} + \pi_n f_{\mathbf{t}_n,\alpha_n} - f_{\mathbf{0},\alpha_0}}{f_{\mathbf{0},\alpha_0}} / \left\| \frac{(1 - \pi_n)f_{\mathbf{0},\alpha_n} + \pi_n f_{\mathbf{t}_n,\alpha_n} - f_{\mathbf{0},\alpha_0}}{f_{\mathbf{0},\alpha_0}} \right\|_2$$

tends to some  $d_{\mathbf{t}_0,\rho_0,\delta_0,\tau_0}$  in  $\mathcal{D}$ , with  $\lim_{n \rightarrow +\infty} \sqrt{n} \left\| \frac{(1 - \pi_n)f_{\mathbf{0},\alpha_n} + \pi_n f_{\mathbf{t}_n,\alpha_n} - f_{\mathbf{0},\alpha_0}}{f_{\mathbf{0},\alpha_0}} \right\|_2 = c$  for some positive constant  $c$ . Then, using the same tricks as for proving Theorem 6.2.2 :

### Theorem 6.2.3

Assume (CMN). Then  $(f_{\mathbf{0},\alpha_0} \cdot \nu)^{\otimes n}$  and  $[((1 - \pi_n)f_{\mathbf{0},\alpha_n} + \pi_n f_{\mathbf{t}_n,\alpha_n}) \cdot \nu]^{\otimes n}$  are mutually contiguous,  $2\lambda_n$  converges under  $(f_{\mathbf{0}} \cdot \nu)^{\otimes n}$  in distribution to

$$\sup_{\mathbf{t} \in \mathbb{T}} \left\{ \left( Z(\mathbf{t}) - \frac{U(\mathbf{t})^T N(\mathbf{t})}{1 + U(\mathbf{t})^T N(\mathbf{t}) U(\mathbf{t})} \begin{pmatrix} W \\ V \end{pmatrix} \right)^2 (1 + U(\mathbf{t})^T N(\mathbf{t}) U(\mathbf{t})) \right\} \\ + \begin{pmatrix} W \\ V \end{pmatrix}^T \begin{pmatrix} \Sigma & S^T \\ S & \Gamma \end{pmatrix}^{-1} \begin{pmatrix} W \\ V \end{pmatrix} - V^T \Gamma^{-1} V,$$

and under  $[((1 - \pi_n)f_{\mathbf{0},\alpha_n} + \pi_n f_{\mathbf{t}_n,\alpha_n}) \cdot \nu]^{\otimes n}$  to

$$\sup_{\mathbf{t} \in \mathbb{T}} \left\{ \left( Z(\mathbf{t}) + c\rho_0 r(\mathbf{t}, \mathbf{t}_0) + cC(\mathbf{t})^T \tau_0 + cG(\mathbf{t})^T \delta_0 \right. \right. \\ \left. \left. - \frac{U(\mathbf{t})^T N(\mathbf{t})}{1 + U(\mathbf{t})^T N(\mathbf{t}) U(\mathbf{t})} \begin{pmatrix} W + c\Sigma\tau_0 + c\rho_0 C(\mathbf{t}_0) \\ V + c\Gamma\delta_0 + c\rho_0 G(\mathbf{t}_0) \end{pmatrix} \right)^2 (1 + U(\mathbf{t})^T N(\mathbf{t}) U(\mathbf{t})) \right\} \\ + \begin{pmatrix} W + c\Sigma\tau_0 + c\rho_0 C(\mathbf{t}_0) \\ V + c\Gamma\delta_0 + c\rho_0 G(\mathbf{t}_0) \end{pmatrix}^T \begin{pmatrix} \Sigma & S^T \\ S & \Gamma \end{pmatrix}^{-1} \begin{pmatrix} W + c\Sigma\tau_0 + c\rho_0 C(\mathbf{t}_0) \\ V + c\Gamma\delta_0 + c\rho_0 G(\mathbf{t}_0) \end{pmatrix} \\ - (V + c\Gamma\delta_0 + c\rho_0 G(\mathbf{t}_0))^T \Gamma^{-1} (V + c\Gamma\delta_0 + c\rho_0 G(\mathbf{t}_0)),$$

where  $\rho_0 = 0$  when  $\mathbf{t}_0 = \mathbf{0}$ .

### Translation mixtures with unknown scale parameter

Here  $h = 1$ ,  $\nu$  is the Lebesgue measure and

$$f_{\mathbf{t},\alpha}(\cdot) = \alpha f_{\mathbf{0},1}(\alpha(\cdot - \mathbf{t})),$$

with  $\mathbb{A} = [a, A]$  for some  $a > 0$ . Then, it is easy to see that Theorem 6.2.3 applies as soon as the following Assumptions (CTMN) hold :

(CTMN)

- $f_{\mathbf{0},1}$  is positive on  $\mathbb{R}^k$ ,
- $x \rightarrow f_{\mathbf{0},1}(x)$  is twice continuously differentiable  $\nu$  a.e.,
- There exists a function  $B \in L^2(f_{\mathbf{0},1} \cdot \nu)$  that upper bounds all following functions :

$$\frac{f_{\mathbf{0},1}(x - \mathbf{t})}{f_{\mathbf{0},1}(x)}, \frac{1}{f_{\mathbf{0},1}(x)} \left| (1 + |x_i|) \frac{\partial f_{\mathbf{0},1}}{\partial x_i}(x - \mathbf{t}) \right|, i = 1, \dots, k, \mathbf{t} \in \mathbb{T},$$

$$\frac{1}{f_{\mathbf{0},1}(x)} \left| (1 + |x_i||x_j|) \frac{\partial^2 f_{\mathbf{0},1}}{\partial x_i \partial x_j}(x - \mathbf{t}) \right|, i, j = 1, \dots, k, \mathbf{t} \in \mathbb{T}, \|\mathbf{t}\| \leq \eta.$$

These assumptions are met when  $f_{\mathbf{0},1}$  is the inverse of a polynomial with degree at least 2, among which the Cauchy density, or the Gaussian densities and the normalization of  $\text{ch}(x)^{-1}$ .

### Gaussian mixtures with unknown variance

Here  $h = k(k+1)/2$  since  $\alpha$  is the unknown variance. It is easy to see that Assumptions (CTMN) hold, and Theorem 6.2.3 applies, as soon as the  $f_{\mathbf{t},\alpha}$  are the Gaussian distributions  $\mathcal{N}(\mathbf{t}, \alpha)$  on  $\mathbb{R}^k$ ,  $\mathbb{T}$  is compact,  $\mathbb{A}$  is a compact subset of symmetric matrices that are positive definite.

#### 6.2.4 General mixtures with unknown nuisance parameter

Let  $\mathcal{F} = \{f_{\mathbf{t},\alpha}, \mathbf{t} \in \mathbb{T}, \alpha \in \mathbb{A}\}$  be a set of densities with respect to some dominating measure  $\nu$ , where  $\mathbb{T}$  is a compact subset of  $\mathbb{R}^k$  and  $\mathbb{A}$  is a compact subset of  $\mathbb{R}^h$ . The parameter  $\mathbf{t}$  will characterize the population in the mixture, the parameter  $\alpha$  will be the same for all populations. As a simple example one may think to Gaussian distributions (possibly multidimensional) with  $\mathbf{t}$  the mean vector and  $\alpha$  the variance matrix. One may define a mixture with  $p$  populations as

$$(6.16) \quad g_{p,\pi,\mathbf{T},\alpha} = \sum_{i=1}^p \pi_i f_{\mathbf{t}^i,\alpha}.$$

Here  $\pi = (\pi_1, \dots, \pi_p)$  is a vector of non negative real numbers that sum to one,  $\mathbf{T} = (\mathbf{t}^1, \dots, \mathbf{t}^p) \in \mathbb{T}^p$  and  $\alpha \in \mathbb{A}$ . One would like to use the LRT for testing  $(H_0) : "g$

is a mixture of  $p_0$  populations" against  $(H_1)$  : "g is a mixture of  $p$  populations", where  $g$  is the density of i.i.d. observations  $X_1, \dots, X_n$ , and  $p_0 < p$ . This is the case when

$$\mathcal{M}_0 = \left\{ g_{p_0, \pi_0, \mathbf{T}_0, \alpha}, \pi_0 \in [0, 1]^{p_0}, \mathbf{T}_0 \in \mathbb{T}^{p_0}, \alpha \in \mathbb{A}, \sum_{i=1}^{p_0} \pi_i = 1 \right\},$$

$$\mathcal{M} = \left\{ g_{p, \pi, \mathbf{T}, \alpha}, \pi \in [0, 1]^p, \mathbf{T} \in \mathbb{T}^p, \alpha \in \mathbb{A}, \sum_{i=1}^p \pi_i = 1 \right\}.$$

To understand what happens and how to do computations, the main point is to understand how two mixtures with possibly different number of populations may become close.

The main weak identifiability Assumption (WID) will be that  $g_{p, \pi, \mathbf{T}, \alpha} = g_{q, \pi', \mathbf{T}', \alpha'}$  if and only if  $\alpha = \alpha'$  and  $\sum_{i=1}^p \pi_i \delta_{\mathbf{t}^i} = \sum_{i=1}^q \pi'_i \delta_{\mathbf{t}'^i}$  where  $\delta_z$  is the Dirac measure at  $z$ .

Then, if the parameterization  $(\mathbf{t}, \alpha) \rightarrow f_{\mathbf{t}, \alpha}$  is smooth enough, two mixtures become close if their parameter  $\alpha$  becomes close, and their mixing measure becomes close in the weak topology.

Let now  $g_0 = g_{p_0, \pi_0, \mathbf{T}_0, \alpha_0}$  be a particular mixture in  $\mathcal{M}_0$  which has exactly  $p_0$  populations and not fewer, that will denote the true unknown density of the observations. We denote by  $\mathbf{t}^{0,i}$  the elements of  $\mathbf{T}_0$ . Since parameter  $\alpha$  is identifiable, its maximum likelihood estimator is consistent under weak smoothness assumptions, so that to define the sets  $\mathcal{S}$  and  $\mathcal{S}_0$  by (6.2) and (6.3), one may restrict  $\alpha$  by  $\|\alpha - \alpha_0\| \leq \eta$  for some small  $\eta$ . Then, as seen in the previous subsections, the main point is to find  $\mathcal{D}$  and  $\mathcal{D}_0$ , so as to be able to :

- understand how parameterization and smoothness may be used to compute the order of the bracketing entropy,
- define the Gaussian process that is used in the limiting distribution.

For these points, smoothness assumptions and bounding with a square integrable function have to be used together with some non degeneracy of functions that come in the norm appearing in denominator, when this one goes to zero. In fact, if it may be degenerate, it means that one has to go further in the order of the Taylor development until non degeneracy. This, of course, depends on particular examples.

A rather general situation is the following. Let  $q = p - p_0$ . Denote by  $D_{\mathbf{t}} f_{\mathbf{t}, \alpha}$  the  $k$ -dimensional vector of derivatives of  $f_{\mathbf{t}, \alpha}$  with respect to  $\mathbf{t}$ ,  $D_{\alpha} f_{\mathbf{t}, \alpha}$  the  $h$ -dimensional vector of derivatives of  $f_{\mathbf{t}, \alpha}$  with respect to  $\alpha$ ,  $D_{\mathbf{t}}^2 f_{\mathbf{t}, \alpha}$  the  $k \times k$ -dimensional matrix of second derivatives of  $f_{\mathbf{t}, \alpha}$  with respect to  $\mathbf{t}$ . Introduce Assumptions (GM) :

(GM)

- $(\mathbf{t}, \alpha) \rightarrow f_{\mathbf{t}, \alpha}$  is three times continuously differentiable  $\nu$  a.e. at any  $(\mathbf{t}, \alpha) \in \mathbb{T} \times \mathbb{A}$ ,
- $\exists \eta > 0$  such that, for all  $\alpha^i \in \mathbb{R}^h$ ,  $\tilde{\mathbf{t}}^i \in \mathbb{R}^k$ ,  $\tau^i \in \mathbb{R}^k$ ,  $\delta^i \in \mathbb{R}^h$ ,  $\pi_i \in \mathbb{R}$ ,  $i = 1, \dots, p_0$ , for all  $\rho_1, \dots, \rho_q \geq 0$  such that  $\|\alpha^i - \alpha_0\| \leq \eta$ ,  $\|\tilde{\mathbf{t}}^i - \mathbf{t}^{0,i}\| \leq \eta$ ,  $\sum_i \rho_i + \sum_j \pi_j = 0$  then :  $\sum_{i=1}^q \rho_i f_{\mathbf{t}^i, \alpha_0} + \sum_{i=1}^{p_0} \pi_i f_{\tilde{\mathbf{t}}^i, \alpha_0} + \sum_{i=1}^{p_0} \langle \delta^i, D_{\alpha} f_{\tilde{\mathbf{t}}^i, \alpha_0} \rangle + \sum_{i=1}^{p_0} \langle \tau^i, D_{\mathbf{t}} f_{\tilde{\mathbf{t}}^i, \alpha_0} \rangle = 0$   $\nu$  a.e. if and only if  $\sum_{i=1}^q \rho_i f_{\mathbf{t}^i, \alpha_0} + \sum_{i=1}^{p_0} \pi_i f_{\mathbf{t}^{0,i}, \alpha_0} = 0$ ,  $\delta^1 = \mathbf{0}, \dots, \delta^{p_0} = \mathbf{0}$  and  $\tau^1 = \mathbf{0}, \dots, \tau^{p_0} = \mathbf{0}$ ,

- For any subset  $J$  of at most  $\inf\{p_0, q\}$  points in  $\mathbb{T}$  such that for each one there is one of the  $\mathbf{t}^{0,i}$ s at distance at most  $\eta$ , for any  $(\tau^j)_{j \in J}$  of vectors of  $\mathbb{R}^k$ , for any  $\delta^1, \dots, \delta^{p_0}$  in  $\mathbb{R}^h$  :  
 $\sum_{i=1}^{p_0} \langle \delta^i, D_\alpha f_{\mathbf{t}^{0,i}, \alpha_i} \rangle + \sum_{j \in J} (\tau^j)^T D_{\mathbf{t}}^2 f_{j, \alpha_0}(\tau^j) = 0$   $\nu$  a.e. if and only if  
 $\delta^1 = \mathbf{0}, \dots, \delta^{p_0} = \mathbf{0}$  and  $\tau^j = \mathbf{0}, j \in J$ ;
- There exists a function  $B \in L^2(g_0 \cdot \nu)$  that upper bounds all following functions :

$$\begin{aligned} & \frac{f_{\mathbf{t}, \alpha}}{g_0}, \frac{1}{g_0} \left| \frac{\partial f_{\mathbf{t}, \alpha}}{\partial t_i} \right|, i = 1, \dots, k, \frac{1}{g_0} \left| \frac{\partial f_{\mathbf{t}, \alpha}}{\partial \alpha_i} \right|, i = 1, \dots, h, \\ & \frac{1}{g_0} \left| \frac{\partial^2 f_{\mathbf{t}, \alpha}}{\partial t_i \partial t_j} \right|, i, j = 1, \dots, k, (\mathbf{t}, \alpha) \in \mathbb{T} \times \mathbb{A}, \|\alpha - \alpha_0\| \leq \eta \\ & \frac{1}{g_0} \left| \frac{\partial^2 f_{\mathbf{t}, \alpha}}{\partial t_i \partial \alpha_j} \right|, i = 1, \dots, k, j = 1, \dots, h, \frac{1}{g_0} \left| \frac{\partial^2 f_{\mathbf{t}, \alpha}}{\partial \alpha_i \partial \alpha_j} \right|, i, j = 1, \dots, h, \\ & \frac{1}{g_0} \left| \frac{\partial^3 f_{\mathbf{t}, \alpha}}{\partial t_i \partial t_j \partial t_l} \right|, i, j, l = 1, \dots, k, \frac{1}{g_0} \left| \frac{\partial^3 f_{\mathbf{t}, \alpha}}{\partial t_i \partial t_j \partial \alpha_l} \right|, i, j = 1, \dots, k, l = 1, \dots, h \\ & (\mathbf{t}, \alpha) \in \mathbb{T} \times \mathbb{A}, \|\alpha - \alpha_0\| \leq \eta, \|\mathbf{t} - \mathbf{t}^i\| \leq \eta \text{ for some } i. \end{aligned}$$

Set  $\Delta = ((\delta^1)^T, \dots, (\delta^{p_0})^T)$ , with  $\delta_i \in \mathbb{R}^h$ ;  $\Theta = ((\tau^1)^T, \dots, (\tau^{p_0})^T)$ , with  $\tau_i \in \mathbb{R}^d$ ;  
 $\mathbf{T} = (\mathbf{t}^1, \dots, \mathbf{t}^q) \in \mathbb{T}^q$ ;  $\Xi = (\rho_1, \dots, \rho_q) \in \mathbb{R}^q$ ;  $\Pi = (\pi_1, \dots, \pi_{p_0}) \in [0, 1]^{p_0}$ ,

$$H_{\mathbf{T}, \Xi, \Pi, \Delta, \Theta} = \sum_{i=1}^q \rho_i f_{\mathbf{t}^i, \alpha_0} + \sum_{i=1}^{p_0} \pi_i f_{\mathbf{t}^{0,i}, \alpha_0} + \sum_{i=1}^{p_0} \langle \delta^i, D_\alpha f_{\mathbf{t}^{0,i}, \alpha_0} \rangle + \sum_{i=1}^{p_0} \langle \tau^i, D_{\mathbf{t}} f_{\mathbf{t}^{0,i}, \alpha_0} \rangle,$$

and

$$d_{\mathbf{T}, \Xi, \Pi, \Delta, \Theta} = \frac{H_{\mathbf{T}, \Xi, \Pi, \Delta, \Theta} / g_0}{\|H_{\mathbf{T}, \Xi, \Pi, \Delta, \Theta} / g_0\|_2}.$$

Define now :

$$\begin{aligned} \mathcal{K} &= \{(\mathbf{T}, \Xi, \Pi, \Delta, \Theta) : \rho_1, \dots, \rho_q \geq 0; \sum_i \rho_i + \sum_i \pi_i = 0; \\ & \|\delta^1\|^2 + \dots + \|\delta^{p_0}\|^2 + \|\tau^1\|^2 + \dots + \|\tau^{p_0}\|^2 + \sum_i \rho_i^2 + \sum_i \pi_i^2 = 1; H_{(\mathbf{T}, \Xi, \Pi, \Delta, \Theta)} \neq 0\}. \end{aligned}$$

Then :

$$\mathcal{D} = \{d_{\mathbf{T}, \Xi, \Pi, \Delta, \Theta}, (\mathbf{T}, \Xi, \Pi, \Delta, \Theta) \in \mathcal{K}\},$$

and

$$\mathcal{D}_0 = \{d_{\mathbf{0}, \mathbf{0}, \mathbf{0}, \Delta, \Theta}\}.$$

It will be possible to obtain the asymptotic distributions in the same way as in Section 6.2.2 under Assumptions (WID) and (GM). Define the Gaussian field  $\mathcal{Z}(\mathbf{T}, \Xi, \Pi, \Delta, \Theta)$  on  $\mathcal{K}$  with covariance

$$r((\mathbf{T}, \Xi, \Pi, \Delta, \Theta), (\mathbf{T}', \Xi', \Pi', \Delta', \Theta')) = \int d_{\mathbf{T}, \Xi, \Pi, \Delta, \Theta} d_{\mathbf{T}', \Xi', \Pi', \Delta', \Theta'} g_0 d\nu.$$

Notice that, as in previous sections,  $\mathcal{K}$  is not closed, and  $r(\cdot, \cdot)$  is not continuous on some limiting points, but may be extended in some sense, as has been done for instance



in Section 6.2.1.

Let also  $p_n, \pi_n, \mathbf{T}_n, \alpha_n$  be such that  $\sqrt{n} \left\| \frac{g_{p_n, \pi_n, \mathbf{T}_n, \alpha_n} - g_0}{g_0} \right\|_2$  tends to some positive constant  $c$ , with  $\frac{g_{p_n, \pi_n, \mathbf{T}_n, \alpha_n} - g_0}{g_0} / \left\| \frac{g_{p_n, \pi_n, \mathbf{T}_n, \alpha_n} - g_0}{g_0} \right\|_2$  tending to  $\bar{d}$  in the closure of  $\mathcal{D}$ .

**Theorem 6.2.4**

If (WID) and (GM) hold, then  $(g_0 \cdot \nu)^{\otimes n}$  and  $(g_{p_n, \pi_n, \mathbf{T}_n, \alpha_n} \cdot \nu)^{\otimes n}$  are mutually contiguous,  $2\lambda_n$  converges under  $(g_0 \cdot \nu)^{\otimes n}$  in distribution to

$$\left( \sup_{(\mathbf{T}, \Xi, \Pi, \Delta, \Theta) \in \mathcal{K}} \mathcal{Z}(\mathbf{T}, \Xi, \Pi, \Delta, \Theta) \right)^2 - \left( \sup_{(\mathbf{0}, \mathbf{0}, \mathbf{0}, \Delta, \Theta) \in \mathcal{K}} \mathcal{Z}(\mathbf{0}, \mathbf{0}, \mathbf{0}, \Delta, \Theta) \right)^2,$$

and under  $(g_{p_n, \pi_n, \mathbf{T}_n, \alpha_n} \cdot \nu)^{\otimes n}$  to

$$(6.17) \quad \left( \sup_{(\mathbf{T}, \Xi, \Pi, \Delta, \Theta) \in \mathcal{K}} \mathcal{Z}(\mathbf{T}, \Xi, \Pi, \Delta, \Theta) + c \int d_{\mathbf{T}, \Xi, \Pi, \Delta, \Theta} \bar{d} g_0 d\nu \right)^2 - \left( \sup_{(\mathbf{0}, \mathbf{0}, \mathbf{0}, \Delta, \Theta) \in \mathcal{K}} \mathcal{Z}(\mathbf{0}, \mathbf{0}, \mathbf{0}, \Delta, \Theta) + c \int d_{\mathbf{0}, \mathbf{0}, \mathbf{0}, \Delta, \Theta} \bar{d} g_0 d\nu \right)^2.$$

It is possible to reduce the formula of the asymptotic distributions in Theorem 6.2.4 into only one supremum, using linear algebra computations as in previous sections. We shall not give the result for all situations since it involves too long and complicated formula. However, in case  $q = 1$ , the result takes a simpler form that we will give below. For this one needs to define notations. When  $q = 1$ ,  $\Xi$  is reduced to  $\rho$  and  $\mathbf{T}$  reduces to  $\mathbf{t}$  so that elements of  $\mathcal{D}$  may be written as  $d_{\mathbf{t}, \Pi, \Delta, \Theta}$  with

$$H_{\mathbf{t}, \Pi, \Delta, \Theta} = \sum_{i=1}^{p_0} \pi_i (f_{\mathbf{t}, \alpha_0} - f_{\mathbf{t}^{0,i}, \alpha_0}) + \sum_{i=1}^{p_0} \langle \delta^i, D_\alpha f_{\mathbf{t}^{0,i}, \alpha_0} \rangle + \sum_{i=1}^{p_0} \langle \tau^i, D_{\mathbf{t}} f_{\mathbf{t}^{0,i}, \alpha_0} \rangle.$$

where  $\sum_{i=1}^{p_0} \pi_i \geq 0$ .

Let  $W$  be the  $p_0(h+d)$ -dimensional centered Gaussian random variable with variance  $\Sigma$  such that for all  $\Delta$  and  $\Theta$ ,

$$\begin{pmatrix} \Delta \\ \Theta \end{pmatrix}^T \Sigma \begin{pmatrix} \Delta \\ \Theta \end{pmatrix} = \left\| \frac{H_{\mathbf{0}, \mathbf{0}, \Delta, \Theta}}{g_0} \right\|_2^2.$$

Let  $Z(\mathbf{t})$  be the  $(p_0)$ -dimensional centered Gaussian field with covariance the  $p_0 \times p_0$  matrix  $\Gamma(\cdot, \cdot)$  such that for all  $\mathbf{t}_1, \mathbf{t}_2$ ,

$$\Gamma(\mathbf{t}_1, \mathbf{t}_2)_{i,j} = \int \left( \frac{\frac{f_{\mathbf{t}, \alpha_0} - f_{\mathbf{t}^{0,i}, \alpha_0}}{g_0}}{\left\| \frac{f_{\mathbf{t}, \alpha_0} - f_{\mathbf{t}^{0,i}, \alpha_0}}{g_0} \right\|_2} \right) \left( \frac{\frac{f_{\mathbf{t}, \alpha_0} - f_{\mathbf{t}^{0,j}, \alpha_0}}{g_0}}{\left\| \frac{f_{\mathbf{t}, \alpha_0} - f_{\mathbf{t}^{0,j}, \alpha_0}}{g_0} \right\|_2} \right) g_0 d\nu,$$

and let

$$\Gamma = \Gamma(\mathbf{t}) = \Gamma(\mathbf{t}, \mathbf{t}).$$

Define  $C = C(\mathbf{t})$  the  $p_0(h + d) \times p_0$  matrix such that for all  $(\mathbf{t}, \Pi, \Delta, \Theta)$ ,

$$\begin{pmatrix} \Delta \\ \Theta \end{pmatrix}^T C(\mathbf{t})\Pi = \sum_{i=1}^{p_0} \pi_i \left\langle \frac{\frac{f_{\mathbf{t},\alpha_0} - f_{\mathbf{t}^0,i,\alpha_0}}{g_0}}{\left\| \frac{f_{\mathbf{t},\alpha_0} - f_{\mathbf{t}^0,i,\alpha_0}}{g_0} \right\|_2}, \frac{H_{0,0,\Delta,\Theta}}{g_0} \right\rangle_2,$$

and let  $A = A(\mathbf{t})$ ,  $U = U(\mathbf{t})$ ,  $M = M(\mathbf{t})$  be matrices such that

$$(6.18) \quad M = (\Sigma - C\Gamma^{-1}C^T)^{-1}$$

$$(6.19) \quad U = -MC\Gamma^{-1}$$

$$(6.20) \quad A = \Gamma^{-1} + \Gamma^{-1}C^TMC\Gamma^{-1}.$$

Let  $\mathbf{1}$  denote the  $p_0$ -dimensional vector with all coordinates equal to 1. Then :

**Theorem 6.2.5**

Assume (WID) and (GM), and  $p = p_0 + 1$ . Then  $2\lambda_n$  converges under  $(g_0 \cdot \nu)^{\otimes n}$  in distribution to

$$\sup_{\mathbf{t}} (AZ + U^TW)^T \left( A^{-1} - \frac{\mathbf{1}\mathbf{1}^T}{\mathbf{1}^T A \mathbf{1}} \mathbf{1}_{(AZ+U^TW)^T \mathbf{1} < 0} \right) (AZ + U^TW).$$

The distribution under contiguous alternatives is rather difficult to express in its full generality so it is omitted for simplicity. The proof of Theorem 6.2.5 is given in Section 6.5.

In the case of Gaussian mixtures with unknown variance, the assumption

$$“ \sum_{i=1}^{p_0} \langle \delta^i, D_\alpha f_{\mathbf{t}^0,i,\alpha^i} \rangle + \sum_{j \in J} (\tau^j)^T D_{\mathbf{t}}^2 f_{j,\alpha_0}(\tau^j) = 0 \nu \text{ a.e. if and only if} ”$$

$$\delta^1 = \mathbf{0}, \dots, \delta^{p_0} = \mathbf{0} \text{ and } \tau^j = \mathbf{0}, j \in J”$$

does not hold. Indeed, second derivatives with respect to  $\mathbf{t}$  are proportional with derivatives with respect to  $\alpha$ . In this case, it is necessary to go further in the Taylor development : when taking third derivative with respect to  $\mathbf{t}$ , the condition of non degeneracy holds. Also, all derivatives till fourth order may be uniformly upper bounded with some function  $B$  as needed. Since the limiting points of process  $Z$  need not to be known at boundary values of  $\mathcal{K}$  to define the asymptotic distribution of  $\lambda_n$ , the following result holds :

**Theorem 6.2.6**

The asymptotic distributions under the null hypothesis and under contiguous hypotheses given in Theorem 6.2.4 and Theorem 6.2.5 hold for Gaussian mixtures with the same unknown variance matrix.

### 6.3 The LRT for contamination mixtures with large set of parameters

As already said in the introduction, the asymptotic distribution of the LRT for compact  $\mathbb{T}$  and  $\mathbb{A}$  can be used in practice for large data sets. The LRT happens in this case to be more powerful than moment tests as shown in Delmas (2003). Nevertheless it suffers from two drawbacks :

- the distribution is not free from the location of the null hypothesis inside  $\mathbb{T}$ ,
- for testing one population against two (or  $p_0$  against  $p$ ) the LRT with bounded parameter is not invariant by translation or change of scale.

Several solutions to the first point exist. Threshold calculation can be conducted under the “worst” form of the null hypothesis (see Delmas, 2003) or one can use a “Plug-in”, that is an estimate of  $f_0$ . It remains that results would be nicer if one would be able to get rid of the compactness assumption. This section and the next one answer by the negative, showing that in the simplest case : contamination for translation mixtures on  $\mathbb{R}$ , the LRT is theoretically less powerful than moment tests under contiguous alternatives. As already said in the introduction, the convergence to this result is very slow, so it is not so relevant in practice. It mainly shows that it is difficult to construct an unbounded asymptotic theory of the LRT.

We consider in this section the contamination mixture model (6.1) with  $\mathbb{T} = [-T, T]$  for a given positive real number  $T$  and  $\nu$  the Lebesgue measure. We use notations and results of Section 6.2.1. Let  $\pi_n$  and  $t_n$  be sequences such that :

- $\lim_{n \rightarrow +\infty} \sqrt{n} \pi_n \|(f_{t_n} - f_0)/f_0\|_2 = c$  for some positive  $c$ ,
- either  $t_n$  tends to some  $t_0 \neq 0$  and  $\sqrt{n} \pi_n$  tends to some positive constant, or  $t_n$  tends to 0, and  $t_n/\|t_n\|$  converges to some limit  $\tau$ .

Let  $\mathbb{P}_{\pi_n, t_n} = \left(g_{\pi_n, t_n} \cdot \nu\right)^{\otimes n}$  and  $\mathbb{P}_0 = \left(f_0 \cdot \nu\right)^{\otimes n}$ . To evaluate the asymptotic power and the asymptotic level for large values of  $T$ , one has to investigate the behaviour of suprema of the Gaussian processes  $Z(t)$  and  $Z(t) + m(t)$  as defined in Theorem 6.2.1.  $Z$  is the centered Gaussian process defined in Section 6.2.1 with covariance given by (6.9). For simplicity we consider this process as defined on the whole real line  $\mathbb{R}$ . We will use assumptions under which the supremum of  $Z(\cdot)$  over  $[-T, T]$  tends to infinity as  $T$  tends to infinity, and is achieved for some  $t$  tending to infinity. So the study of this supremum on  $[0, T]$  for large  $T$  can be replaced by the study of the supremum on  $[1, T]$ . The discontinuity of the covariance function  $r$  at 0 will have for us no consequence on the extreme behaviour of the process  $Z$ . We shall use Azaïs and Mercadier (2003) to derive the asymptotic distribution of suprema of Gaussian processes. Hence let  $M(a, b) = \sup_{t \in (a, b)} (Z(t) + m(t))$ . Since the asymptotic distribution of  $2 \lambda_n$ , under the null hypothesis or under contiguous alternatives, in Theorem 6.2.1 can be written as  $M(-T, T)^2$  (taking  $m(t) = 0$  under the null hypothesis and  $m(t) = \mu(t)$  as defined by (6.11) under contiguous alternatives), we want to characterize asymptotic behaviours of  $M(-T, T)$  as  $T \rightarrow +\infty$ . We thus introduce further notations and assumptions.

Write  $r_{ij}(s, t)$  instead of  $\frac{\partial^{i+j}}{\partial^i s \partial^j t} r(s, t)$  and define  $R(t) = \int_0^t \sqrt{r_{11}(s, s)} ds$ .

Let  $a_t = \sqrt{2 \log \circ R(t)}$ ,  $b_t = a_t - \frac{\log(2\pi)}{a_t}$  and  $\tilde{b}_t = a_t - \frac{\log(\pi)}{a_t}$ .

Let  $V = \{V(t) = Z(R^{-1}(t)) + m(R^{-1}(t)), t \in \mathbb{R}\}$  be the “unit-speed” transformation of  $Z + m$  in the sense that the variance of  $V'(t)$  equals 1 for all  $t$  in  $\mathbb{R}$ . We denote by  $r^V$  its covariance function.

We shall use the following Assumptions (G) on  $r$  and  $\mu$  :

(G)

(G1)  $\forall t \in \mathbb{R}, r_{11}(t, t) > 0$  and  $\lim_{t \rightarrow +\infty} R(t) = +\infty$ ,

(G2)  $r(s, t) \log |R(s) - R(t)| \rightarrow 0$  as  $|R(s) - R(t)| \rightarrow +\infty$ ,

(G3)  $\forall \varepsilon > 0 \sup_{|R(s)-R(t)| > \varepsilon} |r(s, t)| < 1$ ,

(G4)  $\star r$  is four times continuously differentiable,

$\star s \rightarrow r_{11}(s, s)$  three times continuously differentiable,

$\star \forall \gamma > 0, r_{01}^Y$  and  $r_{04}^Y$  are bounded on  $\{(s, t) \in \mathbb{R}^2, |s| > \gamma \text{ and } |t| > \gamma\}$ ,

(G5)  $\sqrt{\log \circ R(t)} \mu(t) \xrightarrow{t \rightarrow +\infty} 0$ .

We have :

**Theorem 6.3.1**

Assume (CM) and (G). Then, as  $T$  tends to infinity,  $a_T(M(-T, T) - \tilde{b}_T)$  tends in distribution to the Gumbel distribution when  $m(t) = 0$  as well as when  $m(t) = \mu(t)$ . In other words, if  $c_{T,\alpha}$  is the threshold of the test defined by

$$\lim_{n \rightarrow +\infty} \mathbb{P}_0(\lambda_n > c_{T,\alpha}) = \alpha,$$

then for any contiguous alternative, the limiting power of the LRT equals its level :

$$\lim_{T \rightarrow +\infty} \lim_{n \rightarrow +\infty} \mathbb{P}_{\pi_n, t_n}(\lambda_n > c_{T,\alpha}) = \alpha.$$

Theorem 6.3.1 is proved in Section 6.5.

The theorem says that for  $T$  large enough, asymptotically, the LRT cannot distinguish the null hypothesis from any contiguous alternatives.

We shall consider the translation mixture model defined in Section 6.2.1. Let  $f_0$  be a density on  $\mathbb{R}$  satisfying Assumptions (H) where we denote by  $f_0^{(i)}$  the derivative of  $f_0$  of order  $i$ .

(H)

(H1)  $\forall x \in \mathbb{R}, f_0(x) > 0$ ,  $f_0$  four times continuously differentiable,

and  $\forall i = 1, \dots, 4, \exists K_i > 0, \forall x \in \mathbb{R}, \left| \frac{f_0^{(i)}}{f_0}(x) \right| \leq K_i$ ,

(H2)  $\forall x \in \mathbb{R}, \lim_{t \rightarrow +\infty} \frac{f_0(x+t)}{f_0(t)} = \lim_{t \rightarrow -\infty} \frac{f_0(x+t)}{f_0(t)} = 1$ ,

(H3)  $\exists M > 0, \forall x, t \in \mathbb{R}, \frac{f_0(x)f_0(t)}{f_0(x+t)} \leq M$ ,

(H4)  $\exists F \in L^2(\lambda) : \sup_{|t| \geq 1} \log |t| \sqrt{f_0(x+t)} \leq F(x)$ ,

(H5)  $\lim_{t \rightarrow +\infty} \log(t) \sqrt{f_0(t)} = 0$ .

Our result is now :

**Corollary 6.3.1**

Assume (H). Then Theorem 6.3.1 applies to the translation mixture model.

Proof of Corollary 6.3.1 is given in Section 6.5.

**Remarks :**

- Assumptions (H1) to (H5) are essentially conditions on the tail of  $f_0$ . (H4) and (H5) are very weak and hold for all usual distributions. But (H1) to (H3) though rather weak, are more restrictive. They hold for example if  $f_0(t) = O(t^{-\alpha})$  for  $\alpha > 0$  as  $t \rightarrow +\infty$  and  $f_0(t) = O(t^{-\beta})$  for  $\beta > 0$  as  $t \rightarrow -\infty$ . For instance, they hold for  $f_0$  being the inverse of a polynomial and in particular for the Cauchy density.
- The proof relies on the verification of assumptions of Theorem 6.3.1. In particular, asymptotic behaviours of the covariance  $r$  and its derivatives have to be checked. Assumptions (H) only express sufficient conditions under which the asymptotic analysis is done with some generality. However, though (H2) does not hold for the Gaussian density, we also verified that Theorem 6.3.1 holds for other densities such as the Gaussian and the normalization of  $\text{ch}(x)^{-1}$  in spite of different justifications.

LRT has to be compared with other testing procedures such as sample mean or Kolmogorov-Smirnov testing procedures.

- Denote by  $\mu_i = \int x^i f_0(x) d\nu(x)$ . Without loss of generality one can assume that  $\mu_1 = 0$ . If  $\mu_2 < +\infty$  applying Le Cam's third Lemma, that is Theorem 6.6 of van der Vaart (1998),

$\sqrt{n} \bar{X}_n$  converges in distribution, as  $n$  tends to infinity, to the Gaussian  $N(0, \mu_2)$  under  $\mathbb{P}_0$  and to the Gaussian  $N(\gamma, \mu_2)$  under  $\mathbb{P}_{\pi_n, t_n}$ , where  $\gamma = c / \|\frac{f_0'}{f_0}\|_2$  if  $t_n \rightarrow 0$  and  $\gamma = ct_0 / \|\frac{ft_0 - f_0}{f_0}\|_2$  if  $t_n \rightarrow t_0 \neq 0$ .

Consequently the asymptotic power is greater than the level.

- Remark that, when no condition of moment is available, other tests can be also proposed. Define  $\mathbb{F}_n$  the random distribution function and  $F_0$  the distribution function associated to  $f_0$ . Let  $I$  denote the identity function on  $[0, 1]$  and let  $\mathbb{U}$  be a Brownian bridge on  $[0, 1]$ . Let  $\|\cdot\|_\infty$  denotes the supremum norm. The natural normalization of  $\mathbb{F}_n$  leads to the definition of the Kolmogorov-Smirnov statistic  $\mathbb{K}_n$  and the Cramér-von Mises statistic  $\mathbb{W}_n^2$  :

$$\mathbb{K}_n = \sqrt{n} \|\mathbb{F}_n - F_0\|_\infty \text{ and } \mathbb{W}_n^2 = \int_{-\infty}^{+\infty} n [\mathbb{F}_n(x) - F_0(x)]^2 dF_0(x).$$

Set on  $[0, 1]$

$$\Delta(x) = \gamma \lim_{n \rightarrow +\infty} \frac{F_0(F_0^{-1}(x) - t_n) - x}{t_n},$$

where  $t_n$  is the translation parameter of the alternative. Hence  $\Delta$  depends on the asymptotic behaviour of  $t_n$ .

- \*  $\mathbb{K}_n$  converges in distribution, as  $n$  tends to infinity, to  $\|\mathbb{U}\|_\infty$  under  $\mathbb{P}_0$  and  $\|\mathbb{U} + \Delta\|_\infty$  under  $\mathbb{P}_{\pi_n, t_n}$ .

★  $\mathbb{W}_n^2$  converges in distribution, as  $n$  tends to infinity, to  $\int_0^1 \mathbb{U}^2 dI$  under  $\mathbb{P}_0$  and  $\int_0^1 (\mathbb{U} + \Delta)^2 dI$  under  $\mathbb{P}_{\pi_n, t_n}$ .

See Shorack and Wellner (1986) for a version of these convergences. Simulations show that in both cases, the distribution under  $\mathbb{P}_{\pi_n, t_n}$  is stochastically greater than that under  $\mathbb{P}_0$ . Consequently the asymptotic power is greater than the level.

## 6.4 Asymptotic distribution of the LRT for Gaussian contamination mixtures with unbounded mean under contiguous alternatives

Consider  $\mathbb{T} = \mathbb{R}$  (no prior upper bound) and the testing problem (6.1) with

$$f_t(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-t)^2}{2}\right).$$

Set

$$g_0 = f_0 \text{ and } g_{\pi, t} = (1 - \pi)f_0 + \pi f_t, \quad 0 \leq \pi \leq 1, t \in \mathbb{T}.$$

Let  $\pi_n$  and  $t_n$  be sequences such that  $\lim_{n \rightarrow +\infty} \sqrt{n}\pi_n t_n = \gamma \in \mathbb{R}$  and  $\lim_{n \rightarrow +\infty} t_n = t_0 \in \mathbb{R}$ . Note that  $t_0$  can be equal to zero.

$\lambda_n$  is now given by :

$$\lambda_n = \sup_{\pi \in [0,1], t \in \mathbb{R}} \sum_{i=1}^n \log\left(1 + \pi \left(\exp[tX_i - \frac{t^2}{2}] - 1\right)\right).$$

Then :

### Theorem 6.4.1

As  $n$  tends to infinity,  $(2\lambda_n - \log \circ \log n + \log(2\pi^2))$  tends in distribution to the Gumbel distribution under  $\mathbb{P}_0$  as well as under  $\mathbb{P}_{\pi_n, t_n}$  for any  $\gamma$  and  $t_0$ . In other words, let us define as rejection values the region :  $(\lambda_n > c_{\alpha, n})$  with

$$c_{\alpha, n} = \frac{1}{2}(G_{1-\alpha} + \log \circ \log n - \log(2\pi^2)),$$

where  $G_{1-\alpha}$  is the  $1 - \alpha$  fractile of the Gumbel distribution. We have by definition

$$\lim_{n \rightarrow +\infty} \mathbb{P}_0(\lambda_n > c_{\alpha, n}) = \alpha.$$

Then for any  $\gamma$  and  $t_0$ , the limit power of the LRT is

$$\lim_{n \rightarrow +\infty} \mathbb{P}_{\pi_n, t_n}(\lambda_n > c_{\alpha, n}) = \alpha.$$

The theorem says that asymptotically, the LRT cannot distinguish the null hypothesis from any contiguous alternative. Indeed, this has to be compared with other testing procedures such as moment testing procedures. For example, if  $\bar{X}_n$  is the sample mean, applying Le Cam's third Lemma,  $\sqrt{n}\bar{X}_n$  converges in distribution, under  $\mathbb{P}_{\pi_n, t_n}$  as  $n$  tends to infinity, to the Gaussian  $N(\gamma, 1)$ . Thus the test based on the statistic  $\sqrt{n}\bar{X}_n$  has an asymptotic power that is strictly greater than the level. As mentioned in the introduction this makes sense in practice only for very large data sets.

### Proof of Theorem 6.4.1

The separation of the hypotheses is greater when  $\gamma \neq 0$ . Using Lemma 14.31 of van der Vaart (1998) it is easy to see that this is the only case to consider. Moreover by symmetry we can suppose also that  $\gamma > 0$ . Let us introduce  $S_n$  the empirical process defined by

$$S_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \exp[tX_i - t^2] - \exp\left(-\frac{t^2}{2}\right) \right\}.$$

Liu and Shao (2000, Theorem 1) recall results obtained by Bickel and Chernoff (1993) on the process  $S_n$  :

$$(6.21) \quad \sup_{t \in \mathbb{R}} S_n(t) = \sup_{|t| \in A_{2,n}} S_n(t) + o_{\mathbb{P}_0}(1)$$

where  $A_{2,n} = [\alpha_n, \beta_n]$ ,  $\alpha_n = 2\sqrt{\log \circ \log \circ \log n}$  and  $\beta_n = \sqrt{\log n/2} - 2\sqrt{\log \circ \log n}$ .

Through the proof of their Theorem 2 Liu and Shao (2000) state that

$$2\lambda_n = \sup_{t \in \mathbb{R}} S_n(t)^2 + o_{\mathbb{P}_0}(1).$$

Combining with (6.21), the last equality becomes

$$2\lambda_n = \sup_{|t| \in A_{2,n}} S_n(t)^2 + o_{\mathbb{P}_0}(1).$$

Let us denote  $\tilde{\mathbb{P}}_0$  the extension of  $\mathbb{P}_0$  constructed by Bickel and Chernoff (1993) by Hungarian construction. According to their formula (39), we get

$$(6.22) \quad 2\lambda_n = \sup_{|t| \in A_{2,n}} S_0(t)^2 + o_{\tilde{\mathbb{P}}_0}(1)$$

where  $S_0$  is the zero mean non-stationary Gaussian process with covariance function

$$\rho(s, t) = \exp\left[-\frac{(s-t)^2}{2}\right] - \exp\left[-\frac{s^2}{2} - \frac{t^2}{2}\right].$$

In their paper, Bickel and Chernoff remark that this process is very close to a stationary process namely  $\tilde{S}_0$ . Because we need it later, we will use here an other way. We define the standardized version of  $S_0$

$$Y_0(t) = \frac{S_0(t)}{\sqrt{\rho(t, t)}} = \frac{S_0(t)}{\sqrt{1 - e^{-t^2}}},$$

in order to be able to apply the Normal Comparison Lemma (Li and Shao, 2002, Theorem 2.1).  $Y_0$  is a zero mean non-stationary Gaussian process, with unit variance and covariance function

$$(6.23) \quad r(s, t) = \frac{\exp(st) - 1}{\sqrt{\exp(s^2) - 1} \sqrt{\exp(t^2) - 1}}.$$

We have

$$0 \leq \sup_{|t| \in A_{2,n}} |Y_0(t) - S_0(t)| \leq \sup_{|t| \in A_{2,n}} (1 - \sqrt{\rho(t,t)}) \sup_{|t| \in A_{2,n}} |Y_0(t)|.$$

Now the function  $r$  satisfies conditions of Corollary 1 of Azaïs and Mercadier (2003). Consequently we know the exact order of the maximum

$$\sup_{|t| \in A_{2,n}} |Y_0(t)| = O_{\tilde{\mathbb{P}}_0}((\log \circ \log n)^{\frac{1}{2}}).$$

This last equation can also be deduced from standard result on the maximum of stationary Gaussian processes using the process  $\tilde{S}_0$  introduced by Bickel and Chernoff (1993).

On the other side, the maximum of  $1 - \sqrt{\rho(t,t)}$  on  $A_{2,n}$  is obtained at  $\alpha_n$ . This permits us to write

$$0 \leq \sup_{|t| \in A_{2,n}} |Y_0(t) - S_0(t)| \leq O_{\tilde{\mathbb{P}}_0}((\log \circ \log n)^{\frac{1}{2}-4}).$$

Finally this approximation allows us to replace  $S_0$  by  $Y_0$  in (6.22) to get

$$(6.24) \quad 2 \lambda_n = \sup_{|t| \in A_{2,n}} Y_0(t)^2 + o_{\tilde{\mathbb{P}}_0}(1).$$

With the same idea as before, we define

$$Y_n(t) = \frac{S_n(t)}{\sqrt{1 - e^{-t^2}}}.$$

For all  $t_0$  and all  $\gamma$ , using argument close to those that lead to formula (7) in Gassiat (2002) we have

$$(6.25) \quad \log \frac{d\mathbb{P}_{\pi_n, t_n}}{d\mathbb{P}_0}(X_1, \dots, X_n) = C(\gamma, t_0) Y_n(t_n) - \frac{C(\gamma, t_0)^2}{2} + o_{\mathbb{P}_0}(1)$$

with  $C(\gamma, t_0) = \gamma$  if  $t_0 = 0$  and  $C(\gamma, t_0) = \gamma \frac{\sqrt{e^{t_0^2} - 1}}{t_0}$  if  $t_0 > 0$ . Since  $\gamma$  can be supposed positive,  $t_0$  is positive. A detailed proof of formula (6.25) is given in Section 6.5. Using the formula (39) of Bickel and Chernoff (1993), we can replace  $Y_n$  by  $Y_0$  to get

$$(6.26) \quad \log \frac{d\mathbb{P}_{\pi_n, t_n}}{d\mathbb{P}_0}(X_1, \dots, X_n) = C(\gamma, t_0) Y_0(t_n) - \frac{C(\gamma, t_0)^2}{2} + o_{\tilde{\mathbb{P}}_0}(1).$$

We next use the following lemma, its proof is given in Section 6.5.

**Lemma 6.4.1**

For all  $t_0$ ,  $2 \lambda_n - \log \circ \log n + \log(2\pi^2)$  and  $\log \frac{d\mathbb{P}_{\pi_n, t_n}}{d\mathbb{P}_0}(X_1, \dots, X_n)$  are asymptotically independent under  $\mathbb{P}_0$ .

Then, as soon as one proves Lemma 6.4.1 the theorem follows from a generalization of Le Cam’s third Lemma. The proof of Lemma 6.4.1 relies on a suitably chosen discretization, following ideas in Azaïs and Mercadier (2003), and an application of the normal comparison lemma as refined in Li and Shao (2002).



## 6.5 Proofs

### 6.5.1 Proof of Theorem 6.2.5

To obtain the supremum in the first limit of Theorem 6.2.4, one has to compute the supremum of :

$$(6.27) \quad \left( \left( \begin{array}{c} \Pi \\ \Delta \\ \Theta \end{array} \right)^T \left( \begin{array}{c} Z \\ W \end{array} \right) \right)^2$$

under the constraints

$$(6.28) \quad \left( \begin{array}{c} \Pi \\ \Delta \\ \Theta \end{array} \right)^T \tilde{\Sigma} \left( \begin{array}{c} \Pi \\ \Delta \\ \Theta \end{array} \right) = 1, \quad \Pi^T \mathbf{1} \geq 0,$$

where

$$\tilde{\Sigma} = \left( \begin{array}{cc} \Gamma & C^T \\ C & \Sigma \end{array} \right).$$

Consider the supremum under the first constraint. Then, similarly to the proof of Theorem 6.2.2, the value of the supremum is

$$(AZ + U^T W)^T A^{-1} (AZ + U^T W) + W^T \Sigma^{-1} W$$

and it is attained on some  $\Pi$  such that  $\Pi^T \mathbf{1}$  has the same sign as  $(AZ + U^T W)^T \mathbf{1}$ . If  $(AZ + U^T W)^T \mathbf{1} < 0$ , then the supremum of (6.27) under (6.28) equals the supremum of (6.27) under the constraints

$$(6.29) \quad \left( \begin{array}{c} \Pi \\ \Delta \\ \Theta \end{array} \right)^T \tilde{\Sigma} \left( \begin{array}{c} \Pi \\ \Delta \\ \Theta \end{array} \right) = 1, \quad \Pi^T \mathbf{1} = 0.$$

Computation of this supremum using Lagrange multipliers leads to the fact that it is equal to

$$(AZ + U^T W)^T \left( A^{-1} - \frac{\mathbf{1}\mathbf{1}^T}{\mathbf{1}^T A \mathbf{1}} \right) (AZ + U^T W) + W^T \Sigma^{-1} W$$

and the Theorem is proved.

### 6.5.2 Proof of Theorem 6.3.1

Set  $u_{T,x} = \frac{x}{a_T} + \tilde{b}_T$  and  $M^V(a, b) = \sup_{t \in (a,b)} V_t$  for  $V$  the unit-speed transformation of  $Z + m$ .

★ We have

$$\mathbb{P}\left(M(-T, T) \leq u_{T,x}\right) = \mathbb{P}\left(M^V(-R(T), R(T)) \leq u_{T,x}\right).$$

Now, applying with  $p = 2$ ,  $D_1 = (-R(T), -\sqrt{R(T)})$  and  $D_2 = (\sqrt{R(T)}, R(T))$  Proposition 4 of Azaïs and Mercadier (2003), we obtain

$$\mathbb{P}(M^V(D_1 \cup D_2) \leq u_{T,x}) = \mathbb{P}(M^V(D_1) \leq u_{T,x})\mathbb{P}(M^V(D_2) \leq u_{T,x}) + o(1).$$

Remark that in Azaïs and Mercadier (2003) sizes of intervals are defined as functions of the level, here it is the opposite which is made. Furthermore, repeated application of Corollary 1 of Azaïs and Mercadier (2003) enables us to state for  $\tau = \sqrt{R(T)}$  and  $\tau = R(T)$  the convergence of  $a_\tau(M^V(0, \tau) - b_\tau)$  and  $a_\tau(M^V(-\tau, 0) - b_\tau)$  to the Gumbel. It follows that  $M^V(-\sqrt{R(T)}, \sqrt{R(T)})$  is stochastically negligible compared with  $M^V(-R(T), R(T))$  and also that  $M(0, \sqrt{R(T)})$  (resp.  $M(-\sqrt{R(T)}, 0)$ ) is stochastically negligible compared with  $M^V(0, R(T))$  (resp.  $M^V(-R(T), 0)$ ). By combining what precedes, we get

$$\begin{aligned} & \mathbb{P}\left(M^V(-R(T), R(T)) \leq u_{T,x}\right) \\ &= \mathbb{P}\left(M^V(0, R(T)) \leq u_{T,x}\right)\mathbb{P}\left(M^V(-R(T), 0) \leq u_{T,x}\right) + o(1), \end{aligned}$$

as  $T$  tends to infinity, and which becomes

$$\mathbb{P}\left(M(-T, T) \leq u_{T,x}\right) = \mathbb{P}\left(M(0, T) \leq u_{T,x}\right)\mathbb{P}\left(M(-T, 0) \leq u_{T,x}\right) + o(1)$$

when we return to the initial process  $Z + m$ .

Let  $G(x) = \exp(-\exp(-x))$  denote the distribution function of the Gumbel. Corollary 1 of Azaïs and Mercadier (2003) yields, as  $T$  tends to infinity,

$$\begin{aligned} \mathbb{P}\left(M(0, T) \leq u_{T,x}\right) &= \mathbb{P}\left(a_T(M(0, T) - \tilde{b}_T) \leq x\right) + o(1) \\ &= \mathbb{P}\left(a_T(M(0, T) - b_T) \leq x + \log(2)\right) + o(1) \\ &= G(x + \log(2)) + o(1). \end{aligned}$$

Since the same equality holds on  $(-T, 0)$ , one can conclude that

$$\mathbb{P}\left(M(-T, T) \leq u_{T,x}\right) = G(x + \log(2))^2 + o(1) = G(x) + o(1).$$

### 6.5.3 Proof of Corollary 6.3.1

The proof relies on the verification of assumptions of Theorem 6.3.1.

*Proof of (CM) :* since  $f_0$  is continuous and positive, for any real  $T$ ,

$$\inf_{t \in [-T, T]} f_0(t) = \delta_T > 0.$$

Using (H3), for all  $t \in [-T, T]$  and  $x \in \mathbb{R}$ ,

$$\left| \frac{f_t - f_0}{f_0}(x) \right| \leq \sup_{x \in \mathbb{R}} \left| \frac{f_0(x-t)f_0(t)}{f_0(x)} \right| \frac{1}{f_0(t)} + 1 \leq \frac{M}{\delta_T} + 1,$$

and using (H1) and (H3)

$$\left| \frac{f'_t(x)}{f_0(x)} \right| \leq K_1 \frac{M}{\delta_T}, \quad \left| \frac{f''_t(x)}{f_0(x)} \right| \leq K_2 \frac{M}{\delta_T}.$$

Let us now prove Assumptions (G). Set

$$N(s, t) = \int \frac{f_0(x-t)f_0(x-s)}{f_0(x)} d\nu(x).$$

Differentiation of  $r$ , for  $s$  and  $t$  in  $\mathbb{R} \setminus \{0\}$ , is a consequence of that of  $N(s, t)$ . Now, for any integers  $i \leq 4$  and  $j \leq 4$ , using (H1) and (H3)

$$\frac{f_0^{(i)}(x-t)f_0^{(j)}(x-s)}{f_0(x)} \leq K_i K_j \frac{f_0(x-t)f_0(x-s)}{f_0(x)} \leq K_i K_j M^2 \frac{f_0(x)}{f_0(t)f_0(s)}$$

and  $f_0(t)f_0(s)$  is positively lower bounded on the neighbourhood of any  $(s_0, t_0)$ , which proves that  $N$  is differentiable at any  $(s, t) \in (\mathbb{R} \setminus \{0\})^2$  with

$$\frac{\partial^{i+j} N}{\partial^i t \partial^j s}(s, t) = (-1)^{i+j} \int \frac{f_0^{(i)}(x-t)f_0^{(j)}(x-s)}{f_0(x)} d\nu(x).$$

*Proof of (G1) :* we thus have for  $t \neq 0$

$$\begin{aligned} r_{11}(t, t) &= \frac{\int \frac{f_0'^2(x-t)}{f_0(x)} d\nu(x) \left( \int \frac{f_0^2(x-t)}{f_0(x)} d\nu(x) - 1 \right) - \left( \int \frac{f_0(x-t)f_0'(x-t)}{f_0(x)} d\nu(x) \right)^2}{\left( \int \frac{f_0(x-t)^2}{f_0(x)} d\nu(x) - 1 \right)^2} \\ &= \frac{\left\| \frac{f_0'(\cdot-t)}{f_0(\cdot)} \right\|_2^2 \left\| \frac{f_0(\cdot-t)-f_0(\cdot)}{f_0(\cdot)} \right\|_2^2 - \left( \left\langle \frac{f_0'(\cdot-t)}{f_0(\cdot)}, \frac{f_0(\cdot-t)-f_0(\cdot)}{f_0(\cdot)} \right\rangle_2 \right)^2}{\left\| \frac{f_0(\cdot-t)-f_0(\cdot)}{f_0(\cdot)} \right\|_2^4} \end{aligned}$$

which is positive by Cauchy-Schwarz inequality. Now,

$$\lim_{t \rightarrow +\infty} r_{11}(t, t) = \frac{\int f_0'^2 d\nu \int f_0^2 d\nu - \left( \int f_0 f_0' d\nu \right)^2}{\left( \int f_0'^2 d\nu \right)^2}.$$

Indeed, define the functions

$$\begin{aligned} A(t) &= \int \frac{f_0^2(x)}{f_0(x+t)} d\nu(x) \\ B(t) &= \int \frac{f_0'^2(x)}{f_0(x+t)} d\nu(x) \\ C(t) &= \int \frac{f_0(x)f_0'(x)}{f_0(x+t)} d\nu(x). \end{aligned}$$

Then write the function  $r_{11}$  under the following form :

$$r_{11}(t, t) = \frac{B(t)f_0(t)(A(t)f_0(t) - f_0(t)) - (C(t)f_0(t))^2}{(A(t)f_0(t) - f_0(t))^2}.$$

Thanks to (H1) and (H3), integrands of  $Af_0$ ,  $Bf_0$  and  $Cf_0$  are respectively dominated by

$$Mf_0(x), K_1^2 Mf_0(x), K_1 Mf_0(x).$$

By application of (H2) and Lebesgue Theorem, we conclude using the following convergences :

$$\begin{aligned} \lim_{t \rightarrow +\infty} A(t)f_0(t) &= \int f_0^2(x) d\nu(x) \\ \lim_{t \rightarrow +\infty} B(t)f_0(t) &= \int f_0'^2(x) d\nu(x) \\ \lim_{t \rightarrow +\infty} C(t)f_0(t) &= \int f_0(x)f_0'(x) d\nu(x). \end{aligned}$$

Thus for a positive constant  $R$

$$(6.30) \quad R(t) \sim_{t \rightarrow +\infty} Rt.$$

*Proof of (G2) :* considering (6.30), we have to prove that

$$(6.31) \quad \lim_{|s-t| \rightarrow +\infty} r(s, t) \log |s - t| = 0.$$

Using (H3),

$$\frac{f_0(t)f_0^2(x)}{f_0(x+t)} \leq Mf_0(x),$$

so that using (H2),

$$\lim_{t \rightarrow +\infty} \int \frac{f_0(t)f_0^2(x)}{f_0(x+t)} d\nu(x) = \int f_0^2(x) d\nu(x),$$

and there exists a constant  $C$  such that for  $|s - t|$  large enough,

$$r(s, t) \leq C \int \sqrt{f_0(t)} \sqrt{f_0(s)} \frac{f_0(x-t)f_0(x-s)}{f_0(x)} d\nu(x).$$

Then, using (H3),

$$r(s, t) \leq \int CM \sqrt{f_0(x)} \sqrt{f_0(x+s-t)} d\nu(x).$$

But according to (H5) for any  $x \in \mathbb{R}$ ,

$$\lim_{|s-t| \rightarrow +\infty} \log |s - t| \sqrt{f_0(x+s-t)} = 0,$$

and so, one may apply Lebesgue Theorem using (H4) to obtain (6.31).

*Proof of (G5) :* (G5) is a consequence of (G2) and formula (6.11) giving  $\mu(t)$ .

*Proof of (G3)* : Using (6.30) and the fact that  $r_{11} > 0$ , one just has to prove that for any  $\varepsilon > 0$ ,

$$(6.32) \quad \sup_{|s-t|>\varepsilon} |r(s, t)| < 1.$$

First of all,  $r(s, t)$  is a continuous function of  $(s, t)$  and  $|r(s, t)| < 1$  as soon as  $s \neq t$  by Cauchy-Schwarz inequality. Thus for any  $\varepsilon > 0$ , for any compact set  $K$ ,

$$\sup_{|s-t|>\varepsilon, t \in K, s \in K} |r(s, t)| < 1.$$

On the other hand because of (G2) for  $|s - t|$  sufficiently large  $r(s, t)$  is bounded away from 1, so we may suppose that  $|s - t|$  is bounded. Suppose that there exists  $s_n$  and  $t_n$  such that  $|s_n - t_n|$  is bounded,  $|s_n - t_n| > \varepsilon$  and  $r(s_n, t_n) \rightarrow 1$ . By compactness it would be possible to choose subsequences  $s_{\varphi(n)}$  and  $t_{\varphi(n)}$  such that  $s_{\varphi(n)} - t_{\varphi(n)} \rightarrow c$ . But using the same tricks as before (using (H2), (H3) and Lebesgue Theorem),

$$\lim_{n \rightarrow +\infty} r(s_{\varphi(n)}, t_{\varphi(n)}) = \frac{\int f_0(x) f_0(x+c) d\nu(x)}{\int f_0^2(x) d\nu(x)}.$$

Since  $|c| \geq \varepsilon > 0$  this value differs from 1. Hence we get a contradiction with assumptions made on sequences  $s_n$  and  $t_n$  and (6.32) is true.

*Proof of (G4)* : The first part of (G4) has been already proved. We use same arguments to prove that  $s \mapsto r_{11}(s, s)$  is three times continuously differentiable. Now, this last regularity associated to (6.30) allow us to reduce our study to that of functions  $r_{01}$  and  $r_{04}$ .

★ The first derivative  $r_{01}(s, t)$  can be written as :

$$\frac{-\langle \frac{f_0'(\cdot-t)}{f_0}, \frac{f_0(\cdot-s)-f_0}{f_0} \rangle_2}{\| \frac{f_0(\cdot-s)-f_0}{f_0} \|_2 \| \frac{f_0(\cdot-t)-f_0}{f_0} \|_2} + \frac{\langle \frac{f_0(\cdot-t)-f_0}{f_0}, \frac{f_0(\cdot-s)-f_0}{f_0} \rangle_{f_0} \langle \frac{f_0'(\cdot-t)}{f_0}, \frac{f_0(\cdot-t)-f_0}{f_0} \rangle_{f_0}}{\| \frac{f_0(\cdot-s)-f_0}{f_0} \|_2 \| \frac{f_0(\cdot-t)-f_0}{f_0} \|_2^3}$$

then Cauchy-Schwarz inequality leads to

$$|r_{01}(s, t)| \leq 2 \frac{\| \frac{f_0'(\cdot-t)}{f_0} \|_2}{\| \frac{f_0(\cdot-t)-f_0}{f_0} \|_2}.$$

This upper bound is a continuous function on  $t$ . By making appear  $f_0(t)$ , it is easily seen that it converges, as  $t$  tends to infinity, to

$$2 \frac{\int f_0'^2 d\nu}{\int f_0^2 d\nu}.$$

Moreover for any  $\delta > 0$ , the denominator is lower bounded on  $D_\delta = \{(s, t), s \in \mathbb{R}, |t| > \delta\}$ . Consequently for any  $\delta > 0$ ,  $(s, t) \mapsto r_{01}(s, t)$  is bounded on  $\mathbb{R}^2 \setminus D_\delta$ .

★ Using easy but tedious computations and Cauchy-Schwarz inequality once more, we have :

$$|r_{04}(s, t)| \leq \sum_{i \geq 1} \sum_{j \geq 1} \frac{\prod_{k=1}^4 \| \frac{f_0^{(k)}(\cdot-t)}{f_0} \|_2^{\alpha_{ijk}}}{\| \frac{f_0(\cdot-t)-f_0}{f_0} \|_2^i}$$

where the sums on  $i$  and  $j$  are finite and where for any  $i$  and  $j : \sum_{k=1}^4 \alpha_{ijk} = i$ . Previous arguments run again and permit us to assert that for any  $\delta > 0$  the function  $(s, t) \mapsto r_{04}(s, t)$  is bounded on  $\mathbb{R}^2 \setminus D_\delta$ .

### 6.5.4 Proof of formula (6.25)

Define the functions  $L$  and  $\tilde{L}$  by  $\log(1 + u) = u - u^2/2 + u^2L(u)$  and  $\tilde{L}(u) = \sup_{|v| < u} |L(v)|$ . It is clear that  $\tilde{L}(u) \rightarrow 0$  as  $u \rightarrow 0$  and we have

$$(6.33) \quad \log \frac{d\mathbb{P}_{\pi_n, t_n}}{d\mathbb{P}_0}(X_1, \dots, X_n) = \sum_{i=1}^n \log(1 + \pi_n(e^{t_n X_i - \frac{t_n^2}{2}} - 1)) \\ = \pi_n \sum_{i=1}^n (e^{t_n X_i - \frac{t_n^2}{2}} - 1) + \frac{\pi_n^2}{2} \sum_{i=1}^n (e^{t_n X_i - \frac{t_n^2}{2}} - 1)^2 + S,$$

with

$$|S| \leq n\pi_n^2 t_n^2 \frac{1}{n} \sum_{i=1}^n \left( \frac{e^{t_n X_i - \frac{t_n^2}{2}} - 1}{t_n} \right)^2 \tilde{L} \left( \pi_n t_n \max_{i=1, \dots, n} \left( \frac{e^{t_n X_i - \frac{t_n^2}{2}} - 1}{t_n} \right) \right).$$

Now it suffices to remark that the random variables  $\left( \frac{e^{t_n X - \frac{t_n^2}{2}} - 1}{t_n} \right)$ ,  $n = 1, 2, \dots$  for  $X$  of distribution  $N(0, 1)$  have bounded third moment. Applying the Markov inequality

$$\max_{i=1, \dots, n} \left( \frac{e^{t_n X_i - \frac{t_n^2}{2}} - 1}{t_n} \right) = o_{\mathbb{P}_0}(\sqrt{n}).$$

Moreover, the class of functions

$$x \rightarrow \left( \frac{e^{t_n x - \frac{t_n^2}{2}} - 1}{t_n} \right)^2$$

is Glivenko-Cantelli in probability (indeed, it is the square of a Donsker class, as a consequence of Section 6.2.1), we get  $S = o_{\mathbb{P}_0}(1)$  and

$$\frac{1}{n} \sum_{i=1}^n [(e^{t_n X_i - \frac{t_n^2}{2}} - 1)^2 - (e^{t_n^2} - 1)] = o_{\mathbb{P}_0}(1),$$

so that

$$(6.34) \quad \log \frac{d\mathbb{P}_{\pi_n, t_n}}{d\mathbb{P}_0}(X_1, \dots, X_n) = \sqrt{n}\pi_n \sqrt{e^{t_n^2} - 1} Y_n(t_n) + n \frac{\pi_n^2}{2} \frac{1}{n} \sum_{i=1}^n (e^{t_n X_i - \frac{t_n^2}{2}} - 1)^2 + o_{\mathbb{P}_0}(1) \\ = \sqrt{n}\pi_n \sqrt{e^{t_n^2} - 1} Y_n(t_n) + n \frac{\pi_n^2}{2} (e^{t_n^2} - 1) + o_{\mathbb{P}_0}(1).$$

Now setting  $C(\gamma, t_0) = \gamma$  if  $t_0 = 0$  and  $C(\gamma, t_0) = \gamma \frac{\sqrt{e^{t_0^2} - 1}}{t_0}$  we have

$$\sqrt{n}\pi_n \sqrt{e^{t_n^2} - 1} = C(\gamma, t_0) + o(1)$$

and

$$\log \frac{d\mathbb{P}_{\pi_n, t_n}}{d\mathbb{P}_0}(X_1, \dots, X_n) = C(\gamma, t_0)Y_n(t_n) - \frac{C(\gamma, t_0)^2}{2} + o_{\mathbb{P}_0}(1).$$

### 6.5.5 Proof of Lemma 6.4.1

Beforehand we set  $c_n = (\log \circ \log n)^{\frac{1}{2}}$  and we recall that  $A_{2,n} = [\alpha_n, \beta_n]$  with  $\alpha_n = 2\sqrt{\log \circ \log \circ \log n}$  and  $\beta_n = \sqrt{\log n/2} - 2\sqrt{\log \circ \log n}$ .

According to (6.24) and (6.26) we need to prove that  $\sup_{t \in A_{2,n}} (Y_0(t) - c_n)$  and  $Y_0(t_0)$  are asymptotically independent. To this end, we consider the discretized process  $\{Y_0(q_n k), k \in \mathbb{Z}\}$  with a step of discretization  $q_n$  depending on  $n$  in a sense which has to be defined. Let us gather the discretized points of  $A_{2,n}$  in  $A_{2,n}^{q_n} = \{d_1, \dots, d_{N(n)}\}$ . By triangular inequalities and simplifications we have for any  $x$  and  $y$

$$\begin{aligned} & \left| \mathbb{P}\left(\sup_{t \in A_{2,n}} Y_0(t) - c_n \leq x; Y_0(t_0) \leq y\right) - \mathbb{P}\left(\sup_{t \in A_{2,n}} Y_0(t) - c_n \leq x\right)\mathbb{P}(Y_0(t_0) \leq y) \right| \\ (6.35) \quad & \leq 2\mathbb{P}\left(\sup_{d \in A_{2,n}^{q_n}} Y_0(d) - c_n \leq x; \sup_{t \in A_{2,n}} Y_0(t) - c_n > x\right) + \\ & \left| \mathbb{P}\left(\sup_{d \in A_{2,n}^{q_n}} Y_0(d) - c_n \leq x; Y_0(t_0) \leq y\right) - \mathbb{P}\left(\sup_{d \in A_{2,n}^{q_n}} Y_0(d) - c_n \leq x\right)\mathbb{P}(Y_0(t_0) \leq y) \right| \end{aligned}$$

The task is now to prove that for fixed  $x$  and  $y$  each component of the upper bound converges to 0.

★ We define the following modification of the function  $r$

$$\begin{aligned} \tilde{r}(t_0, t) &= 0 & t \in A_{2,n}^{q_n}, t \neq t_0, \\ \tilde{r}(s, t) &= r(s, t) & s, t \in A_{2,n}^{q_n}. \end{aligned}$$

Note that under the Gaussian distribution defined by  $\tilde{r}$  the value of the process at  $t_0$  is independent of the values of the process at other locations whose distribution does not changes. This proves that  $\tilde{r}$  is a covariance function. We define  $\xi(t) = \sup_{u, |u-t_0|>t} |r(u, t_0)|$ . From (6.23) we have

$$\xi(t) = O\left(\exp\left(-\frac{t^2}{2}\right)\right).$$

We restrict our attention to  $n$ 's such that

$$c_n > 2|x| \quad ; \quad \xi(\alpha_n) < 1/2 \text{ so that } \frac{(x + c_n)^2}{2(1 + \xi(\alpha_n))} \geq \frac{c_n^2}{12}$$

The Normal Comparison Lemma (Li and Shao, 2002, Theorem 2.1) gives bounds to terms of the type

$$\mathbb{P}\left(Y_1 \leq u_1, \dots, Y_n \leq u_n\right) - \mathbb{P}\left(\tilde{Y}_1 \leq u_1, \dots, \tilde{Y}_n \leq u_n\right)$$

where  $Y$  and  $\tilde{Y}$  are two centered Gaussian vectors with the same variance and possibly different covariances  $\rho_{ij}$  and  $\tilde{\rho}_{ij}$ ,  $i, j = 1, \dots, n$ . It says that

$$(6.36) \quad \mathbb{P}\left(Y_1 \leq u_1, \dots, Y_n \leq u_n\right) - \mathbb{P}\left(\tilde{Y}_1 \leq u_1, \dots, \tilde{Y}_n \leq u_n\right) \\ \leq \frac{1}{2\pi} \sum_{1 \leq i < j \leq n} \left(\arcsin(\rho_{ij}) - \arcsin(\tilde{\rho}_{ij})\right)^+ \exp\left(-\frac{u_i^2 + u_j^2}{2(1 + \bar{\rho}_{ij})}\right)$$

where  $z^+ = \max\{z, 0\}$ ,  $\bar{\rho}_{ij} = \max\{|\rho_{ij}|, |\tilde{\rho}_{ij}|\}$ . Let  $(Const)$  represents a generic positive constant. Since  $\arcsin(x) \leq x\pi/2$  for  $0 \leq x \leq 1$ , applying inequality (6.36) in both directions to the vector  $Y_0$  with covariance  $r$  and to the vector  $\tilde{Y}_0$  with covariance  $\tilde{r}$  corresponding to the points belonging to  $\{t_0\} \cup A_{2,n}^{q_n}$  we get :

$$\left| \mathbb{P}\left(\sup_{d \in A_{2,n}^{q_n}} Y_0(d) - c_n \leq x; Y_0(t_0) \leq y\right) - \mathbb{P}\left(\sup_{d \in A_{2,n}^{q_n}} Y_0(d) - c_n \leq x\right) \mathbb{P}\left(Y_0(t_0) \leq y\right) \right| \\ \leq (Const) \sum_{d \in A_{2,n}^{q_n}} |r(d, t_0)| \exp\left(-\frac{(x + c_n)^2 + y^2}{2(1 + |r(d, t_0)|)}\right) \\ \leq (Const) \sum_{d \in A_{2,n}^{q_n}} |r(d, t_0)| \exp\left(-\frac{c_n^2}{12}\right) \\ \leq \frac{(Const)}{q_n} \exp\left(-\frac{c_n^2}{12}\right) \int_{\alpha_n - q_n}^{+\infty} \xi(t) dt = \frac{(Const)}{q_n} \exp\left(-\frac{c_n^2}{12}\right).$$

which tends to zero if , for example,  $q_n = (\log \circ \log n)^{-\theta}$  as soon as  $\theta > 0$ .

★ To deal with the first term of (6.35), we denote by  $U_z$  and  $U_z^{q_n}$  the point processes of up-crossings of level  $z$  for  $Y_0$  and its  $q_n$ -polygonal approximation (linear interpolation) respectively. For any subset  $B$  of  $\mathbb{R}$ ,

$$U_z(B) = \#\{t \in B, Y_0(t) = z, Y_0'(t) > 0\}, \\ U_z^{q_n}(B) = \#\left\{l \in \mathbb{Z}, q_n(l-1) \in B, q_n l \in B, Y_0(q_n(l-1)) < z < Y_0(q_n l)\right\}.$$

Set  $\Phi$  the distribution function of the standard Gaussian and  $\bar{\Phi} = 1 - \Phi$ .

$$\mathbb{P}\left(\sup_{d \in A_{2,n}^{q_n}} Y_0(d) - c_n \leq x; \sup_{t \in A_{2,n}} Y_0(t) - c_n > x\right) \\ \leq \mathbb{P}\left(Y_0(\alpha_n) > x + c_n\right) + \mathbb{P}\left(Y_0(\alpha_n) \leq x + c_n, U_{x+c_n}(A_{2,n}) \geq 1, U_{x+c_n}^{q_n}(A_{2,n}) = 0\right) \\ \leq \bar{\Phi}(x + c_n) + \mathbb{E}\left(U_{x+c_n}(A_{2,n}) - U_{x+c_n}^{q_n}(A_{2,n})\right)$$

where the last upper bound is due to Markov inequality. The first term above tends trivially to zero, as for the second if we set  $q_n = (\log \circ \log n)^{-\theta}$  with  $\theta > \frac{1}{2}$ , Condition (U7) of Lemma 2 of Azaïs and Mercadier (2003) is met. It is easy to check that since  $\mathbb{E}\left(U_{x+c_n}(A_{2,n})\right)$  is bounded we are in the condition of application of that lemma and

$$\mathbb{E}\left(U_{x+c_n}(A_{2,n}) - U_{x+c_n}^{q_n}(A_{2,n})\right) = o(1).$$



## Bibliography of Chapter 6

- Azaïs, J.-M., Gassiat, É., and Mercadier, C.** (2004). Asymptotic distribution and power of the likelihood ratio test for mixtures : bounded and unbounded cases. *Submitted*.
- Azaïs, J.-M. and Mercadier, C.** (2003). Asymptotic poisson character of extremes in non-stationary gaussian models. *Extremes*, 6 :301–318.
- Bickel, P. and Chernoff, H.** (1993). Asymptotic distribution of the likelihood ratio statistic in a prototypical non regular problem. *Statistics and Probability : A Raghu Raj Bahadur Festschrift. Ghosh et al. Eds. Wiley Eastern Limited, Publishers. 83-96*.
- Boistard, H.** (2003). Test of goodness-of-fits for mixture of population. Technical report, Toulouse III, France, Université Toulouse III, France.
- Chen, H. and Chen, J.** (2001). Large sample distribution of the likelihood ratio test for normal mixtures. *Statist. Probab. Lett.*, 52(2) :125–133.
- Chernoff, H. and Lander, E.** (1995). Asymptotic distribution of the likelihood ratio test that a mixture of two binomials is a single binomial. *J. Statist. Plann. Inference*, 43(1-2) :19–40.
- Dacunha-Castelle, D. and Gassiat, É.** (1997). Testing in locally conic models, and application to mixture models. *ESAIM Probab. Statist.*, 1 :285–317.
- Dacunha-Castelle, D. and Gassiat, É.** (1999). Testing the order of a model using locally conic parametrization : population mixtures and stationary ARMA processes. *Ann. Statist.*, 27(4) :1178–1209.
- Delmas, C.** (2003). On likelihood ratio tests in Gaussian mixture models. *Sankhyā*, 65(3) :513–531.
- Garel, B.** (2001). Likelihood ratio test for univariate Gaussian mixture. *J. Statist. Plann. Inference*, 96(2) :325–350.
- Gassiat, É.** (2002). Likelihood ratio inequalities with applications to various mixtures. *Ann. Inst. H. Poincaré Probab. Statist.*, 6 :897–906.
- Ghosh, J. and Sen, P.** (1985). On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983)*, Wadsworth Statist./Probab. Ser., pages 789–806, Belmont, CA. Wadsworth.
- Goffinet, B., Loisel, P., and Laurent, B.** (1992). Testing in normal mixture models when the proportions are known. *Biometrika*, 79(4) :842–846.
- Hall, P. and Stewart, M.** (2004). Theoretical analysis of power in a two-component normal mixture model. Private communication.

- Hartigan, J. A.** (1985). A failure of likelihood asymptotics for normal mixtures. In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983)*, Wadsworth Statist./Probab. Ser., pages 807–810, Belmont, CA. Wadsworth.
- James, L. F., Priebe, C. E., and Marchette, D. J.** (2001). Consistent estimation of mixture complexity. *Ann. Statist.*, 29(5) :1281–1296.
- Lemdani, M. and Pons, O.** (1997). Likelihood ratio tests for genetic linkage. *Statist. Probab. Lett.*, 33(1) :15–22.
- Lemdani, M. and Pons, O.** (1999). Likelihood ratio tests in contamination models. *Bernoulli*, 5(4) :705–719.
- Li, W. V. and Shao, Q.-M.** (2002). A normal comparison inequality and its applications. *Probab. Theory Related Fields*, 122(4) :494–508.
- Lindsay, B. G.** (1995). *Mixture Models : Theory, Geometry and Applications*. NSF-CBMS Series. Institute of Mathematical Statistics, Hayward.
- Liu, X. and Shao, Y.** (2000). Asymptotic distributions of the likelihood ratio test in two-component normal mixture models. *Ann. Statist.*
- McLachlan, G. and Peel, D.** (2000). *Finite mixture models*. Wiley Series in Probability and Statistics : Applied Probability and Statistics. Wiley-Interscience, New York.
- Mercadier, C.** (2005). Numerical bounds for the distribution of one- and two-parameter processes. *Submitted*. Software <http://www.lsp.ups-tlse.fr/Fp/Mercadier/MAGP.html>.
- Mosler, K. and Seidel, W.** (2001). Testing for homogeneity in an exponential mixture model. *Aust. N. Z. J. Stat.*, 43(2) :231–247.
- Seidel, W., Mosler, K., and Alker, M.** (2000). A cautionary note on likelihood ratio tests in mixture models. *Ann. Inst. Statist. Math.*, 52(3) :481–487.
- Shorack, G. R. and Wellner, J. A.** (1986). *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics : Probability and Mathematical Statistics. John Wiley & Sons Inc., New York.
- Titterton, D. M., Smith, A. F. M., and Makov, U. E.** (1985). *Statistical analysis of finite mixture distributions*. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics. John Wiley & Sons Ltd., Chichester.
- van der Vaart, A. W.** (1998). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.



# Chapitre 7

---

## TEST D'UNE HYPOTHÈSE SUR LA MATRICE DE COVARIANCE

Les statisticiens analysent fréquemment des tableaux de données de grandes dimensions en regardant les matrices de covariance et leurs décompositions spectrales pour en déduire les composantes principales. L'objectif de ce chapitre est de mettre en relation ces principes avec la théorie des matrices aléatoires. Nous expliquons notre intérêt pour le nombre de condition.<sup>1</sup>

### 7.1 Introduction

La réduction par analyse en composantes principales de certains jeux de données conduit à l'étude spectrale des matrices de Wishart.

#### 7.1.1 Analyse en composantes principales

L'analyse en composantes principales est utilisée dans de nombreux champs d'application mais particulièrement en psychologie et autres sciences humaines ou sociales. Un observateur dispose fréquemment de nombreuses données associées à de nombreuses variables. C'est dans le but de réduire ce dernier nombre de paramètres qu'une méthode statistique comme l'analyse en composantes principales s'utilise. Dans cette analyse les axes du repère, dans lequel est représenté le nuage de points, sont modifiés par rotation. Les nouveaux axes sont choisis suivant une méthode de maximisation de la variance correspondant à une transformation orthogonale particulière des variables. Supposons que nous disposons d'un  $n$ -échantillon d'une population et de  $m$  mesures pour chaque individu.

L'analyse en composantes principales s'effectue à l'aide de l'analyse spectrale de la matrice de corrélation des données. Nous évoquons ici sa version non-standardisée basée sur l'analyse de la matrice de covariance notée  $S$ . Cette matrice étant symétrique, elle se décompose comme

$$S = PDP^T$$

---

<sup>1</sup>Cette étude préliminaire a été réalisée aux côtés du Professeur Mario Wschebor dans le laboratoire de Mathématiques de la faculté des Sciences de l'Université de la République de Montevideo (Uruguay). Ce séjour a été financé par une bourse ATUPS accordée par l'Université Paul Sabatier.

avec

$$\begin{aligned} P &= (p_1 | \dots | p_m), \text{ une matrice orthogonale} \\ D &= \text{diag}(\lambda_1, \dots, \lambda_m), \\ \lambda_1 &\geq \lambda_2 \geq \dots \geq \lambda_m \geq 0. \end{aligned}$$

En définissant le vecteur aléatoire des  $m$  caractères observés par  $V^T = (V_1, \dots, V_m)$ , la  $k$ ème composante principale est la combinaison linéaire

$$c_k = p_k^T V$$

qui vérifie  $\text{Var}(c_k) = \lambda_k$  et  $\text{Cov}(c_k, c_l) = 0$  dès que  $k \neq l$ . De plus, la projection de  $V$  dans un sous-espace de dimension  $k$  (au sens d'une minimisation de la variance de l'erreur) est donnée par

$$p_{[1,k]}^T V$$

où  $p_{[1,k]} = (p_1 | \dots | p_k)$  est la matrice de taille  $m \times k$  dont les vecteurs colonnes sont les vecteurs propres associés aux  $k$  plus grandes valeurs propres.

Nous supposons maintenant que nous disposons d'une matrice de données dans  $M_{n,m}(\mathbb{R})$  notée

$$A = (a_{ij})_{i=1, \dots, n; j=1, \dots, m}.$$

Par analogie avec ce qui précède, nous supposons que les lignes de  $A$  sont indépendantes. Cependant nous faisons une hypothèse sur le vecteur  $V$  regroupant les  $m$  caractères observés. Nous supposons qu'il est de loi normale centrée  $N(0, \Sigma)$  où  $\Sigma$  est une matrice de  $M_m(\mathbb{R})$ . Nous pourrions donc désirer connaître la structure de ces données au sens de l'analyse en composantes principales issue de la matrice de covariance empirique  $S = n^{-1}B$  avec

$$B = A^T A.$$

Lorsque la valeur de  $m$  est fixée, la structure spectrale de  $n^{-1}B$  est une bonne approximation de celle de  $\Sigma$ . Et les résultats associés à ce contexte sont rassemblés dans les références classiques comme Anderson (1963) ou Muirhead (1982). Nous nous intéressons au cas où la valeur de  $m$  est comparable à celle de  $n$ . C'est d'ailleurs le cas dans beaucoup de données actuelles.

Les matrices construites de la même manière que  $B$  sont en fait des matrices qui suscitent un large intérêt théorique.

### 7.1.2 Matrice de Wishart

Soit  $A$  une matrice de  $M_{n,m}(\mathbb{R})$  dont les lignes sont indépendantes et telle que la  $i$ ème ligne soit de loi normale  $\mathcal{N}_m(0, \Sigma)$ . La matrice symétrique de  $M_m(\mathbb{R})$

$$B = A^T A$$

a une loi appelée distribution de Wishart de dimension  $m$  avec  $n$  degrés de liberté, de matrice de covariance  $\Sigma$ . La distribution de  $B$  est notée

$$W_m(n, \Sigma).$$

Notons les  $m$  valeurs propres de  $B$  de la manière suivante

$$\lambda_{\max} = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{m-1} \geq \lambda_m = \lambda_{\min} \geq 0.$$

Différents résultats sur la distribution asymptotique des valeurs propres de matrices de Wishart ont été obtenus. Nous proposons de les décrire en se restreignant au cadre suivant

$$\left\{ \begin{array}{l} \Sigma = I_m, \text{ la matrice identité dans } M_m(\mathbb{R}) \\ \frac{m}{n} \xrightarrow{n \rightarrow +\infty} \gamma \in ]0, 1] \end{array} \right.$$

Nous commençons par énoncer des résultats de convergence presque sûre.

**Théorème 7.1.1 (Marčenko et Pastur; Wachter)**

Définissons la distribution empirique  $F_m$  des valeurs propres de  $n^{-1}B$  par

$$F_m(x) = \frac{1}{m} \sum_{k=1}^m \mathbb{1}_{n^{-1}\lambda_k \leq x}.$$

$F_m(x)$  converge presque sûrement vers  $F_\gamma(x)$  où

$$F'_\gamma(x) = \frac{1}{2\pi x \gamma} \sqrt{(d-x)(x-c)} \mathbb{1}_{\{c \leq x \leq d\}}$$

pour  $c = (1 - \sqrt{\gamma})^2$  et  $d = (1 + \sqrt{\gamma})^2$ .

Les deux résultats suivants prouvent que les bornes du support sont atteintes.

**Théorème 7.1.2 (Geman, Silverstein)**

- a)  $n^{-1}\lambda_{\max} \xrightarrow[n \rightarrow +\infty]{ps} (1 + \sqrt{\gamma})^2,$
- b)  $n^{-1}\lambda_{\min} \xrightarrow[n \rightarrow +\infty]{ps} (1 - \sqrt{\gamma})^2.$

Après les travaux de Johansson (2000) dans le cas complexe, Johnstone (2001) obtient les fluctuations de la plus grande valeur propre.

**Théorème 7.1.3 (Johnstone)**

$\frac{\lambda_{\max} - \beta_{nm}}{\gamma_{nm}}$  converge en distribution vers la loi de Tracy-Widom notée  $F_1$  avec

$$\begin{aligned} \beta_{nm} &= (\sqrt{n-1} + \sqrt{m})^2 \\ \gamma_{nm} &= (\sqrt{n-1} + \sqrt{m})(1/\sqrt{n-1} + 1/\sqrt{m})^{1/3}. \end{aligned}$$

Ce résultat a d'ailleurs été étendu par Soshnikov (2002) pour la distribution jointe des  $k$  plus grandes valeurs propres.

Les fluctuations de la plus petite valeur propre ne sont pas établies (dans le cas réel, comme ici). Cependant, nous disposons dans plusieurs cas d'une expression explicite

de sa densité notée  $f_{\lambda_{\min}}$ . Nous introduisons, lorsque  $a > 0$  et  $b < 1$ , la fonction de Tricomi  $U(a, b, z)$  qui est l'unique solution de l'équation de Kummer

$$zu''(z) + (b - z)u'(z) - au(z) = 0$$

satisfaisant  $U(a, b, 0) = \Gamma(1 - b)/\Gamma(1 + a - b)$  et  $U(a, b, \infty) = 0$ .

De plus, nous notons  $L_m^{(\alpha)}(-x) = \sum_{k=0}^m C_{m+\alpha}^{m-k} x^k$  le polynôme généralisé de Laguerre.

### Théorème 7.1.4 (Edelman)

– lorsque  $n = m$ ,

$$f_{\lambda_{\min}}(\lambda) = \frac{m}{\sqrt{2\pi}} \Gamma\left(\frac{m+1}{2}\right) \lambda^{-1/2} e^{-\lambda m/2} U\left(\frac{m-1}{2}, -\frac{1}{2}, \frac{\lambda}{2}\right),$$

$$f_{m\lambda_{\min}}(x) \xrightarrow{m \rightarrow +\infty} \frac{1 + \sqrt{x}}{2\sqrt{x}} e^{-(x/2 + \sqrt{x})},$$

– lorsque  $n = m + 1$ ,  $m\lambda_{\min}$  suit une loi exponentielle

$$f_{\lambda_{\min}}(\lambda) = \frac{m}{2} e^{-\lambda m/2},$$

– lorsque  $n = m + 2$ ,

$$f_{\lambda_{\min}}(\lambda) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{2\pi}} \left[ L_{m-1}^{(2)}(-\lambda) U\left(\frac{m-1}{2}, -\frac{1}{2}, \frac{\lambda}{2}\right) + \frac{\lambda}{2} L_{m-2}^{(3)}(-\lambda) U\left(\frac{m+1}{2}, \frac{1}{2}, \frac{\lambda}{2}\right) \right],$$

– lorsque  $n = m + 3$ ,

$$f_{\lambda_{\min}}(\lambda) = \frac{1}{2(m+1)} L_{m-1}^{(3)}(-\lambda).$$

## 7.2 Deux tests pour la matrice de covariance

Dans la littérature, nous trouvons plusieurs tests d'hypothèses sur les matrices de covariance. Nous nous sommes intéressés à deux tests qui s'écrivent de la manière suivante,

$$\begin{aligned} \text{TEST 1} \quad (H_0) : \Sigma = I_m \quad \text{contre} \quad (H_1) : \Sigma \neq I_m, \\ \text{TEST 2} \quad (H_0) : \Sigma = \theta I_m \quad \text{contre} \quad (H_1) : \Sigma \neq \theta I_m. \end{aligned}$$

Ces deux tests permettent en fait de traiter des hypothèses plus générales à savoir pour le premier  $(H_0) : \Sigma = \Sigma_0$  et  $(H_0) : \Sigma = \theta \Sigma_0$  pour le second. Il suffit pour cela de renormaliser correctement les données pour présenter un test sous la forme précédente.

### 7.2.1 Test du "Spiked model"

Ce premier modèle est un cas particulier du test 1 puisqu'il considère la matrice de covariance particulière suivante

$$\Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r, 1, \dots, 1)$$

où

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 1.$$

La matrice  $B$  qui est associée à ce modèle suit par conséquent une loi de Wishart  $W_m(n, \Sigma_r)$ .

Johnstone (2001) démontre que pour  $r < m$ , la  $(r + 1)$ ième plus grande valeur propre d'une  $W_m(n, \Sigma_r)$  est stochastiquement plus petite que la plus grande valeur propre d'une  $W_{m-r}(n, I_{m-r})$ .

Il propose alors d'utiliser le résultat sur les fluctuations de la plus grande valeur propre, rappelée dans la section précédente (théorème 7.1.3), pour construire un test basé sur la statistique  $\lambda_{r+1}$ . La comparaison stochastique implique que ce test est conservatif. Il faut cependant noter que Johnstone donne une idée "empirique" sur l'écart des lois de la  $(r + 1)$ ième plus grande valeur propre d'une  $W_m(n, \Sigma_r)$  et de la plus grande valeur propre d'une  $W_{m-r}(n, I_{m-r})$ .

Depuis ce travail, la connaissance autour du "Spiked model" s'est élargie.

**Théorème 7.2.1 (Paul)**

Si  $k \leq r$  et  $\sigma_k \leq 1 + \sqrt{\gamma}$ , nous avons

$$n^{-1}\lambda_k \xrightarrow[n \rightarrow +\infty]{ps} (1 + \sqrt{\gamma})^2,$$

et si  $\sigma_k > 1 + \sqrt{\gamma}$ , nous avons

$$n^{-1}\lambda_k \xrightarrow[n \rightarrow +\infty]{ps} \sigma_k \left(1 + \frac{\gamma}{\sigma_k - 1}\right).$$

Il existe là encore un résultat sur les fluctuations de la  $k$ ième plus grande valeur propre.

**Théorème 7.2.2 (Paul)**

Supposons les hypothèses suivantes :

- (i)  $\sigma_k > 1 + \sqrt{\gamma}$ ,
- (ii)  $\sigma_k$  est de multiplicité 1,
- (iii)  $m/n - \gamma = o(n^{-1/2})$ .

Alors, nous avons

$$\sqrt{n} \left[ n^{-1}\lambda_k - \sigma_k \left(1 + \frac{\gamma}{\sigma_k - 1}\right) \right] \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, V^2(\sigma_k))$$

avec  $V^2(\sigma_k) = 2\sigma_k^2 \left(1 - \frac{\gamma}{(\sigma_k - 1)^2}\right)$ .

Il existe enfin des résultats sur la convergence de l'angle formé entre les vecteurs propres associés à  $\sigma_k$  et  $\lambda_k$  respectivement. Nous référons à Johnstone et Lu (2004) et Paul (2004) pour de plus amples détails.

Nous n'avons pas trouvé d'étude s'appuyant sur ces dernières avancées pour la construction d'un test. Cependant, il faut savoir que la consistance de la statistique de test

$$V = \frac{1}{m} \text{tr}[(n^{-1}B - I)^2]$$

(et d'une fine modification de celle-ci) est étudiée dans Ledoit et Wolf (2002).



### 7.2.2 Test de Sphéricité

Revenons maintenant sur un test du type 2, qui constitue la motivation de ce chapitre. Le cas  $\Sigma = \theta I_m$  correspond à un test de sphéricité. Pour établir un lien avec l'analyse en composantes principales, cette hypothèse revient à affirmer qu'il n'y a pas de direction privilégiée dans la structure des données recueillies dans  $A$ . Autrement dit, les  $m$  caractères étudiés sont, comparativement les uns aux autres, tous autant explicatifs.

Les auteurs Ledoit et Wolf (2002) s'intéressent dans ce cas à la consistance du test basé sur la statistique

$$U = \frac{1}{m} \text{tr} \left[ \left( \frac{n^{-1}B}{\text{tr}(n^{-1}B)/m} - I_m \right)^2 \right].$$

Nous avons souhaité considérer une statistique dont l'interprétation théorique mais aussi géométrique est claire. Beaucoup de problèmes d'inférence en statistique ont des propriétés inhérentes de symétrie ou d'invariance et imposent naturellement des restrictions sur les procédures envisageables. Sans rentrer dans les détails de cette notion, nous nous sommes intéressés aux statistiques de test invariantes sous l'action d'un groupe de transformations orthogonales. Plus précisément, nous recherchons une fonction qui attribue une valeur identique aux éléments  $\Sigma$  et  $a^2 H \Sigma H^T$  lorsque  $a$  est un réel non nul et  $H$  est une matrice orthogonale de taille  $m$  puisque le test de sphéricité reste invariant sous cette transformation. Nous pouvons démontrer que cette application est obligatoirement une fonction du vecteur

$$\left( \frac{\sigma_1}{\sigma_m}, \frac{\sigma_2}{\sigma_m}, \dots, \frac{\sigma_{m-1}}{\sigma_m} \right)$$

où  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_m > 0$  sont les valeurs propres de  $\Sigma$ . Il suit que toute statistique invariante pour le test de sphéricité est construite à l'aide des rapports

$$\left( \frac{\lambda_{\max}}{\lambda_{\min}}, \frac{\lambda_2}{\lambda_{\min}}, \dots, \frac{\lambda_{m-1}}{\lambda_{\min}} \right).$$

Pour plus de détails sur cette affirmation, nous conseillons de lire par exemple Muirhead (1982, pages 196 et 333). La première variable aléatoire de ce vecteur semble être intuitivement très bien appropriée pour l'étude de la sphéricité. Dans le cas où  $m$  est fixé, quelques résultats sont donnés par Sugiyama (1970), Waikar et Schuurmann (1973) et Krishnaiah et Schuurmann (1974).

## 7.3 Distribution du nombre de condition

Supposons pour commencer que

$$n = m.$$

Définissons la norme de la matrice  $A$  comme

$$\|A\| = \sup_{\|s\|=1} \|As\|,$$

avec la norme euclidienne sur  $\mathbb{R}^n$ . Lorsque  $A$  n'est pas singulière (ce que nous supposons par la suite), son nombre de condition est défini par

$$\kappa(A) = \|A\| \|A^{-1}\|.$$

Le rôle de  $\kappa(A)$  est bien établi dans une variété de problèmes d'analyse numérique pour le contrôle de la sensibilité aux perturbations.

Définissons maintenant le polynôme quadratique  $X : S^{m-1} \rightarrow \mathbb{R}$  par

$$X(s) = s^T B s.$$

Tout vecteur propre  $s_\lambda$  normalisé associé à la valeur propre  $\lambda$  de  $B$  satisfait  $X(s_\lambda) = \lambda$ .

Il suit que

- $\lambda_{\max} = \max_{s \in S^{m-1}} X(s) = \|A\|^2,$
- $\lambda_{\min} = \min_{s \in S^{m-1}} X(s) = 1/\|A^{-1}\|^2,$
- $\kappa(A) = (\lambda_{\max}/\lambda_{\min})^{1/2}.$

Lorsque  $\lambda_{\min} = 0$ , on prend pour convention  $\kappa(A) = +\infty$ .

Par conséquent, notre statistique d'intérêt pour la construction d'un test de sphéricité est le carré de  $\kappa(A)$ . Lorsque  $n$  et  $m$  ne sont pas égaux, nous définissons le nombre de condition de  $A$  à travers l'égalité qu'il vérifie dans le cas carré

$$\kappa(A) = \left( \frac{\lambda_{\max}}{\lambda_{\min}} \right)^{1/2}.$$

Malheureusement, peu de résultats autour de la distribution de  $\kappa(A)$  sont connus.

En conséquence du théorème 7.1.2, nous obtenons

$$(7.1) \quad \kappa(A) \xrightarrow[n \rightarrow +\infty]{ps} \frac{1 + \sqrt{\gamma}}{1 - \sqrt{\gamma}}.$$

En utilisant l'expression de la densité de  $f_{\lambda_{\min}}$  et la première partie du théorème 7.1.2, Edelman (1989) montre que si  $n = m$  la densité de  $\kappa(A)$  vérifie

$$(7.2) \quad f_{\kappa(A)/m}(x) \xrightarrow[m \rightarrow +\infty]{} \frac{2x + 4}{x^3} e^{-2(1/x + 1/x^2)}.$$

Logiquement, nous avons souhaité calculer la densité de  $\lambda_{\min}$  pour le cas rectangulaire général.

### Lemme 7.3.1

*L'expression de la valeur de la densité  $f_{\lambda_{\min}}$  au point  $\lambda$  est reliée à la solution de l'équation différentielle, d'ordre au plus 2, avec un second membre ne dépendant que  $n - m$  et de la variable  $\lambda$ .*

*Le second membre est linéaire dans les cas  $n = m$ ,  $n = m + 1$  et  $n = m + 3$  et non linéaire dans les autres cas.*

L'utilisation de cette méthode semble donc limitée à ces trois cas. La preuve de ce résultat n'est qu'une extension de la démonstration d'Edelman dans le cas  $n = m$ . Elle est reportée à la section 7.4.

Dans le cas carré, nous connaissons asymptotiquement la queue de distribution du nombre de solution. Sachant que son support est  $\mathbb{R}_+$  nous avons, en intégrant la densité limite de  $\kappa(A)/m$  obtenue par Edelman,

$$\lim_{m \rightarrow +\infty} \mathbb{P}(\kappa(A) > mx) = 1 - e^{-2(1/x+1/x^2)}.$$

Il existe cependant un résultat non asymptotique sur la queue de distribution de  $\kappa(A)$ .

**Théorème 7.3.2 (Azaïs et Wschebor)**

Si  $n = m \geq 3$ , il existe des constantes universelles positives  $c$  et  $C$  telles que pour tout  $x > 1$ ,

$$\frac{c}{x} < \mathbb{P}(\kappa(A) > mx) < \frac{C}{x}.$$

De plus nous savons que  $c \geq 0.140$  et  $C \leq 5.55$ .

Leur démonstration s'appuie sur la méthode de Rice puisque les valeurs propres  $\lambda_{\max}$  et  $\lambda_{\min}$  sont vues comme les valeurs extrêmes d'un polynôme quadratique.

Une extension dans le cas décentré est donnée par Wschebor (2004) :  $A$  est une matrice réelle aux entrées indépendantes qui suivent une loi gaussienne de même variance notée  $\sigma$ . Notons  $M$  la matrice des espérances des entrées de  $A$ .

**Théorème 7.3.3 (Wschebor)**

Si  $n = m \geq 3$ , nous avons pour tout  $x > 0$

$$\mathbb{P}(\kappa(A) > mx) < \frac{1}{x} \left( \frac{1}{4\sqrt{2\pi m}} + C(M, \sigma, m) \right)$$

où  $C(M, \sigma, m) = 7 \left( 5 + \frac{4\|M\|^2(1 + \log(m))}{\sigma m} \right)^{\frac{1}{2}}$ .

En particulier lorsque  $0 < \sigma \leq 1$  et  $\|M\| \leq 1$ , il existe une constante universelle  $K$  telle que pour tout  $x > 0$ ,

$$\mathbb{P}(\kappa(A) > mx) < \frac{K}{\sigma^{1/2}x}.$$

Il existe d'autre part un certain nombre de résultats sur le logarithme du nombre de condition qui utilisent une approche semblable (Cuesta-Albertos et Wschebor, 2003; Cucker et Wschebor, 2003; Cuesta-Albertos et Wschebor, 2004; Wschebor, 2004).

Voici une description de la méthode utilisée pour démontrer le théorème 7.3.2.

- BORNE SUPÉRIEURE. Pour tout  $x > 1$ , nous avons

$$\mathbb{P}(\kappa(A) > x) = \mathbb{P}\left(\frac{\lambda_{\max}}{\lambda_{\min}} > x^2\right) \leq \mathbb{P}\left(\lambda_{\min} < \frac{L^2 m}{x^2}\right) + \mathbb{P}\left(\frac{\lambda_{\max}}{\lambda_{\min}} > x^2, \lambda_{\min} \geq \frac{L^2 m}{x^2}\right)$$

où  $L$  est une constante positive quelconque.

- BORNE INFÉRIEURE. Pour tout  $\gamma > 0$  et  $x > 1$ , nous avons

$$\begin{aligned} \mathbb{P}(\kappa(A) > x) &= \mathbb{P}\left(\frac{\lambda_{\max}}{\lambda_{\min}} > x^2\right) \geq \mathbb{P}\left(\frac{\lambda_{\max}}{\lambda_{\min}} > x^2, \lambda_{\min} < \frac{\gamma^2 m}{x^2}\right) \\ &= \mathbb{P}\left(\lambda_{\min} < \frac{\gamma^2 m}{x^2}\right) - \mathbb{P}\left(\frac{\lambda_{\max}}{\lambda_{\min}} \leq x^2, \lambda_{\min} < \frac{\gamma^2 m}{x^2}\right). \end{aligned}$$

Par conséquent, il faut

1. Calculer une majoration de la densité jointe  $f_{\lambda_{\min}, \lambda_{\max}}$  pour majorer les probabilités d'intersection.
2. Calculer une majoration de la loi de  $\lambda_{\min}$  pour conclure dans la partie "borne supérieure".
3. Calculer une minoration de la loi de  $\lambda_{\min}$  pour conclure dans la partie "borne inférieure".

Nous nous sommes alors interrogés sur la possibilité d'extension de ce résultat dans le cas rectangulaire ( $n \neq m$ ) et/ou non nul correspondant à une matrice de covariance

$$\Sigma^* = \text{diag}(\sigma_1, \dots, \sigma_m).$$

Nous présentons les bornes obtenues pour les valeurs propres extrêmes d'une matrice  $B$  de loi de Wishart  $W_m(n, \Sigma^*)$ .

#### Lemme 7.3.4 (Majoration loi jointe)

En notant  $\sigma_+ = \max_{i=1, \dots, m} \sigma_i$ , la densité jointe de  $\lambda_{\min}$  et  $\lambda_{\max}$  vérifie

$$f_{\lambda_{\min}, \lambda_{\max}}(\lambda, \Lambda) \leq \frac{\sqrt{2}}{4(m-2)!(n-m)!} \Lambda^{\frac{n+m-3}{2}} \lambda^{\frac{n-m-1}{2}} \exp\left(-\frac{\lambda + \Lambda}{2\sigma_+}\right).$$

#### Lemme 7.3.5 (Majoration pour $\lambda_{\min}$ )

En notant  $\sigma_+ = \max_{k=1, \dots, m} \sigma_k$  et  $\sigma_- = \min_{k=1, \dots, m} \sigma_k$ , la densité de  $\lambda_{\min}$  est majorée par

$$f_{\lambda_{\min}}(\lambda) \leq \frac{\sqrt{\pi}}{2^{\frac{n+m-1}{2}}} \frac{\Gamma(n)}{\Gamma(n-m+1)\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \frac{\lambda^{\frac{n-m-1}{2}}}{\sigma_-^{n/2}} \exp\left(-\frac{\lambda}{2\sigma_+}\right).$$

La plus grande difficulté ici est de n'avoir aucune idée a priori sur une possible vitesse de convergence, si convergence il y a. L'obtention d'une minoration de la loi de  $\lambda_{\min}$  avec la méthode d'Azaïs et Wschebor est, là encore, fortement dépendante de la vitesse de convergence. Nous avons donc simplifié notre cadre en supposant la matrice  $A$  carrée. Autrement dit, nous regardons la plus petite valeur propre de la matrice  $B$  qui suit une  $W_m(m, \Sigma^*)$ .

**Lemme 7.3.6 (Minoration pour  $\lambda_{\min}$ )**

Lorsque  $0 < \lambda < m^{-1}$  et  $\sigma_- > 1/2$ , nous avons  $\mathbb{P}(\lambda_{\min} < \lambda) \geq C(\sigma_-, \sigma_+) \sqrt{m\lambda}$  avec

$$C(\sigma_-, \sigma_+) = \frac{2^{\frac{3}{2}}}{\sqrt{3\pi}\sigma_+^{n/2}} \left( \frac{2}{\sigma_-} - \frac{2}{3} \right).$$

Les preuves de ces lemmes sont reportées à la section 7.4.

## COMMENTAIRES.

★ Dans les preuves des résultats asymptotiques sur le nombre de condition  $\kappa(A)$ , à savoir (7.1) et (7.2),  $\lambda_{\max}$  est caractérisé par sa convergence presque sûre. Pour atteindre une distribution (non dégénérée) de  $\kappa(A)$  dans le cas où  $m/n$  converge vers  $\gamma \in ]0, 1[$ , il faudrait aussi un résultat du type Tracy-Widom pour  $\lambda_{\min}$ , connu jusque là uniquement dans le cas complexe.

★ La méthode utilisée dans les lemmes 7.3.4, 7.3.5 et 7.3.6 semble tout de même peu puissante pour traiter ce type de recherche. Cependant elle est justifiée par le fait que le cas non-centré considéré par Wschebor (2004) ne connaît pas d'équivalent dans la littérature dite classique.

★ Il apparaît une zone d'analyse statistique intéressante pour la distribution asymptotique de  $\kappa(A)$  lorsque  $\gamma = 1$ . Supposons alors que  $n$  s'écrit de la manière suivante

$$n = m + Cm^\beta$$

où le paramètre  $C$  est strictement positif et  $\beta$  est inférieur à 1. Nous avons différentes vitesses de convergence pour nos variables d'intérêt

–  $C = 0$  ou  $\beta = -\infty$ ,

$$n\lambda_{\min} \xrightarrow{\mathcal{L}} \text{ et } \kappa(A)/n \xrightarrow{\mathcal{L}},$$

–  $C = 1$  et  $\beta = 0$ ,

$$n\lambda_{\min} \xrightarrow{\mathcal{L}} \text{ et } \kappa(A) \xrightarrow{ps},$$

–  $\beta = 1$ ,

$$\lambda_{\min}/n \xrightarrow{ps} \text{ et } \kappa(A) \xrightarrow{ps}.$$

Nous pouvons donc penser qu'il existe une petite plage de paramètres  $\beta$  contenue dans l'intervalle  $]0, 1[$  pour laquelle le comportement asymptotique de ces variables change beaucoup.

★ La construction d'un test basé sur  $\kappa(A)$  nécessite clairement une connaissance plus complète sur la distribution de cette variable. Ce problème est donc une question ouverte.

## 7.4 Preuves des lemmes

### Preuve du lemme 7.3.1

En intégrant la densité jointe des  $m$  valeurs propres de  $B$ , nous savons que  $\lambda_{\min}$  a pour densité

$$(7.3) \quad f_{\lambda_{\min}}(\lambda) = \frac{K_{m,n}}{(m-1)!} \lambda^{\frac{n-m-1}{2}} e^{-\frac{m\lambda}{2}} \int_{\mathbb{R}_+^{m-1}} \prod_{i=1}^{m-1} (x_i + \lambda)^{\frac{n-m-1}{2}} \Delta d\Omega$$

avec

- $K_{m,n}^{-1} = (2^n/\pi)^{m/2} \prod_{i=1}^m \Gamma(\frac{n-i+1}{2}) \Gamma(\frac{m-1+1}{2})$ ,
- $\Delta = \prod_{1 \leq i < j \leq m-1} |x_i - x_j|$ ,
- $d\Omega = d\mu(x_1) \dots d\mu(x_{m-1})$ ,
- $d\mu(x) = x e^{-x/2} dx$ .

Posons

$$\omega(\lambda) = \int_{\mathbb{R}_+^{m-1}} \prod_{i=1}^{m-1} (x_i + \lambda)^{\frac{n-m-1}{2}} \Delta d\Omega.$$

Dans le cas  $n = m + 1$ , il est clair que  $\omega(\lambda)$  ne dépend pas  $\lambda$ .

Nous allons prouver que  $\omega$  vérifie une équation différentielle. Pour cela, nous avons besoin des notations suivantes :

- $\Delta = \delta \Delta_2$ ,
- $\delta = \prod_{i=2}^{m-1} |x_1 - x_i|$ ,
- $\Delta_2 = \prod_{2 \leq i < j \leq m-1} |x_i - x_j|$ ,
- $f_j^{a,b} = x_j^a (x_j + \lambda)^b$ ,
- $g_j = \prod_{i=j}^{m-1} (x_i + \lambda)^{\frac{n-m-1}{2}}$ ,
- $d\Omega_2 = d\mu(x_2) \dots d\mu(x_{m-1})$ ,
- $I(\lambda) = \int_{\mathbb{R}_+^{m-1}} f_1^{0, \frac{n-m-3}{2}} f_2^{0, \frac{n-m-3}{2}} g_3 \Delta d\Omega$ .

DÉRIVÉE PREMIÈRE DE  $\omega$ .

Nous pouvons écrire  $\omega(\lambda) = \int_{\mathbb{R}_+^{m-1}} g_1 \Delta d\Omega$ , de plus

$$g_1'(\lambda) = \sum_{i=1}^{m-1} \frac{n-m-1}{2} (x_i + \lambda)^{\frac{n-m-3}{2}} \prod_{j=1, j \neq i}^{m-1} (x_j + \lambda)^{\frac{n-m-1}{2}}.$$

Par un argument de symétrie, nous obtenons

$$\omega'(\lambda) = (m-1) \frac{n-m-1}{2} \int_{\mathbb{R}_+^{m-1}} (x_1 + \lambda)^{\frac{n-m-3}{2}} \prod_{j=2}^{m-1} (x_j + \lambda)^{\frac{n-m-1}{2}} \Delta d\Omega$$

qui s'exprime, avec nos notations, de la manière suivante

$$(7.4) \quad \omega'(\lambda) = (m-1) \frac{n-m-1}{2} \int_{\mathbb{R}_+^{m-1}} f_1^{0, \frac{n-m-3}{2}} g_2 \Delta d\Omega.$$

Dans le cas  $n = m + 3$ , il est clair que  $\omega'(\lambda)$  ne dépend pas  $\lambda$ .

DÉRIVÉE SECONDE DE  $\omega$ .

Calculons la dérivée de l'intégrand précédent,

$$\begin{aligned} \left( f_1^{0, \frac{n-m-3}{2}} g_2 \right)'(\lambda) &= \frac{n-m-3}{2} (x_1 + \lambda)^{\frac{n-m-5}{2}} g_2(\lambda) \\ &+ \sum_{i=2}^{m-1} \frac{n-m-1}{2} (x_i + \lambda)^{\frac{n-m-3}{2}} \prod_{j=1, j \neq i}^{m-1} (x_j + \lambda)^{\frac{n-m-1}{2}} f_1^{0, \frac{n-m-3}{2}}(\lambda). \end{aligned}$$

De nouveau un argument de symétrie nous conduit à

$$(7.5) \quad \omega''(\lambda) = (m-1) \frac{n-m-1}{2} \left\{ \frac{n-m-3}{2} \int_{\mathbb{R}_+^{m-1}} f_1^{0, \frac{n-m-5}{2}} g_2 \Delta d\Omega + (m-2) \frac{n-m-1}{2} I(\lambda) \right\}.$$

ÉCRITURE DE  $\omega$  EN FONCTION DE  $\omega'$ .

Pour cela il suffit de calculer  $g_1$  en fonction  $f_1$  et  $g_2$

$$g_1(\lambda) = (\lambda + x_1) f_1^{0, \frac{n-m-3}{2}} g_2,$$

donc

$$\omega(\lambda) = \lambda \int_{\mathbb{R}_+^{m-1}} f_1^{0, \frac{n-m-3}{2}} g_2 \Delta d\Omega + \int_{\mathbb{R}_+^{m-1}} f_1^{1, \frac{n-m-3}{2}} g_2 \Delta d\Omega.$$

Autrement dit, nous avons

$$(7.6) \quad \omega(\lambda) = \int_{\mathbb{R}_+^{m-1}} f_1^{1, \frac{n-m-3}{2}} g_2 \Delta d\Omega + \frac{2\lambda\omega'(\lambda)}{(m-1)(n-m-1)}.$$

TRANSFORMATION DE LA PREMIÈRE INTÉGRALE DE (7.6).

• Nous cherchons à écrire différemment son intégrand en isolant la variable  $x_1$

$$\begin{aligned} \int_{\mathbb{R}_+^{m-1}} f_1^{1, \frac{n-m-3}{2}} g_2 \Delta d\Omega &= \int_{\mathbb{R}_+^{m-1}} x_1^2 (x_1 + \lambda)^{\frac{n-m-3}{2}} g_2 e^{-\frac{x_1}{2}} \Delta d\Omega_2 dx_1, \\ &= \int_{\mathbb{R}_+^{m-1}} f_1^{2, \frac{n-m-3}{2}} g_2 \left\{ -2 \frac{\partial}{\partial x_1} (e^{-x_1/2}) \right\} \Delta dx_1 d\Omega_2, \\ &= \int_{\mathbb{R}_+^{m-1}} f_1^{2, \frac{n-m-3}{2}} \delta \left\{ -2 \frac{\partial}{\partial x_1} (e^{-x_1/2}) \right\} g_2 \Delta_2 dx_1 d\Omega_2, \\ (7.7) \quad &= 2 \int_{\mathbb{R}_+^{m-1}} \frac{\partial}{\partial x_1} \left\{ f_1^{2, \frac{n-m-3}{2}} \delta \right\} e^{-x_1/2} g_2 \Delta_2 dx_1 d\Omega_2. \end{aligned}$$

- Calculons la dérivée partielle qui apparaît dans l'intégrand de (7.7). Nous avons

$$\begin{aligned} \frac{\partial}{\partial x_1} \left\{ x_1^2 (x_1 + \lambda)^{\frac{n-m-3}{2}} \prod_{i=2}^{m-1} |x_1 - x_i| \right\} = \\ 2x_1 (x_1 + \lambda)^{\frac{n-m-3}{2}} \prod_{i=2}^{m-1} |x_1 - x_i| + x_1^2 \frac{n-m-3}{2} (x_1 + \lambda)^{\frac{n-m-5}{2}} \prod_{i=2}^{m-1} |x_1 - x_i| \\ + x_1^2 (x_1 + \lambda)^{\frac{n-m-3}{2}} \sum_{i=2}^{m-1} (\mathbb{1}_{x_1 > x_i} - \mathbb{1}_{x_1 < x_i}) \prod_{j=2, j \neq i}^{m-1} |x_1 - x_j|. \end{aligned}$$

- La première intégrale de (7.6) devient

$$(7.8) \quad \int_{\mathbb{R}_+^{m-1}} f_1^{1, \frac{n-m-3}{2}} g_2 \Delta d\Omega = \frac{8\omega'(\lambda)}{(m-1)(n-m-1)} + (n-m-3) \int_{\mathbb{R}_+^{m-1}} f_1^{1, \frac{n-m-5}{2}} g_2 \Delta d\Omega \\ + 2(m-2) \int_{\mathbb{R}_+^{m-1}} f_1^{0, \frac{n-m-3}{2}} \frac{x_1}{x_1 - x_2} g_2 \Delta d\Omega.$$

- En isolant de nouveau la variable  $x_1$ , nous parvenons à faire apparaître  $\omega'$  dans la seconde intégrale de (7.8)

$$(7.9) \quad \int_{\mathbb{R}_+^{m-1}} f_1^{1, \frac{n-m-5}{2}} g_2 \Delta d\Omega = \int_{\mathbb{R}_+^{m-1}} (x_1 + \lambda - \lambda) (x_1 + \lambda)^{\frac{n-m-5}{2}} g_2 \Delta d\Omega \\ = \int_{\mathbb{R}_+^{m-1}} (x_1 + \lambda)^{\frac{n-m-3}{2}} g_2 \Delta d\Omega - \lambda \int_{\mathbb{R}_+^{m-1}} (x_1 + \lambda)^{\frac{n-m-5}{2}} g_2 \Delta d\Omega \\ = \frac{2\omega'(\lambda)}{(m-1)(n-m-1)} - \lambda \int_{\mathbb{R}_+^{m-1}} f_1^{0, \frac{n-m-5}{2}} g_2 \Delta d\Omega.$$

- Regardons enfin la troisième intégrale de (7.8).

$$(7.10) \quad \int_{\mathbb{R}_+^{m-1}} f_1^{0, \frac{n-m-3}{2}} \frac{x_1}{x_1 - x_2} g_2 \Delta d\Omega = \int_{\mathbb{R}_+^{m-1}} \frac{x_1(x_2 + \lambda)}{x_1 - x_2} f_1^{0, \frac{n-m-3}{2}} f_2^{0, \frac{n-m-3}{2}} g_3 \Delta d\Omega \\ = \lambda \int_{\mathbb{R}_+^{m-1}} \frac{x_1}{x_1 - x_2} f_1^{0, \frac{n-m-3}{2}} f_2^{0, \frac{n-m-3}{2}} g_3 \Delta d\Omega \\ = \frac{\lambda}{2} I(\lambda),$$

où nous avons successivement utilisé l'antisymétrie de  $(x_1, x_2) \mapsto \frac{x_1 x_2}{x_1 - x_2}$  et le fait que  $x_1/(x_1 - x_2) + x_2/(x_2 - x_1) = 1$ .



CONCLUSION.

En reprenant (7.6) et en appliquant les égalités (7.8), (7.9) et (7.10), nous avons

$$\begin{aligned}
\omega(\lambda) &= \frac{2\lambda\omega'(\lambda)}{(m-1)(n-m-1)} + \int_{\mathbb{R}_+^{m-1}} f_1^{1, \frac{n-m-3}{2}} g_2 \Delta d\Omega \\
&= \frac{2\lambda\omega'(\lambda)}{(m-1)(n-m-1)} + \frac{8+2(n-m-3)}{(m-1)(n-m-1)}\omega'(\lambda) \\
&\quad - \lambda(n-m-3) \left[ \frac{4\omega''(\lambda)}{(m-1)(n-m-1)(n-m-3)} - \frac{(m-2)(n-m-1)}{n-m-3} I(\lambda) \right] \\
&\quad + \lambda(m-2)I(\lambda) \\
&= \frac{2(\lambda+n-m+1)}{(m-1)(n-m-1)}\omega'(\lambda) - \frac{4\lambda}{(m-1)(n-m-1)}\omega''(\lambda) + \lambda(m-2)(n-m)I(\lambda)
\end{aligned}$$

Posons maintenant  $\omega(\lambda) = u(z)$  avec  $\lambda = 2z$ . L'équation précédente devient

$$(7.11) \quad zu''(z) + (b-z)u'(z) - au(z) = czI(2z),$$

avec

$$b = -(n-m+1)/2,$$

$$a = -(m-1)(n-m-1)/2,$$

$$c = (m-1)(m-2)(n-m)(n-m-1).$$

Le cas  $n = m$  conduit à l'équation différentielle linéaire homogène

$$zu''(z) + (b-z)u'(z) - au(z) = 0.$$

Dans les autres cas le second membre dépend de la valeur de la fonction

$$I(\lambda) = \int_{\mathbb{R}_+^{m-1}} [(x_1 + \lambda)(x_2 + \lambda)]^{\frac{n-m-3}{2}} \prod_{i=3}^{m-1} (x_i + \lambda)^{\frac{n-m-1}{2}} \prod_{1 \leq i < j \leq m-1} |x_i - x_j| \prod_{i=1}^{m-1} x_i e^{-x_i/2} dx_i$$

au point  $\lambda = 2z$ . □

### Preuve du lemme 7.3.4

Définissons sur  $S^{m-1}$  le polynôme quadratique défini par

$$X(s) = s^T B s.$$

Pour  $s \in S^{m-1}$ , nous notons  $\pi_s$  la projection sur  $s^\perp$ .

Soit  $(\lambda, \Lambda) \in \mathbb{R}^2$  avec  $\lambda < \Lambda$ , nous avons presque sûrement

$$(7.12) \quad \left\{ \begin{array}{l} \lambda_{\min} \in (\lambda, \lambda + d\lambda), \lambda_{\max} \in (\Lambda, \Lambda + d\Lambda) \\ \exists s, t \in S^{m-1}, \langle s, t \rangle = 0, X(s) \in (\lambda, \lambda + d\lambda), X(t) \in (\Lambda, \Lambda + d\Lambda), \\ \pi_s(Bs) = 0, \pi_t(Bt) = 0, X''(s) \prec 0, X''(t) \succ 0 \end{array} \right\}$$

En notant  $N_{\lambda, \Lambda, d\lambda, d\Lambda}$  le nombre de paires  $(s, t)$  de l'ensemble précédent, nous avons, par des considération sur les couples  $(\pm s, \pm t)$

$$\mathbb{P}(\lambda_{\min} \in (\lambda, \lambda + d\lambda), \lambda_{\max} \in (\Lambda, \Lambda + d\Lambda)) = \frac{1}{4} \mathbb{E}(N_{\lambda, \Lambda, d\lambda, d\Lambda}).$$

Notons

$$V = \{(s, t) \in (S^{m-1})^2, \langle s, t \rangle = 0\}.$$

D'après Azaïs et Wschebor,  $V$  est une variété sans bord contenue dans  $\mathbb{R}^{2m}$  de dimension  $2m - 3$  de mesure géométrique  $\theta_V(V) = \sqrt{2}\theta_{m-1}\theta_{m-2}$  où  $\theta_{m-1}$  représente l'aire de la surface de  $S^{m-1} \subset \mathbb{R}^m$ , soit

$$\theta_{m-1} = \frac{2\pi^{m/2}}{\Gamma(m/2)}.$$

Sur  $V$  nous définissons le champ aléatoire

$$Y : V \rightarrow \mathbb{R}^{2m} \quad Y(s, t) = \begin{pmatrix} \pi_s(Bs) \\ \pi_t(Bt) \end{pmatrix}.$$

Donc en utilisant exactement les arguments d'Azaïs et Wschebor (2003), qui ne prennent pas jusque là en compte la valeur de  $n$  ni celle de  $\Sigma$ , nous pouvons démontrer que

$$(7.13) \quad \mathbb{E}(N_{\lambda, \Lambda, d\lambda, d\Lambda}) \leq \int_V \mathbb{E}_{\lambda, \Lambda, 0} [\Delta(s, t) \mathbb{1}_{\{X''(s) \prec 0, X''(t) \succ 0\}}] p_{X(s), X(t), Y(s, t)}(\lambda, \Lambda, 0) \theta_V(d(s, t)),$$

avec

$$\Delta(s, t) = \left( \det \left[ (Y'(s, t))^T Y'(s, t) \right] \right)^{\frac{1}{2}}$$

et  $\mathbb{E}_{\lambda, \Lambda, 0}$  l'espérance sous la condition  $X(s) = \lambda, X(t) = \Lambda, Y(s, t) = 0$ .

Ils appliquent alors un argument d'invariance de la loi de  $A$  sous l'effet d'isométries, leur permettant d'enlever l'intégrale. Cet argument n'est plus valide dans notre cadre. L'idée est alors de caractériser les éléments apparaissant dans (7.13).

Construisons  $\{w_3, \dots, w_m\}$  dans  $(s, t)^\perp$  de sorte que  $\{s, t, w_3, \dots, w_m\}$  soit une base de  $\mathbb{R}^m$  et  $\{w_3, \dots, w_m\}$  est orthonormée pour la métrique induite par  $\Sigma$ . C'est-à-dire

$$(7.14) \quad \langle w_i, w_j \rangle_\Sigma = \sum_{k=1}^m (w_i)_k (w_j)_k \sigma_k = \delta_{ij}.$$

Notons la matrice de  $M_{m,m}(\mathbb{R})$  représentant  $B$  dans la base  $\{s, t, w_3, \dots, w_m\}$  par

$$B^{s,t,w} = (b^{s,t,w})_{i,j=1,\dots,m}.$$

Nous avons

$$X(s) = s^T B s = b_{11}^{s,t,w},$$

$$X(t) = t^T B t = b_{22}^{s,t,w},$$

$$X''(s) = 2(B_1^{s,t,w} - b_{11}^{s,t,w} I_{m-1}),$$

$$X''(t) = 2(B_2^{s,t,w} - b_{22}^{s,t,w} I_{m-1}),$$

$$Y(s, t) = \sum_{i=3}^m [b_{i1}^{s,t,w}(w_i, 0) + b_{i2}^{s,t,w}(w_i, 0)] + \sqrt{2} b_{12}^{s,t,w} [(s, t)/\sqrt{2}],$$

où  $B_i^{s,t,w}$  est obtenue à partir de  $B^{s,t,w}$  en supprimant la  $i$ ème ligne et la  $i$ ème colonne.

LOI DU VECTEUR  $(X(s), X(t), Y(s, t))$

Notons la représentation matricielle de  $A$  dans les bases  $\{s, t, w_3, \dots, w_m\}$  de  $\mathbb{R}^m$  et  $\{e_1, \dots, e_n\}$  de  $\mathbb{R}^n$  par

$$A^{s,t,w} = (a_{i,j}^{s,t,w})_{i=1,\dots,n; j=1,\dots,m}$$

et sa  $i$ ème colonne par  $\mathcal{A}_i^{s,t,w}$ . On peut montrer que

$$b_{ij}^{s,t,w} = \langle \mathcal{A}_i^{s,t,w}, \mathcal{A}_j^{s,t,w} \rangle.$$

• Nous recherchons la densité de

$$b_{31}^{s,t,w}, \dots, b_{m1}^{s,t,w}, b_{32}^{s,t,w}, \dots, b_{m2}^{s,t,w}$$

sachant

$$\mathcal{A}_1^{s,t,w}, \mathcal{A}_2^{s,t,w}, \|\mathcal{A}_1^{s,t,w}\|^2 = \lambda, \|\mathcal{A}_2^{s,t,w}\|^2 = \Lambda, \langle \mathcal{A}_1^{s,t,w}, \mathcal{A}_2^{s,t,w} \rangle = 0.$$

En utilisant les caractéristiques des vecteurs  $w_i$ , autrement dit (7.14), nous obtenons qu'il s'agit d'un vecteur gaussien centré de matrice de covariance

$$\Xi = \left( \begin{array}{c|c} \lambda I_{m-2} & 0 \\ \hline 0 & \Lambda I_{m-2} \end{array} \right).$$

La valeur de la densité conditionnelle au point 0 est

$$(7.15) \quad (2\pi)^{2-m} (\lambda\Lambda)^{\frac{2-m}{2}}.$$

• Nous recherchons maintenant la densité  $p$  de  $\|\mathcal{A}_1^{s,t,w}\|^2, \|\mathcal{A}_2^{s,t,w}\|^2, \langle \mathcal{A}_1^{s,t,w}, \mathcal{A}_2^{s,t,w} \rangle$  au point  $(\lambda, \Lambda, 0)$ . La loi jointe de  $\mathcal{A}_1^{s,t,w}$  et  $\mathcal{A}_2^{s,t,w}$  est gaussienne centrée de matrice de covariance

$$\Upsilon = \left( \begin{array}{c|c} \rho(s, s)I_n & \rho(s, t)I_n \\ \rho(s, t)I_n & \rho(t, t)I_n \end{array} \right)$$

où  $\rho(u, v) = \sum_{k=1}^m u_k v_k \sigma_k$ .

Définissons, en suivant la méthode de Wschebor (2004), l'application

$$F : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^3 \quad F(u, v) = (\|u\|^2, \|v\|^2, \langle u, v \rangle).$$

D'après la formule de la coaire de Federer, la densité du triplet au point  $(\lambda, \Lambda, 0)$  peut s'écrire comme

$$(7.16) \quad p(\lambda, \Lambda, 0) = \int_{F^{-1}(\lambda, \Lambda, 0)} (\det[F'(u, v)F'(u, v)^T])^{-\frac{1}{2}} (2\pi)^{-n} e^{-\frac{1}{2}(u, v)\Upsilon^{-1}(u, v)^T} d\gamma(u, v)$$

où  $\gamma$  désigne la mesure géométrique sur  $F^{-1}(\lambda, \Lambda, 0)$ . Il s'agit d'une variété infiniment différentiable donnée par les équations  $\|u\|^2 = \lambda, \|v\|^2 = \Lambda$  et  $\langle u, v \rangle = 0$  donc de dimension  $2n - 3$ . On peut vérifier que

$$\gamma(F^{-1}(\lambda, \Lambda, 0)) = (\lambda + \Lambda)^{\frac{1}{2}} (\lambda\Lambda)^{\frac{n-2}{2}} \theta_{n-1} \theta_{n-2}.$$

D'un autre côté, nous avons

$$F'(u, v) = \begin{pmatrix} 2u^T & 0 \\ 0 & 2v^T \\ v^T & u^T \end{pmatrix},$$

par suite, sous les conditions  $(u, v) \in F^{-1}(\lambda, \Lambda, 0)$ , nous obtenons

$$(7.17) \quad \det[F'(u, v)F'(u, v)^T] = 16\lambda\Lambda(\lambda + \Lambda).$$

Sous la condition  $(u, v) \in F^{-1}(\lambda, \Lambda, 0)$ , nous avons aussi

$$(7.18) \quad (u, v)\Upsilon^{-1}(u, v)^T = \frac{\rho(t, t)\lambda + \rho(s, s)\Lambda}{\rho(s, s)\rho(t, t) - \rho(s, t)^2}.$$

Le regroupement de (7.16) avec (7.17) et (7.18) conduit à

$$p(\lambda, \Lambda, 0) = \frac{(2\pi)^{-n} (\lambda\Lambda)^{\frac{n-3}{2}}}{4} \theta_{n-1} \theta_{n-2} \exp\left(-\frac{1}{2} \frac{\rho(t, t)\lambda + \rho(s, s)\Lambda}{\rho(s, s)\rho(t, t) - \rho(s, t)^2}\right)$$

En résumé nous avons montré que la densité de  $(X(s), X(t), Y(s, t))$  au point  $(\lambda, \Lambda, 0)$  vaut

$$p_{X(s), X(t), Y(s, t)}(\lambda, \Lambda, 0) = \frac{(\lambda\Lambda)^{\frac{n-m-1}{2}}}{2^{n+m} \pi^{m+n-2}} \theta_{n-1} \theta_{n-2} \exp\left(-\frac{1}{2} \frac{\rho(t, t)\lambda + \rho(s, s)\Lambda}{\rho(s, s)\rho(t, t) - \rho(s, t)^2}\right).$$

DÉTERMINATION DE  $\Delta(s, t)$ .

$Y'(s, t)$  est une application linéaire de l'espace tangent  $V$  au point  $(s, t)$  à valeurs dans  $\mathbb{R}^{2m}$ . Commençons par calculer la représentation matricielle de  $Y'(s, t)$  vue comme une application linéaire  $\tilde{Y}'(s, t)$  de  $\mathbb{R}^{2m}$  à valeurs dans  $\mathbb{R}^{2m}$ .

$$\tilde{Y}'(s, t) = \left( \begin{array}{c|c} B - \langle Bs, s \rangle I_n - 2ss^T B & 0 \\ \hline 0 & B - \langle Bt, t \rangle I_n - 2tt^T B \end{array} \right)$$

La représentation matricielle de  $Y'(s, t)$  s'obtient en regardant l'image de la base  $\{(w_3, 0), \dots, (w_m, 0), (0, w_3), \dots, (0, w_m), (s, t)/\sqrt{2}\}$  par  $\tilde{Y}'(s, t)$ . Après de longs mais simples calculs, nous trouvons sous la condition  $X(s) = \lambda, X(t) = \Lambda$  et  $Y(s, t) = 0$ ,

$$Y'(s, t) = \begin{pmatrix} 0_{1,m-2} & 0_{1,m-2} & 0 \\ 0_{1,m-2} & 0_{1,m-2} & (\lambda - \Lambda)/\sqrt{2} \\ B_{12}^{s,t,w} - \lambda I_{m-2} & 0_{m-2,m-2} & 0_{m-2,1} \\ 0_{1,m-2} & 0_{1,m-2} & (\lambda - \Lambda)/\sqrt{2} \\ 0_{1,m-2} & 0_{1,m-2} & 0 \\ 0_{m-2,m-2} & B_{12}^{s,t,w} - \Lambda I_{m-2} & 0_{m-2,1} \end{pmatrix}.$$

Par conséquent

$$\Delta(s, t) = |\det(B_{12}^{s,t,w} - \lambda I_{m-2})| |\det(B_{12}^{s,t,w} - \Lambda I_{m-2})| (\Lambda - \lambda).$$

Ainsi l'espérance conditionnelle de la relation (7.13) devient

$$\mathbb{E}_{\lambda, \Lambda, 0} \left( |\det(B_{12}^{s,t,w} - \lambda I_{m-2})| |\det(B_{12}^{s,t,w} - \Lambda I_{m-2})| (\Lambda - \lambda) \mathbb{1}_{\{B_1^{s,t,w} - \lambda I_{m-1} < 0, B_2^{s,t,w} - \Lambda I_{m-1} > 0\}} \right)$$

qui est majorée par

$$\Lambda^{m-2} (\Lambda - \lambda) \mathbb{E}_{x,y,0} \left( |\det(B_{12}^{s,t,w} - \lambda I_{m-2})| \mathbb{1}_{\{B_2^{s,t,w} - \lambda I_{m-1} > 0\}} \right).$$

En conditionnant par rapport à  $\mathcal{A}_1^{s,t,w}$  et  $\mathcal{A}_2^{s,t,w}$ , les vecteurs  $\mathcal{A}_i^{s,t,w}$  où  $i \geq 3$  sont gaussiens standards et indépendants dans  $\mathbb{R}^{n-2}$ . Plus précisément nous pouvons écrire cette dernière espérance comme

$$\mathbb{E}(|\det(M - \lambda I_{m-2})| \mathbb{1}_{\{M - \lambda I_{m-2} > 0\}})$$

où  $M$  est une matrice de  $M_{m-2}(\mathbb{R})$  et  $(M)_{ij} = \langle v_i, v_j \rangle$  pour des vecteurs indépendants gaussiens standards dans  $\mathbb{R}^{n-2}$ . Enfin cette dernière espérance est majorée par

$$\mathbb{E}(\det M) = \frac{(n-2)!}{(n-m)!}.$$

En résumé, l'espérance conditionnelle qui apparaît dans (7.13) est majorée par

$$(7.19) \quad \mathbb{E}_{\lambda, \Lambda, 0} \left( \Delta(s, t) \mathbb{1}_{\{X''(s) < 0, X''(t) > 0\}} \right) \leq \Lambda^{m-1} \frac{(n-2)!}{(n-m)!}.$$

Nous pouvons donc conclure que  $\mathbb{E}(N_{\lambda, \Lambda, d\lambda, d\Lambda})$  est majorée par

$$\Lambda^{m-1} \frac{(n-2)!}{(n-m)!} \frac{(\lambda\Lambda)^{\frac{n-m-1}{2}}}{2^{n+m}\pi^{m+n-2}} \theta_{n-1}\theta_{n-2} \int_V \exp\left(-\frac{1}{2} \frac{\rho(t,t)\lambda + \rho(s,s)\Lambda}{\rho(s,s)\rho(t,t) - \rho(s,t)^2}\right) \theta_V(d(s,t)).$$

Majorons le dénominateur dans l'exponentielle par  $\rho(s,s)\rho(t,t)$  et  $\rho(s,s)$  par  $\sigma_+$  de telle sorte que

$$\int_V \exp\left(-\frac{1}{2} \frac{\rho(t,t)\lambda + \rho(s,s)\Lambda}{\rho(s,s)\rho(t,t) - \rho(s,t)^2}\right) \theta_V(d(s,t)) \leq \sqrt{2}\theta_{m-1}\theta_{m-2} \exp\left(-\frac{\lambda + \Lambda}{2\sigma_+}\right)$$

et

$$\mathbb{E}(N_{\lambda, \Lambda, d\lambda, d\Lambda}) \leq \frac{\sqrt{2}\theta_{m-1}\theta_{m-2}\theta_{n-1}\theta_{n-2}}{2^{n+m}\pi^{n+m-2}} \frac{(n-2)!}{(n-m)!} \Lambda^{\frac{n+m-3}{2}} \lambda^{\frac{n-m-1}{2}} \exp\left(-\frac{\lambda + \Lambda}{2\sigma_+}\right).$$

Or il existe une formule de duplication de la fonction  $\Gamma$  (voir par exemple Abramowitz et Stegun, 1992),

$$\Gamma(2z) = (2\pi)^{-1/2} 2^{2z-1/2} \Gamma(z)\Gamma(z-1).$$

En l'application à  $z = (n-1)/2$  et  $z = (m-1)/2$ , nous avons

$$\theta_{m-1}\theta_{m-2}\theta_{n-1}\theta_{n-2} = \frac{\pi^{m+n-2}}{(m-2)!(n-2)!2^{-m-n}},$$

d'où

$$f_{\lambda_{\min}, \lambda_{\max}}(\lambda, \Lambda) \leq \frac{\sqrt{2}}{4(m-2)!(n-m)!} \Lambda^{\frac{n+m-3}{2}} \lambda^{\frac{n-m-1}{2}} \exp\left(-\frac{\lambda + \Lambda}{2\sigma_+}\right),$$

qui est bien le résultat annoncé. □

### Preuve du lemme 7.3.5

Pour  $\lambda > 0$ , nous avons presque sûrement

$$\begin{aligned} \{\lambda_{\min} \in (\lambda, \lambda + d\lambda)\} &= \{\exists s \in S^{m-1}, X(s) \in (\lambda, \lambda + d\lambda), \pi_s(Bs) = 0, X''(s) \succ 0\} \\ &\subset \{\exists s \in S^{m-1}, X(s) \in (\lambda, \lambda + d\lambda), X'(s) = 0, X''(s) \succ 0\}, \end{aligned}$$

en utilisant les mêmes notations que celles utilisées dans la démonstration du lemme 7.3.4. Cette fois en notant  $N_{\lambda, d\lambda}$  le nombre de point  $s$  de l'ensemble précédent, par des considérations sur le signe de  $s$ , nous avons

$$\mathbb{P}(\lambda_{\min} \in (\lambda, \lambda + d\lambda)) \leq \frac{1}{2} \mathbb{E}(N_{\lambda, d\lambda}),$$

et

$$f_{\lambda_{\min}}(\lambda) \leq \frac{1}{2} \int_{S^{m-1}} \mathbb{E}_{\lambda, 0} (|\det[X''(s)]| \mathbb{1}_{\{X''(s) \succ 0\}}) p_{X(s), X'(s)}(\lambda, 0) d\theta_{m-1}(s),$$

où  $\mathbb{E}_{\lambda, 0}$  est l'espérance conditionnée par  $X(s) = \lambda$  et  $X'(s) = 0$ .

De nouveau nous construisons une base  $\{s, w_2, \dots, w_m\}$  de  $\mathbb{R}^m$  avec  $w_i$  une famille dans  $s^\perp$  orthonormée pour la métrique induite par  $\Sigma$ . Nous reprenons les notations du lemme 7.3.4 en les adaptant à notre contexte. Comme  $X(s) = b_{11}^{s,w} = \|\mathcal{A}_1^{s,w}\|_n^2$ , nous pouvons montrer que  $X(s)$  s'écrit à la constante près  $\rho(s, s)$  comme une  $\chi^2(n)$  d'où

$$p_{X(s)}(\lambda) = \frac{1}{\rho(s, s)} \frac{1}{2^{n/2} \sqrt{\pi}} \exp\left(-\frac{\lambda}{2\rho(s, s)}\right) \left(\frac{\lambda}{\rho(s, s)}\right)^{n/2-1} \frac{\Gamma(1/2)}{\Gamma(n/2)}.$$

D'autre part, le vecteur  $(b_{21}^{s,w}, \dots, b_{m,1}^{s,w})$  sachant  $\mathcal{A}_1^{s,w}$  suit une loi normale centrée de matrice de covariance  $\|\mathcal{A}_1^{s,w}\|_n^2 I_{m-1}$ . Il suit

$$p_{X'(s)/X(s)=\lambda}(z) = \frac{1}{2^{m-1}} \frac{1}{(2\pi)^{\frac{m-1}{2}}} \frac{1}{\lambda^{\frac{m-1}{2}}} \exp\left(-\frac{\|z\|_n^2}{4\lambda}\right)$$

et

$$p_{X(s), X'(s)}(\lambda, 0) = \frac{1}{2^{\frac{n+3m-3}{2}} \pi^{\frac{m-1}{2}} \Gamma(n/2)} \frac{\lambda^{\frac{n-m-1}{2}}}{\rho(s, s)^{n/2}} \exp\left(-\frac{\lambda}{2\rho(s, s)}\right).$$

Rappelons que  $X''(s) = 2(B_1^{s,w} - X(s)I_{m-1})$  et que les éléments de  $B_1^{s,w}$  sont les produits scalaires des colonnes de  $A^{s,w}$ . Or  $\mathcal{A}_i^{s,w}$  pour  $i \geq 2$  sachant  $X(s) = \lambda$  et  $X'(s) = 0$  est de loi normale standard dans  $\mathbb{R}^{n-1}$ . Nous pouvons construire une matrice  $M$  de  $M_{m-1}(\mathbb{R})$  où  $(M)_{ij} = \langle v_i, v_j \rangle$  avec des vecteurs  $v$  indépendants gaussiens standard dans  $\mathbb{R}^{n-1}$  telle que

$$\mathbb{E}_{\lambda, 0} (|\det[X''(s)]| \mathbb{1}_{\{X''(s) \succ 0\}}) = \mathbb{E}(|\det[2(M - \lambda I_{m-1})]| \mathbb{1}_{M - \lambda I_{m-1} \succ 0}) \leq \frac{2^{m-1}(n-1)!}{(n-m)!}.$$

De plus

$$\int_{S^{m-1}} \frac{1}{\rho(s, s)^{n/2}} \exp\left(-\frac{\lambda}{2\rho(s, s)}\right) \theta_{m-1}(ds) \leq \frac{\theta_{m-1}}{\sigma_-^{n/2}} \exp\left(-\frac{\lambda}{2\sigma_+}\right).$$

La conclusion suit en remplaçant  $\theta_{m-1}$  par son expression.  $\square$

### Preuve du lemme 7.3.6

Pour tout  $t \in S^{m-1}$  tel que  $X'(t) = 0$ , notons  $i_X(t)$  le nombre de valeurs propres négatives de  $X''(t)$ . Pour chaque  $\lambda > 0$  et  $i \in \{0, \dots, m-1\}$  posons

$$N_i(\lambda) = \#\{t \in S^{m-1} : X(t) < \lambda, X'(t) = 0, i_X(t) = i\},$$

$$M(\lambda) = \sum_{i=0}^{m-1} (-1)^i N_i(\lambda).$$

- Si  $\lambda \leq \lambda_{\min}$  alors  $N_i(\lambda) = 0$  pour tout  $i \in \{0, \dots, m-1\}$ .
- Si  $\lambda_k \leq \lambda \leq \lambda_{k+1}$  alors  $N_i(\lambda) = 2$  pour  $i \in \{0, \dots, k-1\}$  et  $N_i(\lambda) = 0$  pour  $i \in \{k, \dots, m-1\}$ .
- Si  $\lambda_{\max} \leq \lambda$  alors  $N_i(\lambda) = 2$  pour tout  $i \in \{0, \dots, m-1\}$ .

Par conséquent, si  $N_0(\lambda) = 0$  alors  $M(\lambda) = 0$  et si  $N_0(\lambda) = 2$  alors  $M(\lambda) \leq 2$ . Nous pouvons en déduire que

$$\mathbb{P}(\lambda_{\min} < \lambda) = \mathbb{P}(N_0(\lambda) = 2) = \frac{1}{2} \mathbb{E}(N_0(\lambda)) \geq \frac{1}{2} \mathbb{E}(M(\lambda)).$$

En appliquant les résultats d'Adler (1981) sur la caractéristique d'Euler  $M(\lambda)$ , nous savons que

$$\mathbb{E}(M(\lambda)) = \int_0^\lambda dy \int_{S^{m-1}} \mathbb{E}[\det(X''(t)/X(t) = y, X'(t) = 0)] p_{X(t), X'(t)}(\lambda, 0) \theta_{m-1}(dt).$$

En reprenant les arguments donnés dans la démonstration du lemme 7.3.4, nous avons

$$\mathbb{E}[\det(X''(t)/X(t) = y, X'(t) = 0)] = \mathbb{E}[\det(2(M - yI_{m-1}))],$$

$M$  étant une matrice de  $M_{m-1}(\mathbb{R})$  où  $(M)_{ij} = \langle v_i, v_j \rangle$  avec des vecteurs  $v$  indépendants gaussiens standard dans  $\mathbb{R}^{m-1}$ . Or nous savons que

$$\mathbb{E}[\det(M - yI_{m-1})] = (m-1)! \sum_{k=0}^{m-1} (-1)^k C_{m-1}^k \frac{y^k}{k!}.$$

Sous la condition  $0 < \lambda < (m-1)^{-1}$ , nous remarquons que le terme de la série est en valeur absolue une fonction décroissante de  $k$ , donc

$$\mathbb{E}[\det(X''(t)/X(t) = y, X'(t) = 0)] \geq 2^{m-1} (m-1)! [1 - (m-1)y].$$

D'après la démonstration du lemme 7.3.5, nous connaissons la densité jointe de  $X(t)$  et  $X'(t)$ , à l'aide de minoration évidentes nous en déduisons que

$$p_{X(s), X'(s)}(y, 0) \geq \frac{1}{2^{2m-\frac{3}{2}} \pi^{\frac{m-1}{2}} \Gamma(\frac{m}{2}) \sigma_+^{n/2}} \frac{y^{-\frac{1}{2}}}{\sigma_-} \exp\left(-\frac{y}{2\sigma_-}\right).$$



En regroupant les différentes inégalités recueillies, nous avons

$$\mathbb{P}(\lambda_{\min} < \lambda) \geq \frac{2^{m-1}(m-1)!\theta_{m-1}}{2^{2m-\frac{3}{2}}\pi^{\frac{m-1}{2}}\Gamma(\frac{m}{2})\sigma_+^{n/2}} \int_0^\lambda y^{\frac{-1}{2}} [1 - (m-1)y] \exp\left(-\frac{y}{2\sigma_-}\right) dy,$$

soit encore

$$\mathbb{P}(\lambda_{\min} < \lambda) \geq \frac{\sqrt{2\pi}(m-1)!}{2^{m-1}\Gamma^2(\frac{m}{2})\sigma_+^{n/2}} \int_0^\lambda y^{\frac{-1}{2}} [1 - (m-1)y] \exp\left(-\frac{y}{2\sigma_-}\right) dy.$$

Étudions maintenant l'intégrale de la borne précédente que nous désignons par

$$I(\lambda, \sigma_-, m) = \int_0^\lambda y^{\frac{-1}{2}} [1 - (m-1)y] \exp\left(-\frac{y}{2\sigma_-}\right) dy.$$

Nous avons, en utilisant le fait que  $\lambda m < 1$  et  $\sigma > 1/2$

$$\begin{aligned} I(\lambda, \sigma, m) &\geq \int_0^\lambda y^{\frac{-1}{2}} [1 - (m-1)y] \left(1 - \frac{y}{2\sigma}\right) dy, \\ &\geq \frac{\lambda^{\frac{5}{2}}(m-1)}{5\sigma} + \frac{1}{3} \frac{\lambda^{\frac{3}{2}}(2\sigma-1-2m\sigma)}{\sigma} + 2\frac{\lambda^{\frac{1}{2}}}{\sigma}, \\ &\geq \frac{\lambda^{\frac{5}{2}}(m-1)}{5\sigma} + \frac{\lambda^{\frac{3}{2}}}{3} \frac{2\sigma-1}{\sigma} + \left(\frac{2}{\sigma} - \frac{2}{3}\right) \lambda^{\frac{1}{2}}, \\ &\geq \left(\frac{2}{\sigma} - \frac{2}{3}\right) \lambda^{\frac{1}{2}}. \end{aligned}$$

La fonction de la variable  $m$

$$\frac{1}{\sqrt{m}} \frac{\sqrt{2\pi}(m-1)!}{2^{m-1}\Gamma^2(\frac{m}{2})}$$

étant croissante, nous la minorons par sa valeur en  $m = 3$ , à savoir

$$\frac{2^{\frac{3}{2}}}{\sqrt{3\pi}}.$$

En regroupant nos résultats nous pouvons conclure que

$$\mathbb{P}(\lambda_{\min} < \lambda) \geq \frac{2^{\frac{3}{2}}}{\sqrt{3\pi}\sigma_+^{n/2}} \left(\frac{2}{\sigma_-} - \frac{2}{3}\right) \sqrt{m\lambda}.$$

□

## Bibliographie du chapitre 7

**Abramowitz, M. et Stegun, I. A.** (1992). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover Publications Inc., New York.

**Adler, R. J.** (1981). *The geometry of random fields*. John Wiley & Sons Ltd., Chichester. Wiley Series in Probability and Mathematical Statistics.

- Anderson, T. W.** (1963). Asymptotic theory for principal component analysis. *Ann. Math. Statist.*, 34 :122–148.
- Azaïs, J.-M. et Wschebor, M.** (2003). Upper and lower bounds for the tails of the distribution of the condition number of a gaussian matrix. Submitted.
- Cucker, F. et Wschebor, M.** (2003). On the expected condition number of linear programming problems. *Numer. Math.*, 94(3) :419–478.
- Cuesta-Albertos, J. A. et Wschebor, M.** (2003). Some remarks on the condition number of a real random square matrix. *J. Complexity*, 19(4) :548–554.
- Cuesta-Albertos, J. A. et Wschebor, M.** (2004). Condition numbers and extrema of random fields. In *Seminar on Stochastic Analysis, Random Fields and Applications IV*, volume 58 of *Progr. Probab.*, pages 69–82. Birkhäuser, Basel.
- Paul, D.** (2004). Asymptotics of the leading sample eigenvalues for a spiked covariance model. Preprint.
- Edelman, A.** (1988). Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(4) :543–560.
- Edelman, A.** (1989). *Eigenvalues and condition numbers of random matrices*. PhD thesis, Yale University, Massachusetts institute of technology.
- Geman, S.** (1980). A limit theorem for the norm of random matrices. *Ann. Probab.*, 8(2) :252–261.
- Johansson, K.** (2000). Shape fluctuations and random matrices. *Comm. Math. Phys.*, 209(2) :437–476.
- Johnstone, I. M.** (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2) :295–327.
- Johnstone, I. M. et Lu, A. Y.** (2004). Sparse Principal Components Analysis. To appear in *J. Amer. Statist. Assoc.*
- Krishnaiah, P. R. et Schuurmann, F. J.** (1974). On the evaluation of some distributions that arise in simultaneous tests for the equality of the latent roots of the covariance matrix. *J. Multivariate Anal.*, 4 :265–282.
- Ledoit, O. et Wolf, M.** (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.*, 30(4) :1081–1102.
- Marčenko, V. A. et Pastur, L. A.** (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72 (114) :507–536.
- Muirhead, R. J.** (1982). *Aspects of multivariate statistical theory*. John Wiley & Sons Inc., New York. Wiley Series in Probability and Mathematical Statistics.
- Silverstein, J. W.** (1985). The smallest eigenvalue of a large-dimensional Wishart matrix. *Ann. Probab.*, 13(4) :1364–1368.

- Soshnikov, A.** (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Statist. Phys.*, 108(5-6) :1033–1056. Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays.
- Sugiyama, T.** (1970). Joint distribution of the extreme roots of a covariance matrix. *Ann. Math. Statist.*, 41 :655–657.
- Wachter, K. W.** (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probability*, 6(1) :1–18.
- Waikar, V. B. et Schuurmann, F. J.** (1973). Exact joint density of the largest and smallest roots of the Wishart and Manova matrices. *Utilitas Math.*, 4 :253–260.
- Wschebor, M.** (2004). Smoothed analysis of  $\kappa(A)$ . *J. Complexity*, 20(1) :97–107.

---

---

QUATRIÈME PARTIE

RÉSULTATS ANNEXES

---

---



# Chapitre 8

---

## PROCESSUS DE POISSON

Nous proposons de décrire dans ce chapitre quelques concepts basiques sur la théorie des processus de Poisson. Leur connaissance est une aide précieuse dans plusieurs résultats de la théorie des extrêmes puisque le nombre d'événements rares observés au cours du temps est généralement modélisé par un processus de Poisson.

### 8.1 Définition

Dès le premier chapitre de ce manuscrit, nous avons rencontré des processus ponctuels comme le processus des upcrossings. Notons l'ensemble des mesures ponctuelles

$$M_p(\mathbb{R}) = \left\{ m = \sum_{i \in \mathbb{N}} \delta_{x_i} \text{ pour } (x_i)_{i \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \right\},$$

et l'ensemble des processus ponctuels

$$M_+(\mathbb{R}) = \{ N, \text{ variable aléatoire à valeurs dans } M_p(\mathbb{R}) \}.$$

Voici une définition du processus de Poisson. Ce processus stochastique est un élément de modélisation incontournable.

#### Définition 8.1

Soit  $\mu$  une mesure positive  $\sigma$ -finie sur  $\mathbb{R}$ . Un processus de Poisson  $N$  d'intensité  $\mu$  vérifie

a) pour tout borélien  $B$  de  $\mathbb{R}$  et tout entier positif ou nul  $k$

$$\mathbb{P}(N(B) = k) = \begin{cases} \frac{\mu(B)^k \exp(-\mu(B))}{k!} & \text{si } \mu(B) < \infty \\ 0 & \text{si } \mu(B) = \infty, \end{cases}$$

b) pour tout entier positif  $r$  et  $B_1, \dots, B_r$  des boréliens de  $\mathbb{R}$  disjoints, les variables  $N(B_1), \dots, N(B_r)$  sont indépendantes.

Lorsque la mesure  $\mu$  est la mesure de Lebesgue notée  $\lambda$ , le processus est appelé processus de Poisson standard.

Nous présentons maintenant une caractérisation du processus de Poisson. Il s'agit d'une version modifiée d'un résultat de Khintchine (1960) grâce à une propriété de Dobrushin (1960).

**Proposition 8.2**

Un processus ponctuel  $N$  satisfaisant les conditions

- i) pour tout intervalle  $I$  et tout entier  $k$  positif ou nul,  $\mathbb{P}(N(I) = k)$  ne dépend que de  $\mu(I)$  et de  $k$ ,
- ii) pour tout entier  $r$  positif et  $I_1, \dots, I_r$  des intervalles disjoints, les variables  $N(I_1), \dots, N(I_r)$  sont indépendantes,
- iii) pour tout  $t$  dans  $\mathbb{R}$ , il existe  $\varepsilon > 0$  tel que  $\mathbb{P}(N(t - \varepsilon, t + \varepsilon) > 1) = 0$ ,
- iv)  $\mathbb{E}(N(0, 1)) < \infty$ .

Le processus  $N$  est un processus de Poisson de mesure  $\mu(I) = \mathbb{E}(N(0, 1))\lambda(I)$ .

## 8.2 Convergence faible vers un processus de Poisson

Nous allons maintenant caractériser la convergence en loi d'un processus ponctuel vers le processus de Poisson standard. Nous avons déjà remarqué dans la section 2.1.2 l'importance de la connaissance asymptotique des processus de upcrossings pour obtenir la loi du maximum d'un processus stochastique. Sous certaines conditions, explicitées dans cette même section, le processus limite n'est autre qu'un processus de Poisson.

Introduisons la famille  $\mathcal{J}$  composée des réunions finies d'intervalles bornés disjoints de  $\mathbb{R}$ . La définition d'un processus de Poisson peut en fait se réduire à l'étude du processus ponctuel sur cette famille.

**Proposition 8.3 (Rényi (1967))**

Un processus ponctuel vérifiant, sur tout élément de  $\mathcal{J}$ , la condition a) de la définition 8.1 est un processus de Poisson de mesure  $\mu$ .

Commençons par rappeler qu'une suite  $N_n$  converge en distribution vers  $N$  si et seulement si de toute sous-suite  $N_{n'}$  on peut extraire une sous-suite  $N_{n''}$  qui converge vers  $N$ . Comme  $M_p(\mathbb{R})$  est un espace polonais (i.e. métrique, séparable et complet), la relative compacité d'une suite de processus ponctuels est équivalente à sa tension. Il nous reste donc à énoncer la condition nécessaire et suffisante pour obtenir la tension d'une suite de processus.

**Rappel 8.4**

Soit  $N_n$  une suite de  $M_+(\mathbb{R})$ . Elle est tendue si et seulement si pour tout borélien  $B$  relativement compact dans  $\mathbb{R}$ ,

$$\lim_{t \rightarrow \infty} \limsup \mathbb{P}(N_n(B) > t) = 0.$$

**Proposition 8.5 (Kallenberg (1973))**

Soit  $N_n$  une suite de  $M_+(\mathbb{R})$  vérifiant pour tout élément  $B$  de la famille  $\mathcal{J}$ ,

- a)  $\mathbb{E}(N_n(B)) \xrightarrow{n \rightarrow \infty} \lambda(B)$ ,
- b)  $\mathbb{P}(N_n(B) = 0) \xrightarrow{n \rightarrow \infty} \exp(-\lambda(B))$ .

Alors,  $N_n$  converge en distribution vers le processus de Poisson standard.

*Preuve de la proposition 8.3.* La condition a) de la définition 8.1 étant vérifiée, il est clair que le processus est tendu d'après la condition nécessaire et suffisante donnée par la proposition 8.4. Ainsi, il suffit d'étudier les répartitions finies du processus, autrement dit, la restriction du processus sur  $\mathcal{J}$ .

Soit  $E$  un élément de  $\mathcal{J}$ , on considère l'événement  $A(E) = \{N(E) = 0\}$ . Si  $(E_i)_{i \in \{1, \dots, n\}}$  est une partition de  $E$ , nous savons que

$$N(E) = \sum_{i=1}^n N(E_i)$$

d'où :

$$\mathbb{P}(A(E)) = \mathbb{P}(A(E_1), \dots, A(E_n)).$$

D'autre part,

$$\mathbb{P}(A(E)) = \mathbb{P}(N(E) = 0) = \exp(-\mu(E)) = \prod_{i=1}^n \exp(-\mu(E_i)) = \prod_{i=1}^n \mathbb{P}(A(E_i)).$$

L'indépendance des événements  $A(E_i)$  entraîne à son tour l'indépendance des indicatrices  $\mathbb{1}_{A(E_i)^c}$ .

Soit  $E$  et  $F$  des éléments de  $\mathcal{J}$  disjoints et  $\varepsilon > 0$ . Découpons  $E$  en  $n$  intervalles  $E_i$  de longueur inférieure ou égale à  $\varepsilon$ . De même,  $F$  est divisé en  $m$  intervalles  $F_j$  de longueur inférieure ou égale à  $\varepsilon$ .

Par hypothèse, pour tout élément  $B$  de  $\mathcal{J}$ ,

$$\mathbb{P}(N(B) \geq 2) = \sum_{k \geq 2} \exp(-\mu(B)) \frac{\mu(B)^k}{k!} \leq \mu(B)^2.$$

Cette inégalité nous permet d'écrire :

$$\mathbb{P}(N(E) \neq \sum_{i=1}^n \mathbb{1}_{A(E_i)^c}) \leq \sum_{i=1}^n \mathbb{P}(N(E_i) \geq 2) \leq \varepsilon \mu(E).$$

De la même manière, on trouve :

$$\mathbb{P}(N(F) \neq \sum_{j=1}^m \mathbb{1}_{A(F_j)^c}) \leq \varepsilon \mu(F).$$

Il s'agit de démontrer que les variables  $N(E)$  et  $N(F)$  sont indépendantes. Puisque les ensembles  $E_i$  et  $F_j$  sont disjoints, on sait déjà que les variables

$$\sum_{j=1}^m \mathbb{1}_{A(F_j)^c} \quad \text{et} \quad \sum_{i=1}^n \mathbb{1}_{A(E_i)^c}$$

sont indépendantes.

$$|\mathbb{P}(N(E) = n, N(F) = m) - \mathbb{P}(N(E) = n)\mathbb{P}(N(F) = m)| \leq |1| + |2| + |3|$$



avec

$$\begin{aligned}
|1| &= |\mathbb{P}(N(E) = n, N(F) = m) - \mathbb{P}(N(E) = n, \sum_{j=1}^m \mathbb{1}_{A(F_j)^c} = m)|, \\
|2| &= |\mathbb{P}(N(E) = n, \sum_{j=1}^m \mathbb{1}_{A(F_j)^c} = m) - \mathbb{P}(\sum_{i=1}^n \mathbb{1}_{A(E_i)^c} = n, \sum_{j=1}^m \mathbb{1}_{A(F_j)^c} = m)|, \\
|3| &= |\mathbb{P}(\sum_{i=1}^n \mathbb{1}_{A(E_i)^c} = n, \sum_{j=1}^m \mathbb{1}_{A(F_j)^c} = m) - \mathbb{P}(N(E) = n)\mathbb{P}(N(F) = m)|.
\end{aligned}$$

Étudions chacun de ces trois termes séparément.

$$\begin{aligned}
|1| &= \mathbb{P}\left(N(E) = n, N(F) = m, \sum_{j=1}^m \mathbb{1}_{A(F_j) \neq m}\right) \leq \varepsilon\mu(F), \\
|2| &= \mathbb{P}\left(N(E) = n, \sum_{j=1}^m \mathbb{1}_{A(F_j)} = m, \sum_{i=1}^n \mathbb{1}_{A(E_i) \neq n}\right) \leq \varepsilon\mu(E), \\
|3| &= \left| \mathbb{P}\left(\sum_{i=1}^n \mathbb{1}_{A(E_i)} = n\right) \mathbb{P}\left(\sum_{j=1}^m \mathbb{1}_{A(F_j)} = m\right) - \mathbb{P}(N(E) = n) \mathbb{P}(N(F) = m) \right| \\
&= \left| \mathbb{P}\left(\sum_{i=1}^n \mathbb{1}_{A(E_i)} = n\right) \left\{ \mathbb{P}\left(\sum_{j=1}^m \mathbb{1}_{A(F_j)} = m, N(F) = m\right) + \mathbb{P}\left(\sum_{j=1}^m \mathbb{1}_{A(F_j)} = m, N(F) \neq m\right) \right\} \right. \\
&\quad \left. - \mathbb{P}(N(E) = n) \mathbb{P}(N(F) = m) \right| \\
&\leq \left| \mathbb{P}\left(\sum_{i=1}^n \mathbb{1}_{A(E_i)} = n\right) \mathbb{P}(N(F) = m) - \mathbb{P}(N(E) = n) \mathbb{P}(N(F) = m) \right| \\
&\quad + \left| \mathbb{P}(N(F) \neq \sum_{j=1}^m \mathbb{1}_{A(F_j)}) \right| \\
&\leq \varepsilon\mu(E) + \varepsilon\mu(F)
\end{aligned}$$

Finalemeent nous obtenons

$$|\mathbb{P}(N(E) = n, N(F) = m) - \mathbb{P}(N(E) = n) \mathbb{P}(N(F) = m)| \leq 2\varepsilon(\mu(E) + \mu(F)),$$

$\varepsilon$  étant arbitrairement petit, cela implique l'indépendance des variables aléatoires  $N(E)$  et  $N(F)$ . On démontre de la même manière l'indépendance des  $r$  variables  $N(E_k)$  pour  $k \in \{1, \dots, r\}$  lorsque les ensembles  $E_k$  de  $\mathcal{J}$  sont disjoints.  $\square$

*Preuve de la proposition 8.5.* Soit  $B$  un élément de  $\mathcal{J}$  :  
Considérons  $K$  une partie relativement compacte de  $B$  :

$$\lim_{t \rightarrow \infty} \limsup_n \mathbb{P}(N_n(K) > t) \leq \lim_{t \rightarrow \infty} \limsup_n \frac{\mathbb{E}(N_n(K))}{t} = \lim_{t \rightarrow \infty} \frac{\lambda(K)}{t} = 0$$

Par conséquent la suite  $(N_n)_{n \in \mathbb{N}}$  est tendue sur  $B$ . Quitte à prendre une sous-suite convergente en distribution, on suppose que la restriction sur  $B$  de la suite  $N_n$  converge

en distribution vers la restriction sur  $B$  de  $N$ . Il s'agit de montrer que  $N(B)$  suit une loi de Poisson de paramètre  $\lambda(B)$ .

D'après le lemme de Portmanteau,  $\liminf \mathbb{E}(N_n(B)) \geq \mathbb{E}(N(B))$  car l'application identité est positive et continue sur  $\mathbb{R}_+$ .

Les variables  $N_n(B)$  étant entières, la convergence en distribution du processus implique en particulier que

$$\mathbb{P}(N_n(B) = 0) \xrightarrow[n \rightarrow \infty]{} \mathbb{P}(N(B) = 0).$$

Par hypothèse et unicité de la limite, on a  $\mathbb{P}(N(B) = 0) = \exp(-\lambda(B))$ .

Découpons  $B$  en  $n$  intervalles ou union d'intervalles  $B_i$  de longueurs  $\frac{\lambda(B)}{n}$ . L'argument précédent permet de dire que  $\mathbb{P}(N(B_i) = 0) = \exp(-\lambda(B_i))$ . En se rapportant à la démonstration de la proposition de Rényi 8.3, on se rend compte que cette dernière égalité suffit pour affirmer l'indépendance des variables  $\mathbb{1}_{N(B_i) > 0}$ .

Presque sûrement nous avons  $N(B) \geq \sum_{i=1}^n \mathbb{1}_{N(B_i) > 0}$ , où le deuxième membre est une variable Binomiale  $B(n, p_n)$ . Comme

$$p_n = 1 - \exp\left(-\frac{\lambda(B)}{n}\right) \sim \frac{\lambda(B)}{n},$$

cela implique

$$\triangleright \sum_{i=1}^n \mathbb{1}_{N(B_i) > 0} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} Z(B) \text{ suit une loi de Poisson de paramètre } \lambda(B),$$

$\triangleright$  presque sûrement  $N(B) \geq Z(B)$ .

En regroupant les différents résultats obtenus, nous avons :

$\mu(B) = \liminf \mathbb{E}(N_n(B)) \geq \mathbb{E}(N(B)) \geq \mathbb{E}(Z(B)) = \lambda(B)$ . Donc presque sûrement  $N(B) = Z(B)$ .

La propriété de Rényi 8.3 permet de conclure que  $N$  est un processus de Poisson standard.  $\square$

## Bibliographie du chapitre 8

**Dobrushin, R. L.** (1960). Properties of sample functions of a stationary gaussian process. *Teoriya Veroyatnostei i ee Primeneniya*, 5 :132–134.

**Kallenberg, O.** (1973). Characterization and convergence of random measures and point processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 27 :9–21.

**Khinchine, Y. A.** (1960). *Mathematical methods in the theory of queueing*. Griffin, London.

**Rényi, A.** (1967). Remarks on the Poisson process. *Studia Sci. Math. Hungar*, 2 :119–123.



# Chapitre 9

---

## FORMULES DE RICE

Dans ce chapitre nous reprenons la littérature sur le calcul du nombre de crossings définis dans la section 1.2 du chapitre 1. Plusieurs méthodes se dessinent, en particulier comme conséquences de différents résultats analytiques. Nous terminons en démontrant une extension de la formule de Rice pour des champs aléatoires gaussiens.

### 9.1 Introduction

L'origine du nom des formules présentées dans ce chapitre provient du mathématicien américain Stephen O. Rice (1907 - 1986). En 1945, il s'est en particulier intéressé dans "Mathematical Analysis of Random Noise" à plusieurs caractéristiques du bruit aléatoire d'un courant  $I(t)$ . Il le représente comme la somme de  $N$  "cosine processes", c'est-à-dire sous la forme

$$a_k \cos(\omega_k t) + b_k \sin(\omega_k t)$$

où les  $a_k, b_k$  sont des variables aléatoires indépendantes de loi normale centrée. Il calcule le nombre moyen de zéros de  $I(t)$  par unité de temps. Il s'interroge déjà sur la loi de l'intervalle qui sépare deux zéros successifs de  $I(t)$  et sur celle du maximum.

De son côté, Kac donne en 1943 l'espérance du nombre de zéros d'un polynôme à coefficients aléatoires

$$X_0 + X_1 x + \dots + X_{n-1} x^{n-1} = 0,$$

où les  $X_i$  sont indépendantes de loi gaussiennes standard.

Cette remarque historique permet de voir que le nombre de solutions d'une équation aléatoire n'est plus un problème récent. Malgré cela, ce type de calculs n'en est pas moins difficile. La formule de Rice est utilisée pour déterminer l'intensité d'événements rares. D'après le chapitre 8, nous savons qu'elle permet de connaître les paramètres des modèles de Poisson. C'est donc pour nous un outil précieux pour l'étude de la loi du maximum d'un processus.

Considérons  $X$  un processus stochastique réel défini sur un intervalle  $B$  dont la longueur est notée  $|B|$ . Rappelons qu'une trajectoire de  $X$  admet un crossing de niveau  $u$  au point  $t$  si dans tout voisinage de  $t$  on peut trouver deux points  $t_1$  et  $t_2$  tels que

$$(X(t_1) - u)(X(t_2) - u) < 0.$$

Le nombre de crossings de  $X$  de niveau  $u$  sur  $B$  est noté  $N_u^X(B)$ . Introduisons le nombre de solutions

$$M_u^X(B) = \#\{t \in B, X(t) = u\}.$$

Clairement nous avons  $M_u^X(B) \geq N_u^X(B)$ .

Deux situations se présentent lorsque  $t$  est solution de  $X(t) = u$  :

- $u$  est un maximum local atteint au point  $t$ ,
- la trajectoire de  $X$  admet un crossing au point  $t$ .

Nous ne nous intéressons ici qu'à des processus n'ayant presque sûrement pas de maximum local égal à un niveau  $u$  fixé préalablement. Autrement dit, les nombres  $N_u^X(B)$  et  $M_u^X(B)$  sont presque sûrement égaux. Citons l'article de Marcus (1977) qui en fait la preuve sous certaines conditions dont l'absolue continuité des trajectoires. Il montre que le premier moment d'une fine modification de la variable  $M_u^X(B)$  est majoré par une expression asymptotiquement égale au premier moment de  $N_u^X(B)$ . Nous donnons maintenant des conditions suffisantes pour obtenir cette égalité presque sûre dans le cas gaussien.

### **Théorème 9.1 (Ylvisaker)**

Soit  $\{X(t), t \in \mathcal{T}\}$  un processus gaussien à valeurs réelles indexé sur un espace topologique  $\mathcal{T}$  séparable et compact. Supposons que

(Y1) les trajectoires de  $X$  sont presque sûrement continues,

(Y2)  $\text{Var}(X(t)) > 0$  pour chaque  $t$  dans  $\mathcal{T}$ .

Alors, pour tout  $u$  dans  $\mathbb{R}$

$$\mathbb{P}((M_u - N_u)(B) > 0) = 0.$$

On appelle formule de Rice, toute formule donnant l'espérance de la variable aléatoire discrète  $N_u^X(B)$ . Remarquons que ce nombre peut se définir lorsque  $X$  est un champ aléatoire défini sur une partie de  $\mathbb{R}^d$  à valeurs dans  $\mathbb{R}^d$  et  $u$  un élément de  $\mathbb{R}^d$ . Notons aussi que la méthode décrite dans ce chapitre est valable pour le calcul des moments relatifs aux upcrossings ou encore aux downcrossings.

• Soit  $\{X(t), t \in B\}$  un processus gaussien stationnaire, centré, de fonction de covariance  $\rho$  vérifiant  $\rho(0) = 1$ , à trajectoires presque sûrement continues. La fonction  $\rho$  étant continue avec  $\rho(0) \neq 0$ , nous savons qu'elle s'écrit d'après le théorème de Bôchner comme

$$\rho(s - t) = \int_{\mathbb{R}} \cos[(s - t)\lambda] dF(\lambda)$$

pour tout  $s$  et  $t$  dans  $B$ , où  $F$  est une mesure sur  $\mathbb{R}$  de masse totale égale à  $\rho(0)$ .  $F$  est appelée la mesure spectrale du processus. Lorsque l'intégrale

$$\int_{\mathbb{R}} \lambda^k dF(\lambda)$$

existe, nous notons  $\lambda_k$  sa valeur et l'appelons  $k$ ième moment spectral. Dans ce cadre et même lorsque  $\lambda_2$  n'est pas finie, nous avons d'après Ylvisaker (1965)

$$(9.1) \quad \mathbb{E}(N_u^X(B)) = \frac{\sqrt{\lambda_2}}{\pi} e^{-u^2/2} |B|.$$

Ainsi c'est dans l'article de Rice (1945) que l'on trouve pour la première fois une telle formule. Ce n'est que durant les années soixante que ce problème est résolu dans un cadre plus large. Voici une synthèse de l'évolution des hypothèses requises :

- Ivanov (1960) avec  $\int_{\mathbb{R}} \lambda^4 dF(\lambda) < +\infty$ ,
- Bulinskaya (1961) lorsque  $X$  admet une dérivée continue, par exemple ceci est impliqué par  $\int_{\mathbb{R}} \lambda^2 \log(1 + |\lambda|)^{1+\delta} dF(\lambda) < +\infty$ ,  $\delta > 0$ ,
- Itô (1964) et Ylvisaker (1965) dans le cadre décrit pour (9.1).
- Cramér and Leadbetter (1965) donnent les moments d'ordres supérieurs dans le cas gaussien stationnaire,
- Leadbetter (1966) s'intéresse au cas non-gaussien et non-stationnaire,
- Marcus (1977) considère un processus stochastique non nécessairement gaussien, non-stationnaire à trajectoires absolument continues ; il étudie aussi les moments d'ordres supérieurs,
- Wschebor (1985) décrit les conditions générales associées à un processus stochastique à plusieurs paramètres non nécessairement gaussien pour obtenir les moments de tous ordres.

• Considérons maintenant  $\{X(t), t \in B\}$  un processus stochastique à trajectoires presque sûrement absolument continues. On note  $X'(t)$  la dérivée de  $X$  dans le sens de l'absolue continuité. Supposons que  $X(t)$  et  $X'(t)$  admettent une densité jointe  $p_{X(t), X'(t)}(\cdot, \cdot)$ . On note  $p_{X(t)}(\cdot)$  la densité marginale de  $X(t)$ .

Sous des hypothèses données par Marcus (1977) nous savons que pour tout réel  $u$

$$(9.2) \quad \mathbb{E}(N_u^X(B)) = \int_B \mathbb{E}(|X'(t)| | X(t) = u) p_{X(t)}(u) dt.$$

Suivant la régularité du processus, on peut obtenir

$$\mathbb{E}(N_u^X(B)) = \int_B \int_{\mathbb{R}} |x| p_{X(t), X'(t)}(u, x) dx dt.$$

Cette liste de références n'est pas tout à fait exhaustive et d'autre part, la recherche d'extensions de ces formules reste clairement un sujet ouvert. Nous conseillons les articles de Rychlik (2000) et Leadbetter and Spaniolo (2004) pour une présentation plus complète de la bibliographie des formules de Rice.

## 9.2 Résultats analytiques

Nous présentons ici le pendant analytique de ce type de calculs et nous décrivons son utilisation dans un cadre stochastique.

### 9.2.1 Pour un niveau fixé

Soit  $f$  une fonction réelle mesurable définie sur un intervalle compact  $B$  de  $\mathbb{R}$ . On note

$$M_u^f(B) = \#\{t \in B, f(t) = u\}.$$

Le but est de trouver une expression (même implicite) de ce nombre.

**Lemme 9.1 (Kac)**

Supposons que  $f$  est continûment différentiable, avec  $f(t) \neq u$  pour  $t \in \partial B$  et de plus  $\{t \in B, f(t) = u, f'(t) = 0\} = \emptyset$ . Nous avons

$$M_u^f(B) = \lim_{\delta \rightarrow 0} \frac{1}{2\delta} \int_B \mathbb{1}_{\{|f(t)-u| < \delta\}} |f'(t)| dt.$$

Remarquons que la dernière hypothèse signifie qu'il n'y a pas de point tangent au niveau  $u$  donc  $M_u^f(B) = N_u^f(B)$ . La preuve du lemme 9.1 est relativement simple. Elle utilise la compacité pour borner le nombre de solutions; puis, au voisinage de chaque solution, elle s'appuie sur la formule de changement de variable. En fait, la version originale du lemme de Kac donne une formule plus générale sur le même dénombrement avec la condition supplémentaire  $g(t) \in ]c, d]$  pour une fonction mesurable  $g$  définie sur l'intervalle  $B$ .

Si on suppose  $f$  continûment différentiable et  $g(t) = f'(t)$  on peut obtenir une identité pour le nombre de upcrossings ou de downcrossings.

En particulier, le lemme de Kac permet de déduire une borne supérieure pour  $N_u^X(B)$  (quand  $X$  est un processus stochastique dont les trajectoires vérifient presque sûrement les hypothèses du lemme) par application du Lemme de Fatou. On obtient facilement une égalité lorsque  $X$  est un processus gaussien polygonal d'ordre  $k$  car on peut borner  $N_u^X(B)$  par le nombre de fois où le processus change de direction, à savoir  $2^k$ . Cette approximation polygonale est utilisée par exemple dans Ylvisaker (1965) ou Leadbetter (1966).

### 9.2.2 Pour presque tout niveau

Ces résultats peuvent s'obtenir sous des hypothèses plus faibles lorsqu'on ne cherche qu'une égalité presque partout dans (9.2). Cependant, étant donné que l'on diminue les hypothèses nécessaires et que cette égalité est vraie sur un ensemble dense, on aimerait pouvoir l'appliquer à n'importe quel niveau donné. La mise en oeuvre pratique de ce raisonnement est décrite par Rychlik (2000) à travers trois exemples précis dont la prédiction du temps de fissure d'un matériel sous pression ou l'analyse de la hauteur de crête des vagues. Il développe cette idée en faisant soit une modélisation bayésienne sur la position incertaine du niveau  $u$  fixé, soit en montrant que seule l'analyse au-dessus d'un niveau donné est nécessaire.

Commençons par rappeler un résultat analytique.

**Théorème 9.2 (Banach)**

Soit  $f : B \subset \mathbb{R} \rightarrow \mathbb{R}$  une fonction absolument continue sur  $B$ , un intervalle. Pour tout borélien  $I \subset B$ , nous avons

$$\int_{\mathbb{R}} N_u^f(I) du = \int_I |f'(t)| dt.$$

Geman and Horowitz (1973) sont les premiers à appliquer ce résultat aux trajectoires d'un processus stochastique; voici leur résultat.

**Théorème 9.3 (Geman and Horowitz)**

Soit  $\{X(t), t \in [0, 1]\}$  un processus strictement stationnaire admettant une dérivée en moyenne quadratique. Si  $\mathbb{P}(X'(0) = 0) = 0$  alors pour presque tout  $u$  réel

$$\mathbb{E}(N_u^X([0, 1])) = \mathbb{E}(|X'(0)|/X(0) = u)p_{X(0)}(u).$$

Pour le voir on applique le théorème de Banach à  $I = \{t \in B, X(t) \leq x\}$ .

Regardons maintenant une écriture intégrale de ce nombre de solutions lorsque la fonction n'est plus réelle. Il s'agit d'un résultat de Federer (1969) présenté ici dans une version plus faible.

**Théorème 9.4 (Federer)**

Soit  $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$  une fonction Lipschitzienne. Nous avons

$$\int_B h[f(x)] \left| \det \left( f'(x) \right) \right| dx = \int_{\mathbb{R}^d} h(u) M_u^f(B) du$$

où  $B$  est un ensemble mesurable,  $h : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$  et

- soit  $h$  est mesurable,
- soit  $N_u^f(B) < +\infty$  pour presque tout  $u$ ,
- soit  $\mathbb{1}_B \cdot (h \circ f) \cdot |\det(f')|$  est mesurable.

Supposons que  $Z$  est un champ aléatoire dont les trajectoires sont presque sûrement lipschitziennes. Sous des conditions raisonnables, on peut obtenir une formule intégrale pour l'espérance du nombre de solutions de  $Z(t) = u$  dans  $B$ . L'intérêt du théorème de Federer est donc d'avoir

$$\mathbb{E}(N_u^Z(B)) = \int_B \mathbb{E} \left( \left| \det \left( Z'(t) \right) \right| / Z(t) = u \right) p_{Z(t)}(u) dt$$

pour presque tout  $u$ . Pour ces dimensions quelconques, nous référons à l'article de Brillinger (1972), qui est le premier à appliquer ce théorème aux trajectoires d'un champ aléatoire.

Dans le cas gaussien, l'égalité partout peut être obtenue en étudiant la continuité en  $u$  de chacun des deux membres. Pour cela, les formules de régression sont un outils important.



### 9.3 Formule de Rice avec des poids

Dans cette section nous présentons un type d'extension de la formule de Rice. En effet, nous cherchons à modifier le poids attribué à chacune des solutions de l'équation  $Z(t) = u$ . Jusqu'à présent nous avons regardé des éléments de la forme  $\sum_{t \in B \cap Z^{-1}(u)} 1$ . Le problème considéré maintenant est le calcul de l'espérance de

$$\sum_{t \in B \cap Z^{-1}(u)} g(Y^t)$$

dans le cas où  $g$  n'est pas identiquement égale à 1 et  $Y^t$  est un processus stochastique défini sur  $B$  pour tout  $t$  dans  $B$ . Nous commençons par rappeler la formule de Rice obtenue pour des champs gaussiens par Azaïs and Wschebor (2005).

#### **Théorème 9.5 (Azaïs and Wschebor)**

Soit  $Z : B \rightarrow \mathbb{R}^d$  un champ aléatoire défini sur  $B$ , un sous ensemble compact de  $\mathbb{R}^d$ , et soit  $u$  un élément de  $\mathbb{R}^d$ . Supposons que

A0 :  $Z$  est gaussien,

A1 :  $t \mapsto Z(t)$  est presque sûrement continûment différentiable,

A2 : pour tout  $t \in B$ ,  $Z(t)$  a une distribution non dégénérée (i.e.  $\text{Var}(Z(t)) \succ 0$ ),

A3 :  $\mathbb{P}\{\exists t \in \dot{B}, Z(t) = u, \det(Z'(t)) = 0\} = 0$ ,

A4 :  $\lambda(\partial B) = 0$ .

Alors

$$\mathbb{E}(N_u^Z(B)) = \int_B \mathbb{E}(|\det(Z'(t))| / Z(t) = u) p_{Z(t)}(u) dt,$$

où les deux membres sont finis.

En fait le même résultat peut s'écrire pour la somme de poids présentée au début de cette section.

#### **Théorème 9.6**

Soit  $Z : B \rightarrow \mathbb{R}^d$  un champ aléatoire défini sur  $B$ , un sous ensemble compact de  $\mathbb{R}^d$ , et soit  $u$  un élément de  $\mathbb{R}^d$ . Supposons que

H0 : pour tout  $t \in B$ ,  $s \mapsto (Z(s), Y^t(s))$  est gaussien,

H1 :  $t \mapsto Z(t)$  est presque sûrement continûment différentiable,

H2 : pour tout  $t \in B$ ,  $s \mapsto Y^t(s)$  est presque sûrement continue,

H3 : pour tout  $t \in B$ ,  $Z(t)$  a une distribution non dégénérée (i.e.  $\text{Var}(Z(t)) \succ 0$ ),

H4 :  $\mathbb{P}\{\exists t \in \dot{B}, Z(t) = u, \det(Z'(t)) = 0\} = 0$ ,

H5 :  $\lambda(\partial B) = 0$ ,

H6 :  $g : \mathcal{C}(B, \mathbb{R}) \rightarrow \mathbb{R}$  est une fonction continue et bornée, lorsqu'on muni  $\mathcal{C}(B, \mathbb{R})$  de la norme de la topologie de la convergence uniforme  $\|f\|_\infty = \sup_{x \in B} |f(x)|$ .

Alors

$$\mathbb{E} \left( \sum_{t \in B \cap Z^{-1}(u)} g(Y^t) \right) = \int_B \mathbb{E}(|\det(Z'(t))| g(Y^t) / Z(t) = u) p_{Z(t)}(u) dt,$$

où les deux membres sont finis.

La démonstration de ce résultat, donnée maintenant, n'est qu'une adaptation de celle du Théorème 9.5.

*Preuve du Théorème 9.6.* Considérons une fonction continue croissante  $F$  telle que  $F(x) = 0$  pour  $x \leq 1/2$  et  $F(x) = 1$  pour  $x \geq 1$ . Notons  $\lambda_{\min}(Z'(s))$  la plus petite valeur propre de  $Z'(s)$  et  $\omega_{Z'}$  le module de continuité de  $Z'$ . Soit  $\Delta$  et  $\eta$  deux réels positifs pour lesquels nous définissons la fonction aléatoire

$$\alpha_{\Delta,\eta}(u) = F\left(\frac{1}{2\Delta} \inf_{s \in B} [\lambda_{\min}(Z'(s)) + \|Z(s) - u\|]\right) \times \left(1 - F\left(\frac{d}{\Delta} \omega_{Z'}(\eta)\right)\right),$$

et l'ensemble  $B_{-\eta} = \{t \in B : \|t - s\| \geq \eta, \forall s \notin B\}$ .

Si  $\alpha_{\Delta,\eta}(u) > 0$  et si  $N_u^Z(B_{-\eta})$  ne s'annule pas, on peut démontrer que dans chaque boule de diamètre  $\frac{\eta}{2}$  centrée en un point de  $B_{-\eta}$  il y a au plus une solution de l'équation  $Z(t) = u$ . Un argument de compacité permet de dire que  $N_u^Z(B_{-\eta})$  est borné par une constante  $C(\eta, B)$ , qui ne dépend que de  $\eta$  et de l'ensemble  $B$ . En utilisant le fait que  $g$  est bornée, on en déduit que

$$\sum_{t \in B_{-\eta} \cap Z^{-1}(u)} g(Y^t)$$

est borné par une constante  $C(\eta, B, g)$  ne dépendant que de  $\eta$ , de l'ensemble  $B$  et de  $g$ .

Prenons maintenant une fonction déterministe continue  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  à support compact. En utilisant la formule de Federer (1969), rappelée dans la section 9.2, le caractère Lipschitz des trajectoires de  $Z$  et l'intégrabilité de  $f(u)\alpha_{\Delta,\eta}(u)$ , nous avons pour tout sous-ensemble compact  $K$  de  $B$

$$\mu(K) := \int_{\mathbb{R}^d} f(u) N_u^Z(K) \alpha_{\Delta,\eta}(u) du = \int_K f(Z(t)) |\det(Z'(t))| \alpha_{\Delta,\eta}(Z(t)) dt.$$

En regardant  $\mu$  comme une mesure positive sur  $B$  et en calculant de deux manières différentes  $\int_{B_{-\eta}} g(Y^t) \mu(dt)$ , on obtient l'égalité

$$\int_{\mathbb{R}^d} f(u) \sum_{t \in B_{-\eta} \cap Z^{-1}(u)} g(Y^t) \alpha_{\Delta,\eta}(u) du = \int_{B_{-\eta}} f(Z(t)) |\det(Z'(t))| g(Y^t) \alpha_{\Delta,\eta}(Z(t)) dt.$$

Prenons les espérances de chaque membre,

$$\int_{\mathbb{R}^d} f(u) \mathbb{E} \left( \sum_{t \in B_{-\eta} \cap Z^{-1}(u)} g(Y^t) \alpha_{\Delta,\eta}(u) \right) du = \int_{\mathbb{R}^d} f(u) du \int_{B_{-\eta}} \mathbb{E} (|\det(Z'(t))| g(Y^t) \alpha_{\Delta,\eta}(u) / Z(t) = u) p_{Z(t)}(u) dt.$$

On obtient que les deux fonctions

$$(i) : \mathbb{E} \left( \sum_{t \in B_{-\eta} \cap Z^{-1}(u)} g(Y^t) \alpha_{\Delta,\eta}(u) \right)$$

$$(ii) : \int_{B_{-\eta}} \mathbb{E} (|\det(Z'(t))|g(Y^t)\alpha_{\Delta,\eta}(u)/Z(t) = u) p_{Z(t)}(u)dt,$$

coïncident presque partout comme fonctions de  $u$ . Il nous suffit de prouver la continuité de chacune d'elles pour conclure à une égalité sur  $\mathbb{R}^d$ .

Montrons que la fonction (i) est continue au point  $u = u_0$  fixé. Considérons la variable aléatoire à l'intérieur de l'espérance dans (i). Presque sûrement, il n'y a pas de point  $t$  dans  $Z^{-1}(u_0)$  tel que  $\det(Z'(t)) = 0$ . Par le théorème d'inversion locale,  $Z(\cdot)$  est inversible dans un voisinage de chaque point de  $Z^{-1}(u_0)$  et la distance de  $Z(t)$  à  $u_0$  est minorée par un réel strictement positif pour  $t \in B_{-\eta}$  à l'extérieur de cette union de voisinages. Cela implique que presque sûrement, comme fonction de  $u$ ,  $N_u^Z(B_{-\eta})$  est constant dans un voisinage (aléatoire) de  $u_0$ . D'un autre côté, il est clair par sa définition que la fonction  $u \mapsto \alpha_{\Delta,\eta}(u)$  est bornée. Nous pouvons appliquer le théorème de convergence dominée quand  $u \rightarrow u_0$ , puisque

$$\sum_{t \in B_{-\eta} \cap Z^{-1}(u)} g(Y^t)\alpha_{\Delta,\eta}(u)$$

est borné par une constante qui ne dépend pas de  $u$ .

Pour la continuité de (ii), il est suffisant de montrer que, pour chaque  $t$  dans  $B$  l'espérance conditionnelle dans l'intégrand est une fonction continue de  $u$ . Remarquons que la variable aléatoire  $|\det(Z'(t))|g(Y^t)\alpha_{\Delta,\eta}(u)$  est une fonctionnelle de  $\{(Z(s), Y^t(s), Z'(s)) : s \in B\}$ . Écrivons la formule de régression gaussienne par rapport à  $Z(t)$ , il s'agit des expressions

$$\begin{aligned} Z(s) &= U^t(s) + \alpha^t(s)Z(t), \\ Y^t(s) &= V^t(s) + \beta^t(s)Z(t), \\ Z'_j(s) &= U_j^t(s) + \gamma_j^t(s)Z(t), \quad j = 1, \dots, d \end{aligned}$$

où  $Z'_j(s)$  ( $j = 1, \dots, d$ ) décrivent les colonnes de  $Z'(s)$ ,  $U^t(s)$ ,  $V^t(s)$  et  $U_j^t(s)$  sont gaussiens, indépendants de  $Z(t)$  pour chaque  $s \in B$ , et les matrices de régression  $\alpha^t(s)$ ,  $\beta^t(s)$ ,  $\gamma_j^t(s)$  ( $j = 1, \dots, d$ ) sont des fonctions continues des variables  $s$  et  $t$  (en prenant en compte l'hypothèse H3). Par substitution dans l'espérance conditionnelle nous disposons maintenant d'une espérance sans conditionnement. En utilisant le fait que les moments du supremum d'un processus gaussien presque sûrement borné sont finis, la continuité en  $u$  s'en suit par convergence dominée.

Fixons maintenant  $u \in \mathbb{R}^d$  et faisons  $\eta \downarrow 0$ ,  $\Delta \downarrow 0$  dans cet ordre, à la fois dans (i) et (ii). Pour (i) on peut utiliser le théorème de Beppo Levi. Remarquons que presque sûrement

$$\sum_{t \in B_{-\eta} \cap Z^{-1}(u)} g(Y^t) \nearrow \sum_{t \in \dot{B} \cap Z^{-1}(u)} g(Y^t) = \sum_{t \in B \cap Z^{-1}(u)} g(Y^t).$$

D'un autre côté on peut aussi prouver (à l'aide de H4) que presque sûrement

$$\inf_{s \in B} \left[ \lambda_{\min}(Z'(s)) + \|Z(s) - u\| \right] > 0$$

de sorte que le premier facteur du membre de droite de l'expression de  $\alpha_{\Delta,\eta}(u)$  croît vers 1 lorsque  $\Delta$  décroît vers 0. Ainsi par le théorème de Beppo Levi

$$\lim_{\Delta \downarrow 0} \lim_{\eta \downarrow 0} \mathbb{E} \left( \sum_{t \in B_{-\eta} \cap Z^{-1}(u)} g(Y^t) \alpha_{\Delta,\eta}(u) \right) = \mathbb{E} \left( \sum_{t \in B \cap Z^{-1}(u)} g(Y^t) \right).$$

Pour (ii), on peut procéder de la même manière après avoir déconditionné pour obtenir le membre de droite de la formule du théorème 9.6. Pour terminer la preuve, il suffit de remarquer que les calculs gaussiens classiques montrent que ce membre de droite est fini.  $\square$

## Bibliographie du chapitre 9

- Azaïs, J.-M. and Wschebor, M.** (2005). On the distribution of the maximum of a gaussian field with d parameters. *Ann. Appl. Probab.*, 15(1) :254–278.
- Brillinger, D. R.** (1972). On the number of solutions of systems of random equations. *Ann. Math. Statist.*, 43 :534–540.
- Bulinskaya, E. V.** (1961). On the mean number of crossings of a level by a stationary gaussian process. *Teor. Veroyatnost. i Primenen*, 6 :474–478.
- Cramér, H. and Leadbetter, M. R.** (1965). The moments of the number of crossings of a level by a stationary normal process. *Ann. Math. Statist.*, 36 :1656–1663.
- Federer, H.** (1969). *Geometric measure theory*. Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York.
- Geman, D. and Horowitz, J.** (1973). Occupation times for smooth stationary processes. *Ann. Probability*, 1(1) :131–137.
- Itô, K.** (1963/1964). The expected number of zeros of continuous stationary Gaussian processes. *J. Math. Kyoto Univ.*, 3 :207–216.
- Ivanov, V. A.** (1960). On the average number of crossings of a level by the sample functions of a stochastic process. *Teor. Veroyatnost. i Primenen*, 5(5) :319–323.
- Kac, M.** (1943). On the average number of crossings of a level by the sample functions of a stochastic process. *Bull. Amer. Math. Soc.*, 49 :314–320.
- Leadbetter, M. R.** (1966). On crossings of levels and curves by a wide class of stochastic processes. *Ann. Math. Statist.*, 37 :260–267.
- Leadbetter, M. R. and Spaniolo, G. V.** (2004). Reflections on Rice's formulae for level crossings—history, extensions and use. *Aust. N. Z. J. Stat.*, 46(1) :173–180. Festschrift in honour of Daryl Daley.

- Marcus, M. B.** (1977). Level crossings of stochastic process with absolutely continuous sample paths. *Ann. Probability*, 5(1) :52–71.
- Rice, S. O.** (1945). Mathematical analysis of random noise. *Bell Systeme Tech. J.*, 24 :46–156.
- Rychlik, I.** (2000). On some reliability applications of Rice’s formula for the intensity of level crossings. *Extremes*, 3(4) :331–348(2001).
- Wschebor, M.** (1985). *Surfaces aléatoires*, volume 1147 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin. Mesure géométrique des ensembles de niveau. [The geometric measure of level sets].
- Ylvisaker, D.** (1965). The expected number of zeros of a stationary gaussian process. *Ann. Math. Statist.*, 36 :1043–1046.
- Ylvisaker, D.** (1968). A note on the absence of tangencies in Gaussian sample paths. *Ann. Math. Statist.*, 39 :261–262.





# Index des auteurs cités

- Abramowitz and Stegun, 1992, *169*  
Adler and Taylor, 2005, *38–41, 79*  
Adler, 1981, *38, 39, 78, 79, 81, 171*  
Anderson, 1963, *152*  
Azaïs, Cierco and al., 1999, *86*  
Azaïs and Delmas, 2002, *39, 78, 79*  
Azaïs and Mercadier, 2003, *57, 113, 134, 139, 141, 147*  
Azaïs and Wschebor  
    2002, *36, 40, 78, 80, 81, 88*  
    2003, *158, 159, 165*  
    2005, *95, 97, 188*  
Azaïs, Cierco and al., 1999, *65*  
Azaïs, Gassiat and Mercadier, 2004, *62, 79, 85, 111*
- Berman  
    1964, *62*  
    1971, *28, 31, 36, 58, 62, 86*  
Bickel and Chernoff, 1993, *53, 113, 138, 139*  
Boistard, 2003, *113*  
Brillinger, 1972, *187*  
Brodtkorb et al., 2000, *42, 80, 85*  
Bulinskaya, 1961, *185*
- Castillo, 1988, *24*  
Chen and Chen, 2001, *112*  
Chernoff and Lander, 1995, *112, 122*  
Cramér and Leadbetter  
    1965, *185*  
    1967, *19, 31, 58*  
Cressie, 1980, *35*  
Cucker and Wschebor, 2003, *158*  
Cuesta-Albertos and Wschebor  
    2003, *158*  
    2004, *158*
- Dacunha-Castelle and Gassiat  
    1997, *21, 48, 49, 52, 61, 112*  
    1999, *61, 112*  
Delmas  
    2001, *39*  
    2003, *36, 86, 114, 134*  
DeLong, 1981, *35, 92, 93*  
Deloubrière and Hilhorst, 2000, *20*  
Dembo et al., 2002, *89*  
Dobrushin, 1960, *177*  
Durbin, 1985, *91*
- Edelman, 1989, *154, 157, 158*  
Ehrhardt et al., 2002, *20, 79*
- Federer, 1969, *79, 97, 187, 189*  
Finkenstädt and Rootzén, 2004, *21, 24*  
Fisher and Tippett, 1928, *24*  
Fréchet, 1927, *23*
- Garel, 2001, *112, 114*  
Gassiat, 2002, *45, 49, 50, 52, 79, 85, 112–117, 139*  
Geman and Horowitz, 1973, *153, 187*  
Genz, 1992, *42, 85*  
Ghosh and Sen, 1985, *45, 48, 85, 112*  
Gnedenko, 1943, *24*  
Goffinet, Loisel and Laurent, 1992, *51, 112, 114*  
Goldman, 1971, *35*
- Hüsler, 1990, 1995, *29*  
Hall and Stewart, 2004, *113*  
Hartigan, 1985, *113*
- Itô, 1963, *185*  
Ivanov, 1960, *185*
- James et al., 2001, *112*  
Johansson, 2000, *153*  
Johnstone and Lu, 2004, *155*  
Johnstone, 2001, *153, 155*
- Kac, 1943, *183*  
Kallenberg, 1973, *62, 178*  
Khintchine, 1960, *177*  
Krishnaiah and Schuurmann, 1974, *156*
- Leadbetter and Spaniolo, 2004, *185*  
Leadbetter, 1966, *185, 186*



- Leadbetter, Lindgren and Rootzén, 1983,  
24, 27, 31, 58, 60, 68
- Ledoit and Wolf, 2002, 155, 156
- Lemdani and Pons  
1997, 112  
1999, 112
- Li and Shao  
2002, 58, 62, 89, 113, 138, 139, 146  
2004, 89
- Lindsay, 1995, 46, 112
- Liu and Shao, 2004, 52, 89, 113, 114,  
138
- Loisel et al., 1994, 51
- Marcus, 1977, 184, 185
- Marčenko and Pastur, 1967, 153
- McKean, 1963, 35
- McLachlan and Peel, 2000, 112, 114
- Mercadier, 2005, 77, 85, 92, 100, 114
- Molchan and Khokhlov, 2004, 20, 79
- Mosler and Seidel, 2001, 112
- Muirhead, 1982, 152, 156
- Paul, 2004, 155
- Piterbarg and Konakov, 1995, 42, 58
- Piterbarg, 1996, 38, 39, 78, 79
- Plackett, 1954, 62
- Resnick, 1987, 24
- Rice, 1945, 183, 185
- Rychlik  
1987, 37, 40, 80, 83  
2000, 185, 186
- Rényi, 1967, 178, 181
- Seidel, Mosler and Alker, 2000, 114
- Shepp  
1971, 34  
1979, 35
- Shepp and Slepian, 1976, 35
- Shorack and Wellner, 1986, 137
- Silverstein, 1985, 153
- Slepian  
1961, 34  
1962, 62, 89
- Soshnikov, 2002, 153
- Sugiyama, 1970, 156
- Sun, 1993, 38, 78
- Takemura and Kuriki, 2002, 38, 78
- Taylor et al., 2004, 79, 81
- Thode et al., 1988, 51
- Titterington et al., 1985, 46, 112
- van der Vaart, 1998, 47, 48, 53, 113,  
114, 136, 138
- Volkonskiĭ and Rozanov, 1961, 30, 31,  
58
- Wachter, 1978, 153
- Waikar and Schuurmann, 1973, 156
- Worsley, 1995, 38, 79
- Wschebor  
1985, 31, 58, 185  
2004, 158, 160, 167
- Ylvisaker  
1965, 184–186  
1968, 84, 97, 99, 184

# Summary of the thesis

On the one hand, the interdisciplinary connections of Statistics are increasingly numerous and more and more varied. They can either play a motivation part in the introduction of models or serve as applications fields. The list of these domains contains finance, insurance, genetics, informatics, petroleum research, ... On the other hand, Statistics are a tool of formalism where the set of models is very extensive. This mathematics speciality, like any other one, is growing richer also by itself.

My thesis work, described through this manuscript, is a short illustration of this double expansion. Indeed, I present theoretical results on *statistics of processes* or *testing theory* as well as numerical methods that answer physical or genetical problems.

Many physical problems are continuous by nature and lead to a particular mathematical modelling : the continuous time stochastic processes. A large variety of characteristics of such processes are of statistical interest.

PART II : DISTRIBUTION OF MAXIMUM OF PROCESSES (page 57)

The second part of this document reports the analysis of the distribution of the global maximum (and in an analogical way, that of the minimum). It precisely contains two different works.

## - Chapter 4, summarized in Section 2.1.2 -

In collaboration with J.-M. Azaïs, we extended a result of weak convergence of the maximum on  $[0, t]$  of a process, denoted by  $M_t$ . The contribution of this work is the non-stationary character of the Gaussian process. Under a reasonable set of hypotheses, we prove the existence and clarify the expression of constants  $a_t$  and  $b_t$  such that  $a_t(M_t - b_t)$  admits the Gumbel law as an asymptotic distribution.

In short, by a simple change of variable, we have an approximation of the law of the maximum  $M_t$  which is accurate for  $t$  large enough. This work is part of

the *theory of extremes* (introduced in Section 2.1.1).

This result is obtained as a direct consequence of the weak convergence of the upcrossings point process (suitably normalized) to a standard Poisson process.

**- Chapter 5, summarized in Section 2.2.3 -**

The second work gives the distribution of the maximum of a process defined on a compact set, by means of an implicit formula. Unlike traditional studies which enumerate the upcrossings, only existence of an upcrossing has been retained here. This method is developed for one and two parameter processes. In this second case, I obtain the first upper bound for the survival function of a Gaussian, unit-speed process. Its expression is explicit and its computation does not need the use of numerical tools.

The practical implementation of the general method requires the discretization of the definition space on several occasions. It led to the creation of a toolbox whose functions make it possible to compute an upper or lower bound of the distribution of the maximum.

Finally, we make a comparative study between these new estimations and known elements such as the bound of Rice, the series of Rice or the use of the Euler characteristic (presented in Section 2.2.2).

These studies belonging to the domain of theoretical statistics apply to different environmental problems. The result presented in Chapter 4 enables the computation of the probability that a concentration exceeds a critical level at least once on a large interval of time. The estimations deduced from the method given in Chapter 5 constitute a precious help for the determination of critical zones in different examples of testing theory.

PART III : STATISTICAL APPLICATIONS (page 111)

This is illustrated through the study of the *mixture models* which are for instance used in genetics to assert the existence of a major gene in the expression of a qualitative feature.

**- Chapter 6, summarized in Sections 3.3 et 3.4 -**

In collaboration with É. Gassiat and J.-M. Azaïs, we summarize and complete the asymptotic properties of the test based on the statistics of maximum likelihood in mixture models (introduced in Section 3.1). Moreover, this test makes it possible to determine the value of the number of components in the mixture.

In a first part, we go back over the asymptotic distribution of this statistics. In the considered framework, it relies on a functional of the maximum of one or several Gaussian processes. For different usual laws, we minimize the set of necessary hypotheses and when possible, we reduce the asymptotic expression

to a functional depending only on one Gaussian process.

In a second time, we are interested by models of infinite horizon or in others words, by models whose parameters are not a priori bounded. We again characterize the asymptotic law but also the asymptotic power of this statistics.

### - Chapter 7 -

In this chapter I expose a summary of those results on Wishart matrices which are relevant to my motivations. Only extreme eigenvalues are retained in order to construct a *test of sphericity* based on the ratio of the largest and the smallest eigenvalue.

It also deals with a preliminary study about the possible extensions of related results on this ratio. On this account, we give some upper bounds of marginal densities or of joint density of variables appearing in this ratio.

Unfortunately some theoretical results are available only in the complex case, what puts a brake on the success or the hope of success of such a construction.

### PART IV : ANNEX RESULTS (page 177)

The studies investigated in this thesis rely, for a certain part, on some background material. In Chapter 8 we recall some characterizations of the Poisson process. Then in Chapter 9 we give a bibliographic description of the Rice method. The latter contains an extension of the Rice formula for a Gaussian field in which we assign a weight to each crossing.





**Auteur :** Cécile Mercadier.

**Titre de la thèse :** Extrema de processus stochastiques.

Propriétés asymptotiques de tests d'hypothèses.

**Directeur de thèse :** M. Jean-Marc AZAÏS.

**Soutenance :** Le 1er juillet 2005 à l'Université Paul Sabatier, Toulouse III.

---

## Résumé

Cette thèse se divise en deux parties.

La première partie s'inscrit dans la lignée des résultats composant la théorie des valeurs extrêmes. Ces analyses se destinent au calcul de probabilité des événements rares. Le premier travail donne l'ordre asymptotique du maximum d'un processus gaussien, non-stationnaire à variance constante. Le second travail caractérise la loi du maximum en temps fini, et donc pour des niveaux de tous ordres. La procédure d'estimation a d'ailleurs donné naissance à une boîte à outils Matlab appelée MAGP. La seconde partie regroupe deux applications statistiques. D'une part, la distribution et la puissance du test, basé sur le maximum de vraisemblance, sont étudiées pour les modèles de mélange. D'autre part, la construction d'un test de sphéricité est envisagée à l'aide des valeurs propres extrêmes des matrices de covariance.

## Abstract

The following thesis falls into two parts.

The first part is concerned with the extreme value theory. This domain intends to compute the probability of rare events. The first work gives the asymptotic order of the maximum of a non-stationary Gaussian process with constant variance. The second work characterizes the law of the maximum in finite time and consequently for all sorts of levels. Moreover, the estimating procedure led to the creation of a Matlab toolbox called MAGP.

The second part contains two statistical applications. On the one hand, the distribution and power of the likelihood ratio test statistics are studied for mixture models. On the other hand, the construction of a test of sphericity is analyzed with extreme eigenvalues of covariance matrices.

---

**Discipline :** Mathématiques Appliquées.

**Spécialité :** Statistique.

**Mots clés :** Théorie des extrêmes - Processus gaussiens - Processus de Poisson - Loi de Gumbel - Modèles de mélange - Maximum de vraisemblance - Puissance asymptotique - Matrices de Wishart - Test de sphéricité.

**Laboratoire de Statistique et Probabilités**  
**UFR MIG - UMR CNRS 5883**  
**Université Paul Sabatier, Toulouse III**