Un théorème limite conditionnel. Applications à l'inférence conditionnelle et aux méthodes d'Importance Sampling.

Virgile Caron

UPMC SORBONNE UNIVERSITÉS

École Doctorale de sciences mathématiques de Paris Centre

Thèse de doctorat

Spécialité : Mathématiques
Option : Statistique

présentée par

Virgile Caron

Un théorème limite conditionnel.

Applications à l'inférence conditionnelle et aux méthodes d'Importance Sampling.

---

sous la direction du Professeur : MICHEL BRONIATOWSKI

Soutenue le 16 Octobre 2012 devant le jury composé de :

M. Eric MOULINES	Télécom ParisTech	président
M. François LE GLAND	INRIA Rennes	rapporteur
M. Christian P. ROBERT	Université Paris Dauphine	rapporteur
M. Gérard BIAU	Université Paris VI	examinateur
M. Gilles PAGES	Université Paris VI	examinateur
M. Michel BRONIATOWSKI	Université Paris VI	directeur

Laboratoire de Statistique Théorique et Appliquée (LSTA)
Université Pierre et Marie Curie - Paris 6
Tour 15-25, 2ème étage
4 place Jussieu
75252 Paris cedex 05

École doctorale de sciences mathématiques de Paris centre Case 188
4 place Jussieu
75 252 Paris cedex 05

Autant laisser ça à des fous, ça leur éviterait le désagrément de le devenir.

Remerciements

Au moment où s'achève ma thèse, je tiens à exprimer ma profonde reconnaissance à Michel Broniatowski, qui l'a dirigée. Ce fut un grand privilège d'apprendre avec lui et je le remercie vivement pour ses conseils avisés, sa rigueur scientifique, sa disponibilité et son soutien bienveillant.

François Le Gland et Christian Robert ont accepté d'être les rapporteurs de ce travail. J'imagine le travail que représente la lecture d'une thèse et les en remercie sincèrement.

J'adresse mes sincères remerciements à Monsieur Eric Moulines d'avoir accepté de présider mon jury de thèse, à Messieurs Gérard Biau et Gilles Pagès pour l'intérêt qu'ils ont accordé à mon travail et pour avoir accepté de participer au jury.

J'adresse également mes remerciements à Paul Deheuvels, directeur du LSTA, pour m'avoir accueilli dans ce laboratoire au sein duquel j'ai disposé de conditions idéales pour écrire cette thèse.

Un grand merci à toute l'équipe administrative du laboratoire, Louise Lamart, Anne Durrande et Corinne Van Lieberghe pour leur gentillesse et leurs compétences. Parce que la recherche en mathématiques ne se conçoit pas sans bibliothèque, merci à Pascal Epron.

L'ambiance chaleureuse et la solidarité de l'équipe des doctorants du LSTA ont été importantes pour moi au cours de cette thèse. Merci à Zhansheng, Sarah, Manu, Olivier, Assia. Un grand merci aussi à tous les doctorants qui ont refusé de me donner cinq euros pour voir leur nom apparaître ici. Merci à tous mes amis pour leurs encouragements et tous ces bons moments partagés. Enfin, je voulais remercier ma famille qui a toujours cru en moi. Leur soutien permanent et irremplaçable m'a permis de surmonter les moments difficiles.

Pour finir, je remercie ma femme Emilie, l'amour de ma vie, qui a été à mes côtés durant toutes ces années. 1052019135

Résumé

Cette thèse présente une approximation fine de la densité de longues sous-suites d'une marche aléatoire conditionnée par la valeur de son extrémité, ou par une moyenne d'une fonction de ses incréments, lorsque sa taille tend vers l'infini. Dans le domaine d'un conditionnement de type grande déviation, ce résultat généralise le principe conditionnel de Gibbs au sens où il décrit les sous suites de la marche aléatoire, et non son comportement marginal. Une approximation est aussi obtenue lorsque l'événement conditionnant énonce que la valeur terminale de la marche aléatoire appartient à un ensemble mince, ou gros, d'intérieur non vide. Les approximations proposées ont lieu soit en probabilité sous la loi conditionnelle, soit en distance de la variation totale. Deux applications sont développées ; la première porte sur l'estimation de probabilités de certains événements rares par une nouvelle technique d'échantillonnage d'importance ; ce cas correspond à un conditionnement de type grande déviation. Une seconde application explore des méthodes constructives d'amélioration d'estimateurs dans l'esprit du théorème de Rao-Blackwell, et d'inférence conditionnelle sous paramètre de nuisance ; l'événement conditionnant est alors dans la gamme du théorème de la limite centrale. On traite en détail du choix effectif de la longueur maximale de la sous suite pour laquelle une erreur relative maximale fixée est atteinte par l'approximation ; des algorithmes explicites permettent la mise en oeuvre effective de cette approximation et de ses conséquences.

Mots-clefs

Principe de Gibbs ; marche aléatoire conditionnée ; grande déviation ; moyenne déviation ; Importance sampling ; inférence conditionnelle ; Théorème de Rao-Blackwell ; familles exponentielles ; paramètre de nuisance.

Abstract

This thesis presents a sharp approximation of the density of long runs of a random walk conditioned on its end value or by an average of a functions of its summands as their number tends to infinity. In the large deviation range of the conditioning event it extends the Gibbs conditional principle in the sense that it provides a description of the distribution of the random walk on long subsequences. An extension for the approximation of the conditional density in the multivariate case is provided. Approximation of the density of the runs is also obtained when the conditioning event states that the end value of the random walk belongs to a thin or a thick set with non void interior. The approximations hold either in probability under the conditional distribution of the random walk, or in total variation norm between measures. Application of the approximation scheme to the evaluation of rare event probabilities through Importance Sampling is provided. When the conditioning event is in the zone of the central limit theorem it provides a tool for statistical inference in the sense that it produces an effective way to implement the Rao-Blackwell theorem for the improvement of estimators; it also leads to conditional inference procedures in models with nuisance parameters. An algorithm for the simulation of such long runs is presented, together with an algorithm determining the maximal length for which the approximation is valid up to a prescribed accuracy.

Keywords

Gibbs principle; conditioned random walk; large deviation; moderate deviation; simulation; Importance sampling; conditional inference; Rao-Blackwell Theorem; exponential families; nuisance parameter.

Table des matières

Introduction **11**

1 Long runs under a conditional limit distribution. **17**
1.1 Context and scope . . . 17
1.2 Random walks conditioned on their sum . . . 19
1.3 Random walks conditioned by a function of their summands . . . 25
1.4 Conditioning on large sets . . . 40
1.5 Appendix . . . 45

2 Towards zero variance estimators for rare events probabilities **59**
2.1 Introduction . . . 59
2.2 Adaptive IS Estimator for rare event probability . . . 61
2.3 Compared efficiencies of IS estimators . . . 62
2.4 Algorithms . . . 68
2.5 Simulation results . . . 69
2.6 A comparison study . . . 70

3 Conditional inference in parametric models. **77**
3.1 Introduction and context . . . 77
3.2 The approximate conditional density of the sample . . . 78
3.3 Sufficient statistics and approximated conditional density . . . 82
3.4 Exponential models with nuisance parameters . . . 86

4 Extension au cas d-dimensionnel **93**
4.1 Introduction . . . 93
4.2 Hypothèses et Notations . . . 93
4.3 Marche aléatoire d-dimensionnelle conditionnée par sa somme. . . . 98
4.4 Extension à un cas plus général de conditionnement . . . 102
4.5 Comment choisir k ? . . . 104
4.6 Exemples de Trajectoires dans un cas simple . . . 106
4.7 Familles exponentielles courbes et conditionnement. . . . 110
4.8 Démonstration du Théorème 1 du Chapitre 4. . . . 118

A Exemples de simulation **137**
A.1 Loi Normale. . . . 137
A.2 Loi Gamma . . . 139
A.3 Loi Weibull . . . 141
A.4 Loi de Gumbel . . . 141
A.5 Loi Inverse-Gaussienne . . . 141

Bibliographie **143**

Introduction

Cette thèse présente un résultat de nature probabiliste, en lien avec le principe conditionnel de Gibbs, et deux applications. La première est liée à l'évaluation numérique de probabilités d'événements rares. La seconde est de nature directement statistique et explore des méthodes constructives d'amélioration d'estimateurs dans l'esprit du théorème de Rao-Blackwell, et d'inférence conditionnelle sous paramètre de nuisance, dans la lignée des travaux de l'école danoise.

Approximations de lois conditionnelles

Notons $\mathbf{X}_1^n = (\mathbf{X}_1, ..., \mathbf{X}_n)$, n copies i.i.d. d'une variable aléatoire $\mathbf{X}$ de densité p et de loi P sur $\mathbb{R}$ et $\mathbf{S}_{1,n}$ leur somme, $\mathbf{S}_{1,n} := \mathbf{X}_1 + ... + \mathbf{X}_n$. On considère une contrainte sur $\mathbf{S}_{1,n}$ de la forme $(\mathbf{S}_{1,n} = na_n)$ où la suite a_n est supposée convergente. Pour $k = k_n < n$, on s'intéresse à la densité du vecteur $\mathbf{X}_1^k = (\mathbf{X}_1, ..., \mathbf{X}_k)$ sous la contrainte locale $(\mathbf{S}_{1,n} = na_n)$ pour divers choix de la suite a_n et de la taille k_n du vecteur $\mathbf{X}_1^k$. Ensuite nous considérons des contraintes globales de la forme $(\mathbf{S}_{1,n} \in nA)$ où A est un ensemble borélien de $\mathbb{R}$ d'intérieur non vide. Les contraintes sont ensuite remplacées par des conditions de la forme $(\mathbf{U}_{1,n} = na_n)$ et $(\mathbf{U}_{1,n} \in nA)$ où $\mathbf{U}_{1,n} := u(\mathbf{X}_1) + ... + u(\mathbf{X}_n)$ et u est une fonction à valeurs réelles telle que $u(\mathbf{X}_1)$ admette une densité. Enfin, on considère le cas où les $\mathbf{X}_i$ sont des vecteurs aléatoires de $\mathbb{R}^d$ et on s'intéresse à la densité du vecteur $\mathbf{X}_1^k$ sous une contrainte locale.

Cette classe de problèmes conditionnels limites porte le nom de *principes conditionnels de Gibbs* et trouve sa source dans l'étude de systèmes thermodynamiques développée à la fin du 19ème siècle et au début du 20ème siècle par Boltzmann, puis par Gibbs, sous le nom de lois microcanoniques lorsque $k = 1$. Des approches directement probabilistes ont été développées dans les années 1970 en lien avec la théorie des grandes déviations, et développées, entre autres, dans le cadre du théorème de Sanov par [26], et par de nombreux auteurs.

Le contexte de notre étude est caractérisé par le fait que la longueur k du vecteur $\mathbf{X}_1^k$ peut être très grande, puisque nous imposons les conditions

$$0 \leq \limsup_{n\to\infty} k/n \leq 1 \tag{K1}$$

et

$$\lim_{n\to\infty} n - k = \infty. \tag{K2}$$

La suite a_n est supposée converger soit vers $E\mathbf{X}_1$ (resp. vers $Eu(\mathbf{X}_1)$) ou vers une valeur différente. Ce second cas est définit une contrainte locale de type "grande déviation" ; le premier cas peut correspondre à plusieurs situations distinctes : moyenne déviation si $\sqrt{n}(a_n - Eu(\mathbf{X}_1)) \to \infty$, ou relevant du domaine du théorème de la limite centrale si $\sqrt{n}(a_n - Eu(\mathbf{X}_1)) = 0(1)$.

Contexte bibliographique succint

Quand k est indépendant de n (typiquement $k = 1$), les résultats que nous présentons sont des généralisations du *Principe Conditionnel de Gibbs* qui a été étudiée pour a_n fixé, différent de $E\mathbf{X}_1$, sous une condition de grande déviation. Diaconis et Freedman [33] ont considéré ce problème dans un cas un peu plus général, $k/n \to \theta$ pour $0 \leq \theta < 1$, en lien avec le théorème de de Finetti pour des suites finies échangeables (voir [88] pour détails). Leur intérêt est lié à l'approximation de la densité de $\mathbf{X}_1^k$ par le produit des densités des variables aléatoires $\mathbf{X}_i$, c'est à dire à la permanence de l'indépendance asymptotique sous le conditionnement local de la somme des variables.

Leur résultat, dans l'esprit de [93], est à mettre en parallèle avec le travail de [26] sur l'indépendance conditionnelle asymptotique lorsque l'événement conditionnant est $(\mathbf{S}_{1,n} > na)$ où $a > E\mathbf{X}$. Dembo et Zeitouni [29] considèrent des problèmes semblables.

En physique statistique, de nombreux articles se rapportant aux propriétés structurelles des polymères traitent de l'approximation de la densité conditionnelle ; voir [30] et [31]. Dans le cas des moyennes déviations, Ermakov [44] considère aussi un problème similaire pour $k = 1$.

Le cas d'un conditionnement de type central limite (lorsque $\sqrt{n}\,(a_n - Eu\,(\mathbf{X}_1)) = 0(1)$) a reçu peu d'attention dans la bibliographie liée à l'approximation des lois conditionnelles, puisque asymptotiquement la loi conditionnelle de $\mathbf{S}_{1,k}$ est de l'ordre de celle sans conditionnement (voir par exemple [94]), lorsque k ne dépend pas de n. Cependant cette question s'avère d'une grande importance pour l'inférence statistique. Nous indiquons brièvement quelques approches dans ce domaine ; le chapitre 3 présente une bibliographie plus étoffée. Pedersen [76] approxime la densité conditionnelle, non pas, du vecteur $\mathbf{X}_1^k$ mais directement la densité de la statistique de test conditionnellement à une contrainte locale. L'intégration cette approximation lui permet d'obtenir des valeurs critiques (voir [9] ou [75]). D'autres approches ont été récemment proposées (voir [71], [42], [67] et [68]). Leur but n'est pas d'obtenir une approximation de la densité conditionnelle, mais de développer des algorithmes permettant directement la simulation d'échantillon sous la densité conditionnelle. Ces méthodes algorithmiques sont valides sous certains modèles, et non valides sous d'autres. Elles ne permettent pas d'obtenir une méthode unifiée pour les distributions continues les plus courantes. Pour chaque loi étudiée, une nouvelle méthode doit être fabriquée.

Méthodes

Dans son livre "Saddlepoint approximation", Jensen [59] commence son chapitre 1 par la comparaison des trois méthodes les plus classiques pour approximer la densité d'une somme de variables aléatoires indépendantes de même loi :

1. l'approximation résultant de l'utilisation de la densité normale de même espérance et variance que la densité de départ
2. l'utilisation de l'approximation d'Edgeworth améliorant la méthode précédente
3. l'approximation de point selle, qui utilise une famille de lois conjuguées remplaçant l'approximation normale directe, liée ensuite à l'utilisation de développements d'Edgeworth.

En comparant sur un exemple ces trois méthodes, il remarque que l'approximation d'Edgeworth améliore la première méthode au mode de la densité seulement, mais pas aux endroits où la densité est faible, alors que l'approximation point selle est bonne partout. L'approximation point selle est basée sur la relation entre la densité de départ et

la densité conjuguée (ou *tiltée*) qui est obtenue en multipliant la densité de départ par $\exp(tx)$, pour un t appartenant au support de P, et en la renormalisant. Ce que Jensen nomme "classical saddlepoint approximation" dans son chapitre 2, est une méthode d'approximation effectuée en deux étapes. Il relie, tout d'abord, la densité de départ à sa densité conjuguée, puis il utilise un lemme fondamental, qui énonce la commutativité de la conjugaison et de la convolution ; l'approximation d'Edgeworth est appliquée à la convolution des conjuguées, ce qui fournit finalement l'approximation de la densité de la somme normalisée des variables initiales.

Pour l'obtention de l'approximation de la densité conditionnelle proposée au Chapitre 1, nous appliquons une méthode utilisant des ingrédients semblables. En effet, la démonstration principale repose sur plusieurs éléments clés en plus de celui dont nous venons d'évoquer. Pour obtenir cette approximation, sont utilisés dans la preuve : la formule de Bayes, un lemme d'invariance sous la densité tiltée et des développements d'Edgeworth ainsi qu'un contrôle précis des caractéristiques des chemins typiques. La méthode de démonstration s'applique de façon identique quel que soit le type de conditionnement. Des différences importantes apparaitraient si l'on s'intéressait au cas où $a_n \to \infty$, hors du contexte de cette thèse.

Amélioration proposée

Le Chapitre 1 présente un théorème d'approximation de la loi conditionnelle d'une longue sous trajectoire. Ce théorème est ensuite travaillé dans divers contextes, menant à des versions diverses. En effet, un changement dans la structure de dépendance des $\mathbf{X}_i$ sous le conditionnement, quand (K2) et (K1) sont satisfaites, est exhibé et une solution constructive pour le schéma d'approximation est fournie, à l'aide d'algorithmes et d'exemples. La densité approximante est obtenue comme une version légèrement modifiée du tilting classique, combiné à un changement dans la variance. Les résultats améliorent les résultats mentionnés puisque ils procurent une approximation plus fine de la densité conditionnelle même quand k est très proche de n. Le résultat présenté coïncide dans le cas gaussien avec la loi conditionnelle exacte sur toute la trajectoire ; en ce sens le résultat obtenu est optimal. Dans le domaine d'un conditionnement de type grande déviation, ce résultat généralise le principe conditionnel de Gibbs au sens où il décrit les sous suites de la marche aléatoire, et non son comportement marginal.

L'aspect crucial de notre résultat est le suivant. L'approximation de la densité de $\mathbf{X}_1^k$ conditionnée à $(\mathbf{U}_{1,n} = na_n)$ n'est pas obtenue sur $\mathbb{R}^k$ tout entier, mais plutôt sur une suite de sous-ensemble de $\mathbb{R}^k$ qui contient les trajectoires simulées sous la densité conditionnelle (de même, pour les trajectoires simulées sous son approximation) avec probabilité tendant vers 1 quand n tend vers l'infini. Ainsi, l'approximation est obtenue sur les *chemins typiques*. Les raisons, pour lesquelles nous considérons l'approximation de cette manière sont de deux sortes.

1. En inférence conditionnelle, l'approximation doit être valide sur les trajectoires simulées sous la densité conditionnelle, tandis que, pour des applications en Importance Sampling, elle doit être valide sur les trajectoires simulées sous la densité approximante.
2. Un nombre important de conditions techniques nécessaires pour obtenir une approximation sur tout l'ensemble $\mathbb{R}^k$ et centrales dans les articles mentionnés ci-dessus est ainsi évité. Ces conditions portent sur la fonction caractéristique de la densité p de $\mathbf{X}$ pour obtenir une bonne approximation sur tout $\mathbb{R}^k$. Puisque l'approximation conditionnelle est obtenue uniquement sur les chemins typiques sous la densité condi-

tionnelle de $\mathbf{X}_1^k$ et sous le conditionnement ponctuel, les régions de $\mathbb{R}^k$ ainsi atteintes sont mieux connues. Cependant, même sous ces hypothèses plus faibles, un résultat d'approximation sur $\mathbb{R}^k$ tout entier est obtenu. En effet, la convergence de l'erreur relative sur les grands ensembles implique que la distance en variation totale entre la loi conditionnelle et son approximation tend vers 0 sur $\mathbb{R}^k$ tout entier. Ce résultat généralise ceux de [33] et [29], portant sur le cas où k est petit devant n. Néanmoins, notre résultat ne procure pas de vitesse de convergence.

Application au calcul de probabilités d'événements rares

Le résultat d'approximation présenté dans le premier chapitre permet, quand l'événement conditionnement est de type moyenne ou grande déviation, de développer des méthodes d'Importance Sampling ; voir [20] pour un aperçu des techniques classiques. On s'intéresse à l'estimation de $P_n = P[\mathbf{U}_{1,n} \in nA]$ pour n grand mais fixé.

Le cas d'un ensemble A d'intérieur non vide est traité en détail lorsqu'il est de la forme (a, ∞) puisqu'il apparaît naturellement dans les méthodes d'Importance Sampling liées à l'évaluation des probabilités d'événements rares. On aborde également l'étude de conditionnements définis par des ensembles A de dimension locale inférieure à 1 en leur infimum essentiel.

L'estimateur d'Importance Sampling pour P_n sous la densité d'échantillonage g sur $\mathbb{R}^n$ est défini par

$$P_g^{(n)}(A) := \frac{1}{L} \sum_{l=1}^{L} \frac{\prod_{i=1}^{n} p\left(Y_i(l)\right)}{g\left(Y_1^n(l)\right)} \mathbb{1}_{nA}\left(U_{1,n}(l)\right)$$

où les L échantillons $Y_1^n(l) := (Y_1(l), ..., Y_n(l))$ sont i.i.d. simulés sous g. On sait que l'estimateur "parfait" (à variance nulle) correspond à un choix de la densité d'échantillonnage g coïncidant précisément avec la densité de $\mathbf{X}_1^n$ conditionné à l'événement $(\mathbf{U}_{1,n} \in nA)$. Grâce à l'approximation obtenue dans le chapitre 1, une densité d'échantillonnage qui approxime cette densité optimale sur les k premières composantes quand k satisfait (K1) est présentée. On présente un algorithme et une argumentation analytique pour le choix maximal de k_n. Des simulations illustrent le gain en variance de l'estimateur de la probabilité de l'événement rare. De façon plus importante on démontre que la variance de cet estimateur est très fortement réduite (au moins théoriquement), par rapport aux méthodes standard. On montre en effet que le gain relatif en termes de nombres de simulations nécessaires pour obtenir une erreur relative de l'ordre de $\alpha\%$ pour $P(\mathbf{S}_{1,n} > na_n)$ diminue d'un facteur $\sqrt{n-k}/\sqrt{n}$ par rapport au schéma classique d'Importance Sampling. De plus, on montre numériquement sur des exemples que le schéma d'approximation proposé augmente la stabilité du facteur d'importance et que le taux de succès (proportion du nombre de fois où la cible est atteinte par l'échantillon simulé) est proche de 100%. Des algorithmes précis et utilisables sont présentés.

Inférence conditionnelle

Basu [10] propose différentes méthodes d'inférence conditionnelle pour des modèles incluant un paramètre de nuisance. Le conditionnement de la loi de l'échantillon par une statistique exhaustive de ce paramètre (si elle existe) est la méthode la plus naturelle, au moins en théorie. Cependant, deux problèmes majeurs apparaissent.

1. La densité conditionnelle est, en général, difficile (voire impossible) à obtenir.

2. En utilisant ces méthodes, la densité approximante obtenue est toujours dépendante du paramètre de nuisance, quoique la loi conditionnelle ne le soit pas.

On montre que lorsque la statistique observée est exhaustive pour un paramètre, l'approximation de la densité conditionnelle de l'échantillon (ou plutôt d'un grand sous échantillon) relativement à cette statistique garde l'indépendance vis à vis du paramètre. Ainsi cette approximation permet la simulation d'échantillons semblables à ceux qui auraient été simulés sous la densité conditionnelle. Nous développons alors deux types d'applications.

Récemment, le fameux théorème de Rao-Blackwell a fait l'objet d'une toute nouvelle attention (voir, par exemple, [22] ou [34]). Ce théorème permet de réduire la variance d'un estimateur sans biais en conditionnant par une statistique exhaustive pour un paramètre. Quand cette statistique est, de plus, complète, la réduction est optimale comme énoncé par le théorème de Lehmann-Scheffé. Un exemple est présenté.

Dans la même veine, on propose des tests de type Monte Carlo pour des modèles à paramètres de nuisance ; le conditionnement est ici réalisé par une statistique exhaustive du paramètre de nuisance, dans le cadre classique des modèles exponentiels. L'intérêt pour les familles exponentielles vient du fait que parmi toutes les familles de distributions absolument continues à support fixé, les familles exponentielles sont les seules permettant de réduire la dimension du paramétrage de l'échantillonnage par exhaustivité. Enfin l'estimation de paramètre d'intérêt dans des modèles où la vraisemblance du paramètre de nuisance n'est pas unimodale peut être réalisée en utilisant cette approche conditionnelle. La multimodalité de cette vraisemblance (voir [91] pour des exemples) peut amener des difficultés sérieuses pour l'inférence ; la méthode que nous proposons, illustrée par un exemple, semble répondre à ces questions.

Résumé des chapitres

Le premier chapitre présente une approximation fine de la densité de longues sous-suites d'une marche aléatoire conditionnée par la valeur de son extrémité, ou par une moyenne d'une fonction de ses incréments, lorsque sa taille tend vers l'infini. Une approximation est aussi obtenue lorsque l'événement conditionnant énonce que la valeur terminale de la marche aléatoire appartient à un ensemble mince, ou gros, d'intérieur non vide. Les approximations proposées ont lieu soit en probabilité sous la loi conditionnelle, soit en distance de la variation totale. On traite en détail du choix effectif de la longueur maximale de la sous suite pour laquelle une erreur relative maximale fixée est atteinte par l'approximation ; des algorithmes explicites permettent la mise en oeuvre effective de cette approximation et de ses conséquences.

Le deuxième chapitre traite des méthodes d'estimation des probabilités de certains événements rares par des techniques d'échantillonnage d'importance. Les événements considérés sont de la forme $(\mathbf{U}_{1,n} \in nA)$. L'approximation de la densité de la loi conditionnelle du vecteur $(X_1, ..., X_{k_n})$ relativement à $(\mathbf{U}_{1,n} \in nA)$ est utilisée lorsque k_n est de l'ordre de n, comme conséquence d'un résultat du premier chapitre.

Le troisième chapitre traite de questions de nature strictement inférentielles en lien avec le concept d'exhaustivité. Une nouvelle approche à l'inférence conditionnelle est proposée, basée sur la simulation d'échantillons conditionnellement à une statistique observée sur les données. Ce résultat permet la Rao-Blackwellisation d'estimateurs. Des méthodes d'inférence conditionnelle sous paramètre de nuisance sont aussi étudiées. Des exemples sont présentés.

Le dernier chapitre généralise une partie des résultats obtenus au chapitre 1 dans le cas où les $\mathbf{X}_i$ sont des vecteurs aléatoires de $\mathbb{R}^d$; la fonction u définie ci-dessus a pour

domaine $\mathbb{R}^d$ et pour image $\mathbb{R}^s$. Après avoir introduit les notations utilisées pour effectuer les développements d'Edgeworth dans $\mathbb{R}^d$ (voir [7]), le théorème principal est exposé. Une généralisation de la règle permettant l'obtention d'un k optimal et des exemples de trajectoires typiques sont finalement proposés.

Afin de faciliter la lecture, certaines démonstrations très longues et les lemmes techniques sont laissés à la fin de chaque chapitre. En annexe, une liste d'exemples de calculs permettant l'application des méthodes exposées est proposée.

A l'exception du dernier chapitre, les résultats de cette thèse ont été obtenus par un travail en commun avec Monsieur le Professeur Michel Broniatowski, directeur de ma thèse.

Chapter 1

Long runs under a conditional limit distribution.

1.1 Context and scope

This paper explores the asymptotic distribution of a random walk conditioned on its final value as the number of summands increases. Denote $\mathbf{X}_1^n := (\mathbf{X}_1, .., \mathbf{X}_n)$ a set of n independent copies of a real random variable $\mathbf{X}$ with density $p_{\mathbf{X}}$ on $\mathbb{R}$ and $\mathbf{S}_{1,n} := \mathbf{X}_1 + ...+\mathbf{X}_n$. We consider approximations of the density of the vector $\mathbf{X}_1^k=(\mathbf{X}_1, .., \mathbf{X}_k)$ on $\mathbb{R}^k$ when $\mathbf{S}_{1,n} = na_n$ and a_n is a convergent sequence. The integer valued sequence $k := k_n$ is such that

$$0 \leq \lim \sup_{n\to\infty} k/n \leq 1 \tag{K1}$$

together with

$$\lim_{n\to\infty} n - k = \infty. \tag{K2}$$

Therefore we may consider the asymptotic behavior of the density of the trajectory of the random walk on long runs. For sake of applications we also address the case when $\mathbf{S}_{1,n}$ is substituted by $\mathbf{U}_{1,n} := u(\mathbf{X}_1) + ... + u(\mathbf{X}_1)$ for some real valued measurable function u, and when the conditioning event writes $(\mathbf{U}_{1,n} = u_{1,n})$ where $u_{1,n}/n$ converges as n tends to infinity. A complementary result provides an estimation for the case when the conditioning event is a large set in the large deviation range, $(\mathbf{U}_{1,n} \in nA)$ where A is a Borel set with non void interior with $Eu\mathbf{X} <\texttt{essinf}A$; two cases are considered, according to the local dimension of A at its essential infimum point $\texttt{essinf}A$.

The interest in this question stems from various sources. When k is fixed (typically $k = 1$) this is a version of the *Gibbs Conditional Principle* which has been studied extensively for fixed $a_n \neq E\mathbf{X}$, therefore under a *large deviation* condition. [33] have considered this issue also in the case $k/n \to \theta$ for $0 \leq \theta < 1$, in connection with de Finetti's Theorem for exchangeable finite sequences. Their interest was related to the approximation of the density of $\mathbf{X}_1^k$ by the *product density* of the summands $\mathbf{X}_i$'s, therefore on the permanence of the independence of the $\mathbf{X}_i$'s under conditioning. Their result is in the spirit of [93] and to be paralleled with [26] asymptotic conditional independence result, when the conditioning event is $(\mathbf{S}_{1,n} > na_n)$ with a_n fixed and larger than $E\mathbf{X}$. In the same vein and under the same *large deviation* condition [29] considered similar problems. This question is also of importance in Statistical Physics. Numerous papers pertaining to structural properties of polymers deal with this issue, and we refer to [30] and [31] for a description of those problems and related results. In the moderate deviation case, [44] also considered a similar problem when $k = 1$.

Approximation of conditional densities is the basic ingredient for the numerical estimation of integrals through improved Monte Carlo techniques. Rare event probabilities may be evaluated through Importance sampling techniques; efficient sampling schemes consist in the simulation of random variables under a proxy of a conditional density, often pertaining to conditioning events of the form $(\mathbf{U}_{1,n} > na_n)$; optimizing these schemes has been a motivation for this work; see Chapter 2.

In parametric statistical inference conditioning on the observed value of a statistics leads to a reduction of the mean square error of some estimate of the parameter; the celebrated Rao-Blackwell and Lehmann-Scheffé Theorems can be implemented when a simulation technique produces samples according to the distribution of the data conditioned on the value of some observed statistics. In these applications the conditioning event is local and when the statistics is of the form $\mathbf{U}_{1,n}$ then the observed value $u_{1,n}$ satisfies $\lim_{n\to\infty} u_{1,n}/n = Eu(\mathbf{X})$. Such is the case in exponential families when $\mathbf{U}_{1,n}$ is a sufficient statistics for the parameter. Other fields of applications pertain to parametric estimation where conditioning by the observed value of a sufficient statistics for a nuisance parameter produces optimal inference through maximum likelihood in the conditioned model; in general this conditional density is unknown; the approximation produced in this paper provides a tool for the solution of these problems; see Chapter 3.

Both for Importance Sampling and for the improvement of estimators, the approximation of the conditional density of $\mathbf{X}_1^k$ on long runs should be of a very special form: it has to be a density on $\mathbb{R}^k$, easy to simulate, and the approximation should be sharp. For these applications the relative error of the approximation should be small on the simulated paths only. Also for inference through maximum likelihood under nuisance parameter the approximation has to be accurate on the sample itself and not on the entire space.

Our first set of results provides a very sharp approximation scheme; numerical evidence on exponential runs with length $n = 1000$ provide a *relative error* of the approximation of order less than 100% for the density of the first 800 terms when evaluated on the sample paths themselves, thus on the significant part of the support of the conditional density; this very sharp approximation rate is surprising in such a large dimensional space, and it illustrates the fact that the conditioned measure occupies a very small part of the entire space. Therefore the approximation of the density of $\mathbf{X}_1^k$ is not performed on the sequence of entire spaces $\mathbb{R}^k$ but merely on a sequence of subsets of $\mathbb{R}^k$ which bear the trajectories of the conditioned random walk with probability going to 1 as n tends to infinity; the approximation is performed on *typical paths*.

The extension of our results from typical paths to the whole space $\mathbb{R}^k$ holds: convergence of the relative error on large sets imply that the total variation distance between the conditioned measure and its approximation goes to 0 on the entire space. So our results provide an extension of [33] and [29] who considered the case when k is of small order with respect to n; the conditions which are assumed in the present paper are weaker than those assumed in the just cited works; however, in contrast with their results, we do not provide explicit rates for the convergence to 0 of the total variation distance on $\mathbb{R}^k$.

It would have been of interest to consider sharper convergence criteria than the total variation distance; the χ^2-distance, which is the mean square relative error, cannot be bounded through our approach on the entire space $\mathbb{R}^k$, since it is only handled on large sets of trajectories (whose probability goes to 1 an n increases); this is not sufficient to bound its expected value under the conditional sampling.

This paper is organized as follows. Section 2 presents the approximation scheme for the conditional density of $\mathbf{X}_1^k$ under the point conditioning sequence $(\mathbf{S}_{1,n} = na_n)$. In section 3, it is extended to the case when the conditioning family of events writes $(\mathbf{U}_{1,n} = u_{1,n})$.

The value of k for which this approximation is fair is discussed; an algorithm for the implementation of this rule is proposed. Algorithms for the simulation of random variables under the approximating scheme are also presented. Section 4 extends the results of Section 3 when conditioning on large sets.

The main steps of the proofs are in the core of the paper; some of the technicalities is left to the Appendix.

1.2 Random walks conditioned on their sum

1.2.1 Notation and hypothesis

In this section the point conditioning event writes

$$\mathcal{E}_n := (\mathbf{S}_{1,n} = na_n)\,.$$

We assume that $\mathbf{X}$ satisfies the Cramer condition, i.e. $\mathbf{X}$ has a finite moment generating function $\Phi(t) := E \exp t\mathbf{X}$ in a non void neighborhood of 0. Denote

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

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

$$\mu_3(t) := \frac{d}{dt} s^2(t)$$

The values of $m(t)$, s^2 and $\mu_3(t)$ are the expectation, the variance and the kurtosis of the *tilted* density

$$\pi^\alpha(x) := \frac{\exp tx}{\Phi(t)} p(x) \tag{1.1}$$

where t is the only solution of the equation $m(t) = \alpha$ when α belongs to the support of $\mathbf{X}$. Conditions on $\Phi(t)$ which ensure existence and uniqueness of t are referred to as *steepness properties;* we refer to [6] , p.153 and followings for all properties of moment generating functions used in this paper. Denote Π^α the probability measure with density π^α.

Remark 1. *In many cases the function $\phi_{\mathbf{U}}(t)$ is not available in closed form, and consequently the same holds for the functions m, s^2 and μ_3. However tail probabilities are needed in safety surveys or in reliability studies; henceforth the real need is not on tail probabilities under the true (usually unknown) model, but merely on some proxy of it, under which the tail probability estimate should have some conservative property. In a number of cases this may lead to a known model with rather simple forms for these functions. In other cases expansions of the function $\phi_{\mathbf{U}}(t)$ are available, leading to some approximation for the resulting derivatives. Inversion of m is feasible; inversion must be realized in a small interval around a_n. Hence the values of $s^2(t)$ and $\mu_3(t)$ for $m(t)$ close to a_n can be calculated with good accuracy.*

We also assume that the characteristic function of $\mathbf{X}$ is in L^r for some $r \geq 1$ which is necessary for the Edgeworth expansions to be performed.

The probability measure of the random vector $\mathbf{X}_1^n$ on $\mathbb{R}^n$ conditioned upon $\mathcal{E}_n$ is denoted P_{na_n}. We also denote P_{na_n} the corresponding distribution of $\mathbf{X}_1^k$ conditioned upon $\mathcal{E}_n$; the vector $\mathbf{X}_1^k$ then has a density with respect to the Lebesgue measure on $\mathbb{R}^k$ for

$1 \leq k < n$,which will be denoted p_{na_n}. For a generic r.v. $\mathbf{Z}$ with density p, $p(\mathbf{Z} = z)$ denotes the value of p at point z. Hence, $p_{na_n}\left(x_1^k\right) = p\left(\mathbf{X}_1^k = x_1^k | \mathbf{S}_{1,n} = na_n\right)$. The normal density function on $\mathbb{R}$ with mean μ and variance τ at x is denoted $\mathfrak{n}(\mu, \tau, x)$. When $\mu = 0$ and $\tau = 1$, the standard notation $\mathfrak{n}(x)$ is used.

1.2.2 A first approximation result

We first put forwards a simple result which provides an approximation of the density p_{na_n} of the measure P_{na_n} on $\mathbb{R}^k$ when k satisfies (K1) and (K2) . For $i \leq j$ denote

$$s_{i,j} := x_i + ... + x_j.$$

Denote $a := a_n$ omitting the index n for clearness.

We make use of the following property which states the invariance of conditional densities under the tilting: For $1 \leq i \leq j \leq n$, for all a in the range of $\mathbf{X}$, for all u and s

$$p(\mathbf{S}_{i,j} = u | \mathbf{S}_{1,n} = s) = \pi^a(\mathbf{S}_{i,j} = u | \mathbf{S}_{1,n} = s) \tag{1.2}$$

where $\mathbf{S}_{i,j} := \mathbf{X}_i + ... + \mathbf{X}_j$ together with $\mathbf{S}_{1,0} = s_{1,0} = 0$. By Bayes formula it holds

$$\begin{aligned} p_{na}\left(x_1^k\right) &= \prod_{i=0}^{k-1} p(\mathbf{X}_{i+1} = x_{i+1} | \mathbf{S}_{i+1,n} = na - s_{1,i}) && (1.3) \\ &= \prod_{i=0}^{k-1} \pi^a(\mathbf{X}_{i+1} = x_{i+1}) \frac{\pi^a(\mathbf{S}_{i+2,n} = na - s_{1,i+1})}{\pi^a(\mathbf{S}_{i+1,n} = na - s_{1,i})} \\ &= \left[\prod_{i=0}^{k-1} \pi^a(\mathbf{X}_{i+1} = x_{i+1})\right] \frac{\pi^a(\mathbf{S}_{k+1,n} = na - s_{1,k})}{\pi^a(\mathbf{S}_{1,n} = na)}. && (1.4) \end{aligned}$$

Denote $\overline{\mathbf{S}_{k+1,n}}$ and $\overline{\mathbf{S}_{1,n}}$ the normalized versions of $\mathbf{S}_{k+1,n}$ and $\mathbf{S}_{1,n}$ under the sampling distribution Π^a. By (1.4)

$$p_{na}\left(x_1^k\right) = \left[\prod_{i=0}^{k-1} \pi^a(\mathbf{X}_{i+1} = x_{i+1})\right] \frac{\sqrt{n}}{\sqrt{n-k}} \frac{\pi^a\left(\overline{\mathbf{S}_{k+1,n}} = \frac{ka - s_{1,k}}{s_a\sqrt{n-k}}\right)}{\pi^a\left(\overline{\mathbf{S}_{1,n}} = 0\right)}.$$

A first order Edgeworth expansion is performed in both terms of the ratio in the above display; see Remark 6 hereunder. This yields, assuming (K1) and (K2)

Proposition 2. *For all x_1^k in $\mathbb{R}^k$*

$$\begin{aligned} p_{na}\left(x_1^k\right) = &\left[\prod_{i=0}^{k-1} \pi^a(\mathbf{X}_{i+1} = x_{i+1})\right] \left[\frac{\mathfrak{n}\left(\frac{ka - s_{1,k}}{s(t^a)\sqrt{n-k}}\right)}{\mathfrak{n}(0)} \sqrt{\frac{n}{n-k}} \right. \\ &\left. \left(1 + \frac{\mu_3(t^a)}{6s^3(t^a)\sqrt{n-k}} H_3\left(\frac{ka - s_{1,k}}{s(t^a)\sqrt{n-k}}\right)\right) + O\left(\frac{1}{\sqrt{n}}\right)\right] \end{aligned} \tag{1.5}$$

where $H_3(x) := x^3 - 3x$. The value of t^a is defined through $m(t^a) = a$.

Despite its appealing aspect, (1.5) is of poor value for applications, since it does not yield an explicit way to simulate samples under a proxy of p_{na} for large values of k. The other way is to construct the approximation of p_{na} step by step, approximating the terms

in (1.3) one by one and using the invariance under the tilting at each step, which introduces a product of different tilted densities in (1.4). This method produces a valid approximation of p_{na} on subsets of $\mathbb{R}^k$ which bear the trajectories of the condition random walk with larger and larger probability, going to 1 as n tends to infinity.

This introduces the main focus of this paper.

1.2.3 A recursive approximation scheme

We introduce a positive sequence ϵ_n which satisfies

$$\lim_{n\to\infty} \epsilon_n \sqrt{n-k} = \infty \tag{E1}$$

$$\lim_{n\to\infty} \epsilon_n \left(\log n\right)^2 = 0. \tag{E2}$$

It will be shown that $\epsilon_n \left(\log n\right)^2$ is the rate of accuracy of the approximating scheme.

We denote a the generic term of the convergent sequence $(a_n)_{n\geq 1}$. For clearness the dependence in n of all quantities involved in the coming development is omitted in the notation.

Approximation of the density of the runs

Define a density $g_{na}(y_1^k)$ on $\mathbb{R}^k$ as follows. Set

$$g_0(y_1|\, y_0) := \pi^a(y_1)$$

with y_0 arbitrary, and for $1 \leq i \leq k-1$ define $g(y_{i+1}|\, y_1^i)$ recursively.

Set t_i the unique solution of the equation

$$m_i := m(t_i) = \frac{n}{n-i}\left(a - \frac{s_{1,i}}{n}\right) \tag{1.6}$$

where $s_{1,i} := y_1 + \ldots + y_i$. The tilted adaptive family of densities π^{m_i} is the basic ingredient of the derivation of approximating scheme. Let

$$s_i^2 := \frac{d^2}{dt^2}\left(\log E_{\pi^{m_i}} \exp t\mathbf{X}\right)(0)$$

and

$$\mu_j^i := \frac{d^j}{dt^j}\left(\log E_{\pi^{m_i}} \exp t\mathbf{X}\right)(0)\,,\ j = 3, 4$$

which are the second, third and fourth cumulants of π^{m_i}. Let

$$g(y_{i+1}|\, y_1^i) = C_i p_{\mathbf{X}}(y_{i+1}) \mathfrak{n}\left(\alpha\beta + a, \beta, y_{i+1}\right) \tag{1.7}$$

be a density where

$$\alpha = t_i + \frac{\mu_3^i}{2s_i^4\left(n-i-1\right)} \tag{1.8}$$

$$\beta = s_i^2\left(n-i-1\right) \tag{1.9}$$

and C_i is a normalizing constant.

Define

$$g_{na}(y_1^k) := g_0(y_1|\, y_0).\prod_{i=1}^{k-1} g(y_{i+1}|\, y_1^i). \tag{1.10}$$

We then have

Theorem 3. *Assume (K1) and (K2) together with (E1) and (E2). Let Y_1^n be a sample with density p_{na}. Then*

$$p_{na}\left(Y_1^k\right) := p(\mathbf{X}_1^k = Y_1^k \Big| \mathbf{S}_{1,n} = na) = g_{na}(Y_1^k)(1 + o_{P_{na}}(\epsilon_n (\log n)^2)). \tag{1.11}$$

Proof. The proof uses Bayes formula to write $p(\mathbf{X}_1^k = Y_1^k \Big| \mathbf{S}_{1,n} = na)$ as a product of k conditional densities of individual terms of the trajectory evaluated at Y_1^k. Each term of this product is approximated through an Edgeworth expansion which together with the properties of Y_1^k under P_{na} concludes the proof. This proof is rather long and we have differed its technical steps to the Appendix.

Denote $S_{1,0} = 0$ and $S_{1,i} := S_{1,i-1} + Y_i$. It holds

$$\begin{aligned} p(\mathbf{X}_1^k = Y_1^k \Big| \mathbf{S}_{1,n} = na) &= p(\mathbf{X}_1 = Y_1 | \, \mathbf{S}_{1,n} = na) \\ \prod_{i=1}^{k-1} p(\mathbf{X}_{i+1} = Y_{i+1} | \, \mathbf{X}_1^i = Y_1^i, \mathbf{S}_{1,n} = na) & \\ = \prod_{i=0}^{k-1} p\left(\mathbf{X}_{i+1} = Y_{i+1} | \, \mathbf{S}_{i+1,n} = na - S_{1,i}\right) & \end{aligned} \tag{1.12}$$

by independence of the r.v's $\mathbf{X}_i's$.

Define t_i through

$$m(t_i) = \frac{n}{n-i}\left(a - \frac{S_{1,i}}{n}\right)$$

a function of the past r.v's Y_1^i and set $m_i := m(t_i)$ and $s_i^2 := s^2(t_i)$. By (1.2)

$$\begin{aligned} & p\left(\mathbf{X}_{i+1} = Y_{i+1} | \, \mathbf{S}_{i+1,n} = na - S_{1,i}\right) \\ & = \pi^{m_i}\left(\mathbf{X}_{i+1} = Y_{i+1} | \, \mathbf{S}_{i+1,n} = na - S_{1,i}\right) \\ & = \pi^{m_i}\left(\mathbf{X}_{i+1} = Y_{i+1}\right) \frac{\pi^{m_i}\left(\mathbf{S}_{i+2,n} = na - S_{1,i+1}\right)}{\pi^{m_i}\left(\mathbf{S}_{i+1,n} = na - S_{1,i}\right)} \end{aligned}$$

where we used the independence of the $\mathbf{X}_j$'s under π^{m_i}. A precise evaluation of the dominating terms in this latest expression is needed in order to handle the product (1.12).

Under the sequence of densities π^{m_i} the i.i.d. r.v's $\mathbf{X}_{i+1}, ..., \mathbf{X}_n$ define a triangular array which satisfies a local central limit theorem, and an Edgeworth expansion. Under π^{m_i}, $\mathbf{X}_{i+1}$ has expectation m_i and variance s_i^2. Center and normalize both the numerator and denominator in the fraction which appears in the last display. Denote $\overline{\pi_{n-i-1}}$ the density of the normalized sum $\left(\mathbf{S}_{i+2,n} - (n-i-1)m_i\right) / \left(s_i\sqrt{n-i-1}\right)$ when the summands are i.i.d. with common density π^{m_i}. Accordingly $\overline{\pi_{n-i}}$ is the density of the normalized sum $\left(\mathbf{S}_{i+1,n} - (n-i)m_i\right) / \left(s_i\sqrt{n-i}\right)$ under i.i.d. π^{m_i} sampling. Hence, evaluating both $\overline{\pi_{n-i-1}}$ and its normal approximation at point Y_{i+1},

$$\begin{aligned} & p\left(\mathbf{X}_{i+1} = Y_{i+1} | \, \mathbf{S}_{i+1,n} = na - S_{1,i}\right) \\ & = \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \pi^{m_i}\left(\mathbf{X}_{i+1} = Y_{i+1}\right) \frac{\overline{\pi_{n-i-1}}\left(\left(m_i - Y_{i+1}\right) / s_i\sqrt{n-i-1}\right)}{\overline{\pi_{n-i}}(0)} \\ & := \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \pi^{m_i}\left(\mathbf{X}_{i+1} = Y_{i+1}\right) \frac{N_i}{D_i}. \end{aligned} \tag{1.13}$$

The sequence of densities $\overline{\pi_{n-i-1}}$ converges pointwise to the standard normal density under (E1) which implies that $n - i$ tends to infinity for all $1 \leq i \leq k$, and an Edgeworth

expansion to the order 5 is performed for the numerator and the denominator. The main arguments used in order to obtain the order of magnitude of the involved quantities are (i) a maximal inequality which controls the magnitude of m_i for all i between 0 and $k-1$ (Lemma 24), (ii) the order of the maximum of the $Y_i's$ (Lemma 25). As proved in the Appendix,

$$N_i = \mathfrak{n}\left(-Y_{i+1}/s_i\sqrt{n-i-1}\right).A.B + O_{P_{na}}\left(\frac{1}{(n-i-1)^{3/2}}\right) \tag{1.14}$$

where

$$A := \left(1 + \frac{aY_{i+1}}{s_i^2(n-i-1)} - \frac{a^2}{2s_i^2(n-i-1)} + \frac{o_{P_{na}}(\epsilon_n \log n)}{n-i-1}\right) \tag{1.15}$$

and

$$B := \begin{pmatrix} 1 - \frac{\mu_3^i}{2s_i^4(n-i-1)}(a - Y_{i+1}) \\ -\frac{\mu_3^i - s_i^4}{8s_i^4(n-i-1)} - \frac{15(\mu_3^i)^2}{72s_i^6(n-i-1)} + \frac{O_{P_{na}}\left((\log n)^2\right)}{(n-i-1)^2} \end{pmatrix} \tag{1.16}$$

The $O_{P_{na}}\left(\frac{1}{(n-i-1)^{3/2}}\right)$ term in (1.14) is uniform upon $(m_i - Y_{i+1})/s_i\sqrt{n-i-1}$. Turn back to (1.13) and perform the same Edgeworth expansion in the denominator, which writes

$$D_i = \mathfrak{n}(0)\left(1 - \frac{\mu_3^i - s_i^4}{8s_i^4(n-i)} - \frac{15(\mu_3^i)^2}{72s_i^6(n-i)}\right) + O_{P_{na}}\left(\frac{1}{(n-i)^{3/2}}\right). \tag{1.17}$$

The terms in $g(Y_{i+1}|\,Y_1^i)$ follow from an expansion in the ratio of the two expressions (1.14) and (1.17) above. The gaussian contribution is explicit in (1.14) while the term $\exp\frac{\mu_3^i}{2s_i^4(n-i-1)}Y_{i+1}$ is the dominant term in B. Turning to (1.13) and comparing with (1.11) it appears that the normalizing factor C_i in $g(Y_{i+1}|\,Y_1^i)$ compensates the term $\frac{\sqrt{n-i}}{\Phi(t_i)\sqrt{n-i-1}}\exp\left(\frac{-a\mu_3^i}{2s_i^2(n-i-1)}\right)$, where the term $\Phi(t_i)$ comes from $\pi^{m_i}(\mathbf{X}_{i+1} = Y_{i+1})$. Further the product of the remaining terms in the above approximations in (1.14) and (1.17) turn to build the $1 + o_{P_{na}}\left(\epsilon_n (\log n)^2\right)$ approximation rate, as claimed. Details are differed to the Appendix. This yields

$$p(\mathbf{X}_1^k = Y_1^k \,\big|\, \mathbf{S}_{1,n} = na) = \left(1 + o_{P_{na}}\left(\epsilon_n (\log n)^2\right)\right) g_0(Y_1|\,Y_0) \prod_{i=1}^{k-1} g(Y_{i+1}|\,Y_1^i)$$

which completes the proof of the Theorem. □

That the variation distance between P_{na_n} and G_{na_n} tends to 0 as $n \to \infty$ is stated in Section 3.

Remark 4. *When the $\mathbf{X}_i$'s are i.i.d. with a standard normal density, then the result in the above approximation Theorem holds with $k = n-1$ stating that $p(\mathbf{X}_1^{n-1} = x_1^{n-1} \,\big|\, \mathbf{S}_{1,n} = na) = g_{na}\left(x_1^{n-1}\right)$ for all x_1^{n-1} in $\mathbb{R}^{n-1}$. This extends to the case when they have an infinitely divisible distribution. However formula (1.11) holds true without the error term only in the gaussian case. Similar exact formulas can be obtained for infinitely divisible distributions using (1.12) making no use of tilting. Such formula is used to produce Figures 1.1, 1.2, 1.3 and 1.4 in order to assess the validity of the selection rule for k in the exponential case.*

Remark 5. *The density in (1.7) is a slight modification of* π^{m_i}. *The modification from* $\pi^{m_i}(y_{i+1})$ *to* $g(y_{i+1}|y_1^i)$ *is a small shift in the location parameter depending both on* a *and on the skewness of* p, *and a change in the variance : large values of* $\mathbf{X}_{i+1}$ *have smaller weight for large* i, *so that the distribution of* $\mathbf{X}_{i+1}$ *tends to concentrate around* m_i *as* i *approaches* k.

Remark 6. *In Theorem 3 , as in Proposition 2, as in Theorem 9 or as in Lemma 25, we use an Edgeworth expansion for the density of the normalized sum of the* $n-ith$ *row of some triangular array of row-wise independent r.v's with common density. Consider the i.i.d. r.v's* $\mathbf{X}_1, ..., \mathbf{X}_n$ *with common density* $\pi^a(x)$ *where* a *may depend on* n *but remains bounded. The Edgeworth expansion pertaining to the normalized density of* $\mathbf{S}_{1,n}$ *under* π^a *can be derived following closely the proof given for example in [46], p.532 and followings substituting the cumulants of* p *by those of* π^a. *Denote* $\varphi_a(z)$ *the characteristic function of* $\pi^a(x)$. *Clearly for any* $\delta > 0$ *there exists* $q_{a,\delta} < 1$ *such that* $|\varphi_a(z)| < q_{a,\delta}$ *and since* a *is bounded,* $\sup_n q_{a,\delta} < 1$. *Therefore the inequality (2.5) in [46] p.533 holds. With* ψ_n *defined as in [46], (2.6) holds with* φ *replaced by* φ_a *and* σ *by* $s(t^a)$; *(2.9) holds, which completes the proof of the Edgeworth expansion in the simple case. The proof goes in the same way for higher order expansions.*

Sampling under the approximation

Applications of Theorem 3 in Importance Sampling procedures and in Statistics require a reverse result. So assume that Y_1^k is a random vector generated under G_{na} with density g_{na}. Can we state that $g_{na}\left(Y_1^k\right)$ is a good approximation for $p_{na}\left(Y_1^k\right)$? This holds true. We state a simple Lemma in this direction.

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

Lemma 7. *Suppose that for some sequence* ε_n *which tends to* 0 *as* n *tends to infinity*

$$\mathfrak{r}_n\left(Y_1^n\right) = \mathfrak{s}_n\left(Y_1^n\right)\left(1 + o_{\mathfrak{R}_n}(\varepsilon_n)\right) \tag{1.18}$$

as n *tends to* ∞. *Then*

$$\mathfrak{s}_n\left(Y_1^n\right) = \mathfrak{r}_n\left(Y_1^n\right)\left(1 + o_{\mathfrak{S}_n}(\varepsilon_n)\right). \tag{1.19}$$

Proof. Denote

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

It holds for all positive δ

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

Write

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

Since

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

it follows that

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

which proves the claim. □

As a direct by-product of Theorem 3 and Lemma 7 we obtain

Theorem 8. *Assume (K1) and (K2) together with (E1) and (E2). Let Y_1^k be a sample with density g_{na}. It holds*

$$p_{na}\left(Y_1^k\right) = g_{na}(Y_1^k)(1 + o_{G_{na}}(\epsilon_n (\log n)^2)).$$

1.3 Random walks conditioned by a function of their summands

This section extends the above results to the case when the conditioning event writes

$$\mathbf{U}_{1,n} := u_{1,n} \tag{1.20}$$

with

$$\mathbf{U}_{1,n} := u(\mathbf{X}_1) + ... + u(\mathbf{X}_n)$$

where the function u is real valued and the sequence $u_{1,n}/n$ converges. The characteristic function of the random variable $u(\mathbf{X})$ is assumed to belong to L^r for some $r \geq 1$. Let $p_{\mathbf{U}}$ denote the density of $\mathbf{U} = u(\mathbf{X})$.

Assume

$$\phi_{\mathbf{U}}(t) := E \exp t\mathbf{U} < \infty \tag{1.21}$$

for t in a non void neighborhood of 0. Define the functions $m(t)$, $s^2(t)$ and $\mu_3(t)$ as the first, second and third derivatives of $\log \phi_{\mathbf{U}}(t)$.

Denote

$$\pi_{\mathbf{U}}^{\alpha}(u) := \frac{\exp tu}{\phi_{\mathbf{U}}(t)} p_{\mathbf{U}}(u) \tag{1.22}$$

with $m(t) = \alpha$ and α belongs to the support of $P_{\mathbf{U}}$, the distribution of $\mathbf{U}$.

We also introduce the family of densities

$$\pi_u^{\alpha}(x) := \frac{\exp tu(x)}{\phi_{\mathbf{U}}(t)} p_{\mathbf{X}}(x). \tag{1.23}$$

1.3.1 Approximation of the density of the runs

Assume that the sequence ϵ_n satisfies (E1) and (E2).

Define a density $g_{u_{1,n}}(y_1^k)$ on $\mathbb{R}^k$ as follows. Set

$$m_0 := u_{1,n}/n$$

and

$$g_0(y_1 | y_0) := \pi_u^{m_0}(y_1) \tag{1.24}$$

with y_0 arbitrary and, for $1 \leq i \leq k-1$, define $g(y_{i+1} | y_1^i)$ recursively. Denote $u_{1,i} := u(y_1) + ... + u(y_i)$.

Set t_i the unique solution of the equation

$$m_i := m(t_i) = \frac{u_{1,n} - u_{1,i}}{n - i} \tag{1.25}$$

and, let

$$s_i^2 := \frac{d^2}{dt^2}\left(\log E_{\pi_{\mathbf{U}}^{m_i}} \exp t\mathbf{U}\right)(0)$$

and

$$\mu_j^i := \frac{d^j}{dt^j}\left(\log E_{\pi_{\mathbf{U}}^{m_i}} \exp t\mathbf{U}\right)(0)\,,\ j=3,4$$

which are the second, third and fourth cumulants of $\pi_{\mathbf{U}}^{m_i}$. A density $g(y_{i+1}|\,y_1^i)$ is defined through

$$g(y_{i+1}|\,y_1^i) = C_i p_{\mathbf{X}}(y_{i+1})\mathfrak{n}\left(\alpha\beta + m_0, \beta, u\left(y_{i+1}\right)\right). \tag{1.26}$$

Here

$$\alpha = t_i + \frac{\mu_3^i}{2s_i^4\,(n-i-1)} \tag{1.27}$$

$$\beta = s_i^2\,(n-i-1) \tag{1.28}$$

and the C_i is a normalizing constant.

Set

$$g_{u_{1,n}}\left(y_1^k\right) := g_0(y_1|\,y_0)\prod_{i=1}^{k-1} g(y_{i+1}|\,y_1^i). \tag{1.29}$$

Theorem 9. *Assume (K1) and (K2) together with (E1) and (E2). Then*

1. *Let Y_1^k be a sample with density $p_{u_{1,n}}$.*

$$p_{u_{1,n}}\left(Y_1^k\right) := p\left(\mathbf{X}_1^k = Y_1^k\,|\mathbf{U}_{1,n} = u_{1,n}\right) = g_{u_{1,n}}(Y_1^k)(1 + o_{P_{u_{1,n}}}(\epsilon_n\,(\log n)^2))$$

2. *Let Y_1^k be a sample with density $g_{u_{1,n}}$.*

$$p_{u_{1,n}}\left(Y_1^k\right) = g_{u_{1,n}}(Y_1^k)(1 + o_{G_{u_{1,n}}}(\epsilon_n\,(\log n)^2)).$$

Proof. We only sketch the initial step of the proof of (i), which rapidly follows the same track as that in Theorem 3 .

As in the proof of Theorem 3 evaluate

$$\begin{aligned}
&p\left(\mathbf{X}_{i+1} = Y_{i+1}|\,\mathbf{U}_{i+1,n} = u_{1,n} - U_{1,i}\right)\\
&= p_{\mathbf{X}}\left(\mathbf{X}_{i+1} = Y_{i+1}\right)\frac{p_{\mathbf{U}}\left(\mathbf{U}_{i+2,n} = u_{1,n} - U_{1,i+1}\right)}{p_{\mathbf{U}}\left(\mathbf{U}_{i+1,n} = u_{1,n} - U_{1,i}\right)}\\
&= \frac{p_{\mathbf{X}}\left(\mathbf{X}_{i+1} = Y_{i+1}\right)}{p_{\mathbf{U}}\left(\mathbf{U}_{i+1} = u\left(Y_{i+1}\right)\right)}p_{\mathbf{U}}\left(\mathbf{U}_{i+1} = u\left(Y_{i+1}\right)\right)\frac{p_{\mathbf{U}}\left(\mathbf{U}_{i+2,n} = u_{1,n} - U_{1,i+1}\right)}{p_{\mathbf{U}}\left(\mathbf{U}_{i+1,n} = u_{1,n} - U_{1,i}\right)}.
\end{aligned}$$

Use the invariance of the conditional density with respect to the change of sampling defined by $\pi_{\mathbf{U}}^{m_i}$ to obtain

$$\begin{aligned}
&p\left(\mathbf{X}_{i+1} = Y_{i+1}|\,\mathbf{U}_{i+1,n} = u_{1,n} - U_{1,i}\right)\\
&= \frac{p_{\mathbf{X}}\left(\mathbf{X}_{i+1} = Y_{i+1}\right)}{p_{\mathbf{U}}\left(\mathbf{U}_{i+1} = u\left(Y_{i+1}\right)\right)}\pi_{\mathbf{U}}^{m_i}\left(\mathbf{U}_{i+1} = u\left(Y_{i+1}\right)\right)\frac{\pi_{\mathbf{U}}^{m_i}\left(\mathbf{U}_{i+2,n} = u_{1,n} - U_{1,i+1}\right)}{\pi_{\mathbf{U}}^{m_i}\left(\mathbf{U}_{i+1,n} = u_{1,n} - U_{1,i}\right)}\\
&= p_{\mathbf{X}}\left(\mathbf{X}_{i+1} = Y_{i+1}\right)\frac{e^{t_i u(Y_{i+1})}}{\phi_{\mathbf{U}}(t_i)}\frac{\pi_{\mathbf{U}}^{m_i}\left(\mathbf{U}_{i+2,n} = u_{1,n} - U_{1,i+1}\right)}{\pi_{\mathbf{U}}^{m_i}\left(\mathbf{U}_{i+1,n} = u_{1,n} - U_{1,i}\right)}
\end{aligned}$$

and proceed through the Edgeworth expansions in the above expression, following verbatim the proof of Theorem 3. We omit details. The proof of (ii) follows from Lemma 7 □

We turn to a consequence of Theorem 9
For all $\delta > 0$, let

$$E_{k,\delta} := \left\{ y_1^k \in \mathbb{R}^k : \left| \frac{p_{u_{1,n}}\left(y_1^k\right) - g_{u_{1,n}}\left(y_1^k\right)}{g_{u_{1,n}}\left(y_1^k\right)} \right| < \delta \right\}$$

which by Theorem 9 satisfies

$$\lim_{n\to\infty} P_{u_{1,n}}\left(E_{k,\delta}\right) = \lim_{n\to\infty} G_{u_{1,n}}\left(E_{k,\delta}\right) = 1. \tag{1.30}$$

It holds

$$\sup_{C\in\mathcal{B}\left(\mathbb{R}^k\right)} \left|P_{u_{1,n}}\left(C\cap E_{k,\delta}\right) - G_{u_{1,n}}\left(C\cap E_{k,\delta}\right)\right|$$

$$\leq \delta \sup_{C\in\mathcal{B}\left(\mathbb{R}^k\right)} \int_{C\cap E_{k,\delta}} g_{u_{1,n}}\left(y_1^k\right) dy_1^k \leq \delta.$$

By (1.30)

$$\sup_{C\in\mathcal{B}\left(\mathbb{R}^k\right)} \left|P_{u_{1,n}}\left(C\cap E_{k,\delta}\right) - P_{u_{1,n}}\left(C\right)\right| < \eta_n$$

and

$$\sup_{C\in\mathcal{B}\left(\mathbb{R}^k\right)} \left|G_{u_{1,n}}\left(C\cap E_{k,\delta}\right) - G_{u_{1,n}}\left(C\right)\right| < \eta_n$$

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

$$\sup_{C\in\mathcal{B}\left(\mathbb{R}^k\right)} \left|P_{u_{1,n}}\left(C\right) - G_{u_{1,n}}\left(C\right)\right| < \delta + 2\eta_n$$

for all positive δ. Applying Scheffé's Lemma, we have proved

Theorem 10. *Under the hypotheses of Theorem 9 the total variation distance between $P_{u_{1,n}}$ and $G_{u_{1,n}}$ goes to 0 as n tends to infinity, and*

$$\lim_{n\to\infty} \int \left|p_{u_{1,n}}\left(y_1^k\right) - g_{u_{1,n}}\left(y_1^k\right)\right| dy_1^k = 0.$$

Remark 11. *This result is to be paralleled with Theorem 1.6 in [33] and Theorem 2.15 in [29] which provides a rate for this convergence for small k's under some additional conditions on the moment generating function of* **U**.

Approximation under other sampling schemes

In statistical applications the r.v.'s Y_i's in Theorems 3 and 9 may at time be sampled under some other distribution than P_{na} or G_{na}.

Consider the following situation.

The model consists in an exponential family $\mathcal{P} := \{P_{\theta,\eta}, (\theta,\eta) \in \mathcal{N}\}$ defined on $\mathbb{R}$ with canonical parametrization (θ,η) and sufficient statistics (t,u) defined on $\mathbb{R}$ through the densities

$$p_{\theta,\eta}(x) := \frac{dP_{\theta,\eta}\left(x\right)}{dx} = \exp\left[\theta t(x) + \eta u(x) - K(\theta,\eta)\right] h(x). \tag{1.31}$$

We assume that both θ and η belong to $\mathbb{R}$. The natural parameter space $\mathcal{N}$ is a convex set in $\mathbb{R}^2$ defined as the domain of

$$k(\theta,\eta) := \exp\left[K(\theta,\eta)\right] = \int \exp\left[\theta t(x) + \eta u(x)\right] h(x)dx.$$

For the statistician, θ is the parameter of interest whereas η is a nuisance one. The unknown parameter of the i.i.d. sample $\mathbf{X}_1^n := (\mathbf{X}_1, ..., \mathbf{X}_n)$ observed as $X_1^n := (X_1, ..., X_n)$ is (θ_T, η_T).

Conditioning on a sufficient statistics for the nuisance parameter produces a new exponential family which is free of η. For any θ denote $\widehat{\eta}_\theta$ the MLE of η_T in model (1.31) parametrized in η, when θ is fixed. A classical solution for the estimation of θ_T consists in maximizing the likelihood

$$L\left(\theta|\, X_1^n\right) := \prod_{i=1}^n p_{\theta,\widehat{\eta}_\theta}(X_i)$$

upon θ. This approach produces satisfactory results when $\widehat{\eta}_\theta$ is a consistent estimator of η_θ . However for curved exponential families, it may happens that for some θ the likelihood

$$L_\theta\left(\eta|\, X_1^n\right) := \prod_{i=1}^n p_{\theta,\eta}(X_i)$$

is multimodal with respect to η which may produce misestimation in $\widehat{\eta}_\theta$, leading in turn unconsistency in the resulting estimates of θ_T, see [91].

Consider $g_{u_{1,n},(\theta,\eta)}$ defined through (1.29) for fixed (θ,η), with $u_{1,n} := u(X_1) + ... + u(X_n)$. Since $u_{1,n}$ is sufficient for η , $p_{u_{1,n},(\theta,\eta)}$ is independent upon η for all k. Assume at present that the density $g_{u_{1,n},(\theta,\eta)}$ on $\mathbb{R}^k$ approximates $p_{u_{1,n},(\theta,\eta)}$ on the sample X_1^n generated under (θ_T, η_T); it follows then that inserting any value η_0 in (1.29) does not change the value of the resulting likelihood

$$L_{\eta_0}\left(\theta|\, X_1^k\right) := g_{u_{1,n},(\theta,\eta_0)}(X_i).$$

Optimizing $L_{\eta_0}\left(\theta|\, X_1^k\right)$ upon θ produces a consistent estimator of θ_T. We refer to Chapter 3 for examples and discussion.

Let $\mathbf{Y}_1^n$ be i.i.d. copies of $\mathbf{Z}$ with distribution Q and density q; assume that Q satisfies the Cramer condition $\int (\exp tx)\, q(x)dx < \infty$ for t in a non void neighborhood of 0. Let $\mathbf{V}_{1,n} := u\left(\mathbf{Y}_1\right) + ... + u\left(\mathbf{Y}_n\right)$ and define

$$q_{u_{1,n}}\left(y_1^k\right) := q\left(\mathbf{Y}_1^k = y_1^k \middle|\, \mathbf{V}_{1,n} = u_{1,n}\right)$$

with distribution $Q_{u_{1,n}}$. It then holds

Theorem 12. *Assume (K1) and (K2) together with (E1) and (E2). Then, with the same hypotheses and notation as in Theorem 9,*

$$p\left(\mathbf{X}_1^k = Y_1^k \,|\mathbf{U}_{1,n} = u_{1,n}\right) = g_{u_{1,n}}(Y_1^k)(1 + o_{Q_{u_{1,n}}}(\epsilon_n (\log n)^2)).$$

Also the total variation distance between $Q_{u_{1,n}}$ and $P_{u_{1,n}}$ goes to 0 as n tends to infinity.

Proof. It is enough to check that Lemmas 23, 24 and 25 hold when $\mathbf{Y}$ satisfies the Cramer condition. □

Remark 13. *In the previous discussion $Q = P_{\theta_T,\eta_T}$ and $\mathbf{X}_1^n$ are independent copies of $\mathbf{X}$ with distribution P_{θ,η_0}.*

1.3.2 How far is the approximation valid?

This section provides a rule leading to an effective choice of the crucial parameter k in order to achieve a given accuracy bound for the relative error in Theorem 9 (ii). The accuracy of the approximation is measured through

$$ERE(k) := E_{G_{u_{1,n}}} 1_{D_k}\left(Y_1^k\right) \frac{p_{u_{1,n}}\left(Y_1^k\right) - g_{u_{1,n}}\left(Y_1^k\right)}{p_{u_{1,n}}\left(Y_1^k\right)} \tag{1.32}$$

and

$$VRE(k) := Var_{G_{u_{1,n}}} 1_{D_k}\left(Y_1^k\right) \frac{p_{u_{1,n}}\left(Y_1^k\right) - g_{u_{1,n}}\left(Y_1^k\right)}{p_{u_{1,n}}\left(Y_1^k\right)} \tag{1.33}$$

respectively the expectation and the variance of the relative error of the approximating scheme when evaluated on

$$D_k := \left\{ y_1^k \in \mathbb{R}^k \text{ such that } \left| g_{u_{1,n}}(y_1^k)/p_{u_{1,n}}\left(y_1^k\right) - 1 \right| < \delta_n \right\}$$

with $\epsilon_n \left(\log n\right)^2 / \delta_n \to 0$ and $\delta_n \to 0$; therefore $G_{u_{1,n}}\left(D_k\right) \to 1$. The r.v's Y_1^k are sampled under $g_{u_{1,n}}$. Note that the density $p_{u_{1,n}}$ is usually unknown. The argument is somehow heuristic and informal; nevertheless the rule is simple to implement and provides good results. We assume that the set D_k can be substituted by $\mathbb{R}^k$ in the above formulas, therefore assuming that the relative error has bounded variance, which would require quite a lot of work to be proved under appropriate conditions, but which seems to hold, at least in all cases considered by the authors. We keep the above notation omitting therefore any reference to D_k.

Consider a two-sigma confidence bound for the relative accuracy for a given k, defining

$$CI(k) := \left[ERE(k) - 2\sqrt{VRE(k)}, ERE(k) + 2\sqrt{VRE(k)} \right].$$

Let δ denote an acceptance level for the relative accuracy. Accept k until δ belongs to $CI(k)$. For such k the relative accuracy is certified up to the level 5% roughly.

The calculation of $VRE(k)$ and $ERE(k)$ should be done as follows.

Write

$$\begin{aligned} VRE(k)^2 &= E_{P_{\mathbf{X}}} \left(\frac{g_{u_{1,n}}^3\left(Y_1^k\right)}{p_{u_{1,n}}\left(Y_1^k\right)^2 p_{\mathbf{X}}\left(Y_1^k\right)} \right) \\ &\quad - E_{P_{\mathbf{X}}} \left(\frac{g_{u_{1,n}}^2\left(Y_1^k\right)}{p_{u_{1,n}}\left(Y_1^k\right) p_{\mathbf{X}}\left(Y_1^k\right)} \right)^2 \\ &=: A - B^2. \end{aligned}$$

By Bayes formula

$$p_{u_{1,n}}\left(Y_1^k\right) = p_{\mathbf{X}}\left(Y_1^k\right) \frac{np\left(\mathbf{U}_{k+1,n}/(n-k) = m(t_k)\right)}{(n-k)\, p\left(\mathbf{U}_{1,n}/n = u_{1,n}/n\right)}. \tag{1.34}$$

The following Lemma holds; see [59] and [79].

Lemma 14. *Let $\mathbf{U}_1, ..., \mathbf{U}_n$ be i.i.d. random variables with common density $p_{\mathbf{U}}$ on $\mathbb{R}$ and satisfying the Cramer conditions with m.g.f. $\phi_{\mathbf{U}}$. Then with $m(t) = u$*

$$p\left(\mathbf{U}_{1,n}/n = u\right) = \frac{\sqrt{n}\phi_{\mathbf{U}}^n(t)\exp -ntu}{s(t)\sqrt{2\pi}}\left(1 + o(1)\right)$$

when $|u|$ is bounded.

Introduce

$$D := \left[\frac{\pi_{\mathbf{U}}^{m_0}(m_0)}{p_{\mathbf{U}}(m_0)}\right]^n$$

and

$$N := \left[\frac{\pi_{\mathbf{U}}^{m_k}(m_k)}{p_{\mathbf{U}}(m_k)}\right]^{(n-k)}$$

with m_k defined in (1.25) and $m_0 = u_{1,n}/n$. Define t by $m(t) = m_0$. By (1.34) and Lemma 14 it holds

$$p_{u_{1,n}}\left(Y_1^k\right) = \sqrt{\frac{n}{n-k}}p_{\mathbf{X}}\left(Y_1^k\right)\frac{D}{N}\frac{s(t)}{s(t_k)}\left(1 + o_{P_{u_{1,n}}}(1)\right).$$

The approximation of A is obtained through Monte Carlo simulation. Define

$$A\left(Y_1^k\right) := \frac{n-k}{n}\left(\frac{g_{u_{1,n}}\left(Y_1^k\right)}{p_{\mathbf{X}}\left(Y_1^k\right)}\right)^3\left(\frac{N}{D}\right)^2\frac{s^2(t_k)}{s^2(t)} \tag{1.35}$$

and simulate L i.i.d. samples $Y_1^k(l)$, each one made of k i.i.d. replications under $p_{\mathbf{X}}$. Set

$$\widehat{A} := \frac{1}{L}\sum_{l=1}^{L} A\left(Y_1^k(l)\right).$$

We use the same approximation for B. Define

$$B\left(Y_1^k\right) := \sqrt{\frac{n-k}{n}}\left(\frac{g_{u_{1,n}}\left(Y_1^k\right)}{p_{\mathbf{X}}\left(Y_1^k\right)}\right)^2\left(\frac{N}{D}\right)\frac{s(t_k)}{s(t)} \tag{1.36}$$

and

$$\widehat{B} := \frac{1}{L}\sum_{l=1}^{L} B\left(Y_1^k(l)\right)$$

with the same $Y_1^k(l)'s$ as above.

Set

$$\overline{VRE}(k) := \widehat{A} - \left(\widehat{B}\right)^2 \tag{1.37}$$

which is a fair approximation of $VRE(k)$.

The curve $k \to \overline{ERE}(k)$ is a proxy for (1.32) and is obtained through

$$\overline{ERE}(k) := 1 - \widehat{B}.$$

A proxy of $CI(k)$ can now be defined through

$$\overline{CI}(k) := \left[\overline{ERE}(k) - 2\sqrt{\overline{VRE}(k)}, \overline{ERE}(k) + 2\sqrt{\overline{VRE}(k)}\right]. \tag{1.38}$$

We now check the validity of the just above approximation, comparing $\overline{CI}(k)$ with $CI(k)$ on a toy case.

Consider $u(x) = x$. The case when $p_{\mathbf{X}}$ is a centered exponential distribution with variance 1 allows for an explicity evaluation of $CI(k)$ making no use of Lemma 14. The conditional density p_{na} is calculated analytically, the density g_{na} is obtained through (1.10), hence providing a benchmark for our proposal. The terms $\widehat{A}$ and $\widehat{B}$ are obtained by Monte Carlo simulation following the algorithm presented hereunder. Figures 1.1, 1.2 and 1.3, 1.4 show the increase in δ w.r.t. k in the large deviation range, with a such that $P(\mathbf{S}_{1,n} > na) \simeq 10^{-8}$. We have considered two cases, when $n = 100$ and when $n = 1000$. These figures show that the approximation scheme is quite accurate, since the relative error is fairly small. Also they show that $\overline{ERE}$ et $\overline{CI}$ provide good tools for the assessing the value of k.

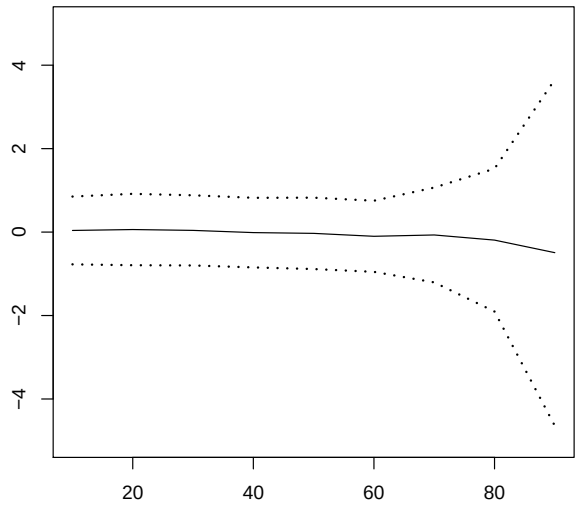

Figure 1.1: $\overline{ERE}(k)$(solid line) along with upper and lower bound of $\overline{CI}(k)$(dotted line) as a function of k with $n = 100$ and a such that $P_n \simeq 10^{-8}$.

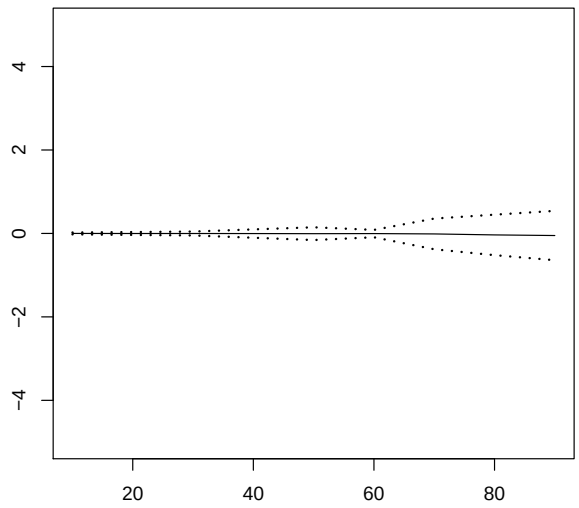

Figure 1.2: $ERE(k)$(solid line) along with upper and lower bound of $CI(k)$(dotted line) as a function of k with $n = 100$ and a such that $P_n \simeq 10^{-8}$.

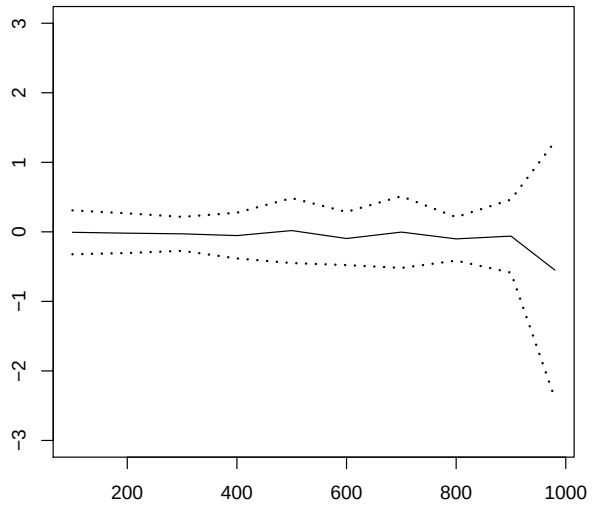

Figure 1.3: $\overline{ERE}(k)$(solid line) along with upper and lower bound of $\overline{CI}(k)$(dotted line) as a function of k with $n = 1000$ and a such that $P_n \simeq 10^{-8}$.

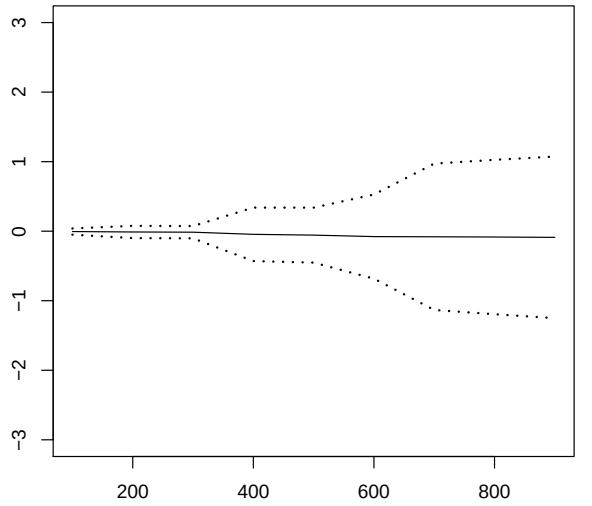

Figure 1.4: $ERE(k)$(solid line) along with upper and lower bound of $CI(k)$(dotted line) as a function of k with $n = 1000$ and a such that $P_n \simeq 10^{-8}$.

Algorithms 1 and 2 produce the curve $k \to \overline{CI}(k)$. The resulting $k = k_\delta$ is the longest size of the runs which makes $g_{u_{1,n}}$ a good proxy for $p_{u_{1,n}}$.

Input : y_1^k, $p_{\mathbf{X}}$, n, $u_{1,n}$
Output : $g_{u_{1,n}}\left(y_1^k\right)$
Initialization:
$t_0 \leftarrow m^{-1}(m_0)$;
$g_0(y_1|\, y_0) \leftarrow$ (1.24);
Procedure :
for $i \leftarrow 1$ **to** $k-1$ **do**
$m_i \leftarrow$ (1.25);
$t_i \leftarrow m^{-1}(m_i)$ $*$;
$\alpha \leftarrow$(1.27);
$\beta \leftarrow$(1.28);
Calculate C_i;
$g(y_{i+1}|\, y_1^i) \leftarrow$(1.26);
end
Compute $g_{u_{1,n}}\left(y_1^k\right) \leftarrow$(1.29);
Return : $g_{u_{1,n}}(y_1^k)$

Algorithm 1: Evaluation of $g_{u_{1,n}}(y_1^k)$

The calculation of $g_{u_{1,n}}\left(y_1^k\right)$ above requires the value of

$$C_i = \left(\int p_{\mathbf{X}}(x) \mathfrak{n}\left(\alpha\beta + m_0, \beta, u(x)\right) dx \right)^{-1}$$

This can be done through Monte Carlo simulation.

Remark 15. *Solving $t_i = m^{-1}(m_i)$ might be difficult. It may happen that the reciprocal function of m is at hand, but even when $p_{\mathbf{X}}$ is the Weibull density and $u(x) = x$, such is not the case. We can replace step $*$ by*

$$t_{i+1} := t_i - \frac{\left(m\left(t_i\right) + u_i\right)}{\left(n-i\right) s^2\left(t_i\right)}.$$

Indeed since

$$m(t_{i+1}) - m(t_i) = -\frac{1}{n-i}\left(m(t_i) + u_i\right)$$

use a first order approximation to derive that t_{i+1} can be substituted by τ_{i+1} defined through

$$\tau_{i+1} := t_i - \frac{1}{\left(n-i\right) s^2(t_i)}\left(m(t_i) + u_i\right).$$

When $\lim_{n\to\infty} u_{1,n}/n = Eu\left(\mathbf{X}\right)$, the values of the function $s^2(.)$ are close to $Var[u\left(\mathbf{X}\right)]$ and the above approximation is fair. For the large deviation case, the same argument applies, since $s^2(t_i)$ keeps close to $s^2(t^a)$.

Input : $p_{\mathbf{X}}$, δ, n, $u_{1,n}$, L
Output : k_δ
Initialization: $k = 1$
Procedure :
  while $\delta \notin \overline{CI}(k)$ **do**
    for $l \leftarrow 1$ **to** L **do**
      Simulate $Y_1^k(l)$ i.i.d. with density $p_{\mathbf{X}}$;
      $A\left(Y_1^k(l)\right)$:=(1.35) using Algorithm 1 ;
      $B\left(Y_1^k(l)\right)$:=(1.36) using Algorithm 1 ;
    end
    Calculate $\overline{CI}(k) \leftarrow$(1.38);
    $k := k + 1$;
  end
Return : $k_\delta := k$

Algorithm 2: Calculation of k_δ

1.3.3 Simulation of typical paths of a random walk under a point conditioning

By Theorem 9 (ii), $g_{u_{1,n}}$ and the density of $p_{u_{1,n}}$ get closer and closer on a family of subsets of $\mathbb{R}^k$ which bear the typical paths of the random walk under the conditional density with probability going to 1 as n increases. By Lemma 7 large sets under $P_{u_{1,n}}$ are also large sets under $G_{u_{1,n}}$. It follows that longs runs of typical paths under $p_{u_{1,n}}$ can be simulated as typical paths under $g_{u_{1,n}}$ defined in (1.29) at least for large n.

The simulation of a sample X_1^k with $g_{u_{1,n}}$ can be fast as easy when $\lim_{n\to\infty} u_{1,n}/n = Eu(\mathbf{X})$. Indeed the r.v. $\mathbf{X}_{i+1}$ with density $g\left(x_{i+1}|x_1^i\right)$ is obtained through a standard acceptance -rejection algorithm. The values of the parameters which appear in the gaussian component of $g\left(x_{i+1}|x_1^i\right)$ in (1.7) are easily calculated, and the dominating density can be chosen for all i as $p_{\mathbf{X}}$. The constant in the acceptance rejection algorithm is then $1/\sqrt{2\pi\beta}$. This is in contrast with the case when the conditioning value is in the range of a large deviation event, i.e. $\lim_{n\to\infty} u_{1,n}/n \neq Eu(\mathbf{X})$, which appears in a natural way in Importance sampling estimation for rare event probabilities; then MCMC techniques can be used.

Denote $\mathfrak{N}$ the c.d.f. of a normal variate with parameter (μ, σ^2) ,and $\mathfrak{N}^{-1}$ its inverse.

Remark 16. *Simulation of Y_1 can be performed through the method suggested in [2].*

Figures 1.5, 1.6, 1.7 and 1.8 present a number of simulations of random walks conditioned on their sum with $n = 1000$ when $u(x) = x$. In the gaussian case, when the approximating scheme is known to be optimal up to $k = n-1$, the simulation is performed with $k = 999$ and two cases are considered: the moderate deviation case is supposed to be modeled when $P(\mathbf{S}_{1,n} > na) = 10^{-2}$ (Figure 1.5); that this range of probability is in the "moderate deviation" range is a commonly assessed statement among statisticians; the large deviation case pertains to $P(\mathbf{S}_{1,n} > na) = 10^{-8}$ (Figure 1.6). The centered exponential case with $n = 1000$ and $k = 800$ is presented in Figures 1.7 and 1.8, under the same events.

Input : p, μ, σ^2
Output : Y
Initialization:
Select a density f on $[0,1]$ and a positive constant K such that $p\left(\mathfrak{N}^{-1}(x)\right) \leq Kf(x)$ for all x in $[0,1]$
Procedure : **while** $Z > p\left(\mathfrak{N}^{-1}(X)\right)$ **do**
  Simulate X with density f;
  Simulate U uniform on $[0,1]$ independent of X;
  Compute $Z := KUf(X)$;
end
Return : $Y := \mathfrak{N}^{-1}(X)$

Algorithm 3: Simulation of Y with density proportional to $p(x)\mathfrak{n}\left(\mu, \sigma^2, x\right)$

Input : $p_{\mathbf{X}}$, δ, n, $u_{1,n}$
Output : Y_1^k
Initialization:
Set $k \leftarrow k_\delta$ with Algorithm 2;
$t_0 \leftarrow= m^{-1}(m_0)$;
Procedure :
Simulate Y_1 with density (1.24);
$u_{1,1} \leftarrow u(Y_1)$;
for $i \leftarrow 1$ **to** $k-1$ **do**
  $m_i \leftarrow$(1.25);
  $t_i \leftarrow m^{-1}(m_i)$;
  $\alpha \leftarrow$(1.27);
  $\beta \leftarrow$(1.28);
  Simulate Y_{i+1} with density $g(y_{i+1}|\, y_1^i)$ using Algorithm 3;
  $u_{1,i+1} \leftarrow u_{1,i} + u(Y_{i+1})$;
end
Return : Y_1^k

Algorithm 4: Simulation of a sample Y_1^k with density $g_{u_{1,n}}$

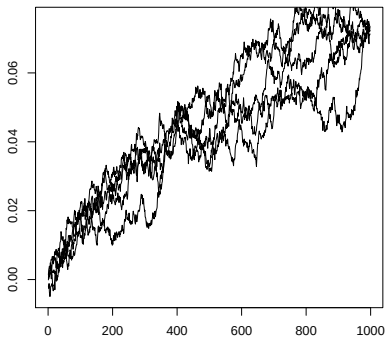

Figure 1.5: Trajectories in the normal case for $P_n = 10^{-2}$

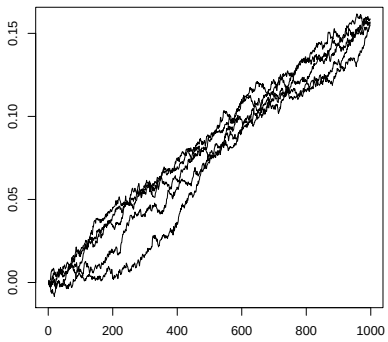

Figure 1.6: Trajectories in the normal case for $P_n = 10^{-8}$

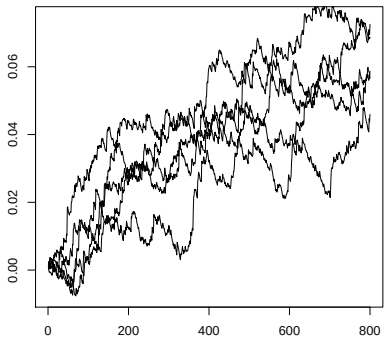

Figure 1.7: Trajectories in the exponential case for $P_n = 10^{-2}$

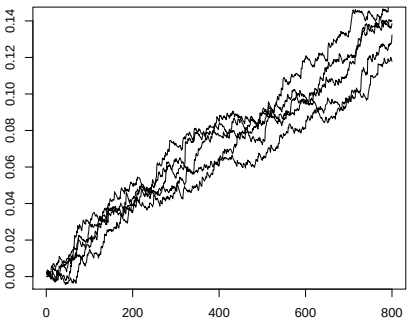

Figure 1.8: Trajectories in the exponential case for $P_n = 10^{-8}$

In order to check the accuracy of the approximation, Figures 1.9, 1.10 (normal case, n=1000, k=999) and Figures 1.11, 1.12 (centered exponential case, n=1000, k=800) present the histograms of the simulated $\mathbf{X}_i's$ together with the tilted densities at point a which are known to be the limit density of $\mathbf{X}_1$ conditioned on $\mathcal{E}_n$ in the large deviation case, and to be equivalent to the same density in the moderate deviation case, as can be deduced from [44]. The tilted density in the gaussian case is the normal with mean a and variance 1; in the centered exponential case the tilted density is an exponential density on $(-1, \infty)$ with parameter $1/(1+a)$.

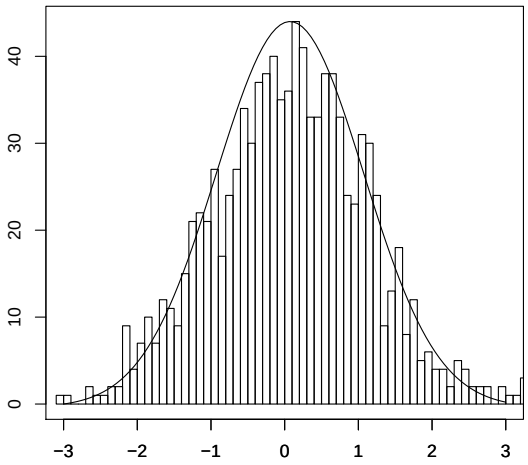

Figure 1.9: Histogram of the $\mathbf{X}_i's$ in the normal case with $n = 1000$ and $k = 999$ for $P_n = 10^{-2}$. The curve represents the associated tilted density.

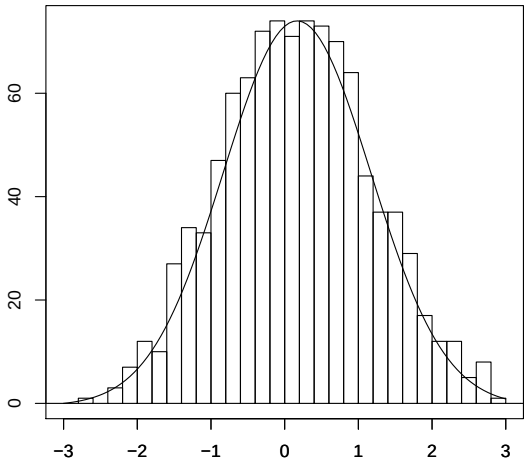

Figure 1.10: Histogram of the $\mathbf{X}_i's$ in the normal case with $n = 1000$ and $k = 999$ for $P_n = 10^{-8}$. The curve represents the associated tilted density.

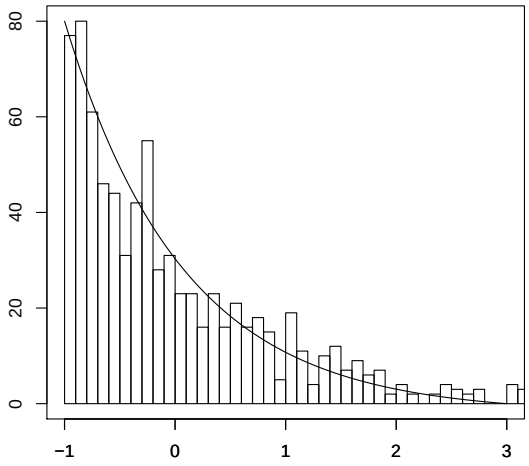

Figure 1.11: Histogram of the $\mathbf{X}_i's$ in the exponential case with $n = 1000$ and $k = 800$ for $P_n = 10^{-2}$. The curve represents the associated tilted density.

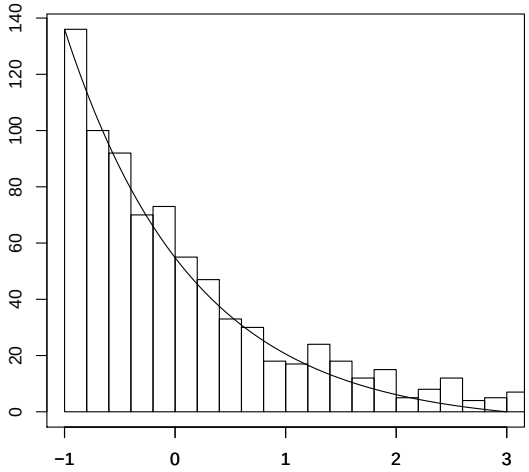

Figure 1.12: Histogram of the $\mathbf{X}_i's$ in the exponential case with $n = 1000$ and $k = 800$ for $P_n = 10^{-8}$. The curve represents the associated tilted density.

Consider now the case when $u(x) = x^2$. Figure 1.13 presents the case when $\mathbf{X}$ is $N(0,1)$, $n = 1000, k = 800$, $P(\mathbf{U}_{1,n} = u_{1,n}) \simeq 10^{-2}$. We present the histograms of the X_i's together with the graph of the corresponding tilted density; when $\mathbf{X}$ is $N(0,1)$ then $\mathbf{X}^2$ is χ^2. It is well known that when $u_{1,n}/n$ is fixed larger than 1 then the limit distribution of $\mathbf{X}_1$ conditioned on $(\mathbf{U}_{1,n} = u_{1,n})$ tends to $N(0,a)$ which is the Kullback-Leibler projection of $N(0,1)$ on the set of all probability measures Q on $\mathbb{R}$ with $\int x^2 dQ(x) = a := \lim_{n\to\infty} u_{1,n}/n$. This distribution is precisely $g_0(y_1|\, y_0)$ defined hereabove. Also consider (1.26); expansion using the definitions (1.27) and (1.28) prove that as $n \to \infty$ the dominating term in $g_i(y_{i+1}|\, y_1^i)$ is precisely $N(0, m_0)$, and the terms including y_{i+1}^4 in the exponential stemming from $\mathfrak{n}(\alpha\beta + m_0, \beta, u(y_{i+1}))$ are of order $O(1/(n-i))$; the terms depending on y_1^i are of smaller order. The fit which is observed in Figure 1.13 is in accordance with the above statement in the LDP range (when $\lim_{n\to\infty} u_{1,n}/n \neq 1$), and with the MDP approximation when $\lim_{n\to\infty} u_{1,n}/n = 1$ and $\liminf_{n\to\infty} (u_{1,n} - n)/\sqrt{n} \neq 0$, following [44] .

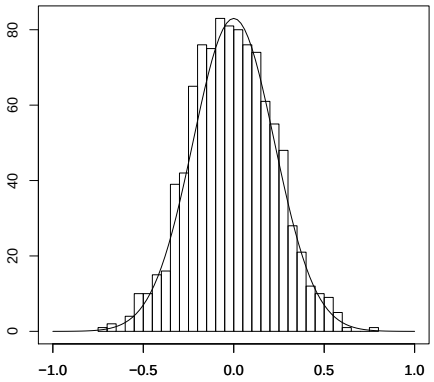

Figure 1.13: Histogram of the $\mathbf{X}_i's$ in the normal case with $n = 1000$, $k = 800$ and $u(x) = x^2$ for $P_n = 10^{-2}$. The curve represents the associated tilted density.

1.4 Conditioning on large sets

Approximation of the density

$$p_{A_n}\left(\mathbf{X}_1^k = Y_1^k\right) := p\left(\mathbf{X}_1^k = Y_1^k \middle|\, \mathbf{U}_{1,n} \in A_n\right)$$

of the runs $\mathbf{X}_1^k$ under large sets $(\mathbf{U}_{1,n} \in A_n)$ for Borel sets A_n with non void interior follows from the above results through integration. Here, in the same vein as previously, Y_1^k is generated under P_{A_n}. Applications of this result for the evaluation of rare event probabilities through Importance Sampling are presented in Chapter 2. The present section pertains to the large deviation case.

We focus on cases when $(\mathbf{U}_{1,n} \in A_n)$ writes $(\mathbf{U}_{1,n}/n \in A)$ where A is a fixed Borel set (independent on n) with essential infimum α larger than $E\mathbf{U}$ and which can be described as a "thin" or "thick" Borel set according to its local density at point α.

The starting point is the approximation of p_{nv} on $\mathbb{R}^k$ for large values of k under the point condition

$$\mathbf{U}_{1,n}/n = v$$

when v belongs to A. Denote g_{nv} the corresponding approximation defined in (1.29). It holds

$$p_{nA}(x_1^k) = \int_A p_{nv}\left(\mathbf{X}_1^k = x_1^k\right) p(\mathbf{U}_{1,n}/n = v | \mathbf{U}_{1,n} \in nA) dv. \tag{1.39}$$

In contrast with the classical Importance Sampling approach for this problem we do not consider the dominating point approach but merely realize a sharp approximation of the integrand at any point of the domain A and consider the dominating contribution of all those distributions in the evaluation of the conditional density p_{nA}. A similar point of view has been considered in [4] for sharp approximations of Laplace type integrals in $\mathbb{R}^d$.

Turning to (1.39) it appears that what is needed is a sharp approximation for

$$p(\mathbf{U}_{1,n}/n = v | \mathbf{U}_{1,n} \in nA) = \frac{p(\mathbf{U}_{1,n}/n = v)\mathbb{1}_A(v)}{P(\mathbf{U}_{1,n} \in nA)} \tag{1.40}$$

with some uniformity for v in A. We will assume that A is bounded above in order to avoid further regularity assumptions on the distribution of $\mathbf{U}$.

1.4.1 Conditioning on a large set defined through the density of its dominating point

Recall that the *essential infimum* $\texttt{essinf}A = \alpha$ of the set A with respect to the Lebesgue measure is defined through

$$\alpha := \inf \{x : \text{for all } \varepsilon > 0, |[x, x+\varepsilon] \cap A| > 0\}$$

with $\inf \varnothing := -\infty$.

We assume that $\alpha > -\infty$, which amounts to say that we do not consider very thin sets (for example not Cantor-type sets).

The density of the point α in A will not be measured in the ordinary way, through

$$d(\alpha) := \lim_{\varepsilon \to 0} \frac{|A \cap [\alpha - \varepsilon, \alpha + \varepsilon]|}{\varepsilon}$$

but merely through the more appropriate quantity

$$M(t) := t \int_{A-\alpha} e^{-ty} dy \ , \qquad t > 0.$$

For any set A, $0 \le M(t) \le 1$. If there exists an interval $[\alpha, \alpha + \varepsilon] \subset A$ then $\lim_{t\to\infty} M(t) = 1$. As an example, for a self similar set $A := A_p$ defined through $A_p := \bigcup_{n\in\mathbb{Z}} p^n I_p$ where $p > 2$ and $I_p := [(p-1)/p, 1]$ it holds $0 = \text{essinf} A_p$ and $pA_p = A_p$. Consequently for any $t \ge 0$, $M(tp) = M(t)$ and $M(tp) = M(t)$ for all $t \ge 0$; it follows that

$$\inf_{1\le u\le p} M(u) = \lim \inf_{t\to\infty} M(t) \le \lim \sup_{t\to\infty} M(t) = \sup_{1\le u\le p} M(u).$$

Define

$$M_n(t) := M(nt)/t = \int_{A-\alpha} e^{-ty} dy$$

and

$$\Psi_n(t) := n \log \phi_{\mathbf{U}}(t) + \log M_n(t) - n\alpha t$$

for all $t > 0$ such that $\phi_{\mathbf{U}}(t)$ is finite. We borrow from [3] the following results.

Define $\mu_n(t) := (1/n) \log M_n(t)$ which is for all $n \ge 1$ a decreasing function of t on $(0, \infty)$, and which is negative for large n. Also $\mu_n'(t) = \mu_1'(nt)$ and μ_1' is non decreasing on $(0, \infty)$.

Let $\overline{\mu} := \lim_{t\to\infty} \mu_1'(t)$ and $\underline{\mu} := \lim_{t\to 0} \mu_1'(t)$. Then [3] it holds

Lemma 17. *Under the above notation and hypotheses, the equation* $\Psi_n'(t) = 0$ *has a unique solution* t_n *in* $(0, t_0)$ *for* α *in* $(E\mathbf{U}+\overline{\mu}, \infty)$ *where* $t_0 := \sup\{t : \phi_{\mathbf{U}}(t) < \infty\}$. *Furthermore if* $\alpha > EU + \underline{\mu}$ *then there exists a compact set* $K \subset (0, t_0)$ *such that* $t_n \in K$ *for all* n.

Assume that $\alpha > E\mathbf{U}+\underline{\mu}$. Define $\psi_n(t) := \Psi_n''(t)$ and suppose that for any $\lambda > 0$

$$\lim_{n\to\infty} \sup_{|u|<\lambda} \frac{\psi_n\left(t_n + \frac{u}{\sqrt{\psi_n(t_n)}}\right)}{\psi_n(t_n)} = 1 \tag{1.41}$$

where t_n solves $\Psi_n'(t) = 0$ in the range $(0, t_0)$. It can be proved that (1.41) holds, for example, when $t \to \log M(t)/t$ is a regularly varying function at infinity with index $\rho \in (0, 1)$,i.e. $\log M(t)/t \in \mathcal{R}_\rho(\infty)$; see [3], Lemma 2.2.

We also assume

$$\limsup_{t\to\infty} t(\log M(t))'' < \infty \tag{1.42}$$

which holds for example when $\log(M(t)/t) \in \mathcal{R}_\rho(\infty)$, for $0 \le \rho < 1$.

Theorem 2.1 in [3] provides a general result to be inserted in (1.40); we take the occasion to correct a misprint in this result.

Theorem 18. *Assume (1.41) and (1.42) together with the aforementioned conditions on the r.v.* $\mathbf{U}$. *Then for* $\alpha > E\mathbf{U}+\underline{\mu}$

$$P(\mathbf{U}_{1,n} \in nA) = \frac{\phi_{\mathbf{U}}^n(t_n) M_n(t_n) e^{-nt_n\alpha}}{\sqrt{\psi_n(t_n)}\sqrt{2\pi}} (1 + o(1)) \quad \text{as } n \to \infty, \tag{1.43}$$

with t_n *satisfying* $\Psi_n'(t) = 0$ *provided that the function* $x \to P(\mathbf{U}_{1,n} \in nA + x)$ *is nonincreasing for* n *large enough. In particular, this last condition holds if*

(i) (Petrov) : $A = (\alpha, \infty)$ *or* $A = [\alpha, \infty)$; *in this case* $M_n(t) = 1/t$; *note that in this case the classical result is slightly different, since*

$$P(\mathbf{U}_{1,n} > na) = \frac{\phi_{\mathbf{U}}^n(t^a) e^{-nt^a a}}{t^a s(t^a)\sqrt{2\pi}} (1 + o(1)) \quad \text{as } n \to \infty$$

with $m(t^a) = a$ *and* $a > E\mathbf{U}$; *this is readily seen to be equivalent to (1.43) when* $A = (a, \infty)$.

(ii)$\mathbf{U}$ *has a symmetric unimodal distribution*

(iii)$\mathbf{U}$ *has a strongly unimodal distribution.*

The shape of A near α is reflected in the behavior of the function $M(t)$ for large values of t. As such, the larger n, the more relevant is the shape of A near α.

Note further that $M_n(t)e^{-nt\alpha} = \int_A e^{-nty}dy$ from which we see that α plays no role in (1.43). Hence α can be replaced by any number γ such that $\int_{A-\gamma} e^{-ty}dy$ converges. Further t_n is independent on α. The so-called dominating point α of A can therefore be defined through

$$\alpha := \lim_{t\to\infty} \log \int_A e^{-ty}dy.$$

In order to examine further the role played in (1.43) by the regularity of A near its essential infimum α introduce the pointwise Hölder dimension of A at α as

$$\delta(\alpha) := \frac{\log G(\varepsilon)}{-\log \varepsilon}$$

where

$$G(\varepsilon) := |A \cap [\alpha, \alpha + \varepsilon]| \quad \text{for positive } \varepsilon.$$

We refer to Proposition 2.1 in [3] for a set of Abel-Tauber type results which link the properties of $M(t)$ at infinity with those of G at 0. For example it follows that $G(\varepsilon) \sim \varepsilon^{\delta(\alpha)}$ (as $\varepsilon \to 0$) iff $M(t) \sim ct^{-\delta(\alpha)+1}\Gamma(1+\delta(\alpha))$ (as $t \to \infty$). Consequently if $M_n(t) \to 1$ as $t \to \infty$ then $M(t) \sim t$ as $t \to \infty$ and $G(\varepsilon) \sim \varepsilon$ as $\varepsilon \to 0$.

Asymptotic formulas for the numerator in (1.40) are well known and have a long history, going back to [79]. It holds

$$p(\mathbf{U}_{1,n}/n = v) = \frac{\sqrt{n}e^{nvt^v}\phi_{\mathbf{U}}(t^v)}{\sqrt{2\pi}s(t^v)}(1 + o(1)) \quad \text{as } n \to \infty \tag{1.44}$$

with t^v defined through $m(t^v) = v$.

Plug-in (1.44) and (1.43) in (1.39) provides an expression for the density of the runs. For applications the only relevant case is developed in the following paragraph.

1.4.2 Conditioning on a thick set

In the case when $A = (a, \infty)$ or with $a > Eu(\mathbf{X})$ or, more generally, when A is a thick set in a neighborhood of its essential infimum (i.e. when $\lim_{t\to\infty} M(t) = 1$) a simple asymptotic evaluation for (1.40) when A is unbounded can be obtained. Indeed a development in the ratio yields

$$p(\mathbf{U}_{1,n}/n = v | \mathbf{U}_{1,n} > na) = (nt \exp -nt(v-a))\mathbb{1}_A(v)(1 + o(1)) \tag{1.45}$$

with $m(t) = a$, indicating that $\mathbf{U}_{1,n}/n$ is roughly exponentially distributed on A with expectation $a + 1/nt$. This result is used in Chapter 3 in order to derive estimators of some rare event probabilities through Importance sampling.

In order to get a sharp approximation for $p_{nA}\left(\mathbf{X}_1^k = Y_1^k\right)$ it is necessary to introduce an interval $(a, a + c_n)$ which bears the principal part of the integral (1.39).

Let c_n denote a positive sequence such that the following condition (C) holds

$$\lim_{n\to\infty} nc_n m^{-1}(a) = \infty$$
$$\sup_{n\geq 1} \frac{nc_n}{(n-k)} < \infty$$

and denote c the current term c_n.

Denote (A) the following set of conditions

$$\lim_{n\to\infty} (n-k)\left(m^{-1}(a)\right)^2 = \infty$$
$$\lim_{n\to\infty} \frac{m^{-1}(a)}{\epsilon_n} = \infty$$

which trivially holds when $\lim_{n\to\infty} a_n > E[\mathbf{U}]$.

Define on $\mathbb{R}^k$ the density

$$g_{nA}(y_1^k) := \frac{nm^{-1}(a)\int_a^{a+c} g_{nv}(y_1^k)\left(\exp -nm^{-1}(a)(v-a)\right)dv}{1 - \exp -nm^{-1}(a)c}. \tag{1.46}$$

The density

$$\frac{nm^{-1}(a)\left(\exp -nm^{-1}(a)(v-a)\right)\mathbb{1}_{(a,a+c)}(v)}{1-\exp -nm^{-1}(a)\,c} \tag{1.47}$$

which appears in (1.46) approximates $p(\mathbf{U}_{1,n}/n = v \,|\, a < \mathbf{U}_{1,n}/n < a+c)$, as follows from standard limit results; see [59], Corollary 6.4.1 and [79] in the LD range, and in [53] in the MD range.

The *variance function* V of the distribution of $\mathbf{U}$ is defined on the span of $\mathbf{U}$ through

$$v \to V(v) := s^2(m^{-1}(v))$$

Denote (V) the condition

$$\sup_{n\geq 1}\left(\sqrt{n}m^{-1}(a)\int_a^\infty V'(v)\left(\exp\left(-nm^{-1}(a)(v-a)\right)\right)dv\right) < \infty. \tag{V}$$

Theorem 19. *Assume (E1),(E2), (C), (V) and (A). Then for any positive $\delta < 1$*

(i)

$$p_{nA}\left(\mathbf{X}_1^k = Y_1^k\right) = g_{nA}(Y_1^k)(1 + o_{P_{nA}}(\delta_n)) \tag{1.48}$$

and (ii)

$$p_{nA}\left(\mathbf{X}_1^k = Y_1^k\right) = g_{nA}(Y_1^k)(1 + o_{G_{nA}}(\delta_n)) \tag{1.49}$$

where

$$\delta_n := \max\left(\epsilon_n (\log n)^2, \left(\exp\left(-ncm^{-1}(a)\right)\right)^\delta\right). \tag{1.50}$$

Proof. See Appendix. □

Remark 20. *Most distributions used in statistics satisfy (V); numerous papers have focused on the properties of variance functions and classification of distributions; see e.g.[66] and references therein.*

Remark 21. *When a is fixed then (A) holds. In the case when a_n converges to $E[\mathbf{U}]$, the CLT zone is not in the range of applicability of the present approach; the classical normal approximation should obviously be used in this range. In the MD or LD range, defining k and ϵ_n according to (A), (C), (E1) and (E2) is always possible. In the MD range, the slowest a_n converges to $E[\mathbf{U}]$, the largest k can be chosen.*

Corollary 22. *Under the hypotheses of Theorem 19 the total variation distance between P_{nA} and G_{nA} goes to 0 as n tends to infinity, i.e.*

$$\lim_{n\to\infty}\int\left|p_{nA}\left(y_1^k\right) - g_{nA}\left(y_1^k\right)\right|dy_1^k = 0.$$

1.4.3 How far is the approximation valid?

In Section 1.3.2, a similar question is addressed and a proxy of the curve $\delta \to k_\delta$ is provided in order to define the maximal k leading to a given relative accuracy under the point condition $(\mathbf{U}_{1,n} = na)$, namely when p_{nA} is replaced by p_{na} and g_{nA} by g_{na}.

Consider the ratio $g_{nA}(Y_1^k)/p_{nA}\left(Y_1^k\right)$ and use Cauchy's mean value theorem to obtain

$$g_{nA}(Y_1^k)/p_{nA}\left(Y_1^k\right)$$

$$= \frac{\int_a^{a+c} g_{nv}(\mathbf{X}_1^k = Y_1^k) \left(\exp\left(-nm^{-1}(a)(v-a)\right)\right) dv}{\int_a^{a+c} p_{nv}(\mathbf{X}_1^k = Y_1^k) \left(\exp\left(-nm^{-1}(a)(v-a)\right)\right) ds} \\ (1 + o_{G_{nA}}(1)) \\ = \frac{g_{n\alpha}(\mathbf{X}_1^k = Y_1^k)}{p_{n\alpha}(\mathbf{X}_1^k = Y_1^k)} (1 + o_{G_{nA}}(1))$$

for some α between a and $a + c$. Since c is small (see condition (C)), it is reasonable to substitute α by a in order to evaluate the accuracy of the approximation. We thus inherit of the relative efficiency curve in Section 1.3.2, to which we refer for definition and derivation.

1.5 Appendix

For clearness the current term a_n is denoted a in all proofs.

1.5.1 Three Lemmas pertaining to the partial sum under its final value

We state three lemmas which describe some functions of the random vector $\mathbf{X}_1^n$ conditioned on $\mathcal{E}_n$. The r.v. $\mathbf{X}$ is assumed to have expectation 0 and variance 1.

Lemma 23. *It holds* $E_{P_{na}}(\mathbf{X}_1) = a, E_{P_{na}}(\mathbf{X}_1\mathbf{X}_2) = a^2 + 0\left(\frac{1}{n}\right), E_{P_{na}}(\mathbf{X}_1^2) = s^2(t) + a^2 + 0\left(\frac{1}{n}\right)$ *where* $m(t) = a$.

Proof. Using

$$p_{na}(\mathbf{X}_1 = x) = \frac{p_{\mathbf{S}_{2,n}}(na - x)\, p_{\mathbf{X}_1}(x)}{p_{\mathbf{S}_{1,n}}(na)} = \frac{\pi^a_{\mathbf{S}_{2,n}}(na - x)\, \pi^a_{\mathbf{X}_1}(x)}{\pi^a_{\mathbf{S}_{1,n}}(na)}$$

normalizing both $\pi^a_{\mathbf{S}_{2,n}}(na - x)$ and $\pi^a_{\mathbf{S}_{1,n}}(na)$ and making use of a first order Edgeworth expansion in those expressions yields $E_{P_{na}}(\mathbf{X}_1^2) = s^2(t) + a^2 + 0\left(\frac{1}{n}\right)$. A similar development for the joint density $p_{na}(\mathbf{X}_1 = x, \mathbf{X}_2 = y)$, with the same tilted distribution π^a produces the limit expression of $E_{P_{na}}(\mathbf{X}_1\mathbf{X}_2)$. □

Lemma 24. *Assume (E1). Then (i)* $\max_{1\le i\le k} |m_i| = a + o_{P_{na}}(\epsilon_n)$. *Also (ii)* $\max_{1\le i\le k} s_i^2$, $\max_{1\le i\le k} \mu_3^i$ *and* $\max_{1\le i\le k} \mu_4^i$ *tend in* P_{na} *probability to the variance, skewness and kurtosis of* $\pi^{\underline{a}}$ *where* $\underline{a} := \lim_{n\to\infty} a_n$.

Proof. (i) Define

$$V_{i+1} := m(t_i) - a \\ = \frac{S_{i+1,n}}{n - i} - a.$$

We state that

$$\max_{0\le i\le k-1} |V_{i+1}| = o_{P_{na}}(\epsilon_n), \tag{1.51}$$

namely for all positive δ

$$\lim_{n\to\infty} P_{na}\left(\max_{0\le i\le k-1} |V_{i+1}| > \delta\epsilon_n\right) = 0$$

which we obtain following the proof of Kolmogorov maximal inequality. Define

$$A_i := ((|V_{i+1}| \geq \delta\epsilon_n) \text{ and } (|V_j| < \delta\epsilon_n \text{ for all } j < i+1)).$$

from which

$$\left(\max_{0\leq i\leq k-1} |V_{i+1}| > \delta\epsilon_n\right) = \bigcup_{i=0}^{k-1} A_i.$$

It holds

$$\begin{aligned}
E_{P_{na}} V_k^2 &= \int_{\cup A_i} V_k^2 dP_{na} + \int_{(\cup A_i)^c} V_k^2 dP_{na} \\
&\geq \int_{\cup A_i} \left(V_i^2 + 2(V_k - V_i) V_i\right) dP_{na} + \int_{(\cup A_i)^c} \left(V_i^2 + 2(V_k - V_i) V_i\right) dP_{na} \\
&\geq \int_{\cup A_i} V_i^2 dP_{na} \\
&\geq \delta^2\epsilon_n^2 \sum_{j=0}^{k-1} P_{na}(A_j) \\
&= \delta^2\epsilon_n^2 P_{na}\left(\max_{0\leq i\leq k-1} |V_{i+1}| > \delta\epsilon_n\right).
\end{aligned}$$

The third line above follows from $EV_i(V_k - V_i) = 0$ which is proved hereunder. Hence

$$P_{na}\left(\max_{0\leq i\leq k-1} |V_{i+1}| > \delta\epsilon_n\right) \leq \frac{Var_{P_{na}}(V_k)}{\delta^2\epsilon_n^2} = \frac{1}{\delta^2\epsilon_n^2 (n-k)}(1 + o(1))$$

where we used Lemma 23; therefore (1.51) holds under (E1). Direct calculation yields $E_{P_{na}}(V_i(V_k - V_i)) = 0$, which achieves the proof of (i).

(ii) follows from (i) since $\lim_{n\to\infty} \max_{1\leq i\leq k} m(t_i) = \underline{a}$. □

We also need the order of magnitude of $\max(|\mathbf{X}_1|, ..., |\mathbf{X}_k|)$ under P_{na} which is stated in the following result.

Lemma 25. *It holds* $\max(|\mathbf{X}_1|, ..., |\mathbf{X}_n|) = O_{P_{na}}(\log n)$.

Proof. Set $|\mathbf{X}_1| := \mathbf{X}_i^- + \mathbf{X}_i^+$ with $\mathbf{X}_i^- := -\min(0, \mathbf{X}_i)$, $\mathbf{X}_i^+ := \max(0, \mathbf{X}_i)$; it is enough to prove that $\max_i \mathbf{X}_i^- = O_{P_{na}}(\log n)$ and $\max_i \mathbf{X}_i^+ = O_{P_{na}}(\log n)$. Since $E\exp t\mathbf{X}$ is finite in a non void neighborhood of 0 so are $E\exp t\mathbf{X}^-$ and $E\exp t\mathbf{X}^+$. We hence prove the Lemma for positive r.v's $\mathbf{X}_i$'s only.

Denote a the current term of the sequence a. For all t it holds

$$\begin{aligned}
P_{na}(\max(\mathbf{X}_1, ..., \mathbf{X}_n) > t) &\leq nP_{na}(\mathbf{X}_n > t) \\
&= n\int_t^\infty \pi^a(\mathbf{X}_n = u) \frac{\pi^a(\mathbf{S}_{1,n-1} = na - u)}{\pi^a(\mathbf{S}_{1,n} = na)} du.
\end{aligned}$$

Let τ be such that $m(\tau) = a$. Denote $s := s(\tau)$. Center and normalize both $\mathbf{S}_{1,n}$ and $\mathbf{S}_{1,n-1}$with respect to the density π^a in the last line above, denoting $\overline{\pi_n^a}$ the density of $\overline{\mathbf{S}_{1,n}} := (\mathbf{S}_{1,n} - na)/s\sqrt{n}$ when $\mathbf{X}$ has density π^a with mean a and variance s^2, we get

$$\begin{aligned}
P_{na}(\max(\mathbf{X}_1, ..., \mathbf{X}_n) > t) \leq n\frac{\sqrt{n}}{\sqrt{n-1}} \int_t^\infty \pi^a(\mathbf{X}_n = u) \\
\frac{\overline{\pi_{n-1}^a}\left(\overline{\mathbf{S}_{1,n-1}} = (na - u - (n-1)a)/(s\sqrt{n-1})\right)}{\overline{\pi_n^a}\left(\overline{\mathbf{S}_{1,n}} = 0\right)} du.
\end{aligned}$$

Under the sequence of densities π^a the triangular array $(\mathbf{X}_1, ..., \mathbf{X}_n)$ obeys a first order Edgeworth expansion

$$\begin{aligned} P_{na}\left(\max\left(\mathbf{X}_1, ..., \mathbf{X}_n\right) > t\right) \leq n\frac{\sqrt{n}}{\sqrt{n-1}}\int_t^\infty \pi^a\left(\mathbf{X}_n = u\right) \\ \frac{\mathfrak{n}\left(\left(a-u\right)/s\sqrt{n-1}\right)\mathbf{P}\left(u,i,n\right)+o(1)}{\mathfrak{n}\left(0\right)+o(1)}du \\ \leq nCst\int_t^\infty \pi^a\left(\mathbf{X}_n = u\right)du. \end{aligned}$$

for some constant Cst independent of n and τ and

$$\mathbf{P}\left(u,i,n\right) := 1 + P_3\left(\left(a-u\right)/s\sqrt{n-1}\right)$$

where $P_3(x) = \frac{\mu_3}{6s^3}\left(x^3 - 3x\right)$ is the third Hermite polynomial; s^2 and μ_3 are the second and third centered moments of π^a. We have used the fact that the sequence a converges to bound all moments of the tilted densities π^a. We used uniformity upon u in the remaining term of the Edgeworth expansions. Making use of Chernoff Inequality to bound $\Pi^a\left(\mathbf{X}_n > t\right)$,

$$P_{na}\left(\max\left(\mathbf{X}_1, ..., \mathbf{X}_n\right) > t\right) \leq nCst\frac{\Phi(t+\lambda)}{\Phi(t)}e^{-\lambda t}$$

for any λ such that $\phi(t+\lambda)$ is finite. For t such that

$$t/\log n \to \infty$$

it holds

$$P_{na}\left(\max\left(\mathbf{X}_1, ..., \mathbf{X}_n\right) < t\right) \to 1,$$

which proves the lemma. □

1.5.2 Proof of the approximations resulting from Edgeworth expansions in Theorem 1

We complete the calculation leading to (1.15) and (1.16).
Set $Z_{i+1} := \left(m_i - Y_{i+1}\right)/s_i\sqrt{n-i-1}$.
It then holds

$$\begin{aligned} \overline{\pi_{n-i-1}}\left(Z_{i+1}\right) = \mathfrak{n}(Z_{i+1})\left[\begin{array}{c} 1 + \frac{1}{\sqrt{n-i-1}}P_3(Z_{i+1}) + \frac{1}{n-i-1}P_4(Z_{i+1}) \\ + \frac{1}{(n-i-1)^{3/2}}P_5(Z_{i+1}) \end{array}\right] \\ + O_{P_{na}}\left(\frac{P_5(Z_{i+1})}{(n-i-1)^{3/2}}\right). \end{aligned} \tag{1.52}$$

We perform an expansion in $\mathfrak{n}(Z_{i+1})$ up to the order 3, with a first order term $\mathfrak{n}\left(-Y_{i+1}/\left(s_i\sqrt{n-i-1}\right)\right)$, namely

$$\begin{aligned} \mathfrak{n}(Z_{i+1}) = \mathfrak{n}\left(-Y_{i+1}/\left(s_i\sqrt{n-i-1}\right)\right) \\ \left(\begin{array}{c} 1 + \frac{Y_{i+1}m_i}{s_i^2(n-i-1)} + \frac{m_i^2}{2s_i^2(n-i-1)}\left(\frac{Y_{i+1}^2}{s_i^2(n-i-1)} - 1\right) \\ + \frac{m_i^3}{6s_i^3(n-i-1)^{3/2}}\frac{\mathfrak{n}^{(3)}\left(\frac{Y^*}{(s_i\sqrt{n-i-1})}\right)}{\mathfrak{n}\left(-Y_{i+1}/\left(s_i\sqrt{n-i-1}\right)\right)} \end{array}\right) \end{aligned} \tag{1.53}$$

where $Y^* = \frac{1}{s_i\sqrt{n-i-1}}(-Y_{i+1} + \theta m_i)$ with $|\theta| < 1$.

Lemmas 24 and 25 provide the orders of magnitude of the random terms in the above displays when sampling under P_{na}.

Use those lemmas to obtain

$$\frac{Y_{i+1}m_i}{s_i^2\,(n-i-1)} = \frac{Y_{i+1}}{n-i-1}\,(a + o_{P_{na}}(\epsilon_n)) \tag{1.54}$$

and

$$\frac{m_i^2}{s_i^2\,(n-i-1)} = \frac{1}{n-i-1}\,(a + o_{P_{na}}(\epsilon_n))^2\,.$$

Also when (A) holds then the dominant terms in the bracket in (1.53) are precisely those in the two displays just above. This yields

$$\mathfrak{n}(Z_{i+1}) = \mathfrak{n}\left(\frac{-Y_{i+1}}{s_i\sqrt{n-i-1}}\right)\left(\begin{array}{c} 1 + \frac{aY_{i+1}}{s_i^2(n-i-1)} - \frac{a^2}{2s_i^2(n-i-1)} \\ + \frac{o_{P_{na}}(\epsilon_n \log n)}{n-i-1} \end{array}\right).$$

We now need a precise evaluation of the terms in the Hermite polynomials in (1.52). This is achieved using Lemmas 24 and 25 which provide uniformity upon i between 1 and $k = k_n$ in all terms depending on the sample path Y_1^k. The Hermite polynomials depend upon the moments of the underlying density π^{m_i}. Since $\overline{\pi_1^{m_i}}$ has expectation 0 and variance 1 the terms corresponding to P_1 and P_2 vanish. Up to the order 4 the polynomials write $P_3(x) = \frac{\mu_3^{(i)}}{6(s_i)^3}H_3(x)$, $P_4(x) = \frac{(\mu_3^i)^2}{72(s_i)^6}H_6(x) + \frac{\mu_4^{(i,n)}-3(s_i)^4}{24(s_i)^4}H_4(x)$ with $H_3(x) := x^3 - 3x$, $H_4(x) := x^4 + 6x^2 - 3$ and $H_6(x) := x^6 - 15x^4 + 45x^2 - 15$.

Using Lemma 24 it appears that the terms in x^j, $j \geq 3$ in P_3 and P_4 will play no role in the asymptotic behavior in (1.52) with respect to the constant term in P_4 and the term in x from P_3 . Indeed substituting x by Z_{i+1} and dividing by $n-i-1$, the term in x^2 in P_4 writes $O_{P_{na}}(\log n)^2/(n-i)^2$ where we used Lemma 24. These terms are of smaller order than the term $-3x$ in P_3 which writes $-\frac{\mu_3^i}{2s_i^4(n-i-1)}(a - Y_{i+1}) = \frac{1}{n-i-1}O_{P_{na}}(\log n)$.

It holds

$$\begin{aligned}\frac{P_3(Z_{i+1})}{\sqrt{n-i-1}} &= -\frac{\mu_3^i}{2s_i^4\,(n-i-1)}\,(m_i - Y_{i+1}) \\ &\quad + \frac{\mu_3^i\,(m_i - Y_{i+1})^3}{6\,(s_i)^6\,(n-i-1)^2}\end{aligned}$$

which yields

$$\frac{P_3(Z_{i+1})}{\sqrt{n-i-1}} = -\frac{\mu_3^i}{2s_i^4\,(n-i-1)}\,(a - Y_{i+1}) + \frac{1}{(n-i-1)^2}O_{P_{na}}(\log n)^3\,. \tag{1.55}$$

For the term of order 4 it holds

$$\frac{P_4(Z_{i+1})}{n-i-1} = \frac{1}{n-i-1}\left(\frac{(\mu_3^i)^2}{72s_i^6}H_6(Z_{i+1}) + \frac{\mu_4^i - 3s_i^4}{24s_i^4}H_4(Z_{i+1})\right)$$

which yields

$$\frac{P_4(Z_{i+1})}{n-i-1} = -\frac{\mu_4^i - 3s_i^4}{8s_i^4\,(n-i-1)} - \frac{15(\mu_3^i)^2}{72s_i^6(n-i-1)} + \frac{O_{P_{na}}((\log n)^2)}{(n-i-1)^2}. \tag{1.56}$$

The fifth term in the expansion plays no role in the asymptotic.

To sum up and comparing the remainder terms in (1.55) and (1.56), we get

$$\overline{\pi_{n-i-1}}\left(Z_{i+1}\right)=\mathfrak{n}\left(-Y_{i+1}/\left(s_i\sqrt{n-i-1}\right)\right).A.B+O_{P_{na}}\left(\frac{P_5(Z_{i+1})}{(n-i-1)^{3/2}}\right)$$

where A and B are given in (1.15) and (1.16).

1.5.3 Final step of the proof of Theorem 1

We make use of the following version of the law of large numbers for triangular arrays (see [86] Theorem 3.1.3).

Theorem 26. *Let $X_{i,n}$,$1\leq i\leq k$ denote an array of row-wise real exchangeable r.v's and $\lim_{n\to\infty}k=\infty$. Let $\rho_n:=EX_{1,n}X_{2,n}$. Assume that for some finite Γ , $EX_{1,n}^2\leq\Gamma$. If for some doubly indexed sequence $(a_{i,n})$ such that $\lim_{n\to\infty}\sum_{i=1}^k a_{i,n}^2=0$ it holds*

$$\lim_{n\to\infty}\rho_n\left(\sum_{i=1}^k a_{i,n}^2\right)^2=0$$

then

$$\lim_{n\to\infty}\sum_{i=1}^k a_{i,n}X_{i,n}=0$$

in probability.

Denote

$$\kappa_1^i:=\frac{\mu_3^i}{2s_i^4}\quad,\quad\kappa_2^i:=\frac{\mu_3^i-s_i^4}{8s_i^4}+\frac{15(\mu_3^i)^2}{72s_i^6},$$
$$\mu_1^*:=\kappa_1^i+\frac{a}{s_i^2}\quad,\quad\mu_2^*:=\kappa_1^i-\frac{a}{2s_i^2}.$$

By (1.13), (1.14) and (1.17)

$$p(\mathbf{X}_{i+1}=Y_{i+1}|S_{i+1,n}=na-S_{1,i})=$$
$$\frac{\sqrt{n-i}}{\sqrt{n-i-1}}\pi^{m_i}\left(\mathbf{X}_{i+1}=Y_{i+1}\right)\frac{\mathfrak{n}\left(\frac{-Y_{i+1}}{s_i\sqrt{n-i-1}}\right)}{\mathfrak{n}(0)}A(i)$$

with

$$A(i):=\frac{1+\frac{\mu_1^*Y_{i+1}}{n-i-1}-\frac{\mu_2^*a}{n-i-1}-\frac{\kappa_2^i}{n-i-1}+\frac{o_{P_{na}}(\epsilon_n\log n)}{n-i-1}}{1-\frac{\kappa_2^i}{n-i}+O_{P_{na}}\left(\frac{1}{(n-i)^{3/2}}\right)}.$$

We perform a second order expansion in both the numerator and the denominator of the above expression, which yields

$$A(i)=\exp\left(\frac{\mu_1^*Y_{i+1}}{n-i-1}-\frac{a}{2s_i^2(n-i-1)}-\frac{a\kappa_1^i}{n-i-1}+\frac{o_{P_{na}}\left(\epsilon_n\log n\right)}{n-i-1}\right)A'(i).\qquad(1.57)$$

The term $\exp\left(\frac{\mu_1^*Y_{i+1}}{n-i-1}+\frac{a}{2s_i^2(n-i-1)}\right)$ in (1.57) is captured in $g(Y_{i+1}|\,Y_1^i)$.

The term $A'(i)$ in (1.57) writes

$$A'(i) := Q_1^i.Q_2^i$$

with

$$Q_1^i := \exp\left(-\left(\frac{\kappa_2^i}{(n-i-1)(n-i)} + \frac{(\kappa_2^i)^2}{2(n-i)^2} + \frac{1}{2}\left(\frac{\mu_1^* Y_{i+1}}{n-i-1} - \frac{a\mu_2^*}{n-i-1} - \frac{\kappa_2^i}{n-i-1}\right)^2\right)\right)$$

and

$$Q_2^i := \frac{\exp B_1}{\exp B_2}$$

where

$$B_1 := \frac{o_{P_{na}}(\epsilon_n^2(\log n)^2)}{(n-i-1)^2} + \frac{\mu_1^* Y_{i+1}}{(n-i-1)^2} o_{P_{na}}(\epsilon_n \log n)$$
$$+ \frac{\mu_2^* a}{(n-i-1)^2} o_{P_{na}}(\epsilon_n \log n) + \frac{o_{P_{na}}(\epsilon_n^2 (\log n))^2}{(n-i-1)^2} + o(u_1^2)$$

$$B_2 := \frac{\kappa_2^i}{n-i} O_{P_{na}}\left(\frac{1}{(n-i)^{3/2}}\right) + O_{P_{na}}\left(\frac{1}{(n-i)^3}\right) +$$

$$O_{P_{na}}\left(\frac{1}{(n-i)^{3/2}}\right) + o\left(\left(\frac{\kappa_2^i}{n-i} + O_{P_{na}}\left(\frac{1}{(n-i)^{3/2}}\right)\right)^2\right).$$

with

$$u_1 = \frac{\mu_1^* Y_{i+1}}{n-i-1} - \frac{\mu_2^* a}{n-i-1} - \frac{\kappa_2^i}{n-i-1} + \frac{o_{P_{na}}(\epsilon_n \log n)}{n-i-1}.$$

We first prove that

$$\prod_{i=0}^{k-1} A'(i) = 1 + o_{P_{na}}(\epsilon_n (\log n)^2) \tag{1.58}$$

as n tends to infinity.

Since

$$p(\mathbf{X}_1^k = Y_1^k | S_{i+1}^n = na) = g_0(Y_1 | Y_0) \prod_{i=0}^{k-1} g\left(Y_{i+1} | Y_1^i\right) \prod_{i=0}^{k-1} A'(i) \prod_{i=0}^{k-1} L_i$$

where

$$L_i := \frac{C_i^{-1}}{\Phi(t_i)} \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \exp\left(-\frac{a\kappa_1^i}{n-i-1}\right)$$

the completion of the proof will follow from

$$\prod_{i=0}^{k-1} L_i = 1 + o_{P_{na}}(\epsilon_n (\log n)^2). \tag{1.59}$$

The proof of (1.58) is achieved in two steps.

Claim 27. $\prod_{i=0}^{k-1} Q_1^i = 1 + o_{P_{na}}(\epsilon_n (\log n)^2)$.

By Lemma 24 the random terms μ_j^i deriving from π^{m_i} satisfy

$$max_{1\leq i\leq k}\left|\mu_j^i-\mu_j\right|=o_{P_{na}}(1)$$

as n tends to ∞, where μ_j is the j-th cumulants of $\pi^{\underline{a}}$ where $\underline{a}:=\lim_{n\to\infty}a$ is finite. Therefore we may substitute μ_j^i by μ_j in order to check the convergence of all subsequent series.

Developing $Q1$ define, for any positive β_1, β_2, β_3 and β_4

$$A_n^1:=\left\{\frac{1}{\epsilon_n\left(\log n\right)^2}\sum_{i=0}^{k-1}\left|\frac{\kappa_2^i}{(n-i-1)(n-i)}\right|<\beta_1\right\},$$

$$A_n^2:=\left\{\frac{1}{\epsilon_n\left(\log n\right)^2}\sum_{i=0}^{k-1}\left|\frac{(\kappa_2^i)^2}{(n-i-1)^2}\right|<\beta_2\right\},$$

$$A_n^3:=\left\{\frac{1}{\epsilon_n\left(\log n\right)^2}\sum_{i=0}^{k-1}\left|\frac{(\mu_2^*a)^2}{(n-i-1)^2}\right|<\beta_3\right\}$$

and

$$A_n^4:=\left\{\frac{1}{\epsilon_n\left(\log n\right)^2}\sum_{i=0}^{k-1}\left|\frac{\mu_2^*\kappa_2^ia}{(n-i-1)^2}\right|<\beta_4\right\}.$$

It clearly holds that

$$\lim_{n\to\infty}P_{na}\left(A_n^j\right)=1;\ j=1,...,4.$$

Let for any positive β_5

$$A_n^5:=\left\{\frac{1}{\epsilon_n\left(\log n\right)^2}\sum_{i=0}^{k-1}\left|\frac{\kappa_1^i\kappa_2^iY_{i+1}}{(n-i-1)^2}\right|<\beta_5\right\}.$$

If $\lim_{n\to\infty}P_{na}\left(A_n^5\right)=1$, then $\lim_{n\to\infty}P_{na}\left(A_n^j\right),\ j=6,7$ where

$$A_n^6:=\left\{\frac{1}{\epsilon_n\left(\log n\right)^2}\sum_{i=0}^{k-1}\left|\frac{\mu_1^*\kappa_2^iY_{i+1}}{(n-i-1)^2}\right|<\beta_6\right\}$$

$$A_n^7:=\left\{\frac{1}{\epsilon_n\left(\log n\right)^2}\sum_{i=0}^{k-1}\left|\frac{\mu_1^*\mu_2^*aY_{i+1}}{(n-i-1)^2}\right|<\beta_7\right\}.$$

Apply Theorem 26 with $X_{i,n}=Y_{i+1}$ and $a_{i,n}=\frac{1}{\epsilon_n(\log n)^2(n-i-1)^2}$. By Lemma 23

$$E_{P_{na}}Y_1^2=s^2(0)+a+O\left(\frac{1}{n}\right).$$

Hence $E_{P_{na}}[Y_1^2]\leq\Gamma$ for some finite Γ. Further $\rho_n=a^2+O\left(\frac{1}{n}\right)$. Both conditions in Theorem 26 are fullfilled. Indeed

$$\lim_{n\to\infty}\sum_{i=1}^{k}a_{n,i}^2=\lim_{n\to\infty}\frac{1}{\epsilon_n^2\left(\log n\right)^4(n-k)^3}=0$$

which holds under (E1), as holds

$$\lim_{n\to\infty}\rho_n\left(\sum_{i=1}^{k}a_{n,i}\right)^2=\lim_{n\to\infty}\frac{a^2}{\epsilon_n^2\left(\log n\right)^4(n-k)^2}=0.$$

Therefore, for $i = 5, 6, 7$

$$\lim_{n\to\infty} P_{na}\left(A_n^i\right) = 1.$$

Define for any positive β_8

$$A_n^8 := \left\{ \frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \frac{(\mu_1^*)^2 Y_{i+1}^2}{(n-i-1)^2} < \beta_8 \right\}.$$

Apply Theorem 26 with $X_{i,n} = Y_{i+1}^2$ and $a_{i,n} = \frac{1}{\epsilon_n (\log n)^2 (n-i-1)^2}$.
It holds

$$\lim_{n\to\infty} \sum_{i=1}^{k} a_{n,i}^2 = 0$$

when (E1) holds.

By Lemma 23,

$$E_{P_{na}} Y_1^4 = E_{\pi^a} Y_1^4 + O\left(\frac{1}{n}\right)$$

which entails that such that $EY_1^4 \leq \Gamma < \infty$ for some Γ. Also

$$E_{P_{na}}\left(Y_1^2 Y_2^2\right) = \left(s^2(0) + a\right)\left(s^2(0) + a\right) + O\left(\frac{1}{n}\right)$$

and

$$\lim_{n\to\infty} \rho_n \left(\frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \frac{1}{(n-i-1)^2} \right)^2 = 0$$

under (E1). Hence

$$\lim_{n\to\infty} P_{na}\left(A_n^8\right) = 1.$$

It follows that, noting A_n the intersection of the events A_n^i , $j = 1, ..., 8$

$$\lim_{n\to\infty} P_{na}\left(A_n\right) = 1.$$

To sum up, we have proved that, under (E1),

$$Q1 = 1 + o_{P_{na}}\left(\epsilon_n (\log n)^2\right).$$

Claim 28. $\prod_{i=0}^{k-1} Q_2^i = 1 + o_{P_{na}}\left(\epsilon_n (\log n)^2\right)$.

This amounts to prove that the sum of the terms in $B1$ (resp in $B2$) is of order $o_{P_{na}}\left(\epsilon_n (\log n)^2\right)$.

The four terms in the the sum of the terms in $B1$ are respectively of order $o_{P_{na}}\left(\epsilon_n^2 (\log n)^4\right)/(n-k)$, $o_{P_{na}}\left(\epsilon_n (\log n)^3\right)/(n-k)$, $o_{P_{na}}\left(a\epsilon_n (\log n)^2\right)/(n-k)$ and $o_{P_{na}}\left(\epsilon_n (\log n)^2\right)/(n-k)$ using Lemma 24. The sum of the terms $o\left(u_1^2\right)$ is of order less than those ones. Assuming (E1) all those terms are $o_{P_{na}}\left(\epsilon_n (\log n)^2\right)$.

For the sum of terms $B2$, by uniformity of the Edgeworth expansion with respect to Y_1^k it holds $\sum_{i=1}^{k} B2 = O_{P_{na}}\left((n-k)^{-1/2}\right) = o_{P_{na}}\left(\epsilon_n (\log n)^2\right)$ by (E1).

We now turn to the proof of (1.59)

Define

$$u := -x \frac{\mu_3^i}{2s_i^4 (n-i-1)} + \frac{(x-a)^2}{2s_i^2 (n-i-1)}.$$

Use the classical bounds

$$1-u+\frac{u^2}{2}-\frac{u^3}{6}\leq e^{-u}\leq 1-u+\frac{u^2}{2}$$

to obtain on both sides of the above inequalities the second order approximation of C_i^{-1} through integration with respect to p. The upper bound yields

$$C_i^{-1}\leq \Phi(t_i)+\frac{\kappa_1^i}{n-i-1}\Phi'(t_i)+\frac{1}{s_i^2(n-i-1)}\left(\Phi"(t_i)-2a\Phi'\left(t_i\right)+a^2\right)+O_{P_{na}}\left(\frac{1}{(n-i-1)^2}\right)$$

from which

$$L_i\leq \frac{\sqrt{n-i}}{\sqrt{n-i-1}}\exp\left(-\frac{a\kappa_1^i}{n-i-1}\right)\begin{pmatrix}1+\frac{\kappa_1^i}{n-i-1}m_i\\ -\frac{s_i^2+m_i^2-2am_i+a^2}{2s_i^2(n-i-1)}+O_{P_{na}}\left(\frac{1}{(n-i-1)^2}\right)\end{pmatrix}$$

where the approximation term is uniform on the Y_1^k.

Substituting $\frac{\sqrt{n-i}}{\sqrt{n-i-1}}$ and $\exp\left(-\frac{a\kappa_1^i}{n-i-1}\right)$ by their expansion $1+\frac{1}{2(n-i-1)}+O\left(\frac{1}{(n-i-1)^2}\right)$ and $1-\frac{a\kappa_1^i}{n-i-1}+\frac{(a\kappa_1^i)^2}{(n-i-1)^2}+O\left(\frac{a^2}{(n-i-1)^2}\right)$ in the upper bound of L_i above yields

$$L_i\leq \left(1+\frac{1}{2(n-i-1)}-\frac{a\kappa_1^i}{n-i-1}+\frac{(a\kappa_1^i)^2}{2(n-i-1)^2}+o\left(\frac{1}{(n-i-1)^2}\right)\right)\left(1+\frac{\kappa_1^i m_i}{n-i-1}-\frac{s_i^2+m_i^2-2am_i+a^2}{2s_i^2(n-i-1)}+O_{P_{na}}\left(\frac{1}{(n-i-1)^2}\right)\right).$$

Using Lemma 24, $m_i^2-2am_i+a^2=o_{P_{na}}(a\epsilon_n)$ and therefore

$$L_i\leq \left(1+\frac{1}{2(n-i-1)}-\frac{a\kappa_1^i}{n-i-1}+\frac{(a\kappa_1^i)^2}{(n-i-1)^2}+o\left(\frac{1}{(n-i-1)^2}\right)\right)\left(1+\frac{\kappa_1^i a}{n-i-1}-\frac{1}{2(n-i-1)}+\frac{o_{P_{na}}(a\epsilon_n)}{n-i-1}\right).$$

Write

$$\prod_{i=1}^{k}L_i\leq\prod_{i=1}^{k}(1+M_i)$$

with

$$M_i=\frac{(a\kappa_1^i)^2}{(n-i-1)^2}+\frac{o_{P_{na}}(a\epsilon_n)}{n-i-1}.$$

Under (E1), $\sum_{i=0}^{k-1}M_i$ is $o_{P_{na}}\left(\epsilon_n\left(\log n\right)^2\right)$. This closes the proof of the Theorem.

1.5.4 Proof of Theorem 19

The following lemma provide asymptotic formula for the tail probability of $\mathbf{U}_{1,n}$ under the hypothesis and notations of section 3. Define

$$I_{\mathbf{U}}(x) := xm^{-1}(x) - \log \phi_{\mathbf{U}}\left(m^{-1}(x)\right)$$

Lemma 29. *(see [59], Corollary 6.4.1) Under the same hypotheses and notations as section 3,*

$$P\left(\frac{\mathbf{U}_{1,n}}{n} > a\right) = \frac{\exp -nI_{\mathbf{U}}(a)}{\sqrt{2\pi}\sqrt{n}\psi(a)}\left(1 + O(\frac{1}{\sqrt{n}})\right)$$

where $\psi(a) := m^{-1}(a)s(m^{-1}(a))$.

Two Lemmas pertaining to the partial sum under its final value

Lemma 30. *Suppose that (V) holds. Then (i)*$E_{P_{nA}}\mathbf{U}_1 = a + o(1)$, *(ii)* $E_{P_{nA}}\mathbf{U}_1^2 = a^2 + s^2\left(m^{-1}(a)\right) + o(1)$ *and (iii)* $E_{P_{nA}}\mathbf{U}_1\mathbf{U}_2 = a^2 + o(1)$.

Proof. We make use of Lemma 23, meaning $E_{P_{nv}}[\mathbf{U}_1] = v$. It holds

$$E_{P_{nA}}\mathbf{U}_1 = \int_a^\infty \left(E_{P_{nv}}\mathbf{U}_1\right) p\left(\mathbf{U}_{1,n}/n = v | \, \mathbf{U}_{1,n} > na\right) dv.$$

Integration by parts yields,

$$E_{P_{nA}}\mathbf{U}_1 = a + \int_a^\infty P\left(\mathbf{U}_{1,n}/n > v | \, \mathbf{U}_{1,n} > na\right) dv.$$

Using Lemma 29 and Chernoff inequality,

$$\int_a^\infty P\left(\mathbf{U}_{1,n}/n > v | \, \mathbf{U}_{1,n} > na\right) dv \leq \sqrt{2\pi}\psi(a)\sqrt{n}\int_a^\infty \exp[n\left(I_{\mathbf{U}}(a) - I_{\mathbf{U}}(v)\right)]$$

where $\psi(a)$ is defined in Lemma 29.

Finally, using $I_{\mathbf{U}}(v) > I'_{\mathbf{U}}(a)v + I_{\mathbf{U}}(a) - aI'_{\mathbf{U}}(a)$, and integrating

$$\int_a^\infty P\left(\mathbf{U}_{1,n}/n > v | \, \mathbf{U}_{1,n} > na\right) dv \leq \frac{\sqrt{2\pi}s(m^{-1}(a))}{\sqrt{n}}.$$

Hence, $E_{P_{nA}}\mathbf{U}_1 = a + o(1)$.

Insert $E_{P_{nv}}\mathbf{U}_1^2 = v^2 + s^2\left(m^{-1}(a)\right) + O\left(\frac{1}{n}\right)$ in

$$E_{p_{nA}}\mathbf{U}_1^2 = \int_a^\infty E_{P_{nv}}\mathbf{U}_1^2 p\left(\mathbf{U}_{1,n}/n = v | \, \mathbf{U}_{1,n} > na\right) dv.$$

Firstly, by integration by parts, Lemma 29 and Chernoff inequality,

$$\int_a^\infty v^2 p\left(\mathbf{U}_{1,n}/n = v | \, \mathbf{U}_{1,n} > na\right) dv = a^2 + o(1)$$

Indeed, since (C) implies $nm^{-1}(a) \to \infty$ when n tends to ∞, it holds

$$\int_a^\infty vp\left(\mathbf{U}_{1,n}/n > v | \, \mathbf{U}_{1,n} > na\right) dv \leq \frac{s(m^{-1}(a))}{\sqrt{n}}\left(a + \frac{1}{nm^{-1}(a)}\right).$$

Secondly,

$$\int_a^\infty V(v)p\left(\mathbf{U}_{1,n}/n = v|\,\mathbf{U}_{1,n} > na\right)dv = s^2(m^{-1}(a)) + 2\int_a^\infty V'(v)P\left(\mathbf{U}_{1,n}/n > v|\,\mathbf{U}_{1,n} > na\right)dv.$$

Using Lemma 29, Chernoff inequality and $I_{\mathbf{U}}(v) > I'_{\mathbf{U}}(a)v + I_{\mathbf{U}}(a) - aI'_{\mathbf{U}}(a)$, it holds under condition (V),

$$\begin{aligned}&\int_a^\infty V'(v)P\left(\mathbf{U}_{1,n}/n > v|\,\mathbf{U}_{1,n} > na\right)dv\\ \leq\ & s(m^{-1}(a))\left(\sqrt{n}m^{-1}(a)\int_a^\infty V'(v)\exp\left(-nm^{-1}(a)(v-a)\right)dv\right)\end{aligned}$$

and

$$\int_a^\infty V(v)p\left(\mathbf{U}_{1,n}/n = v|\,\mathbf{U}_{1,n} > na\right)dv = s^2(m^{-1}(a)) + o(1).$$

The third term is handled similarly due to the fact that the $O(1/n)$ consists in a sum of powers of v.

For $E_{P_{nA}}\mathbf{U}_1\mathbf{U}_2 = a^2 + o(1)$, the proof is similar. □

Lemma 30 yields the maximal inequality stated in Lemma 24 under the condition $(\mathbf{U}_{1,n} > na)$. We also need the order of magnitude of the maximum of $(|\mathbf{U}_1|, ..., |\mathbf{U}_k|)$ under P_{nA} which is stated in the following result.

Lemma 31. *It holds for all k between 1 and n*

$$\max\left(|\mathbf{U}_1|, ..., |\mathbf{U}_k|\right) = O_{P_{nA}}(\log n).$$

Proof. Using the same argument as in Lemma 25, we consider the case when the r.v's $\mathbf{U}_i$ take non negative values. We prove that

$$\lim_{n\to\infty} P_{nA}\left(\max\left(\mathbf{U}_1, ..., \mathbf{U}_k\right) > t_n\right) = 0$$

when

$$\lim_{n\to\infty}\frac{t_n}{\log n} = \infty.$$

It holds

$$\begin{aligned}P_{nA}\left(\max\left(\mathbf{U}_1, ..., \mathbf{U}_k\right) > t_n\right) &= \int_a^{a+c} P_{nv}\left(\max\left(\mathbf{U}_1, ..., \mathbf{U}_k\right) > t_n|\,\mathbf{U}_{1,n}/n = v\right)\\ &\quad p\left(\mathbf{U}_{1,n}/n = v|\,\mathbf{U}_{1,n} > na\right)dv\\ &\quad + \int_{a+c}^\infty P_{nv}\left(\max\left(\mathbf{U}_1, ..., \mathbf{U}_k\right) > t_n|\,\mathbf{U}_{1,n}/n = v\right)\\ &\quad p\left(\mathbf{U}_{1,n}/n = v|\,\mathbf{U}_{1,n} > na\right)dv\\ &= : I + II.\end{aligned}$$

Now, using the same arguments as before,

$$II \leq \frac{P\left(\mathbf{U}_{1,n}/n > a + c\right)}{P\left(\mathbf{U}_{1,n}/n > a\right)} \leq \frac{m^{-1}(a)s(m^{-1}(a))}{m^{-1}(a+c)s(m^{-1}(a+c))}\exp\left(-ncm^{-1}(a)\right)$$

Since c is fixed and $m^{-1}(a)$ is bounded , $II \to 0$ under (C).

Furthermore by Lemma 25, $\lim_{n\to\infty} P\left(\max\left(\mathbf{U}_1, ..., \mathbf{U}_n\right) > t_n|\,\mathbf{U}_{1,n}/n = v\right) =: \lim_{n\to\infty} r_n = 0$ when $v \in (a, a + c)$. Hence

$$I \leq r_n(1 + o(1)) \to 0.$$

This proves the Lemma. □

We now prove Theorem 19(i).

*Step 1.*We first prove that the integral (1.39) can be reduced to its principal part, namely that

$$p_{nA}(Y_1^k) = (1 + o_{P_{nA}}(1))$$

$$\int_a^{a+c} p(\mathbf{X}_1^k = Y_1^k \Big| \mathbf{U}_{1,n}/n = v) p(\mathbf{U}_{1,n}/n = v | \mathbf{U}_{1,n} > na) dv \tag{1.60}$$

holds for any fixed $c > 0$.

Apply Bayes formula to obtain

$$p_{nA}(Y_1^k) = \frac{np_{\mathbf{X}}\left(Y_1^k\right)}{(n-k)}$$

$$\frac{\int_a^\infty p\left(\frac{\mathbf{U}_{k+1,n}}{n-k} = \frac{n}{n-k}\left(t - \frac{k\overline{U_{1,k}}}{n}\right)\right) dt}{P\left(\mathbf{U}_{1,n} > na\right)}$$

where $\overline{U_{1,k}} := \frac{U_{1,k}}{k}$.

Denote

$$I := \frac{P\left(\frac{\mathbf{U}_{k+1,n}}{n-k} > m_k + \frac{nc}{n-k}\right)}{P\left(\frac{\mathbf{U}_{k+1,n}}{n-k} > m_k\right)}.$$

with

$$m_k = \frac{n}{n-k}\left(a - \frac{k\overline{U_{1,k}}}{n}\right).$$

Then (1.60) holds whenever $I \to 0$ (under P_{nA}).

Under P_{nA} it holds

$$\overline{U_{1,n}} = a + O_{P_{nA}}\left(\frac{1}{nm^{-1}(a)}\right).$$

A similar result as Lemma 24 holds under condition $(\mathbf{U}_{1,n} > na)$, using Lemma 30; namely it holds

$$\max_{0 \le i \le k-1} \left|\overline{U_{i+1,n}}\right| = a + o_{P_{nA}}(\epsilon_n).$$

Using both results, it holds

$$m_k = a + O_{P_{nA}}(v_n) \tag{1.61}$$

with $v_n = \max\left(\epsilon_n, \frac{1}{(n-k)m^{-1}(a)}\right)$ which tends to 0 under (C).

We now prove that $I \to 0$. Using once more Lemma 29 yields

$$I \le \frac{m^{-1}(m_k)s(m^{-1}(m_k))}{m^{-1}(m_k + \frac{nc}{n-k})s(m^{-1}(m_k) + \frac{nc}{n-k})}$$

$$\exp\left(-(n-k)\left(I_{\mathbf{U}}\left(m_k + \frac{nc}{n-k}\right) - I_{\mathbf{U}}(m_k)\right)\right).$$

Now by convexity of the function $I_{\mathbf{U}}$, and (1.61),

$$\begin{aligned} & \exp - (n-k)\left(I_{\mathbf{U}}\left(m_k + \frac{nc}{n-k}\right) - I_{\mathbf{U}}(m_k)\right) \\ \le\ & \exp - ncm^{-1}(m_k) = \exp - nc\left[m^{-1}(a) + \frac{1}{V(a + \theta O_{P_{nA}}(v_n))} O_{P_{nA}}(v_n)\right] \end{aligned}$$

for some θ in $(0,1)$. which tends to 0 under P_{nA} when (A) and (C) hold. By monotonicity of $t \to m(t)$ and condition (C) the ratio in I is bounded.

We have proved that

$$I = O_{P_{nA}}\left(\exp -ncm^{-1}(a)\right).$$

Step 2. Theorem (19)(i) holds uniformly in v in $(a, a+c)$ where Y_1^k is generated under P_{nA}. This result follows from a similar argument as used in Theorem 3 where (1.48) is proved under the local sampling P_{nv}. A close look at the proof shows that (1.48) holds whenever Lemmas 24 and 25 stated for the variables $\mathbf{U}_i$'s instead of $\mathbf{X}_i$'s hold under P_{nA}. Those lemmas are substituted by Lemmas 30 and 31 here above.

Inserting (1.48) in (1.60) yields

$$p_{nA}(Y_1^k) = \left(\int_a^{a+c} g_{nv}(Y_1^k) p(\,\mathbf{U}_{1,n}/n = v|\,\mathbf{U}_{1,n} > na) dv\right) \left(1 + o_{p_{nA}}\left(\max\left(\epsilon_n \left(\log n\right)^2, \left(\exp\left(-ncm^{-1}(a)\right)\right)^{\delta}\right)\right)\right)$$

for any positive $\delta < 1$.

The conditional density of $\mathbf{U}_{1,n}/n$ given $(\mathbf{U}_{1,n} > na)$ is given in (2.7) which holds uniformly in v on $(a, a+c)$.

Summing up we have proved

$$p_{nA}(Y_1^k) =$$

$$\left(nm^{-1}(a) \int_a^{a+c} g_{nv}(Y_1^k) \left(\exp -nm^{-1}(a)(v-a)\right) dv\right) \left(1 + o_{p_{nA}}\left(\max\left(\epsilon_n \left(\log n\right)^2, \left(\exp\left(-ncm^{-1}(a)\right)\right)^{\delta}\right)\right)\right)$$

as $n \to \infty$ for any positive δ.

In order to get the approximation of p_{nA} by the density g_{nA} it is enough to observe that

$$nm^{-1}(a) \int_a^{a+c} g_{nv}(Y_1^k) \left(\exp -nm^{-1}(a)(v-a)\right) dv = 1 + o_{P_{nA}}\left(\exp -ncm^{-1}(a)\right)$$

as $n \to \infty$ which completes the proof of (1.48). The proof of (1.49) follows from (1.48) and Lemma 7 cited hereunder.

Chapter 2

Towards zero variance estimators for rare events probabilities

2.1 Introduction

Importance Sampling procedures aim at reducing the calculation time which is necessary in order to evaluate integrals, often in large dimension. We consider the case when the integral to be numerically computed is the probability of an event defined by a large number of random components; this case has received quite a lot of attention, above all when the event is of *small* probability. The present paper proposes estimators for both large and moderate deviation probabilities.

The classical IS scheme for rare event estimation rests on two basic ingredients: the sampling distribution is fitted to the so-called dominating point of the set to be measured; independent and identically distributed replications under this sampling distribution are performed. Our concern is related to the performance of these methods for an ill-behaved set, when this set is non connected, and bears some asymmetry with respect to the underlying probability model. This case leads to overwhelming loss in relative accuracy for the probability to be estimated. Our proposal hints on two choices: first do not make use of the notion of dominating point and explore all the target set instead; secondly, do not use i.i.d. replications, but merely sample long runs of variables under a proxy of the optimal sampling scheme. It will be seen in the example in Section 2.6 that a net improvement in the relative error stems from those two factors. This has been the main motivation for this paper; note also that accurate evaluation for moderate deviation probabilities is important in statistics, when handling power functions; see Chapter 3 where the present technique has been used in the context of Monte Carlo tests.

The situation which is considered is the following.

The r.v's $\mathbf{X}, \mathbf{X}_i's$ are i.i.d. with known common density $p_{\mathbf{X}}$ on $\mathbb{R}$, and u is a real valued measurable function defined on $\mathbb{R}$. Define $\mathbf{U} := u(\mathbf{X})$ with density $p_{\mathbf{U}}$ and

$$\mathbf{U}_{1,n} := \sum_{i=1}^{n} \mathbf{U}_i.$$

We intend to estimate

$$P_n := P\left(\mathbf{U}_{1,n} \in nA\right)$$

for large but fixed n where

$$A := (a_n, \infty) \tag{2.1}$$

and a_n is a convergent sequence. The limit of this sequence either equals $E[\mathbf{U}]$ or is assumed to be larger than $E[\mathbf{U}]$. In the first case it will be assumed that a_n converges slowly in such a way that $P(\mathbf{U}_{1,n} \in nA)$ is not obtainable through the central limit theorem; we may call this case a moderate deviation case. The second situation is classically referred to as a large deviation case.

The case when A is a union of intervals can be deduced from the present one; see the example in Section 2.6.

The basic estimate of P_n is defined as follows: generate L i.i.d. samples $X_1^n(l) := (X_1(l), .., X_n(l))$ with underlying density $p_{\mathbf{X}}$ and define

$$P^{(n)}(A) := \frac{1}{L}\sum_{l=1}^{L} \mathbb{1}_{\mathcal{E}_n}\left(X_1^n(l)\right)$$

where

$$\mathcal{E}_n := \{(x_1, ..., x_n) \in \mathbb{R}^n : (u(x_1) + ... + u(x_n)) \in nA\} \tag{2.2}$$

with $u_i := u(x_i)$. The Importance Sampling estimator of P_n with sampling density g on $\mathbb{R}^n$ is

$$P_g^{(n)}(A) := \frac{1}{L}\sum_{l=1}^{L} P_n(l)\mathbb{1}_{\mathcal{E}_n}\left(Y_1^n(l)\right) \tag{2.3}$$

where $P_n(l)$ is called "importance factor" and can be written

$$P_n(l) := \frac{\prod_{i=1}^{n} p_{\mathbf{X}}(Y_i(l))}{g\left(Y_1^n(l)\right)} \tag{2.4}$$

and where the L samples $Y_1^n(l) := (Y_1(l), ..., Y_n(l))$ are i.i.d. with common density g; the coordinates of $Y_1^n(l)$ however need not be i.i.d.

The *classical* IS scheme consists in the simulation of n i.i.d. replications $Y_1^{(l)}, ..., Y_n^{(l)}$ with density π^{a_n} on $\mathbb{R}$ and therefore $g(y_1, ..., y_n) = \pi^{a_n}(y_1)...\pi^{a_n}(y_n)$. The density π^{a_n} is the so-called *tilted* (or *twisted*) density at point a_n which, in case when $a_n = a$ is fixed, is called the *dominating point* of the set (a, ∞); see [20]. In spite of the fact that this terminology is usually used in the large deviation case, we adopt it also in the moderate deviation one, for reasons to be stated later on. This approach produces efficient IS schemes since it can be proved that in the large deviation range the variance of the classical IS is proportional to $P_n^2\sqrt{n}$. Adaptive IS schemes have been considered in various contexts; for first passage times, see [37]; splitting techniques have also been considered in this field; see [16] and [28], where connections between IS and particle splitting is considered; similar problems as developped in this paper for heavy tailed distribution are presented in [36]. An overview on various topics in IS is presented in [83].

The numerator in the expression (2.4) is the product of the $p_{\mathbf{X}_1}(Y_i)$'s while the denominator need not be a density of i.i.d. copies evaluated on the $Y_i's$. Indeed the optimal choice for g is the density of $\mathbf{X}_1^n := (\mathbf{X}_1, ..., \mathbf{X}_n)$ conditioned upon $(\mathbf{X}_1^n \in \mathcal{E}_n)$, leading to a zero variance estimator. We will propose an IS sampling density which approximates this conditional density very sharply on its first components $y_1, ..., y_k$ where $k = k_n$ is very large, namely $k/n \to 1$. This motivates the title of this paper; however, but in the gaussian case, k should satisfy $\lim_{n\to\infty} n - k = \infty$ by the very construction of the approximation.

Let us introduce a toy case in order to define the main step of the procedure, namely the simulation of a sample under a proxy of the conditional density. Assume $\mathbf{X}_1^n$ is a

vector of n i.i.d. standard normal real valued random variables and $P_n := P(\mathbf{S}_{1,n} > na)$ with $\mathbf{S}_{1,n} := \mathbf{X}_1 + ... + \mathbf{X}_n$ and $a > 0$.

1- For any $v > a$ the joint density p_{nv} of $\mathbf{X}_1, ..., \mathbf{X}_{n-1}$ conditionally upon $(\mathbf{S}_{1,n} = nv)$ is known analytically and simulation under p_{nv} is easy for any v. A general form of this statement is Theorem 9 of Chapter 1.

2-The optimal sampling density g is similar to p_{nv} with conditioning event $(\mathbf{S}_{1,n} > na)$. The density g is obtained by integrating p_{nv} with respect to the the conditional distribution of $\mathbf{S}_{1,n}/n$ under $(\mathbf{S}_{1,n} > na)$ which is well approximated by an exponential distribution on (a, ∞) with expectation $a + (1/na)$; this follows from standard asymptotics for large and moderate deviations. The corresponding general statement is Theorem 19 of Chapter 1. Therefore samples under a proxy of g are obtained through Monte Carlo simulation as follows: draw Y_1^n with density $p_{n\mathbf{V}}$ where $\mathbf{V}$ follows the just cited exponential density. Insert these terms in (2.4) repeatedly to get $P_g^{(n)}$.

In the general case the joint distribution p_{nv} cannot be approximated sharply on the very long run $1, ..., n-1$, but merely on $1, ..., k_n$ with k_n close to n. The approximation provided in Theorem 9 and, as a consequence in Theorem 19, is valid on the first k_n coordinates; the IS density on $\mathbb{R}^n$ is obtained multiplying this proxy by a product of $n - k_n$ identical densities π^{a_n}. It will be shown in Section 2.3 that the *relative accuracy* drops by a factor $\sqrt{n-k}/\sqrt{n}$ with respect to the classical IS scheme. A precise tuning of k_n is provided in Section 1.4.3, extending the rule of Section 1.3.2 of Chapter 1. Algorithms are in Section 2.4. Section 2.5 contains a simulation study in two simple cases. Since v is simulated on the whole set $(a, +\infty)$, no search is done in order to identify dominating points and no part of the target set $(a, +\infty)$ is neglected in the simulation of runs; the example in section 2.6, where the classical IS scheme is compared to the present one, is illuminating in this respect. It also contains graphs pertaining to the hit rate. Other performance indices have been considered, which focus on the stability of the estimate; see [13].

It may seem that we could have reduced this paper to the case when u is the identity function, hence simulating runs $\mathbf{U}_1^k := (u(\mathbf{X}_1), ..., u(\mathbf{X}_k))$ under $(\mathbf{U}_{1,n} > na)$. However it often occurs that the conditioning event is defined through a joint set of conditions, say

$$u(\mathbf{X}_1) + ... + u(\mathbf{X}_n) > na \tag{2.5}$$

and

$$h(\mathbf{X}_1^n) \in B_n \tag{2.6}$$

for some function h and some measurable set B_n. Clearly in most cases the approximation of the density of $\mathbf{X}_1^k$ under both constraints is intractable and the approximation of the density of $\mathbf{X}_1^k$ conditioned upon $(\mathbf{X}_1^n \in \mathcal{E}_n)$ and extended on $\mathbb{R}^n$ as indicated above, provides a good IS sampling scheme for the estimation of

$$P(u(\mathbf{X}_1) + ... + u(\mathbf{X}_n) > na \cap h(\mathbf{X}_1^n) \in B_n).$$

We rely on Chapter 1 where the basic approximation can be found. A first draft in the direction of the present work is in [18].

2.2 Adaptive IS Estimator for rare event probability

The IS scheme produces samples $Y := (Y_1, ..., Y_k)$ distributed under g_{nA}, which is a continuous mixture of densities g_{nv}, with exponential mixing measure with parameter

$nm^{-1}(a)$ on (a,∞)

$$\mathbb{1}_{(a,\infty)}(v)nm^{-1}(a)\exp\left(-nm^{-1}(a)(v-a)\right) \tag{2.7}$$

Since all IS schemes produce unbiased estimators, and since the truncation parameter c in (1.46) is immaterial, we consider untruncated versions of g_{nA} defined in (1.46) integrating on (a,∞) instead of $(a,a+c)$. This avoids a number of computational and programming questions, a difficult choice of an extra parameter c, and does not change the numerical results; this point has been checked carefully by the authors. We keep the notation g_{nA} for the untruncated density.

The density g_{nA} is extended from $\mathbb{R}^k$ onto $\mathbb{R}^n$ completing the $n-k$ remaining coordinates with i.i.d. copies of r.v's $Y_{k+1},...,Y_n$ with common tilted density

$$g_{nA}\left(y_{k+1}^n|\,y_1^k\right) := \prod_{i=k+1}^{n}\pi_u^{m_k}(y_i) \tag{2.8}$$

with $m_k := m(t^k) = \frac{n}{n-k}\left(v-\frac{u_{1,k}}{n}\right)$ and

$$u_{1,k} = \sum_{i=1}^{k} u(y_i)$$

The last $n-k$ r.v's $\mathbf{Y}_i$'s are therefore drawn according to the classical i.i.d. scheme in phase with [87] or [45] schemes in the large or moderate deviation setting.

We now define our IS estimator of $P_n := P(\mathbf{U}_{1,n} > na)$.

Let $Y_1^n(l) := Y_1(l),...,Y_n(l)$ be generated under g_{nA}. Let

$$\widehat{P_n}(l) := \frac{\prod_{i=0}^{n} p_{\mathbf{X}}(Y_i(l))}{g_{nA}(Y_1^n(l))}\mathbb{1}_{\mathcal{E}_n}(Y_1^n(l)) \tag{2.9}$$

and define

$$\widehat{P_n} := \frac{1}{L}\sum_{l=1}^{L}\widehat{P_n}(l). \tag{2.10}$$

2.3 Compared efficiencies of IS estimators

The situation which we face with our proposal lacks the possibility to provide an order of magnitude of the variance our our IS estimate, since the properties necessary to define it have been obtained only on *typical paths* under the sampling density g_{nA} and not on the whole space $\mathbb{R}^n$. This leads to a quasi-MSE measure for the performance of the proposed estimator, which quantifies the variability evaluated on classes of subsets of $\mathbb{R}^n$ whose probability goes to 1 under the sampling g_{nA}. Not surprisingly the loss of performance with respect to the optimal sampling density is due to the $n-k$ last i.i.d. simulations, leading a quasi- MSE of the estimate proportional to $\sqrt{n-k}$.

2.3.1 The efficiency of the classical IS scheme

We first recall the definition of the classical IS sampling scheme and its asymptotic performance. The r.v.'s Y_i's in (2.4) are i.i.d. and have density $g = \pi_u^a$, hence with $m(t) = a$. See [87] in the LDP case and [45] in the MDP case. The reason for this

sampling scheme is the fact that in the large deviation case, a is the "dominating point" of the set (a,∞) i.e. a is such that the proxy of the conditional distribution of $\mathbf{X}_1$ given $(\mathbf{U}_{1,n} > na)$ is Π_u^a; this is the basic form of the Gibbs conditioning principle.

Although developed for the large deviation case, the classical IS applies for the moderate deviation case since for $a \to E[u(\mathbf{X})]$ and $(a - E[u(\mathbf{X})])\sqrt{n} \to \infty$ it holds

$$P(\mathbf{X}_1 \in B | \mathbf{U}_{1,n} > na) = (1 + o(1))\,\Pi_u^a(B) \tag{2.11}$$

for any Borel set B as $n \to \infty$. This follows as a consequence of Sanov Theorem for moderate deviations (see [45] and [27]) and justifies the classical IS scheme in this range.

The classical IS is defined simulating L times a random sample of n i.i.d. r.v's $Y_1^n(l)$, $1 \le l \le L$, with tilted density π_u^a. The standard IS estimate is defined through

$$\overline{P_n} := \frac{1}{L}\sum_{l=1}^{L} \mathbb{1}_{\mathcal{E}_n}(Y_1^n(l)) \frac{\prod_{i=1}^n p_{\mathbf{X}}(Y_i(l))}{\prod_{i=1}^n \pi_u^a(Y_i(l))}$$

where the $X_i(l)$ are i.i.d. with density π_u^a and $\mathbb{1}_{\mathcal{E}_n}(Y_1^n(l))$ is as in (2.2). Set

$$\overline{P_n}(l) := \mathbb{1}_{\mathcal{E}_n}(Y_1^n(l)) \frac{\prod_{i=1}^n p_{\mathbf{X}}(Y_i(l))}{\prod_{i=1}^n \pi_u^a(Y_i(l))}.$$

The variance of $\overline{P_n}$ is given by

$$Var\overline{P_n} = \frac{1}{L}\left(E_{\Pi_u^a}\left[\overline{P_n}(l)^2\right] - P_n^2\right).$$

The *relative accuracy* of the estimate $\overline{P_n}$ is defined through

$$RE(\overline{P_n}) := \frac{Var\overline{P_n}}{P_n^2} = \frac{1}{L}\left(\frac{E_{\Pi_u^a}\left[\overline{P_n}(l)^2\right]}{P_n^2} - 1\right).$$

The following result holds.

Proposition 32. *The relative accuracy of the estimate $\overline{P_n}$ is given by*

$$RE(\overline{P_n}) = \frac{\sqrt{2\pi}\sqrt{n}}{L} a(1 + o(1))$$

as n tends to infinity.

We will prove that no reduction of the variance of the estimator can be achieved on subsets B_n of $\mathbb{R}^n$ such that $\Pi_u^a(B_n) \to 1$.

The easy case when $\mathbf{U}_1, ..., \mathbf{U}_n$ are i.i.d. with standard normal distribution and $u(x) = x$ is sufficient for our need.

The variance of the IS estimate of $P(\mathbf{U}_{1,n} > na)$ is proportional to

$$\begin{aligned} V &:= E_{P_{\mathbf{U}}}\left[\mathbb{1}_{(a,\infty)}\left(\frac{\mathbf{U}_{1,n}}{n}\right)\frac{p_{\mathbf{U}}(\mathbf{U}_1^n)}{\pi_{\mathbf{U}}^a(\mathbf{U}_1^n)}\right] - P_n^2 \\ &= E_{P_{\mathbf{U}}}\left[\mathbb{1}_{(a,\infty)}\left(\frac{\mathbf{U}_{1,n}}{n}\right)\left(\exp\left(\frac{na^2}{2}\right)\right)(\exp(-a\mathbf{U}_{1,n}))\right] - P_n^2 \end{aligned}$$

A set B_n resulting as reducing the MSE should penalize large values of $-(\mathbf{U}_1 + ... + \mathbf{U}_n)$ while bearing nearly all the realizations of $\mathbf{U}_1 + ... + \mathbf{U}_n$ under the i.i.d. sampling scheme $\pi_{\mathbf{U}}^a$ as n tends to infinity. It should therefore be of the form (b,∞) for some $b = b_n$ so that

(a)

$$\lim_{n\to\infty} E_{\Pi^a_{\mathbf{U}}}\left[\mathbb{1}_{(b,\infty)}\left(\frac{\mathbf{U}_{1,n}}{n}\right)\right] = 1$$

and

(b)

$$\lim_{n\to\infty}\sup \frac{E_{P_{\mathbf{U}}}\left[\mathbb{1}_{(a,\infty)\cap(b,\infty)}\left(\frac{\mathbf{U}_{1,n}}{n}\right)\frac{p_{\mathbf{U}}(\mathbf{U}_1^n)}{\pi^a_{\mathbf{U}}(\mathbf{U}_1^n)}\right]}{V} < 1$$

which means that the IS sampling density $\pi^a_{\mathbf{U}}$ can lead a MSE defined by

$$MSE(B_n) := E_{P_{\mathbf{U}}}\left[\mathbb{1}_{(na,\infty)\cap(nb,\infty)}\frac{p_{\mathbf{U}}(\mathbf{U}_1^n)}{\pi^a_{\mathbf{U}}(\mathbf{U}_1^n)}\right] - P_n^2$$

with a clear gain over the variance indicator. However when $b \leq a$, (b) does not hold and, when $b > a$, (a) does not hold.

So no reduction of this variance can be obtained by taking into account the properties of the *typical paths* generated under the sampling density: a reduction of the variance is possible only by conditioning on "small" subsets of the sample paths space. On no classes of subsets of $\mathbb{R}^n$ with probability going to 1 under the sampling is it possible to reduce the variability of the estimate, whose rate is definitely proportional to $\sqrt{n}$, imposing a burden of order $L\sqrt{n}\alpha$ in order to achieve a relative efficiency of $\alpha\%$ with respect to P_n.

2.3.2 Efficiency of the adaptive twisted scheme

The gain in efficiency of the present proposal can be evaluated when some regularity condition holds pertaining to the density function of the r.v. $\mathbf{U}$. For such density the approximation of the density of $\mathbf{U}_{1,n}$ in the moderate and large deviation range, as well as the tail approximation of its distribution function is uniform. See [59], Chapter 6. These conditions are satisfied for most densities used in statistical modelling.

Regularity assumptions

1. Log-concave and almost Log-concave densities: $p_{\mathbf{U}}$ can be written as

$$p_{\mathbf{U}}(x) = c(x)\exp\left(-h(x)\right), \qquad x < \infty$$

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

$$c_1 < c(x) < c_2 \text{ for } x_0 < x < \infty.$$

Examples of densities which satisfy the above conditions include the Normal, the hyperbolic density, etc. An other example is when $\mathbf{U} := (\mathbf{X}-\psi)^2$ and $\mathbf{X}$ has log-normal distribution, $\mathbf{X} = \exp(\mathbf{Z})$ with $\mathbf{Z} \sim \mathcal{N}(\mu, \tau^2)$ and $\psi = E\mathbf{X} = \mu + \frac{1}{2}\tau$. Then

$$p_{\mathbf{U}}(x) = \frac{\exp(-\frac{3}{8}\tau)}{\sqrt{6\pi\tau^3}}$$

$$\left\{1 - \exp\left(-\sqrt{3x}\right)\right\}\exp\left\{\frac{\sqrt{3x}}{2} - \frac{x}{2\tau}\right\}$$

which is log-concave.

2. Gamma-like densities: the density of the r.v. $\mathbf{U}$ satisfies

$$p_{\mathbf{U}}(x) = c(x) \exp(-h(x))$$

for all x with $0 < c_1 < c(x) < c_2 \leq \infty$ when x is larger than some $x_0 > 0$ and $h(x)$ is a convex function which satisfies $h(x) = \tau + h_1(x)$ with, for $x_1 < x_2$,

$$a_1 \log \frac{x_2}{x_1} - b_1 < h_1(x_2) - h_1(x_1) < a_2 \log \frac{x_2}{x_1} - b_2$$

where a_1, a_2, b_1 and b_2 are positive constants with $a_2 < 1$.
A wide class of densities for which our results apply is when there exist constants $x_0 > 0$, $\alpha > 0$, $\tau > 0$ and A such that

$$p_{\mathbf{U}}(x) = A x^{\alpha-1} l(x) \exp(-\tau x) \qquad x > x_0$$

where $l(x)$ is slowly varying at infinity.

3. Densities defined through conditions on their characteristic function. We refer to see [59] Chapter 6 for an exhaustive set of such conditions which imply uniformity in the local and tail approximations of the distribution of the sample mean in the large (and, indeed, moderate) deviation case.

We denote by (R) any of the above condition pertaining to the regularity of the density $p_{\mathbf{U}}(x)$.

Gain of the IS procedure

We first put forwards a Lemma which assesses that large sets under the sampling distribution g_{nA} bear all what is needed to achieve a dramatic improvement of the relative efficiency of the IS procedure.

Lemma 33. *Assume $k/n \to 1$. It then holds,*

1. *There exist sets C_n in $\mathbb{R}^n$ such that*
   - $\lim_{n\to\infty} G_{nA}(C_n) = 1$
   - *for any y_1^n in C_n, $|\frac{p_{nA}}{g_{nA}}\left(y_1^k\right) - 1| < \delta_n$ with δ_n as in (1.50).*
   - *when $a \to E[\mathbf{U}]$ (moderate deviation),*

$$t^k s(t^k) = a\,(1 + o(1)) \tag{2.12}$$

   - *when $\lim_{n\to\infty} a_n$ is larger than $E[\mathbf{U}]$ (large deviation) , $t^k s(t^k)$ remains bounded away from 0 and infinity.*

Proof. Assume $k/n \to 1$. Let C_n in $\mathbb{R}^n$ such that for all y_1^n in C_n,

$$\left| \frac{p_{nA}(y_1^k)}{g_{nA}\left(y_1^k\right)} - 1 \right| < \delta_n$$

with δ_n as in (24) and

$$\left| \frac{m(t^k)}{a} - 1 \right| < \alpha_n$$

where t^k is defined through

$$m(t^k) := \frac{n}{n-k}\left(a - \frac{u_{1,k}}{n}\right)$$

with $u_{1,k} := \sum_{i=1}^{k} u(y_i)$ and α_n satisfies

$$\lim_{n\to\infty} \alpha_n = 0 \tag{2.13}$$

together with

$$\lim_{n\to\infty} \alpha_n a\sqrt{n-k} = \infty. \tag{2.14}$$

We prove that

$$\lim_{n\to\infty} G_{nA}\left(C_n\right) = 1.$$

Let

$$A_{n,\varepsilon_n} := A^k_{\varepsilon_n} \times \mathbb{R}^{n-k}$$

with

$$A^k_{\varepsilon_n} := \left\{ x_1^k : \left| \frac{p_{nA}(x_1^k)}{g_{nA}\left(x_1^k\right)} - 1 \right| < \delta_n \right\}.$$

By the above definition

$$\lim_{n\to\infty} P_{nA}\left(A_{n,\varepsilon_n}\right) = 1. \tag{2.15}$$

Note also that

$$\begin{aligned} G_{nA}\left(A_{n,\varepsilon_n}\right) &:= \int \mathbb{1}_{A_{n,\varepsilon_n}}(x_1^n) g_{nA}\left(x_1^n\right) dx_1^n \\ &= \int \mathbb{1}_{A^k_{\varepsilon_n}}(x_1^k) g_{nA}\left(x_1^k\right) dx_1^n \\ &\geq \frac{1}{1+\delta_n} \int \mathbb{1}_{A^k_{\varepsilon_n}}(x_1^k) p_{nA}(x_1^k) dx_1^k \\ &= \frac{1}{1+\delta_n}\left(1+o(1)\right) \end{aligned}$$

which goes to 1 as n tends to ∞. We have just proved that the sequence of sets A_{n,ε_n} contains roughly all the sample paths X_1^n under the importance sampling density g_{nA}.

We use the fact that t^k defined through

$$m(t^k) = \frac{n}{n-k}\left(a - \frac{u_{1,k}}{n}\right)$$

is close to a under p_{nv} uniformly upon v in $(a, a+c)$ and integrate out with respect to the distribution of $\mathbf{U}_{1,n}/n$ conditionally on $\mathbf{U}_{1,n}/n \in (a, a+c)$.

Let δ_n tend to 0 and $\lim_{n\to\infty} a\alpha_n\sqrt{n-k} = \infty$ and

$$B_n := \left\{ x_1^n : \left| \frac{m(t^k)}{a} - 1 \right| < \alpha_n \right\}.$$

We prove that on B_n

$$t^k s(t^k) = a\left(1+o(1)\right) \tag{2.16}$$

holds.

By Lemma 24 in Chapter 1 and integrating w.r.t. p_{nv} on $(a, a+c)$ it holds, under (2.13) and (2.14)

$$\lim_{n\to\infty} P_{nA}\left(B_n\right) = 1. \tag{2.17}$$

There exists δ'_n such that for any x_1^n in B_n

$$\left| \frac{t^k}{a} - 1 \right| < \delta'_n. \tag{2.18}$$

Indeed

$$\left|\frac{m(t^k)}{a}-1\right| = \left|\frac{t^k\left(1+v_k\right)}{a}-1\right| < \alpha_n$$

and $\lim_{n\to\infty} v_k = 0$. Therefore

$$1-\frac{v_k t^k}{a}-\alpha_n < \frac{t^k}{a} < 1-\frac{v_k t^k}{a}+\alpha_n.$$

Since $\frac{m(t^k)}{a}$ is bounded so is $\frac{t^k}{a}$ and therefore $\frac{v_k t^k}{a} \to 0$ as $n\to\infty$ which implies (2.18).

Further (2.18) implies that there exists δ_n" such that

$$\left|\frac{t^k s(t^k)}{a}-1\right| < \delta_n".$$

Indeed

$$\begin{aligned}\left|\frac{t^k s(t^k)}{a}-1\right| &= \left|\frac{t^k\left(1+u_k\right)}{a}-1\right| \\ &\le \delta_n' + \left(1+\delta_n'\right)u_k = \delta_n"\end{aligned}$$

where $\lim_{n\to\infty} u_k = 0$. Therefore (2.16) holds.

Define

$$C_n := B_n \cap A_{n,\varepsilon_n}$$

Since

$$\int \mathbb{1}_{C_n}(x_1^n) g_{nA}\left(x_1^k\right) dx_1^n \ge \frac{1}{1+\delta_n}\int \mathbb{1}_{C_n} p_{nA}(x_1^n) dx_1^n$$

and by (2.15) and (2.17)

$$\lim_{n\to\infty} P_{nA}\left(C_n\right) = 1$$

we obtain

$$\lim_{n\to\infty} G_{nA}\left(C_n\right) = 1.$$

which concludes the proof of (i) and (ii). □

We now evaluate the relative accuracy of the adaptive twisted IS estimator on this family of sets. Let

$$RE\left(\widehat{P_n}\right) = \frac{1}{L}\left(\frac{E_{G_{nA}}\left[\mathbb{1}_{C_n}\widehat{P_n}(l)^2\right]}{P_n^2}-1\right).$$

We prove that

Proposition 34. *When (R) holds together with the conditions in Theorem 19, the relative accuracy of the estimate $\hat{P}_n$ is given by*

$$RE(\widehat{P_n}) = \frac{\sqrt{2\pi}\sqrt{n-k-1}}{L}a(1+o(1))$$

as n tends to infinity.

Proof. Using the definition of C_n we get

$$
\begin{aligned}
& E_{G_{nA}}[\left(\mathbb{1}_{C_n}\widehat{P_n}(l)\right)^2] \\
&= P_n E_{P_{nA}}[\mathbb{1}_{C_n}(Y_1^n)\frac{p_{\mathbf{X}}(Y_1^k)p_{\mathbf{X}}(Y_{k+1}^n)}{g_{nA}(Y_1^k)g_{nA}(Y_{k+1}^n\big|\, Y_1^k)}] \\
&\leq P_n(1+\delta_n)E_{P_{nA}}[\mathbb{1}_{C_n}(Y_1^n)\frac{p_{\mathbf{X}}(Y_1^k)}{p(Y_1^k|\,\mathcal{E}_n)}\frac{p_{\mathbf{X}}(Y_{k+1}^n)}{g_{nA}(Y_{k+1}^n\big|\, Y_1^k)}] \\
&= P_n^2(1+\delta_n)E_{P_{nA}}[\mathbb{1}_{C_n}(Y_1^n)\frac{1}{p(\mathcal{E}_n|\, Y_1^k)}\frac{p_{\mathbf{X}}(Y_{k+1}^n)}{g_{nA}(Y_{k+1}^n\big|\, Y_1^k)}] \\
&= P_n^2(1+\delta_n)\sqrt{2\pi}\sqrt{n-k-1} \\
& E_{P_{nA}}[\mathbb{1}_{C_n}(Y_1^n)t^k s(t^k)](1+o(1)) \\
&= P_n^2 a\sqrt{2\pi}\sqrt{n-k-1}(1+o(1)).
\end{aligned}
$$

The third line is Bayes formula. The fourth line is Corollary 6.4.1 in [59]. The fifth line uses (2.12) and uniformity in the just above cited Corrollary, whose conditions are easily checked since, in its notation, $J(\theta) = \mathbb{R}$, condition (i) holds for θ in a neighborhood of 0 (Θ_0 indeed is restricted to such a set in our case), (ii) clearly holds and (iii) is a consequence of the assumption on the characteristic function of $u(\mathbf{X}_1)$. □

2.4 Algorithms

The next two algorithms 5 and 6 (associated with the algorithms presented in Chapter 1) provide the estimate of P_n.

Input : y_1^n, $p_{\mathbf{X}}$, n, k, a, M
Output : $g_{nA}(y_1^n)$
Procedure :
  for $m \leftarrow 1$ **to** M **do**
    Simulate v_m with density (2.7);
    Calculate $g_{nv_m}\left(y_1^k\right)$ with Algorithm 1 of Chapter 1;
    Calculate $g_{nv_m}\left(y_{k+1}^n|y_1^k\right) \leftarrow$ (2.8);
    Calculate $g_{nv_m}(y_1^n) \leftarrow g_{nv_m}\left(y_1^k\right) g_{nv_m}\left(y_{k+1}^n|y_1^k\right)$
  end
  Compute $g_{nA}(y_1^n) \leftarrow \frac{1}{M}\sum_{m=1}^{M} g_{nv_m}(y_1^n)$;
Return : $g_{nA}(y_1^n)$

Algorithm 5: Evaluation of $g_{nA}(y_1^n)$

Remark 35. *$\pi_{\mathbf{U}}^{\alpha_l}$ is defined as in (2.8)*

$$\alpha_l := \frac{n}{n-k}\left(v_l - \frac{u_{1,k}}{n}\right)$$

and

$$u_{1,k} = \sum_{i=1}^{k} u(Y_i(l)).$$

Input : $p_{\mathbf{X}}$, δ, n, a, M, L
Output : $\widehat{P_n}$
Initialization:
Set $k \to k_\delta$ with Algorithm 2;
Procedure :
for $l \leftarrow 1$ **to** L **do**
Simulate v_l with density (2.7);
Simulate $Y_1^k(l)$ with density g_{nv_l} with Algorithm 1 of Chapter 1;
Simulate $Y_{k+1}^n(l)$ i.i.d. with density $\pi_u^{\alpha_l}$;
Calculate $g_{nA}\left(Y_1^n(l)\right)$ with Algorithm 5;
Calculate $\widehat{P_n}(l) \leftarrow$ (2.9);
end
Compute $\widehat{P_n} \leftarrow$ (2.10);
Return : $\widehat{P_n}$

Algorithm 6: Calculation of $\hat{P}_n$

2.5 Simulation results

2.5.1 The gaussian case

The random variables $X_i's$ are i.i.d. with normal distribution with mean 0 and variance 1. The case treated here is $P_n = P\left(\frac{\mathbf{S}_{1,n}}{n} > a\right) = 0.009972$ with $n = 100$, and $a = 0.232$. We build the curve of the estimate of P_n (solid lines) and the two sigma confidence interval (dot lines) with respect to k. The value of L is $L = 2000$.

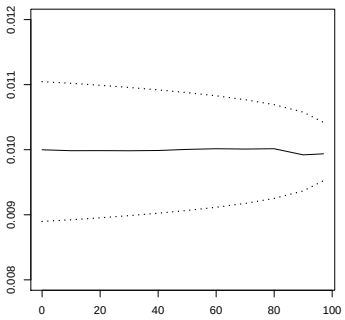

Figure 2.1: Curve of $\widehat{P_n}$ (solid line) in the normal case along with the two sigma confidence interval (dotted lines) as function of k with $n = 100$ for $L = 2000$.

2.5.2 The exponential case

The random variables $X_i's$ are i.i.d. with exponential distribution with parameter 1 on $(-1, \infty)$. The case treated here is $P_n = P\left(\frac{\mathbf{S}_{1,n}}{n} > a\right) = 0.013887$ with $n = 100$, and $a = 0.232$. The solid lines is the estimate of P_n, the dot lines are the two sigma confidence interval. Abscissa is k.

Figure 2.3 shows the ratio of the empirical value of the MSE of the adaptive estimate

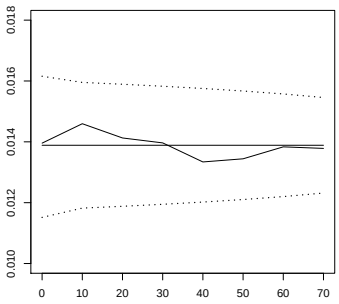

Figure 2.2: Curve of $\widehat{P_n}$ (solid line) in the exponential case along with the two sigma confidence interval (dotted lines) as function of k with $n = 100$ for $L = 2000$.

w.r.t. the empirical MSE of the i.i.d. twisted one, in the exponential case with $P_n = 10^{-2}$ and $n = 100$. The value of k is growing from $k = 1$ (i.i.d. twisted sample) to $k = 70$ (according to the rule of section 1.4.3). This ratio stabilizes to $\sqrt{n-k}/\sqrt{n}$ for $L = 2000$. The abscissa is k and the solid line is $k \to \sqrt{n-k}/\sqrt{n}$.

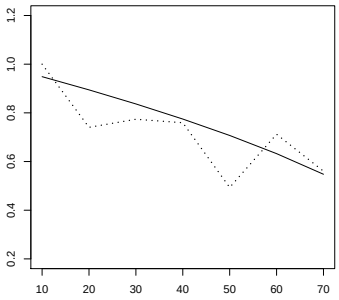

Figure 2.3: Ratio of the empirical value of the MSE of the adaptive estimate w.r.t. the empirical MSE of the i.i.d. twisted one (dotted line) along with the true value of this ratio (solid line) as a function of k.

2.6 A comparison study

2.6.1 With the classical twisted IS scheme

This section compares the performance of the present approach with respect to the standard tilted one as described in Section 2.1. The bilateral target set A considered here leads to similar sampling schemes as developped hereabove, substituting $p(\mathbf{U}_{1,n}/n = v|\,\mathbf{U}_{1,n} > na)$ by $p(\mathbf{U}_{1,n}/n = v|\,\mathbf{U}_{1,n} \in nA)$. Simulation of samples $\mathbf{U}_{1,n}/n$ under this density can be performed through Metropolis-Hastings algorithm, since

$$r(v, v') := \frac{p(\mathbf{U}_{1,n}/n = v|\,\mathbf{U}_{1,n} \in nA)}{p(\mathbf{U}_{1,n}/n = v'|\,\mathbf{U}_{1,n} \in nA)}$$

turns out to be independent upon $P(\mathbf{U}_{1,n} \in nA)$. The proposal distribution of the algorithm should be supported by A.

Consider a random sample $X_1, ..., X_{100}$ where X_1 has a normal distribution $N(0.05, 1)$ and let

$$\mathcal{E}_{100} := \left\{ x_1^{100} : \frac{|x_1 + ... + x_{100}|}{100} > 0.28 \right\}$$

for which

$$P_{100} = P\left((X_1, ..., X_{100}) \in \mathcal{E}_{100}\right) = 0.01120.$$

Our interest is to show that in this simple dissymmetrical case a direct extension of our proposal provides a good estimate, while the standard IS scheme ignores a part of the event $\mathcal{E}_{100}$. The standard i.i.d. IS scheme introduces the dominating point $a = 0.28$ and the family of i.i.d. tilted r.v's with common $N(a, 1)$ distribution. The resulting estimator of P_{100} is $0,01074$ (with $L = 1000$), indicating that the event $S_{1,100}/100 < -0.28$ is ignored in the evaluation of P_{100}, inducing a bias in the estimation. Since the simulated r.v's are independent under the tilted distribution the Importance factor oscillates wildly. Also the hit rate is of order 50%. It can also be seen that $S_1^{100}/100 < -0.28$ is never visited through the procedure.

This example is not as artificial as it may seem; indeed it leads to a two dominating points situation which is quite often met in real life. Exploring at random the set of interest under the distribution of $(x_1 + ... + x_{100})/100$ under $\mathcal{E}_{100}$ avoids any search for dominating points. Drawing L i.i.d. points $v_1, ..., v_L$ according to the distribution of $S_{1,100}/100$ conditionally upon $|S_{1,100}|/100 > 0.28$ we evaluate P_{100} with $k = 99$; note that in the gaussian case Theorem 9 of Chapter 1 provides an exact description of the conditional density of X_1^k for all k between 1 and n, and therefore the same nearly holds in Theorem 19. The resulting value of the estimate is 0.01125 which is fairly close to P_{100}.

As expected the Importance factor is very close to P_{100} for all sample paths X_1^n simulated under G_{nA}; this is in accordance with Theorem 9. Also the hit rate is very close to 100%. This example is in the same vein as the one developped in [38]. Under the present proposal the distribution of the Importance Factor concentrates around P_{100}; hence so-called "rogue path phenomenon" (see [38]) does not occur.

It is interesting to draw the hit rate as a function of k. When $k = 1$ then this rate is close to 50%, since the present algorithms coincides with the classical i.i.d. IS scheme. As k increases, the hit rate approaches 100%; the value of L is 1000.

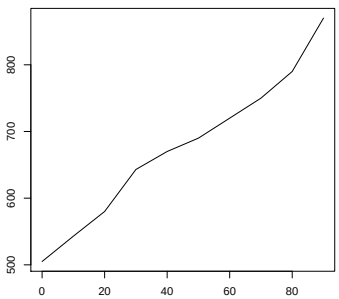

Figure 2.4: Curve of the hit Rate as a function of k with $n = 100$ for $L = 1000$.

In the present case the Importance factors have a slightly smaller variability around P_{100} under the present approach compared to that under the classical IS scheme; this is due to the fact that the contribution of the set $\mathcal{E}_{100}^-$ is roughly 10^{-2} times that of $\mathcal{E}_{100}$.

The relative errors (RE) are to be compared when arguing in terms of run time; the RE for the classical IS is 4% while it drops to .4% with the present proposal. Obviously letting L get larger in the classical IS than in the present approach results in an improved accuracy, without escaping the rogue path phenomenon. Although not discussed here Proposition 4.3 also holds in the present case.

We now explore the gain in relative accuracy when the dimension of the measured set increases. Let therefore$B := (\mathcal{E}_{100})^d$ which is the d-cartesian product of $\mathcal{E}_{100}$. The 100 r.v.'s X_i 's are i.i.d. random vectors in $\mathbb{R}^d$ with common i.i.d.$N(0.05, 1)$ distribution. The dominating point has all coordinates equal 0.28. Rogue path curse produces an overwhelming loss in accuracy, imposing a very large increase in runtime to get reasonable results. The following figure shows the gain in relative accuracy w.r.t. the classical IS scheme according to the growth of d.

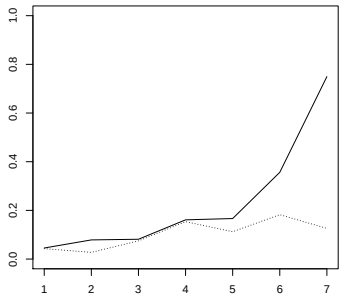

Figure 2.5: Relative Accuracy of the adaptive estimate (dotted line) w.r.t. i.i.d. twisted one (solid line) as a function of the dimension d for $L = 1000$.

2.6.2 With the Cross-Entropy method.

Quick overview of the Cross-Entropy method in our context.

The *Cross-Entropy* method, originally proposed by [84], is a type of importance-sampling technique in which the new distributions are successively calculated by minimizing the cross-entropy with respect to the ideal (but unattainable) zero-variance distribution, denoted g^*, defined in the Introduction of this chapter by

$$g^*(x) = \frac{p_\theta(x)}{P} \mathbb{1}_{S(x)\in A}$$

where $P = P[S(X) \in A]$ is the quantity to be estimated. The CE method proposes an IS sampling density within a parametric family that is in some sense the closest to g^*. Let briefly state the main step of the method (see [14] for a tutorial). Let p_θ the underlying density of X, where we make explicit the dependence of the parameter vector θ. Consider the family of probability densities p_ν indexed by the parameter vector ν within which to obtain the optimal importance density. Define the *Kullback-Leibler divergence*, or *cross-entropy distance*, between a density f and an another density g by

$$\mathcal{D}(f, g) := \int f(x) \log\left(\frac{f(x)}{g(x)}\right) dx$$

We want to locate the density p_{ν^*} such that $D(g^*, p_{\nu^*})$ is minimal. Now the functional minimization problem of finding an optimal importance density g reduces to a parametric minimization problem of finding the optimal parameter vector ν^* such as

$$\nu^* = \arg\min_{\nu} \mathcal{D}(g^*, p_\nu) = \arg\max_{\nu} \int g^*(x) \log\left(p_\nu(x)\right) dx.$$

Inserting the true value of g^*, we obtain

$$\nu^* = \arg\max_{\nu} \int \frac{p_\theta(x)}{P} \mathbb{1}_{S(x)\in A} \log\left(p_\nu(x)\right) dx.$$

The unknown value to be estimated P is independent of v, it holds

$$\nu^* = \arg\max_{\nu} \int p_\theta(x) \mathbb{1}_{S(x)\in A} \log\left(p_\nu(x)\right) dx.$$

This deterministic problem does not always admit an analytic solution. Instead, ν^* can be estimated by finding

$$\hat{\nu}^* = \arg\max_{\nu} \frac{1}{L} \sum_{l=1}^{L} \mathbb{1}_{S(X(l))\in A} \log\left(p_\nu\left(X(l)\right)\right). \tag{2.19}$$

where $X(l)$ is a i.i.d. sample from p_θ the nominal density. When $(S(x) \in A)$ is a rare event, the equation (2.19) can be rewritten, using importance sampling, as

$$\hat{v}^* = \arg\max_{v} \frac{1}{L} \sum_{l=1}^{L} \mathbb{1}_{S(X(l))\in A} W\left(X(l); \theta, v\right) \log\left(p_v\left(X(l)\right)\right). \tag{2.20}$$

where $X(l)$ is a i.i.d. sample from p_w for an arbitrary w and where

$$W\left(X(l); \theta, v\right) := \frac{p_\theta\left(X(l)\right)}{p_w\left(X(l)\right)}.$$

The solution of (2.20) can often be calculated analytically. In particular, this happens if the distribution of the random variables belong to a natural exponential family (see next section for example).

Input : ν_0, p_θ, L
Output : $\hat{\nu^*}$

Initialization:
$t = 1$;

Procedure :
repeat
Generate an i.i.d. sample $X(1), ..., X(L)$ from $p_{\nu_{t-1}}$;
Solve for ν_t equation (2.20);
$t = t + 1$;
until *Stopping Criterion*;

Return : $\hat{\nu^*}$

Algorithm 7: Evaluation of $\hat{\nu^*}$

This simple version of the CE method was improved in two main ways.

1. If the probability to be estimated is very small (typically $P_n \leq 10^{-5}$), a multi-level version of Algorithm 7 can be used to overcome this difficulty (see [14]).
2. In high dimensional settings, the multi-level CE algorithm can go wrong, and an improved version has been developped (see [24] and [54]).

In the example below, we focus only on the comparison of our proposed method with the generic CE algorithm since we study real event for which the probablity to be estimated in order 10^{-2}.

Example

We pursue the study of the dissymmetrical case presented in 2.6.

The sampling distribution is chosen as a normal one with variance 1, as adapted to this situation; the mean is estimated recursively through Algorithm 7. At step t, the update version of $\nu^{(t)}$ writes

$$\nu^{(t)} = \frac{\sum_{l=1}^{L} \mathbb{1}_{S_{1,n}(l)\in nA} W\left(X_1^l(l); \theta, \nu^{(t-1)}\right) S_{1,n}(l)}{\sum_{l=1}^{L} \mathbb{1}_{S_{1,n}(l)\in nA} W\left(X_1^l(l); \theta, \nu^{(t-1)}\right)} \tag{2.21}$$

In this dissymmetrical case, the convergence is very fast. However, the method depends strongly of the choice of $\nu^{(0)}$. In fact, when $\nu^{(0)}$ is close to 0.28 then the performance is similar to the classical IS scheme, since the successive means keep close to 0.28 (see Figure 2.6); at the contrary when $\nu^{(0)}$ is defined close to -0.28 the sequence of sampling distributions tend to concentrate around $N(-0.28, 1)$ (see Figure 2.7) and the resulting estimate produces a relative error of order 100%. Indeed it is roughly $\left|\left(4.8 \times 10^{-4} - 10^{-2}\right)/10^{-2}\right|$ since $P_{\mathcal{E}_{100}^-} \sim 4.8 \times 10^{-4}$ where $\mathcal{E}_{100}^- := \left\{x_1^{100} : \frac{x_1+\ldots+x_{100}}{100} < -0.28\right\}$.

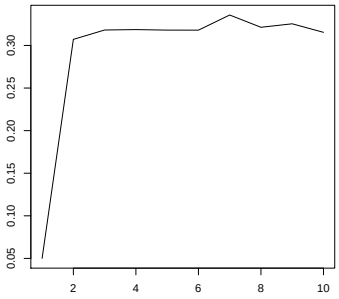

Figure 2.6: Convergence of the sequence $\nu^{(t)}$ for $n = 100$ in the case where $\nu^{(0)} = 0.05$.

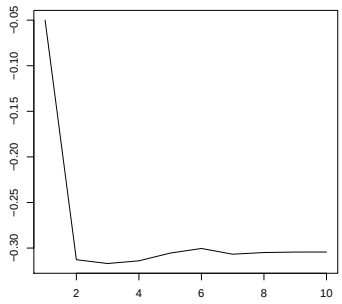

Figure 2.7: Convergence of the sequence $\nu^{(t)}$ for $n = 100$ in the case where $\nu^{(0)} = -0.05$.

Hence, CE method may be at odds even in simple but non standard cases.

Chapter 3

Conditional inference in parametric models.

3.1 Introduction and context

This paper explores conditional inference in parametric models. A comprehensive overview on this area is the illuminating review paper by Reid (1995) [78]. Our starting point is as follows: given a model $\mathcal{P}$ defined as a collection of continuous distributions P_θ on $\mathbb{R}^d$, with density p_θ where the parameter θ belongs to some subset Θ in $\mathbb{R}^s$ and given a sample of independent copies of a random variable with distribution P_{θ_T} for some unknown value θ_T of the parameter, we intend to provide some inference about θ_T conditioning on some observed statistics of the data. The situations which we have in mind are of two different kinds.

The first one is the Rao-Blackwellisation of estimators, which amounts to reduce the variance of an unbiased estimator by conditioning on any sufficient statistics; when the conditioning statistics is complete and sufficient for the parameter then this procedure provides optimal reduction, as stated by Lehmann-Scheffé Theorem. These facts yield the following questions.

1. is it possible to provide good approximations for the density of a sample conditioned on a given statistics, and, when applied for a model where some sufficient statistics for the parameter is known, does sufficiency w.r.t. the parameter still holds for the approximating density?
2. in the case when the first question has positive answer, is it possible to simulate samples according to the approximating density, and to propose some Rao-Blackwellised version for a given preliminary estimator? Also we would hope that the proposed method would be feasible, that the programming burden would be light, that the run time for this method be short, and that the involved techniques would keep in the range of globally known ones by the community of statisticians.

The second application of conditional inference pertains to the role of conditioning in models with nuisance parameters. There is a huge bibliography on this topic, some of which will be considered in details in the sequel. The usual frame for this field of problems is the exponential families one, for reasons related both with the importance of these models in applications and on the role of the concept of sufficiency when dealing with the notion of nuisance parameter. Conditioning on a sufficient statistics for the nuisance parameter produces a new exponential family, which gets free of this parameter, and allows for simple inference on the parameter of interest, at least in simple cases.

This will also be discussed, since the reality, as known, is not that simple, and since so many complementary approaches have been developped over decades in this area. Using the approximation of the conditional density in this context and performing simulations yields Monte Carlo tests for the parameter of interest, free from the nuisance parameter. Also conditional maximum likelihood estimators will be produced. Comparison with the parametric bootstrap will also be discussed.

This paper is organized as follows. Section 2 describes a general approximation scheme for the conditional density of long runs of subsamples conditioned on a statistics, with explicit formulas. The proof of the main result of this section is presented in Chapter 1. Discussion about implementation is provided. Section 3 presents two aspects of the approximating conditional scheme: we first show on examples that sufficiency is kept under the approximating scheme and, second, that this yields to an easy Rao-Blackwellisation procedure. An illustration of Lehmann-Scheffé Theorem is presented. Section 4 deals with models with nuisance parameters in the context of exponential families. We have found it useful to spend a few paragraphs on bibliographical issues. We address Monte Carlo tests based on the simulation scheme; in simple cases its performance is similar to that of parametric bootstrap; however conditional simulation based tests improve clearly over parametric bootstrap procedure when the test pertains to models for which the likelihood is multimodal with respect to the nuisance parameter; an example is provided. Finally we consider conditioned maximum likelihood based on the approximation of the conditional density; in simple cases its performance is similar to that of estimators defined through global likelihood optimization; however when the preliminary estimator of the nuisance is difficult to obtain, for example when it depends strongly on some initial point for a Newton-Raphson routine (this is indeed a very common situation), then, by the very nature of sufficiency, conditional inference based on the proxy of the conditional likelihood performs better; this is illustrated with examples.

3.2 The approximate conditional density of the sample

Most attempts which have been proposed for the approximation of conditional densities stem from arguments developped in [64] for inference on the parameter of interest in models with nuisance parameter; however the proposals in this direction hinge at the approximation of the distribution of the sufficient statistics for the parameter of interest given the observed value of the sufficient statistics of the nuisance parameter. We will present some of these proposals in the section devoted to exponential families. To our knowledge, no attempt has been made to approximate the conditional distribution of a sample (or of a long subsample) given some observed statistics.

However, generating samples from the conditional distribution itself (such samples are often called co-sufficient samples, following [71]) has been considered by many authors; see for example [42], [67] and references therein, and [68].

In [42], simulating exponential or normal samples under the given value of the empirical mean is proposed. For example under the exponential distribution $Exp(\theta)$, the minimal sufficient statistics for θ is the sum of the observations, say t_n; a co-sufficient sample x^* can be created by generating an $x^{'}$-sample from $Exp(1)$ and taking $x_i^* = x_i^{'} t_n / \overline{x'}$. However, this approach may be at odd in simple cases, as for the Gamma density in the non exponential case.

Lockhart et al. [71] proposed a different framework based on the Gibbs sampler, simulating the conditioned sample one at a time through a sequential procedure. The example

which is presented is for the Gamma distribution under the empirical mean; in these examples it seems to perform well for location parameter, when the true parameter is in some range, therefore not uniformly on the model. Their paper contains a comparative study with the parametric bootstrap procedure (introduced in [41]) for similar problems. In a simple case, they argue favorably for both methods. We will turn back to global likelihood maximization in relation with conditional likelihood estimators, in the last section of this paper.

Other techniques have been developped in specific cases: for the inverse gaussian distribution see [74] or [25]; for the Weibull distribution see [69]. No unified technique exists in the literature which would work under general models.

3.2.1 Approximation of conditional densities

Notation and hypotheses

For sake of clearness we consider the case when the model $\mathcal{P}$ is a family of distributions on $\mathbb{R}$. Extension to $\mathbb{R}^d, d > 1$ can be achieved in the same way, using similar results developed in future work.

Denote $\mathbf{X}_1^n := (\mathbf{X}_1, .., \mathbf{X}_n)$ a set of n independent copies of a real random variable $\mathbf{X}$ with density $p_{\mathbf{X},\theta_T}$ on $\mathbb{R}$. Let $\mathbf{x}_1^n := (\mathbf{x}_1, ..., \mathbf{x}_n)$ denote the observed values of the data, each $\mathbf{x}_i$ resulting from the sampling of $\mathbf{X}_i$. Define the r.v. $\mathbf{U} := u(\mathbf{X})$ and $\mathbf{U}_{1,n} := u(\mathbf{X}_1) + ... + u(\mathbf{X}_n)$ where u is a real-valued measurable function on $\mathbb{R}$, and, accordingly, $u_{1,n} := u(\mathbf{x}_1) + ... + u(\mathbf{x}_n)$. Denote $p_{\mathbf{U},\theta_T}$ the density of the r.v. $\mathbf{U}$. We consider approximations of the density of the vector $\mathbf{X}_1^k = (\mathbf{X}_1, .., \mathbf{X}_k)$ on $\mathbb{R}$ when $\mathbf{U}_{1,n} = u_{1,n}$. It will be assumed that the observed value $u_{1,n}$ is "typical", in the sense that it satisfies the law of large numbers. Since the approximation scheme for the conditional density is validated through limit arguments, it will be assumed that the sequence $u_{1,n}$ satisfies

$$\lim_{n\to\infty} \frac{u_{1,n}}{n} = Eu(\mathbf{X}). \tag{3.1}$$

We state a similar version to the Theorem 9 of Chapter 1 in order to make the appearance of the parameter θ_T clear. We propose an approximation for

$$p_{u_{1,n},\theta_T}\left(x_1^k\right) := p_{\theta_T}(x_1^k | \mathbf{U}_{1,n} = u_{1,n})$$

where $\mathbf{X}_1^k := (\mathbf{X}_1, .., \mathbf{X}_k)$ and $k := k_n$ is an integer sequence such that

$$0 \leq \lim\sup_{n\to\infty} k/n \leq 1 \tag{K1}$$

together with

$$\lim_{n\to\infty} n - k = \infty \tag{K2}$$

which is to say that we may approximate $p_{u_{1,n},\theta_T}\left(x_1^k\right)$ on long runs. The rule which defines the value of k for a given accuracy of the approximation is stated in section 3.2 of Chapter 1.
The hypotheses pertaining to the function u and the r.v. $\mathbf{U} = u(\mathbf{X})$ are as follows.

1. u is real valued and the characteristic function of the random variable $\mathbf{U}$ is assumed to belong to L^r for some $r \geq 1$.

2. The r.v. $\mathbf{U}$ is supposed to fulfill the Cramer condition: its moment generating function satisfies
$$\phi_{\mathbf{U}}(t) := E \exp t\mathbf{U} < \infty$$
for t in a non void neighborhood of 0.

Define the functions $m(t), s^2(t)$ and $\mu_3(t)$ as the first, second and third derivatives of $\log \phi_{\mathbf{U}}(t)$. Denote
$$\pi^{\alpha}_{u,\theta_T}(x) := \frac{\exp tu(x)}{\phi_{\mathbf{U}}(t)} p_{\mathbf{X},\theta_T}(x)$$
with $m(t) = \alpha$ and α belongs to the support of $P_{\mathbf{X},\theta_T}$, the distribution of $\mathbf{X}$. The density $\pi^{\alpha}_{u,\theta_T}$ is the *tilted* density with parameter α. Also it is assumed that this latest definition of t makes sense for all α in the support of $\mathbf{X}$. Conditions on $\phi_{\mathbf{U}}(t)$ which ensure this fact are referred to as *steepness properties*, and are exposed in [6], p153 and followings.

We introduce a positive sequence ϵ_n which satisfies

$$\lim_{n\to\infty} \epsilon_n \sqrt{n-k} = \infty \tag{E1}$$
$$\lim_{n\to\infty} \epsilon_n (\log n)^2 = 0. \tag{E2}$$

3.2.2 The proxy of the conditional density of the sample

The density $g_{u_{1,n},\theta_T}(x_1^k)$ on $\mathbb{R}^k$, which approximates $p_{u_{1,n},\theta_T}\left(x_1^k\right)$ sharply with relative error smaller than $\epsilon_n (\log n)^2$ is defined recusively as follows.

Set
$$m_0 := u_{1,n}/n$$
and
$$g_0(x_1|\, x_0) := \pi_u^{m_0}(x_1)$$
with x_0 arbitrary, and for $1 \le i \le k-1$ define the density $g(x_{i+1}|\, x_1^i)$ recursively.

Set t_i the unique solution of the equation
$$m_i := m(t_i) = \frac{u_{1,n} - u_{1,i}}{n-i} \tag{3.2}$$
where $u_{1,i} := u(x_1) + ... + u(x_i)$. The tilted adaptive family of densities $\pi^{m_i}_{u,\theta_T}$ is the basic ingredient of the derivation of approximating scheme. Let
$$s_i^2 := \frac{d^2}{dt^2}\left(\log E_{\pi^{m_i}_{u,\theta_T}} \exp tu(\mathbf{X})\right)(0)$$
and
$$\mu_j^i := \frac{d^j}{dt^j}\left(\log E_{\pi^{m_i}_{u,\theta_T}} \exp tu(\mathbf{X})\right)(0),\ j = 3,4$$
which are the second, third and fourth cumulants of $\pi^{m_i}_{u,\theta_T}$. Let
$$g(x_{i+1}|\, x_1^i) = C_i p_{\mathbf{X},\theta_T}(x_{i+1}) \mathfrak{n}(\alpha\beta + m_0, \beta, u(x_{i+1})) \tag{3.3}$$
where $\mathfrak{n}(\mu, \tau, x)$ is the normal density with mean μ and variance τ at x. Here
$$\beta = s_i^2 (n-i-1) \tag{3.4}$$

$$\alpha = t_i + \frac{\mu_3^i}{2s_i^4\,(n-i-1)} \tag{3.5}$$

and the C_i is a normalizing constant.

Define

$$g_{u_{1,n},\theta_T}(x_1^k) := g_0(x_1|\,x_0)\prod_{i=1}^{k-1} g(x_{i+1}|\,x_1^i). \tag{3.6}$$

It holds

Theorem 36. *Assume (K1,K2) together with (E1,E2). Then*

1. *Let Y_1^k be a sample with density $p_{u_{1,n}}$.*

$$p_{u_{1,n},\theta_T}(x_1^k) = g_{u_{1,n},\theta_T}(x_1^k)(1 + o_{P_{u_{1,n},\theta_T}}(\epsilon_n\,(\log n)^2))$$

2. *Let Y_1^k be a sample with density $g_{u_{1,n}}$.*

$$p_{u_{1,n},\theta_T}(x_1^k) = g_{u_{1,n},\theta_T}(x_1^k)(1 + o_{G_{u_{1,n},\theta_T}}(\epsilon_n\,(\log n)^2)).$$

3. *The total variation distance between $P_{u_{1,n},\theta_T}$ and $G_{u_{1,n},\theta_T}$ goes to 0 as n tends to infinity.*

For the proof, see Chapter 1.

Statement (i) means that the conditional likelihood of any long sample path $\mathbf{X}_1^k$ given $\mathbf{U}_{1,n} = u_{1,n}$ can be approximated by $G_{u_{1,n},\theta_T}(\mathbf{X}_1^k)$ with a small relative error on typical realizations of $\mathbf{X}_1^n$.

The second statement states that simulating $\mathbf{X}_1^k$ under $g_{u_{1,n},\theta_T}$ produces runs which could have been sampled under the conditional density $p_{u_{1,n},\theta_T}$ since $g_{u_{1,n}}$ and $p_{u_{1,n},\theta_T}$ coincide sharply on larger and larger subsets of $\mathbb{R}^k$ as n increases.

Remark 37. *Theorem 36 states that the density $g_{u_{1,n},\theta_T}$ on $\mathbb{R}^k$ approximates $p_{u_{1,n},\theta_T}$ on the sample $\mathbf{X}_1^n$ generated under θ_T. However, in some cases, the r.v.'s $\mathbf{X}_i$'s in Theorem 36 may at time be generated under some other parameters, say θ_0. Indeed, for direct applications developped in this paper, Theorem 36 have to hold when the sample is generated under an other sampling scheme. Theorem 12 in Chapter 1 states that the approximation scheme holds true in this case. Indeed let $\mathbf{Y}_1^n$ be i.i.d. copies with distribution P_{θ_0} . Define*

$$p_{u_{1,n},\theta_0}\left(y_1^k\right) := p_{\theta_0}\left(\mathbf{Y}_1^k = y_1^k \middle|\, \mathbf{U}_{1,n} = u_{1,n}\right)$$

with distribution $P_{u_{1,n},\theta_0}$. It then holds

Theorem 38. *Then, with the same hypotheses and notation as in Theorem 36,*

$$p_{\theta_T}\left(\mathbf{X}_1^k = Y_1^k\,|\mathbf{U}_{1,n} = u_{1,n}\right) = g_{u_{1,n},\theta_T}(Y_1^k)(1 + o_{P_{u_{1,n},\theta_0}}(\epsilon_n\,(\log n)^2)).$$

3.2.3 Comments on implementation

The simulation of a sample X_1^k with density $g_{u_{1,n},\theta_T}$ is fast as easy. Indeed the r.v. X_{i+1} with density $g\,(x_{i+1}|x_1^i)$ is obtained through a standard acceptance -rejection algorithm. When θ_T is unknown, a preliminary estimator may be used. When $\mathbf{U}_{1,n}$ is sufficient for $p_{u_{1,n},\theta}$ it is nearly sufficient for its proxy $g_{u_{1,n},}$ (see next section); indeed changing the value of this preliminary estimator does not alter the value of the likelihood of the sample; as shown in the simulations developped here after, any value of θ can be used; call θ^* the

value of θ chosen as initial value , using henceforth $p_{\mathbf{X},\theta^*}$ instead of $p_{\mathbf{X},\theta_T}$ in (3.3). In exponential families the values of the parameters which appear in the gaussian component of $g\left(x_{i+1}|x_1^i\right)$ in (3.3) are easily calculated; note also that due to (3.1) the parameters in $\mathfrak{n}\left(\alpha\beta,\beta,u\left(x_{i+1}\right)\right)$ are such that the dominating density can be chosen for all i as $p_{\mathbf{X},\theta^*}$. The constant in the acceptance rejection algorithm is then $1/\sqrt{2\pi\alpha}$. This is in contrast with the case when the conditioning value is in the range of a large deviation with respect to $p_{\mathbf{X},\theta_T}$; in this case, which appears in a natural way in Importance sampling estimation for rare event probabilities, the simulation algorithm is more complex ; see Chapter 2.

3.3 Sufficient statistics and approximated conditional density

3.3.1 Keeping sufficiency under the proxy density

The density $g_{u_{1,n},\theta_T}(y_1^k)$ is used in order to handle Rao -Blackellisation of estimators or statistical inference for models with nuisance parameters. The basic property is sufficiency with respect to the nuisance parameter. We show on some examples that the family of densities $g_{u_{1,n},\theta}(y_1^k)$ defined in (3.6), when indexed by θ, inherits of the invariance with respect to the parameter θ when conditioning on a sufficient statistics.

Consider the Gamma density

$$f_{r,\theta}(x) := \frac{\theta^{-(r+1)}}{\Gamma(r+1)} x^r \exp -x/\theta \quad \text{for } x > 0. \tag{3.7}$$

As r varies in $(-1,\infty)$ and θ is positive, the density runs in an exponential family with parameters r and θ, and sufficient statistics $t(x) := \log x$ and $u(x) := x$ respectively for r and θ. Given an i.i.d. sample $X_1^n := (X_1,...,X_n)$ with density f_{r_T,θ_T} the resulting sufficient statistics are respectively $T_{1,n} := \log X_1 + ... + \log X_n$ and $U_{1,n} := X_1 + ... + X_n$. We consider two parametic models $(f_{r_T,\theta}, \theta \geq 0)$ and $(f_{r,\theta_T}\ , r > 0)$ respectively assuming r_T or θ_T known.

We first consider sufficiency of $\mathbf{U}_{1,n}$ in the first model. The density $g_{u_{1,n},(r_T,\theta_T)}(y_1^k)$ should be free of the current value of the true parameter θ_T of the parameter under which the data are drawn. However as appears in (3.6) the unknown value θ_T should be used in its very definition. We show by simulation that whatever the value of θ inserted in place of θ_T in (3.6) the value of the likelihood of X_1^k under $g_{u_{1,n},(r_T,\theta)}$ does not depend upon θ. We thus observe that $\mathbf{U}_{1,n}$ is "sufficient" for θ in the conditional density approximating $p_{u_{1,n},(r_T,\theta)}$, as should hold as a consequence of Theorem 36 . Say that $\mathbf{U}_{1,n}$ is quasi sufficient for θ in $g_{u_{1,n},(r_T,\theta)}$ if this loose invariance holds.

Similarly the same fact occurs in the model $(f_{r,\theta_T}, r > 0)$.

In both cases whatever the value of the parameter θ (Figure 3.1) or r (Figure 3.2), the likelihood of X_1^k remains constant.

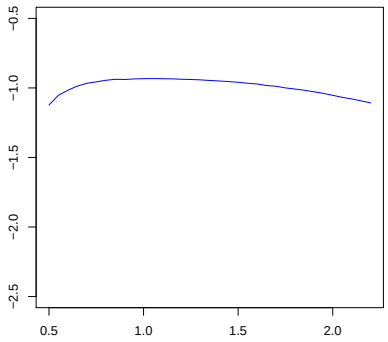

Figure 3.1: Proxy of the conditional likelihood of X_1^k under $g_{T_{1,n}}$ as a function of θ for $n = 100$ and $k = 80$ in the gamma case.

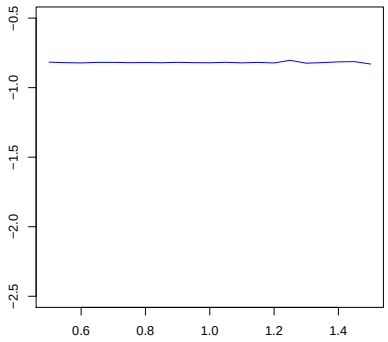

Figure 3.2: Proxy of the conditional likelihood of X_1^k under $g_{U_{1,n}}$ as a function of r for $n = 100$ and $k = 80$ in the gamma case.

We also consider the Inverse Gaussian distribution with density

$$f_{\lambda,\mu}(x) := \left[\frac{\lambda}{2\pi}\right]^{1/2} \exp -\frac{\lambda\,(x-\mu)^2}{2\mu^2 x} \qquad \text{for } x > 0$$

with both parameters λ and μ be positive. Given an i.i.d. sample $X_1^n := (X_1, ..., X_n)$ with density $f_{\mu,\lambda}$, the resulting sufficient statistics are respectively $T_{1,n} := X_1 + ... + X_n$ and $U_{1,n} := X_1^{-1} + ... + X_n^{-1}$. Similarly as for the Gamma case we draw the likelihood of a subsample X_1^k under $g_{u_{1,n},(\lambda,\mu_T)}$ with $T_{1,n} := X_1 + ... + X_n$,which is a sufficient statistics for μ (Figure 3.3), and upon $U_{1,n} := X_1^{-1} + ... + X_n^{-1}$ which is sufficient for λ (Figure 3.4). In either cases the other coefficient is kept fixed at the true value of the parameter generating the sample. As for the Gamma case these curves show the invariance of the proxy of the conditional density with respect to the parameter for which the chosen statistics is sufficient.

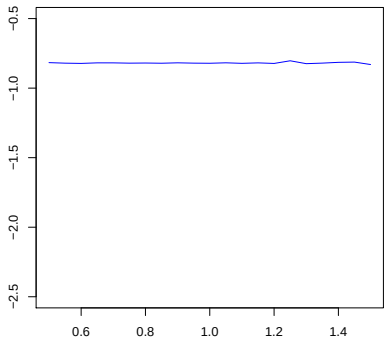

Figure 3.3: Conditional likelihood of X_1^k under $g_{T_{1,n}}$ as a function of μ for $n = 100$ and $k = 80$ in the Inverse Gaussian case.

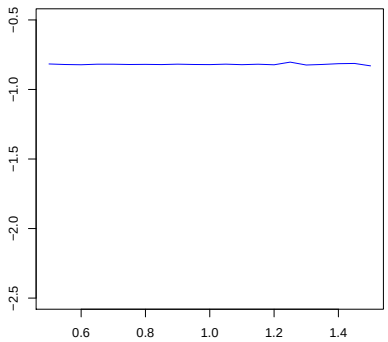

Figure 3.4: Conditional likelihood of X_1^k under $g_{U_{1,n}}$ as a function of λ for $n = 100$ and $k = 80$ in the Inverse Gaussian case.

3.3.2 Rao-Blackwellisation

Rao-Blackwell Theorem holds regardless of whether biased or unbiased estimators are used, since conditioning reduces the MSE (see Remark 39). Although its statement is rather weak, in practice the improvement is often enormous. New interest in Rao-Blackwellisation procedures have risen in the recent years, conditioning on ancillary variables (see [49] for a survey on ancillaries in conditional inference); specific Rao-Blackwellisation schemes have been proposed in [22], [23], [77], [81] and [56], whose purpose is to improve the variance of a given statistics (for example a tail probability) under a *known* distribution, through a simulation scheme under this distribution; the ancillary variables used in the simulation process itself are used as conditioning ones for the Rao-Blackwellisation of the statistics. The present approach is more classical in this respect, since we do not assume that the parent distribution is known; conditioning on a sufficient statistics $\mathbf{U}_{1,n}$ with respect to the parameter θ and simulating samples according to the approximating density $g_{u_{1,n},\theta}$ will produce the improved estimator.

Since $\mathbf{U}_{1,n}$ is sufficient for the parameter θ in $g_{u_{1,n},\theta}$ it can be used in order to obtain improved estimators of θ_T through Rao Blackwellization. We shortly illustrate the procedure and its results on some toy cases. Consider again the Gamma family defined here-above with canonical parameters r and θ.

First the parameter to be estimated is θ_T. A first unbiaised estimator is chosen as

$$\widehat{\theta_2} := \frac{X_1 + X_2}{2r_T}.$$

Given an i.i.d. sample X_1^n with density f_{r_T,θ_T} the Rao-Blackwellised estimator of $\widehat{\theta_2}$ is defined through

$$\theta_{RB,2} := E\left(\widehat{\theta_2}\middle|\, \mathbf{U}_{1,n}\right)$$

whose variance is less than $Var\widehat{\theta_2}$.

Consider $k = 2$ in $g_{u_{1,n},(r_T,\theta_T)}(y_1^k)$ and let (Y_1, Y_2) be distributed according to $g_{u_{1,n},(r_T,\theta_T)}(y_1^2)$; note that any value θ can be used in practice instead of the unknown value θ_T, by quasi sufficiency of $\mathbf{U}_{1,n}$. Replications of (Y_1, Y_2) induce an estimator of $\theta_{RB,2}$ for fixed $u_{1,n}$. Iterating on the simulation of the runs X_1^n produces, for $n = 100$ an i.i.d. sample of $\theta_{RB,2}$'s and the $Var\theta_{RB,2}$ is estimated. The resulting variance shows a net improvement with respect to the estimated variance of $\widehat{\theta_2}$. It is of some interest to confront this gain in variance as the number of terms involved in $\widehat{\theta_k}$ increases together with k. As k approaches n the variance of $\widehat{\theta_k}$ approaches the Cramer-Rao bound. The graph below shows the decay in variance of $\widehat{\theta_k}$. We note that whatever the value of k the estimated value of the variance of $\theta_{RB,k}$ is constant. This is indeed an illustration of Lehmann-Scheffé's theorem.

Remark 39. *It is often assumed that Rao-Blackwell holds for unbiased estimators only. However, the improvment of estimators through conditionning also holds for biased estimators.*

Let θ_1 be an estimator of θ and T any sufficient statistic. Define $\theta_2 := E[\theta_1|T]$. By the law of total variation, it holds

$$Var(\theta_1) = E[Var(\theta_1|T)] + Var(\theta_2) \geq Var(\theta_2) \tag{3.8}$$

Now, $E[\theta_2] = E[\theta_1]$, hence θ_1 and θ_2 have same bias. From 3.8, it results

$$MSE(\theta_2) := Var(\theta_2) + (E[\theta_2] - \theta)^2 \leq Var(\theta_1) + (E[\theta_1] - \theta)^2 := MSE(\theta_1).$$

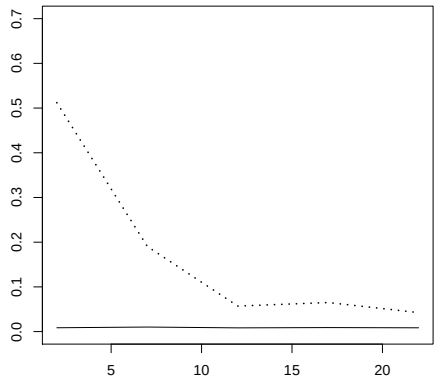

Figure 3.5: Variance of $\widehat{\theta}_k$, the initial estimator (dotted line), along with the variance of $\theta_{RB,k}$, the Rao-Blackwellised estimator (solid line) with $n = 100$ as a function ok k.

Remark 40. *Lockhart and O'Reilly ([70]) establish, under certain conditions and for fixed k, the asymptotic equivalence of the plug-in estimate for the distribution $P_{\theta_{ML}}\left(\mathbf{X}_1^k \in B\right)$ and the Rao-Blackwell estimate $P\left(\mathbf{X}_1^k \in B \middle| \mathbf{U}_{1,n}\right)$ where θ_{ML} is the maximum likelihood estimator of θ_T based on the whole sample $\mathbf{X}_1^n$ (this result is known as Moore's conjecture (see [73])). They also provide rates for this convergence.*

3.4 Exponential models with nuisance parameters

3.4.1 Conditional inference in exponential families

We consider the case when the parameter consists in two distinct subparameters, one of interest denoted θ and a nuisance component denoted η. As is well known, conditioning on a sufficient statistics for the nuisance parameter produces a new exponential family which is free of it. Assuming the observed dataset $\mathbf{x}_1^n := (\mathbf{x}_1, ..., \mathbf{x}_n)$ resulting from sampling of a vector $\mathbf{X}_1^n := (\mathbf{X}_1, ..., \mathbf{X}_n)$ of i.i.d. random variables with distribution in the initial exponential model, and denoting $\mathbf{U}_{1,n}$ a sufficient statistics for η, simulation of samples under the conditional distribution of $\mathbf{X}_1^n$ given $\mathbf{U}_{1,n} = u_{1,n}$ and $\theta = \theta_0$ for some θ_0 produces the basic ingredient for Monte Carlo tests with $H_0 : \theta_T = \theta_0$ where θ_T stands for the true value of the parameter of interest. Changing θ_0 for other values of the parameter of interest produces power curves as functions of the level of the test. This is the well known principle of Monte Carlo tests (see [11] or [12]), which are considered hereunder. We consider a steep but not necessarily regular exponential family exponential family $\mathcal{P} := \{P_{\mathbf{X},(\theta,\eta)}, (\theta, \eta) \in \mathcal{N}\}$ defined on $\mathbb{R}$ with canonical parametrization (θ, η) and minimal sufficient statistics (t, u) defined on $\mathbb{R}$ through the density

$$p_{\mathbf{X},(\theta,\eta)}(x) := \frac{dP_{\mathbf{X},(\theta,\eta)}(x)}{dx} = \exp\left[\theta t(x) + \eta u(x) - K(\theta, \eta)\right] h(x). \tag{3.9}$$

For notational conveniency and without loss of generality both θ and η belong to $\mathbb{R}$. Also the model can be defined on $\mathbb{R}^d$, $d > 1$, at the cost of similar but more envolved tools. The natural parameter space is $\mathcal{N}$ (which is a convex set in $\mathbb{R}^2$) defined as the effective

domain of

$$k(\theta,\eta) := \exp\left[K(\theta,\eta)\right] = \int \exp\left[\theta t(x) + \eta u(x)\right] h(x)dx. \tag{3.10}$$

As above denote $\mathbf{x}_1^n := (\mathbf{x}_1, ..., \mathbf{x}_n)$ be the observed values of n i.i.d. replications of a general random variable $\mathbf{X}$ with density (3.9). Denote

$$T_{1,n} := \sum_{i=1}^{n} t(\mathbf{x}_i) \quad \text{and} \quad U_{1,n} := \sum_{i=1}^{n} u(\mathbf{x}_i). \tag{3.11}$$

Basu [10] discusses ten different ways for eliminating the nuisance parameters, among which conditioning on sufficient statistics and consider UMPU tests pertaining to the parameter of interest. In most cases, the density of $T_{1,n}$ given $U_{1,n} = u_{1,n}$ is unknown. Two main ways have been developped to deal with this issue: approximating this conditional density of a statistics or simulating samples from the conditional density. These two approaches are combined hereunder.

The classical technique is to approximate this conditional density using some expansion. Then integration produces critical values. For example, Pedersen [76] defines the mixed Edgeworth-saddlepoint approximation, or the single saddlepoint approximation. However, the main issue of this technique is that the approximated density still depends on the nuisance parameter. In order to obtain the expansion, some suitable values for the parameter of interest and for the nuisance parameter have to be chosen. In the method developped here, as seen before, the conditional approximated density inherits of the invariance with respect to the nuisance parameter when conditioning on a sufficient statistics pertaining to this parameter.

Rephrasing the notation of Section 2 in the present setting the MLE (θ_{ML}, η_{ML}) satisfies

$$\left.\frac{\partial K(\theta,\eta)}{\partial \eta}\right|_{\theta_{ML},\eta_{ML}} = u_{1,n}/n$$

and therefore $u_{1,n}/n$ converges to $\left(\frac{\partial K(\theta_T,\eta)}{\partial \eta}\right)^{-1}(\eta_T)$.

For notational clearness denote μ the expectation of $u(\mathbf{X}_1)$ and σ^2 its variance under (θ_T, η_T) , hence

$$\mu := \mu_{(\theta_T,\eta_T)} := \partial K(\theta_T,\eta_T)/\partial\eta \qquad \sigma^2 := \sigma^2_{(\theta_T,\eta_T)} := \partial^2 K(\theta_T,\eta_T)/\partial\eta^2$$

Assume at present θ_T and η_T known. It holds

$$\phi(r) := E_{(\theta_T,\eta_T)} \exp[ru(\mathbf{X})] = \exp\left[K(\theta_T,\eta_T + r) - K(\theta_T,\eta_T)\right]$$

and

$$\begin{aligned} m(r) &= \mu_{(\theta_T,\eta_T+r)} \\ s^2(r) &= \sigma^2_{(\theta_T,\eta_T+r)} \\ \mu_3(r) &= \partial^3 K(\theta_T,\eta_T + r)/\partial r^3 \ . \end{aligned}$$

Further

$$\pi^{\alpha}_{u,\theta_T,\eta_T}(x) := \frac{\exp ru(x)}{\phi(r)} p_{\mathbf{X},(\theta_T,\eta_T)}(x) = p_{\mathbf{X},(\theta_T,\eta_T+r)}(x) \tag{3.12}$$

for any given α in the range of $P_{\mathbf{X},(\theta_T,\eta_T)}$. In the above formula (3.12) the parameter r denotes the only solution of the equation

$$m(r) = \alpha.$$

For large k depending on n, using Monte Carlo tests based on runs of length k instead of n does not affect the accuracy of the results.

3.4.2 Application of conditional sampling to MC tests

Consider a test defined through $H0 : \theta_T = \theta_0$ versus $H1 : \theta_T \neq \theta_0$. Monte Carlo (MC) tests aim at obtaining $p-$values through simulation when the distribution of the desired test statistics under $H0$ is either unknown or very cumbersome to obtain; a comprehensive reference is [61].

Recall the principle of thoses tests: denote t the observed value of the studied statistic based on the dataset and let $t_2, .., t_L$ the values of the resulting test statistics obtained through the simulation of $L-1$ samples $\mathbf{X}_1^n$ under $H0$. If t is the Mth largest value of the sample $(t, t_2, ..., t_L)$, H_0 will be rejected at the $\alpha = M/L$ signifiance level, since the rank of t is uniformly distributed on the integer $2, ..., L$ when $H0$ holds. The present MC procedure uses simulated samples under the proxy of $p_{u_{1,n},(\theta_0,\eta_t)}$. Using quasi-sufficiency of $\mathbf{U}_{1,n}$ we may use any value in place of η_T; we have compared this simple choice with the common use, inserting the MLE $\hat{\eta}_{\theta_0}$ in place of η_T in $g_{u_{1,n},(\theta_0,\eta_t)}$. This estimate $\hat{\eta}_{\theta_0}$ is the MLE of η_T in the one parameter family $p_{\mathbf{X},(\theta_0,\eta)}$ defined through (3.9); this choice follows the commonly used one, as advocated for instance in [76] and [75]. Innumerous simulation studies support this choice in various contexts; we found no difference in the resulting procedures.

Consider the problem of testing the null hypothesis $H0 : \theta_T = \theta_0$ against the alternative $H1 : \theta_T > \theta_0$ in model (3.9) where η is the nuisance parameter.

When $p_{u_{1,n},(\theta_0,\eta_T)}$ is known, the classical conditional test $H0 : \theta_T = \theta_0$ versus $H1 : \theta_T > \theta_0$ with level α is UMPU.

Substituting $p_{u_{1,n},(\theta_0,\eta_T)}\left(\mathbf{X}_1^n = x_1^n \,|\mathbf{U}_{1,n} = u_{1,n}\right)$ by $g_{u_{1,n},(\theta_0,\eta_T)}\left(x_1^k\right)$ defined in (3.6), i.e. substituting the test statistics T_1^n by T_1^k and $p_{\theta_0}\left(\mathbf{X}_1^k = x_1^k \,|\mathbf{U}_{1,n} = u_{1,n}\right)$ by $g_{u_{1,n},(\theta_0,\eta_T)}\left(x_1^k\right)$ i.e. changing the model for a proxy while keeping the same parameter of interest θ yields the conditional test with level α

$$\psi_\alpha(x_1^k) := \begin{cases} 1 & \text{if } T_{1,k} > t_\alpha \\ \gamma & \text{if } T_{1,k} = t_\alpha \\ 0 & \text{if } T_{1,k} < t_\alpha \end{cases}$$

and

$$E_{G_{u_{1,n},(\theta_0,\eta_T)}}[\psi_\alpha(X_1^k)] = \alpha$$

i.e. $\alpha := \int \mathbb{1}_{t_{1,k}>t_\alpha} g_{u_{1,n},(\theta_0,\eta_T)}\left(x_1^k\right) dx_1 ... dx_k$. Its power under a simple hypothesis $\theta_T = \theta$ is defined through

$$\beta_{\psi_\alpha}(\theta|u_n) = E_{G_{u_{1,n},(\theta_0,\theta)}}[\psi_\alpha(\mathbf{X}_1^k)].$$

By quasi-sufficiency of $\mathbf{U}_{1,n}$ with respect to η any value can be inserted in $g_{u_{1,n},(\theta_0,\eta_T)}$ in place of η_T.

Recall that the parametric bootstrap produces samples from a parametric model which is fitted to the data, often through maximum likelihood. In the present setting, the parameter θ is set to θ_0 and the nuisance parameter η is replaced by its estimator $\widehat{\eta}_{\theta_0}$ which is the MLE of η_T when the parameter θ is fixed at the value θ_0 defining $H0$. Comparing their exact conditional MC tests with parametric bootstrap ones for Gamma distributions, Lockhart et al [70] conclude that no significant difference can be noticed in

terms of level or in terms of power. We proceed in the same vein, comparing conditional sampling MC tests with the parametric bootstrap ones, obtaining again similar results when the nuisance parameter is estimated accurately. However the results are somehow different when the nuisance parameter cannot be estimated accurately, which may occur in various cases.

3.4.3 Unimodal Likelihood: testing the coefficients of a Gamma distribution

Let $\mathbf{X}_1^n$ be an i.i.d. sample of random variables with Gamma distribution f_{r_T,θ_T} . As r and θ vary this distribution is a two parameter exponential family. The statistics $T_{1,n} := \log \mathbf{X}_1 + ... + \log \mathbf{X}_n$ is sufficient for r and $\mathbf{U}_{1,n} := \mathbf{X}_1 + ... + \mathbf{X}_n$ is sufficient for the parameter θ.

MC conditional test with $H_0 : r_T = r_0$ Denote $u_{1,n} = \sum_{i=1}^n X_i$ and $\widehat{\theta}_{r_0}$ the MLE of θ_T. Calculate for $l \in \{2, L\}$

$$t_l := \sum_{i=0}^{k} \log\left(Y_i(l)\right).$$

where the Y_i' are a sample from $g_{u_{1,n},\left(r_0,\widehat{\theta}_{r_0}\right)}$.

Consider the corresponding parametric bootstrap procedure for the same test, namely simulate $Z_i(l)$, $2 \leq l \leq L$ and $0 \leq i \leq k$ with distribution $f_{r_0,\widehat{\theta}_{r_0}}$; denote

$$s_l := \sum_{i=0}^{k} \log\left(Z_i(l)\right).$$

In this example simulation shows that for any α the Mth largest value of the sample $(t, t_2, ..., t_L)$ is very close to the corresponding empirical M/L-quantile of s_l's. Hence Monte Carlo tests through parametric bootstrap and conditional compete equally. Also in terms of power, irrespectively in terms of α and in terms of alternatives (close to $H0$), the two methods seem to be equivalent.

MC conditional test with $H_0 : \theta_T = \theta_0$ Denote $t_{1,n} = \sum_{i=1}^n \log(X_i)$ and $\widehat{r}_{\theta_0}$ the MLE of r_T. Calculate for $l \in \{2, L\}$

$$t_l := \sum_{i=0}^{k} Y_i(l)$$

where the Y_i' are a sample from $g_{u_{1,n},\left(\widehat{r}_{\theta_0},\theta_0\right)}$ and, as above define accordingly

$$s_l := \sum_{i=0}^{k} \log\left(Z_i(l)\right)$$

where the $Z_i(l)$'s are simulated under $f_{\widehat{r}_{\theta_0},\theta_0}$.

As above, parametric bootstrap and conditional sampling yield equivalent Monte Carlo tests in terms of power function under alternatives close to $H0$.

In the two cases studied above the value of k has been obtained through the rule exposed in section 3.2 of Chapter 1.

3.4.4 Bimodal likelihood: testing the mean of a normal distribution in dimension 2

In contrast with the above mentionned examples, the following case study shows that estimation through the unconditional likelihood may fail to provide consistent estimators when the likelihood surface has multiple critical points.

Sundberg [91] proposes four examples that allow likelihood multimodality. Two of them can also be found in [39] and [40], and in [8], Ch 2. We consider the "Normal parabola" model which is a curved (2, 1) family (see Example 2.35 in [8], Ch 2 and [90], Section 15.10). Two independent Gaussian variates have unknown means and known variances; their means are related by a parabolic relationship.

Let $\mathbf{X}$ and $\mathbf{Y}$ be two independent gaussian r.v.'s with same variance σ_T^2 with expectation ψ_T and ψ_T^2. In the present example $\sigma_T^2 = 1$ and $\psi_T = 2$.

Let $(\mathbf{x}_i, \mathbf{y}_i)$, $1 \leq i \leq n$ be i.i.d. realizations of $(\mathbf{X}_i, \mathbf{Y}_i)$.

The parameter of interest is σ^2 whislt the nuisance parameters is ψ. Derivation of the likelihood function of the observed sample with respect to ψ yields the following equation

$$(u_{1,n} - \psi) + 2\psi \left(v_{1,n} - \psi^2\right) = 0$$

with $u_{1,n} := \mathbf{x}_1 + ... + \mathbf{x}_n$ and $v_{1,n} := \mathbf{y}_1 + ... + \mathbf{y}_n$. Define accordingly $\mathbf{U}_{1,n}$ and $\mathbf{V}_{1,n}$.The following table shows that the likelihood function is bimodal in ψ.

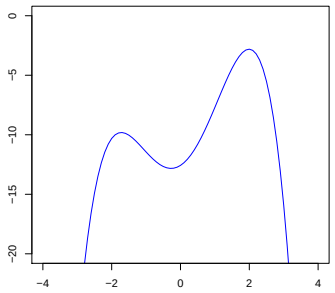

Figure 3.6: Bimodal likelihood in ψ.

Estimation of the nuisance parameter ψ is performed through the standard Newton Raphson method. The Newton-Raphson optimizer of the likelihood function converges to the true value when the initial value is larger than 1 and fails to converge to $\psi_T = 2$ otherwise. Hencefore the parametric bootstrap estimation of the likelihood function of the sample based on this preliminary estimate of the nuisance parameter may lead to erroneous estimates of the parameter of interest. Indeed according to the initial value we obtained estimators of ψ_T close to 2 or to -2. When the estimator of the nuisance parameter is close to its true value 2 then parametric bootstrap yields Monte Carlo tests with power close to 1 for any α and any alternative close to $H0$. At the contrary when this estimate is close to the second maximizer of the likelihood (i.e. close to -2) then the resulting Monte Carlo test based on parametric bootstrap has power close to 0 irrespectively of the value of α and of the alternative, when close to $H0$. In contrast with these results, Monte Carlo tests based on conditional sampling provide powers close to 1 for any α; we have considered alternatives close to $H0$. This result is of course a consequence of quasi sufficiency of the statistics $(\mathbf{U}_{1,n}, \mathbf{V}_{1,n})$ for the parameter ψ of the distribution of the sample $(\mathbf{x}_i, \mathbf{y}_i)_{i=1,...,n}$; see next paragraph for a discussion of this point.

3.4.5 Estimation through conditional likelihood

Considering model (3.9) we intend to perform an estimation of θ_T irrespectively upon the value of η_T. Denote $\widehat{\eta}_\theta$ the MLE of η_T when θ holds; the model $p_{\mathbf{X},(\theta,\widehat{\eta}_\theta)}(x)$ is a one parameter model which is fitted to the data for any peculiar choice of θ.The optimizer in θ of the resulting likelihood function is the global MLE. Properties of the resulting estimators strongly rely on the constistency properties of $\widehat{\eta}_\theta$ at any given θ.

Consider the consequence of Theorem 38. Condition on the value of the sufficient statistics $\mathbf{U}_{1,n}$, and consider the conditional likelihood of the observed subsample $\mathbf{x}_1^k$ under parameter $(\theta, \widehat{\eta}_\theta)$; recall that $\mathbf{x}_1^k$ is generated under (θ_T, η_T) . By Theorem 38 this likelihood is approximated by $g_{u_{1,n},(\theta,\widehat{\eta}_\theta)}\left(\mathbf{x}_1^k\right)$ with a small relative error. Conditioned likelihood estimation is performed optimizing $g_{u_{1,n},(\theta,\widehat{\eta}_\theta)}\left(\mathbf{x}_1^k\right)$ upon θ. Any value of the nuisance parameter η can be used, as seen in Section 3.1.

In most cases, as the normal, gamma or inverse-gaussian, estimations through the unconditional likelihood or through conditional likelihood give a consistent estimator.

We consider the example of the Bimodal likelihood from the above subsection, inheriting of the notation and explore the behaviour of the proxy of the conditional likelihood of the sample $(\mathbf{x}_i, \mathbf{y}_i)$, $1 \leq i \leq n$ when conditioning on $u_{1,n}$ and $v_{1,n}$, as a function of σ^2. The likelihood writes

$$\begin{aligned} &L\left(\sigma^2 \middle| u_{1,n}, v_{1,n}\right) \\ &= p_{u_{1,n}\sigma^2}\left(\mathbf{x}_1^n\right) p_{v_{1,n}\sigma^2}\left(\mathbf{y}_1^n\right) \end{aligned}$$

where we have used the independence of the r.v.'s $\mathbf{X}_i$'s and $\mathbf{Y}_i$'s.

Applying Theorem 1 to the above expression it appears that ψ cancels in the resulting density $g_{u_{1,n},\sigma^2}$ and $g_{v_{1,n},\sigma^2}$. This proves that the proxy of the conditional likelihood provides consistent estimation of σ_T^2 as shown on Figures 3.7 and 3.8 (see the solid lines).

On Figure 3.7, the dot line is the likelihood function

$$L\left(\sigma^2\right) := \sum_{i=1}^{n} \log p_{\mathbf{X},\left(\sigma^2,\widehat{\psi}_{\sigma^2}\right)}(\mathbf{x}_i)$$

where $\widehat{\psi}_{\sigma^2}$ is a consistent estimator of the nuisance parameter; the resulting maximizer in the variable σ^2 is close to $\sigma_T^2 = 1$. At the opposite in Figure 3.8 an inconsistent preliminary estimator of ψ_T obtained through a bad tuning of the initial point in the Newton-Raphson procedure leads to unconsistency in the estimation of σ_T^2, the resulting likelihood function being unbounded.

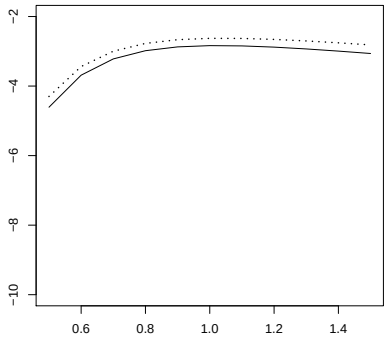

Figure 3.7: Proxy of the conditional likelihood (solid line) along with the classical likelihood (dotted line) as function of σ^2 for $n = 100$ and $k = 99$ in the case where a good initial point in Newton-Raphson procedure is chosen.

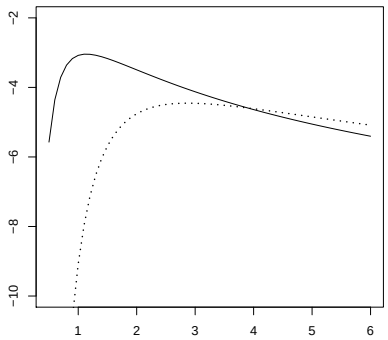

Figure 3.8: Proxy of the conditional likelihood (solid line) along with the classical likelihood (dotted line) as function of σ^2 for $n = 100$ and $k = 99$ in the case where a bad initial point in Newton-Raphson procedure is chosen.

Chapitre 4

Extension au cas d-dimensionnel

4.1 Introduction

L'objectif de ce chapitre est de généraliser les résultats obtenus au chapitre 1 dans le cas où $\mathbf{X}_1^n := (\mathbf{X}_1, ..., \mathbf{X}_n)$ est une suite de vecteurs aléatoires i.i.d. de $\mathbb{R}^d$ d'espérance μ et de matrice de variance-covariance Σ, copies d'un vecteur aléatoire $\mathbf{X} :=^t (\mathbf{X}^{(1)}, ..., \mathbf{X}^{(d)})$ de densité p et de loi P. On munit $\mathbb{R}^d$ de son produit scalaire canonique et de sa norme associée. Nous voulons obtenir une approximation de la densité du vecteur $\mathbf{X}_1^k$ sur $\left(\mathbb{R}^d\right)^k$, quand l'événement conditionnant s'écrit

$$(\mathbf{S}_{1,n} := \mathbf{X}_1 + ... + \mathbf{X}_n = na_n) \tag{4.1}$$

avec $a_n :=^t (a_n^{(1)}, ..., a_n^{(d)})$ une suite convergente et $k := k_n$ tel que

$$\lim_{n\to\infty} \frac{k}{n} = 1 \tag{K1}$$

$$\lim_{n\to\infty} (n - k) = +\infty \tag{K2}$$

La section 4.2 expose les notations et les hypothèses générales. La section 4.3 expose le théorème principal dans le cas où le conditionnement s'écrit (4.1) et les éléments principals de sa démonstration. La section 4.4 généralise le résultat précédent dans le cas où le conditionnement s'écrit $(\mathbf{U}_{1,n} := u(\mathbf{X}_1) + ... + u(\mathbf{X}_n) = u_{1,n})$ pour une fonction mesurable u de $\mathbb{R}^d$ dans $\mathbb{R}^s$ avec d et s des entiers plus grand ou égaux que 1. La section 4.5 propose une méthode effective pour déterminer la valeur adéquate de k pour une erreur donnée. La section 4.6 illustre le théorème principal avec des exemples de trajectoires typiques. La section 4.7 est une ébauche portant sur la relation entre les familles exponentielles courbes et les approximations dévéloppées dans cette thèse .Les lemmes techniques utilisés dans la démonstration ainsi que les détails de celle-ci sont laissés dans l'annexe.

4.2 Hypothèses et Notations

4.2.1 Notations générales

Dans cette section, le conditionnement s'écrit

$$(\mathbf{S}_{1,n} := \mathbf{X}_1 + ... + \mathbf{X}_n = na_n)\,.$$

On suppose que la fonction caractéristique de $\mathbf{X}$ à la puissance r pour $r \geq 1$ est intégrable sur $\mathbb{R}^d$. Cette hypothèse sur la loi de $\mathbf{X}$ nous permettra d'effectuer les développements d'Edgeworth nécéssaires à l'obtention de l'approximation voulue.

On suppose, de plus, que $\mathbf{X}$ satisfait la condition de Cramer, c.a.d. $\mathbf{X}$ a une fonction génératrice des moments finie dans un voisinage d'intérieur non-vide de $\underline{0}$

$$\Phi(t) := E[\exp < t, \mathbf{X} >] < \infty, \quad t \in V(\underline{0}) \subset \mathbb{R}^d \tag{4.2}$$

où $V(\underline{0})$ est un voisinage de $\underline{0}$, le vecteur de $\mathbb{R}^d$ dont toutes les coordonnées sont nulles.

On note :

$$m(t) := \nabla \log(\Phi(t)), \quad t \in V(\underline{0}) \subset \mathbb{R}^d \tag{4.3}$$

et

$$\kappa(t) := {}^t\nabla\nabla m(t), \quad t \in V(\underline{0}) \subset \mathbb{R}^d \tag{4.4}$$

Notons $\mathfrak{N} := \{t \in \mathbb{R}^d : \Phi(t) < \infty\}$ et $\mathcal{H}$ l'enveloppe convexe du support de P. On suppose dans la suite que

$$m \quad \text{est un homeomorphisme de int } \mathfrak{N} \text{ dans int } \mathcal{H} \tag{4.5}$$

Dans le cas des familles exponentielles essentiellement régulières (Chapitre 3), cette condition est valide. Les valeurs $m(t_\alpha) := {}^t\nabla \log \Phi(t_\alpha)$ et $\kappa(t_\alpha) := {}^t\nabla\nabla \log \Phi(t_\alpha)$ sont respectivement l'espérance et la matrice de variance-covariance de la densité tiltée définie par :

$$\pi^\alpha(x) := \frac{\exp < t, x >}{\Phi(t)} p(x) \tag{4.6}$$

où t est l'unique solution de $m(t) = \alpha$ quand α appartient à $\mathcal{H}$.

La densité conditionnelle de $\mathbf{X}_1^k$ sur $\left(\mathbb{R}^d\right)^k$ conditionné par $(\mathbf{S}_{1,n} = na_n)$ sera noté p_{na_n}. Pour un vecteur aléatoire générique $\mathbf{Z}$ avec densité p, on note $p\,(\mathbf{Z} = z)$ la valeur de p au point z. Ainsi, on notera

$$p_{na_n}(Y_1^k) := p(\mathbf{X}_1^k = Y_1^k | \mathbf{S}_{1,n} = na_n) \tag{4.7}$$

La densité normale d-dimensionnelle ayant pour espérance μ et pour matrice de variance-covariance κ au point x sera noté $\mathfrak{n}_d(x; \mu, \kappa)$ et $\mathfrak{n}_d(x; 0, I_d)$ sera noté $\mathfrak{n}_d(x)$.

4.2.2 Cumulants

Les cumulants sont un ensemble de quantités qui apportent une alternative aux traditionnels moments. Les cumulants déterminent les moments, et réciproquement (cf. section 4.2.3 pour des notations spécifiques). Pour déterminer les moments de $\mathbf{X}$ dans le cas réel, on écrit le développement en série entière de (4.2)

$$\Phi(t) = 1 + \sum_{i=0}^{\infty} \mu_i \frac{t^i}{i!}$$

où μ_i est le i-iéme moment de $\mathbf{X}$. Pour obtenir les cumulants, on écrit la fonction génératrice des cumulants comme le logarithme de la fonction génératrice des moments, et de la même façon, on effectue un développement en série entière de cette fonction

$$h(t) := \sum_{i=0}^{\infty} \kappa_i \frac{t^i}{i!}$$

où κ_i est le i-iéme cumulant de $\mathbf{X}$. On obtient alors :

$$\begin{aligned} h(t) &= -\sum_{i=1}^{\infty} \frac{1}{n} \left(-\sum_{j=0}^{\infty} \mu_j \frac{t^j}{j!} \right)^n \\ &= \mu_1 t + \left(\mu_2 - \mu_1^2 \right) \frac{t^2}{2} + \left(\mu_3 - 3\mu_2\mu_1 + 2\mu_1^3 \right) \frac{t^3}{3!} + ... \end{aligned}$$

Par analogie, on récupère ainsi la forme de tous les cumulants de $\mathbf{X}$. A titre d'exemple, le premier cumulant est l'espérance, le second et le troisième cumulant sont respectivement le deuxième et le troisième moment centré. Pour finir, dans le cas réel, une formule simple est proposé pour déterminer le n-ième cumulants en fonction des moments et des cumulants précédents :

$$\kappa_n = \mu_n - \sum_{j=1}^{n-1} C_{n-1}^{j-1} \kappa_j \mu_{n-j}.$$

Les moments μ_n ne doivent pas être confondus avec les moments centrés. Pour obtenir les moments centrés en fonction des cumulants, il suffit d'ignorer tous les termes dans lesquels on trouve le facteur κ_1.

Pour $d \geq 1$, en effectuant un travail identique, on obtient alors la formule suivante pour le cumulant joint de $\left(\mathbf{X}^{(i_1)} ... \mathbf{X}^{(i_\nu)} \right)$:

$$K[\mathbf{X}^{(i_1)} ... \mathbf{X}^{(i_\nu)}] = \sum_{\alpha=1}^{|I|} (-1)^{\alpha-1} (\alpha - 1)! \sum_{I/\alpha} \kappa_1^{I_1} ... \kappa_1^{I_\alpha}.$$

Les sommes internes sont prises sur toutes les partitions $I_1, ..., I_\alpha$ de $I = \{i_1 ... i_\nu\}$ en α blocs. On peut ainsi calculer les cumulants joints en fonction des moments correspondants. Par exemple, en prenant $d = 3$, on a :

$$\begin{aligned} K(\mathbf{X}^{(i_1)}\mathbf{X}^{(i_2)}\mathbf{X}^{(i_3)}) &= E(\mathbf{X}^{(i_1)}\mathbf{X}^{(i_2)}\mathbf{X}^{(i_3)}) - E(\mathbf{X}^{(i_1)}\mathbf{X}^{(i_2)})E(\mathbf{X}^{(i_3)}) - E(\mathbf{X}^{(i_1)}\mathbf{X}^{(i_3)})E(\mathbf{X}^{(i_2)}) \\ &\quad - E(\mathbf{X}^{(i_2)}\mathbf{X}^{(i_3)})E(\mathbf{X}^{(i_1)}) + 2E(\mathbf{X}^{(i_1)})E(\mathbf{X}^{(i_2)})E(\mathbf{X}^{(i_3)}). \end{aligned}$$

4.2.3 Notations spécifiques

Dans cette partie, les notations introduites par Barndorff-Nielsen et Cox (1990) [7] (p.130 et suivantes) seront principalement utilisées. Cependant, certaines notations ont été adaptées et peuvent légérement différer.

Les indices $j, k, l, ...$ avec ou sans suffixes sont des entiers compris entre 1 et d. De plus, nous adopterons la convention de sommation d'Einstein. En d'autres termes, quand un indice apparaît en haut et en bas dans une même expression, alors la sommation s'effectue sur tout l'ensemble des valeurs prises par cet indice. Par exemple, $a^j b_j = \sum_{j=1}^{d} a^j b_j$ ou encore $a^{jlm} b_{jr} = \sum_{j=1}^{d} a^{jlm} b_{jr}$.

Moments et cumulants

Notons $\kappa^{j,l}$ le terme générique de la matrice de variance-covariance. Le terme générique de la matrice inverse sera noté $\kappa_{j,l}$. De manière générale, le moment joint et le cumulant joint de $(X^{(i_1)}...X^{(i_\nu)})$ seront notés, respectivement

$$\kappa^{i_1...i_\nu} = E[X^{(i_1)}...X^{(i_\nu)}] \tag{4.8}$$
$$\kappa^{i_1,...,i_\nu} = K[X^{(i_1)}...X^{(i_\nu)}] \tag{4.9}$$

Dans le cas des variables centrées, $\kappa^{jl} = \kappa^{j,l}$ et $\kappa^{jlm} = \kappa^{j,l,m}$. Néanmoins, ces égalités ne se généralisent pas aux moments supérieurs. Ainsi,

$$\kappa^{j,l,m,r} = \kappa^{jlmr} - \kappa^{j,l}\kappa^{m,r}$$

En général, $[n]$ après un symbole indiquera une somme de n termes déterminée par une permutation adéquate des indices. Par exemple,

$$\kappa^{j,l}x_m[3] = \kappa^{j,l}x_m + \kappa^{j,m}x_l + \kappa^{m,l}x_j.$$

Pour finir, en utilisant la convention de sommation, notons $\kappa_{i_1,...,i_\nu}$ la quantité définie par :

$$\kappa_{i_1,...,i_\nu} = \kappa_{i_1,j_1}...\kappa_{i_\nu,j_\nu}\kappa^{j_1,...,j_\nu} \tag{4.10}$$

où $\kappa_{j,l}$ est le terme générique de la matrice inverse.

Index notation

On donne un exemple particulièrement intéressant de l'utilisation de la notation index (p.136). La convention de sommation explicitée plus haut permet d'écrire la formule de Taylor pour une fonction f de d variables de manière compacte :

$$f(x+t) = \sum_{\nu=0}^{\infty} \frac{1}{\nu!} t^{(i_1)}...t^{(i_\nu)} f_{/i_1...i_\nu}(x)$$

où $x = \left(x^{(1)}, ..., x^{(d)}\right)$, $t = \left(t^{(1)}, ..., t^{(d)}\right)$ et $f_{/i_1...i_\nu}$ est la ν-ième dérivée de f par rapport à $x^{(i_1)}, ..., x^{(i_\nu)}$ et, par la convention de sommation, la somme est possible sur tous les choix possibles de $(i_1, ..., i_\nu)$. Grace aux notations index, on peut proposer une version encore plus compacte :

$$f(x+t) = \sum_{|I|}^{\infty} \frac{1}{|I|!} t^I f_{/I} \tag{4.11}$$

où I est un ensemble index $\{i_1, ..., i_\nu\}$ and $|I| = \nu$ quand $t^I = \left(t^{(i_1)}...t^{(i_\nu)}\right)$.

De la même façon, on peut réécrire les moments joints sous une forme plus compacte, pour un ensemble $I = \{i_1, ..., i_\nu\}$

$$\kappa^I := \kappa^{i_1...i_\nu} \tag{4.12}$$

Les relations exlog

Pour simplifier encore les notations qui peuvent être très lourdes, on propose aussi d'écrire :

$$\kappa_1^I := \kappa_1^{i_1...i_\nu} = E[X^{(i_1)}...X^{(i_\nu)}] \tag{4.13}$$

$$\kappa_0^I := \kappa_0^{i_1,...,i_\nu} = K[X^{(i_1)}...X^{(i_\nu)}] \tag{4.14}$$

Cette notation nous permet alors de proposer dans le cas particulier des notations exlog (p.140), des formules pour déterminer les moments en termes de cumulants et inversement. La formule qui exprime les moments en termes de cumulants et celle qui exprime les cumulants en terme de moments sont, respectivement,

$$\kappa_1^I = \sum_{\alpha=1}^{|I|} \sum_{I/\alpha} \kappa_0^{I_1}...\kappa_0^{I_\alpha},$$

et

$$\kappa_0^I = \sum_{\alpha=1}^{|I|} (-1)^{\alpha-1}\,(\alpha-1)! \sum_{I/\alpha} \kappa_1^{I_1}...\kappa_1^{I_\alpha}.$$

Les sommes internes sont prises sur toutes les partitions $I_1, ..., I_\alpha$ de $I = \{i_1...i_\nu\}$ en α blocs.

Polynômes tensoriels d'Hermite

Dans la démonstration principale, un développement asymptotique d'Edgeworth qui utilise les polynômes d'Hermite sera utilisé. La formule qui permet de déterminer l'expression des polynômes standards (p.18) d'Hermite H_k associé la densité normale standard est la suivante :

$$\mathfrak{n}_1(x)H_k(x) = (-1)^k \frac{d^k}{dx^k}\mathfrak{n}_1(x) \tag{4.15}$$

Dans $\mathbb{R}^d$, de manière équivalente, une formule pour déterminer les polynômes tensoriels d'Hermite $h_{i_1...i_k}$ (p.150) associés à la densité normale standard d-dimensionnelle peut-étre obtenue :

$$\mathfrak{n}_d(x)h_{i_1...i_k}(x) = (-1)^k \partial_{i_1}...\partial_{i_1}\mathfrak{n}_d(x) \tag{4.16}$$

où $\partial_{i_1} = \frac{\partial}{\partial x^{(i_1)}}$ et $x^{(i_1)}$ la i_1-ième coordonnée de x.

Notons à présent : $x_j = \kappa_{j,l}x^{(l)}$.

Ainsi les polynômes h utilisés dans le développement d'Edgeworth multidimensionnel sont :

$$h_{jlm}(x) := x_j x_l x_m - \kappa_{j,l}x_m[3] \tag{4.17}$$

$$h_{jlmq}(x) := x_j x_l x_m x_q - \kappa_{j,l}x_m x_q[6] + \kappa_{j,l}\kappa_{m,q}[3] \tag{4.18}$$

$$h_{jlmqrs}(x) := x_j x_l x_m x_q x_r x_s - \kappa_{j,l} x_m x_q x_r x_s[15] + \kappa_{j,l}\kappa_{m,q} x_r x_s[45] - \kappa_{j,l}\kappa_{m,q}\kappa_{r,s}[15] \tag{4.19}$$

Ils sont clairement l'extension naturelle des polynômes standard d'Hermite :

$$H_3(x) = x^3 - 3x$$
$$H_4(x) = x^4 - 6x^2 + 3$$
$$H_6(x) = x^6 - 15x^4 + 45x^2 - 15$$

Développements d'Egeworth

Le développement d'Edgeworth dans $\mathbb{R}^d$ pour les sommes aléatoires i.i.d. repose sur (4.16) pour déterminer les polynômes d'Hermite et sur le développement de Taylor de la fonction génératrice des moments. On obtient en utilisant la convention de sommation d'Einstein :

$$p_{\frac{\overline{\mathbf{S}_{1,n}}}{\sqrt{n}}}(x) := \mathfrak{n}_d(x)(1 + \frac{\kappa^{j,l,m}}{6\sqrt{n}} h_{jlm}(x) + \frac{\kappa^{j,l,m,q}}{24n} h_{jlmq}(x) + \frac{\kappa^{j,l,m}\kappa^{q,r,s}}{72n} h_{jlmqrs}(x)) \tag{4.20}$$
$$+ O(\tfrac{1}{n^{3/2}})$$

où $\overline{\mathbf{S}_{1,n}} = \Sigma^{-1/2}(\mathbf{S}_{1,n} - n\mu)$.

Notons

$$Q_3(x) := \frac{\kappa^{j,l,m}}{6} h_{jlm}(x) \tag{4.21}$$

et

$$Q_4(x) := \frac{\kappa^{j,l,m,q}}{24} h_{jlmq}(x) + \frac{\kappa^{j,l,m}\kappa^{q,r,s}}{72} h_{jlmqrs}(x). \tag{4.22}$$

Ainsi, (4.20) devient

$$p_{\frac{\overline{\mathbf{S}^n_1}}{\sqrt{n}}}(x) := \mathfrak{n}_d(x)(1 + \frac{1}{\sqrt{n}} Q_3(x) + \frac{1}{n} Q_4(x)) + O(\frac{1}{n^{3/2}}). \tag{4.23}$$

4.3 Marche aléatoire d-dimensionnelle conditionnée par sa somme.

Nous introduisons une suite ϵ_n strictement positive telle que

$$\lim_{n\to\infty} \epsilon_n^2 (n-k) = \infty \tag{E1}$$

$$\lim_{n\to\infty} \epsilon_n (\log n)^2 = 0 \tag{E2}$$

Nous montrerons que $\epsilon_n (\log n)^2$ est le niveau de précision du schéma d'approximation obtenue pour la densité condtionnelle étudiée.

Soit $a := a_n$ une suite convergente de $\mathbb{R}^d$. On définit une densité $g_{na}(y_1^k)$ sur $\left(\mathbb{R}^d\right)^k$ de la façon suivante. Notons :

$$g_0(y_1|y_0) := \pi^a(y_1) \tag{4.24}$$

avec y_0 arbitraire et π^a définit par (4.6). Pour $1 \leq i \leq k-1$, on définit $g(y_{i+1}|y_1^i)$ de manière récursive. Soit $t_i \in \mathbb{R}^d$ l'unique solution de l'équation :

$$m_i := m(t_i) = \frac{n}{n-1}\left(a - \frac{s_{1,i}}{n}\right) \tag{4.25}$$

où $s_{1,i} := y_1 + ... + y_i$.

Soit

$$\kappa_{(i,n)}^{j,l} := \frac{d^2}{dt^{(j)}dt^{(l)}} \left(\log E_{\pi^{m_i}} \exp < t, \mathbf{X} >\right)(0) \tag{4.26}$$

et

$$\kappa_{(i,n)}^{j,l,m} := \frac{d^3}{dt^{(j)}dt^{(l)}dt^{(m)}} \left(\log E_{\pi^{m_i}} \exp < t, \mathbf{X} >\right)(0)\,. \tag{4.27}$$

Alors

$$g(y_{i+1}|y_1^i) := C_i \mathfrak{n}_d(y_{i+1}; \beta\alpha + a, \beta) p(y_{i+1}) \tag{4.28}$$

et

$$\alpha := \left(t_i + \frac{\kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)}\right) \tag{4.29}$$

$$\beta := \kappa_{(i,n)}(n-i-1) \tag{4.30}$$

et γ définie par

$$\gamma := (\sum_{j=1}^{d} \kappa_{(i,n)}^{j,j,p})_{1 \leq p \leq d}. \tag{4.31}$$

et le facteur de normalisation C_i doit assurer à $g(y_{i+1}|y_1^i)$ d'être une densité.

On définit finalement

$$g_{na}(y_1^k) := g_0(y_1|y_0) \prod_{i=1}^{k-1} g(y_{i+1}|y_1^i) \tag{4.32}$$

Théorème 1. *On suppose (E1) et (E2). Soit Y_1^k un échantillon de loi P_{na}. Alors*

$$p(\boldsymbol{X}_1^k = Y_1^k | \mathbf{S}_{1,n} = na) = g_{na}(Y_1^k)(1 + o_{P_{na}}(1 + \epsilon_n(\log n)^2)) \tag{4.33}$$

Démonstration. La démonstration utilise les mêmes arguments que ceux utilisés pour démontrer le Théorème 3 du chapitre 1 : la formule de Bayes pour écrire $p(\mathbf{X}_1^k = Y_1^k | \mathbf{S}_{1,n} = na)$ comme un produit de k vraisemblances conditionnelles de termes individuelles de la trajectoire Y_1^k, et la formule d'invariance exposée dans le Lemme 48. Chaque terme du produit est approximé par un développement d'Edgeworth qui, à l'aide des Lemmes 50, 51 et 52 concluent la démonstration. Ces lemmes techniques permettant de contrôler les termes de reste dans les développements d'Edgeworth portent sur les propriétés des coordonnées du vecteur $\left(Y_1^{(j)}, ..., Y_k^{(j)}\right)$ pour tout j entre 1 et d. Par analogie avec le cas réel où les deux lemmes techniques permettent le contrôle du maximum des m_i et des Y_{i+1} pour tout i entre 0 et $k-1$, dans le cas d-dimensionnel, nous contrôlons, pour tout j entre

1 et d, le maximum des $m_i^{(j)}$ et des $Y_{i+1}^{(j)}$ pour tout i entre 0 et $k-1$. Les démonstrations sont ainsi identiques à celles considérées dans le Chapitre 1.

Notons : $S_{1,0} := 0$, $S_{1,1} = Y_1$ et $S_{1,i} = S_{1,i-1} + Y_i$. Alors

$$p(\mathbf{X}_1^k = Y_1^k | \mathbf{S}_{1,n} = na) =$$

$$p(\mathbf{X}_1 = Y_1 | \mathbf{S}_{1,n} = na) \prod_{i=1}^{k-1} p(\mathbf{X}_{i+1} = Y_{i+1} | \mathbf{X}_1^i = Y_1^i, \mathbf{S}_{1,n} = na)$$

$$= \prod_{i=0}^{k-1} p(\mathbf{X}_{i+1} = Y_{i+1} | \mathbf{S}_{i+1,n} = na - S_{1,i}).$$

Par le Lemme 48,

$$p(\mathbf{X}_{i+1} = Y_{i+1} | \mathbf{S}_{i+1}^n = na - S_{1,i})$$

$$= \pi^{m_i}(\mathbf{X}_{i+1} = Y_{i+1} | \mathbf{S}_{i+1,n} = na - S_{1,i})$$

$$= \pi^{m_i}(\mathbf{X}_{i+1} = Y_{i+1}) \frac{\pi^{m_i}(\mathbf{S}_{i+2,n} = na - S_{1,i+1})}{\pi^{m_i}(\mathbf{S}_{i+1,n} = na - S_{1,i})}$$

où la formule de Bayes et l'indépendance des $\mathbf{X}_j$ sous π^{m_i} sont utilisées. Une évaluation précise du (ou des) termes dominants est nécéssaire pour étudier le produit.

Sous la suite des densités π^{m_i}, les vecteurs aléatoires i.i.d. $\mathbf{X}_{i+1}, ..., \mathbf{X}_n$ définissent un tableau triangulaire qui satisfait un théorème central limite et un développement d'Edgeworth. Sous π^{m_i}, $\mathbf{X}_{i+1}$ a pour espérance m_i et pour matrice de variance-covariance $\kappa_{(i,n)}$.

En centrant et normalisant $\mathbf{S}_{i+1,n}$ et $\mathbf{S}_{i+2,n}$ et en effectuant un développement d'Edgeworth de leurs densités respectives, on obtient, comme prouvé, dans l'annexe sous (E1) et (E2),

$$p(\mathbf{X}_{i+1} = Y_{i+1} | \mathbf{S}_{i+1,n} = na - S_{1,i}) = \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \pi^{m_i}(\mathbf{X}_{i+1} = Y_{i+1}) \frac{N_i}{D_i} \tag{4.34}$$

où N_i et D_i sont définies par

$$N_i := \exp\{-\frac{{}^t(Y_{i+1} - a)\, \kappa_{(i,n)}^{-1}\, (Y_{i+1} - a)}{2(n-i-1)}\} A_i + O_{P_{na}}(\frac{1}{(n-i-1)^{3/2}}) \tag{4.35}$$

avec

$$A_i := 1 + \frac{{}^t(Y_{i+1} - a)\, \kappa_{(i,n)}^{-2} \gamma}{2(n-i-1)} + \frac{\delta^{(i,n)}}{n-i-1} + \frac{o_{P_{na}}(\epsilon_n(\log n))}{n-i-1} \tag{4.36}$$

et

$$D_i := 1 + \frac{\delta^{(i,n)}}{n-i} + O_{P_{na}}(\frac{1}{(n-i)^{3/2}}) \tag{4.37}$$

où l'expression de $\delta^{(i,n)}$ dépendant des cumulants se trouve dans l'Annexe.

Le terme $O_{P_{na}}(\frac{1}{(n-i-1)^{3/2}})$ dans (4.35) est uniforme en Y_{i+1}.

Les termes dans l'expression de la densité approximante viennent d'un développement du ratio entre les deux expressions ci-dessus. La composante gaussienne est explicite dans (4.35) tandis que le terme $\frac{{}^tY_{i+1}\kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)}$ est le terme dominant dans A_i. Le facteur de normalisation C_i dans $g(Y_{i+1}|Y_1^i)$ compense le terme $\frac{\sqrt{n-i}}{\Phi(t_i)\sqrt{n-i-1}}\exp\left(\frac{{}^ta\kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)}\right)$ où $\Phi(t_i)$ provient du terme $\pi^{m_i}\left(\mathbf{X}_{i+1}=Y_{i+1}\right)$. Le produit des termes restants permettent d'obtenir la vitesse d'approximation $1+o_{P_{na}}(1+\epsilon_n(\log n)^2)$. Les détails sont différés dans l'annexe. ■

Remark 41. *Quand les $\mathbf{X}_i$ sont i.i.d. de densité normale standard, le résultat (comme dans le cas réel) de l'approximation dans le théorème précédent est vrai pour $k=n-1$. En effet, on a $p(\mathbf{X}_1^{n-1}=x_1^{n-1}\Big|\,\mathbf{S}_{1,n}=na)=g_a\left(x_1^{n-1}\right)$ pour tout x_1^{n-1} dans $\left(\mathbb{R}^d\right)^{n-1}$.*

Remark 42. *La densité (4.28) est une légère modification de π^{m_i}. La modification qui permet de passer de π^{m_i} à g est*

1. *un léger décalage dans le paramètre de location de la densité normale dépendant à la fois de a à travers m_i, mais aussi de la forme de la densité de $\mathbf{X}$ à travers sa matrice de variance-covariance et ses cumulants joints.*
2. *un changement dans la variance qui permet aux grandes valeurs de $\mathbf{X}_{i+1}$ d'avoir un poids plus petit pour les grands i, afin que la distribution de $\mathbf{X}_{i+1}$ tendent à se concentrer autour de m_i quand i s'approche de k.*

Remark 43. *Les dévéloppements d'Edgeworth effectués dans la démonstration du théorème 4.33 sont valides sous les conditions énoncées dans le théorème suivant (Théorème 6.4, p.205, Barndorff-Nielsen et Cox (1990) [7]) pour a fixé et dans la remarque 6 du Chapitre 1 pour $a=a_n$ bornée.*

Theorem 44. *Soit $\mathbf{X}_1^n$ i.i.d. de densité p avec esperance μ et matrice de variance-covariance Σ. Si tous les moments de $\mathbf{X}$ d'ordre $r+2$ existent, pour un entier $r\geq 0$, alors, en notant $\overline{\mathbf{S}_{1,n}}=\Sigma^{-1/2}\left(\mathbf{S}_{1,n}-n\mu\right)$,*

$$p(x)=f_{\frac{\overline{\mathbf{S}_{1,n}}}{\sqrt{n}}}(x)+o(n^{-r/2}) \tag{4.38}$$

uniformement en x. Dans le cas où les moments d'ordre $r+3$ existent, l'expression de l'erreur dans (4.38) se réécrit $O(n^{-(r+1)/2})$ (pour la démonstration, cf. Bhattacharya et Rao (1986) [17], section 19).

Si on veut appliquer le Théorème 1 afin d'étendre la mise en place des procédures d'Importance Sampling (Chapitre 2) ou d'estimation (Chapitre 3), on doit obtenir un résultat inverse. En effet, en supposant que l'on a un à notre disposition un échantillon de vecteurs aléatoires Y_1^k simulé sous g_{na}, pouvons-nous dire que $g_{na}\left(Y_1^k\right)$ est une bonne approximation pour $p_{na}\left(Y_1^k\right)$? Pour répondre à cette question, le lemme suivant, énoncé de façon identique dans le chapitre 1, permet de répondre de manière favorable à cette question.

Soit $\mathfrak{R}_n$ et $\mathfrak{S}_n$ deux mesures de probabilités $\left(\mathbb{R}^d\right)^n$ avec densité, respectivement, $\mathfrak{r}_n$ et $\mathfrak{s}_n$.

Lemma 45. *On suppose qu'il existe une suite ε_n qui tends vers 0 quand n tends vers l'infini. Alors si on a*

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

quand n tends vers ∞. Alors

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

Démonstration. Voir la démonstration du Lemme 7, p. 24. ■

En utilisant le Théorème 1 et le Lemme 45, on obtient le théorème suivant :

Théorème 2. *On suppose (E1) et (E2). Soit Y_1^k un échantillon de loi G_{na}. Alors*

$$p(\boldsymbol{X}_1^k = Y_1^k | \mathbf{S}_{1,n} = na) = g_{na}(Y_1^k)(1 + o_{G_{na}}(1 + \epsilon_n(\log n)^2)) \tag{4.39}$$

4.4 Extension à un cas plus général de conditionnement

Dans le chapitre précédent, la densité conditionnelle a été approximée quand l'événement conditionnant s'écrivait $(\mathbf{S}_{1,n} = na_n)$. Pour pouvoir utiliser ce résultat dans un cadre pratique, ce résultat doit être étendu à des formes plus générales de conditionnement. Ici, $\mathbf{X}_1^n := (\mathbf{X}_1, ..., \mathbf{X}_n)$ est une suite de vecteurs aléatoires i.i.d., copies d'un vecteur aléatoire $\mathbf{X}$ dans $\mathbb{R}^d$, de densité noté $p_{\mathbf{X}}$ et u est une fonction mesurable définie de $\mathbb{R}^d$ dans $\mathbb{R}^s$, avec $d, s \geq 1$.

Notons

$$\mathbf{U}_{1,n} = \sum_{i=1}^{n} u\left(\mathbf{X}_i\right).$$

On suppose que $\mathbf{U} := u\left(\mathbf{X}\right)$ a une densité $p_{\mathbf{U}}$ (de loi $P_{\mathbf{U}}$) par rapport à la mesure de Lebesgue sur $\mathbb{R}^s$. On s'intéresse à présent à des événements conditionnants qui s'écrivent

$$(\mathbf{U}_{1,n} := u_{1,n}) \tag{4.40}$$

avec $u_{1,n}/n$ une suite convergente. On suppose que la fonction u est telle que la fonction caractéristique de $\mathbf{U} = u\left(\mathbf{X}\right)$ appartient à L^r pour $r \geq 1$.

4.4.1 Approximation de la densité

L'objectif de cette section est d'obtenir une approximation pour la densité des suites de vecteurs aléatoires $(\mathbf{X}_1, ..., \mathbf{X}_k)$ conditionnées par l'événement

$$(\mathbf{U}_{1,n} := u_{1,n}) \tag{4.41}$$

et $u_{1,n}/n$ est une suite convergente de $\mathbb{R}^s$.

On suppose que $\mathbf{U}$ satisfait la condition de Cramer,

$$\Phi_{\mathbf{U}}(t) := E[\exp < t, \mathbf{U} >] < \infty, \quad t \in V(\underline{0}) \subset \mathbb{R}^s.$$

On note :

$$m(t) := {}^t\nabla \log(\Phi_{\mathbf{U}}(t)), \quad t \in V(\underline{0}) \subset \mathbb{R}^s \tag{4.42}$$

et

$$\kappa(t) := {}^t\nabla\nabla \log(\Phi_{\mathbf{U}}(t)), \quad t \in V(\underline{0}) \subset \mathbb{R}^s. \tag{4.43}$$

Les valeurs $m(t)$ et $\kappa(t)$ sont respectivement l'esperance et la matrice de variance-covariance de la densité tiltée définie par :

$$\pi^{\alpha}_{\mathbf{U}}(u) := \frac{\exp < t, u >}{\Phi_{\mathbf{U}}(t)} p_{\mathbf{U}}(u) \tag{4.44}$$

où t est l'unique solution de $m(t) = \alpha$ et α appartient au support de $P_{\mathbf{U}}$.

On suppose, de plus, l'existence de t, définie ci-dessus, pour tout α dans le support convexe de $\mathbf{U}$. Les conditions sur $\Phi_{\mathbf{U}}(t)$ qui l'assurent sont énoncées dans [6], p132.

Nous introduisons une autre famille de densité

$$\pi^{\alpha}_{u}(x) := \frac{\exp < t, u(x) >}{\Phi_{\mathbf{U}}(t)} p_{\mathbf{X}}(x) \tag{4.45}$$

Par extension avec le cas centré réduit, notons $p_{u_{1,n}}$ la densité conditionnelle étudiée et $g_{u_{1,n}}$ son approximation définie de la façon suivante.

Notons $m_0 = u_{1,n}/n$ et

$$g_0(y_1|y_0) := \pi^{m_0}_{u}(y_1) \tag{4.46}$$

avec y_0 arbitraire et $\pi^{m_0}_{u}$ défini par (4.45).

Pour $1 \leq i \leq k-1$, on définit $g(y_{i+1}|y_1^i)$ de manière récursive. On définit alors t_i comme l'unique solution de

$$m(t_i) = m_i := \frac{u_{1,n} - u_{1,i}}{n-i} \tag{4.47}$$

où $u_{1,i} = u(y_1) + ... + u(y_i)$.

Soit

$$\kappa^{j,l}_{(i,n)} := \frac{d^2}{dt^{(j)}dt^{(l)}} \left(\log E_{\pi^{m_i}_{\mathbf{U}}} \exp < t, \mathbf{U} > \right)(0) \tag{4.48}$$

et

$$\kappa^{j,l,m}_{(i,n)} := \frac{d^3}{dt^{(j)}dt^{(l)}dt^{(m)}} \left(\log E_{\pi^{m_i}_{\mathbf{U}}} \exp < t, \mathbf{U} > \right)(0)\,. \tag{4.49}$$

pour j, l et m dans $\{1, ..., s\}$

Notons

$$g(y_{i+1}|y_1^i) := C_i \mathfrak{n}_d \left(u(y_{i+1}); \beta\alpha + m_0, \beta \right) p_{\mathbf{X}}(y_{i+1}) \tag{4.50}$$

où

$$\alpha := \left(t_i + \frac{\kappa^{-2}_{(i,n)}\gamma}{2(n-i-1)} \right) \tag{4.51}$$

$$\beta := \kappa_{(i,n)}(n-i-1) \tag{4.52}$$

et γ définie par

$$\gamma := \left(\sum_{j=1}^{s} \kappa^{j,j,p}_{(i,n)} \right)_{1 \leq p \leq s} \tag{4.53}$$

et C_i est un facteur de normalisation.

Ainsi

$$g_{u_{1,n}}(y_1^k) := g_0(y_1|y_0) \prod_{i=1}^{k-1} g(y_{i+1}|y_1^i) \tag{4.54}$$

Théorème 3. *On suppose (E1) et (E2).*
– *Soit Y_1^k un échantillon de loi $P_{u_{1,n}}$. Alors*

$$p\left(\boldsymbol{X}_1^k = Y_1^k | \mathbf{U}_{1,n} = u_{1,n}\right) = g_{u_{1,n}}(Y_1^k)(1 + o_{P_{u_{1,n}}}(1 + \epsilon_n(\log n)^2)) \tag{4.55}$$

– *Soit Y_1^k un échantillon de loi $G_{u_{1,n}}$. Alors*

$$p\left(\boldsymbol{X}_1^k = Y_1^k | \mathbf{U}_{1,n} = u_{1,n}\right) = g_{u_{1,n}}(Y_1^k)(1 + o_{G_{u_{1,n}}}(1 + \epsilon_n(\log n)^2)) \tag{4.56}$$

Démonstration. On présente uniquement la première partie de la démonstration de 4.55 qui suit rapidement le même chemin que la démonstration du théorème principal de la section précédente.

Notons : $U_{i,j} := u(Y_i) + ... + u(Y_j)$.

Evaluons :

$$p\left(\mathbf{X}_{i+1} = Y_{i+1} | \mathbf{U}_{i+1,n} = u_{1,n} - U_{1,i}\right) = p_{\mathbf{X}}\left(\mathbf{X}_{i+1} = Y_{i+1}\right) \frac{p\left(\mathbf{U}_{i+2,n} = u_{1,n} - U_{1,i+1}\right)}{p\left(\mathbf{U}_{i+1,n} = u_{1,n} - U_{1,i}\right)}.$$

En multipliant et divisant par $p_{\mathbf{U}}\left(\mathbf{U}_{i+1} = u(Y_{i+1})\right)$, on peut utiliser l'invariance par tilting sous $\pi_{\mathbf{U}}^{m_i}$. Ainsi,

$$p\left(\mathbf{X}_{i+1} = Y_{i+1} | \mathbf{U}_{i+1,n} = u_{1,n} - U_{1,i}\right)$$

$$= \frac{p_{\mathbf{X}}\left(\mathbf{X}_{i+1} = Y_{i+1}\right)}{p_{\mathbf{U}}\left(\mathbf{U}_{i+1} = u(Y_{i+1})\right)} \pi_{\mathbf{U}}^{m_i}\left(\mathbf{U}_{i+1} = u(Y_{i+1})\right) \frac{\pi_{\mathbf{U}}^{m_i}\left(\mathbf{U}_{i+2,n} = u_{1,n} - U_{1,i+1}\right)}{\pi_{\mathbf{U}}^{m_i}\left(\mathbf{U}_{i+1,n} = u_{1,n} - U_{1,i}\right)}$$

On procède ensuite à un développement d'Edgeworth en suivant exactement les étapes de la démonstration du Théorème 1. La démonstration de (4.56) suit directement en utilisant le Lemme 45.

■

4.5 Comment choisir k ?

Dans le chapitre 1, nous avons proposé une règle effective du calcul du paramètre k pour un niveau de précision donné. La fabrication d'une règle dans le cas d-dimensionnelle est assez facile. En effet, la règle du Chapitre 1 repose sur un développement asymptotique explicité dans le Lemme 14, Chapitre 1, p.30 (voir [59], Chap.6, p.144, Formule (6.1.2)). Le Lemme ci-dessous en est sa version d-dimensionnelle. Ainsi, la règle obtenue dans cette section est une simple extension du cas réel. Les quantités utilisées dans cette section ont été définies dans la section 29 du chapitre 1 (ainsi que la discussion sur ces définitions).

On rappelle uniquement la définition des trois quantités mesurant la précision de l'approximation :

$$ERE(k) := E_{G_{u_{1,n}}} 1_{D_k}\left(Y_1^k\right) \frac{p_{u_{1,n}}\left(Y_1^k\right) - g_{u_{1,n}}\left(Y_1^k\right)}{p_{u_{1,n}}\left(Y_1^k\right)} \tag{4.57}$$

$$VRE(k) := Var_{G_{u_{1,n}}} 1_{D_k}\left(Y_1^k\right) \frac{p_{u_{1,n}}\left(Y_1^k\right) - g_{u_{1,n}}\left(Y_1^k\right)}{p_{u_{1,n}}\left(Y_1^k\right)} \tag{4.58}$$

et

$$CI(k) := \left[ERE(k) - 2\sqrt{VRE(k)}, ERE(k) + 2\sqrt{VRE(k)} \right]. \tag{4.59}$$

On écrit

$$\begin{aligned} VRE(k)^2 &= E_{P_{\mathbf{X}}} \left(\frac{g_{u_{1,n}}^3 \left(Y_1^k \right)}{p_{u_{1,n}} \left(Y_1^k \right)^2 p_{\mathbf{X}} \left(Y_1^k \right)} \right) \\ &\quad - E_{P_{\mathbf{X}}} \left(\frac{g_{u_{1,n}}^2 \left(Y_1^k \right)}{p_{u_{1,n}} \left(Y_1^k \right) p_{\mathbf{X}} \left(Y_1^k \right)} \right)^2 \\ &=: A - B^2. \end{aligned}$$

Par la formule de Bayes,

$$p_{u_{1,n}} \left(Y_1^k \right) = p_{\mathbf{X}} \left(Y_1^k \right) \frac{np \left(\mathbf{U}_{k+1,n}/(n-k) = (m_k) \right)}{(n-k)\, p \left(\mathbf{U}_{1,n}/n = (m_0) \right)}. \tag{4.60}$$

avec $m_k = m(t_k)$ défini dans (4.25) et $m_0 = m(t) = u_{1,n}/n$.

Lemma 46. *([59], Chap.6, p.144, Formule (6.1.2)) Soit $\mathbf{U}_1, ..., \mathbf{U}_n$ un échantillon de variables aléatoires i.i.d. de densité $p_{\mathbf{U}}$ sur $\mathbb{R}^d$ telle que la fonction génératrice des moments de $\mathbf{U}$ existe. Alors avec $\nabla \left(\log \Phi_{\mathbf{U}} \right)(t) = u$ et $\Sigma(t) :=^t \nabla\nabla \left(\log \Phi_{\mathbf{U}} \right)(t)$,*

$$p_{\mathbf{U}_1^n/n}(u) = \frac{n^{d/2} \Phi_{\mathbf{U}}^n(t) \exp -n < t, u >}{|\Sigma(t)|^{1/2} (2\pi)^{d/2}} (1 + o(1))$$

quand $|u|$ est borné.

On introduit

$$D := \left[\frac{\pi_{\mathbf{U}}^{m_0}(m_0)}{p_{\mathbf{U}}(m_0)} \right]^n$$

et

$$N := \left[\frac{\pi_{\mathbf{U}}^{m_k}(m_k)}{p_{\mathbf{U}}(m_k)} \right]^{(n-k)}$$

Par (4.60) et le Lemme 46, on obtient

$$p_{u_{1,n}} \left(Y_1^k \right) = \left(\frac{n-k}{n} \right)^{\frac{d-2}{2}} p_{\mathbf{X}} \left(Y_1^k \right) \frac{D}{N} \frac{|\Sigma(t)|^{1/2}}{|\Sigma(t_k)|^{1/2}} (1 + o_p(1)).$$

On définit

$$A \left(Y_1^k \right) := \left(\frac{n}{n-k} \right)^{d-2} \left(\frac{g_{u_{1,n}} \left(Y_1^k \right)}{p_{\mathbf{X}} \left(Y_1^k \right)} \right)^3 \left(\frac{N}{D} \right)^2 \frac{|\Sigma(t)|}{|\Sigma(t_k)|}$$

et on simule L échantillons $Y_1^k(l)$ i.i.d., chacun étant un échantillon i.i.d. de taille k sous $p_{\mathbf{X}}$. L'approximation de A est obtenue par la méthode de simulation de Monte Carlo,

$$\widehat{A} := \frac{1}{L} \sum_{l=1}^{L} A \left(Y_1^k(l) \right).$$

En utilisant la même approximation pour B, on définit

$$B\left(Y_1^k\right) := \left(\frac{n}{n-k}\right)^{\frac{d-2}{2}} \left(\frac{g_{u_{1,n}}\left(Y_1^k\right)}{p_{\mathbf{X}}\left(Y_1^k\right)}\right)^2 \left(\frac{N}{D}\right) \frac{|\Sigma(t)|^{1/2}}{|\Sigma(t_k)|^{1/2}}$$

et

$$\widehat{B} := \frac{1}{L}\sum_{l=1}^{L} B\left(Y_1^k(l)\right)$$

pour les mêmes $Y_1^k(l)$ que précedemment.

Ainsi

$$\overline{VRE}(k) := \widehat{A} - \left(\widehat{B}\right)^2$$

$$\overline{ERE}(k) := 1 - \widehat{B}$$

$$\overline{CI}(k) := \left[\overline{ERE}(k) - 2\sqrt{\overline{VRE}(k)}, \overline{ERE}(k) + 2\sqrt{\overline{VRE}(k)}\right].$$

Remark 47. *Les algorithmes présentés dans le Chapitre 1 afin de déterminer k restent inchangées. Cependant, un problème majeur apparaît dans le cas d-dimensionnelle. En effet, résoudre $t_i = m^{-1}(m_i)$ peut poser des difficultés non négligeables, même en utilisant des techniques de résolutions numériques. Il arrive, dans certain cas, que la fonction réciproque de m soit calculable, mais même dans des situations assez courantes, comme la loi de Weibull, ce n'est pas le cas. En pratique, l'utilisation d'une méthode Quasi-Newton (plus généralement, la méthode de Broyden-Fletcher-Goldfarb-Shanno (BFGS)) est pratiquement obligatoire si la dimension de la matrice jacobienne de m (ici, $\kappa_{(i,n)}$) est grande. Cependant, ici, la matrice $\kappa_{(i,n)}$ (et son inverse) est supposée connue, car elle apparaît explicitement dans la forme de $g(y_{i+1}|y_1^i)$. Ainsi, une méthode de Newton paraît dans ce cas la méthode la plus pertinente même si, la convergence peut-être très lente si la valeur initiale est loin du vrai zéro. On peut commencer par localiser une région favorable par une méthode grossière ou par un préconditionnement de la matrice, puis appliquer la méthode de Newton pour peaufiner la précision (voir [1] pour des explications complètes).*

4.6 Exemples de Trajectoires dans un cas simple

Par le Théorème 1 (ii), $g_{u_{1,n}}$ et $p_{u_{1,n}}$ sont de plus en proche sur une famille d'ensemble qui contient les chemins typiques sous le conditionnement $(\mathbf{U}_{1,n} = u_{1,n})$ avec probablité 1 quand n croît. Par le Lemme 45, les ensembles grands sous $P_{u_{1,n}}$ sont aussi grand sous $G_{u_{1,n}}$. On peut en conclure que simuler des "longs" chemins typiques sous $p_{u_{1,n}}$ revient à simuler ces chemins typiques sous $g_{u_{1,n}}$ au moins pour n assez grand.

Etant donné la forme de la densité approximante, il peut-être particulièrement difficile d'obtenir un échantillon $\mathbf{X}_1^k$ simulé sous cette densité approximante. En effet, $g(x_{i+1}|\, x_1^i)$ est le produit normalisé d'une densité normale et de la densité d'origine. Les paramètres de cette densité normale (qui dépendent de $u_{1,n}$ et des cumulants de $u(\mathbf{X})$) peuvent rendre extrêmement compliqué la simulation d'un Y_{i+1}. Deux cas naturels se présentent, et pour chacun d'entre eux, un algorithme de simulation bien différent doit être mis en oeuvre :

- Quand $u_{1,n}$ est telle que $\lim_{n\to\infty} u_{1,n}/n = E[u(\mathbf{X})]$, simuler un échantillon est assez rapide. En effet, $\mathbf{X}_{i+1}$ de densité $g(x_{i+1}|\, x_1^i)$ est obtenue avec un algorithme d'acceptation-rejet avec densité dominante $p_{\mathbf{X}}$. La constante dans l'algorithme d'acceptation-rejet est alors $1/\sqrt{2\pi\beta}$.

– Quand $u_{1,n}$ est telle que $\lim_{n\to\infty} u_{1,n}/n \neq E[u(\mathbf{X})]$, simuler un échantillon est assez difficile. La méthode précédente ne fonctionne pas correctement et le seul algorithme qui permet de résoudre ce problème est l'algorithme de Metropolis-Hastings.

On propose de tracer des trajectoires typiques, pour $n = 100$ et différentes valeurs de a uniquement dans le cas gaussien et quand le conditionnement s'écrit $(\mathbf{S}_{1,n} = na_n)$. Dans ce cas (voir Remarque 41), l'approximation est optimal avec $k = n-1$ et la règle explicitée dans la section 4.5 n'est pas utile. Différentes figures sont présentées, représentant chacune une trajectoire typique $((S_{1,n})^{(1)}/n$ en abscisse et $(S_{1,n})^{(2)}/n$ en ordonnée), ainsi que le boxplot de X_1^{n-1}.

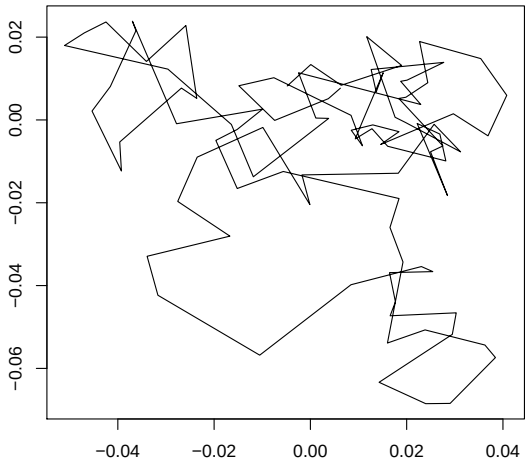

FIGURE 4.1 – Trajectoire typique pour $S_{1,n} = \left(S_{1,n}^{(1)}, S_{1,n}^{(2)}\right)$ pour $a = (0,0)$ et $n = 100$.

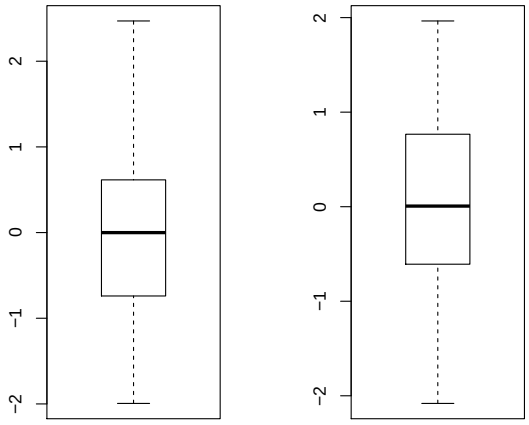

FIGURE 4.2 – Boxplot de X_1^{n-1} simulés sous g_{na} pour $a = (0,0)$ et $n = 100$.

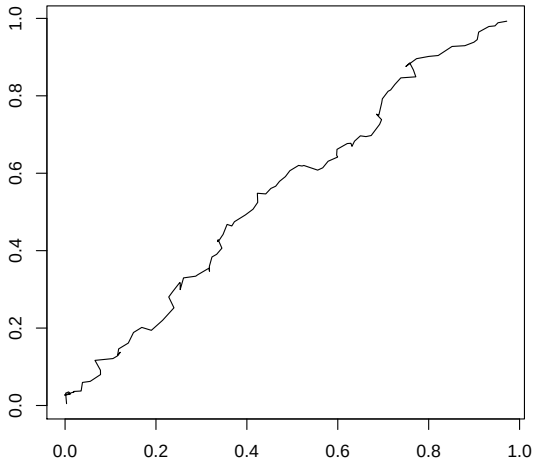

Figure 4.3 – Trajectoire typique pour $S_{1,n} = \left(S_{1,n}^{(1)}, S_{1,n}^{(2)}\right)$ pour $a = (1, 1)$ et $n = 100$.

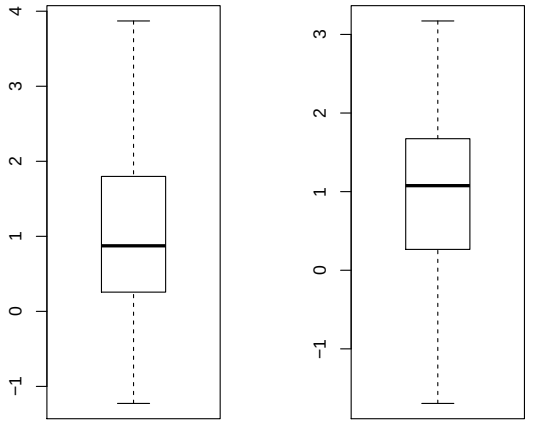

Figure 4.4 – Boxplot de X_1^{n-1} simulés sous g_{na} pour $a = (1, 1)$ et $n = 100$.

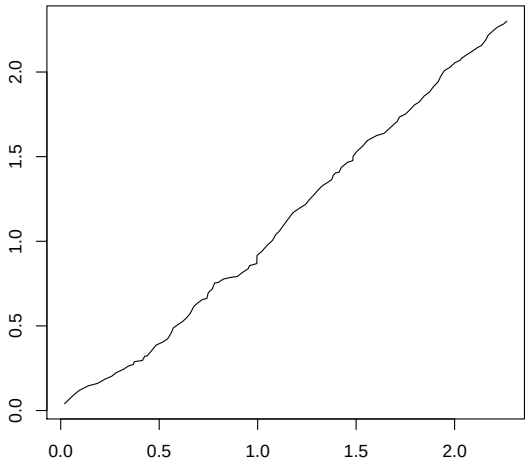

FIGURE 4.5 – Trajectoire typique pour $S_{1,n} = \left(S_{1,n}^{(1)}, S_{1,n}^{(2)}\right)$ pour $a = (2.3, 2.3)$ et $n = 100$.

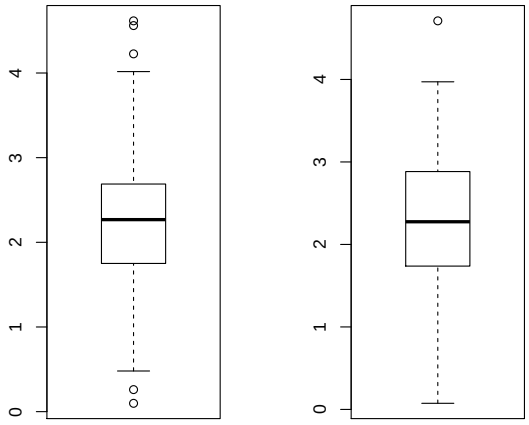

FIGURE 4.6 – Boxplot de X_1^{n-1} simulés sous g_{na} pour $a = (2.3, 2.3)$ et $n = 100$.

4.7 Familles exponentielles courbes et conditionnement.

4.7.1 Généralités

Nous commençons par des définitions sur les familles exponentielles pleines et les familles exponentielles courbes (voir [6], [65]). Une famille de lois $\{P_\theta\}$ est une famille ex-

ponentielle s-dimensionnelle si

$$\frac{dP_\theta}{d\lambda}(x) = p_\theta(x) := \exp\left[\sum_{i=1}^{s} \eta_i(\theta) T_i(x) - C(\theta)\right] h(x) \tag{4.61}$$

où λ est la mesure de Lebesgue.

Les fonctions η_i sont des fonctions indéfiniment dérivables, non-linéaires, de matrice jacobienne de plein rang. $\eta = (\eta_1, ..., \eta_s)$ est le paramètre canonique associé à la statistique exhaustive minimale canonique $T(x) = (T_1(x), ..., T_s(x))$ et $C(\theta)$ une constante de normalisation. La famille est pleine si $\dim(\theta) = s$ et courbe si $\dim(\theta) < s$. On notera une famille courbe, une famille (s, r) avec $s = \dim(T)$ et $r = \dim(\theta)$.

La vraisemblance associée à un échantillon i.i.d. X_1^n d'une famille exponentielle pleine est une fonction concave dans la paramétrisation (4.61), donc elle possède un unique maximum (s'il existe). Dans le cas où X_1^n est un échantillon i.i.d. d'une famille exponentielle courbe, la vraisemblance n'est plus concave et de multiples maximas peuvent apparaître. Récemment, plusieurs auteurs ont étudiés les familles courbes dans ce cadre. Sundberg [92] présente trois exemples de familles courbes, reprenant certains exemples présentés dans [39] et [8]. Dans le cas des modèles SUR (Seemingly Unrelated Regressions), Drton et Richardson [35] prouve que, dans le cas de petits échantillons, la vraisemblance peut aussi être multimodale (voir [35], Table 3, p.18) . Nous traiterons, en détails les trois exemples décrits par Sundberg [92], laissant l'exemple du modèle SUR pour de futures investigations.

Considérons un modèle courbe (s, r) avec $s > r$. Nous nous intéressons à transformer une famille courbe en famille pleine par conditionnement. En conditionnant par une partie de la statistique exhaustive minimale de dimension $(s - r)$ bien choisie (correspondant d'une certaine façon aux paramètres absents), le modèle devient plein (et la vraisemblance unimodale) permettant l'estimation de tous les paramètres du modèle. Afin de bien comprendre la façon dont l'approximation proposée dans les Chapitre 3 et 4, permet de résoudre ce problème, nous proposons une série d'exemples.

4.7.2 Parabole normale (3,2)

Soit (X_1^n, Y_1^n) un échantillon i.i.d. de loi indicée par $\theta = (\psi, \sigma^2)$

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \psi \\ \psi^2 \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}\right).$$

Cette famille de lois est une famille courbe $(3, 2)$ de la famille pleine indicée par (μ_1, μ_2, σ^2)

$$\mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}\right)$$

caractérisée par la statistique exhaustive minimale canonique de dimension 3

$$T(X, Y) = \begin{pmatrix} \frac{X^2+Y^2}{2} \\ X \\ Y \end{pmatrix}$$

et par le paramètre canonique de dimension 3

$$\eta = \begin{pmatrix} -\frac{1}{\sigma^2} \\ \frac{\mu_1}{\sigma^2} \\ \frac{\mu_2}{\sigma^2} \end{pmatrix}$$

Cette exemple a déjà étudié dans le Chapitre 3, section 3.4.3, dans un cadre différent. En effet, le but était d'estimer σ^2 considéré comme le paramètre d'intérêt en présence du paramètre de nuisance ψ. On a montré que la dérivée par rapport à ψ de la vraisemblance permet l'obtention de l'équation suivante

$$(u_{1,n} - \psi) + 2\psi(v_{1,n} - \psi^2) = 0$$

avec $u_{1,n} = X_1 + ... + X_n$ et $v_{1,n} = Y_1 + ... + Y_n$. Pour certaines valeurs de ψ, cette équation admet deux solutions et, alors, la vraisemblance est bimodale.

Afin d'estimer σ^2 indépendamment du paramètre de nuisance ψ, nous avons conditionné par la statistique exhaustive de dimension 2 pour ψ, $(\mathbf{U}_{1,n}, \mathbf{V}_{1,n})$ transformant la famille courbe en famille pleine $(1, 1)$. Ici, nous proposons une alternative à cette méthode permettant de transformer la famille courbe en famille pleine $(2, 2)$. Par indépendance des X_i et des Y_i, on peut écrire

$$p_{\psi,\sigma^2}\left(X_1^{n-1}, Y_1^{n-1} | \mathbf{V}_{1,n} = v_{1,n}\right) = p_{\psi,\sigma^2}\left(X_1^{n-1}\right) p_{\sigma^2}\left(Y_1^{n-1} | \mathbf{V}_{1,n} = v_{1,n}\right)$$

Ainsi

$$p_{\psi,\sigma^2}\left(X_1^{n-1}, Y_1^{n-1} | \mathbf{V}_{1,n} = v_{1,n}\right) = \prod_{i=0}^{n-2} p_{\psi,\sigma^2}\left(X_{i+1}\right) p_{\sigma^2}\left(Y_{1+1} | Y_1^i, \mathbf{V}_{1,n} = v_{1,n}\right)$$

Utilisant le Théorème 1 du Chapitre 3, le paramètre ψ n'apparait plus dans la seconde densité qui s'écrit, pour tout $i \in \{0, ..., k-1\}$,

$$p_{\sigma^2}\left(Y_{i+1} | Y_1^i, \mathbf{V}_{1,n} = v_{1,n}\right) = \mathfrak{n}\left(m_i \frac{n-i-1}{n-i}, \sigma^2 \frac{n-i-1}{n-i}, Y_{i+1}\right).$$

où la dépendance par rapport à Y_i^i est définie par

$$m_i = \frac{v_{1,n} - V_{1,i}}{n-i}.$$

La loi de $(X_{i+1}, Y_{i+1} | X_1^i, Y_1^i, \mathbf{V}_{1,n} = v_{1,n})$ appartient à une famille exponentielle de paramètre canonique

$$\eta = \begin{pmatrix} \frac{\psi}{\sigma^2} \\ -\frac{1}{\sigma^2} \end{pmatrix}$$

et de statistique canonique

$$T(X_{i+1}, Y_{i+1}) = \begin{pmatrix} X_{i+1} \\ \frac{1}{2}\left(X_{i+1}^2 + \frac{n-i}{n-i-1}\left(Y_{i+1} - m_i \frac{n-i-1}{n-i}\right)^2\right) \end{pmatrix}$$

Elle s'écrit

$$p_{\psi,\sigma^2}\left(X_{i+1}, Y_{1+1} | X_1^i Y_1^i, \mathbf{V}_{1,n} = v_{1,n}\right) = \exp\left(< \eta, T(X_{i+1}, Y_{i+1}) > - C(\psi, \sigma^2)\right) h(x_{i+1}, y_{i+1})$$

avec

$$h(x_{i+1}, y_{i+1}) = 1$$

et

$$\exp\left(-C(\psi)\right) = \frac{1}{2\pi\sigma^2}\sqrt{\frac{n-i}{n-i-1}}\exp\left(-\frac{\psi^2}{2\sigma^2}\right)$$

Dans ce premier cas, l'approximation est exacte (voir Remarque 4 du Chapitre 1) et la famille de lois obtenue par conditionnement est une famille exponentielle pleine. De plus, les statistiques canoniques associées au paramètre canonique de la famille courbe ont été modifiées par le conditionnement. La statistique canonique associé à (ψ/σ^2) est resté inchangée tandis que la statistique canonique associé à $(-1/\sigma^2)$ dans le modèle plein

$$\frac{X_{i+1}+Y_{i+1}}{2}$$

est devenue

$$\frac{1}{2}\left(X_{i+1}^2 + \frac{n-i}{n-i-1}\left(Y_{i+1} - m_i\frac{n-i-1}{n-i}\right)^2\right)$$

prenant compte du conditionnement par $(\mathbf{V}_{1,n} = v_{1,n})$.

4.7.3 Modèle de Behrens-Fisher (4,3)

Ce modèle est un des trois étudiés par Sundberg [92] permettant l'obtention d'une vraisemblance mutimodale. Soit $X_1^n, Y_{1,}^n$ un échantillon i.i.d. de loi

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right).$$

C'est une famille courbe $(4,3)$ ($\dim(T(X,Y)) = 4$ et $\dim(\theta) = 3$) indicée par $\theta = (\mu, \sigma_1^2, \sigma_2^2)$ de la famille pleine

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right).$$

de statistique canonique de dimension 4

$$T(X,Y) = \begin{pmatrix} X^2 \\ Y^2 \\ X \\ Y \end{pmatrix}$$

et de paramètre canonique indicé par $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$

$$\eta(\theta) = \begin{pmatrix} -\frac{1}{2\sigma_1^2} \\ -\frac{1}{2\sigma_2^2} \\ \frac{\mu_1}{2\sigma_1^2} \\ \frac{\mu_2}{2\sigma_2^2} \end{pmatrix}$$

En conditionnant par $(\mathbf{S}_{1,n} = \sum_{i=1}^n X_i = u_{1,n})$, la loi conditionnelle de (X_{i+1}, Y_{i+1}) appartient à une famille exponentielle pleine de statistique canonique

$$T(X_{i+1}, Y_{i+1}) = \begin{pmatrix} \frac{\left(X_{i+1} - \frac{n-i-1}{n-i} m_i\right)^2}{\frac{n-i-1}{n-i}} \\ Y_{i+1}^2 \\ Y_{i+1} \end{pmatrix}$$

et de paramètre canonique associé

$$\eta(\theta) = \begin{pmatrix} -\frac{1}{2\sigma_1^2} \\ -\frac{1}{2\sigma_2^2} \\ \frac{\mu}{2\sigma_2^2} \end{pmatrix}.$$

Ainsi, on peut écrire cette densité sous la forme

$$\exp\left(< \eta(\theta), T(X_{i+1}, Y_{i+1}) > - C(\theta)\right) h(x_{i+1}, y_{i+1})$$

avec $h(x) = 1$ et

$$\exp\left(-C(\theta)\right) = \frac{1}{2\pi} \frac{1}{\sqrt{\sigma_1^2 \sigma_2^2 \frac{n-i-1}{n-i}}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right).$$

Nous avons transformé par conditionnement la famille courbe en famille pleine.

4.7.4 Loi Gamma

Nous avons montré numériquement dans la section 3.3.1 du Chapitre 2 que, dans le cas d'une densité gamma conditionnée par une statistique exhaustive pour un des deux paramètres, la loi conditionnelle résultante est indépendante de ce paramètre. Nous allons vérifier explicitement que cette indépendance est effective. Soit X_1^n un échantillon i.i.d. de densité $f_{\theta,\eta}$ définie par

$$f_{\theta,\eta}(x) := \frac{\eta^\theta}{\Gamma(\theta)} x^{\theta-1} \exp\left(-x\eta\right) \ , \ x > 0.$$

Soit $i \in \{0, ..., k-1\}$. La densité de $\left(X_{i+1} | X_i^i, \mathbf{S}_{1,n} = a_n\right)$ est approximée par

$$C_i \exp\left(-\frac{1}{2\beta}\left(X_{i+1} - \alpha\beta\right)^2\right\} \exp\{-\eta X_{i+1} + \theta \log(X_{i+1}) - \log(X_{i+1})\Big) \tag{4.62}$$

où les valeurs α et β sont explicitées en l'annexe. C_i est une constante assurant à (4.62) d'être une densité. En développant ce calcul, on obtient

$$C_i \exp\left(-\frac{\theta}{2m_i^2\left(n-i-1\right)} X_{i+1}^2 + \eta X_{i+1} - \frac{\theta}{m_i} X_{i+1} + \frac{1}{m_i(n-i-1)} - \eta X_{i+1} + \theta \log(X_{i+1}) - \log(X_{i+1})\right)$$

Ainsi, pour tout $i \in \{0, ...k-1\}$, la loi conditionnelle est indépendante du paramètre η. De plus, pour tout $i \in \{0, ...k-1\}$, la loi appartient à une famille exponentielle pleine qui s'écrit

$$\exp\left(\theta T(X_{i+1}) - C(\theta)\right) h(X_{i+1})$$

avec comme paramètre canonique θ, comme statistque canonique

$$T(X_{i+1}) = \frac{X_{i+1}^2}{2m_i^2\left(n-i-1\right)} - \frac{X_{i+1}}{m_i} + \log(X_{i+1})$$

et avec $h(X_{i+1}) = X_{i+1}$ et $\exp\left(-C(\theta)\right) = C_i$.

Une nouvelle fois, le conditionnement par la statistique canonique associée à η a modifié la statistique canonique associée à θ.

4.7.5 Coefficient de corrélation (2,1)

Soit (X_1^n, Y_1^n) un échantillon i.i.d. de loi indicée par $\theta = (\rho)$

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

Cette famille de lois est une famille courbe $(2,1)$ de la famille pleine indicée par (σ^2, ρ)

$$\mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix}\right)$$

caractérisée par la statistique exhaustive minimal canonique de dimension 2

$$T(X,Y) = \begin{pmatrix} \frac{X^2+Y^2}{2} \\ XY \end{pmatrix}$$

et par le paramètre canonique de dimension 2

$$\eta(\rho) = \begin{pmatrix} -\frac{1}{(1-\rho^2)\sigma^2} \\ \frac{\rho}{(1-\rho^2)\sigma^2}. \end{pmatrix}$$

Dans le modèle plein, la dimension du paramètre à estimer est deux tandis que, dans le modèle courbe, un seul paramètre est à estimer pour une statistique canonique de dimension deux. Sundberg [92] remarque qu'il existe des cas où la vraisemblance associée au modèle courbe peut admettre jusqu'à trois maximas. Afin d'éviter de tels cas, même rares, nous nous intéressons à transformer le modèle courbe en un modèle plein par conditionnement. A l'aide de l'approximation dévéloppée dans le Chapitre 4, nous pouvons déterminer la loi du vecteur (X_1^k, Y_1^k) conditionnée par l'événement

$$\left(\mathbf{U}_{1,n} = \sum_{i=1}^{n} u(X_i, Y_i) = u_{1,n}\right)$$

où $u_{1,n}$ est la valeur observée de la statistique et u est la fonction de $\mathbb{R}^2$ dans $\mathbb{R}$ définie par $u(x,y) = xy$.

Soit $i \in \{0, ..., k-1\}$. La densité de $(X_{i+1}, Y_{i+1} | X_1^i, Y_1^i, \mathbf{U}_{1,n} = u_{1,n})$ est approximée, à une constante près indépendante de (X_{i+1}, Y_{i+1}), par

$$\exp\left(-\frac{1}{2\beta}u^2(X_{i+1}, Y_{i+1}) + \alpha u(X_{i+1}, Y_{i+1}) + \eta_2(\rho)u(X_{i+1}, Y_{i+1}) + \eta_1(\rho)v(X_{i+1}, Y_{i+1})\right)$$

Utilisant la Remarque 4 et les expressions de α et β définies dans le Chapitre 4, Section 4.3, le terme $\eta_2(\rho)u(X_{i+1}, Y_{i+1})$ s'annule avec celui venant de $t_i u(X_{i+1}, Y_{i+1})$. Ainsi

$$\exp\left(\eta_1(\rho)\left(-\frac{1}{2\eta_1(\rho)\beta}u^2(X_{i+1}, Y_{i+1}) + \left(\gamma + \frac{\mu_3(t_i)}{2s^4(t_i)(n-i-1)\eta_1(\rho)}\right)u(X_{i+1}, Y_{i+1}) + v(X_{i+1}, Y_{i+1})\right)\right)$$

Notons $T_\rho(X_{i+1}, Y_{i+1})$ la statistique telle que la quantitée précédente s'écrit $\exp\left(\eta_2(\rho)T_\rho(X_{i+1}, Y_{i+1})\right)$.

Il nous faut montrer que $T_\rho(X_{i+1}, Y_{i+1})$ ne dépend pas de $\eta_2(\rho)$. Cependant, le paramètre ρ apparait explicitement en calculant β, μ_3, s^2 et γ à l'aide de la remarque 4. Nous allons montrer que $T_\rho(X_{i+1}, Y_{i+1})$ est *presque* indépendante de ρ et donc la densité

approximante de $(X_{i+1}, Y_{i+1}|X_1^i, Y_1^i, \mathbf{U}_{1,n} = u_{1,n})$ appartient *presque* à une famille exponentielle.

Pour cela, on définit pour tout $i \in \{0, ..., k-1\}$ la fonction $E_A(i)$ suivante

$$E_A(i) = \max_{\rho \in A} T_\rho(x_{i+1}, y_{i+1}) - \min_{\rho \in A} T_\rho(x_{i+1}, y_{i+1})$$

où A est un ensemble strictement inclus dans $[-1, 1]$.

On simule un échantillon (X_1^n, Y_1^n) pour $n = 100$ et $\rho = 0.1$ et on trace ci-dessous la fonction $E(i)$ en fonction de i pour différents intervalles A. En étudiant ces courbes, on peut observer que le modèle décrit ci-dessus forme *presque* une famille exponentielle pleine pour un voisinage assez grand autour de la vrai valeur de ρ.

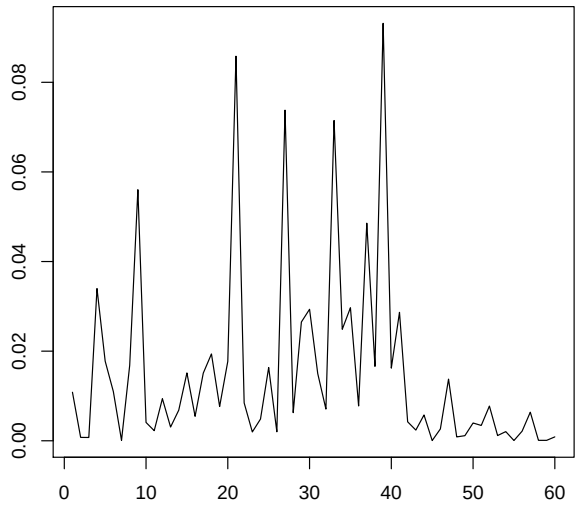

FIGURE 4.7 – $E(i)$ en fonction de i pour $n = 100$, $k = 60$ et $A = [-0.5, 0.5]$.

Pour de futurs développements, il serait intéressant d'étudier la relation théorique entre les approximations dévéloppées dans cette thèse et la transformation des familles courbes en familles pleines.

Remarque 4. *Nous donnons les quantités permettant d'obtenir les paramètres de l'approximation.*

$$\phi(t) = \frac{\sqrt{1-\rho^2}}{1 - (\rho + (1-\rho^2)t)^2}$$

Notons :

$$t_i = -\gamma\eta_1(\rho) - \eta_2(\rho).$$

γ est solution de

$$m_i\gamma^2 - \frac{\gamma}{\eta_1(\rho)} - m_i = 0$$

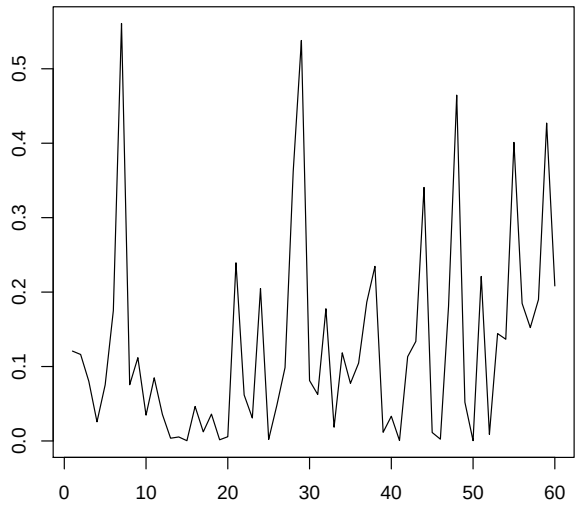

Figure 4.8 – $E(i)$ en fonction de i pour $n = 100$, $k = 60$ et $A = [-0.9, 0.9]$.

et les solutions qui doivent appartenir à $[-1; 1]$

$$\gamma = \frac{1 \pm \sqrt{1 + 4m_i^2\eta_1(\rho)}}{2m_i\eta_1(\rho)}$$

$$s^2(t) = m_i^2 \frac{1+\gamma^2}{\gamma^2}$$

$$\mu_3(t) = 6m_i^3 \frac{1+2\gamma^2}{\gamma^2}$$

4.8 Démonstration du Théorème 1 du Chapitre 4.

4.8.1 Lemmes permettant la démonstration principale.

Lemma 48 (Lemme d'Invariance)**.**

$$\begin{gathered}\forall 1 \leq i \leq j \leq n, \forall a \in supp(P), \forall u, s \\ p\left(\mathbf{S}_{i,j} = u | \mathbf{S}_{1,n} = s\right) = \pi^a\left(\mathbf{S}_{i,j} = u | \mathbf{S}_{1,n} = s\right)\end{gathered} \tag{4.63}$$

Démonstration. Utilisons la formule de Bayes pour obtenir :

$$p\left(\mathbf{S}_{i,j} = u | \mathbf{S}_{1,n} = s\right) = \frac{p\left(\mathbf{S}_{i,j} = u\right) p\left(\mathbf{S}_{1,i-1} + \mathbf{S}_{j+1,n} = s - u\right)}{p\left(\mathbf{S}_{1,n} = s\right)} \tag{4.64}$$

De même,

$$\pi^a\left(\mathbf{S}_{i,j} = u | \mathbf{S}_{1,n} = s\right) = \frac{\pi^a\left(\mathbf{S}_{i,j} = u\right) \pi^a\left(\mathbf{S}_{1,i-1} + \mathbf{S}_{j+1,n} = s - u\right)}{\pi^a\left(\mathbf{S}_{1,n} = s\right)}$$

En utilisant la définition (4.6) de π^a, l'expression précédente devient,

$$\pi^a\left(\mathbf{S}_{i,j} = u | \mathbf{S}_{1,n} = s\right) = \left(\frac{\exp < t_a, u >}{\Phi_{\mathbf{S}_{i,j}}(t_a)} p\left(\mathbf{S}_{i,j} = u\right)\right) \left(\frac{\exp < t_a, s - u >}{\Phi_{\mathbf{S}_{1,i-1}+\mathbf{S}_{j+1,n}}(t_a)} p\left(\mathbf{S}_{1,i-1} + \mathbf{S}_{j+1,n} = s - u\right)\right)$$

$$\left(\frac{\Phi_{\mathbf{S}_{1,n}}(t_a)}{\exp < t_a, s >} \frac{1}{p\left(\mathbf{S}_{1,n} = s\right)}\right)$$

En utilisant (4.64) et la bilinéarité du produit scalaire,

$$\pi^a\left(\mathbf{S}_{i,j} = u | \mathbf{S}_{1,n} = s\right) = p\left(\mathbf{S}_{i,j} = u | \mathbf{S}_{1,n} = s\right)$$

$$\left(\frac{\Phi_{\mathbf{S}_{i,j}}(t_a) \Phi_{\mathbf{S}_{1,i-1}+\mathbf{S}_{j+1,n}}(t_a)}{\Phi_{\mathbf{S}_{1,n}}(t_a)}\right)$$

Il reste à étudier le produit des fonctions génératrices des moments et s'assurer qu'il est égal à 1. Par définition, les $(\mathbf{X}_i)_{1 \leq i \leq n}$ sont i.i.d.. Ainsi

$$\Phi_{\mathbf{S}_{1,n}}(t_a) = E[\exp < t_a, \mathbf{S}_{1,n} >]$$

$$\Phi_{\mathbf{S}_{1,n}}(t_a) = E[\exp < t_a, \mathbf{S}_{1,i-1} + \mathbf{S}_{i,j} + \mathbf{S}_{j+1,n} >]$$

$$\Phi_{\mathbf{S}_{1,n}}(t_a) = E[\exp < t_a, \mathbf{S}_{1,i-1} + \mathbf{S}_{j+1,n} > \exp < t_a, S_{i,j} >]$$

$$\Phi_{\mathbf{S}_{1,n}}(t_a) = E[\exp < t_a, \mathbf{S}_{1,i-1} + \mathbf{S}_{j+1,n} >] E[\exp < t_a, \mathbf{S}_{i,j} >]$$

Ainsi,

$$p\left(\mathbf{S}_{i,j} = u | \mathbf{S}_{1,n} = s\right) = \pi^a\left(\mathbf{S}_{i,j} = u | \mathbf{S}_{1,n} = s\right)$$

■

Lemma 49 (Moments sous p_{na})**.** *Pour tout j, p, q dans $\{1, ..., d$, on a $E_{P_{na}}\left(\mathbf{X}_1^{(j)}\right) = a^{(j)}$, $E_{P_{na}}\left(\mathbf{X}_1^{(p)} \mathbf{X}_2^{(q)}\right) = a^{(p)} a^{(q)} + 0\left(\frac{1}{n}\right)$ et $E_{P_{na}}\left(\mathbf{X}_1^{(p)} \mathbf{X}_1^{(q)}\right) = \kappa_{p,q}(t) + a^{(p)} a^{(q)} + 0\left(\frac{1}{n}\right)$ où $\kappa_{p,q}(t) = \left({}^t\nabla\nabla \log \Phi(t)\right)_{p,q}$ et t tel que $m(t) = a$ avec a une suite convergente.*

Démonstration. Notons $x = {}^t\left(0, ..., 0, x^{(p)}, 0, ..., 0, x^{(q)}, 0, ..., 0\right)$. En utilisant la formule de Bayes, on obtient :

$$p_{na}(\mathbf{X}_1 = x) = \frac{p\left(\mathbf{S}_{2,n} = na - x\right) p\left(\mathbf{X}_1 = x\right)}{p\left(\mathbf{S}_{1,n} = na\right)} = \frac{\pi^a\left(\mathbf{S}_{2,n} = na - x\right) \pi^a\left(\mathbf{X}_1 = x\right)}{\pi^a\left(\mathbf{S}_{1,n} = na\right)}$$

En normalisant à la fois $\mathbf{S}_{2,n}$ et $\mathbf{S}_{1,n}$ et en effectuant un développement d'Edgeworth dans les expressions ci-dessus, on obtient $E_{P_{na}}\left(\mathbf{X}_1^{(p)}\mathbf{X}_1^{(q)}\right) = \kappa_{p,q}(t) + a^{(p)}a^{(q)} + 0\left(\frac{1}{n}\right)$. Une approche similaire pour la densité jointe $p_{na}(\mathbf{X}_1 = x, \mathbf{X}_2 = y)$, utilisant la même densité tilté π^a, permet d'obtenir le résultat. ■

Lemma 50 (Convergence de m_i)**.** *Soient a une suite convergente. Alors sous (E1), pour tout j entre 1 et d, on a :*

$$\max_{1\leq i\leq k}\left|m_i^{(j)}\right| = a_n^{(j)} + o_{P_{na}}\left(\epsilon_n\right). \tag{4.65}$$

Nous avons aussi que, pour tout j, r, s dans $\{1, ..., d\}$, $\max_{1\leq i\leq k}\left|\kappa_{(i,n)}^{j,r}\right|$ et $\max_{1\leq i\leq k}\left|\kappa_{(i,n)}^{j,r,s}\right|$ tendant en P_{na}-probabilité respectivement, vers $(\sigma_{j,r})$, le terme générique de la matrice de variance-covariance de $\mathbf{X}$ sous $P_{\mathbf{X}}$, et vers $(\kappa^{j,r,s})$ le cumulant joint de $\mathbf{X}$ sous $P_{\mathbf{X}}$.

Démonstration. Soit $j \in \{1, ...d\}$. On définit :

$$\begin{aligned} V_{i+1}^{(j)} &:= m(t_i)^{(j)} - a^{(j)} \\ &= \frac{S_{i+1,n}^{(j)}}{n-i} - a^{(j)}. \end{aligned}$$

Nous allons montrer que

$$\max_{0\leq i\leq k-1}|V_{i+1}^{(j)}| = o_{P_{na}}\left(\epsilon_n\right), \tag{4.66}$$

en d'autres termes, pour tout δ positif,

$$\lim_{n\to\infty} P_{na}\left(\max_{0\leq i\leq k-1}|V_{i+1}^{(j)}| > \delta\epsilon_n\right) = 0$$

ce que nous allons prouver en suivant une démonstration semblable à celle de l'inégalité maximal de Kolmogorov.

On définit

$$A_i := \left(\left(|V_{i+1}^{(j)}| \geq \delta\epsilon_n\right) \text{ and } \left(|V_j^{(j)}| < \delta\epsilon_n \text{ for all } j < i+1\right)\right).$$

d'où

$$\left(\max_{0\leq i\leq k-1}|V_{i+1}^{(j)}| > \delta\epsilon_n\right) = \bigcup_{i=0}^{k-1} A_i.$$

On a :

$$\begin{aligned}
E_{P_{na}}(V_k^{(j)})^2 &= \int_{\cup A_i} (V_k^{(j)})^2 dP_{na} + \int_{(\cup A_i)^c} (V_k^{(j)})^2 dP_{na} \\
&\geq \int_{\cup A_i} \left((V_k^{(j)})^2 + 2\left(V_k^{(j)} - V_i^{(j)}\right) V_i^{(j)} \right) dP_{na} \\
&+ \int_{(\cup A_i)^c} \left((V_k^{(j)})^2 + 2\left(V_k^{(j)} - V_i^{(j)}\right) V_i^{(j)} \right) dP_{na} \\
&\geq \int_{\cup A_i} (V_k^{(j)})^2 dP_{na} \\
&\geq \delta^2 \epsilon_n^2 \sum_{j=0}^{k-1} P_{na}(A_j) \\
&= \delta^2 \epsilon_n^2 P_{na}\left(\max_{0\leq i\leq k-1} |V_{i+1}^{(j)}| > \delta\epsilon_n \right).
\end{aligned}$$

La troisième ligne vient du fait que $EV_i^{(j)}\left(V_k^{(j)} - V_i^{(j)}\right) = 0$, ce qu'on peut prouver de manière directe. Ainsi,

$$P_{na}\left(\max_{0\leq i\leq k-1} |V_{i+1}^{(j)}| > \delta\epsilon_n \right) \leq \frac{Var_{P_{na}}(V_k^{(j)})}{\delta^2\epsilon_n^2} = \frac{1}{\delta^2\epsilon_n^2 (n-k)}(1+o(1))$$

où on utilise le Lemme 49. Ainsi (4.66) est vrai sous (E1). ■

Lemma 51 (Etude du maximum des $\mathbf{X}_i^{(j)}$)**.** *Soit a une suite convergente. Alors, pour tout j entre 1 et d, on a :*

$$\max\left(|\mathbf{X}_1^{(j)}|, ..., |\mathbf{X}_k^{(j)}| \right) = O_{P_{na}}(\log n) \tag{4.67}$$

Démonstration. Soit $j \in \{1, ...d\}$.

Soit $\left|\mathbf{X}_i^{(j)}\right| := \mathbf{X}_i^{(j),-} + \mathbf{X}_i^{(j),+}$ avec $\mathbf{X}_i^{(j),-} := -\min\left(0, \mathbf{X}_i^{(j)}\right)$, $\mathbf{X}_i^{(j),+} := \max\left(0, \mathbf{X}_i^{(j)}\right)$; il est suffisant de prouver que $\max_i \mathbf{X}_i^{(j),-} = O_{P_{na}}(\log n)$ et $\max_i \mathbf{X}_i^{(j),+} = O_{P_{na}}(\log n)$. Puisque $E\exp < t, \mathbf{X} >$ est finie dans un voisinage non vide de $\underline{0}$, $E\exp t\mathbf{X}^{(j),-}$ et $E\exp t\mathbf{X}^{(j),+}$ le sont. Il suffit donc de prouver le Lemme pour des $\mathbf{X}_i^{(j)}$ positifs seulement.

Pour tout t, on a

$$\begin{aligned}
P_{na}\left(\max\left(\mathbf{X}_1^{(j)}, ..., \mathbf{X}_k^{(j)}\right) > t^{(j)} \right) &\leq k P_{na}\left(\mathbf{X}_n^{(j)} > t^{(j)}\right) \\
&= k \int_{t^{(j)}}^{\infty} p_{na}\left(\mathbf{X}_n^{(j)} = x^{(j)}\right) dx^{(j)} \\
&= k \int_{t^{(j)}}^{\infty} \left(\int_{\mathbb{R}^{d-1}} p_{na}\left(\mathbf{X}_n = x\right) \prod_{q=1, q\neq j}^{d} dx^{(q)} \right) dx^{(j)} \\
&= k \int_{t^{(j)}}^{\infty} \left(\int_{\mathbb{R}^{d-1}} \pi^a\left(\mathbf{X}_n = x\right) \frac{\pi^a(\mathbf{S}_{1,n-1} = na - x)}{\pi^a\left(\mathbf{S}_{1,n} = na\right)} \prod_{q=1, q\neq j}^{d} dx^{(q)} \right) dx^{(j)}
\end{aligned}$$

En centrant et normalisant à la fois $\mathbf{S}_{1,n}$ et $\mathbf{S}_{1,n-1}$ par rapport à la densité π^a dans les trois lignes ci-dessus et en notant $\overline{\pi_n^a}$ la densité de $\overline{\mathbf{S}_{1,n}} := \kappa_{(a)}^{-1/2}\left(\mathbf{S}_{1,n} - na\right)/\sqrt{n}$ quand $\mathbf{X}$ est de densité π^a d'espérance a et de matrice variance-covariance $\kappa_{(a)}$, on obtient

$$P_a\left(\max\left(\mathbf{X}_1^{(j)}, ..., \mathbf{X}_k^{(j)}\right) > t^{(j)} \right) \leq k\frac{\sqrt{n}}{\sqrt{n-1}}$$

$$\int_{t^{(j)}}^{\infty}\left(\int_{\mathbb{R}^{d-1}}\pi^{a}\left(\mathbf{X}_{n}=u\right)\frac{\overline{\pi_{n-1}^{a}}\left(\overline{\mathbf{S}_{1,n-1}}=\kappa_{(a)}^{-1/2}\left(a-x\right)/\left(\sqrt{n-1}\right)\right)}{\overline{\pi_{n}^{a}}\left(\overline{\mathbf{S}_{1,n}}=0\right)}\prod_{q=1,q\neq j}^{d}x^{(q)}\right)dx^{(j)}.$$

Sous la suite des densités π^{a}, un développement au premier degré d'Edgeworth peut-être effectué,

$$P_{na}\left(\max\left(\mathbf{X}_{1}^{(j)},...,\mathbf{X}_{k}^{(j)}\right)>t^{(j)}\right)\leq k\frac{\sqrt{n}}{\sqrt{n-1}}$$

$$\int_{t^{(j)}}^{\infty}\left(\int_{\mathbb{R}^{d-1}}\pi^{a}\left(\mathbf{X}_{n}=u\right)\frac{\mathfrak{n}\left(\kappa_{(a)}^{-1/2}\left(a-x\right)/\left(\sqrt{n-1}\right)\right)\mathbf{P}\left(u,i,n\right)+o(1)}{\mathfrak{n}\left(0\right)+o(1)}\prod_{q=1,q\neq j}^{d}x^{(q)}\right)dx^{(j)}.$$

$$\leq kC\frac{\sqrt{n}}{\sqrt{n-1}}\int_{t^{(j)}}^{\infty}\left(\int_{\mathbb{R}^{d-1}}\pi^{a}\left(\mathbf{X}_{n}=u\right)\prod_{q=1,q\neq j}^{d}x^{(q)}\right)dx^{(j)}.$$

pour C une constante indépendante de n et de $t^{(j)}$. Ainsi

$$\mathbf{P}\left(u,i,n\right):=1+P_{3}\left(\kappa_{(a)}^{-1/2}\left(a-x\right)/\left(\sqrt{n-1}\right)\right)$$

où $P_{3}(x)=\frac{\kappa_{(a)}^{j,k,l}}{6\sqrt{n}}h_{jkl}(x)$ avec $h_{jkl}(x)=x_{j}x_{k}x_{l}-\kappa_{j,k}^{(a)}x_{l}[3]$ et $x_{i}=\kappa_{i,r}^{(a)}x^{(r)}$; $\kappa_{(a)}^{j,k,l}$ est le troisième cumulant joint sous π^{a}. En utilisant l'uniformité sous u dans le terme restant du développement d'Edgeworth et en utilisant l'inégalité de Chernoff pour borner $\Pi^{a}\left(\mathbf{X}_{n}^{(j)}>t\right)$,

$$P_{na}\left(\max\left(\mathbf{X}_{1}^{(j)},...,\mathbf{X}_{k}^{(j)}\right)>t^{(j)}\right)\leq kCst\frac{\Phi(t^{(j)}+\lambda)}{\Phi(t^{(j)})}e^{-\lambda t^{(j)}}$$

pour tout λ tel que $\Phi(t^{(j)}+\lambda)$ est finie. Pour t tel que

$$t^{(j)}/\log k\to\infty,$$

on a

$$P_{na}\left(\max\left(\mathbf{X}_{1}^{(j)},...,\mathbf{X}_{k}^{(j)}\right)<t^{(j)}\right)\to 1,$$

ce qui prouve le Lemme. ■

Lemma 52 (Etude du maximum des $\mathbf{V}_{i}^{(j)}$)**.** *Soit a_{n} une suite convergente. En notant $\mathbf{V}_{i+1}:=\kappa_{(i,n)}^{-1/2}(\mathbf{X}_{i+1}-m_{i,n})$ pour tout i entre 0 et $k-1$, on a*

$$\forall j\in\{1,...,d\}\quad\max\left(|\mathbf{V}_{1}^{(j)}|,...,|\mathbf{V}_{k}^{(j)}|\right)=O_{P_{na}}(\log n)\tag{4.68}$$

Démonstration. Soit $j\in\{1,...,d\}$.

$$\begin{aligned}|\mathbf{V}_{i}^{(j)}|=|[\kappa_{(i,n)}^{-1/2}(\mathbf{X}_{i+1}-m_{i,n})]^{(j)}|&\\ \leq\sup_{1\leq l\leq d}|\alpha_{j,l}|\sup_{1\leq l\leq d}|\mathbf{X}_{i+1}^{(l)}-m_{i,n}^{(l)}|&\\ \leq\sup_{1\leq l\leq d}|\alpha_{j,l}|(\sup_{1\leq l\leq d}|\mathbf{X}_{i+1}^{(l)}|+\sup_{1\leq l\leq d}|m_{i,n}^{(l)}|)&\\ \leq\sup_{1\leq l\leq d}|\alpha_{j,l}|(O_{P_{na}}(\log n)+a^{(j)}+o_{P_{na}}\left(\epsilon_{n}\right))&\\ \leq CO_{P_{na}}(\log n)&\end{aligned}$$

en utilisant les Lemme 50 et 51 et en notant $\kappa_{(i,n)}^{-1/2} := (\alpha_{j,l})_{1\leq l\leq d}$.

Ainsi

$$\max_{1\leq i\leq k-1} |\mathbf{V}_i^{(j)}| \leq \sup_{1\leq l\leq d} |\alpha_{j,l}|(\max_{1\leq i\leq k-1} \sup_{1\leq l\leq d} |\mathbf{X}_{i+1}^{(l)}| + \max_{1\leq i\leq k-1} \sup_{1\leq l\leq d} |m_{i,n}^{(l)}|)$$

et, finalement,

$$\max_{1\leq i\leq k-1} |\mathbf{V}_i^{(j)}| \leq CO_{P_{na}}(\log n)$$

■

Le théorème suivant (Taylor (1985) [86], Théorème 3.1.3) permet de contrôler les termes de restes dans les développements d'Edgeworth.

Théorème 5 (Théorème de Taylor)**.** *Soit $X_{i,n}$,$1 \leq i \leq k$ une matrice de variables aléatoires réelles échangeables ligne par ligne et $\lim_{n\to\infty} k = \infty$. Soit $\rho_n := EX_{1,n}X_{2,n}$. On suppose qu'il existe une constante finie Γ telle que $EX_{1,n}^2 \leq \Gamma$. Si pour une suite $(a_{i,n})$ telle que $\lim_{n\to\infty} \sum_{i=1}^k a_{i,n}^2 = 0$, on a*

$$\lim_{n\to\infty} \rho_n \left(\sum_{i=1}^k a_{i,n}^2\right)^2 = 0$$

Alors

$$\lim_{n\to\infty} \sum_{i=1}^k a_{i,n}X_{i,n} = 0$$

en probabilité.

4.8.2 Démonstration de l'approximation utilisant le développement d'Edgeworth dans le théorème principal.

Remark 53. *En utilisant les notations index et les notations exlog (section 4.2.3 et 4.2.3), nous pouvons simplifier certaines des notations suivantes, en particulier les cumulants joints des quantités consididérées. Nous préciserons, quand ce sera le cas, l'ensemble index considéré. Ainsi, en notant $I = \{i_1, ..., i_\nu\}$, on a*

$$\kappa_0^I = K\left[\mathbf{X}^{(i_1)}...\mathbf{X}^{(i_\nu)}\right]$$

L'indice 0 sera omis dans la suite.

Notons

$$\mathbf{V}_{i+1} := \kappa_{(i,n)}^{-1/2}(\mathbf{X}_{i+1} - m_{i,n}) \tag{4.69}$$

Sous π^{m_i}, $\mathbf{V}_{i+1}$ a pour espérance 0 et pour matrice de variance-covariance I_d.

Notons $\mathbf{V}_{i+2,n} := \sum_{j=i+2}^n \mathbf{V}_j$

Notons $\overline{\pi_{n-i-1}^{m_{i,n}}}$ la densité des sommes partielles normalisées $\mathbf{V}_{i+2,n}/(\sqrt{n-i-1})$ quand les vecteurs aléatoires sommés sont i.i.d. avec une densité commune π^{m_i}. De plus, nous évaluons $\overline{\pi_{n-i-1}^{m_i}}$ et son approximation normale au point U_{i+1}

$$p(\mathbf{X}_{i+1} = Y_{i+1}|\mathbf{S}_{i+1,n} = na - S_{1,i})$$

$$= \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \pi^{m_i}(\mathbf{X}_{i+1} = Y_{i+1}) \frac{\overline{\pi^{m_i}_{n-i-1}}(-U_{i+1}/\sqrt{n-i-1})}{\overline{\pi^{m_{i,n}}_{n-i}}(0)}.$$

où $U_{i+1} = \kappa_{(i,n)}^{-1/2}(Y_{i+1} - m_i)$.

La suite de densités $\overline{\pi^{m_i}_{n-i}}$ converge localement vers la densité normale standard d-dimensionnelle sous les hypothèses sus-mentionnées, quand $(n-i)$ tends vers l'infini, i.e. quand $(n-k)$ tends vers l'infini, et un développement d'Edgeworth à l'ordre 4 peut être effectué pour le dénominateur et le numérateur.

Notons $Z_{i+1} := -U_{i+1}/\sqrt{n-i-1}$.

Développement d'Edgeworth

Nous allons d'abord étudier le développement du numérateur :

$$\overline{\pi^{m_{i,n}}_{n-i-1}}(Z_{i+1}) = \mathfrak{n}_d(Z_{i+1})[\ 1 + \tfrac{1}{\sqrt{n-i-1}} Q_3(Z_{i+1}) + \tfrac{1}{n-i-1} Q_4(Z_{i+1})\ ] + O_{P_{na}}(\frac{Q_4(Z_{i+1})}{(n-i-1)^{3/2}})$$

uniformément en Z_{i+1} avec Q_3 et Q_4 définies (section 4.2.3) par

$$Q_3(x) := \frac{1}{6} \kappa^{j,l,m} h_{jlm}(x)$$

$$Q_4(x) := \frac{1}{24} \kappa^{j,l,m,q} h_{jlmq}(x) + \frac{1}{72} \kappa^{j,l,m} \kappa^{q,r,s} h_{jlmqrs}(x).$$

Notre but est d'obtenir un développement polynomial en termes de puissances de $(n-i)$. Les moments qui constituent les polynômes tensoriels d'Hermite sont ici les moments de U. En effet, la densité des sommes partielles de U est étudiée. Quand nous changerons de variables à la fin de cette partie, les moments devront aussi être modifiés.

Etude de Q_3

$$Q_3(Z_{i+1}) = \frac{1}{6} \sum_{j=1}^{d} \sum_{l=1}^{d} \sum_{m=1}^{d} \kappa_{(i,n)}^{j,l,m} (\frac{U_{i+1}^{(j)} U_{i+1}^{(l)} U_{i+1}^{(m)}}{(n-i-1)^{3/2}} + \frac{\kappa_{j,l}^{(i,n)} U_{i+1}^{(m)}}{\sqrt{n-i-1}}[3])$$

Donc ,

$$\frac{Q_3(Z_{i+1})}{\sqrt{n-i-1}} = \frac{1}{6} \sum_{j=1}^{d} \sum_{l=1}^{d} \sum_{m=1}^{d} \kappa_{(i,n)}^{j,l,m} (\frac{U_{i+1}^{(j)} U_{i+1}^{(l)} U_{i+1}^{(m)}}{(n-i-1)^2} + \frac{\kappa_{j,l}^{(i,n)} U_{i+1}^{(m)}}{n-i-1}[3])$$

En utilisant le Lemme 51 :

$$\frac{U_{i+1}^{(j)} U_{i+1}^{(l)} U_{i+1}^{(m)}}{(n-i-1)^2} = \frac{O_{P_{na}}((\log n)^3)}{(n-i-1)^2}$$

et

$$\frac{Q_3(Z_{i+1})}{\sqrt{n-i-1}} = \frac{1}{6} \sum_{j=1}^{d} \sum_{l=1}^{d} \sum_{m=1}^{d} \kappa_{(i,n)}^{j,l,m} \frac{\kappa_{j,l}^{(i,n)} U_{i+1}^{(m)}}{n-i-1}[3] + \frac{O_{P_{na}}((\log n)^3)}{(n-i-1)^2}$$

Notons :

$$A := \frac{1}{6}\sum_{j=1}^{d}\sum_{l=1}^{d}\sum_{m=1}^{d}\kappa_{(i,n)}^{j,l,m}\frac{\kappa_{j,l}^{(i,n)}U_{i+1}^{(m)}}{n-i-1}[3]$$

Nous allons à présent étudier A en détail.
D'après les notations présentées dans la section 4.2.3, nous avons :

$$\kappa_{j,l}^{(i,n)}U_{i+1}^{(m)}[3] = \kappa_{j,l}^{(i,n)}U_{i+1}^{(m)} + \kappa_{j,m}^{(i,n)}U_{i+1}^{(l)} + \kappa_{l,m}^{(i,n)}U_{i+1}^{(j)}$$

$$A = \frac{1}{6(n-i-1)}\left(\sum_{j=1}^{d}\sum_{l=1}^{d}\sum_{m=1}^{d}\kappa_{(i,n)}^{j,l,m}\kappa_{j,l}^{(i,n)}U_{i+1}^{(m)} + \sum_{j=1}^{d}\sum_{l=1}^{d}\sum_{m=1}^{d}\kappa_{(i,n)}^{j,l,m}\kappa_{j,m}^{(i,n)}U_{i+1}^{(l)} + \sum_{j=1}^{d}\sum_{l=1}^{d}\sum_{m=1}^{d}\kappa_{(i,n)}^{j,l,m}\kappa_{l,m}^{(i,n)}U_{i+1}^{(j)}\right)$$

Ici,la matrice de variance-covariance est la matrice identitée,

$$A = \frac{1}{6(n-i-1)}\left(\sum_{j=1}^{d}\sum_{m=1}^{d}\kappa_{(i,n)}^{j,j,m}U_{i+1}^{(m)} + \sum_{j=1}^{d}\sum_{l=1}^{d}\kappa_{(i,n)}^{j,l,j}U_{i+1}^{(l)} + \sum_{j=1}^{d}\sum_{m=1}^{d}\kappa_{(i,n)}^{j,m,m}U_{i+1}^{(j)}\right)$$

Ainsi,

$$\frac{Q_3(Z_{i+1})}{\sqrt{n-i-1}} = \frac{1}{2(n-i-1)}{}^{t}U_{i+1}\gamma + \frac{O_{P_{na}}((\log n)^3)}{(n-i-1)^2}$$

en utilisant l'invariance des cumulants par permutation des indices, et en notant,

$$\gamma = (\sum_{j=1}^{d}\kappa_{(i,n)}^{j,j,m})_{1\le m\le d}. \tag{4.70}$$

Etude de Q_4 L'étude de Q_4 s'effectue en deux parties. Notons :

$$A_1 := \kappa_{(i,n)}^{j,l,m,p}h_{jlmp}(Z_{i+1})$$
$$A_2 := \kappa_{(i,n)}^{j,l,m}\kappa_{(i,n)}^{p,q,r}h_{jlmpqr}(Z_{i+1})$$

Avec ces notations, Q_4 devient

$$\frac{Q_4(x)}{n-i-1} = \frac{A_1}{24(n-i-1)} + \frac{A_2}{72(n-i-1)} \tag{4.71}$$

Etude de A_1. En utilisant les notations introduites dans la section 4.2.3,

$$h_{jlmp}(x) = x^{(j)}x^{(l)}x^{(m)}x^{(p)} - \kappa_{j,l}x^{(m)}x^{(p)}[6] + \kappa_{j,l}\kappa_{m,p}[3]$$

$$h_{jlmp}(Z_{i+1}) = \frac{U_{i+1}^{(j)}U_{i+1}^{(l)}U_{i+1}^{(m)}U_{i+1}^{(p)}}{(n-i-1)^2} - \kappa_{j,l}^{(i,n)}\frac{U_{i+1}^{(m)}U_{i+1}^{(p)}[6]}{n-i-1} + \kappa_{j,p}^{(i,n)}\kappa_{m,p}^{(i,n)}[3]$$

$$\frac{A_1}{24(n-i-1)} = \frac{1}{24}\sum_{j=1}^{d}\sum_{l=1}^{d}\sum_{m=1}^{d}\sum_{p=1}^{d}\kappa_{(i,n)}^{I}\left(\frac{U_{i+1}^{(j)}U_{i+1}^{(l)}U_{i+1}^{(m)}U_{i+1}^{(p)}}{(n-i-1)^3} - \kappa_{j,l}^{(i,n)}\frac{U_{i+1}^{(m)}U_{i+1}^{(p)}[6]}{(n-i-1)^2} + \frac{\kappa_{j,l}^{(i,n)}\kappa_{m,p}^{(i,n)}[3]}{n-i-1}\right)$$

avec $I = \{j, l, m, p\}$.

En utilisant le Lemme 51,

$$\frac{U_{i+1}^{(j)} U_{i+1}^{(l)} U_{i+1}^{(m)} U_{i+1}^{(p)}}{(n-i-1)^3} = \frac{O_{P_{na}}((\log n)^4)}{(n-i-1)^3}$$

et

$$\kappa_{j,l} \frac{U_{i+1}^{(m)} U_{i+1}^{(p)}[6]}{(n-i-1)^2} = \frac{O_{P_{na}}((\log n)^2)}{(n-i-1)^2}$$

Notons :

$$\delta_1^{(i,n)} := \frac{1}{8} \sum_{j=1}^{d} \sum_{m=1}^{d} \kappa_{(i,n)}^{j,m} \tag{4.72}$$

Finalement,

$$\frac{A_1}{24(n-i-1)} = \frac{\delta_1^{(i,n)}}{n-i-1} + \frac{O_{P_{na}}((\log n)^2)}{(n-i-1)^2}$$

Etude de A_2. En utilisant les notations introduites dans la section 4.2.3,

$$h_{jlmpqr}(Z_{i+1}) = \frac{U_{i+1}^{(j)} U_{i+1}^{(l)} U_{i+1}^{(m)} U_{i+1}^{(p)} U_{i+1}^{(q)} U_{i+1}^{(r)}}{(n-i-1)^3} - \frac{\kappa_{j,l}^{(i,n)} U_{i+1}^{(m)} U_{i+1}^{(p)} U_{i+1}^{(q)} U_{i+1}^{(r)}[15]}{(n-i-1)^2}$$

$$+ \frac{\kappa_{j,l}^{(i,n)} \kappa_{m,p}^{(i,n)} U_{i+1}^{(q)} U_{i+1}^{(r)}[45]}{n-i-1} - \kappa_{j,l}^{(i,n)} \kappa_{m,p}^{(i,n)} \kappa_{q,r}^{(i,n)}[15]$$

Ainsi, pour A_2 :

$$\frac{A_2}{72(n-i-1)} = \frac{1}{72} \sum_{j=1}^{d} \sum_{l=1}^{d} \sum_{m=1}^{d} \sum_{p=1}^{d} \sum_{q=1}^{d} \sum_{r=1}^{d} \kappa_{(i,n)}^{j,l,m} \kappa_{(i,n)}^{p,q,r}$$

$$\Big(\frac{U_{i+1}^{(j)} U_{i+1}^{(l)} U_{i+1}^{(m)} U_{i+1}^{(p)} U_{i+1}^{(q)} U_{i+1}^{(r)}}{(n-i-1)^4} - \frac{\kappa_{j,l}^{(i,n)} U_{i+1}^{(m)} U_{i+1}^{(p)} U_{i+1}^{(q)} U_{i+1}^{(r)}[15]}{(n-i-1)^3}$$

$$+ \frac{\kappa_{j,l}^{(i,n)} \kappa_{m,p}^{(i,n)} U_{i+1}^{(q)} U_{i+1}^{(r)}[45]}{(n-i-1)^2} - \frac{\kappa_{j,l}^{(i,n)} \kappa_{m,p}^{(i,n)} \kappa_{q,r}^{(i,n)}[15]}{n-i-1} \Big)$$

En appliquant une nouvelle fois le Lemme 51, on obtient :

$$\frac{U_{i+1}^{(j)} U_{i+1}^{(l)} U_{i+1}^{(m)} U_{i+1}^{(p)} U_{i+1}^{(q)} U_{i+1}^{(r)}}{(n-i-1)^4} = \frac{O_{P_{na}}((\log n)^6)}{(n-i-1)^4},$$

$$\frac{\kappa_{j,l}^{(i,n)} U_{i+1}^{(m)} U_{i+1}^{(p)} U_{i+1}^{(q)} U_{i+1}^{(r)}[15]}{(n-i-2)^3} = \frac{O_{P_{na}}((\log n)^4)}{(n-i-1)^3},$$

et

$$\frac{\kappa_{j,l}^{(i,n)} \kappa_{m,p}^{(i,n)} U_{i+1}^{(q)} U_{i+1}^{(r)}[45]}{(n-i-1)^2} = \frac{O_{P_{na}}((\log n)^2)}{(n-i-1)^2}.$$

Notons :

$$\delta_2^{(i,n)} := \frac{15}{72}\sum_{j=1}^{d}\sum_{m=1}^{d}\sum_{q=1}^{d}\kappa_{(i,n)}^{j,j,m}\kappa_{(i,n)}^{m,q,q} \tag{4.73}$$

Donc :

$$\frac{A_2}{72(n-i-1)} = -\frac{\delta_2^{(i,n)}}{n-i-1} + \frac{O_{P_{na}}((\log n)^2)}{(n-i-1)^2}$$

Conclusion sur le développement d'Edgeworth

Finalement, le développement de $\overline{\pi_{n-i-1}^{m_i}}$ s'écrit :

$$\overline{\pi_{n-i-1}^{m_i}}(Z_{i+1}) = \mathfrak{n}_d(Z_{i+1})\left(1 + \frac{1}{2(n-i-1)}\,{}^tU_{i+1}\gamma + \frac{\delta_1^{(i,n)} - \delta_2^{(i,n)}}{n-i-1} + \frac{O_{P_{na}}((\log n)^3)}{(n-i-1)^2}\right)$$

$$+O_{P_{na}}\left(\frac{1}{(n-i-1)^{3/2}}\right)$$

sous (E1) et (E2).

Une forme équivalente est obtenu pour le dénominateur :

$$\overline{\pi_{n-i}^{m_i}}(0) = \mathfrak{n}_d(0)\left(1 + \frac{\delta_1^{(i,n)} - \delta_2^{(i,n)}}{n-i}\right) + O_{P_{na}}\left(\frac{1}{(n-i)^{3/2}}\right) \tag{4.74}$$

Ainsi,

$$\frac{\overline{\pi_{n-i-1}^{m_i}}(Z_{i+1})}{\overline{\pi_{n-i}^{m_i}}(0)} = \frac{\mathfrak{n}_d(Z_{i+1})}{\mathfrak{n}_d(0)}\frac{1 + \frac{1}{2(n-i-1)}\,{}^tU_{i+1}\gamma + \frac{\delta_1^{(i,n)}-\delta_2^{(i,n)}}{n-i-1} + \frac{O_{P_{na}}((\log n)^3)}{(n-i-1)^2}}{1 + \frac{\delta_1^{(i,n)}-\delta_2^{(i,n)}}{n-i} + O_{P_{na}}\left(\frac{1}{(n-i)^{3/2}}\right)} \tag{4.75}$$

Après cette première étude, un changement de variable doit être effectué. γ, $\delta_1^{(i,n)}$ et $\delta_2^{(i,n)}$ dépendent implicitement du vecteur aléatoire U_{i+1}, il faut aussi changer de variable dans ces trois quantités. Dans $\delta_1^{(i,n)}$ et $\delta_2^{(i,n)}$, le changement de variable s'effectue sans faire d'autres commentaires (car ces termes ne feront pas partie des termes principaux).

Etudions le terme ${}^tU_{i+1}\gamma$. En utilisant les formules (4.69) et (4.70),

$${}^tU_{i+1}\gamma = {}^t\left(\kappa_{(i,n)}^{-1/2}(Y_{i+1} - m_{i,n})\right)\gamma$$

On rappelle les deux lemmes classiques (voir par exemple [52]).

Lemma 54. *Une matrice réelle définie positive de dimension $d \times d$ admet 2^n racines carrées, toutes symétriques réelles, dont une unique est définie positive.*

Lemma 55. *Toute matrice définie positive est inversible, et son inverse est elle aussi définie positive. De plus, si cette matrice est symétrique, alors son inverse l'est aussi.*

Ici, $\kappa_{(i,n)}$ est symétrique définie positive réelle de dimension $d \times d$ comme matrice de variance-covariance. Ainsi, en utilisant le Lemme 55, $\kappa_{(i,n)}^{-1}$ est symétrique définie positive et, en utilisant le Lemme 54, $\kappa_{(i,n)}^{-1/2}$ est symétrique. On choisit l'unique définie positive. ${}^tU_{i+1}\gamma$ devient alors :

$$^tU_{i+1}\gamma =^t (Y_{i+1} - m_i)\kappa_{(i,n)}^{-1/2}\gamma$$

et $\gamma = \kappa_{(i,n)}^{-3/2}\gamma_Y$.
Donc

$$^tU_{i+1}\gamma =^t (Y_{i+1} - m_i)\kappa_{(i,n)}^{-2}\gamma_Y$$

$$^tU_{i+1}\gamma =^t Y_{i+1}\kappa_{(i,n)}^{-2}\gamma_Y -^t m_i\kappa_{(i,n)}^{-2}\gamma_Y$$

$$\frac{^tU_{i+1}\gamma}{2(n-i-1)} = \frac{^tY_{i+1}\kappa_{(i,n)}^{-2}\gamma_Y}{2(n-i-1)} - \frac{^tm_i\kappa_{(i,n)}^{-2}\gamma_Y}{2(n-i-1)}$$

En utilisant les Lemmes 50 et 51, on obtient

$$\frac{^tU_{i+1}\gamma}{2(n-i-1)} = \frac{^tY_{i+1}\kappa_{(i,n)}^{-2}\gamma_Y}{2(n-i-1)} - \frac{^ta\kappa_{(i,n)}^{-2}\gamma_Y}{2(n-i-1)} + \frac{o_{P_{na}}(\epsilon_n)}{n-i-1} \tag{4.76}$$

Ainsi,

$$\frac{\overline{\pi_{n-i-1}^{m_{i,n}}}(Z_{i+1})}{\overline{\pi_{n-i}^{m_{i,n}}}(0)} = \frac{\mathfrak{n}_d(Z_{i+1})}{\mathfrak{n}_d(0)} \tag{4.77}$$

$$\frac{1 + \frac{^tY_{i+1}\kappa_{(i,n)}^{-2}\gamma_Y}{2(n-i-1)} - \frac{^ta_n\kappa_{(i,n)}^{-2}\gamma_Y}{2(n-i-1)} + \frac{o_{P_{na}}(\epsilon_n)}{n-i-1} + \frac{\delta_1^{(i,n)}-\delta_2^{(i,n)}}{n-i-1} + \frac{O_{P_{na}}((\log n)^3)}{(n-i-1)^2}}{1 + \frac{\delta_1^{(i,n)}-\delta_2^{(i,n)}}{n-i} + O_{P_{na}}(\frac{1}{(n-i)^{3/2}})}$$

Notons $C := \frac{\mathfrak{n}_d(Z_{i+1})}{\mathfrak{n}_d(0)}$

Développement de Taylor de C

Effectuons un développement de Taylor dans $\mathbb{R}^d$, de $\frac{\mathfrak{n}_d(Z_{i+1})}{\mathfrak{n}_d(0)}$. On rappelle les notations $Z_{i+1} = -\frac{U_{i+1}}{\sqrt{n-i-1}}$ et $U_{i+1} = \kappa_{(i,n)}^{-1/2}(Y_{i+1} - m_i)$. Ainsi,

$$Z_{i+1} = -\frac{\kappa_{(i,n)}^{-1/2}Y_{i+1}}{\sqrt{n-i-1}} + \frac{\kappa_{(i,n)}^{-1/2}m_i}{\sqrt{n-i-1}}$$

Lemma 56. *Soient f une fonction de $\mathbb{R}^d$ dans $\mathbb{R}$ et $\alpha \in \mathbb{R}^d$.*

$$f(x) = f(\alpha) + \nabla f(\alpha)(x-\alpha) + \frac{1}{2}{}^t(x-\alpha)H_f(\alpha)(x-a) + o(\|x-\alpha\|^2) \tag{4.78}$$

où H_f est la matrice Hessienne de f et ∇f son gradient.

Ici, $f(x) := \exp(-\frac{1}{2} < x, x >)$.
Soit $j \in \{1, ..., d\}$. Ainsi,

$$\nabla f(x) = -^t x f(x)$$

et

$$H_f(x) = (-I_d + x^t x) f(x)$$

Finalement,

$$f(x) = f(\alpha)(1 - {}^t\alpha(x-\alpha) + \frac{1}{2}{}^t(x-\alpha)H_f(\alpha)(x-\alpha)) + o(\|x-\alpha\|^2)$$

Donc,

$$f(x) = f(\alpha)(1 - {}^t\alpha(x-\alpha) - \frac{1}{2}{}^t(x-\alpha)I_d(x-\alpha) + \frac{1}{2}{}^t(x-\alpha)\alpha^t\alpha(x-\alpha)) + o(\|x-a\|^2) \quad (4.79)$$

Dans notre cas, $x = Z_{i+1}$ et $\alpha := \kappa_{(i,n)}^{-1/2} Y_{i+1}$. Etudions les termes principaux de (4.79). En utilisant les Lemmes 50 et 51 :

$$\begin{aligned} -{}^t\alpha(x-\alpha) &= -{}^t(-\frac{\kappa_{(i,n)}^{-1/2} Y_{i+1}}{\sqrt{n-i-1}})(\frac{\kappa_{(i,n)}^{-1/2} m_i}{\sqrt{n-i-1}}) \\ &= \frac{{}^t Y_{i+1}\kappa_{(i,n)}^{-1} m_i}{n-i-1} \\ &= \frac{{}^t Y_{i+1}\kappa_{(i,n)}^{-1} a}{n-i-1} + \frac{o_{P_{na}}(\epsilon_n(\log n))}{n-i-1} \end{aligned}$$

Pour le deuxième terme, en utilisant les Lemmes 50 et 51,

$$\begin{aligned} -{}^t(x-\alpha)I_d(x-\alpha) &= -{}^t(x-\alpha)(x-\alpha) \\ &= -{}^t(\frac{\kappa_{(i,n)}^{-1/2} m_i}{\sqrt{n-i-1}})(\frac{\kappa_{(i,n)}^{-1/2} m_i}{\sqrt{n-i-1}}) \\ &= -\frac{1}{n-i-1}{}^t m_i \kappa_{(i,n)}^{-1} m_i \\ &= -\frac{1}{n-i-1}{}^t a \kappa_{(i,n)}^{-1} a + \frac{o_{P_{na}}((\epsilon_n)^2)}{n-i-1}. \end{aligned}$$

Finalement, le dernier terme s'écrit, en utilisant les Lemmes 50 et 51,

$$\begin{aligned} {}^t(x-\alpha)\alpha^t\alpha(x-\alpha) &= {}^t({}^t\alpha(x-\alpha))({}^t\alpha(x-\alpha)) \\ &= ({}^t\alpha(x-\alpha))^2 \\ &= (\frac{{}^t Y_{i+1}\kappa_{(i,n)}^{-1} m_i}{n-i-1})^2 \\ &= \frac{o_{P_{na}}((\epsilon_n \log n)^2)}{(n-i-1)^2}. \end{aligned}$$

On obtient sous (E1) et (E2) :

$$f(Z_{i+1}) = f(-\frac{\kappa_{(i,n)}^{-1/2} Y_{i+1}}{\sqrt{n-i-1}})(1 + \frac{{}^t Y_{i+1}\kappa_{(i,n)}^{-1} a}{n-i-1} - \frac{{}^t a \kappa_{(i,n)}^{-1} a}{2(n-i-1)} + \frac{o_{P_{na}}(\epsilon_n(\log n))}{n-i-1}) \quad (4.80)$$

Résultat Final

Les deux résultats principaux obtenus ((4.77) et (4.80)) sont alors réunis

$$\frac{\overline{\pi_{n-i-1}^{m_i}}(Z_{i+1})}{\overline{\pi_{n-i}^{m_i}}(0)} = \exp\{-\frac{{}^tY_{i+1}\kappa_{(i,n)}^{-1}Y_{i+1}}{2(n-i-1)}\}$$

$$\frac{(1+\frac{{}^tY_{i+1}\kappa_{(i,n)}^{-2}\gamma_Y}{2(n-i-1)}-\frac{{}^ta\kappa_{(i,n)}^{-2}\gamma_Y}{2(n-i-1)}+\frac{\delta_1^{(i,n)}-\delta_2^{(i,n)}}{n-i-1}+\frac{O_{P_{na}}((\log n)^3)}{(n-i-1)^2})(1+\frac{{}^tY_{i+1}\kappa_{(i,n)}^{-1}a}{n-i-1}-\frac{{}^ta\kappa_{(i,n)}^{-1}a}{2(n-i-1)}+\frac{o_{P_{na}}(\epsilon_n(\log n))}{n-i-1})}{1+\frac{\delta_1^{(i,n)}-\delta_2^{(i,n)}}{n-i}+O_{P_{na}}(\frac{1}{(n-i)^{3/2}})}$$

Pour simplifier les notations, on note $\delta^{(i,n)} = \delta_1^{(i,n)} - \delta_2^{(i,n)}$ et $\gamma_Y = \gamma$. Sous (E1) et (E2), on obtient

$$\frac{\overline{\pi_{n-i-1}^{m_i}}(Z_{i+1})}{\overline{\pi_{n-i}^{m_i}}(0)} = \exp\{-\frac{{}^tY_{i+1}\kappa_{(i,n)}^{-1}Y_{i+1}}{2(n-i-1)}\}$$

$$\frac{1+\frac{{}^tY_{i+1}\kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)}-\frac{{}^ta\kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)}+\frac{\delta^{(i,n)}}{n-i-1}+\frac{{}^tY_{i+1}\kappa_{(i,n)}^{-1}a}{n-i-1}-\frac{{}^ta\kappa_{(i,n)}^{-1}a}{2(n-i-1)}+\frac{o_{P_{na}}(\epsilon_n(\log n))}{n-i-1}}{1+\frac{\delta^{(i,n)}}{n-i}+O_{P_{na}}(\frac{1}{(n-i)^{3/2}})}$$

Notons :

$$u_1 = \frac{{}^tY_{i+1}\kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)} - \frac{{}^ta\kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)} + \frac{\delta^{(i,n)}}{n-i-1} + \frac{{}^tY_{i+1}\kappa_{(i,n)}^{-1}a}{n-i-1} - \frac{{}^ta\kappa_{(i,n)}^{-1}a}{2(n-i-1)} + \frac{o_{P_{na}}(\epsilon_n(\log n))}{n-i-1} \quad (4.81)$$

et

$$u_2 = \frac{\delta^{(i,n)}}{n-i} + O_{P_{na}}(\frac{1}{(n-i)^{3/2}}). \quad (4.82)$$

Un développement du deuxième ordre au numérateur et au dénominateur est effectué, ce qui permet d'obtenir

$$p(\mathbf{X}_{i+1} = Y_{i+1}|\mathbf{S}_{i+1,n} = na - S_{1,i}) = L_iC_i$$

$$\exp\{{}^tY_{i+1}(t_i + \frac{\kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)})\}\exp\{-\frac{{}^t(Y_{i+1}-a)\kappa_{(i,n)}^{-1}(Y_{i+1}-a)}{2(n-i-1)}\}\exp\{o_{P_{na}}(\frac{\epsilon_n(\log n)}{n-i-1})\}A(i)$$

avec

$$A(i) := \frac{\exp\{\frac{u_1^2}{2}+o(u_1^2)\}}{\exp\{O_{P_{na}}(\frac{1}{(n-i-1)^2})+\frac{u_2^2}{2}+o(u_2^2)\}}\exp\{\frac{\delta^{(i,n)}}{n-i-1}-\frac{\delta^{(i,n)}}{n-i}\}$$

et$L_i := \frac{\sqrt{n-i}}{\sqrt{n-i-1}}\frac{C_i^{-1}}{\Phi(t_{i,n})}\exp\{{}^ta\frac{\kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)}\}$

Ainsi

$$p(\mathbf{X}_1^k = Y_1^k|\mathbf{S}_{1,n} = na) = g_0(Y_1|Y_0)\prod_{i=1}^{k-1}g(Y_{i+1}|Y_1^i)\prod_{i=0}^{k-1}A(i)\prod_{i=0}^{k-1}L(i) \quad (4.83)$$

Il reste à prouver deux choses :

1. $\prod_{i=0}^{k-1} A(i) = 1 + o_{P_{na}}(\epsilon_n(\log n)^2)$
2. $\prod_{i=0}^{k-1} L(i) = 1 + o_{P_{na}}(\epsilon_n(\log n)^2)$

Etude de $L(i)$

On rappelle l'expression de C_i^{-1} :

$$C_i^{-1} = \int \exp\{\frac{{}^t x(t_i + \kappa_{(i,n)}^{-2}\gamma)}{2(n-i-1)}\} \exp\{-\frac{{}^t(x-a)\kappa_{(i,n)}^{-1}(x-a)}{2(n-i-1)}\} p(x)dx \tag{4.84}$$

Notons

$$u_x := \frac{{}^t x\kappa_{(i,n)}^{-2}\gamma}{2\,(n-i-1)} + \frac{{}^t(x-a)\kappa_{(i,n)}^{-1}(x-a)}{2\,(n-i-1)}.$$

On utilise alors les bornes classiques suivantes pour u_x :

$$1 - u + \frac{u_x^2}{2} - \frac{u_x^3}{6} \leq e^{-u_x} \leq 1 - u_x + \frac{u_x^2}{2}$$

pour obtenir des deux cotés de l'inégalités précédentes une approximation d'ordre 2 de C_i^{-1} en intégrant selon p. La borne supérieure donne alors :

$$\begin{aligned} C_i^{-1} &= \int_{\mathbb{R}^d} \exp\{< x, t_i >\} \exp\{-u_x\} p(x)dx \\ &\leq \int_{\mathbb{R}^d} \exp\{< x, t_i >\}(1 - u_x + \frac{u_x^2}{2}) p(x)dx \\ \leq \Phi(t_i) - \int_{\mathbb{R}^d} u_x \exp\{< x, t_i >\} p(x)dx &+ \int_{\mathbb{R}^d} \frac{u_x^2}{2} \exp\{< x, t_i >\} p(x)dx \end{aligned}$$

$$\begin{aligned} \leq \Phi(t_i)(1 + \int_{\mathbb{R}^d} &\frac{{}^t x\kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)} \frac{\exp\{< x, t_i >\} p(x)}{\Phi(t_i)} dx \\ - \int_{\mathbb{R}^d} &\frac{{}^t(x-a)\kappa_{(i,n)}^{-1}(x-a)}{2(n-i-1)} \frac{\exp\{< x, t_i >\} p(x)}{\Phi(t_i)} \\ &+ O_{P_{na}}\left(\frac{1}{(n-i-1)^2}\right)) \end{aligned}$$

où l'approximation est uniforme en Y_1^k.
On étudie les termes principaux de ce développement.

Remark 57. *Soit $v := {}^t(u^{(1)}, ..., u^{(d)})$ et $M = (b_{i,j})_{1 \leq i,j \leq d}$. Alors*

$${}^t vMv = \sum_{j=1}^{d} \sum_{k=1}^{d} v^{(j)} v^{(k)} b_{j,k}$$

Etudions le premier terme.
On a par définition, pour tout j et k entre 1 et d :

$$m_i^{(j)} = \frac{1}{\Phi(t_i)} \int_{\mathbb{R}^d} x^{(j)} \exp\{< x, t_i >\} p(x)dx$$

et

$$\kappa_{(i,n)}^{j,k} = \frac{1}{\Phi(t_i)} \int_{\mathbb{R}^d} x^{(j)} x^{(k)} \exp\{< x, t_i >\} p(x) dx$$

Si, de plus, on note (momentanément) $\alpha := {}^t(\alpha^{(1)}, ..., \alpha^{(d)}) = \kappa_{(i,n)}^{-2}\gamma$
Ainsi

$$\begin{aligned}
&\int_{\mathbb{R}^d} \frac{{}^t x \kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)} \frac{\exp\{< x, t_i >\} p(x)}{\Phi(t_i)} dx \\
&= \sum_{j=1}^{d} \alpha^{(j)} \int_{\mathbb{R}^d} \frac{x^{(j)}}{2(n-i-1)} \frac{\exp\{< x, t_i >\} p(x)}{\Phi(t_i)} dx \\
&= \frac{1}{2(n-i-1)} \sum_{j=1}^{d} \alpha^{(j)} m_i^{(j)} \\
&= \frac{{}^t m_i \kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)}.
\end{aligned}$$

En effectuant un calcul identique dans le second terme, on obtient :

$$\begin{aligned}
&- \int_{\mathbb{R}^d} \frac{{}^t(x-a)\kappa_{(i,n)}^{-1}(x-a)}{2(n-i-1)} \frac{exp\{< x, t_i >\} p(x)}{\Phi(t_i)} \\
&= -\frac{1}{2(n-i-1)} (1 + {}^t(m_i - a)\kappa_{(i,n)}^{-1}(m_i - a))
\end{aligned}$$

Nous obtenons finalement pour la borne supérieure :

$$C_i^{-1} \leq \Phi(t_i) \left(1 + \frac{{}^t m_i \kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)} - \frac{1}{2(n-i-1)} - \frac{{}^t(m_i - a)\kappa_{(i,n)}^{-1}(m_i - a)}{2(n-i-1)} + O_{P_{na}}\left(\frac{1}{(n-i-1)^2} \right) \right)$$

En utilisant le lemme 50, l'égalité précédente devient :

$$C_i^{-1} \leq \Phi(t_i) \left(1 + \frac{{}^t a \kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)} - \frac{1}{2(n-i-1)} + \frac{o_{P_{na}}(\epsilon_n)}{n-i-1} \right) \tag{4.85}$$

Cette égalité nous permet d'obtenir :

$$L_i \leq \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \exp\{\frac{{}^t a \kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)}\} (1 + \frac{{}^t a \kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)} - \frac{1}{2(n-i-1)} + \frac{o_{P_{na}}(\epsilon_n)}{n-i-1}) \tag{4.86}$$

En remplaçant $\frac{\sqrt{n-i}}{\sqrt{n-i-1}}$ et $\exp\left(-\frac{{}^t a \kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)} \right)$ par leurs développements de Taylor respectifs $1 + \frac{1}{2(n-i-1)} + O\left(\frac{1}{(n-i-1)^2}\right)$ and $1 - \frac{{}^t a \kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)} + O\left(\frac{||a||^2}{(n-i-1)^2}\right)$ dans la borne supérieure de L_i ci-dessus, nous avons alors :

$$\begin{aligned}
L_i \leq \ & \left(1 + \frac{1}{2(n-i-1)} + O\left(\frac{1}{(n-i-1)^2} \right) \right) \left(1 - \frac{{}^t a \kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)} + O\left(\frac{||a||^2}{(n-i-1)^2} \right) \right) \\
& (1 + \frac{{}^t a \kappa_{(i,n)}^{-2}\gamma}{2(n-i-1)} - \frac{1}{2(n-i-1)} + \frac{o_{P_{na}}(\epsilon_n)}{n-i-1})
\end{aligned}$$

Ecrivons

$$\prod_{i=1}^{k} L_i \leq \prod_{i=1}^{k} (1 + M_i)$$

avec

$$M_i = -\frac{({}^t a \kappa_{(i,n)}^{-2} \gamma)^2}{4(n-i-1)^2} + \frac{o_{P_{na}}(\epsilon_n)}{n-i-1}.$$

Sous (E1) and (E2), $\sum_{i=0}^{k-1} M_i$ est $o_{P_{na}}\left(\epsilon_n (\log n)^2\right)$.

Etude de $A(i)$

On rappelle les notations définies précedemment :

$$u_1 = \frac{{}^t Y_{i+1} \kappa_{(i,n)}^{-2} \gamma}{2(n-i-1)} - \frac{{}^t a \kappa_{(i,n)}^{-2} \gamma}{2(n-i-1)} + \frac{\delta^{(i,n)}}{n-i-1} + \frac{{}^t Y_{i+1} \kappa_{(i,n)}^{-1} a}{n-i-1} - \frac{{}^t a \kappa_{(i,n)}^{-1} a}{2(n-i-1)} + \frac{o_{P_{na}}(\epsilon_n (\log n))}{n-i-1}$$

et

$$u_2 = \frac{\delta^{(i,n)}}{n-i} + O_{P_{na}}(\frac{1}{(n-i)^{3/2}}).$$

avec

$$A(i) := \frac{\exp\{\frac{u_1^2}{2} + o(u_1^2)\}}{\exp\{O_{P_{na}}(\frac{1}{(n-i-1)^2}) + \frac{u_2^2}{2} + o(u_2^2)\}} \exp\{\frac{\delta^{(i,n)}}{n-i-1} - \frac{\delta^{(i,n)}}{n-i}\}$$

On veut prouver $\prod_{i=0}^{k-1} A(i) = 1 + o_{P_{na}}(\epsilon_n (\log n)^2)$.
Pour tout nombre positif β_j pour $j \in \{1, ..., 17\}$, on note :

$$A_n^1 := \left\{ \frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{\delta^{(i,n)}}{(n-i-1)(n-i)} \right| < \beta_1 \right\}$$

$$A_n^2 := \left\{ \frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{(\delta^{(i,n)})^2}{(n-i)^2} \right| < \beta_2 \right\}$$

$$A_n^3 := \left\{ \frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{(\delta^{(i,n)})^2}{(n-i-1)^2} \right| < \beta_3 \right\}$$

$$A_n^4 := \left\{ \frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{({}^t a \kappa_{(i,n)}^{-2} \gamma)^2}{(n-i-1)^2} \right| < \beta_4 \right\}$$

$$A_n^5 := \left\{ \frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{({}^t a \kappa_{(i,n)}^{-1} a)^2}{(n-i-1)^2} \right| < \beta_5 \right\}$$

$$A_n^6 := \left\{ \frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{\delta^{(i,n)} ({}^t a \kappa_{(i,n)}^{-2} \gamma)}{(n-i-1)^2} \right| < \beta_6 \right\}$$

$$A_n^7 := \left\{ \frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{\delta^{(i,n)} ({}^t a \kappa_{(i,n)}^{-1} a)}{(n-i-1)^2} \right| < \beta_7 \right\}$$

$$A_n^8 := \left\{ \frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{({}^t a \kappa_{(i,n)}^{-2} \gamma)({}^t a \kappa_{(i,n)}^{-1} a)}{(n-i-1)^2} \right| < \beta_8 \right\}$$

$$A_n^9 := \left\{ \frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{({}^t Y_{i+1} \kappa_{(i,n)}^{-2} \gamma)^2}{(n-i-1)^2} \right| < \beta_9 \right\}$$

$$A_n^{10} := \left\{ \frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{({}^t Y_{i+1} \kappa_{(i,n)}^{-1} a)^2}{(n-i-1)^2} \right| < \beta_{10} \right\}$$

$$A_n^{11} := \left\{ \frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{({}^t Y_{i+1} \kappa_{(i,n)}^{-2} \gamma)({}^t a \kappa_{(i,n)}^{-2} \gamma)}{(n-i-1)^2} \right| < \beta_{11} \right\}$$

$$A_n^{12} := \left\{ \frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{\delta^{(i,n)}({}^t Y_{i+1} \kappa_{(i,n)}^{-2} \gamma)}{(n-i-1)^2} \right| < \beta_{12} \right\}$$

$$A_n^{13} := \left\{ \frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{({}^t Y_{i+1} \kappa_{(i,n)}^{-2} \gamma)({}^t a \kappa_{(i,n)}^{-1} a)}{(n-i-1)^2} \right| < \beta_{13} \right\}$$

$$A_n^{14} := \left\{ \frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{({}^t Y_{i+1} \kappa_{(i,n)}^{-1} a)({}^t a \kappa_{(i,n)}^{-2} \gamma)}{(n-i-1)^2} \right| < \beta_{14} \right\}$$

$$A_n^{15} := \left\{ \frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{\delta^{(i,n)}({}^t Y_{i+1} \kappa_{(i,n)}^{-1} a)}{(n-i-1)^2} \right| < \beta_{15} \right\}$$

$$A_n^{16} := \left\{ \frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{({}^t Y_{i+1} \kappa_{(i,n)}^{-1} a)({}^t a \kappa_{(i,n)}^{-1} a)}{(n-i-1)^2} \right| < \beta_{16} \right\}$$

$$A_n^{17} := \left\{ \frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{({}^t Y_{i+1} \kappa_{(i,n)}^{-1} a)({}^t Y_{i+1} \kappa_{(i,n)}^{-2} \gamma)}{(n-i-1)^2} \right| < \beta_{17} \right\}$$

On peut séparer l'étude de ces dix-sept ensembles en trois groupes. Pour chacun de ces groupes, nous allons étudier un exemple. Les autres se traitant de manière équivalente. Pour tout $j \in \{1, ..., 8\}$, on a immédiatement que,

$$\lim_{n \to \infty} P_{na} \left(A_n^j \right) = 1$$

Le deuxième groupe d'ensembles (de A_n^9 à A_n^{16}) dépendent de Y_{i+1}. Etudions A_n^{16}. Pour démontrer que $\lim_{n\to\infty} P_{na} \left(A_n^{16} \right) = 1$, le Théorème 5 doit être utilisé. Réécrivons la somme définie dans A_n^{16}.

$$\begin{aligned} |({}^t Y_{i+1} \kappa_{(i,n)}^{-1} a)({}^t a \kappa_{(i,n)}^{-1} a)| &= | < Y_{i+1}, \kappa_{(i,n)}^{-1} a >< , \kappa_{(i,n)}^{-1} a > | \\ &\leq ||Y_{i+1}|| ||\kappa_{(i,n)}^{-1} a|| ||a|| ||\kappa_{(i,n)}^{-1} a|| \\ &\leq \lambda_1^2 ||Y_{i+1}|| ||a||^3 \end{aligned}$$

où λ_1 est la plus grande valeur singulière de $\kappa_{(i,n)}^{-1}$.

Ainsi, on peut appliquer le théorème 5 dans le cas où $X_{i,n} := ||Y_{i+1}||$ et $a_{i,n} := \frac{||a||^3}{\epsilon_n (\log n)^2 (n-i-1)^2}$.

Vérifions les hypothèses de ce théorème.

1

$$E[X_{1,n}^2] = E[||Y_1||^2] = E[\sum_{j=1}^{d} [Y_1^{(j)}]^2] = \sum_{j=1}^{d} E[[Y_1^{(j)}]^2].$$

En utilisant le Lemme 49, on obtient :

$$E[||Y_1||^2] = \sum_{j=1}^{d} (\kappa_{j,j} + [a^{(j)}]^2) = tr(\kappa_{(i,n)}) + ||a||^2.$$

Ainsi il existe Γ tel que $E[X_{1,n}^2] \leq \Gamma$.

2

$$\rho_n := E[X_{1,n} X_{2,n}] \leq \sqrt{E[X_{1,n}^2] E[X_{1,n}^2]} = \sqrt{E[||Y_1||^2] E[||Y_2||^2]}$$

Ainsi

$$E[||Y_1||^2 ||Y_2||^2] = E[(\sum_{j=1}^{d} [Y_1^{(j)}]^2)(\sum_{j=1}^{d} [Y_2^{(j)}]^2)]$$

or, comme démontré ci-dessus,

$$E[||Y_1||^2] = tr(\kappa_{(i,n)}) + ||a||^2.$$

Donc,

$$\rho_n \leq tr(\kappa_{(i,n)}) + ||a||^2$$

3 Nous pouvons à présent vérifier les deux dernières hypothèses portant sur $a_{i,n}$.

$$\lim_{n\to\infty} \sum_{i=0}^{k-1} a_{i,n}^2 = \lim_{n\to\infty} \frac{||a||^6}{\epsilon_n^2 (\log n)^4 (n-k)^3} = 0 \tag{4.87}$$

sous (E1). De même, sous (E1), la deuxième condition s'écrit :

$$\lim_{n\to\infty} \rho_n \left(\sum_{i=0}^{k-1} a_{i,n} \right)^2 \leq \lim_{n\to\infty} \frac{\rho_n ||a||^6}{\epsilon_n^2 (\log n)^4 (n-k)^2} = 0 \tag{4.88}$$

Ainsi, le théorème 5 permet de conclure que $\lim_{n\to\infty} P_{na}\left(A_n^{16}\right) = 1$.

Etudions pour finir A_n^{17} en suivant les mêmes étapes que précédemment.

$$A_n^{17} := \left\{ \frac{1}{\epsilon_n (\log n)^2} \sum_{i=0}^{k-1} \left| \frac{({}^t Y_{i+1} \kappa_{(i,n)}^{-1} a)({}^t Y_{i+1} \kappa_{(i,n)}^{-2} \gamma)}{(n-i-1)^2} \right| < \beta_{17} \right\}$$

On réécrit la somme définie dans A_n^{17}.

$$|({}^t Y_{i+1} \kappa_{(i,n)}^{-1} a)({}^t Y_{i+1} \kappa_{(i,n)}^{-2} \gamma)| \leq \lambda_1^3 ||Y_{i+1}||^2 ||a||$$

où λ_1 est la plus grande valeur singulière de $\kappa_{(i,n)}^{-1}$.

Ainsi, nous voulons appliquer le théorème 5 dans le cas où $X_{i,n} := ||Y_{i+1}||^2$ et $a_{i,n} := \frac{||a||}{\epsilon_n (\log n)^2 (n-i-1)^2}$.

1 Par le Lemme 49, Il est clair qu'il existe une constante strictement positive Γ telle que $E[X_{1,n}^2] \leq \Gamma$.

2

$$\rho_n = E[X_{1,n}X_{2,n}] = E[||Y_1||^2||Y_2||^2] = \sum_{j=1}^{d} E[[Y_1^{(j)}]^2[Y_2^{(j)}]^2] + \sum_{j=1}^{d}\sum_{l\neq j} E[[Y_1^{(j)}]^2[Y_2^{(l)}]^2] \quad (4.89)$$

En utilisant des arguments équivalents au Lemme 49, on peut montrer qu'il existe une constante Γ telle que

$$E[[Y_1^{(j)}]^2[Y_2^{(k)}]^2] =\leq \Gamma$$

3 Sous (E1),

$$\lim_{n\to\infty} \sum_{i=0}^{k-1} a_{i,n}^2 = \lim_{n\to\infty} \frac{||a||^2}{\epsilon_n^2 (\log n)^4 (n-k)^3} = 0 \quad (4.90)$$

et

$$\lim_{n\to\infty} \rho_n \left(\sum_{i=0}^{k-1} a_{i,n}\right)^2 \leq \lim_{n\to\infty} \frac{\rho_n ||a||^2}{\epsilon_n^2 (\log n)^4 (n-k)^2} = 0 \quad (4.91)$$

Ainsi, le théorème 5 permet de conclure que $\lim_{n\to\infty} P_{na}\left(A_n^{16}\right) = 1$.
En notant $A_n := \bigcup_{j=1}^{17} A_n^j$, nous avons $\lim_{n\to\infty} P_{na}\left(A_n\right) = 1$.
Ainsi, $\prod_{i=0}^{k-1} A(i) = 1 + o_{P_{na}}(\epsilon_n (\log n)^2)$.

Annexe A

Exemples de simulation

Nous proposons dans cette annexe une liste d'exemples permettant d'utiliser les algorithmes énoncés dans la partie principale de la thèse. Nous donnons la forme de la densité tiltée associée et les paramètres α et β de la densité approximante g_{na}.

On rappelle briévement la forme de g_{na} dans la cas réel puis dans le cas d-dimensionnel quand l'évènement conditionnant s'écrit $(\mathbf{U}_{1,n} = na)$. Sous les hypothèses et notations du chapitre 1,

$$g_{na}(y_1^k) = g_0(y_1|y_0) \prod_{i=1}^{k-1} g(y_{i+1}|y_1^i)$$

avec

$$g(y_{i+1}|y_1^i) = C_i \mathfrak{n}(\alpha\beta + a, \beta, u\,(y_{i+1}))p_{\mathbf{X}}(y_{i+1})$$

et

$$\alpha := t_i + \frac{\mu_3^i}{2s_i^4(n-i-1)}$$
$$\beta = (n-i-1)s_i^2$$

Sous les hypothèses et notations du chapitre 4,

$$g_{na}(y_1^k) = g_0(y_1|y_0) \prod_{i=1}^{k-1} g(y_{i+1}|y_1^i)$$

avec

$$g(y_{i+1}|y_1^i) = C_i \mathfrak{n}_d(\alpha\beta + a, \beta, u\,(y_{i+1}))p_{\mathbf{X}}(y_{i+1})$$

et

$$\alpha := t_i + \frac{\kappa_{(i,n)}^{-2}\gamma_{(i,n)}}{2(n-i-1)}$$
$$\beta = (n-i-1)\kappa_{(i,n)}$$

A.1 Loi Normale.

Soit $\mathbf{X}_1^n$ de loi $\mathcal{N}(\mu, \sigma^2)$.

Conditionnement par $(\mathbf{S}_{1,n} = na)$**.** La densité tiltée $\pi_{\mathbf{X}}^{m_i}$ est

$$\mathcal{N}(m_i, \sigma^2).$$

Les paramètres de $g(y_{i+1}|y_1^i)$ sont

$$\alpha = \frac{m_i - \mu}{\sigma^2}$$

et

$$\beta = (n - i - 1)\sigma^2.$$

On peut aussi dans ce cas précis définir $g(y_{i+1}|y_1^i)$ comme la densité

$$\mathcal{N}(\mu_*, \sigma_*^2)$$

avec

$$\mu_* = \frac{n - i - 1}{n - i} m_i$$

et

$$\sigma_*^2 = \frac{n - i - 1}{n - i} \sigma^2.$$

Conditionnement par $(\mathbf{U}_{1,n} = na)$ **avec** $u(x) = x^2$ La densité tiltée $\pi_u^{m_i}$ est

$$\mathcal{N}(\frac{\mu}{1 - 2t_i\sigma^2}, \frac{\sigma^2}{1 - 2t_i\sigma^2})$$

Pour déterminer t_i, on résoud en t l'équation suivante

$$\sigma^2 \left(1 - 2t\sigma^2\right) + \mu^2 - m_i \left(1 - 2t\sigma^2\right)^2 = 0.$$

Les paramètres de $g(y_{i+1}|y_1^i)$ sont

$$\alpha = \left(t_i + \frac{\mu_3^i}{2s_i^2(n - i - 1)}\right)$$

et

$$\beta = s_i^2(n - i - 1)$$

avec

$$\mu_3^i = \frac{8\sigma^6}{(1 - 2t_i\sigma^2)^3} + \frac{24\sigma^4\mu^2}{(1 - 2t_i\sigma^2)^4}$$

et

$$s_i^2 = \frac{2\sigma^4}{(1 - 2t_i\sigma^2)^2} + \frac{4\sigma^2\mu^2}{(1 - 2t_i\sigma^2)^3}.$$

Conditionnement par $(\mathbf{U}_{1,n} = na)$ **avec** $u(x) :=^t (x, x^2)$. La densité tiltée $\pi_u^{m_i}$ est

$$\mathcal{N}\left(m_i^{(1)}, m_i^{(2)} - [m_i^{(1)}]^2\right).$$

Les paramètres de $g(y_{i+1}|y_1^i)$ sont

$$\begin{aligned}\alpha &:= t_i + \frac{\kappa_{(i,n)}^{-2}\gamma_{(i,n)}}{2(n-i-1)}\\ \beta &= (n-i-1)\kappa_{(i,n)}\end{aligned}$$

avec

$$\kappa_{(i,n)} = \begin{pmatrix} m_i^{(2)} - \left(m_i^{(1)}\right)^2 & 2m_i^{(1)}\left(m_i^{(2)} - \left(m_i^{(1)}\right)^2\right) \\ 2m_i^{(1)}\left(m_i^{(2)} - \left(m_i^{(1)}\right)^2\right) & 2\left(m_i^{(2)} - \left(m_i^{(1)}\right)^2\right)\left(m_i^{(2)} + m_i^{(1)}\right)\end{pmatrix}$$

et

$$\gamma_{(i,n)} = \begin{pmatrix} 8m_i^{(1)}\left(m_i^{(2)} - \left(m_i^{(1)}\right)^2\right) \\ 2\left(m_i^{(2)} - \left(m_i^{(1)}\right)^2\right)\left(4\left(m_i^{(2)} - \left(m_i^{(1)}\right)^2\right)\left(m_i^{(2)} - 2\left(m_i^{(1)}\right)^2\right) + 1\right)\end{pmatrix}$$

et $t_i :=^t \left(t_i^{(1)}, t_i^{(2)}\right)$ solution de

$$\begin{aligned} m_i^{(1)} &= \frac{\mu + \sigma^2 t_i^{(1)}}{1 - 2\sigma^2 t_i^{(2)}}\\ m_i^{(2)} &= \frac{\sigma^2}{1 - 2\sigma^2 t_i^{(2)}} + \left(m_i^{(1)}\right)^2\end{aligned}$$

A.2 Loi Gamma

Soit $\mathbf{X}_1^n$ de loi $\Gamma(\theta, \eta)$.

On définit les fonctions digamma, trigamma et polygamma d'ordre m pour $m > 2$ par :

$$\begin{aligned} digamma(x) := \psi(x) &= \frac{d}{dx}\ln\Gamma(x)\\ trigamma(x) &:= \frac{d}{dx}\psi(x)\\ psigamma(x, m) &:= \frac{d^m}{dx^m}\psi(x)\end{aligned}$$

Conditionnement par $(\mathbf{S}_{1,n} = na)$. La densité tiltée $\pi_{\mathbf{X}}^{m_i}$ est

$$\Gamma(\theta, \frac{\theta}{m_i})$$

Les paramètres de $g(y_{i+1}|y_1^i)$ sont

$$\beta = \frac{m_i^2}{\theta}(n-i-1).$$

et

$$\alpha = \eta - \frac{\theta}{m_i} + \frac{1}{m_i(n-i-1)}$$

Conditionnement par $(\mathbf{U}_{1,n} = na)$ **avec** $u(x) = \log(x)$. La densité tiltée $\pi_u^{m_i}$ est

$$\Gamma(\theta + t_i, \eta)$$

Pour déterminer t_i, on résoud en t l'équation suivante

$$digamma(t+\theta) - \log(\eta) - m_i = 0.$$

Les paramètres de $g(y_{i+1}|y_1^i)$ sont

$$\beta = (n-i-1) * trigamma(t_i + \theta)$$
$$\alpha = t_i + \frac{psigamma(t_i+\theta, 2)}{2(n-i-1)(trigamma(t_i+\theta))^2}$$

Conditionnement par $(\mathbf{U}_1^n = na)$ **avec** $u(x) =^t (x, \log(x))$. La densité tiltée $\pi_u^{m_i}$ est

$$\Gamma(\theta + t_i^{(2)}, \eta - t_i^{(1)}).$$

Pour déterminer $t_i :=^t \left(t_i^{(1)}, t_i^{(2)}\right)$, on résoud en t les équations suivantes

$$m_i^{(1)} = \frac{\theta + t_i^{(2)}}{\eta - t_i^{(1)}}$$

$$digamma(\theta + t_i^{(2)}) - \log(\eta - t_i^{(1)}) - m_i^{(2)} = 0$$

Les paramètres de $g(y_{i+1}|y_1^i)$ sont

$$\alpha := t_i + \frac{\kappa_{(i,n)}^{-2}\gamma_{(i,n)}}{2(n-i-1)}$$
$$\beta = (n-i-1)\kappa_{(i,n)}$$

avec

$$\kappa_{(i,n)} = \begin{pmatrix} \frac{\theta + t_i^{(2)}}{\left(\eta - t_i^{(1)}\right)^2} & \frac{1}{\eta - t_i^{(1)}} \\ \frac{1}{\eta - t_i^{(1)}} & trigamma(\theta + t_i^{(2)}) \end{pmatrix}$$

et

$$\gamma_{(i,n)} = \begin{pmatrix} 2\frac{\theta + t_i^{(2)}}{\left(\eta - t_i^{(1)}\right)^3} \\ \frac{1}{\left(\eta - t_i^{(1)}\right)^2} + psigamma(\theta + t_i^{(2)}, 2). \end{pmatrix}$$

A.3 Loi Weibull

Soit $\mathbf{X}_1^n$ de loi $W(b,c)$.

Conditionnement par $(\mathbf{U}_{1,n} = na)$ **avec** $u(x) := \log(x)$. La densité tiltée $\pi_u^{m_i}$ est

$$\pi_u^{m_i}(x) = \frac{1}{\Gamma(\frac{t_i}{c}+1)}\left(\frac{c}{b}\left(\frac{x}{b}\right)^{t_i+c+1}\exp\{-\left(\frac{x}{c}\right)^c\}\right)$$

On détermine t_i en résolvant l'équation suivante en t :

$$\log(b) + \frac{1}{c}digamma(\frac{t}{c}+1) - m_i = 0$$

Les paramètres de $g(y_{i+1}|y_1^i)$ sont

$$\alpha = t_i + c\frac{psigamma(\frac{t_i}{c}+1,2)}{2(n-i-1)\left(trigamma(\frac{t_i}{c}+1)\right)^2}$$
$$\beta = \frac{(n-i-1)trigamma(\frac{t_i}{c}+1)}{c^2}$$

A.4 Loi de Gumbel

Soit $\mathbf{X}_1^n$ de loi $V(\theta,\eta)$.

Conditionnement par $(\mathbf{S}_{1,n} = na)$. La densité tiltée $\pi_{\mathbf{X}}^{m_i}$ est

$$V\left(\theta, \frac{\eta}{1-\eta t_i}\right)$$

On détermine t_i en résolvant l'équation suivante en t :

$$\theta - \eta digamma(1-\eta t) - m_i = 0$$

Les paramètres de $g(y_{i+1}|y_1^i)$ sont

$$\alpha = t_i - \frac{psigamma(1-\eta t_i,2)}{2\eta(n-i-1)\left(trigamma(1-\eta t_i)\right)^2}$$
$$\beta = (n-i-1)\eta^2 trigamma(1-\eta t_i)$$

A.5 Loi Inverse-Gaussienne

Soit $\mathbf{X}_1^n$ de loi $IG(\mu,\lambda)$.

Conditionnement par $(\mathbf{S}_{1,n} = na)$. La densité tiltée $\pi_{\mathbf{X}}^{m_i}$ est

$$IG(\frac{\mu^2}{m_i},\lambda)$$

Les paramètres de $g(y_{i+1}|y_1^i)$ sont

$$\alpha = \lambda\left(\frac{1}{2\mu^2} - \frac{1}{2m_i^2} + \frac{3m_i^2}{2\mu^3(n-i-1)}\right)$$
$$\beta = \frac{m_i^3}{\lambda}(n-i-1)$$

Conditionnement par $(\mathbf{U}_1^n = na)$ **avec** $u(x) = 1/x$. La densité tiltée $\pi_u^{m_i}$ est

$$IG(\mu\sqrt{1-\frac{2t_i}{\lambda}}, \lambda\sqrt{1-\frac{2t_i}{\lambda}})$$

On détermine t_i par

$$t_i = \frac{1}{2}\left(\lambda - \frac{4\mu^2}{\left(-2\sqrt{\lambda}+\sqrt{4\lambda+4m_i\mu}\right)^2}\right)$$

Les paramètres de $g(y_{i+1}|y_1^i)$ sont

$$\alpha := t_i + \frac{\mu_3^i}{2s_i^4(n-i-1)}$$
$$\beta = (n-i-1)s_i^2$$

avec

$$s_i^2 = \frac{2}{(\lambda-2t_i)^2} - \frac{\sqrt{\lambda}}{\mu}(\lambda-2t_i)^{-3/2}$$
$$\mu_3^i = \frac{8}{(\lambda-2t_i)^3} - 3\frac{\sqrt{\lambda}}{\mu}(\lambda-2t_i)^{-5/2}$$

Bibliographie

[1] Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive MCMC. Stat. Comput., 18 :343-373.

[2] Barbe, P. and Broniatowski, M. (1999). Simulation in exponential families. ACM Transactions on Modeling and Computer Simulation (TOMACS), 9 :203–223.

[3] Barbe, P. and Broniatowski, M. (2000). Large-deviation probability and the local dimension of sets. Proceedings of the 19th Seminar on Stability Problems for Stochastic Models, Part I (Vologda, 1998). J. Math. Sci. (New York) 99(3) :1225–1233

[4] Barbe, P. and Broniatowski M. (2004). On sharp large deviations for sums of random vectors and multidimensional Laplace approximation. Teor Veroyatn Primen 49(4) :743–774

[5] Bar-Lev S.K., Bshouty D. (1982). Rational variance functions. Ann. of Stat., 17(2), 741-748.

[6] Barndorff-Nielsen, O.E. (1978). Information and exponential families in statistical theory. Wiley Series in Probability and Mathematical Statistics. Chichester : John Wiley & Sons.

[7] Barndorff-Nielsen, O.E. and Cox, R.R. (1990). Asymptotics Techniques for Use in Statisticals. Chapman and Hall.

[8] Barndorff-Nielsen, O.E. and Cox, R.R. (1994). Inference and Asymptotics. Chapman & Hall, London.

[9] Barndorff-Nielsen, O.E., Pedersen, K. (1978). Sufficient data reduction and exponential families. Math. Scand., 22 :197–202.

[10] Basu D. (1977). On the elimination of nuisance parameters. J. Am. Statist. Assoc. 72 :355–366.

[11] Besag, J. and Clifford, P. (1989). Generalized Monte Carlo Significance Tests. Biometrika, 76 :633–642.

[12] Besag, J. and Clifford, P. (1991). Sequential Monte Carlo p-Values. Biometrika 78 :301–304.

[13] Blanchet, J., Glynn, P., L'ecuyer, P., Sandmann, W., and Tuffin, B. (2007). Asymptotic Robustness of Estimators in Rare-Event Simulation. In Proceedings of the 2007 INFORMS Simulation Society (I-Sim) Research Workshop, Fontainebleau, France, July 2007.

[14] de Boer, P. T., Kroese, D. P., Mannor, S., and Rubinstein, R. Y. (2005). A tutorial on the cross-entropy method. Annals of Operations Research, 134(1) :19-67.

[15] Bolviken, E. and Skovlund, E. (1989). Confidence Intervals from Monte Carlo tests. J. Amer. Statist. Assoc., 91 :1071–1077.

[16] Botev Z.I. and Kroese D.P. (2012). Efficient Monte Carlo simulation via the Generalized Splitting Method. Stat. and Comp, 22 :1-16.

[17] Bhattacharya R.N. et Rao R. R. (1986). Normal Approximation and asymptotic expansions. R.E.Krieger

[18] Broniatowski, M. and Ritov, Y. (2009). Importance Sampling for rare events and conditioned random walks. *arXiv :0910.1819*

[19] Brown L.D. (1986). Fundamentals of statistical exponential families : with applicationin statitical theory. Hayward, Calif., Institute of Mathematical Statistics.

[20] Bucklew J.A. (2004). Introduction to rare event simulation. Springer Series in Statistics, Springer-Verlag, New York

[21] Buehler R.J., (1982). Some Ancillary Statistics and Their Properties. J. of the Am. Statis. Assoc., 77(379) :581–589

[22] Casella, G. and Robert, C. (1996). Rao-Blackwellisation of sampling schemes. Biometrika, 83(1) :81–94.

[23] Casella, G. and Robert, C. (1998). Post-processing accept-reject samples : recycling and rescaling. J. Comput. Graph. Statist., 7(2) :139–157.

[24] Chan, J. C. C. and Kroese, D. P. (2009). Rare-event probability estimation with conditional Monte Carlo. Annals of Operations Research. To appear.

[25] Cheng, R.C.H. (1984). Generation of inverse Gaussian variates with given sample mean and dispersion. Appl. Statist. 33 :309-316.

[26] Csiszár, I.(1984). Sanov property, generalized I-projection and a conditional limit theorem. Ann. Probab., 12 :768–793.

[27] de Acosta A. (1992). Moderate deviations and associated Laplace approximations for sums of independent random vectors. Trans. Amer. Math. Soc., 329(1) :357–375

[28] Dean, T., and Dupuis, P. (2009). Splitting for rare event simulation : A large deviation approach to design and analysis. Stoch. Proc. App. 119 :562–587

[29] Dembo, A. and Zetouni, O. (1996). Refinements of the Gibbs conditioning principle. Probab. Theory Related Fields, 104 :1–14.

[30] den Hollander, W.Th.F. and Weiss, G.H. (1988). On the range of a constrained random walk. J. Appl. Probab., 25 :451–463.

[31] den Hollander, W.Th.F. and Weiss, G.H. (1988). A note on configurational properties of constrained random walks. J. Phys. A., 21 :2405–2415.

[32] Devroye, L. (1987). Non-Uniform random variate generation. Springer-Verlag, New-York.

[33] Diaconis, P. and Freedman, D.A. (1988). Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. J. Theoret. Probab., 1 :381–410.

[34] Douc R. and Robert C.P. (2011). A vanilla Rao-Blackwellisation of Metropolis-Hastings algorithms. ArXiv : stat.CO/0904.2144

[35] Drton, M., and Richardson, T.S., (2004). Multimodality of the likelihood in the bivariate seemingly unrelated regressions model. Biometrika,91(2) :383-392..

[36] Dupuis, P., Leder, K. and Wang, H. 2007. Importance Sampling for Sums of Random Variables with Regularly Varying Tails. ACM Trans. Model. Comput. Simul. 17 :14.

[37] Dupuis, P., Sezer, A.D., and Wang, H. 2007. Dynamic importance sampling for queueing networks. Ann. Appl. Probab. 17 :1306-1346.

[38] Dupuis, P. and Wang, H. (2004). Importance sampling, large deviations, and differential games. Stoch. Stoch. Rep., 76 :481–508.

[39] EFRON, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency) (with discussion). Ann. Statist. 3 :1189-1242.

[40] EFRON, B. (1978). The geometry of exponential families. Ann. Statist. 6 :362-376.

[41] EFRON, B. (1979), Bootstrap methods : another look at the jackknife. Ann. Statist. 7(1) :1–26.

[42] Engen, S. and Lillegard, M. (1997). Stochastic simulations conditioned on sufficient statistics. Biometrika. 84 :235–240.

[43] Ermakov M.S. (2003). Asymptotically efficient statistical inferences for moderate deviation probabilities. Teor. Veroyatn. Primen., 48(4) :676–700

[44] Ermakov, M.S. (2006). The importance sampling method for modeling the probabilities of moderate and large deviations of estimates and tests. Teor. Veroyatn. Primen. 51 :319–332.

[45] Ermakov, M.S. (2007). Importance sampling for simulations of moderate deviation probabilities of statistics. Statist. Decisions, 25(4) :265–284.

[46] Feller, W. (1971). An introduction to probability theory and its applications. Vol.I. Second edition, John Wiley & Sons Inc., New York.

[47] Feller, W. (1971). An introduction to probability theory and its applications. Vol.II. Second edition, John Wiley & Sons Inc., New York.

[48] Fisher, R. A. (1925). Theory of statistical estimation. Proceedings of the Cambridge Philosophical Society, 22 :700–725.

[49] FRASER, D. A. S. (2004). Ancillaries and conditional inference. With comments by Ronald W. Butler, Ib M. Skovgaard, Rudolf Beran and a rejoinder by the author. Statist. Sci. 19(2) :333–369

[50] Glynn P.W. and Whitt W. (1992) The asymptotic efficiency of simulation estimators. Oper. Res., 40(3) :505–520.

[51] Holst, L., (1981). Some conditional limit theorems in exponential families. Ann. Prob., 9 :818–830.

[52] Hefferon J. (2010). Linear Algebra. http ://joshua.smcvt.edu/linearalgebra/book.pdf

[53] Höglund, T. (1979). A unified formulation of the central limit theorem for small and large deviations from the mean. Z. Wahrsch. Verw. Gebiete 49(1) :105–117

[54] Homem-de-Mello, T. (2007). A study on the cross-entropy method for rare event probability estimation. INFORMS Journal on Computing, 19(3) :381-394.

[55] Hope, A.C.A. (1968). A simplified Monte Carlo signifiance test procedure. J. R. Statist. Soc. B, 30 :582–598.

[56] IACOBUCCI, A., MARIN, J.-M., ROBERT, C. (2010). On variance stabilisation in population Monte Carlo by double Rao-Blackwellisation. Comput. Statist. Data Anal. 54(3) :698–710.

[57] Inglot T., Kallenberg W.C.M., Ledwina T. (1992). Strong moderate deviation theorems. Ann Probab 20(2) :987–1003

[58] Jensen J.L., (1986). Similar tests and the standardized log likelihood statistic. Biometrika, 73(3) :567–572.

[59] Jensen, J.L. (1995). Saddlepoint approximations. Oxford Statistical Science Series.

[60] Jöckel, K.-H. (1984). Computational aspects of Monte Carlo tests. Proceedings of COMPSTAT, 84 :183–188.

[61] Jöckel, K.-H. (1986). Finite sample properties and asymptotic efficiency of Monte Carlo tests. The Ann. of Stat., 14 :336–347.

[62] Jørgensen, B., (1986). Some Properties of Exponential Dispersion Models. Scand. Jour. of Stat., 13(3) :187–197

[63] Jørgensen, B., (1987). Exponential Dispersion Models. J. of the Ro. Statistical Society. Series B (Methodological), 49(2) :127–162.

[64] Lehmann, E.L. (1986). Testing Statistical Hypotheses. Springer.

[65] Lehman E.L., Casella G., (1998). Theory of Point Estimation, Second Edition. Springer-Verlag, New-York.

[66] Letac, G., Mora, M. (1990). Natural real exponential families with cubic variance functions. Ann. Statist. 18(1) :1–37.

[67] Lindqvist, B.H., Taraldsen, G., Lillegard, M., Engen, S. (2003). A counterexample to a claim about stochastic simulations. Biometrika, 90 :489-490.

[68] Lindqvist, B.H. and Taraldsen, G. (2005). Monte Carlo conditioning on a sufficient statistic. Biometrika, 92 :451-464.

[69] Lockhart, R.A. and Stephens, M. A. (1994). Estimation and tests of fit for the three–parameter Weibull distribution. J. Roy. Statist. Soc.. B. 56 :491–500.

[70] LOCKHART, R. AND O'REILLY, F. (2005) A note on Moore's conjecture. Statist. Probab. Lett. 74 , no. 2, 212–220.

[71] Lockhart, R.A., O'Reilly F.J., and Stephens M.A. (2007). Use of the Gibbs sampler to obtain conditional tests, with applications. Biometrika, 94(4) :992-998.

[72] Michel R. (1979). On the Asymptotic Efficiency of Conditional Tests for Exponential Families. Ann. of Statis., 7(6) :1256–1263

[73] MOORE, DAVID S. (1973). A note on Srinivasan's goodness-of-fit test. Biometrika, 60 :209–211.

[74] O'REILLY, F. AND GRAVIA-MEDRANO, L. (2006). On the conditional distribution of goodness-of-fit tests. Commun. Statist. A, 35 :541–9.

[75] Pace L. and Salvan A., (1992). A note on conditional cumulants in canonical exponential families. Scand. J. Statist., 19 :185–191.

[76] Pedersen, B.V., (1979). Approximating conditional distributions by the mixed Edgeworth-saddlepoint expansion. Biometrika, 66(3) :597–604.

[77] Perron, F. (1999). Beyond accept-reject sampling. Biometrika, 86(4) :803–813.

[78] REID, N. (1995). The roles of conditioning in inference. With comments by V. P. Godambe, Bruce G. Lindsay and Bing Li, Peter McCullagh, George Casella, Thomas J. DiCiccio and Martin T. Wells, A. P. Dawid and C. Goutis and Thomas Severini. With a rejoinder by the author. Statist. Sci. 10(2) :138–157, 173–189, 193–196.

[79] Rihter, V. (1957). Local limit theorems for large deviations. Dokl. Akad. Nauk SSSR (N.S.), 115 :53–56.

[80] Rihter, V. (1958). Multidimensional local limit theorems for large deviations. Theory Prob. Appl., 3 :100–106.

[81] DOUC, R. AND ROBERT, C.P.(2011). A vanilla Rao-Blackwellization of Metropolis-Hastings algorithms. Ann. Statist. 39(1) :261–277.

[82] Rockafellar R.T., (1970). Convex Analysis, Princeton Landmarks In Mathematics.

[83] Rubino G., Tuffin B. (2009). Rare Event Simulation Using Monte Carlo Methods. Wiley, Chichester, 2009.

[84] Rubinstein, R. Y. (1997). Optimization of computer simulation models with rare events. European Journal of Operational Research, 99 :89-112.

[85] Serfling, R.J., (1980). Approximation theorems of mathematical statistics. John Wiley & Sons Inc., New-York.

[86] Taylor, R.L., Daffer, P.Z. and Patterson, R.F. (1985). Limit theorems for sums of exchangeable random variables. Rowman & Allanheld Probability and Statistics Series,Totowa, NJ.

[87] Sadowsky J.S. and Bucklew J.A. (1990). On large deviations theory and asymptotically efficient Monte Carlo estimation. IEEE Trans Inform Theory 36(3) :579–588

[88] Shiryaev, A. (1996). Probability Theory (Second ed.). New York : Springer-Verlag.

[89] Severini T.A. (1994). On the approximate elimination of nuisance parameters by conditioning. Biometrika, 81(4) :649–661.

[90] Snedecor, G.W. and Cochran, W.G. (1989). Statistical Methods, Eighth Edition. Ames, IA : Iowa State University Press.

[91] Sundberg, R. (2009). Flat and multimodal likelihoods and model lack of fit in curved exponential families. Research Report 2009 :1, http ://www.math.su.se/matstat.

[92] Sundberg, R. (2010). Flat and multimodal likelihoods and model lack of fit in curved exponential families. Scand. Jour. of Stat, 37(4) :632-643.

[93] Van Campenhout, J.M. and Cover, T.M. (1981). Maximum entropy and conditional probability. IEEE Trans. Inform. Theory, 27 :483–489.

[94] Zabell S.L. (1980). Rates of convergence for conditional expectations. *Ann. Probab.* 8 :5, 928-941.