



HAL
open science

Asymptotic approaches in financial risk management

Adrien Genin

► **To cite this version:**

Adrien Genin. Asymptotic approaches in financial risk management. General Mathematics [math.GM]. Université Sorbonne Paris Cité, 2018. English. NNT : 2018USPCC120 . tel-02361422

HAL Id: tel-02361422

<https://theses.hal.science/tel-02361422>

Submitted on 13 Nov 2019

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

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

**UNIVERSITÉ PARIS DIDEROT (PARIS 7)
SORBONNE PARIS CITE**

**ÉCOLE DOCTORALE DE SCIENCES MATHÉMATIQUES
DE PARIS CENTRE**

**Laboratoire de Probabilités, Statistique et Modélisation
CNRS UMR 8001**

THÈSE DE DOCTORAT

Discipline : Mathématiques Appliquées

Présentée par **Adrien GENIN**

**APPROCHES ASYMPTOTIQUES EN
GESTION DES RISQUES FINANCIERS**

**ASYMPTOTIC APPROACHES IN FINANCIAL
RISK MANAGEMENT**

Sous la direction de **Peter TANKOV**

Rapporteurs :

M. Damien Lamberton Université Paris Est Marne La Vallée

M. Henrik Hult KTH Stockholm

Soutenue publiquement le **21/09/2018** devant le jury composé de :

M. Jean-François Chassagneux	Professeur	Université Paris Diderot	Président du jury
Mme Zorana Grbac	Maître de conférences	Université Paris Diderot	Examinatrice
Mme Ying Jiao	Maître de conférences	Université Claude Bernard Lyon 1	Examinatrice
M. Henrik Hult	Professeur	KTH Stockholm	Rapporteur
M. Damien Lamberton	Professeur	Université Paris Est Marne La Vallée	Rapporteur
M. Stefano De Marco	Maître de conférences	Ecole Polytechnique	Examineur
M. Peter Tankov	Professeur	ENSAE	Directeur de thèse

À mes grands parents
À Nana

Remerciements

Je souhaite ici adresser mes remerciements à l'ensemble des personnes qui ont directement ou indirectement permis le bon déroulement de ma thèse.

Tout d'abord je souhaite remercier tout particulièrement mon directeur de thèse, Peter Tankov, pour son encadrement, son aide précieuse, sa patience et ses encouragements qui ont largement permis à cette thèse d'aboutir. Je tiens également à remercier Damien Lambertson et Henrik Hult, qui ont accepté de rapporter cette thèse ainsi que pour leur lecture attentive du manuscrit. Enfin je suis reconnaissant que Jean-François Chassagneux, Zorana Grbac, Ying Jiao et Stefano De Marco fassent partis du jury et les remercie d'avoir accepté.

Je tiens également à remercier l'ensemble du LMPA (d'sormais LPSM) qui est un environnement idéal pour effectuer une thèse. Je souhaite donc remercier ici son directeur Francis Comets ainsi que l'ensemble de l'équipe administrative pour leur aide tout au long de ces années. Mon expérience au sein du LPMA a été largement marquée par les doctorants qui ont permis d'installer une ambiance amicale et bienveillante. Ainsi un grand merci à tous ceux que j'ai croisé durant ces années, et en particulier: Anna, Arturo, Aser, Jiayu, Guillaume, Lorick, Marc-Antoine, Sophie et Vu-Lan. Je tiens à adresser un mot spécial à Maud pour son soutien et ses encouragements ainsi que son amitié. Enfin je remercie également Noufel qui, bien que n'étant plus doctorant, a accepté de partager déjeuner et pause café avec les "petits".

Je souhaite remercier l'ensemble de l'équipe d'Opus Finance pour m'avoir donné l'opportunité d'effectuer cette thèse tout en m'ayant permis de me confronter au monde professionnel dans lequel j'ai énormément appris. Merci donc à Philippe, Henk, Olivier, Vincent ainsi qu'à l'ensemble des personnes que j'ai eu l'opportunité de rencontrer durant mon parcours.

Je terminerai en remerciant mes proches, amis et famille, pour leur soutien et les merveilleux moments passés et à venir.

Abstract

Cette thèse se propose de traiter de trois problèmes de gestion des risques financiers en utilisant différentes approches asymptotiques. La première partie présente un algorithme Monte Carlo d'échantillonnage d'importance pour la valorisation d'options asiatiques dans des modèles exponentiels de Lévy. La mesure optimale d'échantillonnage d'importance est obtenue grâce à la théorie des grandes déviations. La seconde partie présente une étude du comportement asymptotique de la somme de n variables aléatoires positives et dépendantes dont la distribution est un *mélange log-normal* ainsi que des applications en gestion des risque de portefeuille d'actifs. Enfin, la dernière partie, présente une application de la notion de variations régulières pour l'analyse du comportement des queues de distribution d'un vecteur aléatoire dont les composantes suivent des distributions à queues épaisses et dont la structure de dépendance est modélisée par une copule Gaussienne. Ces résultats sont ensuite appliqués au comportement asymptotique d'un portefeuille d'options dans le modèle de Black-Scholes.

Abstract

This thesis focuses on three problems from the area of financial risk management, using various asymptotic approaches. The first part presents an importance sampling algorithm for Monte Carlo pricing of exotic options in exponential Lévy models. The optimal importance sampling measure is computed using techniques from the theory of large deviations. The second part uses the Laplace method to study the tail behavior of the sum of n dependent positive random variables, following a *log-normal mixture* distribution, with applications to portfolio risk management. Finally, the last part employs the notion of multivariate regular variation to analyze the tail behavior of a random vector with heavy-tailed components, whose dependence structure is modeled by a Gaussian copula. As application, we consider the tail behavior of a portfolio of options in the Black-Scholes model.

Contents

1	Introduction (en français)	13
1	Les défis et méthodes de gestion des risques financiers	14
1.1	Valorisation d'instruments dérivés	14
1.2	Les mesures de risque	17
1.3	Simulation d'évènements rares et analyse asymptotique	20
2	Principales contributions de cette thèse	26
2.1	Echantillonnage d'importance dans le contexte de processus de Lévy	27
2.2	Analyse asymptotique des portefeuilles suivants des mélanges log-normaux	31
2.3	Asymptotiques de distributions à queues épaisses et structure de dépendance à copule Gaussienne	34
2	Introduction	37
1	Challenges and methods of financial risk management	38
1.1	Pricing of derivative instruments	38
1.2	Measuring the risk	41
1.3	Rare events simulation and tail analysis	44
2	Main contributions of this thesis	50
2.1	Optimal importance sampling for Lévy processes	50
2.2	Tail asymptotics of log-normal mixture portfolios	54
2.3	Asymptotics of heavy-tailed risks with Gaussian copula dependence	57
3	Optimal importance sampling for Lévy processes	61
1	Introduction	62
2	Pathwise large deviations for Lévy processes	65
3	Main results	70
4	Examples	79
5	Numerical illustrations	84

4	Tail asymptotics of log-normal mixture portfolios	93
1	Introduction	94
2	Notation and preliminaries	95
3	Asymptotic behavior of the left tail	97
4	Examples	103
4.1	Variance Gamma process	103
4.2	Heston model	103
5	Variance Reduction of Monte Carlo estimates	105
6	Proof of Theorem 4.3	109
5	Asymptotics of heavy-tailed risks with Gaussian copula dependence	119
1	Introduction	120
2	Main results	121
3	Examples	127
3.1	Portfolio of options in the Black-Scholes model	127
3.2	Aggregate production of a wind farm	130

Chapter 1

Introduction (en français)

Contents

1	Les défis et méthodes de gestion des risques financiers .	14
1.1	Valorisation d'instruments dérivés	14
1.2	Les mesures de risque	17
1.3	Simulation d'évènements rares et analyse asymptotique .	20
2	Principales contributions de cette thèse	26
2.1	Echantillonnage d'importance dans le contexte de processus de Lévy	27
2.2	Analyse asymptotique des portefeuilles suivants des mélanges log-normaux	31
2.3	Asymptotiques de distributions à queues épaisses et struc- ture de dépendance à copule Gaussienne	34

Cette thèse propose 3 contributions à la théorie de la gestion des risques financiers en se basant sur des méthodes asymptotiques. Ce chapitre introductif présente les difficultés majeures de la gestion des risques ainsi que les méthodes couramment utilisées. Il se propose également de montrer dans quelles mesures les méthodes asymptotiques peuvent être performantes notamment dans le cadre d'évènements rares. Enfin, les principaux résultats développés dans la thèse sont présentés dans les parties suivantes.

1 Les défis et méthodes de gestion des risques financiers

1.1 Valorisation d'instruments dérivés

La valorisation d'instruments dérivés est le fondement de la gestion des risques. En effet, le prix des actifs illiquides détenus dans un portefeuille, nécessaires pour calculer la charge en capital règlementaire, est calculé via un modèle. Par conséquent, la sélection d'un modèle approprié est un élément clé du processus de valorisation. Un modèle approprié doit reproduire les faits stylisés observés sur les marchés afin d'estimer de façon adéquate les risques. De plus, il doit également fournir un cadre opérationnel pour calculer les prix et les mesures de risques. Les principaux modèles utilisés pour la valorisation d'instruments dérivés sont présentés ci-dessous.

Le modèle de Black-Scholes Le célèbre modèle de Black-Scholes développé en 1973 permet d'obtenir une formule explicite de valorisation des options européennes. Si le prix d'un actif S suit un mouvement Brownien géométrique de volatilité σ alors, le prix du call européen de maturité T et de strike K au temps t et avec le taux d'intérêt r est donné par la formule suivante:

$$C(t, \sigma; T, K) = S_t N(d_1) - K e^{-r(T-t)} N(d_2)$$

où S_t est le prix de l'actif sous-jacent au temps t , N est la fonction de répartition de la loi normale standard et

$$d_1 = \frac{\log \frac{S_t}{K e^{-r(T-t)}} + (T-t) \frac{\sigma^2}{2}}{\sigma \sqrt{T-t}}$$

$$d_2 = d_1 - \sigma \sqrt{T-t}.$$

D'après ce modèle, le prix de l'option dépend de deux types de paramètres: des paramètres de marchés tels que la volatilité et éventuellement le taux d'intérêt et les dividendes (ici considérés comme nuls) ainsi que des paramètres contractuels tels

que le strike et la maturité de l'option. Tous ces paramètres sont observables sauf la volatilité. Ainsi la volatilité est telle que le prix de marché est égal au prix donné par le modèle. Le modèle de Black-Scholes est une fonction bijective de la volatilité, il est ainsi possible de résoudre le problème d'inversion à partir du prix de marché afin d'obtenir la volatilité, appelée volatilité implicite. Formellement, la volatilité implicite est définie comme le paramètre $\sigma_{T,K}^{imp}$ tel que

$$C(t, \sigma_{T,K}^{imp}; T, K) = C^{mkt}(t; T, K)$$

où $C^{mkt}(t; T, K)$ est le prix de marché de l'option. Ce paramètre, constant pour tous les strikes et maturités, peut ensuite être utilisé pour valoriser des options non-cotées. Le principal avantage de ce modèle est qu'il permet d'obtenir une formule explicite et que sa calibration est relativement simple à implémenter.

Néanmoins, les observations empiriques du marché des options montrent que la volatilité implicite n'est pas constante pour tous les strikes et maturités. En effet, le mapping $T, K \mapsto \sigma_{T,K}^{imp}$ appelé surface de volatilité implicite, n'est pas constant ce qui contredit une des hypothèses importantes du modèle. De plus, le modèle de Black-Scholes s'appuie sur l'hypothèse que les rendements du sous-jacent sont normalement distribués mais cette hypothèse est également contredite par l'observation des marchés qui mettent en évidence des distributions à queues épaisses pour les rendements.

Afin de prendre en compte les observations sur la volatilité implicite et les distributions des rendements, le modèle de Black-Scholes a été étendu. Certaines de ces extensions sont présentées ci-après.

Le modèle à volatilité locale Le modèle à volatilité locale suppose que la volatilité est une fonction déterministe du temps et du prix du sous-jacent. Il est défini de la façon suivante:

$$\frac{dS_t}{S_t} = \mu dt + \sigma(t, S_t) dW_t$$

où $\mu \in \mathbb{R}$ et W est un mouvement Brownien standard. Ce modèle permet de reproduire la surface de volatilité implicite. En effet, étant donnée une surface de volatilité implicite au temps $t = 0$, il existe une unique fonction $\sigma(t, x)$ qui reproduit le prix des options [?]. Cependant, ce modèle prévoit une mauvaise dynamique de cette surface dans le temps. Ainsi, le smile de volatilité futur prédit par le modèle est souvent plus aplati que le smile observé. Les modèles à volatilité stochastique permettent de prévoir une meilleure dynamique de la volatilité implicite.

Les modèles à volatilité stochastique Dans les modèles à volatilité stochastiques, le prix du sous-jacent est modélisé comme la première composante d'un processus 2-dimensionnel (S, σ) dirigé par un mouvement Brownien (W^1, W^2) bi-dimensionnel,

$$\frac{dS_t}{S_t} = \mu dt + \sigma_t dW_t^1 \quad (1.1)$$

$$\sigma_t = f(Y_t) \quad dY_t = a_t dt + b_t dW_t^2. \quad (1.2)$$

La volatilité est supposée être un processus aléatoire possiblement corrélé avec les trajectoires du sous-jacent. Deux modèles à volatilité stochastique couramment utilisés sont le modèle SABR [?] et le modèle de Heston [?].

Les modèles à volatilité stochastique permettent en général de répliquer le smile de volatilité même si le skew (la pente du smile de volatilité) pour les maturités courtes est souvent trop faible. En effet, un choix approprié des coefficients de diffusion permet d'obtenir une distribution à queues arbitrairement épaisses. Néanmoins, une telle calibration pour les maturités courtes a pour conséquence des valeurs irréalistes pour la volatilité de la volatilité (paramètre b dans (1.2)). Ces modèles ne sont pas toujours adaptés pour la valorisation d'options. Un important skew de volatilité pour les options de courtes maturités signifie que les prix des options ne sont pas compatibles avec le modèle de diffusion. Ce phénomène reflète la crainte d'un mouvement baissier important.

Dans ce contexte, ajouter des sauts au modèle de diffusion permet de répondre aux variations importantes des rendements ainsi qu'aux faits stylisés observés sur la volatilité implicite. Les modèles de Lévy permettent de modéliser les trajectoires des prix du sous-jacent avec des sauts.

Les processus de Lévy

Les processus de Lévy sont des processus stochastiques dont les incréments sont indépendants et stationnaires. Le livre [?] propose une description détaillée des processus de Lévy et de leurs applications en finance. La loi d'un processus de Lévy $(X_t)_{t \geq 0}$ est déterminée par la loi de X_t au temps $t \geq 0$ et est caractérisée par sa fonction caractéristique donnée par la formule de Lévy-Khintchine: pour $s \in \mathbb{R}$,

$$\mathbb{E} [e^{isX_t}] = e^{t\Psi(s)}$$

$$\Psi(s) = is\gamma - \frac{\sigma^2 s^2}{2} + \int_{\mathbb{R}} (e^{isx} - 1 - isx\mathbf{1}_{|x| \leq 1}) \nu(dx).$$

Le triplet (σ^2, ν, γ) est appelé le triplet caractéristique du processus de Lévy où $\gamma \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}_+$, et ν une mesure positive qui satisfait la condition d'intégrabilité $\int_{\mathbb{R}} (1 \wedge |x|^2) \nu(dx) < \infty$.

La décomposition de Lévy-Itô permet d'obtenir la décomposition de la trajectoire de X :

$$X_t = \gamma t + \sigma B_t + \int_0^t \int_{|x| \leq 1} x \tilde{J}(ds \times dx) + \int_0^t \int_{|x| > 1} x J(ds \times dx)$$

où B est un mouvement Brownien standard, J est une mesure de Poisson d'intensité $dt \times \nu(dx)$ et \tilde{J} est la version centrée de J .

Le modèle Variance-Gamma [?] et le modèle Normal Inverse Gaussien [?] sont deux exemples de modèles construits à partir de processus de Lévy et utilisés pour la modélisation des actifs avec des sauts.

1.2 Les mesures de risque

La gestion des risques permet notamment d'atteindre trois objectifs. Elle permet de (i) réduire la volatilité des rendements d'un portefeuille, (ii) diversifier les positions d'un portefeuille et limiter les effets de concentration, (iii) mesurer l'exposition d'un portefeuille à des scénarios économiques et identifier ceux pouvant entraîner des pertes de valeur importantes du portefeuille.

Value-at-Risk La Value-at-Risk permet d'analyser le risque de marché d'un portefeuille à partir d'hypothèses sur les rendements. Son avantage principal est qu'elle constitue une mesure synthétique du risque d'un portefeuille. La Value-at-Risk correspond au quantile de la distribution des pertes et profits du portefeuille considéré. Elle représente le niveau de pertes qui pourrait être atteint avec une certaine probabilité et pour un horizon de temps donné. Plus formellement, si V_t représente la valorisation (connue) du portefeuille au temps t , τ l'horizon de temps considéré, alors la Value-at-Risk au niveau $\alpha \in (0, 1)$, notée $\text{VaR}_{\alpha, \tau}$, est définie comme la solution de l'équation suivante:

$$\text{VaR}_{\alpha, \tau} = \inf\{x : \mathbb{P}[V_{t+\tau} - V_t \leq -x] \leq 1 - \alpha\}.$$

Il existe principalement trois approches pour calculer la Value-at-Risk décrites ci-dessous.

La méthode historique L'approche par simulation historique consiste à déduire la distribution des pertes et profits du portefeuille à partir de l'observation des mouvements historiques des facteurs de risques. Les mouvements historiques sont répliqués sur le portefeuille actuel donnant ainsi une distribution des pertes et profits et permettant ainsi de calculer le quantile. Bien que simple et rapide à implémenter, cette méthode dépend fortement de l'historique considéré qui peut ne pas contenir d'importants mouvements baissiers et ainsi minimiser la Value-at-Risk. En pratique, cette méthode est essentiellement utilisée pour calculer une Value-at-Risk stressée, intégrant intentionnellement des historiques contenant des mouvements baissiers importants des facteurs de risque.

Les méthodes delta-normale et delta-gamma La valorisation d'un portefeuille peut s'écrire comme une fonction d'un ensemble de n facteurs de risque $(x^{(1)}, \dots, x^{(n)})$. Ainsi,

$$V = f(x^{(1)}, \dots, x^{(n)})$$

La méthode delta-normale suppose que les variations de valeur du portefeuille peuvent être approchée de la façon suivante:

$$V_{i+1} - V_i \approx \sum_{j=1}^n \frac{\partial f}{\partial x_i^{(j)}} (x_{i+1}^{(j)} - x_i^{(j)}) = \sum_{j=1}^n \frac{\partial f}{\partial x_i^{(j)}} x_i^{(j)} r_{i+1}^{(j)}$$

où $r_{i+1}^{(j)} = \frac{x_{i+1}^{(j)}}{x_i^{(j)}} - 1$ représente le rendement du facteur j . De plus, en supposant que les rendements sont distribués selon une loi normale centrée et de matrice de variance-covariance constante en fonction du temps, alors la variance du portefeuille, notée σ_V^2 , peut être calculée de la façon suivante:

$$\sigma_{V,i+1}^2 = \sum_{j=1}^n \sum_{k=1}^n \sigma^{(j)} \sigma^{(k)} \rho^{(j,k)} \frac{\partial f}{\partial x_{i+1}^{(j)}} x_i^{(j)} \frac{\partial f}{\partial x_{i+1}^{(k)}} x_i^{(k)}.$$

La VaR du portefeuille au niveau p sur une période temps se calcule donc:

$$VaR_{i,i+1} = \sigma_{V,i+1} N^{-1}(1-p)$$

où N est la fonction de répartition de la loi normale standard.

Cette méthode présente l'avantage d'être relativement simple à implémenter, cependant l'approximation à l'ordre 1 se révèle inexacte pour les portefeuilles contenant des options. En effet, la dérivée de second ordre a un impact direct sur la queue de distribution du portefeuille et donc sur la VaR. Ainsi, une extension de cette méthode, considérant l'approximation à l'ordre 2, a été développée. Elle est

appelée la méthode delta-gamma. Néanmoins, cette méthode peut être relativement complexe à implémenter notamment pour calculer les dérivées d'ordre 2.

Ces 2 méthodes reposent sur l'hypothèse que les rendements du portefeuille sont normalement distribués, ce qui est contredit par les observations empiriques. Cette approximation peut conduire à une sous-estimation du quantile de la distribution et donc à une sous-estimation de la VaR. Même si ces méthodes peuvent être répliquées dans un contexte non-gaussien, leur mise en oeuvre peut s'avérer difficile et coûteuse en temps de calcul. Ainsi, en pratique, l'analyse des queues de distribution d'un portefeuille et le calcul de la VaR sont généralement conduits au travers de la méthode de Monte Carlo, plus flexible.

Méthode de Monte Carlo. La méthode de Monte Carlo consiste à générer des scénarios pour les facteurs de risques afin d'en déduire le quantile de la distribution des variations de prix du portefeuille. En fonction des facteurs de risques, le portefeuille s'écrit $V = f(x^{(1)}, \dots, x^{(n)})$ où f est une fonction connue (e.g. payoff). La flexibilité de la méthode repose sur le fait qu'elle s'applique à tous les modèles de diffusion des facteurs de risques (tant qu'il est simulable) permettant ainsi de prendre en compte des distributions à queues épaisses. Dans ce contexte, cette méthode permet de mesurer de façon plus précise les queues de distribution, et par conséquent les risques associés au portefeuille. Au-delà des indicateurs de risques, cette méthode est aussi largement utilisée pour le pricing d'instruments dérivés. Dans cette thèse, les méthodes de Monte Carlo sont appliquées aux deux situations.

Autres mesures de risque D'autres mesures de risque ont été développées afin de mesurer d'autres types de risques contenus dans les portefeuilles d'actifs financiers. En effet, la VaR indique le niveau de pertes potentielles d'un portefeuille pour une probabilité donnée, cependant elle ne fournit pas d'indications sur les montants de pertes au-delà de ce niveau ni d'analyse sur le comportement du portefeuille en cas de scénarios défavorables.

Afin de compenser une potentielle sous-estimation des risques mesurés au travers la VaR, le régulateur impose le calcul d'une VaR stressée incluant une période stressée des marchés. Ceci permet de forcer la prise en compte de mouvements extrêmes et réalistes afin de quantifier l'impact de ces mouvements sur les portefeuilles.

De plus, l'*Expected Shortfall* permet de calculer le montant moyen des pertes subies pour les événements ayant une probabilité d'occurrence inférieures au niveau

de la VaR. Elle est définie de la façon suivante:

$$ES_\alpha = \frac{1}{\alpha} \int_0^\alpha \text{VaR}_\gamma d\gamma.$$

Cependant, le calcul de l'Expected Shortfall n'est pas immédiat car il est nécessaire de pouvoir calculer l'espérance des valeurs du portefeuilles pour des probabilités faibles.

Outre les mesures de risque, les stress tests sont un autre outil largement utilisés pour la gestion des risques. Le stress testing consiste à simuler différents scénarios adverses et analyser leurs impacts sur le portefeuille (cf. [?] du Comité de Bale pour la supervision Bancaire, BCBS). Les stress tests jouent un rôle important notamment pour compenser les limitations des modèles utilisés ou le manque de données historiques, ainsi que pour tester la validité des hypothèses d'un modèle en cas de scénario adverse. De plus, les modèles de risque sont généralement calibrés sur des données historiques. De fait, dans des conditions stressées, les schémas usuellement observés et les dépendances entre les facteurs de risque peuvent être modifiés. Par exemple, les méthodes de stress tests utilisés avant la crise de 2007, n'ont pas permis d'anticiper les effets de corrélations entre le risque de marché et de crédit ou encore entre le financement et la liquidité. Le comité de Bale souligne notamment les faiblesses des scénarios sélectionnés jusqu'alors (cf. [?]).

Ainsi, la sélection de scénarios pour les stress tests est essentielle. En effet, ils doivent être à la fois suffisamment sévères pour reproduire un choc brutal sur les marchés et être réalistes pour correspondre à une possible situation de tension économique.

1.3 Simulation d'évènements rares et analyse asymptotique

Comme décrit dans les sections précédentes, le calcul des mesure de risque nécessite d'évaluer des faibles quantiles de la distribution des portefeuilles et par conséquent le calcul de probabilité d'occurrence d'évènements rares. Dans cette thèse, nous proposons de s'appuyer sur (i) les méthodes de Monte Carlo avec réduction de variance pour calculer les probabilités d'évènements rares de manière efficiente ainsi que (ii) des méthodes asymptotiques basées sur la théorie des variations régulières et des grandes déviations.

La méthode de Monte Carlo pour les évènements rares La valorisation d'instruments financiers complexes par méthode de Monte Carlo requiert d'être capable de simuler des évènements rares. Par exemple, pour valoriser une option barrière, les évènements rares à simuler correspondent à l'évènement où le prix du

sous-jacent dépasse un certain niveau. Dans ce contexte, l'application de la méthode de Monte Carlo brute peut se révéler inappropriée comme décrit ci-dessous.

Description du problème Soit X une variable aléatoire. Supposons que l'on cherche à estimer la probabilité $p = \mathbb{P}[X < x]$, avec p petit. En utilisant un échantillon de taille n , (X_1, \dots, X_n) , il est possible d'estimer p via la moyenne empirique:

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \leq x}.$$

D'après le théorème central limite, quand $n \rightarrow \infty$, la distribution de la variable

$$\frac{\sqrt{n}(\hat{p} - p)}{\sqrt{\text{Var} \mathbf{1}_{X \leq x}}} = \frac{\sqrt{n}(\hat{p} - p)}{\sqrt{p - p^2}}$$

tend vers une loi normale standard $\mathcal{N}(0, 1)$. Ainsi, pour n suffisamment grand, on obtient

$$\mathbb{P} \left[p \in \left(\hat{p} - \frac{z\sqrt{p-p^2}}{\sqrt{n}}; \hat{p} + \frac{z\sqrt{p-p^2}}{\sqrt{n}} \right) \right] \approx 2N(z) - 1, \quad z > 0.$$

où N est la fonction de répartition de la loi normale standard. Autrement dit, la quantité $\frac{z\sqrt{p-p^2}}{\sqrt{n}}$ mesure la largeur de l'intervalle de confiance. Néanmoins, l'erreur relative est une mesure plus appropriée de la précision de la simulation. Elle est définie comme:

$$RE = \frac{z\sqrt{p-p^2}}{p\sqrt{n}}$$

Ainsi, pour un événement rare, p est petit et

$$RE \approx \frac{z}{\sqrt{pn}},$$

devient très grand à moins que n soit suffisamment grand par rapport à $\frac{1}{p}$. Cependant, augmenter la taille de l'échantillon augmente le temps de calcul nécessaire ce qui peut rendre la méthode inappropriée pour des applications pratiques. Dans ce contexte, la réduction de variance de l'estimateur peut s'avérer une méthode efficace pour ne pas augmenter la taille de l'échantillon.

Echantillonnage d'importance pour la réduction de variance des estimateurs Monte Carlo Dans cette section, nous ne présentons pas les détails techniques liés à la méthode mais nous nous concentrerons sur les idées principales de la méthode. Ainsi, comme le problème vient de la faible probabilité d'occurrence des événements que nous cherchons à simuler, l'objectif de la méthode est de leur donner une plus grande probabilité d'occurrence. Supposons que l'on souhaite estimer

l'espérance de la variable aléatoire X , $\theta = \mathbb{E}_{\mathbb{P}}[X]$. Alors, étant donnée une autre mesure de probabilité \mathbb{Q} , équivalente à \mathbb{P} , on peut écrire

$$\mathbb{E}_{\mathbb{P}}[X] = \mathbb{E}_{\mathbb{Q}} \left[X \frac{d\mathbb{P}}{d\mathbb{Q}} \right].$$

et donc, $\theta = \mathbb{E}_{\mathbb{Q}} \left[X \frac{d\mathbb{P}}{d\mathbb{Q}} \right]$. L'estimateur devient donc

$$\hat{\theta}_{\mathbb{Q}} = \frac{1}{n} \sum_{i=1}^n \hat{X}_i \frac{d\mathbb{P}}{d\mathbb{Q}} \Big|_i$$

où $\hat{X}_1, \dots, \hat{X}_n$ sont des copies indépendantes et identiquement distribuées de X sous la probabilité \mathbb{Q} et $\frac{d\mathbb{P}}{d\mathbb{Q}} \Big|_i$ sont les réalisations correspondantes de la densité du changement de mesure. La variance de cet estimateur s'écrit

$$\sigma_{\mathbb{Q}}^2 = \mathbb{E}_{\mathbb{Q}} \left[X^2 \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right)^2 \right] - \theta^2$$

L'objectif est alors de trouver le changement de mesure $\mathbb{P} \rightsquigarrow \mathbb{Q}$ qui minimise la variance de l'estimateur $\sigma_{\mathbb{Q}}^2$. Il s'agit de minimiser

$$\mathbb{E}_{\mathbb{Q}} \left[X^2 \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right)^2 \right] = \mathbb{E}_{\mathbb{P}} \left[X^2 \frac{d\mathbb{P}}{d\mathbb{Q}} \right].$$

En supposant que $X > 0$ et en choisissant $\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{X}{\mathbb{E}_{\mathbb{P}}[X]}$, nous obtenons

$$\mathbb{E}_{\mathbb{P}} \left[X^2 \frac{d\mathbb{P}}{d\mathbb{Q}} \right] = \mathbb{E}_{\mathbb{P}}[X]^2$$

et ainsi $\sigma_{\mathbb{Q}}^2 = 0$. Bien que ce choix de \mathbb{Q} est optimal, il ne peut être atteint car cela requiert la connaissance de $\mathbb{E}_{\mathbb{P}}[X]$, qui est précisément la quantité que nous cherchons à estimer. Cependant, dans certains cas le changement de mesure optimal peut être calculé de façon numérique.

Les méthodes d'échantillonnage d'importance ont été largement utilisées pour la valorisation d'options notamment dans les articles [?, ?, ?, ?, ?, ?, ?]. Dans le Chapitre 2, nous présenterons une application de la méthode d'échantillonnage d'importance pour valoriser les options exotiques dans des modèles basés sur les processus de Lévy.

Méthodes asymptotiques.

L'étude du comportement asymptotique des portefeuilles considérés peut constituer une alternative aux méthodes de Monte Carlo pour la simulation d'évènements

rares. En effet, ceci peut permettre de calculer rapidement (sans effectuer des simulations coûteuses en temps de calcul) les mesures de risque associées à des portefeuilles contenant un nombre important d'actifs.

Dans cette thèse, nous nous baserons sur la théorie des variations régulières ainsi que sur la théorie des grandes déviations. Les paragraphes suivants proposent une vue d'ensemble de ces méthodes et de leurs applications en mathématiques financières.

La théorie des variations régulières La théorie des variations régulières constitue une méthode intéressante pour l'étude asymptotique des distributions. Dans cette section, nous présentons les principales définitions et concepts de cette théorie. Pour plus de détails, le lecteur peut se référer à [?].

Definition 1.1. Une fonction mesurable $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ est dite à variations régulières à l'infini d'indice $\rho \in \mathbb{R}$ (noté $f \in RV_\rho$) si pour $x > 0$,

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = x^\rho.$$

Si $\rho = 0$, f est dite à variations lentes. Une conséquence directe de cette définition est que si $f \in RV_\rho$, alors $\frac{f(x)}{x^\rho} \in RV_0$ et nous nous avons la proposition suivante.

Proposition 1.2. Si $f \in RV_\rho$, alors il existe une fonction à variations lentes l telle que $f(x) = x^\rho l(x)$.

Definition 1.3. Une variable aléatoire positive X de fonction de répartition F à une queue à variations régulières à l'infini d'indice α si sa fonction de survie s'écrit de la façon suivante

$$\bar{F}(r) := 1 - F(r) = r^{-\alpha} L(r), r > 0, \alpha > 0,$$

où L est une fonction à variations lentes.

Le lecteur peut se référer à [?] pour une description de certaines applications de la théorie des variations régulières aux méthodes asymptotiques pour la valorisation d'options.

L'extension de la théorie des variations régulières dans \mathbb{R} aux situations multivariées décrit dans [?, ?] est un outil important pour l'étude des dépendances asymptotiques et le calcul des mesures de risque.

Soit un vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_n)$ à valeurs dans $[0, \infty)^n$. La distribution de \mathbf{X} est à variations régulières multivariées à l'infini avec mesure limite ν si

il existe une fonction $b(t) \uparrow +\infty$ quand $t \rightarrow +\infty$ et une mesure de Radon positive $\nu \neq 0$ telles que

$$t\mathbb{P}[b(t)^{-1}\mathbf{X} \in \cdot] \xrightarrow[t \rightarrow +\infty]{v} \nu, \quad (1.3)$$

dans $\mathbb{E} = [0, \infty]^n \setminus \{0\}$, où \xrightarrow{v} correspond à la convergence vague des mesures. Dans ce cas, la fonction b est à variations régulières. Si on suppose que $\nu(\{x : x_1 > 1, \dots, x_n > 1\}) > 0$ et que la mesure de la frontière de cette ensemble est nulle, alors

$$\mathbb{P}[X_1 > x, \dots, X_n > x] \sim \frac{\nu(\{x : x_1 > 1, \dots, x_n > 1\})}{\nu(\{x : x_1 > 1\})} \mathbb{P}[X_1 > x], \quad x \rightarrow \infty.$$

Des résultats similaires sont donnés dans [?]. Ainsi, dans ce contexte, les variations régulières multivariées permettent de calculer des asymptotiques pour des distributions jointes de vecteurs aléatoires (qui peuvent représenter les rendements d'un portefeuille par exemple).

L'hypothèse (1.3) de variation régulière multivariée implique que les fonctions de répartition des composants de \mathbf{X} sont asymptotiquement équivalents entre eux. Si cela n'est pas vérifié, il est possible d'imposer cette condition sur la copule de \mathbf{X} .

Lorsque $\nu(\{x : x_1 > 1, \dots, x_n > 1\}) = 0$, les variations régulières ne permettent pas de calculer des asymptotiques de façon pertinente pour les probabilités jointes. Dans ce cas, le degré de dépendance peut être mesuré en utilisant les variations régulières cachées (cf. [?] pour une revue détaillée de la théorie des variations régulières cachées). Cette théorie suppose que, outre (1.3), il existe une fonction croissante $b^*(t) \uparrow +\infty$ telle que $\frac{b(t)}{b^*(t)} \rightarrow +\infty$ quand $t \rightarrow +\infty$, et une mesure de Radon ν^* on \mathbb{E}^0 , telle que

$$t\mathbb{P}[b^*(t)^{-1}\mathbf{X} \in \cdot] \xrightarrow[t \rightarrow +\infty]{v} \nu^*,$$

sur \mathbb{E}^0 , where $\mathbb{E}^0 := \mathbb{E} \setminus \bigcup_{i=1}^n L_i$ où

$$L_i = (0, \dots, 0, (0, \infty], 0, \dots, 0),$$

avec $(0, \infty]$ en position i . Intuitivement, les variations régulières cachées impliquent que la mesure ν est concentrée sur les axes et que les probabilités de la forme

$$\mathbb{P}[X_i > tx_i, X_j > tx_j]$$

pour $i \neq j$ décroissent plus rapidement quand $t \rightarrow \infty$ que les fonctions de répartition des composants de \mathbf{X} .

L'hypothèse de variations régulières cachées imposée sur la distribution de \mathbf{X} permet de calculer des asymptotiques à gauche de la somme des composants de \mathbf{X} de façon appropriée. En effet, si on suppose que $\nu^*(\{x : x_1 > 1, \dots, x_n > 1\}) > 0$ et que cette mesure ne charge pas la frontière de cet ensemble, alors

$$\mathbb{P}[X_1 > x, \dots, X_n > x] \sim \frac{\nu^*(\{x : x_1 > 1, \dots, x_n > 1\})}{b^{*(-1)}(x^{-1})},$$

où $b^{*(-1)}$ est l'inverse asymptotique de b^* . Dans le chapitre 5, nous présentons une application de la théorie de variations régulières cachées pour le calcul des mesures de risque pour des portefeuilles d'options valorisées via le modèle de Black-Scholes.

Les grandes déviations. Au sens strict, les grandes déviations permettent de mesurer la vitesse de convergence de la loi des grands nombres. Soit \mathcal{X} un espace de Hausdorff équipé de sa tribu Borélienne. Une fonction de taux est une fonction à valeurs dans $[0, \infty]$ semi-continue inférieurement sur \mathcal{X} . Elle est dite bonne fonction de taux si l'image réciproque des compacts est compacte. Une famille $\{X^\varepsilon\}$ de variables aléatoires à valeurs dans \mathcal{X} satisfait le principe de grandes déviations (LDP) dans \mathcal{X} avec une *bonne fonction de taux* I si pour tout sous-ensemble ouvert $G \subset \mathcal{X}$ et chaque sous-ensemble fermé $F \subset \mathcal{X}$:

Definition 1.4 (Principe des Grandes Déviations).

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{P}[X^\varepsilon \in F] \leq - \inf_{x \in F} I(x)$$

et

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{P}[X^\varepsilon \in G] \geq - \inf_{x \in G} I(x).$$

Il existe de nombreuses situations dans lesquelles un principe de grandes déviations s'applique. Une des situations les plus connues est caractérisée par le théorème de Cramér's dans le cas de variables aléatoires i.i.d. Dans un effort de simplicité, le théorème est donné dans le cas unidimensionnel.

Theorem 1.5. Soit X_1, \dots, X_n, \dots des variables aléatoires i.i.d. distribuées selon la loi de probabilité μ . La log fonction génératrice des moments associée à cette loi est donnée par

$$\Lambda(\lambda) = \log \mathbb{E}[e^{\lambda X_1}],$$

et sa transformée de Fenchel-Legendre par

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \Lambda(\lambda)\}.$$

Soient les moyennes empiriques

$$S_n = \frac{1}{n} \sum_{j=1}^n X_j.$$

Alors la séquence $\{S_n\}$ satisfait un principe de grandes déviations sur \mathbb{R} et sa fonction de taux est Λ^* .

Le théorème de Gärtner-Ellis est un autre résultat important sur les grandes déviations dans lequel l'hypothèse d'indépendance est affaiblie. Dans le chapitre 3, nous utiliserons des résultats de grandes déviations sur les processus de Lévy sur les espaces de trajectoires cadlag.

Le théorème de Varadhan est un autre résultat important que nous utiliserons dans cette thèse pour analyser le comportement asymptotique de fonctionnelles de l'espérance.

Lemma 1.6 (Varadhan's lemma). *Supposons que $\{X_\varepsilon\}$ satisfait un principe de grandes déviations avec une bonne fonction de taux $I : \mathcal{X} \rightarrow [0, \infty[$ et soit $\phi : \mathcal{X} \rightarrow \mathbb{R}$ une fonction continue. Supposons de plus que pour $\gamma > 1$,*

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{E} \left[e^{\frac{\gamma \phi(X_\varepsilon)}{\varepsilon}} \right] < \infty.$$

Alors,

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{E} \left[e^{\frac{\phi(X_\varepsilon)}{\varepsilon}} \right] = \sup_{x \in \mathcal{X}} \{ \phi(x) - I(x) \}.$$

La théorie des grandes déviations a été utilisée dans la littérature pour le calcul d'événements rares pour des applications en finance, en particulier dans le cas de valorisations (cf. [?, ?]), la mesure du risque de crédit (cf. [?, ?]) et des asymptotiques de modèles à volatilité stochastique (cf. [?, ?]). De plus, [?, ?, ?] ont appliqué les méthodes d'échantillonnage d'importance et de grandes déviations dans le cadre de simulations de Monte Carlo.

2 Principales contributions de cette thèse

Cette thèse propose trois contributions principales à la valorisation et la gestion des risques en mathématiques financières. Dans le chapitre 3, nous développons une méthode de réduction de variance pour la valorisation des options exotiques lorsque le sous-jacent suit un modèle exponentiel de Lévy. Le chapitre 4 propose une méthode d'analyse asymptotique de portefeuilles dont les composantes sont modélisées par un mélange gaussien. Ces analyses couvrent notamment les modèles

de Heston et variance-gamme multi-dimensionnels. Ces résultats asymptotiques permettent de construire des méthodes de réduction de variance pour les simulations de Monte Carlo d'évènements rares. Enfin, le chapitre 5 présente des résultats sur le comportement asymptotique de portefeuilles dont la structure de dépendance est modélisée par une copule gaussienne. Ces résultats couvrent notamment le cas de portefeuilles d'options modélisées par le modèle de Black-Scholes. La section suivante présente les résultats principaux associés à chaque sujet.

2.1 Echantillonnage d'importance dans le contexte de processus de Lévy

Dans cette section, nous proposons une méthode efficace pour implémenter des estimateurs de l'espérance de fonctionnelles des processus de Lévy permettant notamment de valoriser des options dans les modèles exponentiels de Lévy. Soit un marché financier composé d'un actif sans risque $S_t^0 \equiv 1$ et de n actifs risqués S^1, \dots, S^n où

$$S_t^i = S_0^i e^{X_t^i},$$

et (X^1, \dots, X^n) est un processus de Lévy tel que S^i est une martingale pour tout i sous la probabilité risque neutre \mathbb{P} . Soient $T < \infty$ un horizon de temps fixé et un actif dérivé de $(S^i)_{1 \leq i \leq n}$ avec un payoff positif $P(S)$ qui dépend de l'ensemble de la trajectoire du sous-jacent jusqu'au temps T . L'objectif est de calculer le prix de l'actif dérivé donné par l'espérance sous la probabilité risque neutre $\mathbb{E}[P(S)]$.

L'estimateur standard de Monte Carlo de $\mathbb{E}[P(S)]$ est défini de la façon suivante:

$$\widehat{P}_N := \frac{1}{N} \sum_{j=1}^N P(S^{(j)}),$$

avec $S^{(j)}, j = 1, \dots, N$, des tirages i.i.d. de même loi que S . Afin d'améliorer la convergence de l'estimateur standard, nous proposons une méthode d'échantillonnage d'importance basée sur la transformée de Escher des trajectoires.

$$\frac{d\mathbb{P}^\theta}{d\mathbb{P}} = \frac{e^{\int_{[0,T]} \langle X_t, \theta(dt) \rangle}}{\mathbb{E} \left[e^{\int_{[0,T]} \langle X_t, \theta(dt) \rangle} \right]}, \quad (1.4)$$

où θ est une mesure signée (déterministe) bornée de $[0, T]$ à valeur dans \mathbb{R}^n . Le choix optimal de θ doit minimiser la variance de l'estimateur sous \mathbb{P}^θ ,

$$\text{Var}_{\mathbb{P}^\theta} \left(P \frac{d\mathbb{P}}{d\mathbb{P}^\theta} \right) = \mathbb{E}_{\mathbb{P}} \left[P^2 \frac{d\mathbb{P}}{d\mathbb{P}^\theta} \right] - \mathbb{E}[P]^2.$$

En notant H le log-payoff de l'option

$$H(X) = \log P(S), \quad X_t^i = \log \frac{S_t^i}{S_0^i}, \quad 0 \leq t \leq T,$$

et le problème de minimisation s'écrit

$$\inf_{\theta \in M} \mathbb{E}_{\mathbb{P}} \left[\exp \left\{ 2H(X) - \int_{[0,T]} \langle X_t, \theta(dt) \rangle + \int_{[0,T]} G(\theta([t, T])) dt \right\} \right],$$

où

$$G(\theta) = \frac{1}{2} \langle \theta, A\theta \rangle + \langle \theta, \gamma \rangle + \int_{\mathbb{R}^n} (e^{\langle \theta, x \rangle} - 1 - \langle \theta, x \rangle \mathbf{1}_{|x| \leq 1}) \nu(dx),$$

et (A, ν, γ) est le triplet caractéristique du processus X .

Inspiré du travail de Guasoni et Robertson [?] dans le cas Black-Scholes, la principale contribution de ce développement est d'utiliser la théorie des grandes déviations pour obtenir une approximation relativement simple à implémenter du paramètre de réduction de variance optimale θ_{opt}^* . Plus précisément, nous utilisons le lemme de Varadhan et le principe de grandes déviations trajectorien appliqué aux processus de Lévy développé par Leonard [?] pour calculer une approximation du paramètre de réduction de variance optimal.

Dans un premier temps, nous obtenons une expression pour l'approximation de la variance que l'on cherche ensuite à minimiser asymptotiquement pour obtenir le paramètre de réduction de variance optimal.

Pour cela, nous avons besoin de l'hypothèse suivante:

$$(A1) \text{ Il existe } \lambda_0 > 0 \text{ avec } \int_{|x| > 1} e^{\lambda_0 |x|} \nu(dx) < \infty.$$

Dans la proposition suivante et celles qui suivront, D représente l'espace de fonctions continues à droite possédant des limites à gauche sur l'intervalle $[0, T]$, M représente la classe des mesures bornées sur $[0, T]$ à valeurs dans \mathbb{R}^n , et $\sigma(D, M)$ correspond à la topologie sur D définie par

$$\lim_n y_n = y \Leftrightarrow \forall \mu \in M, \lim_n \int_{[0,T]} y_n d\mu = \int_{[0,T]} y d\mu.$$

Proposition 1.7. *Supposons que l'hypothèse (A1) est vérifiée, et que l'ensemble $\{x \in D : H(x) > -\infty\}$ est ouvert pour la topologie $\sigma(D, M)$ et contient 0. Supposons de plus que H est continue sur cet ensemble pour la topologie $\sigma(D, M)$ ainsi que H et $\theta \in M$ vérifient une des hypothèses suivantes:*

(i) Il existe des constantes $C < \infty$ et $B < \lambda_0/4n$ telles que

$$H(x) \leq C + B \sup_{s \in [0, T]} \sum_{i=1}^n |x_s^i|$$

et

$$\max_{0 \leq t \leq T} |\theta([t, T])| < \lambda_0 - 4nB.$$

(ii) H est bornée et

$$\int_{[0, T]} G(\gamma\theta([t, T])) dt < \infty$$

pour certains $\gamma > 1$.

Alors

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{E} \left[e^{\frac{2H(X^\varepsilon) - \int_{[0, T]} \langle X_t^\varepsilon, \theta(dt) \rangle}{\varepsilon}} \right] = \sup_{x \in D} \left\{ 2H(x) - \int_{[0, T]} \langle x_t, \theta(dt) \rangle - J(x) \right\}.$$

où X^ε est défini tel que: $X_t^\varepsilon = \varepsilon X_{t/\varepsilon}$ et la fonction J est définie comme:

$$J(x) = \begin{cases} \sup_{\mu \in M} \left\{ \int_{[0, T]} \langle x_t, \mu(dt) \rangle - \int_{[0, T]} G(\mu([t, T])) dt \right\} & \text{si } x \in V_r, \\ + \infty & \text{sinon.} \end{cases}$$

où V_r représente le sous-espace de D contenant l'ensemble des fonctions dans D à variations bornées.

D'après ce résultat, un candidat pour le paramètre de réduction de variance optimal θ^* s'obtient en minimisant sur $\theta \in M$ l'expression suivante:

$$\sup_{x \in D} \left\{ 2H(x) - \int_{[0, T]} \langle x_t, \theta(dt) \rangle + \int_{[0, T]} G(\theta([t, T])) dt - J(x) \right\}. \quad (1.5)$$

Il est alors naturel de se demander à quel point la mesure obtenue est "éloignée" de la mesure de réduction de variance optimale. Dans ce contexte, le lemme de Varadhan permet de définir la notion d'optimalité asymptotique. Soit une famille de mesures d'échantillonnage d'importance $(\mathbb{Q}(\varepsilon))_{\varepsilon > 0}$. D'après l'inégalité de Jensen,

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E}^{\mathbb{Q}(\varepsilon)} \left[\left(e^{\frac{1}{\varepsilon} H(X^\varepsilon)} \frac{d\mathbb{P}}{d\mathbb{Q}(\varepsilon)} \right)^2 \right] \geq 2 \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[e^{\frac{1}{\varepsilon} H(X^\varepsilon)} \right].$$

Ainsi, nous dirons qu'une famille de changement de mesure d'échantillonnage d'importance est optimale, si, pour cette famille, l'inégalité ci-dessus est une égalité et les liminf/limsup deviennent des limites. En d'autres termes, un changement de mesure optimal doit faire au moins aussi bien que tout autre changement de mesure sur une échelle logarithmique.

Definition 1.8. Soit $(\mathbb{Q}(\varepsilon))_{\varepsilon>0}$ une famille de changement de mesures d'échantillonnage d'importance. Nous dirons que $(\mathbb{Q}(\varepsilon))_{\varepsilon>0}$ est asymptotiquement optimale si

$$\lim_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E}^{\mathbb{Q}(\varepsilon)} \left[\left(e^{\frac{1}{\varepsilon} H(X^\varepsilon)} \frac{d\mathbb{P}}{d\mathbb{Q}(\varepsilon)} \right)^2 \right] = 2 \lim_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[e^{\frac{1}{\varepsilon} H(X^\varepsilon)} \right].$$

Le théorème suivant, qui constitue le deuxième résultat principal, montre que dans le cas d'un log-payoff concave, la mesure d'échantillonnage d'importance θ^* calculée en minimisant l'expression (1.5) est asymptotiquement optimale. En d'autres termes, pour de tels payoffs, il suffit de considérer les mesures de la forme (1.4) avec θ déterministe pour être asymptotiquement optimal. En pratique, les options de type put européens sur des paniers ainsi que différentes options trajectoire-dépendantes de type put ont des payoffs concaves. Nous démontrons ce théorème en supposant l'hypothèse suivante:

(A2) La fonction G est semi-continue inférieurement et son domaine $\mathcal{D} := \{x \in \mathbb{R}^n : G(x) < \infty\}$ est ouvert et borné.

Theorem 1.9. Soit H concave, et nous supposons que l'ensemble $\{x \in V_r : H(x) > -\infty\}$ est non vide et contient 0. Nous supposons également que H est continue sur cet ensemble au sens de la topologie $\sigma(D, M)$. Supposons également que les hypothèses (A1) et (A2) sont vérifiées. Alors,

$$\begin{aligned} \inf_{\theta \in M} \sup_{x \in V_r} \left\{ 2H(x) - \int_{[0,T]} \langle x_t, \theta(dt) \rangle + \int_{[0,T]} G(\theta([t, T])) dt - J(x) \right\} \\ = 2 \inf_{\theta \in M} \left\{ \widehat{H}(\theta) + \int_{[0,T]} G(\theta([t, T])) dt \right\} \quad (1.6) \end{aligned}$$

avec

$$\widehat{H}(\theta) = \sup_{x \in V_r} \left\{ H(x) - \int_{[0,T]} \langle x_t, \theta(dt) \rangle \right\}.$$

De plus, si l'infimum à droite de l'expression (1.6) est atteint par θ^* , alors la même valeur θ^* atteint l'infimum à gauche de l'expression (1.6). Si, de plus, le payoff H et θ^* vérifient (i) ou (ii) de la proposition 3.5 alors la mesure d'échantillonnage d'importance correspondante à θ^* est asymptotiquement optimale.

Ce travail a fait l'objet d'une pré-publication avec P. Tankov qui est en cours de revue pour Stochastic Processes and Their Applications.

2.2 Analyse asymptotique des portefeuilles suivants des mélanges log-normaux

Dans cette section, nous considérons le comportement asymptotique de la somme de n variables aléatoires non indépendantes et positives

$$X = \sum_{i=1}^n X_i.$$

Dans un contexte financier, X peut représenter la valeur d'un portefeuille (long en actifs seulement) contenant n actifs. Comprendre le comportement asymptotique de X est important du point de vue de la gestion des risques notamment dans les contextes de calculs de Value-at-Risk, de mesures d'évènements rares ou encore pour la simulation d'évènements extrêmes. En particulier, les scénarios de stress test peuvent être construits sur la base de simulations des composants X_1, \dots, X_n conditionnellement aux valeurs de X .

Ce problème a été largement traité dans la littérature notamment dans le contexte de l'assurance où les variables aléatoires X_1, \dots, X_n représentent les pertes des actifs et l'étude asymptotique de X permet d'estimer la probabilité de pertes importantes du portefeuille agrégé. Si les distributions des variables aléatoires X_1, \dots, X_n sont à queues suffisamment épaisses (sous-exponentielle), alors sous certaines hypothèses sur la structure de dépendance, il est possible de montrer que le comportement asymptotique à *droite* de X est déterminé par la variable aléatoire dont la distribution a la queue la plus épaisse. Dans cette thèse, nous nous concentrons sur la gestion des risques financiers où l'on s'intéresse aux *petites* valeurs de X . Ainsi, afin d'estimer la probabilité de pertes importantes, nous nous intéressons au comportement asymptotique à *gauche* de X . Par la positivité des variables aléatoires X_1, \dots, X_n , les comportements asymptotiques à droite et à gauche sont très différents.

Dans cette étude, nous calculons l'asymptotique à gauche de la fonction de répartition et de la densité de X sous l'hypothèse que X_1, \dots, X_n sont distribuées selon un mélange log-normal. De plus, nous proposerons des applications de ces développements. Plus précisément, nous supposons que pour $i = 1, \dots, n$, $X_i = e^{Y_i}$, où le vecteur $\mathbf{Y} = (Y_1, \dots, Y_n)$ suit un mélange gaussien.

Un mélange gaussien sur \mathbb{R}^n est défini de la façon suivante:

$$\mathbf{Y} = \sqrt{\Theta} \mathbf{Z} + \Theta \boldsymbol{\mu} + \boldsymbol{\Lambda}$$

où $\boldsymbol{\mu} \in \mathbb{R}^n$ et $\boldsymbol{\Lambda} \in \mathbb{R}^n$ sont des vecteurs constants, $\mathbf{Z} \sim N(0, \mathfrak{B})$ avec $\mathfrak{B} \in \mathbb{M}^{n \times n}$ une matrice symétrique, définie positive et $\Theta \in \mathbb{R}_+$ est la variable aléatoire de mélange.

Dans ce travail, nous supposons que Θ admet une densité, notée $\rho(s)$, qui décroît de façon exponentielle lorsque $s \rightarrow \infty$.

Rappelons quelques notations que nous utiliserons par la suite. Pour deux fonctions f et g , on note $f(x) \sim g(x)$ quand $x \rightarrow a$ lorsque $\frac{f(x)}{g(x)} \rightarrow 1$ quand $x \rightarrow a$. La notation $\mathbf{1}$ représente un vecteur en dimension n dont tous les composants valent 1. De façon générale, les vecteurs seront notés en gras, de sorte que $\mathbf{w} = (w_1, \dots, w_n)^\perp$, où \perp représente la matrice transposée. Pour $\mathbf{v} \in \mathbb{R}_+^n$, on note $\mathcal{E}(\mathbf{v}) = -\sum_{i=1}^n v_i \log v_i$ avec la convention usuelle $0 \log 0 = 0$. Les éléments de la matrice de variance covariance \mathfrak{B} seront notés $(b_{ij})_{1 \leq i, j \leq n}$ et ceux de la matrice inverse \mathfrak{B}^{-1} , $(a_{ij})_{1 \leq i, j \leq n}$. Δ_n représente le simplexe en dimension n défini par:

$$\Delta_n = \left\{ \mathbf{w} \in \mathbb{R}^n : w_i \geq 0, i = 1, \dots, n \text{ and } \sum_{i=1}^n w_i = 1 \right\}.$$

Les fonctions seront appliquées à chaque composants des vecteurs, de sorte que $\log(\mathbf{w})$ équivaut à $(\log w_1, \dots, \log w_n)^\perp$. De façon similaire, les inégalités sur des vecteurs telle que $\mathbf{w} > 0$ doivent être lues composants par composants. Enfin, rappelons que la fonction beta est définie de la façon suivante:

$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt, \quad x, y > 0$$

Pour $\boldsymbol{\alpha} \in \mathbb{R}^n$, $n \geq 2$, avec $\alpha_i > 0$, $i = 1, \dots, n$, la fonction beta multidimensionnelle est définie par:

$$B(\boldsymbol{\alpha}) := \int_{\Delta_n} \prod_{i=1}^n w_i^{\alpha_i-1} d\mathbf{w} = \frac{\prod_{i=1}^n \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^n \alpha_i)}.$$

Par convention, $B(\boldsymbol{\alpha}) = 1$ pour tout $\boldsymbol{\alpha} \in \mathbb{R}$.

Principaux résultats Considérons la fonction suivante:

$$F(t, \mathbf{w}) = \theta t + \frac{(1 + t\boldsymbol{\mu}^\perp \mathbf{w})^2}{2\mathbf{w}^\perp \mathfrak{B} \mathbf{w} t}.$$

Le lemme suivant, démontré dans [?], établit certaines propriétés de cette fonction.

Lemma 1.10. *Il existe un unique couple $(\bar{t}, \bar{\mathbf{w}})$, avec $\bar{t} \in (0, \infty)$ et $\bar{\mathbf{w}} \in \Delta_n$ tel que*

$$F(\bar{t}, \bar{\mathbf{w}}) = \min_{t>0} \max_{\mathbf{w} \in \Delta_n} F(t, \mathbf{w}).$$

De plus, la fonction

$$f(t) = F(t, \bar{\mathbf{w}})$$

possède un unique minimum atteint au point \bar{t} .

Dans la suite, nous ferons les hypothèses suivantes:

(A1) Le point auquel le minimum est atteint \bar{t} satisfait

$$\mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu}) > 0.$$

(A2) $\rho(t)$ admet l'écriture suivante:

$$\rho(t) = e^{-\theta t + m\sqrt{t}} \rho_0(t)$$

avec $\theta > 0$, $m \in \mathbb{R}$ et ρ_0 une fonction à variations régulières d'indice α lorsque $t \rightarrow \infty$.

L'hypothèse (A1) est relativement forte et nous la discutons en détail dans le chapitre dédié. Le théorème suivant constitue le résultat principal qui établit une formule asymptotique pour la fonction de répartition et la densité de la variable aléatoire X .

Theorem 1.11. *Sous les hypothèses (A1)-(A2), la densité de X , $p(z)$, satisfait*

$$\begin{aligned} p(z) &\sim \frac{(2\pi)^{\frac{1-n}{2}} \bar{t}^{\alpha + \frac{3-n}{2}}}{\sqrt{\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1}} \det \mathfrak{B}} B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu})\right) \exp\left(\frac{m^2 \bar{t}^2}{8\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1}} - \frac{1}{\bar{t}} (\mathbf{1} + \bar{t}\boldsymbol{\mu})^\perp \mathfrak{B}^{-1} \boldsymbol{\Lambda}\right) \\ &\times \frac{\rho_0\left(\log \frac{1}{z}\right)}{z \left(\log \frac{1}{z}\right)^{\frac{n-1}{2}}} e^{-\log \frac{1}{z} F(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t} \log \frac{1}{z}}}. \end{aligned}$$

lorsque $z \rightarrow 0$, et la fonction de répartition

$$\begin{aligned} F(x) &\sim \frac{(2\pi)^{\frac{1-n}{2}} \bar{t}^{\alpha + \frac{3-n}{2}}}{F(\bar{t}, \bar{\mathbf{w}}) \sqrt{\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1}} \det \mathfrak{B}} B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu})\right) \\ &\times \exp\left(\frac{m^2 \bar{t}^2}{8\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1}} - \frac{1}{\bar{t}} (\mathbf{1} + \bar{t}\boldsymbol{\mu})^\perp \mathfrak{B}^{-1} \boldsymbol{\Lambda}\right) \\ &\times \rho_0\left(\log \frac{1}{x}\right) \left(\log \frac{1}{x}\right)^{\frac{1-n}{2}} e^{-\log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t} \log \frac{1}{x}}} \end{aligned}$$

lorsque $x \rightarrow 0$

Le théorème 4.3 permet d'estimer les espérances conditionnelles associées. Le corollaire suivant caractérise le comportement asymptotique de la transformée de Laplace conditionnelle de Y_1, \dots, Y_n , étant donné que $X \leq x$.

Corollary 1.12. *Supposons que les hypothèses (A1)-(A2) sont vérifiées et soit $\mathbf{u} \in \mathbb{R}^n$ fixé. On note:*

$$\begin{aligned} C_{\mathbf{u}} &= \boldsymbol{\mu}^\perp \mathbf{u} + \frac{1}{2} \mathbf{u}^\perp \mathfrak{B} \mathbf{u} \\ \boldsymbol{\mu}_{\mathbf{u}} &= \boldsymbol{\mu} + \mathfrak{B} \mathbf{u} \\ \theta_{\mathbf{u}} &= \theta - C_{\mathbf{u}} \\ F_{\mathbf{u}}(t, w) &= \theta_{\mathbf{u}} t + \frac{(1 + t \boldsymbol{\mu}_{\mathbf{u}}^\perp \mathbf{w})^2}{2 \mathbf{w}^\perp \mathfrak{B} \mathbf{w} t} \end{aligned}$$

Si $\theta_{\mathbf{u}} > 0$ et $\mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu}_{\mathbf{u}}) > 0$, alors lorsque $x \rightarrow 0$,

$$\begin{aligned}\mathbb{E} \left[e^{\sum_{i=1}^n u_i Y_i} | X \leq x \right] &\sim x^{1+\mathbf{u}} \frac{F(\bar{t}, \bar{\mathbf{w}})}{F(\bar{t}, \bar{\mathbf{w}}) + \mathbf{1}^\perp \mathbf{u}} \frac{B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu}) + \mathbf{u}\right)}{B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu})\right)} \\ \mathbb{E} \left[e^{\sum_{i=1}^n u_i Y_i} | X = x \right] &\sim x^{1+\mathbf{u}} \frac{B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu}) + \mathbf{u}\right)}{B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu})\right)}\end{aligned}$$

Le corollaire suivant établit une formule asymptotique pour la Value-at-Risk.

Corollary 1.13. *Sous les hypothèses du théorème 4.3, la Value-at-Risk au niveau $1 - y$ satisfait*

$$\text{VaR}_{1-y}(X) \sim \left(y \frac{\left(F^{-1} \log \frac{1}{y}\right)^{\frac{n-1}{2}}}{C \rho_0 \left(F^{-1} \log \frac{1}{y}\right)} \exp \left(-m \sqrt{\bar{t} F^{-1} \log \frac{1}{y} - \frac{m^2 \bar{t}}{2F}} \right) \right)^{\frac{1}{F}},$$

où nous avons noté $F = F(\bar{t}, \bar{\mathbf{w}})$ et

$$C = \frac{(2\pi)^{\frac{1-n}{2}} \bar{t}^{\alpha + \frac{3-n}{2}}}{F \sqrt{\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1} \det \mathfrak{B}}} B \left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu}) \right) \times \exp \left(\frac{m^2 \bar{t}^2}{8 \mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1}} - \frac{1}{\bar{t}} (\mathbf{1} + \bar{t}\boldsymbol{\mu})^\perp \mathfrak{B}^{-1} \boldsymbol{\Lambda} \right)$$

2.3 Asymptotiques de distributions à queues épaisses et structure de dépendance à copule Gaussienne

Dans cette section, nous montrons qu'une distribution multivariée dont les marginales sont à queues épaisses (à variations régulières) et dont la structure de dépendance est décrite par une copule Gaussienne est à variations régulières sur le cône $(0, \infty)^n$, sans qu'il soit nécessaire que les marginales sont identiquement distribuées et que la distribution est à variations régulières sur $[0, \infty)^n$. Ceci nous permettra ensuite d'obtenir des formules asymptotiques de fonctionnelles, telle que la somme des composants, de vecteurs aléatoires à queues épaisses avec une structure de dépendance donnée par une copule Gaussienne. Nous développerons également des applications pour le calcul de mesures de risque sur des portefeuilles d'options modélisées par le modèle de Black-Scholes ainsi que pour des fermes de production d'énergie éolienne.

Soient X_1, \dots, X_n des variables aléatoires à valeurs dans (a, ∞) avec $a \in [-\infty, \infty)$, dont les fonctions de distributions sont notées F_1, \dots, F_n . Nous supposons que la dépendance des éléments du vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_n)$ est décrite par une copule Gaussienne, c'est-à-dire que, $X_k = F_k^{-1} \left(\mathcal{N} \left(\tilde{X}_k \right) \right)$ pour $k = 1, \dots, n$, où $\tilde{\mathbf{X}} = \left(\tilde{X}_1, \dots, \tilde{X}_n \right)$ est un vecteur Gaussien centré réduit et de matrice de corrélation Σ , \mathcal{N} est la fonction de distribution de loi normale standard et F_k^{-1}

représente l'inverse généralisée de F_k .

Notre résultat principal permet de quantifier le comportement asymptotique de X_1, \dots, X_n sous l'hypothèse que les composantes du vecteur sont à variations régulières. Avant d'énoncer le résultat, nous introduisons les notations nécessaires.

Soit $\widetilde{\mathbf{X}}$ un vecteur gaussien dont les marginales sont centrées réduites et dont la matrice de corrélation Σ est définie positive. On note B la matrice inverse de Σ . Pour $\mathbf{x} \in \mathbb{R}_+^n$ fixé, tel que $\mathbf{x} \neq 0$, on note \mathbf{x}^* la solution (le minimiseur) du problème suivant:

$$\min_{\mathbf{z} \geq \mathbf{x}} \langle \mathbf{z}, B\mathbf{z} \rangle. \quad (1.7)$$

Alors (cf. Proposition 2 dans [?]), il existe de façon unique, 2 ensemble d'indices disjoints I et J , avec $I \cup J = \{1, \dots, n\}$, tel que

$$\mathbf{x}_I^* = \mathbf{x}_I, \quad \mathbf{x}_J^* = -B_{JJ}^{-1}B_{JI}\mathbf{x}_I, \quad (1.8)$$

$$\text{et } \langle \mathbf{e}_i, \Sigma_{II}^{-1}\mathbf{x}_I \rangle > 0, \quad \forall i \in I, \quad (1.9)$$

où dans la notation précédente, l'indice de la matrice est choisi avant l'inverse.

Theorem 1.14. *Soient X_1, \dots, X_n des variables aléatoires à valeurs dans (a, ∞) , dont les fonctions de répartition sont notées F_1, \dots, F_n . Supposons que la structure de dépendance du vecteur $\mathbf{X} = (X_1, \dots, X_n)$ est donnée par une copule gaussienne de matrice de corrélation Σ , et que pour tout i , la fonction de survie $\tilde{F}_i = 1 - F_i$ est à variations régulières en $+\infty$ d'indice $-\alpha_i < 0$. Notons $\boldsymbol{\beta} = (\sqrt{\alpha_1}, \dots, \sqrt{\alpha_n})$, $\boldsymbol{\beta}^*$ la solution du problème*

$$\min_{\mathbf{z} \geq \boldsymbol{\beta}} \langle \mathbf{z}, B\mathbf{z} \rangle,$$

ainsi que I et J les ensembles d'indices correspondants. Supposons que $\boldsymbol{\beta}_J^ > \boldsymbol{\beta}_J$. Alors pour $x_1 > 0, \dots, x_n > 0$, lorsque $u \uparrow \infty$,*

$$\begin{aligned} & \mathbb{P}[X_1 > ux_1, \dots, X_n > ux_n] \\ &= \exp \left\{ - \sum_{i,j \in I} b_{ij} \sqrt{\ln \tilde{F}_i(u) \ln \tilde{F}_j(u)} \right\} L(u) \nu(\mathbf{x}) (1 + o(1)) \end{aligned}$$

où

$$\begin{aligned} L(u) &= C \ln u^{\frac{1}{2} \sum_{i,j} b_{ij} \frac{\beta_i}{\beta_j} - \frac{|I|}{2}} \\ C &= \prod_{i \in I} \beta_i^{\frac{1}{\beta_i} \sum_{j \in I} b_{ij} \beta_j} \frac{(4\pi)^{\frac{1}{2} \sum_{i,j \in I} b_{ij} \frac{\beta_i}{\beta_j} - \frac{|I|}{2}}}{|\Sigma_{II}|^{\frac{1}{2}} \prod_{i \in I} \langle \mathbf{e}_i, \Sigma_{II}^{-1} \boldsymbol{\beta}_I \rangle} \\ \nu(\mathbf{x}) &= \prod_{i \in I} x_i^{-\beta_i \sum_{j \in I} b_{ij} \beta_j} \end{aligned}$$

Remark 1.15. Le théorème ci-dessus montre que le vecteur \mathbf{X} possède la propriété de variations régulières sur le cône $(0, \infty)^n$ mais pas nécessairement sur $[0, \infty)^n \setminus \{0\}$. En effet, comme les composantes du vecteur ne sont pas identiquement distribuées la propriété de variations régulières sur $[0, \infty)^n$ n'est pas systématiquement vérifiée.

Le corollaire suivant permet de quantifier le comportement asymptotique à gauche de la somme des composantes d'un vecteur à queues épaisses dont la structure de dépendance est caractérisée par une copule gaussienne.

Corollary 1.16. *Soient X_1, \dots, X_n des variables aléatoires à valeurs dans $(0, \infty)$, dont les fonctions de distribution sont notées F_1, \dots, F_n . Supposons que la structure de dépendance du vecteur $\mathbf{X} = (X_1, \dots, X_n)$ est donnée par une copule gaussienne de matrice de corrélation Σ . Supposons également que pour tout i , les fonctions de répartition F_i sont à variations régulières en 0 d'indice $\alpha_i > 0$. Notons $\boldsymbol{\beta} = (\sqrt{\alpha_1}, \dots, \sqrt{\alpha_n})$, $\boldsymbol{\beta}^*$ la solution du problème*

$$\min_{\mathbf{z} \geq \boldsymbol{\beta}} \langle \mathbf{z}, B\mathbf{z} \rangle,$$

ainsi que I et J les ensembles d'indices correspondants. Supposons que $\beta_J^* > \beta_J$. Alors, lorsque $z \downarrow 0$,

$$\mathbb{P}[X_1 + \dots + X_n < z] = \exp \left\{ - \sum_{i,j \in I} b_{ij} \sqrt{\ln F_i(z) \ln F_j(z)} \right\} L \left(\frac{1}{z} \right) \nu_0(A)$$

où L est la fonction définie dans le théorème 5.2 et

$$\nu_0(A) = \prod_{k=1}^{|I|-1} p_{i_k} B \left(p_{i_k}, 1 + \sum_{j=k+1}^{|I|} p_{i_j} \right).$$

$$p_i = \beta_i \sum_{j \in I} b_{ij} \beta_j.$$

Ici B correspond à la fonction Beta d'Euler, les éléments de l'ensemble I sont notés i_k pour $k = 1, \dots, |I|$ (dans l'ordre croissant) et le produit sur un ensemble vide vaut 1 par convention.

Chapter 2

Introduction

Contents

1	Challenges and methods of financial risk management .	38
1.1	Pricing of derivative instruments	38
1.2	Measuring the risk	41
1.3	Rare events simulation and tail analysis	44
2	Main contributions of this thesis	50
2.1	Optimal importance sampling for Lévy processes	50
2.2	Tail asymptotics of log-normal mixture portfolios	54
2.3	Asymptotics of heavy-tailed risks with Gaussian copula dependence	57

This thesis presents three contributions to the theory of financial risk management, based on asymptotic methods. In this introductory chapter, we first review the main challenges and methods of financial risk management, and show how the asymptotic methods may prove useful to address these challenges in a rare event context. We then present in detail the main contributions of this thesis.

1 Challenges and methods of financial risk management

1.1 Pricing of derivative instruments

Pricing of derivative instruments is a cornerstone of risk management since the pricing of illiquid assets in the bank's portfolio, required for the computation of the regulatory capital charge, is done through a model. The selection of an appropriate model is then key to achieve this goal. A good model should reproduce stylized facts observed on the markets in order to accurately estimate the risks and provide an appropriate operational framework to compute prices and risk measures. Below, we briefly review the main models used in the banking sector for the pricing of derivatives.

Black-Scholes model The well-known pricing model developed by Black, Scholes and Merton in 1973 [?] provides a framework for the pricing of European style options with an explicit formula. Assuming that the underlying asset follows a geometric Brownian motion with volatility σ , the price of a European call option written on an underlying S at time t with maturity T and strike K , with interest rate r , is given by:

$$C(t, \sigma; T, K) = S_t N(d_1) - K e^{-r(T-t)} N(d_2)$$

where S_t is the price of the underlying at time t , N is the standard normal distribution function and

$$d_1 = \frac{\log \frac{S_t}{K e^{-r(T-t)}} + (T-t) \frac{\sigma^2}{2}}{\sigma \sqrt{T-t}}$$

$$d_2 = d_1 - \sigma \sqrt{T-t}.$$

The option price in this model depends on two types of parameters: market parameters such as the volatility and eventually interest rates or dividends (here considered as null) and contractual parameters, the strike and the maturity of the option. All of them are observables except the volatility. In the case of European style options, the associated market is often very liquid which means that, at any time, there are

enough buyers and sellers so that the quoted prices reflect the equilibrium between them. Therefore, for a quoted option, the value of the volatility in the model is to be such that the option's model price equals its market price. In the Black-Scholes model, the price is bijective with respect to the volatility and thus one can solve the inverse problem and find the so-called implied volatility. In other words, the implied volatility is defined as the parameter $\sigma_{T,K}^{imp}$ such that

$$C(t, \sigma_{T,K}^{imp}; T, K) = C^{mkt}(t; T, K)$$

where $C^{mkt}(t; T, K)$ is the quoted price of the same option at time t . This parameter, assumed to be equal for all strikes and maturities, can then be used to price non-quoted options. The main advantage of this formula is that the price is given by an explicit formula and the calibration of the model is relatively straightforward to implement.

However, empirical observations have shown that the implied volatility is not constant across strikes and maturities. Indeed, the mapping $T, K \mapsto \sigma_{T,K}^{imp}$, called the implied volatility surface, is not constant and contradicts one major assumption of the model. The dependence of the implied volatility on strike and maturity, known as the implied volatility smile phenomenon, shows that the Black-Scholes model is not supported by the option market data. In addition, the Black-Scholes model relies on the assumption that the stock returns are normally distributed but this assumption is also contradicted by empirical observations which provide evidence of a fat-tailed distribution of the returns.

In order to deal with the observations of the implied volatility surface and fat-tailed distributions, the Black-Scholes model has been extended. Some of these models are briefly presented hereafter.

Local volatility models The local volatility model assumes that the volatility is a deterministic function of time and stock price. It is defined as follows:

$$\frac{dS_t}{S_t} = \mu dt + \sigma(t, S_t) dW_t$$

where $\mu \in \mathbb{R}$ and W is a standard Brownian motion. This model provides a perfect fit of the implied volatility surface. Indeed, for any given arbitrage-free surface at time $t = 0$, there exists a unique function $\sigma(t, x)$ that matches all the option prices [?]. Nevertheless, this model predicts a wrong dynamic for the surface. Indeed, the forecasted future smile is usually flatter than the observed future smile. A better description of the implied volatility dynamics is provided by the stochastic volatility models.

Stochastic volatility models In stochastic volatility models the stock price process is modelled as the first component of a two-dimensional process (S, σ) driven by a bi-dimensional Brownian motion (W^1, W^2) ,

$$\frac{dS_t}{S_t} = \mu dt + \sigma_t dW_t^1 \quad (2.1)$$

$$\sigma_t = f(Y_t) \quad dY_t = a_t dt + b_t dW_t^2. \quad (2.2)$$

The volatility is assumed to be a random process possibly correlated with the stock trajectories. Two widely used stochastic volatility models for financial risk management are the SABR model [?] and the Heston model [?].

Stochastic volatility models generally provide a good fit of the smile and enjoy better dynamical properties than local volatility but the skew (the slope of the volatility smile) for short maturities is often too low. Indeed, while an appropriate choice of diffusion coefficients can lead to an arbitrary fat-tailed distribution, such calibration, when carried out for short maturities, often results in unrealistic values of volatility of volatility (the parameter b in (2.2)). These models are then not fully adapted for pricing purposes. The strong implied volatility skew of short maturity options means that these options have higher prices than what is compatible with diffusion models. This effect reflects a fear of asymmetric large downward moves of market participants. Adding jumps to the diffusion model could thus be an answer to both high amplitude and peaked returns as well as stylized effects observed on the implied volatility surface. A convenient class of asset price models with jumps is provided by Lévy processes which are briefly described below.

Lévy processes

Lévy processes are stochastic processes with independent and stationary increments. For a detailed review of the main results related to Lévy processes and their applications to financial modelling, the interested reader can refer to [?]. The law of a Lévy process, $(X_t)_{t \geq 0}$ is determined by the law of X_t at any given time $t \geq 0$, and can be characterized by its characteristic function. It is given by the Lévy-Khintchine formula: for $s \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E} [e^{isX_t}] &= e^{t\Psi(s)} \\ \Psi(s) &= is\gamma - \frac{\sigma^2 s^2}{2} + \int_{\mathbb{R}} (e^{isx} - 1 - isx\mathbf{1}_{|x| \leq 1}) \nu(dx). \end{aligned}$$

The triplet (σ^2, ν, γ) is called the characteristic triplet of the Lévy process with $\gamma \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}_+$, and ν a positive measure, which satisfies the integrability condition

$$\int_{\mathbb{R}} (1 \wedge |x|^2) \nu(dx) < \infty.$$

The Lévy-Itô decomposition gives a representation of the paths of X :

$$X_t = \gamma t + \sigma B_t + \int_0^t \int_{|x| \leq 1} x \tilde{J}(ds \times dx) + \int_0^t \int_{|x| > 1} x J(ds \times dx)$$

where B is a standard Brownian motion, J is a Poisson random measure with intensity measure $dt \times \nu(dx)$ and \tilde{J} is the compensated version of J .

The Variance Gamma model [?] and the Normal Inverse Gaussian model [?] are two examples of models based on Lévy processes which are used for financial modelling with jumps.

1.2 Measuring the risk

Risk measures as tools for mitigation and control provide a comprehensive framework to make tradeoffs between risks and returns and between different types of risks. Among many others, risk management has three key objectives. First, it enables to reduce the volatility of portfolio returns, second it allows to diversify the positions in order to limit concentrations, and third, it measures exposures to extreme economic scenarios in order to identify those causing extreme losses.

Value-at-Risk

The well known Value-at-Risk allows to analyze portfolio market risk based on assumptions on the returns. One of its main advantages is that it provides a synthetic view of the risk held even in large and complex portfolios. Value-at-Risk is related to the quantile of the profit and loss distribution. It represents the level of loss that could be exceeded with a given probability and in a given time horizon. More precisely, if V_t denotes the (known) value of the portfolio at time t , τ the time horizon considered, then the Value-at-Risk at level $\alpha \in (0, 1)$, denoted $\text{VaR}_{\alpha, \tau}$, is defined as the solution to the following equation:

$$\text{VaR}_{\alpha, \tau} = \inf\{x : \mathbb{P}[V_{t+\tau} - V_t \leq -x] \leq 1 - \alpha\}.$$

There exists three main approaches to compute the VaR which are briefly described thereafter.

The historical simulation approach. The historical simulation approach consists in deducing the profit and loss distribution of the portfolio from the observations

of the historical movements of the risk factors. These movements are then replicated on the actual portfolio giving an historical distribution from which the desired percentile can be easily computed. Thus, this method is quite simple to implement but its main drawback is its strong dependance upon the chosen period which may not contain extreme negative returns. Hence, it is mainly used to compute the Stressed-VaR, which explicitly integrates adverse scenarios for the risk factors.

The delta-normal method and delta-gamma method. In VaR computations, the value of the portfolio is assumed to be a function of a set of n risk factors denoted $(x^{(1)}, \dots, x^{(n)})$

$$V = f(x^{(1)}, \dots, x^{(n)})$$

The delta-normal method assumes that the variation of the value of the portfolio can be approximated by:

$$V_{i+1} - V_i \approx \sum_{j=1}^n \frac{\partial f}{\partial x_i^{(j)}} (x_{i+1}^{(j)} - x_i^{(j)}) = \sum_{j=1}^n \frac{\partial f}{\partial x_i^{(j)}} x_i^{(j)} r_{i+1}^{(j)}$$

where $r_{i+1}^{(j)} = \frac{x_{i+1}^{(j)}}{x_i^{(j)}} - 1$ is the return of the risk factor j . In addition, if one assumes that the returns are normally distributed with mean 0 and covariance matrix constant over time, then the variance of the portfolio, denoted σ_V^2 , can immediately be deduced through

$$\sigma_{V,i+1}^2 = \sum_{j=1}^n \sum_{k=1}^n \sigma^{(j)} \sigma^{(k)} \rho^{(j,k)} \frac{\partial f}{\partial x_{i+1}^{(j)}} x_i^{(j)} \frac{\partial f}{\partial x_{i+1}^{(k)}} x_i^{(k)}.$$

The VaR at level p and over one period of time of the portfolio can then be computed as

$$VaR_{i,i+1} = \sigma_{V,i+1} N^{-1}(1 - p)$$

where N is the distribution function of the standard Gaussian law. This method has the advantage of being relatively convenient to implement, however the first-order approximation can be inaccurate for a portfolio containing options. Indeed, the second-order derivative has a direct impact on the tail of the distribution of the portfolio and thus on the VaR. To this end, an extension of this method considering the second-order derivative was developed, called the delta-gamma method. However, this method is not straightforward to implement since it needs the computation of the mixed second-order derivatives.

These two methods rely on the assumption of normally distributed returns which is contradicted by empirical studies and thus lead to an underestimation of the quantile of the distribution and, by consequence, to an underestimation of the potential

losses. Using these approximations under a non-Gaussian context leads to more complex computations since there is no explicit formula in general. Hence, these methods are not used in practice for VaR computations or in the more general context of computing tails of portfolio distributions, and practitioners prefer the more flexible Monte Carlo method.

The Monte-Carlo method. This method relies on the generation of scenarios for the risk factors to compute the percentile of the distribution of the price variations. The portfolio writes $V = f(x^{(1)}, \dots, x^{(n)})$ with respect to risk factors and the function f is known (e.g. the payoff). The flexibility of this method relies on the fact that one can choose any stochastic model for the risk factors (it only has to be simulable) and hence it allows to take into account any arbitrary fat-tailed distribution. In this context, this method leads to a more accurate measure of the distribution tail and consequently of the risk. In addition to risk indicators, this method is also widely used for the purpose of pricing financial instruments. In this thesis, we developed several improved methods of Monte Carlo simulation in both situations.

Other risk measures The Value-at-Risk has been complemented with several other risk measures in order to capture additional aspects of the risk held in portfolios. Indeed, the VaR provides insights on the level of potential loss with a fixed probability. However, it provides neither the amount that can be lost with lower probabilities nor how the portfolio behaves in case of adverse events for example.

In order to overcome the potential underestimation of the risks assessed through the usual VaR, the regulator requires the computation of a Stressed-VaR which includes a period of stress observed on the markets. This allows to leverage on brutal and realistic moves in order to quantify potential losses under adverse scenarios.

In addition, the Expect Shortfall allows to compute the expected amount that could be lost with probabilities lower than the confidence level of the VaR. It is defined as:

$$ES_\alpha = \frac{1}{\alpha} \int_0^\alpha \text{VaR}_\gamma d\gamma.$$

It is to be noted that the computation of the Expected Shortfall is not straightforward since it requires to compute the expectation of the portfolio values for scenarios with very small probabilities of occurrence.

Another tool widely used in financial risk management is stress testing. Stress

testing consists in simulating several adverse scenarios, corresponding to huge shocks on the market, and analyzing their impact on the portfolios. As stated in the BCBS report on stress testing [?], *Stress testing alerts bank management to adverse unexpected outcomes related to a variety of risks and provides an indication of how much capital might be needed to absorb losses should large shocks occur*. Stress tests play an important role in overcoming the limitations of models and historical data, give a way to challenge assumptions of the risk models and provide insights about the validity of these models under adverse scenarios. As stress testing allows for the simulation of shocks which have not previously occurred, it should be used to assess the robustness of models to possible changes in the economic and financial environment. Moreover, risk models are generally calibrated on historical data. In stressed conditions, usual features may become inaccurate and strong dependencies between them may appear and lead to non-anticipated system-wide interactions. For example, prior-crisis stress test methodologies did not anticipate new sources of risk, such as the Wrong Way Risk and neglected the strong interaction between funding and liquidity risk. As a consequence, the selection of stress scenarios is also key. The Basel Committee report underlines the weaknesses of previous choices for crisis scenarios: *Most bank stress tests were not designed to capture the extreme market events that were experienced* [?]. Indeed, those scenarios have to be extreme enough to provide a real test of severe losses and realistic enough to highlight a plausible economic situation.

1.3 Rare events simulation and tail analysis

As explained in the previous sections, the computation of risk measures requires the evaluation of the low percentiles of the portfolio distribution and hence the estimation of probabilities of rare events. In this thesis, we propose to rely on (i) the Monte Carlo simulation with variance reduction allowing to estimate the probabilities of rare events efficiently and on (ii) asymptotic methods based either on the regular variation theory or the large deviations approach.

Rare events and Monte Carlo simulation The pricing of certain sophisticated financial instruments may require the simulation of rare events. For example, in order to price barrier options, one needs to compute the potentially very small probability that the stock price exceeds a certain level. In this context, the crude Monte Carlo method can be inaccurate as presented thereafter.

Description of the problem. Let X be a random variable and assume that we need to estimate the probability $p = \mathbb{P}[X < x]$, and that p is small. Using a sample

of size n , (X_1, \dots, X_n) , we can estimate p by the empirical mean:

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \leq x}.$$

By the central limit theorem, as $n \rightarrow \infty$, the distribution of the random variable

$$\frac{\sqrt{n}(\hat{p} - p)}{\sqrt{\text{Var} \mathbf{1}_{X \leq x}}} = \frac{\sqrt{n}(\hat{p} - p)}{\sqrt{p - p^2}}$$

tends to the standard normal distribution $\mathcal{N}(0, 1)$. Therefore, for n sufficiently large, one can write

$$\mathbb{P} \left[p \in \left(\hat{p} - \frac{z\sqrt{p-p^2}}{\sqrt{n}}; \hat{p} + \frac{z\sqrt{p-p^2}}{\sqrt{n}} \right) \right] \approx 2N(z) - 1, \quad z > 0.$$

where N is the cumulative distribution function of the standard normal distribution. In other words, the quantity $\frac{z\sqrt{p-p^2}}{\sqrt{n}}$ measures the width of the confidence interval. However, a more adequate measure of the accuracy of the simulation process is the relative error

$$RE = \frac{z\sqrt{p-p^2}}{p\sqrt{n}}$$

Then, in the case of a rare event, p is small and therefore the relative error satisfies

$$RE \approx \frac{z}{\sqrt{pn}},$$

and will be very large unless n is large compared to $\frac{1}{p}$. However, increasing the size of the sample leads to an increase of the computational time and the method becomes inappropriate for operational applications. Hence, in order to obtain a more efficient simulation algorithm, one needs to find a way to reduce the variance of the estimator without increasing the sample size.

Importance sampling to reduce the variance of Monte Carlo estimators.

Importance sampling is a widely used method for rare events simulation. Here, we do not detail all technical considerations in order to focus on the main ideas underlying the method and its efficiency for rare events simulation. Since the problem of the simulation comes from the small probability of the event considered, the basic idea of the method consists in giving this event a larger probability to occur. We assume that our goal is to estimate the expectation of a random variable X , $\theta = \mathbb{E}_{\mathbb{P}}[X]$. Then, for any other probability measure \mathbb{Q} , equivalent to \mathbb{P} , we can write

$$\mathbb{E}_{\mathbb{P}}[X] = \mathbb{E}_{\mathbb{Q}} \left[X \frac{d\mathbb{P}}{d\mathbb{Q}} \right].$$

and thus $\theta = \mathbb{E}_{\mathbb{Q}} \left[X \frac{d\mathbb{P}}{d\mathbb{Q}} \right]$. The importance sampling estimator is therefore defined by

$$\hat{\theta}_{\mathbb{Q}} = \frac{1}{n} \sum_{i=1}^n \hat{X}_i \frac{d\mathbb{P}}{d\mathbb{Q}} \Big|_i$$

where $\hat{X}_1, \dots, \hat{X}_n$ are independent and identically distributed copies of X under the probability \mathbb{Q} and $\frac{d\mathbb{P}}{d\mathbb{Q}} \Big|_i$ are the corresponding realizations of the density of the measure change. The variance of this estimator writes

$$\sigma_{\mathbb{Q}}^2 = \mathbb{E}_{\mathbb{Q}} \left[X^2 \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right)^2 \right] - \theta^2$$

Then, the purpose of importance sampling is to find the change of measure $\mathbb{P} \rightsquigarrow \mathbb{Q}$ that minimizes the variance of the estimator $\sigma_{\mathbb{Q}}^2$. This is equivalent to minimizing

$$\mathbb{E}_{\mathbb{Q}} \left[X^2 \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right)^2 \right] = \mathbb{E}_{\mathbb{P}} \left[X^2 \frac{d\mathbb{P}}{d\mathbb{Q}} \right].$$

Now, assuming that $X > 0$ and choosing $\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{X}{\mathbb{E}_{\mathbb{P}}[X]}$, we see that

$$\mathbb{E}_{\mathbb{P}} \left[X^2 \frac{d\mathbb{P}}{d\mathbb{Q}} \right] = \mathbb{E}_{\mathbb{P}}[X]^2$$

and therefore $\sigma_{\mathbb{Q}}^2 = 0$. While this choice of \mathbb{Q} is optimal, it cannot be achieved since it requires us to know $\mathbb{E}_{\mathbb{P}}[X]$, which is the very value we want to compute! However, in some specific cases this optimal measure change can be approximated numerically. In addition, it is to be noted that the importance sampling method requires the ability to simulate the random variable and the density of the measure change under the new probability.

Importance sampling has been widely used for options pricing purpose, see for example [?, ?, ?, ?, ?, ?, ?]. In Chapter 2, we will present an application of the importance sampling method to compute prices of exotic options in models based on Lévy processes.

Asymptotic methods.

An alternative to Monte Carlo methods for rare events computation consists in finding approximations exploiting the asymptotic behavior of distributions. For example, such approximations can be used to quickly compute some risk measures even for large portfolios and heavy tailed distributions without requiring simulations.

In this thesis, we mainly rely on 2 methods: regular variation and large deviations. An overview of the basic definitions and concepts as well as some financial applications of these methods are presented hereafter.

Regular variation The regular variation theory provides some interesting tools in the context of measuring tail behaviour. Hence, some basics definitions and properties are presented in this part. For a detailed review of regular variations theory, the interested reader can refer to [?].

Definition 2.1. *A measurable function $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is regularly varying at ∞ with index $\rho \in \mathbb{R}$ (written $f \in RV_\rho$) if for $x > 0$,*

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = x^\rho.$$

If $\rho = 0$, f is said to be slowly varying. A direct consequence of this definition is that for $f \in RV_\rho$, then $\frac{f(x)}{x^\rho} \in RV_0$ and thus we have the following proposition.

Proposition 2.2. *If $f \in RV_\rho$, then there exists a slowly varying function l such that $f(x) = x^\rho l(x)$.*

Definition 2.3. *A non-negative random variable X with distribution function F has a regularly varying right tail at ∞ with heavy-tail index α if its survival function is of the following form,*

$$\bar{F}(r) := 1 - F(r) = r^{-\alpha} L(r), r > 0, \alpha > 0,$$

where L is a slowly varying function.

See e.g., [?] for an application of regular variation to asymptotic approximation option prices and implied volatilities.

The multivariate extension of regular variation provides a powerful tool for the study of joint extremal dependence and the computation of risk measures in the multivariate setting [?, ?]. For a random vector $\mathbf{X} = (X_1, \dots, X_n)$ with values in $[0, \infty)^n$, the distribution of \mathbf{X} is multivariate regularly varying at $+\infty$ with limit measure ν if there exists a function $b(t) \uparrow +\infty$ as $t \rightarrow +\infty$ and a non-negative Radon measure $\nu \neq 0$ such that

$$t\mathbb{P}[b(t)^{-1}\mathbf{X} \in \cdot] \xrightarrow[t \rightarrow +\infty]{v} \nu, \tag{2.3}$$

on $\mathbb{E} = [0, \infty)^n \setminus \{0\}$, where \xrightarrow{v} stands for the vague convergence of measures. In this case, the function b is necessarily regularly varying. Assuming that $\nu(\{x : x_1 >$

$1, \dots, x_n > 1\}) > 0$ and that the measure ν does not charge the boundary of this set, we then get:

$$\mathbb{P}[X_1 > x, \dots, X_n > x] \sim \frac{\nu(\{x : x_1 > 1, \dots, x_n > 1\})}{\nu(\{x : x_1 > 1\})} \mathbb{P}[X_1 > x], \quad x \rightarrow \infty.$$

Similar results are given in [?]. Therefore, in this case the multivariate regular variation allows to compute the sharp asymptotics of the joint tail probability of a random vector (which may, e.g., represent the portfolio returns).

The multivariate regular variation assumption (2.3) implies that the distribution functions of the components of \mathbf{X} are equivalent to each other in the tail. If this is not the case, one may impose this assumption after a marginal transformation, in other words, on the copula of \mathbf{X} .

When $\nu(\{x : x_1 > 1, \dots, x_n > 1\}) = 0$, the regular variation does not allow to compute sharp asymptotics of the joint tail probability. In this case, the precise degree of dependence may be quantified using the concept of hidden regular variation (see [?] for a comprehensive review). This concept assumes that in addition to (2.3), there exists a non-decreasing function $b^*(t) \uparrow +\infty$ such that $\frac{b(t)}{b^*(t)} \rightarrow +\infty$ as $t \rightarrow +\infty$, and a Radon measure ν^* on \mathbb{E}^0 , such that

$$t\mathbb{P}[b^*(t)^{-1}\mathbf{X} \in \cdot] \xrightarrow[t \rightarrow +\infty]{v} \nu^*,$$

on \mathbb{E}^0 , where $\mathbb{E}^0 := \mathbb{E} \setminus \bigcup_{i=1}^n L_i$ with

$$L_i = (0, \dots, 0, (0, \infty], 0, \dots, 0),$$

with $(0, \infty]$ at the i -th position. Intuitively, hidden regular variation implies that the measure ν is concentrated on the coordinate axes, and probabilities of the form

$$\mathbb{P}[X_i > tx_i, X_j > tx_j]$$

for $i \neq j$ decay faster as $t \rightarrow \infty$ than the distribution functions of the components of \mathbf{X} .

The assumption of hidden regular variation imposed at the level of the distribution of \mathbf{X} once again allows to compute sharp asymptotics for the left tail of the sum of the components. Indeed, assuming that $\nu^*(\{x : x_1 > 1, \dots, x_n > 1\}) > 0$ and that this measure does not charge the boundary of this set, we have

$$\mathbb{P}[X_1 > x, \dots, X_n > x] \sim \frac{\nu^*(\{x : x_1 > 1, \dots, x_n > 1\})}{b^{*(-1)}(x^{-1})},$$

where $b^{*(-1)}$ is an asymptotic inverse of b^* . In Chapter 5, we shall present an application of the hidden regular variation to the computation of risk measures for option portfolios in the multidimensional Black-Scholes model.

Large deviations. In a very narrow sense, the large deviations theory provides convergence rates for the law of large numbers. We first formulate the large deviations principle (LDP) on abstract spaces. Let \mathcal{X} be a Hausdorff topological space endowed with its Borel σ -field. A rate function is a $[0, \infty]$ -valued lower semi-continuous function on \mathcal{X} . It is said to be a good rate function if its level sets are compact.

Definition 2.4 (Large Deviation Principle). *A family $\{X^\varepsilon\}$ of \mathcal{X} -valued random variables is said to obey a LDP in \mathcal{X} with rate function I if for each open subset $G \subset \mathcal{X}$ and each closed subset $F \subset \mathcal{X}$*

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{P}[X^\varepsilon \in F] \leq - \inf_{x \in F} I(x)$$

and

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{P}[X^\varepsilon \in G] \geq - \inf_{x \in G} I(x).$$

There are many different contexts in which a large deviations principle holds, the simplest and most famous one being Cramér's theorem, dealing with sequences of i.i.d. random variables. For simplicity we give here the one-dimensional version.

Theorem 2.5. *Let X_1, \dots, X_n, \dots be i.i.d. random variables distributed according to the probability law μ . Define the logarithmic moment generating function associated with this law by*

$$\Lambda(\lambda) = \log \mathbb{E}[e^{\lambda X_1}],$$

and the Fenchel-Legendre transform of Λ by

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \Lambda(\lambda)\}.$$

Consider the empirical means

$$S_n = \frac{1}{n} \sum_{j=1}^n X_j.$$

Then the sequence $\{S_n\}$ satisfies the LDP on \mathbb{R} with the rate function Λ^* .

Another famous large deviations result is the Gärtner-Ellis theorem, where the independence assumption is replaced with a much weaker one on the limiting behavior of the Laplace transforms of S_n . Both Cramer's theorem and Gärtner-Ellis theorem deal with finite-dimensional random variables. In chapter 3 we shall make use of a large deviations result for Lévy processes which holds on the space of càdlàg trajectories.

The LDP allows to quantify the asymptotic behavior of the probability distribution of a family of random variables at the log-scale. Another important result which will be used in this thesis, the Varadhan's lemma, may be used to quantify the asymptotic behavior of more general expectation-type functionals.

Lemma 2.6 (Varadhan’s lemma). *Suppose that $\{X_\varepsilon\}$ satisfies the LDP with a good rate function $I : \mathcal{X} \rightarrow [0, \infty[$ and let $\phi : \mathcal{X} \rightarrow \mathbb{R}$ be a continuous function. Assume further that for some $\gamma > 1$,*

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{E} \left[e^{\frac{\gamma \phi(X_\varepsilon)}{\varepsilon}} \right] < \infty.$$

Then,

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{E} \left[e^{\frac{\phi(X_\varepsilon)}{\varepsilon}} \right] = \sup_{x \in \mathcal{X}} \{ \phi(x) - I(x) \}.$$

The large deviation theory has been used for rare events computation for financial applications, in particular for pricing purposes in [?, ?], measurement of credit risk in [?, ?] and for various asymptotics for stochastic volatility models in [?, ?]. In addition, [?, ?, ?] have applied both importance sampling technics and large deviation theory in order to obtain efficient Monte Carlo simulations.

2 Main contributions of this thesis

This thesis contains three contributions to pricing and risk management problems in mathematical finance using asymptotic methods. Chapter 3 develops a variance reduction technique for Monte Carlo pricing of exotic options when the stock price is modelled as the exponential of a Lévy process. Chapter 4 develops a framework for asymptotic analysis of portfolios whose components are modelled by log-normal mixtures. This includes for instance multidimensional extensions of Heston and variance gamma model. These asymptotic results are applied to construct variance reduction methods for Monte Carlo simulation of rare events associated to such portfolios. Finally, Chapter 5 presents results on asymptotic behavior of portfolios whose dependence structure is modelled by a Gaussian copula. This covers for example a portfolio of options in the multidimensional Black-Scholes model. The following sections give an overview of the main results related to each contribution.

2.1 Optimal importance sampling for Lévy processes

In this part we develop effective and easy to implement importance sampling estimators of expectations of functionals of Lévy processes, corresponding to option prices in exponential Lévy models. To model a financial market with a Lévy process, we assume that the market consists of a risk-free asset $S_t^0 \equiv 1$ and n risky assets S^1, \dots, S^n where

$$S_t^i = S_0^i e^{X_t^i},$$

and (X^1, \dots, X^n) is a Lévy process, such that S^i is a martingale for each i , under the risk-neutral probability \mathbb{P} . We fix a time horizon $T < \infty$ and consider a derivative

written on $(S^i)_{1 \leq i \leq n}$ with a nonnegative pay-off $P(S)$ which depends of the entire trajectory of the stocks up to time T . We are interested in computing the price of this derivative, given by the risk-neutral expectation $\mathbb{E}[P(S)]$.

The standard Monte Carlo estimator of $\mathbb{E}[P(S)]$ is defined as the empirical mean

$$\widehat{P}_N := \frac{1}{N} \sum_{j=1}^N P(S^{(j)}),$$

where $S^{(j)}, j = 1, \dots, N$ are i.i.d. samples with the same law as S . The standard estimator often converges too slowly, and we therefore propose to use an importance sampling estimator based on the path-dependent Esscher transform,

$$\frac{d\mathbb{P}^\theta}{d\mathbb{P}} = \frac{e^{\int_{[0,T]} \langle X_t, \theta(dt) \rangle}}{\mathbb{E} \left[e^{\int_{[0,T]} \langle X_t, \theta(dt) \rangle} \right]}, \quad (2.4)$$

where θ is a (deterministic) bounded \mathbb{R}^n -valued signed measure on $[0, T]$. The optimal choice of θ should minimize the variance of the estimator under \mathbb{P}^θ ,

$$\text{Var}_{\mathbb{P}^\theta} \left(P \frac{d\mathbb{P}}{d\mathbb{P}^\theta} \right) = \mathbb{E}_{\mathbb{P}} \left[P^2 \frac{d\mathbb{P}}{d\mathbb{P}^\theta} \right] - \mathbb{E}[P]^2.$$

Denoting by H the log-payoff of the option expressed as function of the Lévy process,

$$H(X) = \log P(S), \quad X_t^i = \log \frac{S_t^i}{S_0^i}, \quad 0 \leq t \leq T,$$

the minimization problem writes

$$\inf_{\theta \in M} \mathbb{E}_{\mathbb{P}} \left[\exp \left\{ 2H(X) - \int_{[0,T]} \langle X_t, \theta(dt) \rangle + \int_{[0,T]} G(\theta([t, T])) dt \right\} \right],$$

where

$$G(\theta) = \frac{1}{2} \langle \theta, A\theta \rangle + \langle \theta, \gamma \rangle + \int_{\mathbb{R}^n} (e^{\langle \theta, x \rangle} - 1 - \langle \theta, x \rangle 1_{|x| \leq 1}) \nu(dx),$$

and (A, ν, γ) denotes the characteristic triplet of the process X .

The main contribution of this paper, inspired by the work of Guasoni and Robertson [?] in the setting of the Black-Scholes model, is to use the large deviations theory to construct an easily computable approximation to the optimal importance sampling measure parameter θ_{opt}^* . Namely, we use Varadhan's lemma and the pathwise large deviation principle for Lévy processes due to Leonard [?] to derive a proxy for the variance of the importance sampling estimator which is much easier to compute than the true variance. Our first main result provides an expression for such a proxy, which we aim to minimize to obtain an asymptotically optimal variance reduction. It requires the following assumption.

(A1) There exists $\lambda_0 > 0$ with $\int_{|x|>1} e^{\lambda_0|x|} \nu(dx) < \infty$.

In this proposition and below, D denotes the space of right-continuous functions with left limits on the interval $[0, T]$, M denotes the class of bounded \mathbb{R}^n -valued measures on $[0, T]$ and $\sigma(D, M)$ denotes the topology on D defined by

$$\lim_n y_n = y \Leftrightarrow \forall \mu \in M, \lim_n \int_{[0, T]} y_n d\mu = \int_{[0, T]} y d\mu.$$

Proposition 2.7. *Let Assumption (A1) hold true, and suppose that the set $\{x \in D : H(x) > -\infty\}$ is open for the $\sigma(D, M)$ topology and contains the zero element, that H is continuous on this set for the $\sigma(D, M)$ -topology, and that H and $\theta \in M$ satisfy one of the following alternative assumptions.*

(i) *There exist constants $C < \infty$ and $B < \lambda_0/4n$ such that*

$$H(x) \leq C + B \sup_{s \in [0, T]} \sum_{i=1}^n |x_s^i|$$

and

$$\max_{0 \leq t \leq T} |\theta([t, T])| < \lambda_0 - 4nB.$$

(ii) *H is bounded and*

$$\int_{[0, T]} G(\gamma\theta([t, T])) dt < \infty$$

for some $\gamma > 1$.

Then it holds that

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{E} \left[e^{\frac{2H(X^\varepsilon) - \int_{[0, T]} \langle X_t^\varepsilon, \theta(dt) \rangle}{\varepsilon}} \right] = \sup_{x \in D} \left\{ 2H(x) - \int_{[0, T]} \langle x_t, \theta(dt) \rangle - J(x) \right\}.$$

where we have defined

$$X_t^\varepsilon = \varepsilon X_{t/\varepsilon}$$

$$J(x) = \begin{cases} \sup_{\mu \in M} \left\{ \int_{[0, T]} \langle x_t, \mu(dt) \rangle - \int_{[0, T]} G(\mu([t, T])) dt \right\} & \text{if } x \in V_r, \\ + \infty & \text{otherwise.} \end{cases}$$

and V_r denotes the subspace of D containing all functions in D with bounded variation.

In view of this result we suggest to obtain the candidate variance reduction parameter θ^* by minimizing

$$\sup_{x \in D} \left\{ 2H(x) - \int_{[0, T]} \langle x_t, \theta(dt) \rangle + \int_{[0, T]} G(\theta([t, T])) dt - J(x) \right\} \quad (2.5)$$

over $\theta \in M$. It is then natural to ask, how close the corresponding measure change will be to the optimal one which minimizes the variance of the Monte Carlo estimator over all possible measure changes. Varadhan's lemma allows to define a notion of asymptotic optimality, which provides a partial answer to this question. Consider a family of importance sampling measures $(\mathbb{Q}(\varepsilon))_{\varepsilon>0}$. By Jensen's inequality,

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E}^{\mathbb{Q}(\varepsilon)} \left[\left(e^{\frac{1}{\varepsilon} H(X^\varepsilon)} \frac{d\mathbb{P}}{d\mathbb{Q}(\varepsilon)} \right)^2 \right] \geq 2 \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[e^{\frac{1}{\varepsilon} H(X^\varepsilon)} \right].$$

We shall call a family of importance sampling measure changes asymptotically optimal if, for this family, the above inequality becomes an equality and the lim-inf/limsup become limits. In other words, the asymptotically optimal measure change does at least as well as any other measure change at the logarithmic scale of large deviations.

Definition 2.8. *Let $(\mathbb{Q}(\varepsilon))_{\varepsilon>0}$ be a family of importance sampling measure changes. We say that $(\mathbb{Q}(\varepsilon))_{\varepsilon>0}$ is asymptotically optimal if*

$$\lim_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E}^{\mathbb{Q}(\varepsilon)} \left[\left(e^{\frac{1}{\varepsilon} H(X^\varepsilon)} \frac{d\mathbb{P}}{d\mathbb{Q}(\varepsilon)} \right)^2 \right] = 2 \lim_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[e^{\frac{1}{\varepsilon} H(X^\varepsilon)} \right].$$

The following theorem, which is our second main result, shows that for concave log-payoffs, the importance sampling measure associated to the parameter θ^* computed by minimizing the expression (2.5) is asymptotically optimal, and its computation is greatly simplified. In other words, for such pay-offs, it is sufficient to consider only measure changes of the form (2.4) with deterministic θ to achieve asymptotic optimality over all possible measure changes. European basket put options and many path-dependent put-like payoffs encountered in practice are indeed concave. This theorem is proven under the following assumption.

(A2) The function G is lower semicontinuous and its effective domain $\mathcal{D} := \{x \in \mathbb{R}^n : G(x) < \infty\}$ is open and bounded.

Theorem 2.9. *Let H be concave, and assume that the set $\{x \in V_r : H(x) > -\infty\}$ is nonempty and contains the zero element, and that H is continuous on this set for $\sigma(D, M)$ -topology. Let Assumptions (A1) and (A2) be satisfied. Then,*

$$\begin{aligned} \inf_{\theta \in M} \sup_{x \in V_r} \left\{ 2H(x) - \int_{[0,T]} \langle x_t, \theta(dt) \rangle + \int_{[0,T]} G(\theta([t, T])) dt - J(x) \right\} \\ = 2 \inf_{\theta \in M} \left\{ \widehat{H}(\theta) + \int_{[0,T]} G(\theta([t, T])) dt \right\} \quad (2.6) \end{aligned}$$

where

$$\widehat{H}(\theta) = \sup_{x \in V_r} \left\{ H(x) - \int_{[0,T]} \langle x_t, \theta(dt) \rangle \right\}.$$

Moreover, if the infimum in the right-hand side of (2.6) is attained by θ^* then the same value θ^* attains the infimum in the left-hand side of (2.6). If, in addition, the pay-off functional H and the candidate parameter θ^* satisfy the assumptions (i) or (ii) of Proposition 3.5 then the importance sampling measure corresponding to θ^* is asymptotically optimal.

This work led to a pre-publication with P. Tankov which is under second round of review for Stochastic Processes and Their Applications.

2.2 Tail asymptotics of log-normal mixture portfolios

In this part we consider the tail behavior of the sum of n dependent positive random variables

$$X = \sum_{i=1}^n X_i.$$

In financial mathematics, X may represent the value of a long-only portfolio of n assets, and understanding the tail behavior of X is important for risk management applications, such as computing the Value at Risk, evaluating tail event probabilities or designing efficient simulation algorithms for tail events. In particular, stress test scenarios may be constructed in a systematic manner by simulating the values of the components X_1, \dots, X_n conditionnally on the event that X takes a given small value.

This problem has received considerable attention in the literature, but mainly in the insurance context, where the random variables X_1, \dots, X_n represent losses from individual claims, and one is interested in the *right tail* asymptotics of X , so as to estimate the probability of having a very large aggregate loss. In this setting, provided the variables X_1, \dots, X_n are sufficiently fat-tailed (subexponential), under various assumptions on the dependence structure, it can be shown that the right tail behavior of X is determined by the single variable with the fattest tail. In our work, we focus on the context of financial risk management where the extreme event of interest corresponds to a *small* value of the random variable X . In this context, to estimate the probability of a large loss, one needs to focus on the *left tail* asymptotics of X . Owing to the positivity of the variables X_1, \dots, X_n , the asymptotic behavior of the left tail of X turns out to be very different from that of the right tail.

In this work, we compute sharp asymptotics of the distribution function and the density of X in the left tail, and discuss the relevant risk management applications, under the assumption that X_1, \dots, X_n follow a log-normal mixture distribution. That is, we assume that for $i = 1, \dots, n$, $X_i = e^{Y_i}$, where the vector $\mathbf{Y} = (Y_1, \dots, Y_n)$ follows a Gaussian variance-mean mixture distribution.

The Gaussian variance-mean mixture distribution on \mathbb{R}^n is defined as :

$$\mathbf{Y} = \sqrt{\Theta}\mathbf{Z} + \Theta\boldsymbol{\mu} + \boldsymbol{\Lambda}$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\boldsymbol{\Lambda} \in \mathbb{R}^n$ are constant vectors, $\mathbf{Z} \sim N(0, \mathfrak{B})$ where $\mathfrak{B} \in \mathbb{M}^{n \times n}$ is a symmetric, positive definite matrix, and $\Theta \in \mathbb{R}_+$ is the mixing random variable. In this paper, Θ is assumed to admit a density, denoted $\rho(s)$, which decays exponentially fast as $s \rightarrow \infty$.

Let us now recall some necessary notation. For two functions f and g , we denote $f(x) \sim g(x)$ as $x \rightarrow a$ whenever $\frac{f(x)}{g(x)} \rightarrow 1$ as $x \rightarrow a$. The symbol $\mathbf{1}$ stands for the n -dimensional vector with all elements equal to 1. In general, vectors will be denoted by bold-faced symbols, so that $\mathbf{w} = (w_1, \dots, w_n)^\perp$, where \perp stands for matrix transposition. For $\mathbf{v} \in \mathbb{R}_+^n$, we denote $\mathcal{E}(\mathbf{v}) = -\sum_{i=1}^n v_i \log v_i$ with the usual convention $0 \log 0 = 0$. The elements of the covariance matrix \mathfrak{B} will be denoted by $(b_{ij})_{1 \leq i, j \leq n}$ and those of the inverse matrix \mathfrak{B}^{-1} will be denoted $(a_{ij})_{1 \leq i, j \leq n}$. Δ_n will stand for the n -dimensional simplex defined by:

$$\Delta_n = \left\{ \mathbf{w} \in \mathbb{R}^n : w_i \geq 0, i = 1, \dots, n \text{ and } \sum_{i=1}^n w_i = 1 \right\}.$$

Elementary functions are applied to vectors component-wise, so that $\log(\mathbf{w})$ stands for $(\log w_1, \dots, \log w_n)^\perp$. Similarly, inequalities involving vectors like $\mathbf{w} > 0$ are understood component-wise. Recall that the beta function is defined in the following way:

$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt, \quad x, y > 0$$

For $\boldsymbol{\alpha} \in \mathbb{R}^n$, $n \geq 2$, with $\alpha_i > 0$, $i = 1, \dots, n$, the multidimensional Beta function can be defined as:

$$B(\boldsymbol{\alpha}) := \int_{\Delta_n} \prod_{i=1}^n w_i^{\alpha_i-1} d\mathbf{w} = \frac{\prod_{i=1}^n \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^n \alpha_i)}.$$

By convention, we set $B(\boldsymbol{\alpha}) = 1$ for all $\boldsymbol{\alpha} \in \mathbb{R}$.

Main results Consider the following function:

$$F(t, \mathbf{w}) = \theta t + \frac{(1 + t\boldsymbol{\mu}^\perp \mathbf{w})^2}{2\mathbf{w}^\perp \mathfrak{B} \mathbf{w} t}.$$

The following lemma, proved in [?] establishes some properties of this function.

Lemma 2.10. *There exists a unique couple $(\bar{t}, \bar{\mathbf{w}})$, with $\bar{t} \in (0, \infty)$ and $\bar{\mathbf{w}} \in \Delta_n$ such that*

$$F(\bar{t}, \bar{\mathbf{w}}) = \min_{t>0} \max_{\mathbf{w} \in \Delta_n} F(t, \mathbf{w}).$$

In addition, the function

$$f(t) = F(t, \bar{\mathbf{w}})$$

has a unique minimum at the point \bar{t} .

We make the following assumptions:

(A1) The minimizer \bar{t} satisfies

$$\mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu}) > 0.$$

(A2) $\rho(t)$ can be written as :

$$\rho(t) = e^{-\theta t + m\sqrt{t}} \rho_0(t)$$

where $\theta > 0$, $m \in \mathbb{R}$ and ρ_0 is a regularly varying function of order α as $t \rightarrow \infty$.

Assumption (A1) is a relatively strong assumption which is discussed in detail in the main body of the chapter. The following theorem is our main result. It provides a sharp asymptotic formula for the density and the cumulative distribution function of the random variable X .

Theorem 2.11. *Under assumptions (A1)-(A2), the density of X , $p(z)$, satisfies*

$$\begin{aligned} p(z) &\sim \frac{(2\pi)^{\frac{1-n}{2}} \bar{t}^{\alpha + \frac{3-n}{2}}}{\sqrt{\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1} \det \mathfrak{B}}} B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu})\right) \exp\left(\frac{m^2 \bar{t}^2}{8 \mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1}} - \frac{1}{\bar{t}} (\mathbf{1} + \bar{t}\boldsymbol{\mu})^\perp \mathfrak{B}^{-1} \boldsymbol{\Lambda}\right) \\ &\times \frac{\rho_0\left(\log \frac{1}{z}\right)}{z \left(\log \frac{1}{z}\right)^{\frac{n-1}{2}}} e^{-\log \frac{1}{z} F(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t} \log \frac{1}{z}}}. \end{aligned}$$

when $z \rightarrow 0$, and the cumulative distribution satisfies

$$\begin{aligned} F(x) &\sim \frac{(2\pi)^{\frac{1-n}{2}} \bar{t}^{\alpha + \frac{3-n}{2}}}{F(\bar{t}, \bar{\mathbf{w}}) \sqrt{\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1} \det \mathfrak{B}}} B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu})\right) \\ &\times \exp\left(\frac{m^2 \bar{t}^2}{8 \mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1}} - \frac{1}{\bar{t}} (\mathbf{1} + \bar{t}\boldsymbol{\mu})^\perp \mathfrak{B}^{-1} \boldsymbol{\Lambda}\right) \\ &\times \rho_0\left(\log \frac{1}{x}\right) \left(\log \frac{1}{x}\right)^{\frac{1-n}{2}} e^{-\log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t} \log \frac{1}{x}}} \end{aligned}$$

when $x \rightarrow 0$

Theorem 4.3 allows us to estimate various conditional expectations. The next assertion provides a characterization of the limiting conditional law of the Laplace transform of Y_1, \dots, Y_n , given that $X \leq x$.

Corollary 2.12. *Suppose Assumptions (A1)-(A2) hold and $\mathbf{u} \in \mathbb{R}^n$ fixed. We denote :*

$$\begin{aligned} C_{\mathbf{u}} &= \boldsymbol{\mu}^\perp \mathbf{u} + \frac{1}{2} \mathbf{u}^\perp \mathfrak{B} \mathbf{u} \\ \boldsymbol{\mu}_{\mathbf{u}} &= \boldsymbol{\mu} + \mathfrak{B} \mathbf{u} \\ \theta_{\mathbf{u}} &= \theta - C_{\mathbf{u}} \\ F_{\mathbf{u}}(t, w) &= \theta_{\mathbf{u}} t + \frac{(1 + t \boldsymbol{\mu}_{\mathbf{u}}^\perp \mathbf{w})^2}{2 \mathbf{w}^\perp \mathfrak{B} \mathbf{w} t} \end{aligned}$$

If $\theta_{\mathbf{u}} > 0$ and $\mathfrak{B}^{-1}(\mathbf{1} + \bar{t} \boldsymbol{\mu}_{\mathbf{u}}) > 0$, then as $x \rightarrow 0$,

$$\begin{aligned} \mathbb{E} \left[e^{\sum_{i=1}^n u_i Y_i} | X \leq x \right] &\sim x^{\mathbf{1}^\perp \mathbf{u}} \frac{F(\bar{t}, \bar{\mathbf{w}})}{F(\bar{t}, \bar{\mathbf{w}}) + \mathbf{1}^\perp \mathbf{u}} \frac{B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t} \boldsymbol{\mu}) + \mathbf{u}\right)}{B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t} \boldsymbol{\mu})\right)} \\ \mathbb{E} \left[e^{\sum_{i=1}^n u_i Y_i} | X = x \right] &\sim x^{\mathbf{1}^\perp \mathbf{u}} \frac{B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t} \boldsymbol{\mu}) + \mathbf{u}\right)}{B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t} \boldsymbol{\mu})\right)} \end{aligned}$$

The following corollary provides an asymptotic formula for the Value at Risk.

Corollary 2.13. *Under the assumptions of Theorem 4.3, the Value at Risk at the level $1 - y$ satisfies*

$$\text{VaR}_{1-y}(X) \sim \left(y \frac{\left(F^{-1} \log \frac{1}{y}\right)^{\frac{n-1}{2}}}{C \rho_0 \left(F^{-1} \log \frac{1}{y}\right)} \exp \left(-m \sqrt{\bar{t} F^{-1} \log \frac{1}{y} - \frac{m^2 \bar{t}}{2F}} \right) \right)^{\frac{1}{F}},$$

where we denote $F = F(\bar{t}, \bar{\mathbf{w}})$ and

$$C = \frac{(2\pi)^{\frac{1-n}{2}} \bar{t}^{\alpha + \frac{3-n}{2}}}{F \sqrt{\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1} \det \mathfrak{B}}} B \left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t} \boldsymbol{\mu}) \right) \times \exp \left(\frac{m^2 \bar{t}^2}{8 \mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1}} - \frac{1}{\bar{t}} (\mathbf{1} + \bar{t} \boldsymbol{\mu})^\perp \mathfrak{B}^{-1} \boldsymbol{\Lambda} \right)$$

2.3 Asymptotics of heavy-tailed risks with Gaussian copula dependence

In this part, we show that a multivariate distribution with heavy-tailed (regularly varying) univariate margins and Gaussian copula dependence exhibits multivariate regular variation on the cone $(0, \infty)^n$ even if the margins are not identically distributed, and the distribution is not multivariate regularly varying on $[0, \infty)^n$. This enables us to compute sharp tail asymptotics of certain functionals, such as the sum of components, of heavy-tailed random vectors with Gaussian copula dependence, and develop a number of applications such as the computation of risk measures for a portfolio of options in the Black-Scholes model, and of the probability distribution of the aggregate production of a wind farm.

Let X_1, \dots, X_n be random variables with values in (a, ∞) with $a \in [-\infty, \infty)$, whose distribution functions are denoted by F_1, \dots, F_n . We assume that the dependence of the random vector $\mathbf{X} = (X_1, \dots, X_n)$ is specified by a Gaussian copula, that is, $X_k = F_k^{-1} \left(\mathcal{N} \left(\tilde{X}_k \right) \right)$ for $k = 1, \dots, n$, where $\tilde{\mathbf{X}} = (\tilde{X}_1, \dots, \tilde{X}_n)$ is a Gaussian vector with standard normal margins and correlation matrix Σ , \mathcal{N} is the standard normal distribution function and F_k^{-1} is the generalized inverse of F_k .

Our main result quantifies the tail behavior of X_1, \dots, X_n under the assumptions that the components of this vector are regularly varying. To state this result, we first recall the following notation.

Let $\tilde{\mathbf{X}}$ be a Gaussian vector with standard normal margins and a positive definite correlation matrix Σ and denote by B the inverse matrix of Σ . For a fixed $\mathbf{x} \in \mathbb{R}_+^n$ with $\mathbf{x} \neq 0$, denote \mathbf{x}^* the solution (minimizer) of the problem

$$\min_{\mathbf{z} \geq \mathbf{x}} \langle \mathbf{z}, B\mathbf{z} \rangle. \quad (2.7)$$

Then (see Proposition 2 in [?]), there exist unique disjoint index sets I and J , with $I \cup J = \{1, \dots, n\}$, such that

$$\mathbf{x}_I^* = \mathbf{x}_I, \quad \mathbf{x}_J^* = -B_{JJ}^{-1} B_{JI} \mathbf{x}_I, \quad (2.8)$$

$$\text{and } \langle \mathbf{e}_i, \Sigma_{II}^{-1} \mathbf{x}_I \rangle > 0, \quad \forall i \in I, \quad (2.9)$$

where in the above matrix notation indexing is performed before taking the inverse.

Theorem 2.14. *Let X_1, \dots, X_n be random variables with values in (a, ∞) , whose distribution functions are denoted by F_1, \dots, F_n . Assume that the dependence of the random vector $\mathbf{X} = (X_1, \dots, X_n)$ is specified by a Gaussian copula with correlation matrix Σ , and that for each i , the survival function $\tilde{F}_i = 1 - F_i$ is regularly varying at $+\infty$ with index $-\alpha_i < 0$. Denote $\boldsymbol{\beta} = (\sqrt{\alpha_1}, \dots, \sqrt{\alpha_n})$ and let $\boldsymbol{\beta}^*$ be the solution of the problem*

$$\min_{\mathbf{z} \geq \boldsymbol{\beta}} \langle \mathbf{z}, B\mathbf{z} \rangle,$$

and I and J be the corresponding index sets. Assume that $\boldsymbol{\beta}_J^* > \boldsymbol{\beta}_J$. Then for $x_1 > 0, \dots, x_n > 0$ as $u \uparrow \infty$,

$$\begin{aligned} & \mathbb{P}[X_1 > ux_1, \dots, X_n > ux_n] \\ &= \exp \left\{ - \sum_{i,j \in I} b_{ij} \sqrt{\ln \tilde{F}_i(u) \ln \tilde{F}_j(u)} \right\} L(u) \nu(\mathbf{x}) (1 + o(1)) \end{aligned}$$

where

$$\begin{aligned}
L(u) &= C \ln u^{\frac{1}{2} \sum_{i,j} b_{ij} \frac{\beta_i}{\beta_j} - \frac{|I|}{2}} \\
C &= \prod_{i \in I} \beta_i^{\frac{1}{\beta_i} \sum_{j \in I} b_{ij} \beta_j} \frac{(4\pi)^{\frac{1}{2} \sum_{i,j \in I} b_{ij} \frac{\beta_i}{\beta_j} - \frac{|I|}{2}}}{|\Sigma_{II}|^{\frac{1}{2}} \prod_{i \in I} \langle \mathbf{e}_i, \Sigma_{II}^{-1} \boldsymbol{\beta}_I \rangle} \\
\nu(\mathbf{x}) &= \prod_{i \in I} x_i^{-\beta_i \sum_{j \in I} b_{ij} \beta_j}
\end{aligned}$$

Remark 2.15. The above theorem states that the vector \mathbf{X} has the property of multivariate regular variation, but on the cone $(0, \infty)^n$ rather than $[0, \infty)^n \setminus \{0\}$. Since, in our context, the components are not identically distributed, multivariate regular variation on $[0, \infty)^n$ may not hold.

The following corollary identifies the left tail asymptotics of the sum of components, of heavy-tailed random vectors with Gaussian copula dependence

Corollary 2.16. *Let X_1, \dots, X_n be random variables with values in $(0, \infty)$, whose distribution functions are denoted by F_1, \dots, F_n . Assume that the dependence of the random vector $\mathbf{X} = (X_1, \dots, X_n)$ is specified by a Gaussian copula with correlation matrix Σ , and that for each i , the distribution function F_i is regularly varying at 0 with index $\alpha_i > 0$. Denote $\boldsymbol{\beta} = (\sqrt{\alpha_1}, \dots, \sqrt{\alpha_n})$ and let $\boldsymbol{\beta}^*$ be the solution of the problem*

$$\min_{\mathbf{z} \geq \boldsymbol{\beta}} \langle \mathbf{z}, B\mathbf{z} \rangle,$$

and I and J be the corresponding index sets. Assume that $\boldsymbol{\beta}_J^* > \boldsymbol{\beta}_J$. Then, as $z \downarrow 0$,

$$\mathbb{P}[X_1 + \dots + X_n < z] = \exp \left\{ - \sum_{i,j \in I} b_{ij} \sqrt{\ln F_i(z) \ln F_j(z)} \right\} L \left(\frac{1}{z} \right) \nu_0(A)$$

where L is the function defined in Theorem 5.2 and

$$\begin{aligned}
\nu_0(A) &= \prod_{k=1}^{|I|-1} p_{i_k} B \left(p_{i_k}, 1 + \sum_{j=k+1}^{|I|} p_{i_j} \right). \\
p_i &= \beta_i \sum_{j \in I} b_{ij} \beta_j.
\end{aligned}$$

Here B is the Euler beta function, the elements of the set I are denoted by i_k for $k = 1, \dots, |I|$ (in increasing order) and the empty product by convention equals 1.

Chapter 3

Optimal importance sampling for Lévy processes

Contents

1	Introduction	62
2	Pathwise large deviations for Lévy processes	65
3	Main results	70
4	Examples	79
5	Numerical illustrations	84

1 Introduction

The aim of this paper is to develop effective and easy to implement importance sampling estimators of expectations of functionals of Lévy processes, corresponding to option prices in exponential Lévy models.

Lévy processes are stochastic processes with stationary independent increments. They are used as models for asset prices when jump risk is relevant, either directly (as in the variance gamma model [?], normal inverse Gaussian process [?], CGMY model [?]) or as building blocks for other models (affine processes, stochastic volatility models with jumps, local Lévy models [?] etc.). To model a financial market with a Lévy process, we assume that the market consists of a risk-free asset $S_t^0 \equiv 1$ and n risky assets S^1, \dots, S^n where

$$S_t^i = S_0^i e^{X_t^i},$$

and (X^1, \dots, X^n) is a Lévy process, such that S^i is a martingale for each i , under the risk-neutral probability \mathbb{P} . The general results of sections 2 and 3 do not require the martingale property.

We fix a time horizon $T < \infty$ and consider a derivative written on $(S^i)_{1 \leq i \leq n}$ with a nonnegative pay-off $P(S)$ which depends on the entire trajectory of the stocks up to time T . We are interested in computing the price of this derivative, given by the risk-neutral expectation $\mathbb{E}[P(S)]$.

Several methods for computing this expectation are available in the literature. When the price process S is one-dimensional and the pay-off P only depends on the terminal value S_T , the Fourier method of Carr and Madan [?] may be used. When the dimension of S is low (say, up to 3-4), and the pay-off is only weakly path-dependent, such as for barrier or American options, one can use deterministic numerical methods for partial integro-differential equations [?], Fourier time stepping [?, ?] and related deterministic methods. Finally, for high dimensional problems, or in the case of strong path dependence, the Monte Carlo method, on which we focus in this paper is the only one available.

The standard Monte Carlo estimator of $\mathbb{E}[P(S)]$ is defined as the empirical mean

$$\widehat{P}_N := \frac{1}{N} \sum_{j=1}^N P(S^{(j)}),$$

where $S^{(j)}, j = 1, \dots, N$ are i.i.d. samples with the same law as S . Note that simulation methods exist for all parametric Lévy models, including multidimensional Lévy processes (see chapter 6 of [?] for a general overview).

The standard Monte Carlo method often converges too slowly for real-time applications, particularly when $\mathbb{E}[P(S)]$ is small compared to $\sqrt{\text{Var} P(S)}$, and various error reduction techniques must be applied (see e.g., [?] for a general overview and

[?] for a review of simulation methods and variance reduction techniques available in the case of the variance gamma model).

The Multilevel Monte Carlo method (see [?] for a general introduction and [?, ?] for an application to Lévy models) reduces the variance by optimizing the number of discretization steps for each path. In practice, a large number of paths are simulated with a coarse discretization, and only a small number of paths are discretized finely.

The Quasi Monte Carlo method (see [?, ?] for the case of jump diffusions / Lévy processes) replaces the i.i.d. samples of S with well-chosen deterministic samples (low-discrepancy sequences).

Finally, the importance sampling method, which is the focus of this paper, consists of simulating the paths of S under a different probability measure, which allows a better exploration of the region of interest. See e.g., [?] for an application in the context of Gaussian vectors, [?] for the case of path-dependent options in the Black-Scholes model, [?] for an application to stochastic volatility models, and [?, ?, ?] for applications to Lévy processes / jump-diffusions.

Importance sampling is based on the following identity, valid for any probability measure \mathbb{Q} , with respect to which \mathbb{P} is absolutely continuous.

$$\mathbb{E}[P(S)] = \mathbb{E}^{\mathbb{Q}} \left[\frac{d\mathbb{P}}{d\mathbb{Q}} P(S) \right].$$

This allows one to define the importance sampling estimator

$$\widehat{P}_N^{\mathbb{Q}} := \frac{1}{N} \sum_{j=1}^N \left[\frac{d\mathbb{P}}{d\mathbb{Q}} \right]^{(j)} P(S_{\mathbb{Q}}^{(j)}),$$

where $\left[\frac{d\mathbb{P}}{d\mathbb{Q}} \right]^{(j)} P(S_{\mathbb{Q}}^{(j)})$ are i.i.d. samples of $\frac{d\mathbb{P}}{d\mathbb{Q}} P(S)$ under the measure \mathbb{Q} . For efficient variance reduction, one needs then to find a probability measure \mathbb{Q} such that S is easy to simulate under \mathbb{Q} and the variance

$$\text{Var}_{\mathbb{Q}} \left[P(S) \frac{d\mathbb{P}}{d\mathbb{Q}} \right]$$

is considerably smaller than the original variance $\text{Var}_{\mathbb{P}} [P(S)]$.

For Lévy processes, a natural choice of probability measure for importance sampling is given by the Esscher transform

$$\left. \frac{d\mathbb{P}^{\theta}}{d\mathbb{P}} \right|_{\mathcal{F}_T} = \frac{e^{\langle \theta, X_T \rangle}}{\mathbb{E} [e^{\langle \theta, X_T \rangle}]}, \quad (3.1)$$

which is well defined for all $\theta \in \mathbb{R}^n$ such that $\mathbb{E} [e^{\langle \theta, X_T \rangle}] < \infty$. This choice was studied, e.g., in [?, ?], where the optimal variance reduction parameter θ^* was estimated numerically.

In this paper, to allow for more freedom in choosing the importance sampling probability for path-dependent payoffs, we propose to use the path-dependent Esscher transform,

$$\frac{d\mathbb{P}^\theta}{d\mathbb{P}} = \frac{e^{\int_{[0,T]} \langle X_t, \theta(dt) \rangle}}{\mathbb{E} \left[e^{\int_{[0,T]} \langle X_t, \theta(dt) \rangle} \right]},$$

where θ is a (deterministic) bounded \mathbb{R}^n -valued signed measure on $[0, T]$. Under \mathbb{P}^θ , the process X has independent increments (see Theorems III.7.23 and II.4.15 in [?]), which simplifies the development of simulation algorithms. The optimal choice of θ should minimize the variance of the estimator under \mathbb{P}^θ ,

$$\text{Var}_{\mathbb{P}^\theta} \left(P \frac{d\mathbb{P}}{d\mathbb{P}^\theta} \right) = \mathbb{E}_{\mathbb{P}} \left[P^2 \frac{d\mathbb{P}}{d\mathbb{P}^\theta} \right] - \mathbb{E} [P]^2.$$

Importance sampling is most effective in the context of *rare event simulation*, e.g., when the probability that the option will be exercised $\mathbb{P}[P(S) > 0]$ is small. Since the theory of large deviations is concerned with the study of probabilities of rare events, it is natural to use measure changes appearing in or inspired by the large deviations theory for importance sampling. We refer, e.g., to [?] and references therein for a review of this approach and to the above quoted references [?, ?, ?] for specific applications to financial models.

The main contribution of this paper, inspired by the work of Guasoni and Robertson [?] in the setting of the Black-Scholes model, is to use the large deviations theory to construct an easily computable approximation to the optimal importance sampling measure parameter θ_{opt}^* . Namely, we use Varadhan's lemma and the pathwise large deviation principle for Lévy processes due to Leonard [?] to derive a proxy for the variance of the importance sampling estimator which is much easier to compute than the true variance. We propose then to use the measure θ^* , obtained by minimizing this proxy, in the importance sampling estimator. Numerical illustrations in Section 5 show that the variance obtained by using θ^* instead of θ_{opt}^* is very close to the optimal one, and that a considerable variance reduction is obtained in all examples with small computational overhead.

When the logarithm of the pay-off P is concave, which is the case in many applications, the proxy for the variance may be further simplified using convex duality. The computation of the asymptotically optimal measure θ^* then reduces to one finite-dimensional optimization problem for European options and to the solution of one ODE system (Euler-Lagrange equations) for the path-dependent ones. In other words, additional complexity is the same as in the case of the Black-Scholes model studied in [?], even though our model is much more complex.

The rest of this paper is structured as follows. In Section 2 we recall the notation and results from the theory of large deviations which are used in the paper. Section 3

provides a representation for the proxy of the variance, a simplified representation in the case of concave log-payoffs and an easy to verify criterion for concavity. Section 4 presents explicit computations for European basket and Asian options. Numerical illustrations of these examples, in the context of the variance gamma model, are provided in Section 5. Lastly, the Appendix contains two technical lemmas.

2 Pathwise large deviations for Lévy processes

In this section we recall and extend the known results on large deviations for Lévy processes, which will be used in the sequel, and introduce all the necessary notation. We first formulate the large deviations principle (LDP) on abstract spaces. Let \mathcal{X} be a Hausdorff topological space endowed with its Borel σ -field. A rate function is a $[0, \infty]$ -valued lower semi-continuous function on \mathcal{X} . It is said to be a good rate function if its level sets are compact.

Definition 3.1 (Large Deviation Principle). *A family $\{X^\varepsilon\}$ of \mathcal{X} -valued random variables is said to obey a LDP in \mathcal{X} with rate function I if for each open subset $G \subset \mathcal{X}$ and each closed subset $F \subset \mathcal{X}$*

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{P} [X^\varepsilon \in F] \leq - \inf_{x \in F} I(x)$$

and

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{P} [X^\varepsilon \in G] \geq - \inf_{x \in G} I(x).$$

The following result is the famous Varadhan's lemma, which allows to evaluate limits of functions of X^ε in the large deviations asymptotics. More precisely, we need an extension of this lemma, given in [?], which allows the function ϕ to take the value $-\infty$. This is necessary since the pay-off of the option may take zero value on part of the domain, and the function ϕ will contain the log-pay-off in the sequel.

Lemma 3.2 (Varadhan's lemma). *Suppose that $\{X_\varepsilon\}$ satisfies the LDP with a good rate function $I : \mathcal{X} \rightarrow [0, \infty]$ and let $\phi : \mathcal{X} \rightarrow [-\infty, \infty[$ be such that the set $\{\phi > -\infty\}$ is open and ϕ is continuous on this set. Assume further that for some $\gamma > 1$,*

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{E} \left[e^{\frac{\gamma \phi(X_\varepsilon)}{\varepsilon}} \right] < \infty.$$

Then,

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{E} \left[e^{\frac{\phi(X_\varepsilon)}{\varepsilon}} \right] = \sup_{x \in \mathcal{X}} \{\phi(x) - I(x)\}.$$

We shall next recall the pathwise large deviation principle for Lévy processes, but first, following [?], we need to introduce specific topological spaces well suited for this application, and recall some preliminary results on Lévy processes.

Spaces and topologies As usual, D denotes the space of right-continuous functions with left limits (RCLL) $y : [0, T] \rightarrow \mathbb{R}^n$ with $y(0) = 0$. The subspace of D containing all functions in D with bounded variation will be denoted by V_r . Similarly, the space of all left-continuous functions with bounded variation on $[0, T]$ will be denoted by V_l . The symbol M will denote the class of bounded \mathbb{R}^n -valued measures on $[0, T]$. Note that there is a one-to-one correspondence between elements of V_r and the elements of M with no atom at 0: in particular, for every such $\mu \in M$, the function $t \mapsto \mu([0, t])$ belongs to V_r . Let $\sigma(D, M)$ denote the coarsest topology on D for which the mapping $x \mapsto \int_{[0, T]} x d\mu$ is continuous for every $\mu \in M$. This topology is stronger than the topology of pointwise convergence but weaker than the uniform topology.

For future reference, we let V_r^{ac} denote the subspace of V_r consisting of all functions of the form $x_t = \int_{[0, t]} \dot{x}_s ds$ with $\dot{x} \in L^1([0, T])$, equipped with the norm $\|x\| = \int_{[0, T]} |\dot{x}_s| ds$.

Preliminaries on Lévy processes Recall that the law of an \mathbb{R}^n -valued Lévy process X is characterized by its Lévy triplet (A, ν, γ) via the Lévy-Khintchine formula

$$\mathbb{E}[e^{i\langle u, X_t \rangle}] = \exp t \left(i\langle u, \gamma \rangle - \frac{\langle Au, u \rangle}{2} + \int_{\mathbb{R}^n} (e^{i\langle u, x \rangle} - 1 - i\langle u, x \rangle 1_{|x| \leq 1}) \nu(dx) \right)$$

Here, the Lévy measure ν is a positive measure on \mathbb{R}^n , which satisfies

$$\int (|x|^2 \wedge 1) \nu(dx) < \infty$$

and governs the intensity of jumps, the matrix A is a positive definite $n \times n$ matrix, which corresponds to the covariance of the diffusion component and the vector $\gamma \in \mathbb{R}^n$ is related to the deterministic linear component of X . We shall need the following lemma.

Lemma 3.3. *Let $\theta \in M$ and let X be a Lévy process on \mathbb{R}^n with characteristic triplet (A, ν, γ) . Then,*

$$\log \mathbb{E} \left[e^{\int_{[0, T]} \langle X_t, \theta(dt) \rangle} \right] = \int_{[0, T]} G(\theta([t, T])) dt \in [0, \infty],$$

where

$$G(u) = \langle u, \gamma \rangle + \frac{\langle Au, u \rangle}{2} + \int_{\mathbb{R}^n} (e^{\langle u, x \rangle} - 1 - \langle u, x \rangle 1_{|x| \leq 1}) \nu(dx). \quad (3.2)$$

Proof. Step 1. Since the trajectories of X are almost surely càdlàg, they are bounded on $[0, T]$. Since θ is also bounded, the dominated convergence theorem and the right-continuity of X lead to the following almost sure approximation result.

$$\int_{[0, T]} \langle X_t, \theta(dt) \rangle = \lim_{n \rightarrow \infty} \int_{(0, T]} \langle X_{\eta^n(t)}, \theta(dt) \rangle = \lim_{n \rightarrow \infty} \sum_{i=1}^n \langle X_{t_i^n}, \theta((t_{i-1}^n, t_i^n]) \rangle,$$

where we denote $t_i^n = \frac{iT}{n}$ and $\eta^n(t) = t_i^n$ whenever $t \in (t_{i-1}^n, t_i^n]$. Using once again the dominated convergence theorem, we conclude that for all $u \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E} \left[e^{iu \int_{[0,T]} \langle X_t, \theta(dt) \rangle} \right] &= \lim_{n \rightarrow \infty} \mathbb{E} \left[\exp \left(iu \sum_{i=1}^n \langle X_{t_i^n}, \theta((t_{i-1}^n, t_i^n]) \rangle \right) \right] \\ &= \lim_{n \rightarrow \infty} \mathbb{E} \left[\exp \left(iu \sum_{j=1}^n \langle X_{t_j^n} - X_{t_{j-1}^n}, \theta((t_{j-1}^n, T]) \rangle \right) \right] \\ &= \lim_{n \rightarrow \infty} \exp \left(\frac{T}{n} \sum_{j=1}^n G(iu\theta((t_{j-1}^n, T])) \right) \\ &= \exp \left(\int_{[0,T]} G(iu\theta([t, T])) dt \right) = \exp \left(\int_{[0,T]} G(iu\theta([t, T])) dt \right), \end{aligned}$$

where the last line can once again be deduced from the dominated convergence theorem for Lebesgue integrals and the fact that the mapping $t \mapsto \theta([t, T])$ is càdlàg.

Step 2. Let us write $\Theta_t = \theta([t, T])$. Then,

$$\begin{aligned} \int_{[0,T]} G(iu\theta([t, T])) dt &= iu \int_{[0,T]} \langle \gamma, \Theta_t \rangle dt - \frac{u^2}{2} \int_{[0,T]} \langle A\Theta_t, \Theta_t \rangle dt \\ &\quad + iu \int_{[0,T]} dt \int_{\mathbb{R}^n} \langle \Theta_t, x \rangle (\mathbf{1}_{|\langle \Theta_t, x \rangle| \leq 1} - \mathbf{1}_{|x| \leq 1}) \nu(dx) \\ &\quad + \int_{[0,T]} \int_{\mathbb{R}} (e^{iuz} - 1 - iuz \mathbf{1}_{|z| \leq 1}) \nu_t(dz) dt, \end{aligned}$$

where $\nu_t(A) := \nu(\{x \in \mathbb{R} : \langle \Theta_t, x \rangle \in A\})$. The integral in the last line is well defined due to the following estimate (compare with Proposition 11.10 in [?]):

$$\begin{aligned} \int_{[0,T]} \int_{\mathbb{R}} (z^2 \wedge 1) \nu_t(dz) dt &= \int_{[0,T]} \int_{\mathbb{R}^n} (|\langle \Theta_t, x \rangle|^2 \wedge 1) \nu(dx) dt \\ &\leq \int_{[0,T]} (\|\Theta_t\|^2 \vee 1) \int_{\mathbb{R}^n} (\|x\|^2 \wedge 1) \nu(dx) dt < \infty. \end{aligned}$$

Fubini's theorem (justified by the same estimate) and Lévy-Khintchine formula then allow us to conclude that $\int_{[0,T]} \langle X_t, \theta(dt) \rangle$ is an infinitely divisible random variable with characteristic triplet (A^*, ν^*, γ^*) , where

$$\begin{aligned} A^* &= \int_{[0,T]} \langle A\Theta_t, \Theta_t \rangle dt; \\ \nu^*(B) &= \int_{[0,T]} \nu_t(B) dt, \quad B \in \mathcal{B}(\mathbb{R}); \\ \gamma^* &= \int_{[0,T]} \langle \gamma, \Theta_t \rangle dt + \int_{[0,T]} \int_{\mathbb{R}^d} \langle \Theta_t, x \rangle (\mathbf{1}_{|\langle \Theta_t, x \rangle| \leq 1} - \mathbf{1}_{|x| \leq 1}) \nu(dx). \end{aligned}$$

The statement of the lemma now follows from Theorem 25.17 in [?].

□

In the sequel, we shall make use of the following assumption on the Lévy process X .

(A1) There exists $\lambda_0 > 0$ with $\int_{|x|>1} e^{\lambda_0|x|} \nu(dx) < \infty$.

A large deviations principle for Lévy processes In the following, we let X denote a \mathbb{R}^n -valued Lévy process on $[0, T]$ with characteristic triple (A, ν, γ) . We introduce a family of Lévy processes $(X^\varepsilon)_{\varepsilon>0}$ defined by $X_t^\varepsilon = \varepsilon X_{t/\varepsilon}$. The following theorem is a corollary of Theorem 5.3 in [?].

Theorem 3.4. *Suppose that Assumption (A1) holds true. Then the family (X^ε) satisfies the LDP in D for the $\sigma(D, M)$ -topology with the good rate function $J(y)$ where*

$$J(x) = \begin{cases} \sup_{\mu \in M} \left\{ \int_{[0, T]} \langle x_t, \mu(dt) \rangle - \int_{[0, T]} G(\mu([t, T])) dt \right\} & \text{if } x \in V_r, \\ + \infty & \text{otherwise.} \end{cases}$$

Proof. First, assume that $A = 0$. To apply [?, Theorem 5.3], we need to check that (in the notation of this reference) $\theta_0 \in \tilde{\mathcal{L}}^d$. Since X is a Lévy process, $\theta_0(t, a) = \phi_0(a) = a$ (see Equation (5.3) in [?]). Therefore, it is sufficient to check that $a \in L_\tau(\Lambda)$ (see [?, page 104]). In our notation this is equivalent to (see Equation (2.1) in [?]):

$$\exists \alpha > 0 : \int_{\mathbb{R}^n} \tau(a/\alpha) \nu(da) \leq 1, \quad (3.3)$$

where $\tau(x) = e^{|x|} - |x| - 1$. Let R be such that $\nu(\{x : |x| > R\}) < \frac{1}{4}$. We shall use the following decomposition.

$$\begin{aligned} \int_{\mathbb{R}^n} \tau(a/\alpha) \nu(da) &\leq \int_{|x|>R} e^{|a|/\alpha} \nu(da) + \nu(\{x : |x| > R\}) \\ &\quad + \frac{1}{\alpha} \int_{|x|>R} |a| \nu(da) + \int_{|x|\leq R} (e^{|a|/\alpha} - |a|/\alpha - 1) \nu(da). \end{aligned}$$

Using the dominated convergence theorem and Assumption (A1), it is easy to see that the first term in the right-hand side converges to $\nu(\{x : |x| > R\})$ as $\alpha \rightarrow \infty$, while the third and the last terms can be made arbitrarily small by taking α large enough. This means that the bound (3.3) holds true.

We will now extend this result to the general case ($A \neq 0$). Let W be an \mathbb{R}^n -valued standard Brownian motion on $[0, T]$, and let σ be any matrix such that $\sigma^\top \sigma = A$. Since σW is a Lévy process admitting all exponential moments, by

Theorem 1.2 in [?], the family $\{\sigma W^\varepsilon := \varepsilon \sigma W_{./\varepsilon}\}$ satisfies a LDP in D endowed with the uniform topology, with rate function

$$J_A(x) = \begin{cases} \int_{[0,T]} G^*(\dot{x}_s) ds, & \text{if } x \in V_r^{ac}, \\ +\infty & \text{otherwise,} \end{cases}$$

where

$$G^*(\eta) = \sup_{\xi \in \mathbb{R}^n} \left\{ \langle \eta, \xi \rangle - \frac{1}{2} \langle A\xi, \xi \rangle \right\}.$$

By Lemma 3.16, the following representation for J_A also holds true.

$$J_A(x) = \begin{cases} \sup_{\mu \in M} \left\{ \int_{[0,T]} \langle x_t, \mu(dt) \rangle - I_A(\mu) \right\} & \text{if } x \in V_r^{ac} \\ +\infty & \text{otherwise,} \end{cases}$$

with

$$I_A(\mu) = \frac{1}{2} \int_{[0,T]} \langle A\mu([t, T]), \mu([t, T]) \rangle dt.$$

Let the continuous martingale part of the Lévy process X be given by σW , and define $Z^\varepsilon = \varepsilon Z_{./\varepsilon}$ with $Z_t = X_t - \sigma W_t$. In view of the independence of Z and W , and the first part of the proof, the couple $(Z^\varepsilon, \sigma W^\varepsilon)$ satisfies the LDP on the product space $D \times D$ with the rate function $\bar{J}(g) + J_A(f)$, where

$$\bar{J}(x) = \begin{cases} \sup_{\mu \in M} \left\{ \int_{[0,T]} \langle x_t, \mu(dt) \rangle - \bar{I}(\mu) \right\} & \text{if } x \in V_r \\ +\infty & \text{otherwise,} \end{cases}$$

with $\bar{I}(\mu) = \int_{[0,T]} \bar{G}(\mu([t, T])) dt$, and \bar{G} defined by the same formula as G but with $A = 0$.

Then by the contraction principle (Theorem 4.2.1 in [?]) applied to the mapping $(x, y) \mapsto x + y$, the family (X^ε) satisfies the LDP on D with the good rate function

$$J(z) = \inf_{x \in V_r, y \in V_r: x+y=z} \{ \bar{J}(x) + J_A(y) \} = \bar{J} \square J_A(z) = \bar{I}^* \square I_A^*(z),$$

where the star denotes the convex conjugate. By the classical result on inf-convolution, this implies that

$$J^*(\mu) = \bar{I}^{**}(\mu) + I_A^{**}(\mu).$$

Since both \bar{I} and I_A are positive convex lower semicontinuous functions, by Proposition 1.3.3 in [?] they coincide with their bipolars, so that

$$J^*(\mu) = \bar{I}(\mu) + I_A(\mu).$$

Finally, since J is a good rate function, it is convex and lower semicontinuous, and it also clearly admits an affine minorant because both \bar{J} and J_A are nonnegative. Thus, once again by Proposition 1.3.3 in [?],

$$J(x) = J^{**}(x) = (\bar{I} + I_A)^*(x),$$

which proves the theorem. □

Note that De Acosta [?] proves an LDP for the uniform topology under the assumption that all exponential moments are finite. However, this assumption is too strong in practice, since most financial models are based on Lévy processes with exponential tail decay.

3 Main results

As mentioned in the introduction, our importance sampling estimator is based on the path-dependent Esscher transform,

$$\frac{d\mathbb{P}^\theta}{d\mathbb{P}} = \frac{e^{\int_{[0,T]} \langle X_t, \theta(dt) \rangle}}{\mathbb{E} \left[e^{\int_{[0,T]} \langle X_t, \theta(dt) \rangle} \right]} \quad (3.4)$$

where θ is a (deterministic) bounded \mathbb{R}^n -valued signed measure on $[0, T]$.

The optimal choice of θ should minimize the variance of the estimator under \mathbb{P}^θ ,

$$\text{Var}_{\mathbb{P}^\theta} \left(P \frac{d\mathbb{P}}{d\mathbb{P}^\theta} \right) = \mathbb{E}_{\mathbb{P}} \left[P^2 \frac{d\mathbb{P}}{d\mathbb{P}^\theta} \right] - \mathbb{E}[P]^2$$

In the following we shall denote by H the log-payoff of the option expressed as function of the Lévy process, with the convention that $\log 0 = -\infty$:

$$H(X) = \log P(S), \quad X_t^i = \log \frac{S_t^i}{S_0^i}, \quad 0 \leq t \leq T.$$

Then, using Lemma 3.3, the minimization problem writes

$$\inf_{\theta \in M} \mathbb{E}_{\mathbb{P}} \left[\exp \left\{ 2H(X) - \int_{[0,T]} \langle X_t, \theta(dt) \rangle + \int_{[0,T]} G(\theta([t, T])) dt \right\} \right],$$

where

$$G(\theta) = \langle \theta, \gamma \rangle + \int_{\mathbb{R}^n} (e^{\langle \theta, x \rangle} - 1 - \langle \theta, x \rangle 1_{|x| \leq 1}) \nu(dx).$$

Given the possibly complex form of the log-payoff H , the above expression for the variance is difficult to minimize. Our approach is instead to minimize a proxy of the variance, which has a more tractable form. Our first main result provides an expression for such a proxy, which we aim to minimize to obtain an asymptotically optimal variance reduction.

Proposition 3.5. *Let Assumption (A1) hold true, and suppose that the set $\{x \in D : H(x) > -\infty\}$ is open and contains the zero element, that H is continuous on this set for the $\sigma(D, M)$ -topology, and that H and $\theta \in M$ satisfy one of the following alternative assumptions.*

(i) *There exist constants $C < \infty$ and $B < \lambda_0/4n$ such that*

$$H(x) \leq C + B \sup_{s \in [0, T]} \sum_{i=1}^n |x_s^i|$$

and

$$\max_{0 \leq t \leq T} |\theta([t, T])| < \lambda_0 - 4nB.$$

(ii) *H is bounded and*

$$\int_{[0, T]} G(-\gamma\theta([t, T])) dt < \infty$$

for some $\gamma > 1$.

Then it holds that

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{E} \left[e^{\frac{2H(X^\varepsilon) - \int_{[0, T]} \langle X_t^\varepsilon, \theta(dt) \rangle}{\varepsilon}} \right] = \sup_{x \in D} \left\{ 2H(x) - \int_{[0, T]} \langle x_t, \theta(dt) \rangle - J(x) \right\}.$$

Proof. Since the pay-off H is assumed to be continuous on the set $\{x : H(x) > -\infty\}$, the continuity of the mapping

$$x \mapsto 2H(x) - \int_{[0, T]} \langle x_t, \theta(dt) \rangle$$

on the same set for the $\sigma(D, M)$ -topology follows from the definition of this topology. It remains to check the integrability condition of Varadhan's lemma under the alternative assumptions. Under the assumption (ii), this condition follows directly from Lemma 3.3. We shall therefore admit that assumption (i) holds true for the rest of this proof. Then we may choose $p > 1$ and $q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$, as well as $\gamma > 1$, such that

$$q\gamma \max_{0 \leq t \leq T} |\theta([t, T])| < \lambda_0 \quad \text{and} \quad 4\gamma pnB < \lambda_0. \quad (3.5)$$

Moreover, there exists $b > 0$ with

$$\mathbb{E} \left[(X_t^i - bt) e^{4B\gamma pn(X_t^i - bt)} \right] < 0 \quad \text{and} \quad \mathbb{E} \left[(-X_t^i - bt) e^{4B\gamma pn(-X_t^i - bt)} \right] < 0 \quad (3.6)$$

for all $t > 0$ and all i . To see this, remark that, for example, for the first inequality,

$$\mathbb{E} \left[(X_t^i - bt) e^{4B\gamma pn(X_t^i - bt)} \right] = \mathbb{E} \left[e^{4B\gamma pn(X_t^i - bt)} \widehat{\mathbb{E}}[X_t^i - bt] \right],$$

where $\widehat{\mathbb{E}}$ denotes the expectation under the new probability measure $\widehat{\mathbb{P}}$ defined by

$$\frac{d\widehat{\mathbb{P}}}{d\mathbb{P}} \Big|_{\mathcal{F}_t} = \frac{e^{4B\gamma pn X_t^i}}{\mathbb{E} [e^{4B\gamma pn X_t^i}]}.$$

In view of the condition $4\gamma pn B < \lambda_0$, this probability measure is well defined, and the process X is a Lévy process with finite expectation under $\widehat{\mathbb{P}}$ (see e.g., Example 33.14 in [?]). Therefore, the first condition in (3.6) is satisfied by taking $b > \widehat{\mathbb{E}}[X_1^i]$ (this choice does not depend on t). The second condition is ensured in a similar way.

Then, by the assumption on H , Hölder's inequality, and Lemma 3.3, the following estimates hold true:

$$\begin{aligned} & \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{E} \left[e^{\frac{\gamma(2H(X^\varepsilon) - \int_{[0,T]} \langle X_t^\varepsilon, \theta(dt) \rangle)}{\varepsilon}} \right] \\ & \leq 2C\gamma + \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{E} \left[e^{\frac{\gamma(2B \sup_{s \in [0,T]} \sum_{i=1}^n |X_{s/\varepsilon}^{i,\varepsilon}| - \int_{[0,T]} \langle X_t^\varepsilon, \theta(dt) \rangle)}{\varepsilon}} \right] \\ & = 2C\gamma + \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{E} \left[e^{\gamma(2B \sup_{s \in [0,T]} \sum_{i=1}^n |X_{s/\varepsilon}^i| - \int_{[0,T]} \langle X_{t/\varepsilon}, \theta(dt) \rangle)} \right] \\ & \leq 2C\gamma + 2bnT\gamma B + \sum_{i=1}^n \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{E} \left[e^{4B\gamma pn \sup_{s \in [0,T]} (X_{s/\varepsilon}^i - bs/\varepsilon)} \right] \\ & \quad + \sum_{i=1}^n \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{E} \left[e^{4B\gamma pn \sup_{s \in [0,T]} (-X_{s/\varepsilon}^i - bs/\varepsilon)} \right] \\ & \quad + \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{E} \left[e^{-q\gamma \int_{[0,T]} \langle X_{t/\varepsilon}, \theta(dt) \rangle} \right] \\ & \leq 2C\gamma + 2bnT\gamma B + \sum_{i=1}^n \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{E} \left[e^{4B\gamma pn \sup_{s \geq 0} (X_s^i - bs)} \right] \\ & \quad + \sum_{i=1}^n \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbb{E} \left[e^{4B\gamma pn \sup_{s \geq 0} (-X_s^i - bs)} \right] + \int_{[0,T]} G(-q\gamma\theta([t, T])) dt < \infty, \end{aligned}$$

where the finiteness of the two sums follows from Lemma 3.15 applied to processes $X_t^i - bt$ and $-X_t^i - bt$, and that of the last term follows from the first inequality in (3.5). □

In view of Proposition 3.5 we suggest to obtain the candidate variance reduction parameter θ^* by minimizing

$$\sup_{x \in V_r} \left\{ 2H(x) - \int_{[0,T]} \langle x_t, \theta(dt) \rangle + \int_{[0,T]} G(\theta([t, T])) dt - J(x) \right\} \quad (3.7)$$

over $\theta \in M$. It is then natural to ask, how close the corresponding measure change will be to the optimal one which minimizes the variance of the Monte Carlo estimator

over all possible measure changes. Varadhan's lemma allows to define a notion of asymptotic optimality, which provides a partial answer to this question. Consider a family of importance sampling measures $(\mathbb{Q}(\varepsilon))_{\varepsilon>0}$. By Jensen's inequality,

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E}^{\mathbb{Q}(\varepsilon)} \left[\left(e^{\frac{1}{\varepsilon} H(X^\varepsilon)} \frac{d\mathbb{P}}{d\mathbb{Q}(\varepsilon)} \right)^2 \right] \geq 2 \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[e^{\frac{1}{\varepsilon} H(X^\varepsilon)} \right]$$

We shall call a family of importance sampling measure changes asymptotically optimal if, for this family, the above inequality becomes an equality and the lim-inf/limsup become limits. In other words, the asymptotically optimal measure change does at least as well as any other measure change at the logarithmic scale of large deviations.

Definition 3.6. *Let $(\mathbb{Q}(\varepsilon))_{\varepsilon>0}$ be a family of importance sampling measure changes. We say that $(\mathbb{Q}(\varepsilon))_{\varepsilon>0}$ is asymptotically optimal if*

$$\lim_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E}^{\mathbb{Q}(\varepsilon)} \left[\left(e^{\frac{1}{\varepsilon} H(X^\varepsilon)} \frac{d\mathbb{P}}{d\mathbb{Q}(\varepsilon)} \right)^2 \right] = 2 \lim_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[e^{\frac{1}{\varepsilon} H(X^\varepsilon)} \right].$$

The following theorem, which is the main result of this paper, shows that for concave log-payoffs, the importance sampling measure associated to the parameter θ^* computed by minimizing the expression (3.7) is asymptotically optimal, and its computation is greatly simplified. In other words, for such pay-offs, it is sufficient to consider only measure changes of the form (3.4) with deterministic θ to achieve asymptotic optimality over all possible measure changes. European basket put options and many path-dependent put-like payoffs encountered in practice are indeed concave. This theorem shall be proven under the following assumption.

(A2) The function G is lower semicontinuous and its effective domain $\mathcal{D} := \{x \in \mathbb{R}^n : G(x) < \infty\}$ is open and bounded.

Remark 3.7. This assumption may appear quite restrictive. It is satisfied by models where the tail decay is exactly exponential, such as variance gamma, normal inverse gaussian, CGMY and their multidimensional versions. However, it rules out models with faster than exponential tail decay such as the celebrated Merton's model. We expect that for such models a similar result may still be shown, but one would need to use different, and slightly more complex methods (Orlicz spaces instead of L^∞). To keep the length of the proof reasonable, we have chosen to present the argument in the case of a bounded domain.

Note also, that even though Assumption (A2) is required for a rigorous proof of asymptotic optimality of θ^* , even without this assumption the large deviations result suggests that a good candidate importance sampling measure change may

be found by solving the variational problem in the right-hand side of Equation 3.8. The importance sampling optimization problem is ill-posed to begin with, since identifying the optimal measure requires one to know the expectation that one is interested in computing in the first place. As such, the objective is to identify a change of measure which is 'good', and Theorem 3.8 certainly provides us with a candidate measure, which can be used to reduce variance, and whose existence does not require the Assumption (A2).

Theorem 3.8. *Let H be concave, and assume that the set $\{x \in V_r : H(x) > -\infty\}$ is nonempty and contains the zero element, and that H is continuous on this set for $\sigma(D, M)$ -topology. Let Assumptions (A1) and (A2) be satisfied. Then,*

$$\begin{aligned} \inf_{\theta \in M} \sup_{x \in V_r} \left\{ 2H(x) - \int_{[0,T]} \langle x_t, \theta(dt) \rangle + \int_{[0,T]} G(\theta([t, T])) dt - J(x) \right\} \\ = 2 \inf_{\theta \in M} \left\{ \widehat{H}(\theta) + \int_{[0,T]} G(\theta([t, T])) dt \right\} \quad (3.8) \end{aligned}$$

where

$$\widehat{H}(\theta) = \sup_{x \in V_r} \left\{ H(x) - \int_{[0,T]} \langle x_t, \theta(dt) \rangle \right\}.$$

Moreover, if the infimum in the right-hand side of (3.8) is attained by θ^* then the same value θ^* attains the infimum in the left-hand side of (3.8). If, in addition, the pay-off functional H and the candidate parameter θ^* satisfy the assumptions (i) or (ii) of Proposition 3.5 then the importance sampling measure corresponding to θ^* is asymptotically optimal.

The proof of this theorem is based on the following minimax result.

Lemma 3.9. *Let H and G satisfy the assumptions of Theorem 3.8. Then,*

$$\begin{aligned} \sup_{x \in V_r} \inf_{\mu \in M} \left\{ 2H(x) - \int_{[0,T]} \langle x_t, \mu(dt) \rangle + \int_{[0,T]} G(\mu([t, T])) dt \right\} \\ = \inf_{\mu \in M} \sup_{x \in V_r} \left\{ 2H(x) - \int_{[0,T]} \langle x_t, \mu(dt) \rangle + \int_{[0,T]} G(\mu([t, T])) dt \right\}. \end{aligned}$$

Proof. To prove this minimax result, we are going to use the classical Banach space theory. The first step is therefore to show that sup and inf above may be taken over Banach spaces.

Step 1. The goal of this step is to show that

$$\begin{aligned} \sup_{x \in V_r} \inf_{\mu \in M} \left\{ 2H(x) - \int_{[0,T]} \langle x_t, \mu(dt) \rangle + \int_{[0,T]} G(\mu([t, T])) dt \right\} \\ = \sup_{\dot{x} \in L^2([0,T])} \inf_{y \in L^2([0,T])} \left\{ 2H(x) - \int_{[0,T]} \langle \dot{x}_t, y_t \rangle dt + \int_{[0,T]} G(y_t) dt \right\} \quad (3.9) \end{aligned}$$

We shall first prove that the sup may be taken over $L^2([0, T])$, or, in other words, that

$$\sup_{x \in V_r} \{2H(x) - J(x)\} = \sup_{\dot{x} \in L^2([0, T])} \{2H(x) - J(x)\}. \quad (3.10)$$

To prove (3.10) we will establish that for every $x \in V_r$ there is a sequence $\{x_n\}$ with $\dot{x}_n \in L^2([0, T])$ with $J(x_n) \rightarrow J(x)$ and $H(x_n) \rightarrow H(x)$. By Lemma 3.16,

$$J(x) = \begin{cases} \int_{[0, T]} L_a \left(\frac{d\dot{x}_a}{dt}(t) \right) dt + \int_{[0, T]} L_s \left(\frac{d\dot{x}_s}{d\mu}(t) \right) d\mu & \text{if } x \in V_r \\ + \infty & \text{otherwise.} \end{cases} \quad (3.11)$$

Note that since the effective domain of G is bounded, the function L_a is Lipschitz, and we denote its Lipschitz constant by K . This means that also $|L_s(x/\|x\|)| \leq K$ for all $x \in \mathbb{R}^n$. We shall apply Theorem 2.2 in [?], more precisely point ii. of Remark 2.3 after this theorem. Letting $f(x, s) = L_a(s)$, it is easy to check, possibly after adding a constant to f , that f satisfies Assumptions (i)–(vi) of this theorem. Then, let μ_n for all n and μ be equal to the Lebesgue measure on $[0, T]$ and let λ be the vector-valued measure on $[0, T]$ associated to the finite variation function x . We may represent

$$\lambda = \lambda^c + \sum_{i=1}^{\infty} c_i \delta_{x_i},$$

where λ^c has no atom, $\{\|c_i\|\}$ is a decreasing sequence with $\sum_{i=1}^{\infty} \|c_i\| < \infty$, $\{x_i\}$ are distinct points with $x_i \in (0, T]$ for all i and δ_x denotes the Dirac measure at x .

By Theorem 2.2 and Remark 2.3 in [?], in view of the representation (3.11), there exists a sequence $\{\lambda_n^c\}$ of measures on $[0, T]$, with $\lambda_n^c \ll \mu$ for all n , converging to λ^c in the weak* topology such that

$$\lim_n J(x_n^c) = J(x^c), \quad (3.12)$$

where $x_n^c(t) = \int_{[0, t]} \lambda_n^c(ds)$ and $x^c = \int_{[0, t]} \lambda^c(ds)$. It is easy to see that we may assume that $\lambda_n^c \in L^2([0, T])$ for every n . Since x^c is continuous, x_n^c converges to x^c in $\sigma(D, M)$. Indeed, since λ has no atom, for every $t \in [0, T]$, $x_n^c(t)$ converges to $x^c(t)$. Then, by dominated convergence, for every $\mu \in M$, $\int_{[0, T]} x_n^c(t) \mu(dt)$ converges to $\int_{[0, T]} x^c(t) \mu(dt)$.

Fix $\varepsilon > 0$, let n_ε be such that $\sum_{i > n_\varepsilon} (1 + K) \|c_i\| < \varepsilon$, and

$$|J(x_n^c) - J(x^c)| \leq \varepsilon$$

for $n \geq n_\varepsilon$. Let $\delta_\varepsilon < \varepsilon$ be such that $x_i > \delta_\varepsilon$ for $i = 1, \dots, n_\varepsilon$, and the intervals $[x_i - \delta_\varepsilon, x_i]$ for $i = 1, \dots, n_\varepsilon$ are disjoint, and moreover

$$\sum_{i=1}^{n_\varepsilon} \int_{x_i - \delta_\varepsilon}^{x_i} (L_a(0) + 2K|\lambda_{n_\varepsilon}^c|) ds \leq \varepsilon$$

and

$$\left| \sum_{i=1}^{n_\varepsilon} \delta_\varepsilon L_a\left(\frac{c_i}{\delta_\varepsilon}\right) - \sum_{i=1}^{n_\varepsilon} L_s(c_i) \right| \leq \varepsilon.$$

Such a choice of δ_ε is possible in virtue of the integrability of $\lambda_{n_\varepsilon}^c$ and of the definition of L_s .

Define

$$\lambda_\varepsilon = \lambda_{n_\varepsilon}^c + \sum_{i=1}^{n_\varepsilon} \frac{c_i}{\delta_\varepsilon} \mathbf{1}_{x_i - \delta_\varepsilon \leq x \leq x_i} dx,$$

and $x_\varepsilon(t) = \int_{[0,t]} \lambda_\varepsilon(s) ds$, so that $\dot{x}_\varepsilon \in L^2([0, T])$. Then, in view of our choice of n_ε and δ_ε ,

$$\begin{aligned} |J(x) - J(x_\varepsilon)| &= \left| J(x^c) - J(x_\varepsilon^c) + \sum_{i=1}^{\infty} L_s(c_i) + \sum_{i=1}^{n_\varepsilon} \int_{x_i - \delta_\varepsilon}^{x_i} \left\{ L_a\left(\lambda_{n_\varepsilon}^c + \frac{c_i}{\delta_\varepsilon}\right) - L_a(\lambda_{n_\varepsilon}^c) \right\} ds \right| \\ &\leq 2\varepsilon + \left| \sum_{i=1}^{n_\varepsilon} L_s(c_i) + \sum_{i=1}^{n_\varepsilon} \int_{x_i - \delta_\varepsilon}^{x_i} \left\{ L_a\left(\lambda_{n_\varepsilon}^c + \frac{c_i}{\delta_\varepsilon}\right) - L_a(\lambda_{n_\varepsilon}^c) \right\} ds \right| \\ &\leq 4\varepsilon + \left| \sum_{i=1}^{n_\varepsilon} \{L_s(c_i) - \delta_\varepsilon L_a(c_i/\delta_\varepsilon)\} \right| \leq 5\varepsilon. \end{aligned}$$

Now, let (ε_k) be a sequence of positive numbers converging to zero. We have shown that $J(x_{\varepsilon_k}) \rightarrow J(x)$ as $k \rightarrow \infty$. It remains to show that x_{ε_k} converges to x in the $\sigma(D, M)$ topology. By [?, Proposition B1] we may equivalently prove that for every LCRL function $f : [0, T] \rightarrow \mathbb{R}^n$ with bounded variation,

$$\lim_{k \rightarrow \infty} \int_{[0, T]} f(s) \lambda_{\varepsilon_k}(ds) = \int_{[0, T]} f(s) \lambda(ds).$$

For the continuous part this convergence has already been established, so that it remains to study the convergence of

$$\sum_{i=1}^{n_{\varepsilon_k}} \frac{c_i}{\delta_{\varepsilon_k}} \int_{x_i - \delta_{\varepsilon_k}}^{x_i} f(s) ds.$$

By the left continuity of f , each term converges to $c_i f(x_i)$, and the summability of $\sum_{i=1}^{\infty} \|c_i\|$ implies that the sum also converges. We have thus shown that x_{ε_k} converges to x in the $\sigma(D, M)$ topology as $k \rightarrow \infty$, and the proof of (3.10) is completed using the continuity of H .

Now, let $\dot{x} \in L^2([0, T])$. By Proposition B.1 in [?],

$$\begin{aligned} \inf_{\mu \in M} \{2H(x) - \int_{[0, T]} \langle x_t, \mu(dt) \rangle + \int_{[0, T]} G(\mu([t, T])) dt\} \\ = \inf_{y \in V_t} \{2H(x) - \int_{[0, T]} \langle y_t, \dot{x}_t \rangle dt + \int_{[0, T]} G(y_t) dt\} \end{aligned}$$

Let $y \in L^2([0, T])$ such that $\int_{[0, T]} G(y_t) dt < \infty$. By modifying y on a set of Lebesgue measure zero, we may assume that $y_t \in \text{dom } G$ for all $t \in [0, T]$. Then, for every $n \geq 1$, $(1 - 1/n)y_t \in \text{int dom } G$ for all $t \in [0, T]$ and there is a constant K_n with $G((1 - 1/n)y_t) \leq K_n$ for $t \in [0, T]$.

For a fixed n , we can find a sequence of functions $(y^m)_{m \geq 1}$ belonging to V_l with $G(y^m) \leq K_n$ for $t \in [0, T]$, which converges to $(1 - 1/n)y$ in Lebesgue measure on $[0, T]$ (use Lusin's theorem plus a uniform approximation of continuous functions with piecewise constant functions). Then, by the dominated convergence theorem, in view of the continuity of G on the interior of its domain,

$$-\int_{[0, T]} \langle \dot{x}_t, y_t^m \rangle dt + \int_{[0, T]} G(y_t^m) dt \rightarrow (1 - \frac{1}{n}) \int_{[0, T]} \langle \dot{x}_t, y_t \rangle dt - \int_{[0, T]} G((1 - \frac{1}{n})y_t) dt$$

as $m \rightarrow \infty$. On the other hand, by Equation (6.2) in [?] (which is easily seen to be valid even in the case of a nonzero Gaussian component), as $n \rightarrow \infty$,

$$(1 - 1/n) \int_{[0, T]} \langle \dot{x}_t, y_t \rangle dt - \int_{[0, T]} G((1 - 1/n)y_t) dt \rightarrow \int_{[0, T]} \langle \dot{x}_t, y_t \rangle dt - \int_{[0, T]} G(y_t) dt$$

This finishes the proof of Step 1.

Step 2. Our aim in this step is to show that sup and inf in (3.9) may be exchanged. We shall use the classical minimax result ([?], Proposition VI.2.3 on page 175). By our assumption on H , it is upper semicontinuous on V_r for the $\sigma(D, M)$ -topology, and thus also upper semicontinuous on $L^2([0, T])$ for the strong norm. On the other hand, the mapping

$$y \mapsto \int_{[0, T]} G(y_t) dt$$

is convex and lower semicontinuous in the strong topology of L^2 (see [?]). Finally, since the effective domain is bounded, the set of y such that $\int_{[0, T]} G(y_s) ds < \infty$ is bounded in L^2 norm. Therefore, Proposition VI.2.3 applies and

$$\begin{aligned} & \sup_{\dot{x} \in L^2([0, T])} \inf_{y \in L^2([0, T])} \{2H(x) - \int_{[0, T]} \langle \dot{x}_t, y_t \rangle dt + \int_{[0, T]} G(y_t) dt\} \\ &= \inf_{y \in L^2([0, T])} \sup_{\dot{x} \in L^2([0, T])} \{2H(x) - \int_{[0, T]} \langle \dot{x}_t, y_t \rangle dt + \int_{[0, T]} G(y_t) dt\}. \end{aligned} \quad (3.13)$$

Step 3. It remains to be seen that the infimum in (3.13) can be taken over the space V_l and the supremum over V_r . To prove the first point, we shall show that the inner supremum equals $+\infty$ whenever y has unbounded variation; the infimum can then be restricted to V_l since a function of bounded variation has a countable number of points of discontinuity and the integrals are computed with respect to the Lebesgue measure. Since $G(y) \geq 0$, it is enough to consider the

first two terms. By assumptions on H , there exists $\varepsilon > 0$ and $C < \infty$ such that $H(x) > -C$ whenever $x \in B_\varepsilon(0)$, where the ball is defined for the topology of uniform convergence. Assume that y has unbounded variation and let $N > 0$. By definition of the total variation, there exists $\psi \in C_0^\infty([0, T])$ with $\|\psi\|_{L^\infty} \leq 1$ such that $\int_{[0, T]} \langle y_t, \dot{\psi}_t \rangle dt < -\frac{N}{\varepsilon}$. Therefore, letting $x = \varepsilon\psi$, we have $x \in B_\varepsilon(0)$ and so

$$2H(x) - \int_{[0, T]} \langle \dot{x}_t, y_t \rangle dt > -C + N.$$

Using the integration by parts (Proposition B.1 in [?]), this proves that the right-hand side of (3.13) is equal to

$$\inf_{\mu \in M} \sup_{\dot{x} \in L^2([0, T])} \left\{ 2H(x) - \int_{[0, T]} \langle x_t, d\mu_t \rangle + \int_{[0, T]} G(\mu([t, T])) dt \right\}.$$

Finally, using the same approximation argument as in the first step, we may show that the inner supremum may be taken over $x \in V_r$. \square

Proof of Theorem 3.8. Applying Lemma 3.9 to the modified pay-off function

$$\tilde{H}(x) = H(x) - \frac{1}{2} \int_{[0, T]} \langle x_t, \theta(dt) \rangle,$$

which clearly also satisfies the required assumptions, and using the expression of the rate function, J , we get

$$\begin{aligned} & \inf_{\theta \in M} \sup_{x \in V_r} \left\{ 2H(x) - \int_{[0, T]} \langle x_t, \theta(dt) \rangle + \int_{[0, T]} G(\theta([t, T])) dt - J(x) \right\} \\ &= \inf_{\theta \in M} \sup_{x \in V_r} \inf_{\mu \in M} \left\{ 2H(x) - \int_{[0, T]} \langle x_t, \theta(dt) + \mu(dt) \rangle \right. \\ & \quad \left. + \int_{[0, T]} G(\theta([t, T])) dt + \int_{[0, T]} G(\mu([t, T])) dt \right\} \\ &= \inf_{\theta \in M} \inf_{\mu \in M} \sup_{x \in V_r} \left\{ 2H(x) - \int_{[0, T]} \langle x_t, \theta(dt) + \mu(dt) \rangle \right. \\ & \quad \left. + \int_{[0, T]} G(\theta([t, T])) dt + \int_{[0, T]} G(\mu([t, T])) dt \right\} \\ &= \inf_{\theta \in M} \inf_{\mu \in M} \left\{ 2\hat{H} \left(\frac{\theta + \mu}{2} \right) + \int_{[0, T]} G(\theta([t, T])) dt + \int_{[0, T]} G(\mu([t, T])) dt \right\} \quad (3.14) \end{aligned}$$

$$= 2 \inf_{\theta \in M} \left\{ \hat{H}(\theta) + \int_{[0, T]} G(\theta([t, T])) dt \right\}, \quad (3.15)$$

where the last equality follows by convexity of G , taking $\mu = \theta$.

To prove the last statement of the theorem, assume that the infimum in (3.15) is attained by θ^* . Then, using the equality of (3.14) and (3.15), and taking $\mu = \theta^*$ in (3.14), we see that the same value θ^* also attains the infimum in (3.14). Let \mathbb{P}^{θ^*}

be the importance sampling measure change associated to θ^* via (3.4). Then, the following estimates hold true.

$$\begin{aligned}
& \lim_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E}^{\mathbb{P}^{\theta^*}} \left[\left(e^{\frac{1}{\varepsilon} H(X^\varepsilon)} \frac{d\mathbb{P}}{d\mathbb{P}^{\theta^*}} \right)^2 \right] \\
&= \inf_{\theta \in M} \sup_{x \in V_r} \left\{ 2H(x) - \int_{[0,T]} \langle x_t, \theta(dt) \rangle + \int_{[0,T]} G(\theta([t, T])) dt - J(x) \right\} \\
&= \sup_{x \in V_r} \inf_{\theta \in M} \left\{ 2H(x) - \int_{[0,T]} \langle x_t, \theta(dt) \rangle + \int_{[0,T]} G(\theta([t, T])) dt - J(x) \right\} \\
&= \sup_{x \in V_r} \{2H(x) - 2J(x)\} = 2 \lim_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{E} \left[e^{\frac{1}{\varepsilon} H(X^\varepsilon)} \right],
\end{aligned}$$

where the first and the last equalities follow from Proposition 3.5, the second one follows from Lemma 3.9, and the third one follows from the expression of J . \square

Concavity of the log-payoff The concavity of the log-payoff function $H(x)$ may be tested using the following simple lemma. We recall that $X_t^i = \log \frac{S_t^i}{S_0^i}$ for $i = 1, \dots, n$.

Lemma 3.10. *Let $\tilde{P}(X) = P(S)$ and assume that \tilde{P} is concave on the set $\mathcal{X}^+ := \{x \in D : \tilde{P}(x) > 0\}$ and that the set \mathcal{X}^+ is convex. Then the log-payoff $H : D \mapsto \bar{\mathbb{R}}$ is concave in x .*

Proof. Let $0 < \alpha < 1$ and choose $a, b \in \mathcal{X}^+$. Then,

$$\begin{aligned}
\alpha H(a) + (1 - \alpha)H(b) &= \alpha \log \tilde{P}(a) + (1 - \alpha) \log \tilde{P}(b) \\
&\leq \log(\alpha \tilde{P}(a) + (1 - \alpha)\tilde{P}(b)) \leq \log \tilde{P}(\alpha a + (1 - \alpha)b) = H(\alpha a + (1 - \alpha)b),
\end{aligned}$$

which shows that H is concave on \mathcal{X}^+ . Since $H(x) = -\infty$ for $x \notin \mathcal{X}^+$ and the set \mathcal{X}^+ is convex, H is also concave on the whole space. \square

4 Examples

In this section, we specialize the results of the previous section to several option pay-offs encountered in practice. Throughout this section we assume that the Lévy process satisfies the assumptions (A1) and (A2). For each considered pay-off, we need to check the assumptions of Proposition 3.5 to ensure that the asymptotically optimal variance reduction measure θ^* may indeed be defined as in Definition 3.6, and the assumptions of Theorem 3.8, to ensure that one can use the simplified formula (3.8) to compute θ^* .

General European pay-off In the case of European pay-offs, the problem of finding the optimal parameter θ^* is finite-dimensional and the optimal measure change is the standard Esscher transform.

Proposition 3.11. *Assume that $H((x_t)_{0 \leq t \leq T}) = h(x_T)$ with $h : \mathbb{R}^n \rightarrow \mathbb{R}$ concave and continuous on the set $\{x \in \mathbb{R}^n : h(x) > -\infty\}$, which is assumed to be open and contain the point $x = 0$. Let G satisfy the assumptions 1 and 2. Then, assumptions of Theorem 3.8 are satisfied and $\theta^* = \bar{\theta}^* \delta_T$, where δ_T is the Dirac measure at T , and*

$$\bar{\theta}^* = \arg \min_{\theta \in \mathbb{R}^n} \{\hat{h}(\theta) + TG(\theta)\},$$

where $\hat{h}(\theta) = \sup_{v \in \mathbb{R}^n} \{h(v) - \langle v, \theta \rangle\}$. In addition, if h is bounded then the additional assumption (ii) of Proposition 3.5 is satisfied and θ^* is asymptotically optimal.

Proof. Note first that the minimizer exists since G is l.s.c. and coercive by Assumption 2, and \hat{h} is l.s.c. and bounded from below. The log-payoff H clearly satisfies the assumptions of Theorem 3.8. If $\theta([0, T]) \neq 0$, then $\hat{H}(\theta) = +\infty$ since one can choose $x_t = a \mathbf{1}_{t < T}$ with a arbitrary. This means that one can restrict the optimization in (3.8) to measures of the form $\theta \delta_T$ where $\theta \in \mathbb{R}^n$, and the rest of the proof follows easily. Finally, since the effective domain of G is open, θ^* belongs to its interior, and therefore there exists $\gamma > 1$ such that

$$\int_{[0, T]} G(\gamma \theta^*([t, T])) dt = TG(\gamma \theta^*) < \infty,$$

which shows that (ii) of Proposition 3.5 is satisfied. \square

Remark 3.12. We observe that the function $G(\theta)$ is known explicitly in most models. In addition, under the measure \mathbb{P}^θ , X is still a Lévy process which often falls into the same parametric class (see e.g., the variance gamma example in the following section). Thus, the only overhead of using the importance sampling estimator proposed in this paper for European options is due to the additional time needed to solve an explicit convex optimization problem in dimension n , which is usually negligible.

European basket put option Now consider a specific European pay-off of the form $P(S^1, \dots, S^n) = (K - S_T^1 - \dots - S_T^n)^+$. Then, using the notation of the previous paragraph,

$$h(x_1, \dots, x_n) = \log(K - S_0^1 e^{x_1} - \dots - S_0^n e^{x_n})^+.$$

To simplify notation we shall assume that $S_0^1 = \dots = S_0^n = 1$ and $K > n$ (the other situations may be considered similarly). This function is bounded from above, continuous on the set where it is not equal to $-\infty$, and finite at $x = 0$. On the

other hand, the function $\tilde{P} = (K - e^{x_1} - \dots - e^{x_n})^+$ is concave on $\{\tilde{P} > 0\}$ by convexity of the exponential and the set $\{e^{x_1} + \dots + e^{x_n} < K\}$ is convex. Therefore, by Lemma 3.10, the function h is concave. Therefore, the asymptotically optimal variance reduction measure parameter θ^* is given in Proposition 3.11, where the convex conjugate of h is easily shown to be

$$\hat{h}(\theta) = \begin{cases} +\infty & \theta_k \geq 0 \text{ for some } k \\ -\left(1 - \sum_k \theta_k\right) \log \frac{1 - \sum_k \theta_k}{K} - \sum_k \theta_k \log(-\theta_k) & \text{otherwise.} \end{cases}$$

Numerical examples for the European basket put option are given in the next section.

Arithmetic asian put option In this example the Lévy process X is one-dimensional and we consider the Asian option with log-payoff

$$H(x) = \log \left(K - \frac{1}{T} \int_{[0,T]} S_0 e^{x_t} dt \right)^+.$$

Once again, to simplify notation we assume that $S_0 = 1$ and $K > 1$ so that $H(0)$ is finite, otherwise a constant shift may be applied to x . Let us show that H is continuous in the $\sigma(D, M)$ -topology on the set where it is finite. Let $\{x^n\} \subset V_r$ be a sequence converging in the $\sigma(D, M)$ -topology. Clearly, this sequence also converges in the weak* topology of L^∞ , which is the dual of the Banach space L^1 . Therefore it follows from the Banach-Steinhaus theorem that $\{x^n\}$ is uniformly bounded. The continuity of e^H and the continuity of H on the set $\{H > -\infty\}$ now follows from the dominated convergence theorem.

Let us now check the concavity of H . Remark that $K - \frac{1}{T} \int_{[0,T]} e^{x_t} dt$ is concave by convexity of the exponential, and for $x, y \in D$ such that $\frac{1}{T} \int_{[0,T]} e^{x_t} dt < K$ and $\frac{1}{T} \int_{[0,T]} e^{y_t} dt < K$,

$$\frac{1}{T} \int_{[0,T]} e^{\alpha x_t + (1-\alpha)y_t} dt \leq \frac{1}{T} \int_{[0,T]} (\alpha e^{x_t} + (1-\alpha)e^{y_t}) dt < K,$$

which implies that the set $\{\tilde{P} > 0\}$ is convex. By Lemma 3.10, H is then concave.

Therefore, the assumptions of Theorem 3.8 are satisfied by the arithmetic Asian put option. The convex conjugate of H and the candidate optimal parameter θ^* are described by the following proposition.

Proposition 3.13.

- i. If θ is absolutely continuous, with density (also denoted by θ_t) satisfying $\theta_t \leq 0$ for all $t \in [0, T]$, then $\hat{H}(\theta)$ is given by*

$$\hat{H}(\theta) = \log \frac{K}{1 - \int_{[0,T]} \theta_t dt} - \int_{[0,T]} \theta_t \log \frac{-KT\theta_t}{1 - \int_{[0,T]} \theta_s ds} dt.$$

Otherwise $\widehat{H}(\theta) = +\infty$.

ii. There exists an asymptotically optimal measure parameter θ^* which minimizes the right-hand side of (3.8) and has a nonpositive density on $[0, T]$, denoted by $(\theta_t^*)_{0 \leq t \leq T}$.

iii. The function $\psi_t^* = \int_{T-t}^T \theta_s^* ds$ is the solution of the boundary value problem

$$\begin{aligned} \dot{p}_t &= -G'(\psi_t^*), & p_T &= -\log \frac{K}{1 - \psi_T^*} + 1, \\ \dot{\psi}_t^* &= -\frac{1}{T} e^{p_t - 1}, & \psi_0^* &= 0. \end{aligned}$$

Remark 3.14. This system can be integrated explicitly:

$$\ddot{\psi}_t = -\dot{\psi}_t G'(\psi_t) \quad \Rightarrow \quad \dot{\psi}_t = -G(\psi_t) - C \quad \Rightarrow \quad \int_{[0, \psi_t]} \frac{d\phi}{C + G(\phi)} = -t,$$

where the constant C is determined from the terminal condition.

Proof. Part i. By definition,

$$\widehat{H}(\theta) = \sup_{x \in V_r} \left\{ \log \left(K - \frac{1}{T} \int_{[0, T]} e^{x_t} dt \right)^+ - \int_{[0, T]} x_t \theta(dt) \right\}.$$

First, assume that there is an interval $[a, b) \subset [0, T]$ such that $\theta([a, b)) > 0$. Then, letting $x_t = \log K$ for $t \notin [a, b)$ and $x_t = -N$ for $t \in [a, b)$, and making N tend to $+\infty$, we see that $\widehat{H}(\theta) = +\infty$. The case when θ has a positive atom at T is treated similarly. Therefore, from now on we may assume that θ is a negative measure. Assume that it is not absolutely continuous. Then, there exists $\varepsilon > 0$ such that for all $\delta > 0$, there exists a finite sequence of pairwise disjoint sub-intervals $[x_k, y_k)$ of $[0, T]$ satisfying

$$\sum_k (y_k - x_k) < \delta$$

such that

$$\sum_k \theta([x_k, y_k)) < -\varepsilon.$$

We define $I = \cup_k [x_k, y_k)$. Let $x_t = N > \log(K/2)$ when $t \in I$ and $x_t = \log(K/2)$ otherwise. Then,

$$\log \left(K - \frac{1}{T} \int_{[0, T]} e^{x_t} dt \right)^+ - \int_{[0, T]} x_t \theta(dt) \geq \log \left(K/2 - \frac{\delta}{T} e^N \right)^+ - \theta(I^c) \log \frac{K}{2} + N\varepsilon.$$

Taking N sufficiently large and δ sufficiently small, we see that $\widehat{H}(\theta) = +\infty$ in this case as well. We may therefore assume that θ is an absolutely continuous negative

measure, and, with an abuse of notation, its density will also be denoted by θ_t . The computation of $\widehat{H}(\theta)$ reduces to computing the supremum

$$\sup_x \left\{ \log \left(K - \frac{1}{T} \int_{[0,T]} e^{x_t} dt \right) - \int_{[0,T]} x_t \theta_t dt \right\}, \quad (3.16)$$

where with no loss of generality we consider only those $x \in V_r$ for which the expression under the sign of logarithm is positive. For such x , the functional is Gateaux differentiable, and since it is also concave, the sufficient condition for optimality for this optimization problem writes

$$\theta_t = -\frac{1}{T} \frac{e^{x_t}}{K - \frac{1}{T} \int_{[0,T]} e^{x_s} ds}.$$

Integrating this expression from 0 to T , we find

$$K - \frac{1}{T} \int_{[0,T]} e^{x_s} ds = \frac{K}{1 - \int_{[0,T]} \theta_t dt}$$

and so

$$x_t = \log \frac{-KT\theta_t}{1 - \int_{[0,T]} \theta_t dt}.$$

Substituting this into (3.16), we obtain the first part of the proposition.

Part ii. To compute the candidate importance sampling measure parameter θ^* , we need to solve

$$\begin{aligned} & \min_{\theta \in M} \left\{ \widehat{H}(\theta) + \int_{[0,T]} G(\theta([t, T])) dt \right\} \\ & = \min_{\theta_t \leq 0} \left\{ \left(1 - \int_{[0,T]} \theta_s ds \right) \log \frac{K}{1 - \int_{[0,T]} \theta_t dt} - \int_{[0,T]} \theta_t \log(-T\theta_t) dt \right. \\ & \quad \left. + \int_{[0,T]} G\left(\int_t^T \theta_s ds \right) dt \right\} \end{aligned}$$

Introducing the function $\psi_t = \int_{T-t}^T \theta_s ds$, the optimization problem becomes

$$\min_{\psi_t \leq 0} \left(1 - \psi_T \right) \log \frac{K}{1 - \psi_T} - \int_{[0,T]} \dot{\psi}_t \log(-T\dot{\psi}_t) dt + \int_{[0,T]} G(\psi_t) dt. \quad (3.17)$$

Let (a, b) be the effective domain of G , and $\gamma : (-\infty, \infty) \rightarrow (a, b)$ be a strictly increasing C^1 function. The above problem can then be written as the standard deterministic control problem:

$$\begin{aligned} & \min_{u_t \leq 0, 0 \leq t \leq T} \left(1 - \gamma(x_T) \right) \log \frac{K}{1 - \gamma(x_T)} - \int_{[0,T]} u_t \log(-Tu_t) dt \\ & \quad + \int_{[0,T]} G(\gamma(x_t)) dt, \quad \dot{x}_t = \frac{u_t}{\gamma'(x_t)}. \end{aligned} \quad (3.18)$$

The existence of an absolutely continuous solution $x_t^*, 0 \leq t \leq T$ for the problem (3.18) and thus of an absolutely continuous solution $\psi_t^*, 0 \leq t \leq T$ for the problem (3.17) is ensured by Theorem III.4.1 and Corollary III.4.1 in [?]. Moreover, this solution satisfies $a < \psi_t^* < b$ for $t \in [0, T]$, and since ψ^* is continuous and the effective domain of G is open, there exists $\gamma > 1$ such that $\int_{[0, T]} G(\gamma\psi_t^*) dt < \infty$, which means that the condition (ii) of Proposition 3.5 is satisfied and the optimal measure parameter $\theta_t^* = \dot{\psi}_t^*$ is asymptotically optimal.

Part iii. By Pontryagin's maximum principle for deterministic control problems (see Theorem II.11.5 in [?] and the discussion on the equivalent problems in paragraph II.4 of this reference), the necessary condition for ψ^* to be the minimizer of (3.17) is that for every $t \in [0, T]$, $\dot{\psi}_t^*$ is the minimizer of

$$p_t \dot{\psi}_t - \dot{\psi}_t \log(-T \dot{\psi}_t),$$

where p_t is the ‘‘adjoint state’’ satisfying

$$\dot{p}_t = -G'(\psi_t^*), \quad p_T = 1 - \log K + \log(1 - \psi_T^*).$$

□

Numerical examples for the Asian put option are given in the next section.

5 Numerical illustrations

In this section, we illustrate the results of this paper with numerical computations in the multivariate variance gamma model. Let $b \in \mathbb{R}^n$, Σ be a positive definite $n \times n$ matrix, and define

$$X_t = \mu t + b\Gamma_t + \Sigma W_{\Gamma_t},$$

where W is a standard Brownian motion in dimension n , Γ is a gamma process with $\mathbb{E}[\Gamma_t] = t$ and $\text{Var} \Gamma_t = t/\lambda$, and μ is chosen to have $\mathbb{E}[e^{X_t^i}] = 1$ for all t and $i = 1, \dots, n$ so that \mathbb{P} is a risk-neutral measure. Then, the cumulant generating function X_1 under the original measure is given by

$$G(\theta) = \langle \theta, \mu \rangle - \lambda \log \left(1 - \frac{\langle \theta, b \rangle}{\lambda} - \frac{\langle \Sigma \theta, \theta \rangle}{2\lambda} \right), \theta \in \mathbb{R}^n.$$

with

$$\mu^i = \lambda \log \left(1 - \frac{b^i}{\lambda} - \frac{\Sigma_{ii}}{2\lambda} \right), \quad i = 1, \dots, n.$$

For this process, $\mathbb{E}[e^{\langle \theta, X_T \rangle}] < \infty$ if and only if

$$1 - \frac{\langle \theta, b \rangle}{\lambda} - \frac{\langle \Sigma \theta, \theta \rangle}{2\lambda} > 0.$$

Given a θ satisfying this condition, one can define a measure \mathbb{P}^θ as in (3.1). Under \mathbb{P}^θ , the cumulant generating function of X_1 can be written as

$$G^\theta(u) = \langle u, \mu \rangle - \lambda \log \left(1 - \frac{\langle u, b + \Sigma \theta \rangle}{\lambda u^*} - \frac{\langle \Sigma u, u \rangle}{2\lambda u^*} \right), \quad u^* = 1 - \frac{\langle \theta, b \rangle}{\lambda} - \frac{\langle \Sigma \theta, \theta \rangle}{2\lambda}.$$

Therefore, under the measure \mathbb{P}^θ , the process X^1 is also a variance gamma process with parameters $\tilde{\mu} = \mu$, $\tilde{\lambda} = \lambda$, $\tilde{b} = \frac{b + \Sigma \theta}{u^*}$ and $\tilde{\Sigma} = \frac{\Sigma}{u^*}$.

Vanilla put in the variance gamma model In the first example, we let $n = 1$ and price a European put option with pay-off $P(S) = (K - S)^+$. The model parameters are $\lambda = 1$, $b = -0.2$ and $\sqrt{\Sigma} = 0.2$, which corresponds to annualized volatility of 28%, skewness of -1.77 and excess kurtosis of 2.25.

Table 3.1 shows the option prices computed with the importance sampling algorithm over 1,000,000 Monte Carlo trajectories together with the corresponding standard errors, execution times and the optimal importance sampling parameter values as function of strike and maturity. It also presents the ratios of the variance of the standard algorithm to those of the IS algorithm. For the same sample size, it takes about 1.27 times longer to run the IS algorithm. This overhead comes almost exclusively from evaluating the IS estimator, the time needed to compute the optimal importance sampling parameter θ^* is negligible in this case (about 0.002 seconds). To take this extra overhead into account, we also show in the table the adjusted variance ratios, which correspond to variance ratios divided by the overhead factor. An adjusted ratio of, say 5.99 means that the IS algorithm is roughly 6 times faster than the standard algorithm for the same standard error.

We see that the highest ratios are attained for out-of-the-money options, whose exercise is a rare event, but that even for at-the-money options, the variance reduction ratios remain quite significant (and similar to those reported in [?, ?, ?]). It is also important to understand, how close are these ratios to the optimal ones which would have been obtained by minimizing the actual variance of the estimator rather than its asymptotic proxy. This is illustrated in Figure 3.1, which plots the variance of the importance sampling estimator (evaluated by Monte Carlo) as function of the parameter θ . We see that for the chosen parameter values θ^* is very close to optimality.

Basket put in the variance gamma model In this example, we let $n = 3$ and price a European basket put option with pay-off $P(S) = (K - S^1 - S^2 - S^3)^+$. The

Strike	Price	Std. error	Var. ratio	Adj. ratio	Time, s.	θ^*
0.5	0.0036521	5.9×10^{-6}	18.04	14.22	25.59	-2.84
0.7	0.017559	2.3×10^{-5}	7.61	5.99	25.25	-2.565
0.9	0.056557	5.8×10^{-5}	4.47	3.52	24.38	-2.237
1.1	0.143786	0.000103	3.32	2.62	25.27	-1.879
1.3	0.305516	0.000115	3.68	2.89	27.26	-1.535
1.5	0.500944	0.000108	4.55	3.59	22.63	-1.25

T	Price	Std. error	Var. ratio	Adj. ratio	Time, s.	θ^*
0.25	0.037874	5.02×10^{-5}	3.4	2.68	24.59	-2.768
0.5	0.060622	6.39×10^{-5}	3.57	2.81	24.01	-2.447
1	0.092327	8.05×10^{-5}	3.72	2.93	24.81	-2.059
2	0.135914	0.000102	3.73	2.94	27.06	-1.645
3	0.168254	0.000117	3.66	2.88	25.35	-1.409

TABLE 3.1. Single-asset European put option. Option prices, standard errors, execution times (IS algorithm), optimal importance sampling parameter values, and variance ratios, as function of strike (for $T = 1$) and time to maturity (for $K = 1$). The adjusted variance ratios are taking into account the computational overhead of the IS algorithm. The adjustment factor is taken equal to 1.27.

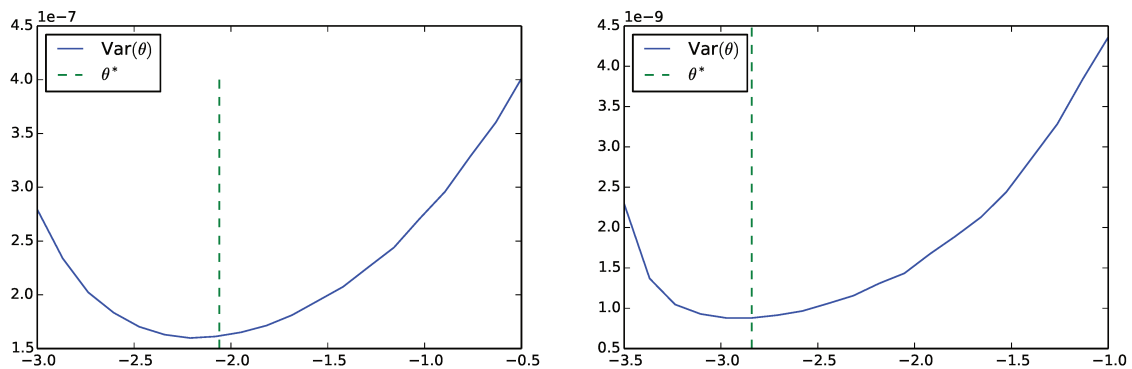


FIGURE 3.1. Optimality of θ^* . Left: $T = 1, K = 1$. Right: $T = 1, K = 0.5$.

model parameters are

$$\lambda = 1, \quad b = \begin{pmatrix} -0.2 \\ -0.2 \\ -0.2 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 0.04 & 0.02 & 0.02 \\ 0.02 & 0.04 & 0.02 \\ 0.02 & 0.02 & 0.04 \end{pmatrix}.$$

Table 3.2 shows the option prices, standard errors, execution times (IS algorithm), optimal importance sampling parameter values, and variance ratios as function of strike and time to maturity. We see that the values are similar to the one-dimensional case.

Asian put in the variance gamma model In this final example we price an Asian put option with pay-off $P(S) = \left(K - \frac{1}{T} \int_{[0,T]} S_t dt\right)^+$, for $T = 1$. The sample size for the Monte Carlo method was 1,000,000 trajectories. The variance reduction algorithm was implemented as follows. First, a discretization grid ($0 = t_0 < t_1 < \dots < t_n = T$) for computing the option's pay-off was fixed. In this example we chose a uniform grid with $n = 100$. The boundary value problem of Proposition 3.13 was then solved on this grid using a standard method (the routine `odeint` of Python for integrating the ODE and the bisection algorithm for solving the boundary value problem). For the numerical evaluation of the importance sampling estimator, ψ^* was then assumed to be constant on the intervals $[t_i, t_{i-1})$, in other words, the measure change applied in the numerical computation is given by

$$\frac{d\mathbb{P}^\psi}{d\mathbb{P}} = \frac{e^{\int_{[0,T]} \psi_t^* dX_t}}{\mathbb{E}[e^{\int_{[0,T]} \psi_t^* dX_t}]} = \frac{e^{\sum_{i=0}^{n-1} \psi_{t_i}^* (X_{t_{i+1}} - X_{t_i})}}{\mathbb{E}[e^{\sum_{i=0}^{n-1} \psi_{t_i}^* (X_{t_{i+1}} - X_{t_i})}]}$$

As observed in the beginning of this section, under the measure \mathbb{P}^ψ , the increment $(X_{t_{i+1}} - X_{t_i})$ follows the variance gamma process with parameters

$$\tilde{\mu}_i = \mu, \quad \tilde{\lambda} = \lambda, \quad \tilde{b} = \frac{b + \sum \psi_{t_i}^*}{u^*} \quad \text{and} \quad \tilde{\Sigma} = \frac{\Sigma}{u^*}$$

with $u^* = 1 - \psi_{t_i}^* b / \lambda - \Sigma(\psi_{t_i}^*)^2 / (2\lambda)$. We then compare the standard Monte Carlo estimator,

$$\hat{P} = \frac{1}{N} \sum_{j=1}^N \left(K - \frac{1}{n} \sum_{i=0}^{n-1} e^{X_{t_i}^{(j)}} \right)^+,$$

where $(X^{(j)})_{j=1}^N$ are i.i.d. samples of the discretized trajectory under the original measure \mathbb{P} , and the importance sampling (IS) estimator,

$$\hat{P}_\psi = \frac{1}{N} \sum_{j=1}^N \left(K - \frac{1}{n} \sum_{i=0}^{n-1} e^{\tilde{X}_{t_i}^{(j)}} \right)^+ \frac{e^{\sum_{i=0}^{n-1} \psi_{t_i}^* (\tilde{X}_{t_{i+1}} - \tilde{X}_{t_i})}}{\mathbb{E}[e^{\sum_{i=0}^{n-1} \psi_{t_i}^* (X_{t_{i+1}} - X_{t_i})}]},$$

T	Price	Std. error	Var. ratio	Adj. ratio	Time, s.
0.25	0.102355	0.000137	3.39	2.66	25.29
0.5	0.163641	0.000174	3.57	2.81	24.80
1	0.250309	0.000219	3.72	2.93	25.0
2	0.367213	0.000278	3.77	2.97	25.7
3	0.455914	0.000320	3.71	2.92	25.67

Strike	Price	Std. error	Var. ratio	Adj. ratio	Time, s.
1.5	0.00749572	1.27×10^{-5}	21.2	16.7	24.06
2.0	0.0321732	4.53×10^{-5}	9.25	7.27	23.65
2.5	0.0994115	0.000114	5.33	4.20	24.37
3.0	0.249647	0.0002198	3.71	2.93	24.59
3.5	0.545304	0.000315	3.29	2.59	25.47
4.0	1.00791	0.000296	4.27	3.36	24.34
4.5	1.50424	0.000269	5.25	4.14	25.21

Strike	Price	Std. error	Var. ratio	Adj. ratio	Time, s.
1.5	0.0350413	5.4×10^{-5}	6.33	4.98	24.26
2.0	0.106674	0.000124	4.70	3.70	23.07
2.5	0.242411	0.000212	4.23	3.33	23.07
3.0	0.455343	0.000321	3.71	2.92	22.95
3.5	0.749882	0.000461	2.91	2.29	23.40
4.0	1.12111	0.000595	2.41	1.89	22.85
4.5	1.55076	0.000687	2.19	1.72	22.77

TABLE 3.2. European basket put option. Top table: Option prices, standard errors, execution times (IS algorithm), optimal importance sampling parameter values, and variance ratios as function of time to maturity T , for $K = 1$. Middle table: same quantities as function of strike for $T = 1$. Bottom table: same quantities as function of strike for $T = 3$.

Strike	Price	Std. error	Var. ratio	Adjusted var. ratio	Time, s.
0.5	0.0003717	1.13×10^{-6}	38.9	27.6	60
0.7	0.003683	7.69×10^{-6}	11.1	7.84	67
0.9	0.02262	3.12×10^{-5}	4.54	3.22	63
1.1	0.10934	6.98×10^{-5}	3.12	2.21	65
1.3	0.30045	5.97×10^{-5}	5.02	3.56	74
1.5	0.500102	5.14×10^{-5}	6.90	4.89	62

TABLE 3.3. Asian put option. Prices (IS algorithm), standard errors (IS algorithm), variance reduction ratios, variance reduction ratios adjusted for overhead of the importance sampling algorithm and computational times. The adjustment factor (ratio of computational time required to run the IS algorithm to that of the standard algorithm with the same sample size) is taken equal to 1.41.

where $(\tilde{X}^{(j)})_{j=1}^N$ are i.i.d. samples of the discretized trajectory under the measure \mathbb{P}^ψ .

The ratio of the time required to compute the IS estimator to that of the standard estimator is approximately 1.41. This overhead comes once again largely from the extra time needed to compute the measure change factor in the estimator, compared to which the time required for solving the boundary value problem is negligible (about 0.1 seconds).

Table 3.3 shows, for different strikes, the option price estimates obtained with the IS estimator, corresponding standard errors, variance reduction ratios (ratio of the variance of the standard estimator to that of the IS estimator), the adjusted variance reduction ratios (divided by the overhead factor 1.41), and the execution times. We see that the variance reduction ratios are even better than the ones obtained for the European call, since for low strikes the exercise probability is smaller for the Asian option than for the European option with the same strike and maturity.

Figure 3.2 plots the “distribution function” $\psi_t^* = \theta^*([t, T])$ of the asymptotically optimal measure θ^* as function of time t for $K = 1$. We see that for the Asian option, the optimal measure change is indeed very different from the standard non path dependent Esscher transform.

Appendix

Lemma 3.15. *Let X be a real-valued Lévy process. We denote $\bar{X}_t = \sup_{0 \leq s \leq t} X_s$. Let $\beta' > 0$ be such that $\mathbb{E}[e^{\beta' X_1}] < \infty$, $\mathbb{E}[e^{\beta' X_1} | X_1|] < \infty$ and $\mathbb{E}[e^{\beta' X_1} X_1] < 0$. Then,*

$$\mathbb{E}[e^{\beta \bar{X}_\infty}] < \infty$$

for all $\beta \in (0, \beta')$.

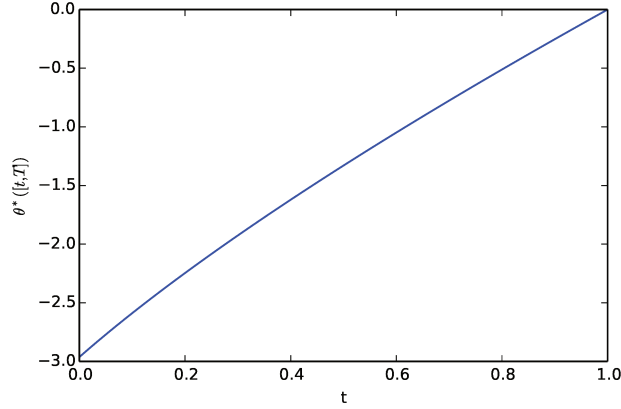


FIGURE 3.2. Asian put option. Distribution function of the asymptotically optimal measure θ^* , for $K = 1$ and $T = 1$.

Proof. By the Wiener-Hopf factorization (Theorem 6.16 in [?], see also Equation (47.9) in [?]),

$$q \int_{[0, \infty]} e^{-qt} \mathbb{E}[e^{\beta \bar{X}_t}] dt = \exp \left(\int_{[0, \infty]} \frac{e^{-qt}}{t} dt \int_{[0, \infty]} (e^{\beta x} - 1) \mathbb{P}[X_t \in dx] \right).$$

Letting $q \rightarrow 0$ and using the Hölder inequality, we get

$$\begin{aligned} \mathbb{E}[e^{\beta \bar{X}_\infty}] &= \exp \left(\int_{[0, \infty]} \frac{1}{t} dt \int_{[0, \infty]} (e^{\beta x} - 1) \mathbb{P}[X_t \in dx] \right) \\ &= \exp \left(\int_{[0, \infty]} \frac{1}{t} dt \mathbb{E}[(e^{\beta X_t} - 1) \mathbf{1}_{X_t \geq 0}] \right) \\ &\leq \exp \left(\int_{[0, \infty]} \frac{\beta}{t} dt \mathbb{E}[X_t e^{\beta X_t} \mathbf{1}_{X_t \geq 0}] \right) \\ &\leq \exp \left(\int_{[0, \infty]} \frac{\beta}{t} dt \mathbb{E}[|X_t|^{\frac{\beta'}{\beta' - \beta}} \mathbf{1}_{X_t \geq 0}]^{1 - \frac{\beta}{\beta'}} \mathbb{E}[e^{\beta' X_t} \mathbf{1}_{X_t \geq 0}]^{\frac{\beta}{\beta'}} \right) \end{aligned}$$

Let $\widehat{X}_t = X_t - \sum_{s \leq t: \Delta X_s < -1} \Delta X_s$. Since $\mathbb{E}[e^{\beta' X_1}] < \infty$, it follows that all moments of \widehat{X}_t are finite (for all $t > 0$) and

$$\mathbb{E}[|X_t|^{\frac{\beta'}{\beta' - \beta}} \mathbf{1}_{X_t \geq 0}] \leq \mathbb{E}[|\widehat{X}_t|^{\frac{\beta'}{\beta' - \beta}}].$$

Let $n^* = \inf\{n \in \mathbb{N} : n \geq \frac{\beta'}{\beta' - \beta}\}$. Then,

$$\mathbb{E}[|\widehat{X}_t|^{\frac{\beta'}{\beta' - \beta}}] \leq \mathbb{E}[|\widehat{X}_t|^{n^*}]^{\frac{\beta'}{n^*(\beta' - \beta)}} \leq C(t + t^{n^*})^{\frac{\beta'}{n^*(\beta' - \beta)}}$$

for some $C < \infty$, where the last inequality follows, e.g., from the Burkholder-Davis-Gundy inequality for jump processes [?, Lemma 2.1]. Thus,

$$\mathbb{E}[e^{\beta \bar{X}_\infty}] \leq \exp \left(\beta C \int_{[0, \infty]} dt (t^{\frac{1-n^*}{n^*}} + 1) \mathbb{E}[e^{\beta' X_t} \mathbf{1}_{X_t \geq 0}]^{\frac{\beta}{\beta'}} \right). \quad (3.19)$$

Further,

$$\mathbb{E}[e^{\beta' X_t} \mathbf{1}_{X_t \geq 0}] \leq \inf_{\lambda \geq 0} \mathbb{E}[e^{(\beta' + \lambda) X_t}] = e^{t \inf_{\lambda \geq 0} \psi(\beta' + \lambda)},$$

where $\psi(\lambda) = \log \mathbb{E}[e^{\lambda X_t}]$. By dominated convergence theorem,

$$\psi'_-(\beta') = \frac{\mathbb{E}[X_1 e^{\beta' X_1}]}{\mathbb{E}[e^{\beta' X_1}]} < 0,$$

so that $\inf_{\lambda \geq 0} \psi(\beta' + \lambda) = \inf_{\lambda \in \mathbb{R}} \psi(\beta' + \lambda)$. Finally, the condition $\mathbb{E}[e^{\beta' X_1} X_1] < 0$ entails that $\mathbb{E}[X_1] < 0$, which means that $\psi'(0) < 0$. This shows that $\inf_{\lambda \in \mathbb{R}} \psi(\beta' + \lambda) < 0$ and the integral in (3.19) converges. \square

Lemma 3.16. *Let $x \in V_r$ and let G be the Laplace exponent of a general Lévy process, defined in (3.2). Then,*

$$\sup_{\mu \in M} \left\{ \int_{[0, T]} \langle x_t, \mu(dt) \rangle - \int_{[0, T]} G(\mu([t, T])) dt \right\} \quad (3.20)$$

$$= \sup_{y \in V_t^{pc}} \left\{ \int_{[0, T]} \langle \dot{x}(dt), y_t \rangle - \int_{[0, T]} G(y_t) dt \right\} \quad (3.21)$$

$$= \sup_{y \in C_0([0, T])} \left\{ \int_{[0, T]} \langle \dot{x}(dt), y_t \rangle - \int_{[0, T]} G(y_t) dt \right\} \quad (3.22)$$

$$= \int_{[0, T]} L_a \left(\frac{d\dot{x}_a}{dt}(t) \right) dt + \int_{[0, T]} L_s \left(\frac{d\dot{x}_s}{d\rho} \right) d\rho, \quad (3.23)$$

where

$$L_a(v) = \sup_{\lambda \in \mathbb{R}^d} \{ \langle \lambda, v \rangle - G(\lambda) \}$$

and

$$L_s(v) = \lim_{u \rightarrow \infty} \frac{L_a(uv)}{u}.$$

Here, V_t^{pc} is the set of left-continuous piecewise constant functions on $[0, T]$, $\dot{x} \in M$ denotes the distributional derivative of the bounded variation function x , $\dot{x} = \dot{x}_a + \dot{x}_s$ is the decomposition of the measure \dot{x} in absolutely continuous and singular parts with respect to the Lebesgue measure on $[0, T]$, $\frac{d\dot{x}_a}{dt}$ denotes the density \dot{x}_a with respect to the Lebesgue measure, ρ is any non-negative measure on $[0, T]$, with respect to which \dot{x}_s is absolutely continuous and $\frac{d\dot{x}_s}{d\rho}$ is the Radon-Nikodym derivative of \dot{x}_s with respect to ρ .

Proof. The equality of (3.21) and (3.22) follows from Lemma 6.1 in [?]. This lemma is stated for a Lévy process without Gaussian component, but it is easy to see that the presence of such component does not alter the validity of equation (6.2) in the proof of this lemma, which is the only place where the structure of the Lévy process

is used. The equality of (3.22) and (3.23) then follows from Theorem 5 in [?]. The conditions of this theorem are satisfied since the set $D(t)$ of this theorem, which in our setting is equal to $\text{dom } G$, does not depend on t . It remains to prove the equality of (3.20) and (3.21), or, more precisely, that (3.20) \leq (3.21).

By integration by parts (Proposition B1 in [?]),

$$\int_{[0,T]} \langle x_t, \mu(dt) \rangle = \int_{[0,T]} \langle \dot{x}(dt), \mu([t, T]) \rangle.$$

Recall that as a convex function, G is continuous on the interior of its effective domain. Let $\mu \in M$ such that $\int_{[0,T]} G(\mu([t, T])) dt < \infty$. We define $y_t = \mu([t, T])$, $t \in [0, T]$, so that $y \in V_l$ and $y_t^n = (1 - 1/n)y_t$, $0 \leq t \leq T$. Therefore, $y_t^n \in \text{int dom } G$ for all $t \in [0, T]$, and by left continuity $\max G(y_t^n) < \infty$. Moreover, y^n converges uniformly to y as $n \rightarrow \infty$, which, together with equation (6.2) in [?] implies that

$$\int_{[0,T]} \langle \dot{x}(dt), y_t^n \rangle - \int_{[0,T]} G(y_t^n) dt \rightarrow \int_{[0,T]} \langle \dot{x}(dt), y_t \rangle - \int_{[0,T]} G(y_t) dt.$$

Therefore, we can and will assume that $\max G(y_t^n) < \infty$.

Let $y_t^c := y_t - \sum_{0 \leq s \leq t: \Delta y_s \neq 0} \Delta y_s$. Fix $\varepsilon > 0$. Since y^c is continuous, we can find n_ε such that $\max_{0 \leq s, t \leq T: |t-s| \leq \frac{1}{n_\varepsilon}} |y_s^c - y_t^c| \leq \frac{\varepsilon}{2}$. Moreover, we can find $\delta > 0$ such that $\sum_{0 \leq s \leq T: 0 < |\Delta y_s| \leq \delta} |\Delta y_s| < \frac{\varepsilon}{2}$. Now, let I be the set of points containing the points $\frac{Ti}{n_\varepsilon}$, $0 \leq i \leq n_\varepsilon$ and all the points $s \in [0, T]$ such that $|\Delta y_s| > \delta$. We let $y_t^\varepsilon = y_{\theta(t)}$, where $\theta_t = \max\{s \in I : s < t\}$. It is clear that y^ε is piecewise constant and $|y_t^\varepsilon - y_t| \leq \varepsilon$ for $0 \leq t \leq T$, which means that y^ε converges uniformly to y as $\varepsilon \rightarrow 0$. In addition, $\max_t G(y_t^\varepsilon) \leq \max_t G(y_t) < \infty$. Then, by dominated convergence,

$$\int_{[0,T]} \langle \dot{x}(dt), y_t^\varepsilon \rangle - \int_{[0,T]} G(y_t^\varepsilon) dt \rightarrow \int_{[0,T]} \langle \dot{x}(dt), y_t \rangle - \int_{[0,T]} G(y_t) dt$$

and so

$$\begin{aligned} \sup_{\mu \in M} \left\{ \int_{[0,T]} \langle x_t, \mu(dt) \rangle - \int_{[0,T]} G(\mu([t, T])) dt \right\} \\ \leq \sup_{y \in V_l^{pc}} \left\{ \int_{[0,T]} \langle \dot{x}(dt), y_t \rangle - \int_{[0,T]} G(y_t) dt \right\} \end{aligned}$$

□

Chapter 4

Tail asymptotics of log-normal mixture portfolios

Contents

1	Introduction	94
2	Notation and preliminaries	95
3	Asymptotic behavior of the left tail	97
4	Examples	103
4.1	Variance Gamma process	103
4.2	Heston model	103
5	Variance Reduction of Monte Carlo estimates	105
6	Proof of Theorem 4.3	109

1 Introduction

In this paper we consider the tail behavior of the sum of n dependent positive random variables

$$X = \sum_{i=1}^n X_i.$$

In financial mathematics, X may represent the value of a long-only portfolio of n assets, and understanding the tail behavior of X is important for risk management applications, such as computing the Value at Risk, evaluating tail event probabilities or designing efficient simulation algorithms for tail events. In particular, stress test scenarios may be constructed in a systematic manner by simulating the values of the components X_1, \dots, X_n conditionally on the event that X takes a given small value.

This problem has received considerable attention in the literature, but mainly in the insurance context, where the random variables X_1, \dots, X_n represent losses from individual claims, and one is interested in the *right tail* asymptotics of X , so as to estimate the probability of having a very large aggregate loss. In this setting, provided the variables X_1, \dots, X_n are sufficiently fat-tailed (subexponential), under various assumptions on the dependence structure, it can be shown that the right tail behavior of X is determined by the single variable with the fattest tail. We refer to [?, ?, ?, ?, ?, ?, ?, ?] and the references therein for precise statements and proofs in various contexts of this result, known as the "principle of single big jump".

In this paper, we focus on the context of financial risk management where the extreme event of interest corresponds to a *small* value of the random variable X . In this context, to estimate the probability of a large loss, one needs to focus on the *left tail* asymptotics of X . Owing to the positivity of the variables X_1, \dots, X_n , the asymptotic behavior of the left tail of X turns out to be very different from that of the right tail.

The problem of left tail asymptotics of the sum of positive random variables has until recently received surprisingly little attention in the literature. The case when X_1, \dots, X_n follow a multidimensional log-normal distribution has been treated in [?] (for $n = 2$), [?] (where sharp asymptotics for the density and the distribution function of X have been computed) and [?] (where efficient Monte Carlo methods have been proposed).

When the variables X_1, \dots, X_n are identically distributed and asymptotically dependent, the tail behavior of X can often be deduced from that of the individual components as, for example, in Wüthrich [?]. However, in financial applications the returns of different assets are not identically distributed, and many models of interest are asymptotically independent in the left tail (although they are not independent).

In the case of asymptotically independent components, only log-scale asymptotics have been computed in [?] which is not sufficient for most applications.

In this paper, we compute sharp asymptotics of the distribution function and the density of X in the left tail, and discuss the relevant risk management applications, under the assumption that X_1, \dots, X_n follow a log-normal mixture distribution. That is, we assume that for $i = 1, \dots, n$, $X_i = e^{Y_i}$, where the vector $\mathbf{Y} = (Y_1, \dots, Y_n)$ follows a Gaussian variance-mean mixture distribution.

The Gaussian variance-mean mixture distribution on \mathbb{R}^n is defined as :

$$\mathbf{Y} = \sqrt{\Theta} \mathbf{Z} + \Theta \boldsymbol{\mu} + \boldsymbol{\Lambda}$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\boldsymbol{\Lambda} \in \mathbb{R}^n$ are constant vectors, $\mathbf{Z} \sim N(0, \mathfrak{B})$ where $\mathfrak{B} \in \mathbb{M}^{n \times n}$ is a symmetric, positive definite matrix, and $\Theta \in \mathbb{R}_+$ is the mixing random variable. In this paper, Θ is assumed to admit a density, denoted $\rho(s)$, which decays exponentially fast as $s \rightarrow \infty$. We refer the interested readers to [?] for more information about Gaussian mean-variance mixtures. Many multidimensional log-return distributions encountered in financial mathematics such as the variance gamma model [?, ?], the normal inverse Gaussian model [?], the generalized hyperbolic model [?, ?] and Heston's model [?] with zero correlation between the volatility and the asset returns have the form of a Gaussian mixture with exponentially decaying mixing variable.

We will next briefly overview the contents of the present paper. Section 2 introduces some notations that will be used through the paper, and recalls the well-known Laplace's method. In section 3, we provide sharp asymptotic formulas for the distribution function and the distribution density of X . We also give a characterization of the Laplace transform of the limiting conditional law and discuss an application of this result to stress testing. In section 4, we apply the previous results to two examples : variance gamma and Heston models and give numerical results. Finally, in section 5 we provide a method to use the previous results for variance reduction in the Monte Carlo method and the last section of the paper contains the proof of the main result.

2 Notation and preliminaries

Notation Let us now introduce some notations that will be used throughout the paper. For two functions f and g , we denote $f(x) \sim g(x)$ as $x \rightarrow a$ whenever $\frac{f(x)}{g(x)} \rightarrow 1$ as $x \rightarrow a$. The symbol $\mathbf{1}$ stands for the n -dimensional vector with all elements equal to 1. In general, vectors will be denoted by bold-faced symbols, so that $\mathbf{w} = (w_1, \dots, w_n)^\perp$, where \perp stands for matrix transposition. For $\mathbf{v} \in \mathbb{R}_+^n$, we

denote $\mathcal{E}(\mathbf{v}) = -\sum_{i=1}^n v_i \log v_i$ with the usual convention $0 \log 0 = 0$. The elements of the covariance matrix \mathfrak{B} will be denoted by $(b_{ij})_{1 \leq i, j \leq n}$ and those of the inverse matrix \mathfrak{B}^{-1} will be denoted $(a_{ij})_{1 \leq i, j \leq n}$. Δ_n will stand for the n -dimensional simplex defined by:

$$\Delta_n = \left\{ \mathbf{w} \in \mathbb{R}^n : w_i \geq 0, i = 1, \dots, n \text{ and } \sum_{i=1}^n w_i = 1 \right\}.$$

Elementary functions are applied to vectors component-wise, so that $\log(\mathbf{w})$ stands for $(\log w_1, \dots, \log w_n)^\perp$. Similarly, inequalities involving vectors like $\mathbf{w} > 0$ are understood component-wise. For a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, we introduce the following notation :

$$\int_{\Delta_n} f(\mathbf{w}) d\mathbf{w} := \int_0^1 dw_1 \int_0^{1-w_1} dw_2 \cdots \int_0^{1-w_1-w_2-\dots-w_{n-2}} f(w_1, w_2, \dots, 1-w_1-w_2-\dots-w_{n-1}) dw_{n-1}$$

Recall that the beta function is defined in the following way:

$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt, \quad x, y > 0$$

For $\boldsymbol{\alpha} \in \mathbb{R}^n$, $n \geq 2$, with $\alpha_i > 0$, $i = 1, \dots, n$, the multidimensional Beta function can be defined as:

$$B(\boldsymbol{\alpha}) = \int_{\Delta_n} \prod_{i=1}^n w_i^{\alpha_i-1} d\mathbf{w} = \frac{\prod_{i=1}^n \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^n \alpha_i)}.$$

By convention, we set $B(\boldsymbol{\alpha}) = 1$ for all $\boldsymbol{\alpha} \in \mathbb{R}$.

Probability density function We want to characterize the asymptotic behavior of the distribution function and the density function, denoted, $p(z)$, of the positive random variable $X = \sum_{i=1}^n e^{Y_i}$ when $z \rightarrow 0$. In particular, $p(z)$ is given by :

$$p(z) = \int_0^\infty \rho(t) p_t(z) dt$$

where $p_t(z)$ is the density function of $\sum_{i=1}^n e^{\sqrt{t} Z_i + t \mu_i + \Lambda_i}$

Starting with the well-known expression for the multivariate normal density, it is easy to see that the density $p_t(z)$ is given by

$$p_t(z) = \frac{1}{z(2\pi t)^{n/2} \sqrt{\det \mathfrak{B}}} \int_{\Delta_n} \frac{d\boldsymbol{\eta}}{\eta_1 \cdots \eta_n} \times \exp \left\{ -\frac{1}{2t} (\log z \mathbf{1} + \log \boldsymbol{\eta} - \boldsymbol{\mu} t - \boldsymbol{\Lambda})^\perp \mathfrak{B}^{-1} (\log z \mathbf{1} + \log \boldsymbol{\eta} - \boldsymbol{\mu} t - \boldsymbol{\Lambda}) \right\}$$

The probability density function p of $X = \sum_{i=1}^n e^{Y_i}$ can then be computed as :

$$\begin{aligned} p(z) &= \int_0^\infty p_t(z) \rho(t) dt \\ &= \int_0^\infty \frac{\rho(t) dt}{z (2\pi t)^{\frac{n}{2}} \sqrt{\det \mathfrak{B}}} \int_{\Delta_n} \frac{d\boldsymbol{\eta}}{\eta_1 \dots \eta_n} \\ &\quad \times \exp \left\{ -\frac{1}{2t} (\log z \mathbf{1} + \log \boldsymbol{\eta} - \boldsymbol{\mu} t - \boldsymbol{\Lambda})^\perp \mathfrak{B}^{-1} (\log z \mathbf{1} + \log \boldsymbol{\eta} - \boldsymbol{\mu} t - \boldsymbol{\Lambda}) \right\} \end{aligned}$$

Laplace's method Laplace's method is used to approximate integrals of a specific form. Let us recall the formula :

Proposition 4.1.

$$I(\lambda) = \int_a^b e^{-\lambda \phi(t)} f(t) dt$$

Assume $\phi \in C^4([a, b])$, $f \in C^2([a, b])$ and the absolute minimum of ϕ on the interval $[a, b]$ is reached at the point $t = t_0$ where $a < t_0 < b$, $\phi'(t_0) = 0$ and $\phi''(t_0) > 0$. Assume also $\phi'(t) \neq 0$ on $[a, b]$ except at the point $t = t_0$. Then, as $\lambda \rightarrow \infty$,

$$I(\lambda) = \sqrt{\frac{2\pi}{\lambda \phi''(t_0)}} f(t_0) e^{-\lambda \phi(t_0)} + O\left(\frac{e^{-\lambda \phi(t_0)}}{\lambda^{\frac{3}{2}}}\right)$$

A detailed proof and more results on asymptotic methods on integrals can be found in [?].

3 Asymptotic behavior of the left tail

The present section studies the left tail asymptotic of the random variable $X = \sum_{i=1}^n e^{Y_i}$.

For $\theta > 0$, consider the following function:

$$F(t, \mathbf{w}) = \theta t + \frac{(1 + t \boldsymbol{\mu}^\perp \mathbf{w})^2}{2 \mathbf{w}^\perp \mathfrak{B} \mathbf{w} t}.$$

The following lemma, proved in [?] establishes some properties of this function.

Lemma 4.2. *There exists a unique couple $(\bar{t}, \bar{\mathbf{w}})$, with $\bar{t} \in (0, \infty)$ and $\bar{\mathbf{w}} \in \Delta_n$ such that*

$$F(\bar{t}, \bar{\mathbf{w}}) = \min_{t > 0} \max_{\mathbf{w} \in \Delta_n} F(t, \mathbf{w}).$$

In addition, the function

$$f(t) = F(t, \bar{\mathbf{w}})$$

has a unique minimum at the point \bar{t} .

We make the following assumptions:

(A1) The minimizer \bar{t} satisfies

$$\mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu}) > 0.$$

(A2) $\rho(t)$ can be written as :

$$\rho(t) = e^{-\theta t + m\sqrt{t}} \rho_0(t)$$

where $\theta > 0$, $m \in \mathbb{R}$ and ρ_0 is a regularly varying function of order α as $t \rightarrow \infty$.

For a complete study of the regular variation theory, the interested reader can see [?].

The assumption (A1) is equivalent to $\bar{\mathbf{w}} > 0$ and it is easy to see that under this assumption the quantities of interest are explicitly given by

$$\bar{t} = \sqrt{\frac{\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1}}{2\theta + \boldsymbol{\mu}^\perp \mathfrak{B}^{-1} \boldsymbol{\mu}}}, \quad \bar{\mathbf{w}} = \frac{\mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu})}{\mathbf{1}^\perp \mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu})} \quad (4.1)$$

$$F(\bar{t}, \bar{\mathbf{w}}) = \theta \bar{t} + \frac{(\mathbf{1} + \bar{t}\boldsymbol{\mu})^\perp \mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu})}{2\bar{t}} = \sqrt{\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1} (2\theta + \boldsymbol{\mu}^\perp \mathfrak{B}^{-1} \boldsymbol{\mu})} + \boldsymbol{\mu}^\perp \mathfrak{B}^{-1} \mathbf{1}. \quad (4.2)$$

The next assertion provides a sharp asymptotic formula for the density and the cumulative distribution function of the random variable X .

Theorem 4.3. *Under assumptions (A1)-(A2), $p(z)$ satisfies*

$$\begin{aligned} p(z) &\sim \frac{(2\pi)^{\frac{1-n}{2}} \bar{t}^{\alpha + \frac{3-n}{2}}}{\sqrt{\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1} \det \mathfrak{B}}} B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu})\right) \exp\left(\frac{m^2 \bar{t}^2}{8\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1}} - \frac{1}{\bar{t}} (\mathbf{1} + \bar{t}\boldsymbol{\mu})^\perp \mathfrak{B}^{-1} \boldsymbol{\Lambda}\right) \\ &\times \frac{\rho_0\left(\log \frac{1}{z}\right)}{z \left(\log \frac{1}{z}\right)^{\frac{n-1}{2}}} e^{-\log \frac{1}{z} F(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t} \log \frac{1}{z}}}. \end{aligned}$$

when $z \rightarrow 0$, and the cumulative distribution satisfies

$$\begin{aligned} F(x) &\sim \frac{(2\pi)^{\frac{1-n}{2}} \bar{t}^{\alpha + \frac{3-n}{2}}}{F(\bar{t}, \bar{\mathbf{w}}) \sqrt{\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1} \det \mathfrak{B}}} B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu})\right) \\ &\times \exp\left(\frac{m^2 \bar{t}^2}{8\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1}} - \frac{1}{\bar{t}} (\mathbf{1} + \bar{t}\boldsymbol{\mu})^\perp \mathfrak{B}^{-1} \boldsymbol{\Lambda}\right) \\ &\times \rho_0\left(\log \frac{1}{x}\right) \left(\log \frac{1}{x}\right)^{\frac{1-n}{2}} e^{-\log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t} \log \frac{1}{x}}} \end{aligned}$$

when $x \rightarrow 0$

Let us now give an interpretation of Assumption (A1) in some special cases:

Symmetric case: When $\boldsymbol{\mu} = 0$, Assumption (A1) simply requires that $\mathfrak{B}^{-1}\mathbf{1} > 0$. Denoting a_{ij} , $1 \leq i, j \leq n$ the elements of the matrix \mathfrak{B}^{-1} , the assumption leads to :

$$\sum_{j=1}^n a_{ij} > 0, \quad \forall 1 \leq i \leq n$$

In other words, the sum over the columns of the inverse of the variance-covariance matrix must be positive. This is a quite simple assumption to verify in practice. For example, when $n = 2$ and the covariance matrix has the form

$$\mathfrak{B} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

Assumption (A1) is satisfied if and only if $\rho < \frac{\min(\sigma_1, \sigma_2)}{\max(\sigma_1, \sigma_2)}$.

Non-symmetric case, $n = 2$: To check Assumption (A1) we need to compute \bar{t} and $\bar{\mathbf{w}}$ and check that $\bar{\mathbf{w}} > 0$. Let $\bar{\mathbf{w}} = (u, 1 - u)^\perp$, with $u \geq 0$. Then we need to compute, with the same parameterization for the covariance matrix as above,

$$\min_{\mathbf{w} \in \Delta_2} \left\{ \sqrt{(\boldsymbol{\mu}^\perp \mathbf{w})^2 + 2\theta \mathbf{w}^\perp \mathfrak{B} \mathbf{w}} - \boldsymbol{\mu}^\perp \mathbf{w} \right\} = \min_{u \in [0, 1]} \varphi(u),$$

with

$$\varphi(u) = \sqrt{(\mu_1 u + \mu_2(1 - u))^2 + 2\theta(\sigma_{11}^2 u^2 + 2\rho\sigma_{11}\sigma_{22}u(1 - u) + \sigma_{22}^2(1 - u)^2) - \mu_1 u - \mu_2(1 - u)}$$

Since the derivative of this function can reach zero only once on the interval $[0, 1]$, $\bar{\mathbf{w}} > 0$ if and only if : $\varphi'(0)\varphi'(1) < 0$ where

$$\begin{aligned} \varphi'(0) &= \frac{\mu_2(\mu_1 - \mu_2) + 2\theta(\rho\sigma_{11}\sigma_{22} - \sigma_{22}^2)}{\sqrt{\mu_2^2 + 2\theta\sigma_{22}^2}} + \mu_2 - \mu_1 \\ \varphi'(1) &= \frac{\mu_1(\mu_1 - \mu_2) + 2\theta(\sigma_{11}^2 - \rho\sigma_{11}\sigma_{22})}{\sqrt{\mu_1^2 + 2\theta\sigma_{11}^2}} + \mu_2 - \mu_1 \end{aligned}$$

Theorem 4.3 allows us to estimate various conditional expectations. The next assertion provides a characterization of the limiting conditional law of the Laplace transform of Y_1, \dots, Y_n , given that $X \leq x$.

Corollary 4.4. *Suppose Assumptions (A1)-(A2) hold and $\mathbf{u} \in \mathbb{R}^n$ fixed. We denote :*

$$\begin{aligned} C_{\mathbf{u}} &= \boldsymbol{\mu}^\perp \mathbf{u} + \frac{1}{2} \mathbf{u}^\perp \mathfrak{B} \mathbf{u} \\ \boldsymbol{\mu}_{\mathbf{u}} &= \boldsymbol{\mu} + \mathfrak{B} \mathbf{u} \\ \theta_{\mathbf{u}} &= \theta - C_{\mathbf{u}} \\ F_{\mathbf{u}}(t, w) &= \theta_{\mathbf{u}} t + \frac{(1 + t \boldsymbol{\mu}_{\mathbf{u}}^\perp \mathbf{w})^2}{2 \mathbf{w}^\perp \mathfrak{B} \mathbf{w} t} \end{aligned}$$

If $\theta_{\mathbf{u}} > 0$ and $\mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu}_{\mathbf{u}}) > 0$, then as $x \rightarrow 0$,

$$\begin{aligned}\mathbb{E}\left[e^{\sum_{i=1}^n u_i Y_i} | X \leq x\right] &\sim x^{1+\mathbf{u}} \frac{F(\bar{t}, \bar{\mathbf{w}})}{F(\bar{t}, \bar{\mathbf{w}}) + \mathbf{1}^{\perp} \mathbf{u}} \frac{B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu}) + \mathbf{u}\right)}{B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu})\right)} \\ \mathbb{E}\left[e^{\sum_{i=1}^n u_i Y_i} | X = x\right] &\sim x^{1+\mathbf{u}} \frac{B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu}) + \mathbf{u}\right)}{B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu})\right)}\end{aligned}$$

Application to VaR computation The Value-at-Risk (VaR) at the level $1 - y$ is a risk measure which answers the following question: given a probability y what is the threshold x such that $\mathbb{P}[X \leq x] = y$. The following result may be obtained from Theorem 4.3 by asymptotic inversion:

Corollary 4.5. *Under the assumptions of Theorem 4.3, the Value at Risk at the level $1 - y$ satisfies*

$$\text{VaR}_{1-y}(X) \sim \left(y \frac{\left(F^{-1} \log \frac{1}{y}\right)^{\frac{n-1}{2}}}{C \rho_0 \left(F^{-1} \log \frac{1}{y}\right)} \exp\left(-m \sqrt{\bar{t} F^{-1} \log \frac{1}{y} - \frac{m^2 \bar{t}}{2F}}\right) \right)^{\frac{1}{F}},$$

where we denote $F = F(\bar{t}, \bar{\mathbf{w}})$ and

$$C = \frac{(2\pi)^{\frac{1-n}{2}} \bar{t}^{\alpha + \frac{3-n}{2}}}{F \sqrt{\mathbf{1}^{\perp} \mathfrak{B}^{-1} \mathbf{1} \det \mathfrak{B}}} B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t}\boldsymbol{\mu})\right) \times \exp\left(\frac{m^2 \bar{t}^2}{8 \mathbf{1}^{\perp} \mathfrak{B}^{-1} \mathbf{1}} - \frac{1}{\bar{t}} (\mathbf{1} + \bar{t}\boldsymbol{\mu})^{\perp} \mathfrak{B}^{-1} \boldsymbol{\Lambda}\right)$$

Proof. Let $v(y) := \text{VaR}_{1-y}(X)$. By Theorem 4.3, as $y \rightarrow 0$,

$$y \sim C \rho_0 \left(\log \frac{1}{v(y)}\right) \left(\log \frac{1}{v(y)}\right)^{\frac{1-n}{2}} e^{-\log \frac{1}{v(y)} F + m \sqrt{\bar{t} \log \frac{1}{v(y)}}}.$$

Taking the logarithm,

$$\begin{aligned}\log \frac{1}{v(y)} &= \frac{1}{F} \log \frac{1}{y} + \frac{\log C}{F} + \frac{1}{F} \log \rho_0 \left(\log \frac{1}{v(y)}\right) + \frac{1-n}{2F} \log \log \frac{1}{v(y)} \\ &\quad + \frac{m}{F} \sqrt{\bar{t} \log \frac{1}{v(y)}} + o(1).\end{aligned}\quad (4.3)$$

This immediately implies that $\frac{\log \frac{1}{y}}{\log \frac{1}{v(y)}} \rightarrow F$ as $y \rightarrow 0$, and substituting this back into (4.3), we now get

$$\log \frac{1}{v(y)} = \frac{1}{F} \log \frac{1}{y} + \frac{m}{F} \sqrt{\frac{\bar{t}}{F} \log \frac{1}{y}} + O(\log \log \frac{1}{y}).$$

Now, substituting the above expression once again into (4.3), and taking into account the fact that ρ_0 is regularly varying, we get

$$\begin{aligned} \log \frac{1}{v(y)} &= \frac{1}{F} \log \frac{1}{y} + \frac{\log C}{F} + \frac{1}{F} \log \rho_0 \left(\frac{1}{F} \log \frac{1}{y} \right) \\ &\quad + \frac{1-n}{2F} \log \left(\frac{1}{F} \log \frac{1}{y} \right) + \frac{m}{F} \sqrt{\bar{t} \left(\frac{1}{F} \log \frac{1}{y} + \frac{m}{F} \sqrt{\frac{\bar{t}}{F}} \log \frac{1}{y} \right)} + o(1), \end{aligned}$$

or, expanding the square root,

$$\begin{aligned} \log \frac{1}{v(y)} &= \frac{1}{F} \left\{ \log \frac{1}{y} + \log C + \log \rho_0 \left(\frac{1}{F} \log \frac{1}{y} \right) \right. \\ &\quad \left. + \frac{1-n}{2} \log \left(\frac{1}{F} \log \frac{1}{y} \right) + m \sqrt{\frac{\bar{t}}{F}} \log \frac{1}{y} + \frac{m^2 \bar{t}}{2F} \right\} + o(1). \end{aligned}$$

□

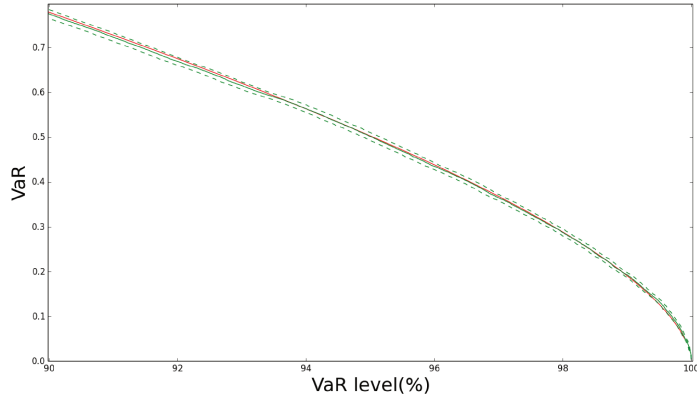


FIGURE 4.1. Value-at-Risk computed via asymptotic inversion versus Value-at-Risk computed via numerical inversion of the asymptotic formula in a two-dimensional Variance Gamma model with $\theta = \kappa = 2$, $\boldsymbol{\mu}^\perp = (0, 0)$, $\boldsymbol{\sigma}^\perp = (1.5, 1.5)$ and $\rho = 0$.

Application to stress testing Consider a benchmark portfolio (market index) containing the assets S^i , $1 \leq i \leq n$ with positive weights ξ_1, \dots, ξ_n and price process denoted by P :

$$P_t = \sum_{i=1}^n \xi_i S_t^i$$

Suppose that an investor holds a portfolio containing asset S^1, \dots, S^n with weights v_1, \dots, v_n . The value of such a portfolio is given by :

$$V_t = \sum_{i=1}^n v_i S_t^i$$

One of the most common stress scenarios is that of equity market fall of a certain magnitude. The benchmark process $\{P_t\}_{t \geq 0}$ under such a scenario is the normalized market index, having the initial value 1, and the adverse event is $\{P_t = x\}$ for some $t \geq 0$ and x which is supposed to be small. The weights ξ_i are then equal to the normalized market capitalizations of the stocks.

Our results allow us to estimate the conditional expected value under the stress scenario of individual stocks

$$e_i(t, x) = \mathbb{E} \left[S_t^i \mid \sum_{k=1}^n \xi_k S_t^k = x \right]$$

and of the entire portfolio

$$\mathbb{E} [V_t \mid P_t = x] = \sum_{i=1}^n v_i e_i(t, x)$$

The quantities $e_i(t, x)$ can be expressed directly using Corollary 4.4 for the conditional Laplace transform, since

$$e_i(t, x) = \frac{1}{\xi_i} \mathbb{E} \left[\exp \{Y^i\} \mid \sum_{k=1}^n \exp \{Y^k\} \leq x \right]$$

where $Y^i = \sqrt{\Theta} Z_i + \Theta \mu_i + \log \xi_i$, for all $1 \leq i \leq n$. Remark that it can be written as :

$$e_i(t, x) = \frac{1}{\xi_i} \mathbb{E} \left[e^{\mathbf{u}^{(i)\top} \mathbf{Y}} \mid \sum_{k=1}^n \exp \{Y^k\} \leq x \right]$$

where $\mathbf{u}^{(i)} \in \mathbb{R}^n$ and $u_j^{(i)} = \delta_{ij}$, $1 \leq j \leq n$. Applying the Corollary, we then get:

$$e_i(t, x) \sim \frac{1}{\xi_i} x \frac{F(\bar{t}, \bar{\mathbf{w}})}{F(\bar{t}, \bar{\mathbf{w}}) + \mathbf{1}^\top \mathbf{u}^{(i)}} \frac{B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t} \boldsymbol{\mu}) + \mathbf{u}^{(i)}\right)}{B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t} \boldsymbol{\mu})\right)} = x \frac{\bar{w}_i}{\xi_i},$$

where we have used the explicit formula (4.1), the definition of the multidimensional beta function and the properties of the gamma function. Therefore, the expected value of i -th stock under the stress scenario is proportional to the value x of the index, with the proportionality factor which is equal to the optimal weight \bar{w}_i . The relationship, $\sum_{i=1}^n \xi_i e_i(t, x) = x$ remains true for the asymptotic formula.

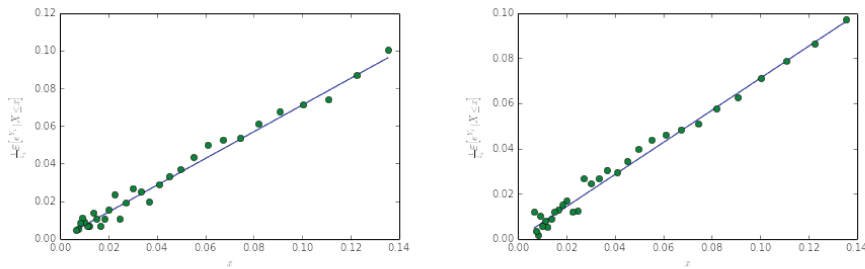


FIGURE 4.2. Two-dimensional stress tests computed via asymptotic formula (blue line) versus stress tests computed via Monte Carlo (green dots) in the variance gamma model with parameters: $\theta = 1.5$, $\kappa = 1$, $\mu^\perp = (0, 0)$, $\mathfrak{B} = Id$ and $\xi^\perp = (0.5, 0.5)$. Left : $x \mapsto \frac{1}{\xi_1} \mathbb{E} \left[e^{Y_1} \mid \sum_{k=1}^2 e^{Y_k} \leq x \right]$. Right : $x \mapsto \frac{1}{\xi_2} \mathbb{E} \left[e^{Y_2} \mid \sum_{k=1}^2 e^{Y_k} \leq x \right]$.

4 Examples

4.1 Variance Gamma process

The variance gamma random variable is defined in the following way :

$$\mathbf{Y}^{\text{VG}} = \boldsymbol{\mu} \Gamma(\theta, \kappa) + \sqrt{\Gamma(\theta, \kappa)} \mathbf{Z}, \quad \mathbf{Z} \sim N(0, \mathfrak{B}),$$

where $\Gamma(\theta, \kappa)$ denotes the gamma random variable with density

$$\rho(t) = \frac{\theta^\kappa}{\Gamma(\kappa)} t^{\kappa-1} e^{-\theta t}, \quad t \geq 0.$$

Setting

$$\rho_0(t) = \frac{\theta^\kappa}{\Gamma(\kappa)} t^{\kappa-1}$$

we clearly see that it is a regularly varying function of order $\kappa - 1$ and since $\theta > 0$, ρ satisfies Assumption (A2).

Figure 4.3 compares the Monte Carlo estimate and the asymptotic formula for the distribution function. The Monte Carlo estimates were produced with 10^6 simulations. The parameter values are given in the caption. Note that for all parameter sets Assumption (A1) is easily seen to be satisfied.

4.2 Heston model

In the Heston model [?], the stock price process S is the solution of the stochastic differential equation

$$dS_t = \mu S_t dt + \sqrt{v_t} S_t (\rho dW_t + \sqrt{1 - \rho^2} dW'_t),$$

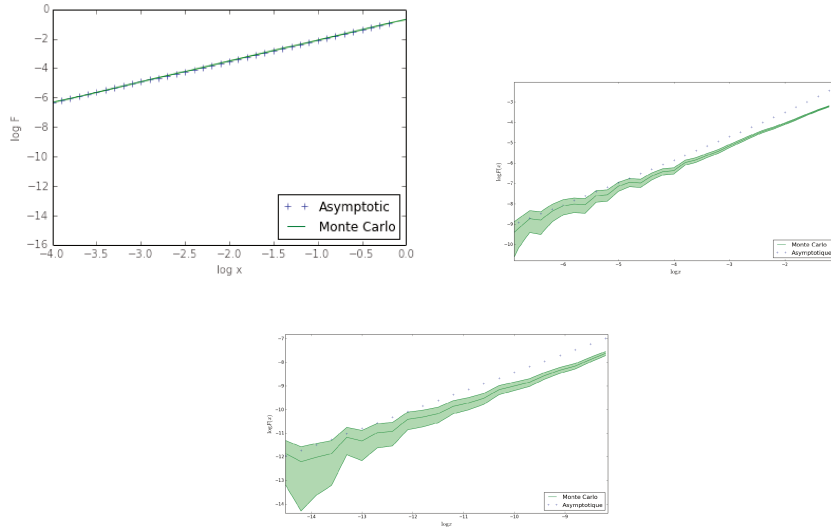


FIGURE 4.3. Variance Gamma distribution function simulated via Monte Carlo versus asymptotic formula represented in log-scale. Left : One-dimensional case with $\mu = 0$, $\sigma = 1$ and $\theta = \kappa = 1$. Right : Two-dimensional case with $\mu^\perp = (0, 0)$, covariance matrix $\mathfrak{B} = \text{Id}_2$ and $\theta = \kappa = 1$. Bottom : Two-dimensional case with $\mu^\perp = (0, 0)$, $\sigma_1 = \sigma_2 = 2$, correlation $\rho = 0.8$ and $\theta = \kappa = 1$.

or, equivalently,

$$d \log S_t = \left(\mu - \frac{v_t}{2} \right) dt + \sqrt{v_t} (\rho dW_t + \sqrt{1 - \rho^2} dW'_t),$$

where (W, W') is a standard two-dimensional Brownian motion and the instantaneous variance follows the CIR process with dynamics

$$dv_t = \kappa (\theta - v_t) dt + \delta \sqrt{v_t} dW_t.$$

Assuming zero correlation between asset returns and volatility ($\rho = 0$), the log-price in Heston's model *at a given date* t can therefore be written as

$$\log S_t = \log S_0 + \mu t - \frac{1}{2} V_t + \sqrt{V_t} Z,$$

where $Z \sim N(0, 1)$ independent from v and V_t stands for the integrated variance $V_t = \int_0^t v_s ds$.

By analogy, a multidimensional version of the Heston model for log prices at a given date t can therefore be defined by:

$$Y_t = \mathbf{\Lambda} t + \boldsymbol{\mu} V_t + \sqrt{V_t} \mathbf{Z}, \quad \mathbf{Z} \sim N(0, \mathfrak{B})$$

where $\boldsymbol{\mu}, \mathbf{\Lambda} \in \mathbb{R}^n$ and \mathbf{Z} and W are independent. Gulisashvili and Stein in [?] derive an asymptotic formula for the probability density function of V_t . At time t , it satisfies :

$$\rho(x) \sim e^{-\frac{C}{t}x} e^{\sqrt{\frac{B}{t}x}} \frac{A}{2} t^{-\frac{1}{4} - \frac{\kappa\theta}{\delta^2}} x^{-\frac{3}{4} + \frac{\kappa\theta}{\delta^2}}, \quad x \rightarrow \infty$$

for some constants A , B and C which are expressed in terms of the model parameters. So if we denote, $\lambda = \frac{C}{t}$, $m = \frac{B}{t}$ and $\rho_0(x) = \frac{A}{2} t^{-\frac{1}{4} - \frac{\kappa\theta}{\delta^2}} x^{-\frac{3}{4} + \frac{\kappa\theta}{\delta^2}}$, it is easy to see that ρ_0 is a regularly varying function of order $\alpha = -\frac{3}{4} + \frac{\kappa\theta}{\delta^2}$ and $\rho(x) = e^{-\lambda x + m\sqrt{x}} \rho_0(x)$ which is the desired form. Thus, we can apply Theorem 4.3 to calculate the cumulative distribution function.

Figure 4.4 compares the Monte Carlo estimate and the asymptotic formula for the distribution function. The Monte Carlo estimates were produced with 10^6 simulations. We have simulated the Heston model at time $t = 1$ with initial volatility $v_0 = 0.2$, mean-reversion speed $\kappa = 0.9$, long-term volatility $\theta = 1$ and volatility of volatility $\delta = 2$.

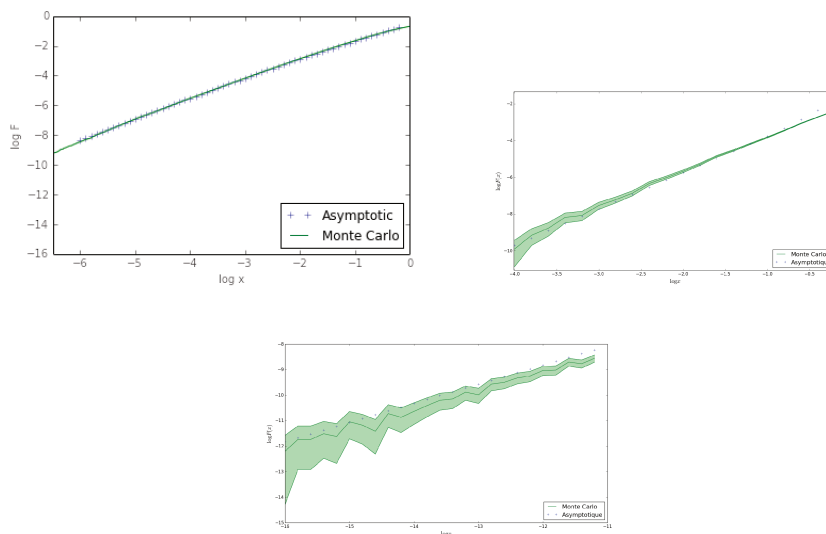


FIGURE 4.4. Cumulative distribution function of the portfolio in the Heston model simulated via Monte Carlo versus asymptotic formula represented in log-scale. Left : one-dimensional case with $\mu = 0$, $\Lambda = 0$, $\sigma = 1$. Right : two-dimensional case with $\mu^\perp = (0, 0)$, $\Lambda^\perp = (0, 0)$ and covariance matrix $\mathfrak{B} = \text{Id}_2$. Bottom : two-dimensional case with $\mu^\perp = (0, 0)$, $\Lambda^\perp = (0, 0)$, $\sigma_1 = \sigma_2 = 1$ and correlation $\rho = 0.8$.

5 Variance Reduction of Monte Carlo estimates

As we have seen from the numerical examples, the asymptotic formulas in Theorem 4.3 typically provide good order-of-magnitude approximations of the distribution function of the sum of exponentials of Gaussian mean-variance mixture random variables. When a more precise estimate is needed, and the dimension n is large, one can use a Monte Carlo estimator. In such a case, as we will next explain, the

asymptotic formulas can be used to construct very efficient variance reduction procedures. To save space, we will only discuss the case of distribution function. Similar ideas can be used to reduce the variance of Monte Carlo estimates of densities, conditional expectations or other quantities of interest.

The following lemmas will be useful to determine the distribution of \mathbf{Y} under a change of probability :

Lemma 4.6. Fix $\mathbf{u} \in \mathbb{R}^n$ and define a new probability $\tilde{\mathbb{P}}$ by :

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} = \frac{e^{\mathbf{u}^\top \mathbf{Y}}}{M_Y(\mathbf{u})}$$

Under this new probability, \mathbf{Y} can be written as :

$$\mathbf{Y} = \sqrt{\Theta_{\mathbf{u}}}\mathbf{Z} + \Theta_{\mathbf{u}}\boldsymbol{\mu}_{\mathbf{u}} + \boldsymbol{\Lambda}$$

where $\mathbf{Z} \sim N(0, \mathfrak{B})$,

$$\boldsymbol{\mu}_{\mathbf{u}} = \boldsymbol{\mu} + \mathfrak{B}\mathbf{u}$$

and $\Theta_{\mathbf{u}}$ has the following probability density:

$$\rho_{\mathbf{u}}(x) = \frac{e^{C_{\mathbf{u}}x}\rho(x)}{\int_0^\infty e^{C_{\mathbf{u}}y}\rho(y)dy}$$

where

$$C_{\mathbf{u}} = \boldsymbol{\mu}^\top \mathbf{u} + \frac{1}{2}\mathbf{u}^\top \mathfrak{B}\mathbf{u}.$$

For the distribution function $F(x) = \mathbb{P}[X \leq x]$, the standard Monte Carlo estimate writes :

$$\hat{F}_N(x) = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\sum_{i=1}^n \exp\{Y_i^{(k)}\} \leq x} \quad (4.4)$$

where $\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(N)}$ are i.i.d. vectors with the Gaussian mean-variance mixture distribution. However, this estimate is not a suitable approximation of the tail of the distribution function. Indeed, the variance of \hat{F}_N is given by

$$\text{Var}\hat{F}_N(x) = \frac{F(x) - F^2(x)}{N} \sim \frac{F(x)}{N}, \quad x \rightarrow 0$$

and the relative error, that is,

$$\frac{\sqrt{\text{Var}\hat{F}_N(x)}}{F(x)} \sim \frac{1}{\sqrt{NF(x)}},$$

explodes very quickly as $x \rightarrow 0$. The usual way to reduce variance is to use importance sampling :

$$F(x) = \mathbb{E} \left[\mathbf{1}_{\sum_{i=1}^n e^{Y_i} \leq x} \right] = \tilde{\mathbb{E}} \left[M_Y(\mathbf{u}) e^{-\mathbf{u}^\top \mathbf{Y}} \mathbf{1}_{\sum_{i=1}^n e^{Y_i} \leq x} \right]$$

where

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} = \frac{e^{\mathbf{u}^\perp \mathbf{Y}}}{M_Y(\mathbf{u})}$$

where $\mathbf{u} \in \mathbb{R}^n$ is a vector that will be chosen later. The goal is to find a non-zero \mathbf{u} such that the corresponding estimate

$$\hat{F}_N^{\mathbf{u}}(x) = \frac{M_Y(\mathbf{u})}{N} \sum_{k=1}^N \exp \left\{ -\mathbf{u}^\perp \tilde{\mathbf{Y}}^{(k)} \right\} \mathbf{1}_{\sum_{i=1}^n e^{\tilde{Y}_i^{(k)}} \leq x} \quad (4.5)$$

has a smaller variance than that of the standard estimate, where $\tilde{\mathbf{Y}}^{(1)}, \dots, \tilde{\mathbf{Y}}^{(N)}$ are i.i.d. random vectors which have the same distribution as \mathbf{Y} under $\tilde{\mathbb{P}}$, given by Lemma 4.6. Simple computations show that the variance of $F_N^{\mathbf{u}}(x)$ is given by:

$$\begin{aligned} \text{Var} \hat{F}_N^{\mathbf{u}}(x) &= \frac{1}{N} \left\{ \tilde{\mathbb{E}} \left[M_Y^2(\mathbf{u}) e^{-2\mathbf{u}^\perp \mathbf{Y}} \mathbf{1}_{\sum_{i=1}^n e^{Y_i} \leq x} \right] - F^2(x) \right\} \\ &= \frac{1}{N} \left\{ \mathbb{E} \left[M_Y(\mathbf{u}) e^{-\mathbf{u}^\perp \mathbf{Y}} \mathbf{1}_{\sum_{i=1}^n e^{Y_i} \leq x} \right] - F^2(x) \right\} \\ &= \frac{1}{N} \left\{ F(x) M_Y(\mathbf{u}) \mathbb{E} \left[e^{-\mathbf{u}^\perp \mathbf{Y}} | X \leq x \right] - F^2(x) \right\} \end{aligned}$$

Let

$$V(\mathbf{u}, x) = M_Y(\mathbf{u}) \mathbb{E} \left[e^{-\mathbf{u}^\perp \mathbf{Y}} | X \leq x \right]$$

Since $F(x)$ does not depend on \mathbf{u} , the optimal variance reduction is obtained by minimizing $V(\mathbf{u}, x)$ as a function of \mathbf{u} . Our idea is to obtain an explicit estimate by replacing the probability in the previous expression by an asymptotically equivalent expression given by Corollary 4.4. Provided that Assumptions (A1) and (A2) are satisfied, and in addition

$$\theta > |\boldsymbol{\mu}^\perp \mathbf{u}| + \frac{1}{2} \mathbf{u}^\perp \mathfrak{B} \mathbf{u}, \quad \text{and} \quad \mathfrak{B}^{-1}(\mathbf{1} + \bar{t} \boldsymbol{\mu}) - \bar{t} \mathbf{u} > 0,$$

we have, as $x \rightarrow 0$,

$$V(\mathbf{u}, x) \sim M_Y(\mathbf{u}) x^{-\mathbf{1}^\perp \mathbf{u}} \frac{F(\bar{t}, \bar{\mathbf{w}})}{F(\bar{t}, \bar{\mathbf{w}}) - \mathbf{1}^\perp \mathbf{u}} \frac{B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t} \boldsymbol{\mu}) - \mathbf{u}\right)}{B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t} \boldsymbol{\mu})\right)}$$

In other words, to optimize the variance, we need to minimize

$$\tilde{V}(\mathbf{u}, x) = e^{\boldsymbol{\Lambda}^\perp \mathbf{u}} M_\rho \left(\boldsymbol{\mu}^\perp \mathbf{u} + \frac{1}{2} \mathbf{u}^\perp \mathfrak{B} \mathbf{u} \right) x^{-\mathbf{1}^\perp \mathbf{u}} \frac{B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1}(\mathbf{1} + \bar{t} \boldsymbol{\mu}) - \mathbf{u}\right)}{F(\bar{t}, \bar{\mathbf{w}}) - \mathbf{1}^\perp \mathbf{u}}$$

as function of \mathbf{u} . This minimization is usually carried out numerically.

Example In the variance gamma model (with $\kappa = \theta$ to avoid redundancy),

$$M_\rho(z) = \left(1 - \frac{z}{\theta}\right)^{-\theta}.$$

In the one-dimensional case, we then simply have

$$\tilde{V}(u, x) = e^{\Lambda u} \left(1 - \frac{\mu}{\theta}u - \frac{\sigma^2 u^2}{2\theta}\right)^{-\theta} x^{-u} \frac{1}{c^* - u} = C \frac{e^{\Lambda u} x^{-u}}{(c^* - u)(c^* + u)^\theta (c_* - u)^\theta},$$

where we have written $c^* = F(\bar{t}, 1) = \frac{\sqrt{2\theta\sigma^2 + \mu^2 + \mu}}{\sigma^2}$ and $c_* = \frac{\sqrt{2\theta\sigma^2 + \mu^2 - \mu}}{\sigma^2}$ and C is a constant. Equivalently, we need to minimize

$$(\Lambda - \log x)u - \log(c^* - u) - \theta \log(c^* + u) - \theta \log(c_* - u).$$

The first order condition for this minimization problem leads to a third degree equation, but when $\mu = 0$, the solution is given explicitly by

$$u^* = \frac{\theta + \frac{1}{2} - \sqrt{\left(\theta + \frac{1}{2}\right)^2 + (\Lambda - \log x)^2 c^2 - (\Lambda - \log x)c}}{\Lambda - \log x},$$

where we write $c = c^* = c_* = \frac{\sqrt{2\theta}}{\sigma}$.

To test the performance of the proposed variance reduction algorithm, we have computed the Monte Carlo estimates with and without variance reduction for different levels x :

- in the two-dimensional Variance Gamma model at one-year horizon with $\mathfrak{B} = Id_2$, $\kappa = \theta = 1$ and $\boldsymbol{\mu} = \boldsymbol{\Lambda} = (0, 0)^\perp$.
- in the two-dimensional Variance Gamma model at one-year horizon with $\sigma_1 = \sigma_2 = 1$, $\rho = 0.8$, $\kappa = \theta = 1$ and $\boldsymbol{\mu} = \boldsymbol{\Lambda} = (0, 0)^\perp$.
- in the two-dimensional Heston model at one-year horizon with $\mathfrak{B} = Id_2$, $\kappa = 3$, $\theta = 0.3$, $\delta = 0.1$, $\boldsymbol{\mu} = (0, 0)^\perp$ and $\boldsymbol{\Lambda} = (0, 0)^\perp$ (see Table 4.3).
- in the two-dimensional Heston model at one-year horizon with $\sigma_1 = \sigma_2 = 1$, $\rho = 0.8$, $\kappa = 3$, $\theta = 0.3$, $\delta = 0.1$, $\boldsymbol{\mu} = (0, 0)^\perp$ and $\boldsymbol{\Lambda} = (0, 0)^\perp$ (see Table 4.4).

Tables 4.1, 4.1 4.3 and 4.4 show the ratio of the standard deviation of the estimate (4.4) to that of the estimate (4.5), together with the optimal value u^* . The reduction factors are greater than one for all values of x and in general quite significant even when the probability of interest is not so small. The method therefore appears quite promising for risk management applications.

x	$\mathbb{P}[X \leq x]$	red. factor	u^*
0.2231	0.003340	28.36	$(-0.827, -0.827)^\perp$
0.2465	0.00414	26.05	$(-0.822, -0.822)^\perp$
0.2725	0.00515	21.31	$(-0.814, -0.814)^\perp$
0.3012	0.00624	19.17	$(-0.807, -0.807)^\perp$
0.3328	0.00782	15.17	$(-0.80, -0.80)^\perp$
0.3678	0.0097	13.03	$(-0.792, -0.792)^\perp$

TABLE 4.1. Standard deviation reduction factors obtained with the variance reduction estimate (4.5).

x	$\mathbb{P}[X \leq x]$	red. factor	u^*
0.2231	0.0166	11.31	$(-0.587, -0.587)^\perp$
0.2465	0.01943	10.10	$(-0.582, -0.582)^\perp$
0.2725	0.02262	8.76	$(-0.577, -0.577)^\perp$
0.3012	0.026244	7.95	$(-0.571, -0.571)^\perp$
0.3328	0.03088	6.93	$(-0.565, -0.565)^\perp$
0.3678	0.03613	6.28	$(-0.559, -0.559)^\perp$

TABLE 4.2. Standard deviation reduction factors obtained with the variance reduction estimate (4.5).

x	$\mathbb{P}[X \leq x]$	red. factor	u^*
0.3679	$4.02 \cdot 10^{-6}$	88649.97	$(-5.76, -5.76)^\perp$
0.4493	$3.07 \cdot 10^{-5}$	1174.16	$(-4.92, -4.92)^\perp$
0.5488	0.00028	636.47	$(-4.58, -4.58)^\perp$
0.6703	0.0016	153.01	$(-3.89, -3.89)^\perp$
0.8187	0.0074	39.33	$(-3.05, -3.05)^\perp$
1.2214	0.0779	5.9	$(-2.44, -2.44)^\perp$

TABLE 4.3. Standard deviation reduction factors obtained with the variance reduction estimate (4.5).

6 Proof of Theorem 4.3

The proof is based on the following estimate for the density $p_t(z)$, which was defined in Section 2.

Lemma 4.7. *For all $\mathbf{w} \in \Delta_n$, the density p_t admits the following upper bound:*

$$p_t(z) \leq \frac{c_q}{z\sqrt{t}} N \left(\frac{\log z - t\boldsymbol{\mu}^\perp \mathbf{w} - \boldsymbol{\Lambda}^\perp \mathbf{w}}{\sqrt{t\mathbf{w}^\perp \boldsymbol{\mathcal{B}} \mathbf{w}}} \right)^{\frac{1}{q}}.$$

x	$\mathbb{P}[X \leq x]$	red. factor	u^*
0.3011	0.00012	1616.77	$(-3.5, -3.5)^\perp$
0.3678	0.00052	530.41	$(-3.11, -3.11)^\perp$
0.4493	0.0018	128.54	$(-2.77, -2.77)^\perp$
0.5488	0.0058	54.65	$(-2.42, -2.42)^\perp$
0.6703	0.016	23.5	$(-2.17, -2.17)^\perp$
0.8187	0.0403	11.86	$(-1.79, -1.79)^\perp$

TABLE 4.4. Standard deviation reduction factors obtained with the variance reduction estimate (4.5).

for every $q > 1$ and for an appropriate constant c_q , where N is the standard normal distribution function.

Proof. The cumulative distribution function F_t associated with the probability density p_t satisfies

$$\begin{aligned}
F_t(z) &= \int_{\{e^{x_1} + \dots + e^{x_n} \leq z\}} \frac{dx_1 \dots dx_n}{(2\pi t)^{\frac{n}{2}} \sqrt{\det \mathfrak{B}}} \exp \left\{ -\frac{1}{2t} (\mathbf{x} - t\boldsymbol{\mu} - \boldsymbol{\Lambda})^\perp \mathfrak{B}^{-1} (\mathbf{x} - t\boldsymbol{\mu} - \boldsymbol{\Lambda}) \right\} \\
&= \int_{-\infty}^{\log z} dx_1 \dots \int_{-\infty}^{\log z} dx_{n-1} \int_{-\infty}^{\log(z - \sum_{i=1}^{n-1} e^{x_i})} \frac{dx_n}{(2\pi t)^{\frac{n}{2}} \sqrt{\det \mathfrak{B}}} \\
&\quad \times \exp \left\{ -\frac{1}{2t} (\mathbf{x} - t\boldsymbol{\mu} - \boldsymbol{\Lambda})^\perp \mathfrak{B}^{-1} (\mathbf{x} - t\boldsymbol{\mu} - \boldsymbol{\Lambda}) \right\} \\
&= \int_{-\infty}^0 \dots \int_{-\infty}^0 \int_{-\infty}^{\log(1 - \sum_{i=1}^{n-1} e^{y_i})} \frac{dy_1 \dots dy_n}{(2\pi t)^{\frac{n}{2}} \sqrt{\det \mathfrak{B}}} \\
&\quad \times \exp \left\{ -\frac{1}{2t} (\mathbf{y} + \log z \mathbf{1} - t\boldsymbol{\mu} - \boldsymbol{\Lambda})^\perp \mathfrak{B}^{-1} (\mathbf{y} + \log z \mathbf{1} - t\boldsymbol{\mu} - \boldsymbol{\Lambda}) \right\}.
\end{aligned}$$

Since $p_t(z) = \frac{\partial}{\partial z} F_t(z)$,

$$\begin{aligned}
p_t(z) &= - \int_{-\infty}^0 \dots \int_{-\infty}^0 \int_{-\infty}^{\log(1 - \sum_{i=1}^{n-1} e^{y_i})} \frac{dy_1 \dots dy_n}{(2\pi t)^{\frac{n}{2}} \sqrt{\det \mathfrak{B}}} \frac{1}{tz} (\mathbf{y} + \log z \mathbf{1} - t\boldsymbol{\mu} - \boldsymbol{\Lambda})^\perp \mathfrak{B}^{-1} \mathbf{1} \\
&\quad \times \exp \left\{ -\frac{1}{2t} (\mathbf{y} + \log z \mathbf{1} - t\boldsymbol{\mu} - \boldsymbol{\Lambda})^\perp \mathfrak{B}^{-1} (\mathbf{y} + \log z \mathbf{1} - t\boldsymbol{\mu} - \boldsymbol{\Lambda}) \right\}
\end{aligned}$$

It can be rewritten as:

$$p_t(z) = -\frac{1}{z\sqrt{t}} \mathbb{E} \left[\mathbf{1}_{\{\sum_{i=1}^n e^{\sqrt{t}Z_i + t\mu_i + \Lambda_i} \leq z\}} \mathbf{Z}^\perp \mathfrak{B}^{-1} \mathbf{1} \right]$$

Let $p, q \in \mathbb{R}^+$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Applying Holder inequality leads to:

$$\begin{aligned} p_t(z) &\leq \frac{1}{z\sqrt{t}} \mathbb{E} \left[\mathbb{1}_{\{\sum_{i=1}^n e^{\sqrt{t}Z_i + t\mu_i + \Lambda_i} \leq z\}} \right]^{\frac{1}{q}} \mathbb{E} \left[|\mathbf{Z}^\perp \mathfrak{B}^{-1} \mathbf{1}|^p \right]^{\frac{1}{p}} \\ &= \frac{1}{z\sqrt{t}} F_t(z)^{\frac{1}{q}} \mathbb{E} \left[|\mathbf{Z}^\perp \mathfrak{B}^{-1} \mathbf{1}|^p \right]^{\frac{1}{p}}. \end{aligned}$$

On the other hand, for every $\mathbf{w} \in \Delta_n$ with $\mathbf{w} > 0$, the distribution function $F_t(z)$ can be bounded from above using Jensen's inequality:

$$\begin{aligned} F_t(z) &= \mathbb{P} \left[\sum_{i=1}^n e^{\sqrt{t}Z_i + t\mu_i + \Lambda_i} \leq z \right] = \mathbb{P} \left[\sum_{i=1}^n w_i e^{\sqrt{t}Z_i + t\mu_i + \Lambda_i - \log w_i} \leq z \right] \\ &\leq \mathbb{P} \left[\exp \left\{ \sum_{i=1}^n w_i (\sqrt{t}Z_i + t\mu_i + \Lambda_i) + \mathcal{E}(\mathbf{w}) \right\} \leq z \right] \\ &= \mathbb{P} \left[\sum_{i=1}^n w_i (\sqrt{t}Z_i + t\mu_i + \Lambda_i) + \mathcal{E}(\mathbf{w}) \leq \log z \right] = N \left(\frac{\log z - \mathcal{E}(\mathbf{w}) - t\boldsymbol{\mu}^\perp \mathbf{w} - \boldsymbol{\Lambda}^\perp \mathbf{w}}{\sqrt{t\mathbf{w}^\perp \mathfrak{B} \mathbf{w}}} \right). \end{aligned}$$

By continuity, this estimate holds for all $\mathbf{w} \in \Delta_n$. We conclude that for every $q > 1$,

$$p_t(z) \leq \frac{1}{z\sqrt{t}} \mathbb{E} \left[|\mathbf{Z}^\perp \mathfrak{B}^{-1} \mathbf{1}|^{\frac{q}{q-1}} \right]^{1-\frac{1}{q}} N \left(\frac{\log z - \mathcal{E}(\mathbf{w}) - t\boldsymbol{\mu}^\perp \mathbf{w} - \boldsymbol{\Lambda}^\perp \mathbf{w}}{\sqrt{t\mathbf{w}^\perp \mathfrak{B} \mathbf{w}}} \right)^{\frac{1}{q}}.$$

Since N is increasing and $\mathcal{E}(\mathbf{w}) \geq 0$, the result follows. \square

Estimation of the density $p(z)$

Without loss of generality, let $z \in (0, 1)$. The density $p(z)$ is given by

$$\begin{aligned} p(z) &= \int_0^\infty \rho(s) p_s(z) ds = \int_0^\infty e^{-\theta s + m\sqrt{s}} \rho_0(s) p_s(z) ds \\ &= \log \frac{1}{z} \int_0^\infty e^{-\theta t \log \frac{1}{z} + m\sqrt{t \log \frac{1}{z}}} \rho_0 \left(t \log \frac{1}{z} \right) p_{t \log \frac{1}{z}}(z) dt. \end{aligned} \quad (4.6)$$

Remark that necessarily,

$$1 + \bar{t} \boldsymbol{\mu}^\perp \bar{\mathbf{w}} > 0.$$

Indeed, if $1 + \bar{t} \boldsymbol{\mu}^\perp \bar{\mathbf{w}} < 0$ then $F(-\frac{1}{\boldsymbol{\mu}^\perp \bar{\mathbf{w}}}, \bar{\mathbf{w}}) < F(\bar{t}, \bar{\mathbf{w}})$ which contradicts the fact that \bar{t} is the minimizer. If $1 + \bar{t} \boldsymbol{\mu}^\perp \bar{\mathbf{w}} = 0$ then $F'_t(\bar{t}, \bar{\mathbf{w}}) = \theta$ which also leads to a contradiction. Therefore, we can introduce

$$\bar{T} = \begin{cases} -\frac{1}{\boldsymbol{\mu}^\perp \bar{\mathbf{w}}}, & \boldsymbol{\mu}^\perp \bar{\mathbf{w}} < 0 \\ +\infty & \text{otherwise,} \end{cases}$$

and choose ε such that $\bar{t} - 2\varepsilon > 0$, $\bar{t} + 2\varepsilon < \bar{T}$ and $\mathfrak{B}^{-1}(\mathbf{1} + t\boldsymbol{\mu}) > 0$ for all $t \in [\bar{t} - 2\varepsilon, \bar{t} + 2\varepsilon]$. We split the interval $[0, \infty)$ into four subintervals:

$$I_1 = [0, \bar{t} - \varepsilon), \quad I_2 = [\bar{t} - \varepsilon, \bar{t} + \varepsilon), \quad I_3 = [\bar{t} + \varepsilon, \bar{T}), \quad I_4 = [\bar{T}, +\infty),$$

and denote for convenience

$$J_k = \int_{I_k} e^{-\theta t \log \frac{1}{z} + m \sqrt{t \log \frac{1}{z}}} \rho_0 \left(t \log \frac{1}{z} \right) p_{t \log \frac{1}{z}}(z) dt$$

Remark also for later use that by Lemma 4.7,

$$p_{t \log \frac{1}{z}}(z) \leq \frac{1}{z \sqrt{t \log \frac{1}{z}}} c_q N \left(-\sqrt{\log \frac{1}{z}} \frac{1 + t \boldsymbol{\mu}^\perp \bar{\boldsymbol{w}} + \boldsymbol{\Lambda}^\perp \boldsymbol{w} \log^{-1} \frac{1}{z}}{\sqrt{\bar{\boldsymbol{w}}^\perp \mathfrak{B} \bar{\boldsymbol{w}} t}} \right)^{1/q}.$$

Estimation of J_2 The interval I_2 gives the main contribution to the asymptotics. It will be estimated using Laplace's method. The density function $p_{t \log \frac{1}{z}}$ can be written as:

$$p_{t \log \frac{1}{z}}(z) = \frac{e^{-\frac{1}{t} \boldsymbol{\Lambda}^\perp \mathfrak{B}^{-1} (\mathbf{1} + t \boldsymbol{\mu})}}{z (2\pi t \log \frac{1}{z})^{\frac{n}{2}} \sqrt{\det \mathfrak{B}}} e^{-h(t) \log \frac{1}{z}} g(t, z)$$

where

$$h(t) = \frac{1}{2t} (\mathbf{1} + t \boldsymbol{\mu})^\perp \mathfrak{B}^{-1} (\mathbf{1} + t \boldsymbol{\mu})$$

and

$$\begin{aligned} g(t, z) &= \int_{\Delta_n} \frac{d\boldsymbol{w}}{w_1 \dots w_n} \exp \left\{ \frac{1}{t} \log \boldsymbol{w}^\perp \mathfrak{B}^{-1} (\mathbf{1} + t \boldsymbol{\mu}) - \frac{1}{2t \log \frac{1}{z}} (\log \boldsymbol{w} - \boldsymbol{\Lambda})^\perp \mathfrak{B}^{-1} (\log \boldsymbol{w} - \boldsymbol{\Lambda}) \right\} \\ &= \int_{\Delta_n} d\boldsymbol{w} \prod_{i=1}^n w_i^{\frac{1}{t} (\mathfrak{B}^{-1} (\mathbf{1} + t \boldsymbol{\mu}))_i - 1} \exp \left\{ -\frac{1}{2t \log \frac{1}{z}} (\log \boldsymbol{w} - \boldsymbol{\Lambda})^\perp \mathfrak{B}^{-1} (\log \boldsymbol{w} - \boldsymbol{\Lambda}) \right\}. \end{aligned}$$

Since for $t \in I_2$, $(\mathfrak{B}^{-1} (\mathbf{1} + t \boldsymbol{\mu})) > 0$, applying the dominated convergence theorem we conclude that,

$$g(t, z) \xrightarrow{z \rightarrow 0} B \left(\frac{1}{t} \mathfrak{B}^{-1} (\mathbf{1} + t \boldsymbol{\mu}) \right)$$

for all $t \in I_2$. Moreover since $g(t, z)$ is decreasing in z , by Dini's theorem the convergence is uniform on t .

We are now going to establish an upper bound on J_2 . Fix $\bar{\varepsilon} > 0$. Then, for all $t \in I_2$ and for sufficiently small values of z ,

$$p_{t \log \frac{1}{z}}(z) \leq \frac{1}{z (2\pi t \log \frac{1}{z})^{\frac{n}{2}} \sqrt{\det \mathfrak{B}}} e^{-\log \frac{1}{z} \frac{1}{2t} (\mathbf{1} + t \boldsymbol{\mu})^\perp \mathfrak{B}^{-1} (\mathbf{1} + t \boldsymbol{\mu})} B \left(\frac{1}{t} \mathfrak{B}^{-1} (\mathbf{1} + t \boldsymbol{\mu}) \right) (1 + \bar{\varepsilon}).$$

Since ρ_0 is regularly varying, we deduce that for z sufficiently small and all $t \in [\bar{t} - \varepsilon, \bar{t} + \varepsilon]$,

$$\rho_0 \left(t \log \frac{1}{z} \right) \leq (1 + \bar{\varepsilon}) t^\alpha \rho_0 \left(\log \frac{1}{z} \right).$$

Thus, for sufficiently small z ,

$$J_2 \leq \frac{(1 + \bar{\varepsilon})^2 \rho_0 \left(\log \frac{1}{z} \right)}{z (2\pi \log \frac{1}{z})^{\frac{n}{2}} \sqrt{\det \mathfrak{B}}} \int_{\bar{t} - \varepsilon}^{\bar{t} + \varepsilon} e^{-\log \frac{1}{z} F(t, \bar{\boldsymbol{w}}) + m \sqrt{t \log \frac{1}{z}} t^{\alpha - \frac{n}{2}}} B \left(\frac{1}{t} \mathfrak{B}^{-1} (\mathbf{1} + t \boldsymbol{\mu}) \right) dt. \quad (4.7)$$

Fix $\delta > 0$. To fix the ideas, assume also that $m \geq 0$ (the case $m < 0$ can be treated in a similar fashion). Since the functions $t \mapsto F(t, \bar{\mathbf{w}})$ and $t \mapsto \sqrt{t}$ are smooth at \bar{t} , and $\frac{\partial F(t, \bar{\mathbf{w}})}{\partial t}|_{t=\bar{t}} = 0$, we may assume, possibly by reducing the value of ε , that for all $t \in [\bar{t} - \varepsilon, \bar{t} + \varepsilon]$

$$F(t, \bar{\mathbf{w}}) \geq F(\bar{t}, \bar{\mathbf{w}}) + \frac{(t - \bar{t})^2}{2} \frac{\partial^2}{\partial t^2} F(t, \bar{\mathbf{w}})|_{t=\bar{t}} (1 - \delta) \quad (4.8)$$

$$m \sqrt{t \log \frac{1}{z}} \leq m \sqrt{\bar{t} \log \frac{1}{z}} \left(1 + \frac{t - \bar{t}}{2\bar{t}} - \frac{(t - \bar{t})^2}{8\bar{t}^2} (1 - \delta) \right) \quad (4.9)$$

and

$$\frac{1}{t} \geq \frac{1}{\bar{t}} - \frac{t - \bar{t}}{\bar{t}^2} + \frac{(t - \bar{t})^2}{8\bar{t}^3} (1 - \delta) \quad (4.10)$$

where we remark that

$$\frac{\partial^2}{\partial t^2} F(t, \bar{\mathbf{w}})|_{t=\bar{t}} = \frac{\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1}}{\bar{t}^3}$$

To simplify the notation, let us denote :

$$\begin{aligned} A(\bar{t}, z) &= \log \frac{1}{z} \frac{\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1}}{2\bar{t}^3} (1 - \delta) + \frac{m \sqrt{\log \frac{1}{z}}}{8\bar{t}^{\frac{3}{2}}} (1 - \delta) + \frac{\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1}}{\bar{t}} (1 - \delta) \\ B(\bar{t}, z) &= \frac{m \sqrt{\log \frac{1}{z}}}{2\sqrt{\bar{t}}} + \frac{\mathbf{1}^\perp \mathfrak{B}^{-1} \Lambda}{\bar{t}^2} \\ g(t) &= t^{\alpha - \frac{n}{2}} B \left(\frac{1}{t} \mathfrak{B}^{-1} (\mathbf{1} + t\boldsymbol{\mu}) \right) \end{aligned}$$

Substituting the estimates (4.8), (4.9) and (4.10) into inequality (4.7), we then get

$$\begin{aligned} J_2 \leq & \frac{(1 + \bar{\varepsilon})^2 \rho_0 \left(\log \frac{1}{z} \right)}{z (2\pi \log \frac{1}{z})^{\frac{n}{2}} \sqrt{\det \mathfrak{B}}} e^{-\log \frac{1}{z} F(\bar{t}, \bar{\mathbf{w}}) + m \sqrt{\bar{t} \log \frac{1}{z}} - \frac{1}{\bar{t}} (\mathbf{1} + \bar{t}\boldsymbol{\mu})^\perp \mathfrak{B}^{-1} \Lambda + \frac{B^2(\bar{t}, z)}{4A(\bar{t}, z)}} \\ & \times \int_{-\varepsilon}^{\varepsilon} e^{-A(\bar{t}, z) \left(s - \frac{B(\bar{t}, z)}{2A(\bar{t}, z)} \right)^2} g(s + \bar{t}) dt \end{aligned}$$

Now observe that on the one hand,

$$\lim_{z \rightarrow 0} \frac{B^2(\bar{t}, z)}{4A(\bar{t}, z)} = \frac{m^2 \bar{t}^2}{8 \mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1}^\perp (1 - \delta)},$$

so that for z sufficiently small,

$$\begin{aligned} J_2 \leq & \frac{(1 + \bar{\varepsilon})^3 \rho_0 \left(\log \frac{1}{z} \right)}{z (2\pi \log \frac{1}{z})^{\frac{n}{2}} \sqrt{\det \mathfrak{B}}} e^{-\log \frac{1}{z} F(\bar{t}, \bar{\mathbf{w}}) + m \sqrt{\bar{t} \log \frac{1}{z}} - \frac{1}{\bar{t}} (\mathbf{1} + \bar{t}\boldsymbol{\mu})^\perp \mathfrak{B}^{-1} \Lambda + \frac{m^2 \bar{t}^2}{8 \mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1}^\perp (1 - \delta)}} \\ & \times \int_{-\varepsilon}^{\varepsilon} e^{-A(\bar{t}, z) \left(s - \frac{B(\bar{t}, z)}{2A(\bar{t}, z)} \right)^2} g(s + \bar{t}) dt \end{aligned}$$

On the other hand, since $\frac{B(\bar{t}, z)}{2A(\bar{t}, z)} \rightarrow 0$, when $z \rightarrow 0$ and the function g is uniformly continuous, for z sufficiently small,

$$\begin{aligned} \int_{-\varepsilon}^{\varepsilon} e^{-A(\bar{t}, z)(s - \frac{B(\bar{t}, z)}{2A(\bar{t}, z)})^2} g(s + \bar{t}) dt &= \int_{-\varepsilon - \frac{B(\bar{t}, z)}{2A(\bar{t}, z)}}^{\varepsilon - \frac{B(\bar{t}, z)}{2A(\bar{t}, z)}} e^{-A(\bar{t}, z)s^2} g\left(s + \frac{B(\bar{t}, z)}{2A(\bar{t}, z)} + \bar{t}\right) dt \\ &\leq (1 + \bar{\varepsilon}) \int_{-2\varepsilon}^{2\varepsilon} e^{-A(\bar{t}, z)s^2} g(s + \bar{t}) dt \end{aligned}$$

The last integral can be estimated using Laplace's method:

$$\int_{-2\varepsilon}^{2\varepsilon} e^{-A(\bar{t}, z)s^2} g(s + \bar{t}) dt \sim \sqrt{\frac{\pi}{A(\bar{t}, z)}} g(\bar{t}), \quad z \rightarrow 0.$$

Combining the above estimates and taking into account the definition of A , we get the following upper bound on J_2 for sufficiently small z :

$$\begin{aligned} J_2 &\leq \frac{(1 + \bar{\varepsilon})^6 \rho_0 \left(\log \frac{1}{z}\right)}{z \left(2\pi \log \frac{1}{z}\right)^{\frac{n}{2}} \sqrt{\det \mathfrak{B}}} \\ &\times e^{-\log \frac{1}{z} F(\bar{t}, \bar{\mathbf{w}}) + m \sqrt{\bar{t} \log \frac{1}{z} - \frac{1}{\bar{t}} (1 + \bar{t}\boldsymbol{\mu})^\perp \mathfrak{B}^{-1} \boldsymbol{\Lambda} + \frac{m^2 \bar{t}^2}{8\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1} (1 - \delta)}}} \sqrt{\frac{2\pi \bar{t}^3}{\log \frac{1}{z} \mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1} (1 - \delta)}} g(\bar{t}). \end{aligned}$$

Proceeding similarly, one can obtain a lower bound of the form

$$\begin{aligned} J_2 &\geq \frac{(1 - \bar{\varepsilon})^6 \rho_0 \left(\log \frac{1}{z}\right)}{z \left(2\pi \log \frac{1}{z}\right)^{\frac{n}{2}} \sqrt{\det \mathfrak{B}}} \\ &\times e^{-\log \frac{1}{z} F(\bar{t}, \bar{\mathbf{w}}) + m \sqrt{\bar{t} \log \frac{1}{z} - \frac{1}{\bar{t}} (1 + \bar{t}\boldsymbol{\mu})^\perp \mathfrak{B}^{-1} \boldsymbol{\Lambda} + \frac{m^2 \bar{t}^2}{8\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1} (1 + \delta)}}} \sqrt{\frac{2\pi \bar{t}^3}{\log \frac{1}{z} \mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1} (1 + \delta)}} g(\bar{t}). \end{aligned}$$

Since $\bar{\varepsilon}$ and δ were arbitrary, this implies that

$$\begin{aligned} J_2 &\sim \frac{(2\pi)^{\frac{1-n}{2}} \bar{t}^{\alpha + \frac{3-n}{2}}}{\sqrt{\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1} \det \mathfrak{B}}} B\left(\frac{1}{\bar{t}} \mathfrak{B}^{-1} (\mathbf{1} + \bar{t}\boldsymbol{\mu})\right) \exp\left(\frac{m^2 \bar{t}^2}{8\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1}} - \frac{1}{\bar{t}} (1 + \bar{t}\boldsymbol{\mu})^\perp \mathfrak{B}^{-1} \boldsymbol{\Lambda}\right) \\ &\times \frac{\rho_0 \left(\log \frac{1}{z}\right)}{z \left(\log \frac{1}{z}\right)^{\frac{n+1}{2}}} e^{-\log \frac{1}{z} F(\bar{t}, \bar{\mathbf{w}}) + m \sqrt{\bar{t} \log \frac{1}{z}}}. \end{aligned}$$

Estimation of J_1 and J_3 For $t \in I_1$ and $t \in I_3$, $1 + t\boldsymbol{\mu}^\perp \bar{\mathbf{w}} > 0$ and also, for z sufficiently small, $1 + t\boldsymbol{\mu}^\perp \bar{\mathbf{w}} + \boldsymbol{\Lambda}^\perp \mathbf{w} \log^{-1} \frac{1}{z} > 0$. Therefore we can estimate the log-normal sum density using that $N(x) \leq e^{-\frac{x^2}{2}}$ for $x < 0$. Then, Lemma 4.7 yields

$$p_{t \log \frac{1}{z}}(z) \leq \frac{c_q}{z \sqrt{t \log \frac{1}{z}}} \exp\left\{-\frac{\log \frac{1}{z} \left(1 + t\boldsymbol{\mu}^\perp \bar{\mathbf{w}} + \boldsymbol{\Lambda}^\perp \mathbf{w} \log^{-1} \frac{1}{z}\right)^2}{2q \bar{\mathbf{w}} \mathfrak{B} \bar{\mathbf{w}} t}\right\},$$

and then, for another constant \bar{c}_q ,

$$p_{t \log \frac{1}{z}}(z) \leq \frac{\bar{c}_q}{z \sqrt{t \log \frac{1}{z}}} \exp \left\{ -\frac{\log \frac{1}{z}}{2q} \frac{(1 + t\mu^\perp \bar{\mathbf{w}})^2}{\bar{\mathbf{w}} \mathfrak{B} \bar{\mathbf{w}} t} \right\}. \quad (4.11)$$

Since for $t \in I_1$,

$$\frac{(1 + t\mu^\perp \bar{\mathbf{w}})^2}{\bar{\mathbf{w}} \mathfrak{B} \bar{\mathbf{w}} t} \geq \frac{(1 + \bar{t}\mu^\perp \bar{\mathbf{w}})^2 \wedge 1}{\bar{\mathbf{w}} \mathfrak{B} \bar{\mathbf{w}} \bar{t}} > 0,$$

and for x large enough, $e^{-x^2} \leq \frac{e^{-\frac{x^2}{q}}}{x}$, we also have

$$p_{t \log \frac{1}{z}}(z) \leq \frac{\tilde{c}_q}{z \log \frac{1}{z}} \exp \left\{ -\frac{\log \frac{1}{z}}{2q^2} \frac{(1 + t\mu^\perp \bar{\mathbf{w}})^2}{\bar{\mathbf{w}} \mathfrak{B} \bar{\mathbf{w}} t} \right\}$$

for a different constant \tilde{c}_q and for z small enough. Therefore,

$$\begin{aligned} J_1 &\leq \frac{\tilde{c}_q}{z \log \frac{1}{z}} \int_{I_1} \rho_0 \left(t \log \frac{1}{z} \right) \\ &\quad \times \exp \left\{ -\frac{\log \frac{1}{z}}{q^2} \left[\theta t + \frac{1}{2} \frac{(1 + t\mu^\perp \bar{\mathbf{w}})^2}{\bar{\mathbf{w}}^\perp \mathfrak{B} \bar{\mathbf{w}} t} \right] + m \sqrt{t \log \frac{1}{z}} \right\} dt \end{aligned}$$

By Lemma 4.2,

$$\min_{t \in I_1} \left\{ \theta t + \frac{1}{2} \frac{(1 + t\mu^\perp \bar{\mathbf{w}})^2}{\bar{\mathbf{w}}^\perp \mathfrak{B} \bar{\mathbf{w}} t} \right\} > F(\bar{t}, \bar{\mathbf{w}}).$$

Since moreover, the function

$$t \mapsto \theta t + \frac{1}{2} \frac{(1 + t\mu^\perp \bar{\mathbf{w}})^2}{\bar{\mathbf{w}}^\perp \mathfrak{B} \bar{\mathbf{w}} t}$$

is (strictly) convex, we can choose $q > 1$, $C^* > F(\bar{t}, \bar{\mathbf{w}})$ and $c^* > 0$ such that

$$\frac{1}{q^2} \left\{ \theta t + \frac{1}{2} \frac{(1 + t\mu^\perp \bar{\mathbf{w}})^2}{\bar{\mathbf{w}}^\perp \mathfrak{B} \bar{\mathbf{w}} t} \right\} \geq C^* - tc^*, \quad t \in I_1.$$

Then,

$$\begin{aligned} J_1 &\leq e^{-C^* \log \frac{1}{z}} \frac{\tilde{c}_q}{z \log \frac{1}{z}} \int_{I_1} e^{-c^* t \log \frac{1}{z} + m \sqrt{t \log \frac{1}{z}}} \rho_0 \left(t \log \frac{1}{z} \right) dt \\ &\leq e^{-C^* \log \frac{1}{z}} \frac{\tilde{c}_q}{(\log \frac{1}{z})^2} \int_0^\infty e^{-c^* t + m \sqrt{t}} \rho_0(t) dt = o(J_2) \quad \text{as } z \rightarrow 0. \end{aligned}$$

The integral J_3 is estimated in a similar way using directly the bound (4.11).

Estimation of J_4 For $t \in I_4$, $p_{t \log \frac{1}{z}}(z) \leq \frac{c_q}{z \sqrt{t \log \frac{1}{z}}}$ and therefore

$$J_4 \leq \frac{c_q}{z \sqrt{\log \frac{1}{z}}} \int_{\bar{T}}^{\infty} e^{-\theta t \log \frac{1}{z} + m \sqrt{t \log \frac{1}{z}}} t^{-\frac{1}{2}} \rho_0 \left(t \log \frac{1}{z} \right) dt.$$

By definition of \bar{T} and Lemma 4.2,

$$\theta \bar{T} > F(\bar{t}, \bar{\mathbf{w}}).$$

Let $C^* \in (F(\bar{t}, \bar{\mathbf{w}}), \theta \bar{T})$. Then, for z small enough,

$$\begin{aligned} J_4 &\leq e^{-C^* \log \frac{1}{z}} \frac{c_q}{z \sqrt{\log \frac{1}{z}}} \int_{\bar{T}}^{\infty} e^{-(\theta - C^*/\bar{T}) t \log \frac{1}{z} + m \sqrt{t \log \frac{1}{z}}} t^{-\frac{1}{2}} \rho_0 \left(t \log \frac{1}{z} \right) dt \\ &\leq e^{-C^* \log \frac{1}{z}} \frac{c_q}{z \log \frac{1}{z}} \int_1^{\infty} e^{-(\theta - C^*/\bar{T}) t + m \sqrt{t}} t^{-\frac{1}{2}} \rho_0(t) dt = o(J_2) \quad \text{as } z \rightarrow 0. \end{aligned}$$

Estimation of the distribution function

Now we want to find an asymptotic equivalent for the cumulative distribution function associated with the density function p . To this end, we only need to integrate the previous asymptotics for $p(x)$:

$$\begin{aligned} F(x) &\sim \frac{(2\pi)^{\frac{1-n}{2}} \bar{t}^{\alpha + \frac{3-n}{2}}}{\sqrt{\mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1} \det \mathfrak{B}}} B \left(\frac{1}{\bar{t}} \mathfrak{B}^{-1} (\mathbf{1} + \bar{t} \boldsymbol{\mu}) \right) \exp \left(\frac{m^2 \bar{t}^2}{8 \mathbf{1}^\perp \mathfrak{B}^{-1} \mathbf{1}} - \frac{1}{\bar{t}} (\mathbf{1} + \bar{t} \boldsymbol{\mu})^\perp \mathfrak{B}^{-1} \boldsymbol{\Lambda} \right) \\ &\quad \int_0^x \frac{\rho_0 \left(\log \frac{1}{z} \right)}{z \left(\log \frac{1}{z} \right)^{\frac{n-1}{2}}} e^{-\log \frac{1}{z} F(\bar{t}, \bar{\mathbf{w}}) + m \sqrt{\bar{t} \log \frac{1}{z}}} dz \end{aligned}$$

We want to find an asymptotic equivalent of the integral term when $x \rightarrow 0$:

$$I(x) = \int_0^x \frac{\rho_0 \left(\log \frac{1}{z} \right)}{z \left(\log \frac{1}{z} \right)^{\frac{n-1}{2}}} e^{-\log \frac{1}{z} F(\bar{t}, \bar{\mathbf{w}}) + m \sqrt{\bar{t} \log \frac{1}{z}}} dz$$

which can be written as :

$$I(x) = \int_{\log \frac{1}{x}}^{\infty} \rho_0(y) y^{\frac{1-n}{2}} e^{-y F(\bar{t}, \bar{\mathbf{w}}) + m \sqrt{\bar{t} y}} dy.$$

For $\delta > 0$, we decompose $I(x)$ as follows.

$$\begin{aligned} I(x) &= \int_{\log \frac{1}{x}}^{(1+\delta) \log \frac{1}{x}} \rho_0(y) y^{\frac{1-n}{2}} e^{-y F(\bar{t}, \bar{\mathbf{w}}) + m \sqrt{\bar{t} y}} dy \\ &\quad + \int_{(1+\delta) \log \frac{1}{x}}^{\infty} \rho_0(y) y^{\frac{1-n}{2}} e^{-y F(\bar{t}, \bar{\mathbf{w}}) + m \sqrt{\bar{t} y}} dy. \end{aligned}$$

We shall compute an asymptotic equivalent for the first term, and show that the second term is negligible with compared to the first one.

$$\begin{aligned} & \int_{\log \frac{1}{x}}^{(1+\delta) \log \frac{1}{x}} \rho_0(y) y^{\frac{1-n}{2}} e^{-yF(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t}y}} dy \\ &= \int_1^{1+\delta} \rho \left(t \log \frac{1}{x} \right) \left(t \log \frac{1}{x} \right)^{\frac{1-n}{2}} e^{-t \log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t}} \sqrt{t \log \frac{1}{x}}} \log \frac{1}{x} dt. \end{aligned}$$

For any $\varepsilon > 0$, $t \in [1, 1 + \delta]$ and x sufficiently small,

$$\rho_0 \left(\log \frac{1}{x} \right) t^\alpha (1 - \varepsilon) \leq \rho_0 \left(t \log \frac{1}{x} \right) \leq \rho_0 \left(\log \frac{1}{x} \right) t^\alpha (1 + \varepsilon).$$

Then,

$$\begin{aligned} & \int_{\log \frac{1}{x}}^{(1+\delta) \log \frac{1}{x}} \rho_0(y) y^{\frac{1-n}{2}} e^{-yF(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t}y}} dy \\ & \leq \rho_0 \left(\log \frac{1}{x} \right) \left(\log \frac{1}{x} \right)^{\frac{3-n}{2}} (1 + \varepsilon) \int_1^{1+\delta} t^{\alpha + \frac{1-n}{2}} e^{-t \log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t}} \sqrt{t \log \frac{1}{x}}} dt \\ & \leq \rho_0 \left(\log \frac{1}{x} \right) \left(\log \frac{1}{x} \right)^{\frac{3-n}{2}} (1 + \varepsilon) (1 + \delta)^{\alpha + \frac{1-n}{2}} \int_1^{1+\delta} e^{-t \log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t}} \sqrt{t \log \frac{1}{x}}} dt. \end{aligned}$$

With a change of variable

$$-t \log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t}} \sqrt{t \log \frac{1}{x}} = -s \log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}}),$$

we get, for x small enough,

$$\begin{aligned} \int_1^{1+\delta} e^{-t \log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t}} \sqrt{t \log \frac{1}{x}}} dt &= \int_{1 - \frac{m\sqrt{\bar{t}}}{F(\bar{t}, \bar{\mathbf{w}}) \sqrt{\log \frac{1}{x}}}}^{1+\delta - \frac{m\sqrt{\bar{t}} \sqrt{1+\delta}}{F(\bar{t}, \bar{\mathbf{w}}) \sqrt{\log \frac{1}{x}}}} e^{-s \log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}})} \frac{ds}{1 - \frac{m\sqrt{\bar{t}}}{2F(\bar{t}, \bar{\mathbf{w}}) \sqrt{\log \frac{1}{x}}}} \\ &\leq \frac{1}{1 - \frac{|m\sqrt{\bar{t}}|}{2F(\bar{t}, \bar{\mathbf{w}}) \sqrt{\log \frac{1}{x}}}} \int_{1 - \frac{m\sqrt{\bar{t}}}{F(\bar{t}, \bar{\mathbf{w}}) \sqrt{\log \frac{1}{x}}}}^{1+\delta - \frac{m\sqrt{\bar{t}} \sqrt{1+\delta}}{F(\bar{t}, \bar{\mathbf{w}}) \sqrt{\log \frac{1}{x}}}} e^{-s \log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}})} ds \\ &\leq \frac{1}{\log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}})} \frac{1}{1 - \frac{|m\sqrt{\bar{t}}|}{2F(\bar{t}, \bar{\mathbf{w}}) \sqrt{\log \frac{1}{x}}}} e^{-\log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t}} \sqrt{\log \frac{1}{x}}} \\ &\leq \frac{1 + \varepsilon}{\log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}})} e^{-\log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t}} \sqrt{\log \frac{1}{x}}}, \end{aligned}$$

so that

$$\begin{aligned} & \int_{\log \frac{1}{x}}^{(1+\delta) \log \frac{1}{x}} \rho_0(y) y^{\frac{1-n}{2}} e^{-yF(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t}y}} dy \\ & \leq \rho_0 \left(\log \frac{1}{x} \right) \left(\log \frac{1}{x} \right)^{\frac{1-n}{2}} (1 + \varepsilon)^2 (1 + \delta)^{\alpha - \frac{1+n}{2}} \frac{1}{F(\bar{t}, \bar{\mathbf{w}})} e^{-\log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t}} \sqrt{\log \frac{1}{x}}}. \end{aligned} \tag{4.12}$$

Proceeding similarly, we can obtain, for x small enough, a lower bound of the following form.

$$\begin{aligned} & \int_{\log \frac{1}{x}}^{(1+\delta) \log \frac{1}{x}} \rho_0(y) y^{\frac{1-n}{2}} e^{-yF(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t}y}} dy \\ & \geq \rho_0 \left(\log \frac{1}{x} \right) \left(\log \frac{1}{x} \right)^{\frac{1-n}{2}} (1-\varepsilon)^3 \frac{1}{F(\bar{t}, \bar{\mathbf{w}})} e^{-\log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t}} \sqrt{\log \frac{1}{x}}}. \end{aligned} \quad (4.13)$$

Our next goal is to analyze

$$\int_{(1+\delta) \log \frac{1}{x}}^{\infty} \rho_0(y) y^{\frac{1-n}{2}} e^{-yF(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t}y}} dy$$

By Potter's theorem (see [?]), we know that for any $A > 1$, $\beta > 0$, there exists $X = X(A, \beta)$ such that:

$$\frac{\rho_0(y)}{\rho_0(x)} \leq A \max \left\{ \left(\frac{y}{x} \right)^{\alpha+\beta}, \left(\frac{y}{x} \right)^{\alpha-\beta} \right\}, \quad x \geq X, y \geq X.$$

Therefore, for x sufficiently small,

$$\begin{aligned} & \int_{(1+\delta) \log \frac{1}{x}}^{\infty} \rho_0(y) y^{\frac{1-n}{2}} e^{-yF(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t}y}} dy \\ & \leq A \left(\log \frac{1}{x} \right)^{\frac{3-n}{2}} \rho_0 \left(\log \frac{1}{x} \right) \int_{1+\delta}^{\infty} t^{\alpha+\beta+\frac{1-n}{2}} e^{-t \log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{t \log \frac{1}{x}}} dt. \end{aligned}$$

For every $\varepsilon > 0$, for $t \geq 1$ and x sufficiently small,

$$F(\bar{t}, \bar{\mathbf{w}}) - \frac{m\sqrt{\bar{t}}}{\sqrt{t \log \frac{1}{x}}} - \left(\alpha + \beta + \frac{1-n}{2} \right) \frac{\log t}{t \log \frac{1}{x}} \geq F(\bar{t}, \bar{\mathbf{w}}) (1-\varepsilon).$$

Therefore,

$$\begin{aligned} & \int_{(1+\delta) \log \frac{1}{x}}^{\infty} \rho_0(y) y^{\frac{1-n}{2}} e^{-yF(\bar{t}, \bar{\mathbf{w}}) + m\sqrt{\bar{t}y}} dy \\ & \leq A \left(\log \frac{1}{x} \right)^{\frac{3-n}{2}} \rho_0 \left(\log \frac{1}{x} \right) \int_{1+\delta}^{\infty} e^{-t \log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}}) (1-\varepsilon)} dt \\ & = \frac{A}{F(\bar{t}, \bar{\mathbf{w}}) (1-\varepsilon)} \left(\log \frac{1}{x} \right)^{\frac{1-n}{2}} \rho_0 \left(\log \frac{1}{x} \right) e^{-(1+\delta) \log \frac{1}{x} F(\bar{t}, \bar{\mathbf{w}}) (1-\varepsilon)} \end{aligned}$$

Taking $\varepsilon < \delta$, we see that the integral over the interval $[1 + \delta, \infty]$ is negligible compared to the integral over the interval $[1, 1 + \delta]$. Combining the above estimate with (4.12) and (4.13), and using the fact that ε and δ are arbitrary, we finally find the asymptotic equivalent for the cumulative distribution function given in the statement of the theorem.

Chapter 5

Asymptotics of heavy-tailed risks with Gaussian copula dependence

Contents

1	Introduction	120
2	Main results	121
3	Examples	127
3.1	Portfolio of options in the Black-Scholes model	127
3.2	Aggregate production of a wind farm	130

1 Introduction

In this note, we show that a multivariate distribution with heavy-tailed (regularly varying) univariate margins and Gaussian copula dependence exhibits multivariate regular variation on the cone $(0, \infty)^n$ even if the margins are not identically distributed, and the distribution is not multivariate regularly varying on $[0, \infty)^n$. This enables us to compute sharp tail asymptotics of certain functionals, such as the sum of components, of heavy-tailed random vectors with Gaussian copula dependence.

It is shown in [?] that the components of such a random vector are asymptotically independent, meaning that the probability that any two components of a random vector are simultaneously large, on a suitable scale, is negligible compared to the probability that any one component is large. Asymptotic independence is a natural property which is often observed in the data coming from many different application domains [?, ?, ?, ?, ?, ?], and is an inherent feature of many widely used models, for example, in finance. In addition to the multivariate Gaussian [?], one can quote, e.g., the multivariate generalized hyperbolic [?], and more generally all Gaussian mixture models with exponentially decaying mixing variable [?].

In the literature, asymptotic independence is often introduced using the property of hidden regular variation [?, ?, ?], which is a refinement of the coefficient of tail dependence introduced in [?, ?]. Recall that a random vector \mathbf{X} with values in $[0, \infty)^n$ is said to be multivariate regularly varying if there exists a function $b(t) \rightarrow \infty$ as $t \rightarrow \infty$ and a non-negative Radon measure $\nu \neq 0$ such that

$$t \Pr \left[\frac{\mathbf{X}}{b(t)} \in \cdot \right] \rightarrow \nu \quad (5.1)$$

as $t \rightarrow \infty$ in the sense of vague convergence of measures on the cone $E = [0, \infty]^n \setminus \{0\}$. Now, \mathbf{X} is said to possess the property of hidden regular variation if in addition to (5.1), there exists a non-decreasing function $b^*(t) \rightarrow \infty$ such that $b(t)/b^*(t) \rightarrow \infty$ as $t \rightarrow \infty$ and a Radon measure $\nu^* \neq 0$ such that

$$t \Pr \left[\frac{\mathbf{X}}{b^*(t)} \in \cdot \right] \rightarrow \nu^* \quad (5.2)$$

on the cone $E^* := E \setminus \cup_{i=1}^n L_i$, where L_i is the i -th coordinate axis. In other words, under hidden regular variation, the probability $\Pr[X_i \geq x_i t, X_j \geq x_j t]$ for any $i \neq j$ decays regularly, but at a faster rate than the tail probabilities of individual components.

The theory of hidden regular variation along with its more recent extensions [?] allows to quantify the asymptotic behavior of the tail probabilities but it is well suited for distributions with tail equivalent margins. In the context of this paper, the components of the random vector are not necessarily tail equivalent,

hence multivariate regular variation on $[0, \infty]^n \setminus \{0\}$ may not hold. Nevertheless, we show that the multivariate random vector \mathbf{X} exhibits multivariate regular variation on the cone $(0, \infty)^n$, which enables us to develop a number of applications such as the computation of risk measures for a portfolio of options in the Black-Scholes model, and of the probability distribution of the aggregate production of a wind farm.

Remarks on notation Throughout this paper, we write $f \sim g$ as x tends to a whenever

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = 1.$$

We recall that a function f is called slowly varying as x tends to 0 whenever

$$\lim_{x \rightarrow 0} \frac{f(\alpha x)}{f(x)} = 1$$

for all $\alpha > 0$.

We also recall that the copula of a random vector (Y_1, \dots, Y_n) is a function $C : [0, 1]^n : [0, 1]$, satisfying the assumptions

- dC is a positive measure in the sense of Lebesgue-Stieltjes integration,
- $C(u_1, \dots, u_n) = 0$ whenever $u_k = 0$ for at least one k ,
- $C(u_1, \dots, u_n) = u_k$ whenever $u_i = 1$ for all $i \neq k$,

and such that

$$\Pr[Y_1 \leq y_1, \dots, Y_n \leq y_n] = C(\Pr[Y_1 \leq y_1], \dots, \Pr[Y_n \leq y_n]), \quad (y_1, \dots, y_n) \in \mathbb{R}^n.$$

A copula exists by Sklar's theorem and is uniquely defined whenever the marginal distributions of Y_1, \dots, Y_n are continuous. We refer to [?] for details on copulas.

2 Main results

Let X_1, \dots, X_n be random variables with values in (a, ∞) with $a \in [-\infty, \infty)$, whose distribution functions are denoted by F_1, \dots, F_n . We assume that the dependence of the random vector $\mathbf{X} = (X_1, \dots, X_n)$ is specified by a Gaussian copula, that is, $X_k = F_k^{-1} \left(\mathcal{N} \left(\tilde{X}_k \right) \right)$ for $k = 1, \dots, n$, where $\tilde{\mathbf{X}} = \left(\tilde{X}_1, \dots, \tilde{X}_n \right)$ is a Gaussian vector with standard normal margins and correlation matrix Σ , \mathcal{N} is the standard normal distribution function and F_k^{-1} is the generalized inverse of F_k .

Our main result quantifies the tail behavior of X_1, \dots, X_n under the assumptions that the components of this vector are regularly varying. To prove it, we shall need

a lemma on the asymptotic behavior of multivariate Gaussian tails, adapted from [?] (see Theorem 4.1). To state this lemma, we first recall the following notation.

Let $\widetilde{\mathbf{X}}$ be a Gaussian vector with standard normal margins and a positive definite correlation matrix Σ and denote by B the inverse matrix of Σ . For a fixed $\mathbf{x} \in \mathbb{R}_+^n$ with $\mathbf{x} \neq 0$, denote \mathbf{x}^* the solution (minimizer) of the problem

$$\min_{\mathbf{z} \geq \mathbf{x}} \langle \mathbf{z}, B\mathbf{z} \rangle. \quad (5.3)$$

Then (see Proposition 2 in [?]), there exist unique disjoint index sets I and J , with $I \cup J = \{1, \dots, n\}$, such that

$$\mathbf{x}_I^* = \mathbf{x}_I, \quad \mathbf{x}_J^* = -B_{JJ}^{-1}B_{JI}\mathbf{x}_I, \quad (5.4)$$

$$\text{and } \langle \mathbf{e}_i, \Sigma_{II}^{-1}\mathbf{x}_I \rangle > 0, \quad \forall i \in I, \quad (5.5)$$

where in the above matrix notation indexing is performed before taking the inverse. Indeed, introduce the Lagrange multipliers corresponding to the problem (5.3):

$$\boldsymbol{\lambda}^* = \arg \max_{\boldsymbol{\lambda} \geq 0} \langle \boldsymbol{\lambda}, \mathbf{x} \rangle - \frac{1}{2} \langle \boldsymbol{\lambda}, \Sigma \boldsymbol{\lambda} \rangle.$$

The optimizer of (5.3) is then given by $\mathbf{x}^* = \Sigma \boldsymbol{\lambda}^*$. Letting $I := \{i : \lambda_i^* > 0\}$, we have that

$$\boldsymbol{\lambda}_I^* = \Sigma_{II}^{-1}\mathbf{x}_I > 0, \quad \mathbf{x}_I^* = \Sigma_{II}\boldsymbol{\lambda}_I^* = \mathbf{x}_I$$

$$\text{and } \mathbf{x}_J^* = \Sigma_{JI}\boldsymbol{\lambda}_I^* = \Sigma_{JI}\Sigma_{II}^{-1}\mathbf{x}_I = -B_{JJ}^{-1}B_{JI}\mathbf{x}_I.$$

Lemma 5.1. *Let $\mathbf{f}(t) : \mathbb{R}_+ \rightarrow \mathbb{R}_+^n$ be such that*

$$\lim_{t \rightarrow \infty} \frac{\mathbf{f}(t)}{t} = \mathbf{x} \in \mathbb{R}_+^n,$$

and assume that \mathbf{x} is such that the solution of the corresponding quadratic optimization problem satisfies $\mathbf{x}_J^ > \mathbf{x}_J$. Then, as $t \rightarrow +\infty$,*

$$\mathbb{P}[\widetilde{\mathbf{X}} > \mathbf{f}(t)] = \frac{\exp(-\alpha_t/2)}{t^{|I|}(2\pi)^{|I|/2}|\Sigma_{II}^{1/2}| \prod_{i \in I} \langle \mathbf{e}_i, \Sigma_{II}^{-1}\mathbf{x}_I \rangle} (1 + o(1)),$$

where

$$\alpha_t = \langle \mathbf{f}(t)_I, \Sigma_{II}^{-1}\mathbf{f}(t)_I \rangle$$

Proof. Our aim is to use Theorem 4.1 in [?]. We are led to consider the family of optimization problems

$$\min_{\mathbf{z} \geq \mathbf{x}_t} \langle \mathbf{z}, B\mathbf{z} \rangle, \quad (5.6)$$

where $\mathbf{x}_t := \mathbf{f}(t)/t$. Let \mathbf{x}_t^* be the minimizer and $\boldsymbol{\lambda}_t^*$ be the vector of Lagrange multipliers corresponding to the problem (5.6) and \mathbf{x}^* and $\boldsymbol{\lambda}^*$ be the corresponding quantities for the limiting problem (with \mathbf{x} instead of \mathbf{x}_t). Introduce the sets $I_t := \{i : \lambda_{t,i}^* > 0\}$ and $I := \{i : \lambda_i^* > 0\}$ as well as the sets $\tilde{J}_t := \{i : x_{t,i}^* > x_{t,i}\}$ and $J := \{i : x_i^* > x_i\}$. Since both Σ and B are nondegenerate, we have that $\boldsymbol{\lambda}_t^* \rightarrow \boldsymbol{\lambda}^*$ and $\mathbf{x}_t^* \rightarrow \mathbf{x}^*$ as $t \rightarrow \infty$. This implies that for sufficiently large t , $I \subseteq I_t$ and $J \subseteq \tilde{J}_t$. Since the sets I_t and \tilde{J}_t are disjoint and $I \cup J = \{1, \dots, n\}$, we conclude that for sufficiently large t , $I_t = I$.

Further, we have that

$$\lim_{t \rightarrow \infty} (\mathbf{x}_t - \mathbf{x}_t^*)_J = (\mathbf{x} - \mathbf{x}^*)_J < 0,$$

which means that the first part of equation (4.1) in [?] is satisfied with $\mathbf{t}_J^* = -\infty$ (componentwise). Moreover the second part of equation (4.1) is satisfied due to (5.5). All assumptions of Theorem 4.1 in [?] are thus satisfied and we conclude that

$$\begin{aligned} \mathbb{P}[\tilde{\mathbf{X}} > \mathbf{f}(t)] &= \frac{\exp(-\alpha_t/2)}{(2\pi)^{|I|/2} |\Sigma_{II}^{1/2}| \prod_{i \in I} \langle \mathbf{e}_i, \Sigma_{II}^{-1} \mathbf{f}(t)_I \rangle} (1 + o(1)) \\ &= \frac{\exp(-\alpha_t/2)}{t^{|I|} (2\pi)^{|I|/2} |\Sigma_{II}^{1/2}| \prod_{i \in I} \langle \mathbf{e}_i, \Sigma_{II}^{-1} \mathbf{x}_I \rangle} (1 + o(1)). \end{aligned}$$

□

The next theorem is the main result of this paper.

Theorem 5.2. *Let X_1, \dots, X_n be random variables with values in (a, ∞) , whose distribution functions are denoted by F_1, \dots, F_n . Assume that the dependence of the random vector $\mathbf{X} = (X_1, \dots, X_n)$ is specified by a Gaussian copula with correlation matrix Σ , and that for each i , the survival function $\tilde{F}_i = 1 - F_i$ is regularly varying at $+\infty$ with index $-\alpha_i < 0$. Denote $\boldsymbol{\beta} = (\sqrt{\alpha_1}, \dots, \sqrt{\alpha_n})$ and let $\boldsymbol{\beta}^*$ be the solution of the problem*

$$\min_{\mathbf{z} \geq \boldsymbol{\beta}} \langle \mathbf{z}, B\mathbf{z} \rangle,$$

and I and J be the corresponding index sets. Assume that $\boldsymbol{\beta}_J^* > \boldsymbol{\beta}_J$. Then for $x_1 > 0, \dots, x_n > 0$ as $u \uparrow \infty$,

$$\begin{aligned} &\mathbb{P}[X_1 > ux_1, \dots, X_n > ux_n] \\ &= \exp \left\{ - \sum_{i,j \in I} b_{ij} \sqrt{\ln \tilde{F}_i(u) \ln \tilde{F}_j(u)} \right\} L(u) \nu(\mathbf{x}) (1 + o(1)) \end{aligned}$$

where

$$\begin{aligned}
L(u) &= C \ln u^{\frac{1}{2} \sum_{i,j} b_{ij} \frac{\beta_i}{\beta_j} - \frac{|I|}{2}} \\
C &= \prod_{i \in I} \beta_i^{\frac{1}{\beta_i} \sum_{j \in I} b_{ij} \beta_j} \frac{(4\pi)^{\frac{1}{2} \sum_{i,j \in I} b_{ij} \frac{\beta_i}{\beta_j} - \frac{|I|}{2}}}{|\Sigma_{II}|^{\frac{1}{2}} \prod_{i \in I} \langle \mathbf{e}_i, \Sigma_{II}^{-1} \boldsymbol{\beta}_I \rangle} \\
\nu(\mathbf{x}) &= \prod_{i \in I} x_i^{-\beta_i \sum_{j \in I} b_{ij} \beta_j}
\end{aligned}$$

Remark 5.3. The above theorem states that the vector \mathbf{X} has the property of multivariate regular variation, but on the cone $(0, \infty)^n$ rather than $[0, \infty)^n \setminus \{0\}$. Since, in our context, the components are not identically distributed, multivariate regular variation on $[0, \infty)^n$ may not hold.

The proof is based on the following lemma.

Lemma 5.4. *Let F be regularly varying at $+\infty$ with index $-\alpha < 0$. Then, as $u \rightarrow +\infty$,*

$$\begin{aligned}
\mathcal{N}^{-1}(F(ux)) &= -\sqrt{2 \ln \frac{1}{F(u)} - \frac{\sqrt{\alpha} \ln x}{\sqrt{2 \ln u}} + \frac{1}{2} \frac{\ln \left(\ln \frac{1}{F(u)} \right)}{\sqrt{2 \ln \frac{1}{F(u)}}} + \frac{\ln \sqrt{4\pi}}{\sqrt{2 \ln \frac{1}{F(u)}}} + o\left(\frac{1}{\sqrt{\ln u}}\right) \\
&= -\sqrt{2 \ln \frac{1}{F(u)} - \frac{\sqrt{\alpha} \ln x}{\sqrt{2 \ln u}} + \frac{1}{2} \frac{\ln(\alpha \ln u)}{\sqrt{2 \ln \frac{1}{F(u)}}} + \frac{\ln \sqrt{4\pi}}{\sqrt{2\alpha \ln u}} + o\left(\frac{1}{\sqrt{\ln u}}\right).
\end{aligned}$$

Proof of Lemma 5.4. From the regular variation property it follows that $F(ux) = x^{-\alpha} F(u)(1+o(1))$ as $u \rightarrow +\infty$ and also that $\ln F(u) = -\alpha \ln u + o(|\ln u|)$. Therefore, as $u \rightarrow +\infty$,

$$\sqrt{2 \ln \frac{1}{F(ux)}} = \sqrt{2\alpha \ln x + 2 \ln \frac{1}{F(u)} + o(1)} = \sqrt{2 \ln \frac{1}{F(u)} + \frac{\alpha \ln x}{\sqrt{2\alpha \ln u}} + o\left(\frac{1}{\sqrt{\ln u}}\right)}.$$

On the other hand, as $u \rightarrow 0$,

$$\mathcal{N}^{-1}(u) = -\sqrt{2 \ln \frac{1}{u}} + \frac{1}{2} \frac{\ln \ln \frac{1}{u}}{\sqrt{2 \ln \frac{1}{u}}} + \frac{\ln \sqrt{4\pi}}{\sqrt{2 \ln \frac{1}{u}}} + o\left(\frac{1}{\ln \frac{1}{u}}\right).$$

Indeed, using the expansion for the inverse error function from [?], we deduce that

$$\mathcal{N}^{-1}(u) = -\sqrt{2 \ln \frac{1}{u} - 2 \ln(\sqrt{4\pi}) - \ln \ln \frac{1}{u}} + o\left(\ln^{-1} \frac{1}{u}\right), \quad u \rightarrow 0,$$

and expanding the square root, we obtain the required formula. \square

Proof of Theorem 5.2. Let us denote

$$\mathbf{f}(u) = \left\{ -N^{-1} \left(\tilde{F}_i(ux_i) \right), i = 1, \dots, n \right\}.$$

and remark that

$$\begin{aligned} & \mathbb{P} [X_1 > ux_1, \dots, X_n > ux_n] \\ &= \mathbb{P} \left[\tilde{X}_1 > -N^{-1} \left(\tilde{F}_1(ux_1) \right), \dots, \tilde{X}_n > -N^{-1} \left(\tilde{F}_n(ux_n) \right) \right] \\ &= \mathbb{P} \left[\tilde{\mathbf{X}} > \mathbf{f}(u) \right] \end{aligned}$$

Then, from Lemma 5.4, we find that

$$\frac{\mathbf{f}(u)}{\sqrt{2 \ln u}} \rightarrow \boldsymbol{\beta} \in \mathbb{R}_+^n$$

Therefore, Lemma 5.1 leads to

$$\mathbb{P}[X_1 > ux_1, \dots, X_n > ux_n] = \frac{e^{-\alpha_u/2}}{(2 \ln u)^{\frac{|I|}{2}} (2\pi)^{\frac{|I|}{2}} |\Sigma_{II}|^{\frac{1}{2}} \prod_{i \in I} \langle \mathbf{e}_i, \Sigma_{II}^{-1} \boldsymbol{\beta}_j \rangle} (1 + o(1))$$

as $u \rightarrow \infty$. The definition of α_u

$$\alpha_u = \langle \mathbf{f}(u), \Sigma_{II}^{-1} \mathbf{f}(u) \rangle = \sum_{i,j \in I} b_{ij} N^{-1} \left(\tilde{F}_i(ux_i) \right) N^{-1} \left(\tilde{F}_j(ux_j) \right)$$

combined with Lemma 5.4 leads to the desired form of the asymptotic:

$$\begin{aligned} e^{-\alpha_u/2} &= \exp \left(-\frac{1}{2} \sum_{i,j \in I} b_{ij} N^{-1} \left(\tilde{F}_i(ux_i) \right) N^{-1} \left(\tilde{F}_j(ux_j) \right) \right) \\ &= \exp \left(-\sum_{i,j \in I} b_{ij} \sqrt{\ln \frac{1}{\tilde{F}_i(u)} \ln \frac{1}{\tilde{F}_j(u)}} \right) \\ &\times \exp \left(-\sum_{i,j \in I} b_{ij} \beta_i \beta_j \ln x_j \right) \\ &\times \exp \left(\frac{\ln \ln u}{2} \sum_{i,j \in I} b_{ij} \frac{\beta_i}{\beta_j} \right) \\ &\times \exp \left(\sum_{i,j \in I} b_{ij} \beta_i \frac{\ln(\beta_j 4\pi)}{\beta_j} \right) \end{aligned}$$

□

We are now interested in studying the asymptotic behavior near zero of a sum of positive heavy-tailed risks. This is given by the following corollary.

Corollary 5.5. *Let X_1, \dots, X_n be random variables with values in $(0, \infty)$, whose distribution functions are denoted by F_1, \dots, F_n . Assume that the dependence of the random vector $\mathbf{X} = (X_1, \dots, X_n)$ is specified by a Gaussian copula with correlation matrix Σ , and that for each i , the distribution function F_i is regularly varying at 0 with index $\alpha_i > 0$. Denote $\boldsymbol{\beta} = (\sqrt{\alpha_1}, \dots, \sqrt{\alpha_n})$ and let $\boldsymbol{\beta}^*$ be the solution of the problem*

$$\min_{\mathbf{z} \geq \boldsymbol{\beta}} \langle \mathbf{z}, B\mathbf{z} \rangle,$$

and I and J be the corresponding index sets. Assume that $\boldsymbol{\beta}_J^* > \boldsymbol{\beta}_J$. Then, as $z \downarrow 0$,

$$\mathbb{P}[X_1 + \dots + X_n < z] = \exp \left\{ - \sum_{i,j \in I} b_{ij} \sqrt{\ln F_i(z) \ln F_j(z)} \right\} L \left(\frac{1}{z} \right) \nu_0(A)$$

where L is the function defined in Theorem 5.2 and

$$\begin{aligned} \nu_0(A) &= \prod_{k=1}^{|I|-1} p_{i_k} B \left(p_{i_k}, 1 + \sum_{j=k+1}^{|I|} p_{i_j} \right). \\ p_i &= \beta_i \sum_{j \in I} b_{ij} \beta_j. \end{aligned}$$

Here B is the Euler beta function, the elements of the set I are denoted by i_k for $k = 1, \dots, |I|$ (in increasing order) and the empty product by convention equals 1.

Proof. Define the set

$$A = \{(x_1, \dots, x_n) \in \mathbb{R}_+^n : x_1^{-1} + \dots + x_n^{-1} < 1\}.$$

and remark that the closure of A belongs to the cone $(0, \infty)^n$, that

$$X_1 + \dots + X_n < z \iff (X_1^{-1}, \dots, X_n^{-1}) \in A,$$

and that the random variables $(X_1^{-1}, \dots, X_n^{-1})$ satisfy the assumptions of Theorem 5.2. Therefore,

$$\begin{aligned} \mathbb{P}[X_1 + \dots + X_n < z] &= \mathbb{P}[(X_1^{-1}, \dots, X_n^{-1}) \in \frac{1}{z}A] \\ &= \exp \left\{ - \sum_{i,j \in I} b_{ij} \sqrt{\ln F_i(z) \ln F_j(z)} \right\} L \left(\frac{1}{z} \right) \nu_0(A) (1 + o(1)), \end{aligned}$$

where ν_0 is the measure defined by

$$\nu_0([x_1, \infty), \dots, [x_n, \infty)) := \prod_{i \in I} x_i^{-\beta_i \sum_{j \in I} b_{ij} \beta_j} = \prod_{i \in I} x_i^{-p_i}.$$

It remains to compute $\nu_0(A)$. Let

$$A_I = \{(x_i)_{i \in I} \in \mathbb{R}_+^{|I|} : \sum_{i \in I} x_i^{-1} < 1\},$$

and let the set I contain the elements $i_k, k = 1, \dots, |I|$. Then,

$$\begin{aligned}\nu_0(A) &= \int_{A_I} \prod_{i \in I} p_i x_i^{-p_i-1} dx_i \\ &= \prod_{i \in I} p_i \int_1^\infty dx_1 \int_{1-\frac{1}{x_1}} dx_2 \cdots \int_{1-\frac{1}{x_1}-\cdots-\frac{1}{x_{|I|-1}}} x_1^{-p_{i_1}-1} \cdots x_{|I|}^{-p_{i_{|I|}}-1} dx_{|I|}\end{aligned}$$

Making the change of variable $y_i = \frac{1}{x_i}$ for $i = 1, \dots, |I|$ leads to

$$\begin{aligned}\nu_0(A) &= \prod_{i \in I} p_i \int_0^1 dy_1 \int_0^{1-y_1} dy_2 \cdots \int_0^{1-y_1-\cdots-y_{|I|-1}} y_1^{p_{i_1}-1} y_2^{p_{i_2}-1} \cdots y_{|I|}^{p_{i_{|I|}}-1} dy_{|I|} \\ &= \prod_{k=1}^{|I|-1} p_{i_k} \int_0^1 dy_1 \int_0^{1-y_2} dy_2 \cdots \\ &\quad \int_0^{1-y_1-\cdots-y_{|I|-2}} y_1^{p_{i_1}-1} y_2^{p_{i_2}-1} \cdots y_{|I|-1}^{p_{i_{|I|-1}}-1} (1-y_1-\cdots-y_{|I|-1})^{p_{i_{|I|}}} dy_{|I|-1} \\ &= B(p_{i_{|I|-1}}, p_{i_{|I|}} + 1) \prod_{k=1}^{|I|-1} p_{i_k} \int_0^1 dy_1 \int_0^{1-y_1} dy_2 \cdots \\ &\quad \int_0^{1-y_1-\cdots-y_{|I|-3}} y_1^{p_{i_1}-1} \cdots y_{|I|-2}^{p_{i_{|I|-2}}-1} (1-y_1-\cdots-y_{|I|-2})^{p_{i_{|I|}}+p_{i_{|I|-1}}} dy_{|I|-2},\end{aligned}$$

where B is the Euler beta function. Continuing the computation iteratively, we find that

$$\nu_0(A) = \prod_{k=1}^{|I|-1} p_{i_k} B\left(p_{i_k}, 1 + \sum_{j=k+1}^{|I|} p_{i_j}\right).$$

□

3 Examples

3.1 Portfolio of options in the Black-Scholes model

Fix a time horizon T and let (X_1, \dots, X_n) denote the vector of logarithmic returns of n risky assets over this time horizon. The risky asset values at date T are then given by $S_i = e^{X_i}$ for $i = 1, \dots, n$ where we have assumed without loss of generality that the initial values of all assets are normalized to 1.

We suppose that the n risky assets follow the multidimensional Black-Scholes model. This means that the distribution of the vector (X_1, \dots, X_n) is Gaussian, and we denote by $\mathfrak{B}T$ its covariance matrix and by $\boldsymbol{\mu}T$ its mean vector.

We are interested in the tail behavior of a long-only portfolio of European call options written on n risky assets. To simplify the discussion we assume that it contains exactly one unit of each option, but the setting can obviously be extended

to an arbitrary number of units. The log-strikes of the options will be denoted by (k_1, \dots, k_n) and the maturity dates by (T_1, \dots, T_n) , where $T_i > T$ for $i = 1, \dots, n$. Assuming that the interest rate is zero, the price of i -th option at date T is given by the Black-Scholes formula:

$$P_i = e^{X_i} \mathcal{N}(d_+) - e^{k_i} \mathcal{N}(d_-), \quad d_{\pm} = \frac{X_i - k_i}{\sigma_i \sqrt{T_i - T}} \pm \frac{\sigma_i \sqrt{T_i - T}}{2}, \quad \sigma_i = \sqrt{\mathfrak{B}_{ii}},$$

where \mathcal{N} is the standard normal distribution function. Let

$$\tilde{X}_i = \frac{X_i - \mu_i T}{\sigma_i \sqrt{T}}$$

for $i = 1, \dots, n$ and define

$$\begin{aligned} f_i(x) &= e^{\mu_i T + x \sigma_i \sqrt{T}} \mathcal{N}\{d_+(x)\} - e^{k_i} \mathcal{N}\{d_-(x)\}, \\ d_{\pm}(x) &= x \sqrt{\frac{T}{T_i - T}} + \frac{\mu_i T - k_i}{\sigma_i \sqrt{T_i - T}} \pm \frac{\sigma_i \sqrt{T_i - T}}{2}. \end{aligned}$$

Then, \tilde{X}_i is a standard normal random variable and $P_i = f_i(\tilde{X}_i)$.

Therefore, to fit in the setting of the preceding section, we set $F_i = \mathcal{N} \circ f_i^{-1}$.

Lemma 5.6. *As $u \rightarrow 0$,*

$$\frac{1}{\tau_i} f_i^{-1}(u) = -\sqrt{2 \ln \frac{1}{u}} + c_i + \frac{\ln \ln \frac{1}{u}}{\sqrt{2 \ln \frac{1}{u}}} - \frac{d_i}{\sqrt{2 \ln \frac{1}{u}}} + O(|\ln u|^{-1})$$

and

$$F_i(u) = u^{\tau_i^2} \exp \left\{ c_i \tau_i^2 \sqrt{2 \ln \frac{1}{u}} + \tau_i^2 \ln \ln \frac{1}{u} \right\} \left(2 \ln \frac{1}{u} \right)^{-\frac{1}{2}} \frac{e^{-\tau_i^2 d_i - \frac{1}{2} \tau_i^2 c_i^2}}{\tau_i \sqrt{2\pi}} (1 + o(1))$$

where

$$c_i = \frac{k_i - \mu_i T}{\sigma_i \sqrt{T_i - T}} + \frac{\sigma_i \sqrt{T_i - T}}{2}, \quad d_i = \ln \frac{e^{k_i} \sigma_i (T_i - T)^{\frac{3}{2}}}{T \sqrt{2\pi}} \quad \text{and} \quad \tau_i = \sqrt{\frac{T_i - T}{T}}.$$

Proof. From the well-known estimate

$$\mathcal{N}(x) = \frac{e^{-\frac{x^2}{2}}}{|x| \sqrt{2\pi}} (1 + O(|x|^{-2})), \quad x \rightarrow -\infty, \quad (5.7)$$

one easily deduces that

$$f_i(x) = \frac{\sigma_i (T_i - T)^{\frac{3}{2}} e^{k_i - d_i^2/2}}{x^2 T \sqrt{2\pi}} (1 + O(|x|^{-2})), \quad x \rightarrow -\infty. \quad (5.8)$$

Taking the logarithm, we obtain

$$\ln f_i(x) = -\frac{1}{2} \left(x \sqrt{\frac{T}{T_i - T}} - c_i \right)^2 - 2 \ln |x| + d_i + O(|x|^2).$$

Therefore,

$$f_i^{-1}(u) \sim -\sqrt{2 \frac{T_i - T}{T} \ln \frac{1}{u}}, \quad u \rightarrow 0,$$

and so

$$\begin{aligned} f_i^{-1}(u) \sqrt{\frac{T}{T_i - T}} &= c_i - \sqrt{2 \ln \frac{1}{u} - 4 \ln |f_i^{-1}(u)| + 2d_i - c_i^2 + O(|\ln u|^{-1})} \\ &= c_i - \sqrt{2 \ln \frac{1}{u} - 4 \ln |f_i^{-1}(u)| + 2d_i - c_i^2 + O(|\ln u|^{-3/2})} \\ &= c_i - \sqrt{2 \ln \frac{1}{u} + \frac{2 \ln |f_i^{-1}(u)|}{\sqrt{2 \ln \frac{1}{u}}} - \frac{d_i - \frac{1}{2}c_i^2}{\sqrt{2 \ln \frac{1}{u}}} + O(\ln \ln \frac{1}{u} |\ln u|^{-3/2})} \\ &= -\sqrt{2 \ln \frac{1}{u}} + c_i + \frac{\ln \ln \frac{1}{u}}{\sqrt{2 \ln \frac{1}{u}}} - \frac{d_i + \frac{1}{2}c_i^2}{\sqrt{2 \ln \frac{1}{u}}} + O(|\ln u|^{-1}). \end{aligned}$$

Using once again the estimate (5.7), we can now compute the asymptotics for F_i . \square

Since $\tau_i^2 > 0$, we can apply Proposition 5.5, to compute $\mathbb{P}[P_1 + \dots + P_n < z]$ as $z \rightarrow 0$.

Proposition 5.7.

$$\mathbb{P}[P_1 + \dots + P_n < z] \sim z^{\tau_I^\perp \Sigma_{II}^{-1} \tau_I} \tilde{L} \left(\frac{1}{z} \right) \nu(A)$$

with

$$\tilde{L} \left(\frac{1}{z} \right) = C \ln \frac{1}{z} \frac{1}{z}^{\tau_I^\perp \Sigma^{-1} \tau_I - \frac{|I|}{2}} \exp \left\{ \sqrt{2 \ln \frac{1}{z}} \sum_{i,j \in I} b_{ij} \tau_i \tau_j c_i \right\}$$

and

$$C = \frac{\exp \left\{ - \sum_{i,j \in I} b_{ij} \tau_i \tau_j (d_i + c_i c_j) \right\}}{(2\pi)^{-|I|/2} |\Sigma_{II}|^{\frac{1}{2}} \prod_{i \in I} \langle e_i, \Sigma_{II}^{-1} \beta_I \rangle}$$

Numerical examples In order to illustrate the accuracy of the asymptotic formula for the left-tail of the distribution of a portfolio containing call options modeled under the Black-Scholes framework, it has been compared to the left-tail behavior

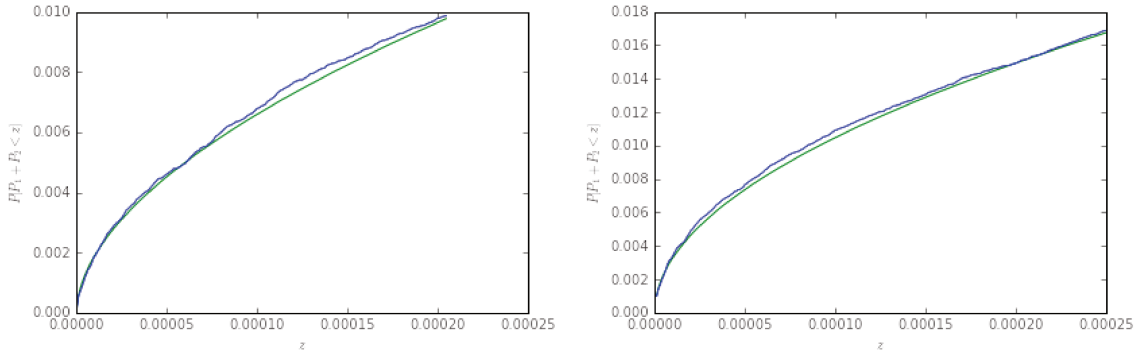


FIGURE 5.1. Left-tail behavior of a portfolio of two calls. The blue line indicates the behavior obtained by Monte Carlo simulation and the green one the behavior obtained by the asymptotic formula. Left graph: the two assets are supposed to have no correlation and $\sigma_1 = \sigma_2 = 0.3$. Right graph: the two assets are suppose to have a positive correlation $\rho = 0.1$ and $\sigma_1 = \sigma_2 = 0.2$.

obtained by Monte Carlo simulations (100000 scenarios) in two different cases. In the first case, we consider a portfolio of two calls with the following parameters : $\boldsymbol{\mu} = (0, 0)^\perp$, $t_0 = 0.8$, $T_1 = T_2 = 1$, and $k = 0$. The results are shown in Figure 5.1

In the second case, we consider a portfolio of 5 calls with the following parameters : $\boldsymbol{\mu} = (0, 0, 0, 0, 0)^\perp$, $t_0 = 0.8$, $T_1 = T_2 = 1$ and $k = 0$. The results are shown in Figure 5.2.

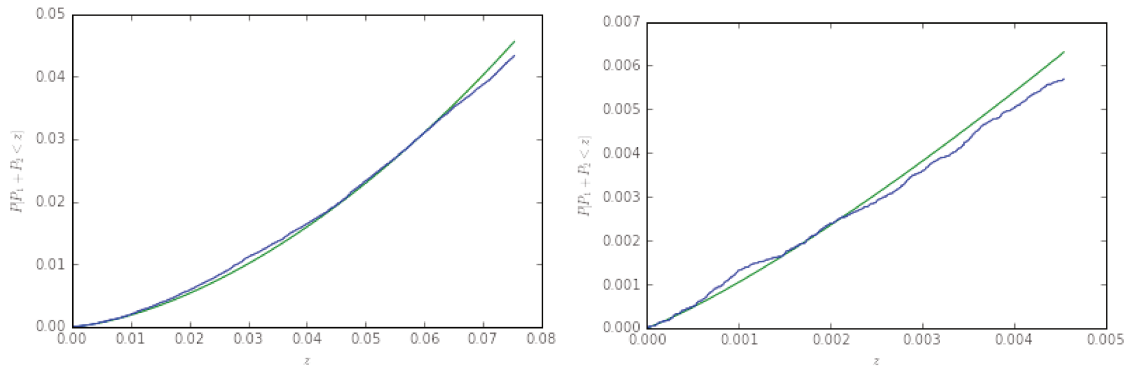


FIGURE 5.2. Left-tail behavior of a portfolio of five calls. The blue line indicates the behavior obtained by Monte Carlo simulation (100000 scenarios) and the green one the behavior obtained by the asymptotic formula. Left graph: the assets are supposed to have no correlation. Right graph: The correlation matrix is given by

3.2 Aggregate production of a wind farm

If we assume that X_i follows a Weibull distribution, then

$$F_i(x) = 1 - e^{-\left(\frac{x}{\lambda_i}\right)^{k_i}}$$

with λ_i and $k_i \in (0, \infty)$. Therefore, F_i is regularly varying at 0 with index k_i and Corollary 5.5 applies.

Proposition 5.8. *As $z \rightarrow 0$,*

$$\mathbb{P}[X_1 + \cdots + X_n < z] \sim z^{\sum_{i,j \in I} b_{ij} \sqrt{k_i k_j}} \tilde{L}\left(\frac{1}{z}\right) \nu(A)$$

where

$$\tilde{L}(z) = \exp \left\{ \frac{1}{2} \sum_{i,j \in I} b_{ij} \sqrt{k_i k_j} \log(\lambda_i \lambda_j) \right\} L(z)$$

Proof. In one hand, from Corollary 5.5, we deduce that as $z \downarrow 0$,

$$\mathbb{P}[X_1 + \cdots + X_n < z] = \exp \left\{ - \sum_{i,j \in I} b_{ij} \sqrt{\ln F_i(z) \ln F_j(z)} \right\} L\left(\frac{1}{z}\right) \nu(A)$$

On the other hand,

$$\begin{aligned} \ln F_i(z) \ln F_j(z) &= (k_i \ln x + k_i \ln \lambda_i + o(1)) (k_j \ln x + k_j \ln \lambda_j + o(1)) \\ &= k_i k_j (\ln x)^2 \left(1 - \frac{\ln(\lambda_i \lambda_j)}{\ln x} - \frac{\ln \lambda_i \ln \lambda_j}{\ln x^2} \right) + o(1) \end{aligned}$$

and finally

$$\sqrt{\ln F_i(z) \ln F_j(z)} = \sqrt{k_i k_j} \ln x - \frac{1}{2} \sqrt{k_i k_j} \ln \lambda_i \lambda_j + o(1)$$

which leads to the desired result. □