



# **Modélisation des rendements financiers à l'aide de la distribution de Laplace asymétrique généralisée**

**Mémoire**

**François Pelletier**

**Maîtrise en actuariat**  
Maître ès sciences (M.Sc.)

Québec, Canada

© François Pelletier, 2013



# Résumé

Les modèles classiques en finance sont basés sur des hypothèses qui ne sont pas toujours vérifiables empiriquement. L'objectif de ce mémoire est de présenter l'utilisation de la distribution de Laplace asymétrique généralisée comme une alternative intéressante à ces derniers. Pour ce faire, on utilise ses différentes propriétés afin de développer des méthodes d'estimation paramétrique, d'approximation et de test, en plus d'élaborer quelques principes d'évaluation de produits dérivés. On présente enfin un exemple d'application numérique afin d'illustrer ces différents concepts.



# Abstract

Classical models in finance are based on a set of hypotheses that are not always empirically verifiable. The main objective behind this master's thesis is to show that the generalized asymmetric Laplace distribution is an interesting alternative to those models. To support this idea, we focus on some of its properties to develop parametric estimation, approximation and testing methods, then we work out some principles of derivatives pricing. Finally, we have a numerical example to illustrate these concepts.



# Table des matières

Résumé	iii
Abstract	v
Table des matières	vii
Liste des tableaux	xi
Liste des figures	xiii
Remerciements	xvii
Introduction	1
<b>1 Les modèles de rendements financiers</b>	<b>3</b>
1.1 L'utilisation de modèles en finance . . . . .	3
1.1.1 Différents types de modèles . . . . .	3
1.1.2 Le risque de modélisation . . . . .	4
1.2 Les rendements financiers . . . . .	5
1.2.1 Définitions et notations . . . . .	5
1.2.2 Rendements cumulés . . . . .	7
1.2.3 Données disponibles . . . . .	8
1.3 Les premiers modèles . . . . .	8
1.3.1 Le modèle de Bachelier . . . . .	8
1.3.2 Proposition de Mandelbrot . . . . .	11
1.3.3 Le modèle de Press . . . . .	12
1.3.4 Le modèle de Praetz . . . . .	15
1.4 Conditions essentielles de Madan et Seneta . . . . .	16
<b>2 La distribution de Laplace asymétrique généralisée</b>	<b>19</b>
2.1 Le processus de Laplace . . . . .	19
2.1.1 Le processus gamma . . . . .	20
2.1.2 Le processus de Wiener . . . . .	21
2.1.3 Le processus de Laplace est un processus subordonné . . . . .	22
2.2 Distribution de Laplace asymétrique généralisée . . . . .	25
2.2.1 Fonction caractéristique . . . . .	25
2.2.2 Invariance d'échelle . . . . .	26

2.2.3	Fonctions génératrices . . . . .	28
2.2.4	Moments et rôle des paramètres . . . . .	28
2.2.5	Changement d'échelle et de localisation . . . . .	30
2.2.6	Représentation alternative et simulation . . . . .	32
2.2.7	Fonction de Bessel et densité . . . . .	33
2.3	Cas particuliers . . . . .	35
2.3.1	Distribution de Laplace asymétrique . . . . .	35
2.4	Relation avec le modèle de Madan et Seneta (1990) . . . . .	37
<b>3</b>	<b>Approximation de la densité et de la fonction de répartition</b>	<b>39</b>
3.1	L'approximation de Laplace . . . . .	39
3.2	L'approximation de Temme . . . . .	41
3.3	La méthode du point de selle . . . . .	42
3.3.1	Approximation de la densité . . . . .	42
3.3.2	Unicité du point de selle . . . . .	44
3.3.3	Approximation de la fonction de répartition . . . . .	44
3.3.4	Quelques propriétés des approximations . . . . .	45
3.4	Application de la méthode du point de selle . . . . .	46
3.4.1	Approximation de la densité . . . . .	46
3.4.2	Approximation de la fonction de répartition . . . . .	47
<b>4</b>	<b>Méthode des moments généralisée</b>	<b>49</b>
4.1	Introduction . . . . .	49
4.1.1	Méthode classique des moments . . . . .	50
4.2	Méthode des moments généralisée . . . . .	50
4.2.1	Définition . . . . .	51
4.2.2	Convergence . . . . .	52
4.2.3	Matrice de pondération optimale . . . . .	53
4.2.4	Méthode des moments généralisée itérative . . . . .	55
4.2.5	Distribution asymptotique des estimateurs . . . . .	55
4.3	Estimation sous contraintes . . . . .	58
4.3.1	Distribution asymptotique des estimateurs contraints . . . . .	59
4.4	Tests d'hypothèses paramétriques . . . . .	61
4.4.1	Test de Wald . . . . .	61
4.4.2	Test du multiplicateur de Lagrange . . . . .	62
4.4.3	Test basé sur la statistique de métrique de distance . . . . .	63
4.4.4	En résumé . . . . .	63
<b>5</b>	<b>Méthode de l'équation d'estimation optimale</b>	<b>65</b>
5.1	Équation d'estimation optimale . . . . .	67
5.2	Équation d'estimation optimale modifiée . . . . .	70
<b>6</b>	<b>Estimation des paramètres de la distribution de Laplace asymétrique gé-</b>	
	<b>néralisée</b>	<b>73</b>
6.1	Vecteur de paramètres initiaux . . . . .	73
6.2	Méthode des moments généralisée . . . . .	74
6.2.1	Matrice de pondération optimale . . . . .	75



6.2.2	Variance-covariance des paramètres . . . . .	76
6.2.3	Contraintes linéaires . . . . .	77
6.3	Méthode de l'équation d'estimation optimale . . . . .	78
<b>7</b>	<b>Tests statistiques</b>	<b>81</b>
7.1	Test de normalité . . . . .	81
7.1.1	Test de Shapiro-Wilk . . . . .	81
7.1.2	Test d'Epps-Pulley . . . . .	82
7.2	Tests d'adéquation . . . . .	82
7.2.1	Test $\chi^2$ de Pearson . . . . .	83
7.2.2	Test de Kolmogorov-Smirnov . . . . .	84
7.2.3	Test de distance minimale basé sur la fonction génératrice des moments	85
<b>8</b>	<b>Évaluation d'options</b>	<b>87</b>
8.1	Définitions . . . . .	87
8.1.1	Équation martingale . . . . .	88
8.1.2	Paramètres neutres au risque . . . . .	89
8.2	Aperçu du modèle de Black-Scholes . . . . .	90
8.3	Méthodes d'évaluation pour options européennes . . . . .	90
8.3.1	Méthode de Heston . . . . .	91
8.3.2	Méthode de Carr et Madan . . . . .	91
8.3.3	Prix d'exercice hors du cours . . . . .	93
8.3.4	Critique de la méthode de Carr-Madan . . . . .	93
8.3.5	Méthode d'Epps . . . . .	93
8.4	Particularités . . . . .	95
8.4.1	Option sur actions avec dividendes . . . . .	95
8.4.2	Options sur contrats à terme et taux de change . . . . .	95
<b>9</b>	<b>Exemple d'application</b>	<b>97</b>
9.1	Description des données . . . . .	97
9.2	Estimation . . . . .	99
9.3	Approximation . . . . .	104
9.4	Graphiques . . . . .	105
9.5	Tests statistiques . . . . .	108
9.6	Évaluation d'options . . . . .	109
	<b>Conclusion</b>	<b>113</b>
<b>A</b>	<b>Éléments de théorie des probabilités</b>	<b>115</b>
A.1	Définitions de base . . . . .	115
A.2	Transformées d'une variable aléatoire . . . . .	116
A.2.1	La fonction caractéristique . . . . .	116
	Transformée de Fourier . . . . .	116
	Définition . . . . .	116
	Les moments . . . . .	117
A.2.2	Inversion de la fonction caractéristique . . . . .	117
	La densité . . . . .	117

La fonction de répartition . . . . .	117
A.2.3 La fonction génératrice des moments . . . . .	118
A.2.4 La fonction génératrice des cumulants . . . . .	119
A.2.5 La transformée d'Esscher . . . . .	120
A.3 La transformée de Fourier rapide . . . . .	120
A.4 Processus de Lévy . . . . .	121
A.4.1 Définition et propriétés . . . . .	121
Représentation de Lévy-Khintchine . . . . .	122
Représentation de Lévy-Itô . . . . .	122
A.4.2 Processus subordonné . . . . .	123
A.5 Théorèmes d'intégration . . . . .	123
A.5.1 Théorème de convergence dominée de Lebesgue . . . . .	124
A.5.2 Théorème de Fubini . . . . .	124
<b>B Éléments de statistique mathématique</b>	<b>125</b>
B.1 Loi faible des grands nombres . . . . .	125
B.2 Théorème central limite . . . . .	125
B.2.1 Cas univarié . . . . .	126
B.2.2 Cas multivarié . . . . .	126
B.3 Méthode delta multivariée . . . . .	127
<b>C Données</b>	<b>129</b>
<b>Bibliographie</b>	<b>131</b>
<b>Contrat de partage</b>	<b>137</b>

# Liste des tableaux

1.1	Domaine et rôle des paramètres de la distribution L stable de Mandelbrot . . . . .	12
1.2	Respect des conditions émises par Madan et Seneta pour les différents modèles présentés . . . . .	17
2.1	Domaine et rôle des paramètres de la distribution de Laplace asymétrique généralisée	30
2.2	Cas spéciaux de la distribution de Laplace asymétrique généralisée . . . . .	35
2.3	Changements de paramétrisation . . . . .	37
4.1	Tests d’hypothèse paramétriques pour la méthode des moments généralisée . . . . .	63
5.1	Méthodes d’estimation représentables par la forme quadratique (5.0.6) . . . . .	67
6.1	Contraintes linéaires pour les cas particuliers de la distribution de Laplace asymétrique généralisée $GAL(\theta, \sigma, \mu, \tau)$ . . . . .	77
9.1	Statistiques d’ordre de l’échantillon $R_1$ . . . . .	97
9.2	Valeurs relatives aux premiers moments de l’échantillon $R_1$ . . . . .	98
9.3	Test de normalité d’Epps-Pulley pour $R_1$ . . . . .	98
9.4	Paramètres $\theta_1$ de la première optimisation . . . . .	100
9.5	Paramètres $\theta_1$ de la première optimisation . . . . .	102
9.6	Paramètres des données $R_1$ . . . . .	103
9.7	Approximation de la densité de $R_1$ . . . . .	104
9.8	Approximation de la fonction de répartition de $R_1$ . . . . .	105
9.9	Courbes de densité . . . . .	105
9.10	Valeur de l’intégrale de l’approximation de la densité par la méthode du point de selle . . . . .	105
9.11	Test du $\chi^2$ . . . . .	108
9.12	Test de Kolmogorov-Smirnov . . . . .	108
9.13	Test de distance minimale basé sur la fonction génératrice des moments . . . . .	109
9.14	Caractéristiques de l’option . . . . .	109
9.15	Paramètres neutres au risque . . . . .	109
C.1	Prix du titre Abbey National (penny sterling) du 31 juillet au 8 octobre 1991 . . . . .	129



# Liste des figures

1.1	Modèle de Bachelier : probabilité composée . . . . .	8
1.2	Distribution des rendements annuels de 40 titres boursiers, de 1890 à 1915, Table XVIII de Mitchell (1916) . . . . .	10
2.1	Premier incrément d'un processus subordonné . . . . .	23
2.2	Simulation d'un processus de Wiener subordonné par un processus gamma . . . . .	24
2.3	Fonction de densité de la distribution Laplace asymétrique généralisée avec différents paramètres : $GAL(y; \theta, \sigma, \kappa, \tau)$ . . . . .	31
2.4	Histogramme et estimateur de densité par noyau de 2500 réalisations de la variable aléatoire $Y \sim GAL(\theta = 0, \sigma = 1, \kappa = 2, \tau = 1)$ . . . . .	34
3.1	Probabilité à l'extrémité du support . . . . .	40
3.2	Distribution normale tronquée à $w_0$ . . . . .	41
9.1	Représentation en série chronologique de l'échantillon $R_1$ . . . . .	98
9.2	Distribution de la variable aléatoire $R_1$ . . . . .	99
9.3	Graphique Quantile-Quantile . . . . .	100
9.4	Équation du point de selle pour $r = 0.01$ . . . . .	104
9.5	Densité de $R_1^*$ selon la méthode des moments généralisée . . . . .	106
9.6	Densité de $R_1^*$ selon la méthode de l'équation d'estimation optimale . . . . .	107
9.7	Prix de l'option selon les paramètres estimés avec la méthode des moments généralisée	110
9.8	Prix de l'option selon les paramètres estimés avec la méthode de l'équation d'estimation optimale . . . . .	111



Dieu ne se soucie pas de nos  
difficultés mathématiques. Il  
intègre empiriquement.

---

Albert Einstein





# Remerciements

Je dois avouer que je ne savais pas trop à quoi m'attendre au moment où j'ai commencé la rédaction de ce mémoire. En essayant de ne pas oublier personne, j'aimerais tout d'abord remercier celui qui m'a accompagné durant ces deux dernières années et qui m'a énormément appris, le professeur Andrew Luong.

J'aimerais aussi remercier mes évaluateurs, les professeurs Claire Bilodeau et Julien Trufin. J'aimerais aussi remercier le professeur Vincent Goulet, qui m'a fortement encouragé de m'inscrire à la maîtrise, et pour les nombreuses références qu'il a pu me fournir, pour la rédaction entre autres, qui m'ont grandement facilité la tâche. Je remercie tout autant le professeur Ghislain Léveillé pour son soutien et pour m'avoir offert la possibilité de présenter mes résultats.

J'aimerais aussi remercier l'ensemble des étudiants gradués, des professeurs et du personnel de l'École d'Actuariat de l'Université Laval, ainsi que mes collègues de travail de la CARRA, pour leur soutien et leurs encouragements.

Je suis reconnaissant envers l'École d'Actuariat et la Chaire d'Actuariat de l'Université Laval pour le soutien financier en m'offrant des postes d'auxiliaires d'enseignement et de recherche.

Enfin, j'aimerais remercier mes parents, ma soeur et l'ensemble de ma famille et de mes amis pour avoir accepté que je passe plus de temps à «écrire des formules bizarres sur l'ordinateur» qu'à être avec eux.



# Introduction

L'utilisation de la distribution de Laplace asymétrique généralisée dans le cadre de la modélisation des rendements financiers est de plus en plus courante dans les milieux académiques et pratiques. Entre autres, elle présente une alternative intéressante au modèle de Black-Scholes pour l'évaluation des options de type européennes. De plus, elle peut être utilisée pour modéliser les rendements obligataires à long terme et les variations du taux de change. Cependant, la littérature actuelle présente peu d'outils qui facilitent l'utilisation en pratique de cette distribution. La principale motivation derrière le travail de recherche et de synthèse présenté dans ce texte a été de développer certains outils qui permettent l'estimation des paramètres de la distribution et l'approximation de celle-ci. De plus, plusieurs tests statistiques sont présentés afin d'évaluer la validité du modèle estimé et de certaines contraintes linéaires qu'on pourrait lui appliquer.

Au chapitre 1, on présente les différents types de modèles financiers ainsi que le risque associé à la modélisation. On introduit aussi différents concepts théoriques entourant les rendements financiers. Puis, on dresse un historique des modèles financiers ayant mené à l'utilisation de la distribution de Laplace asymétrique généralisée.

Au chapitre 2, on introduit les processus gamma et de Wiener qui sont les deux composantes essentielles du processus de Laplace. Puis, on présente les principales caractéristiques de la distribution de Laplace asymétrique généralisée. Enfin, on présente quelques cas particuliers de celle-ci et on fait le lien avec le modèle variance-gamma de Madan et Seneta.

Au chapitre 3, on présente la méthode du point de selle qui permet d'effectuer l'approximation de la densité et de la fonction de répartition lorsqu'elles n'ont pas de forme analytique. On applique ensuite cette méthode à la distribution de Laplace asymétrique généralisée.

Au chapitre 4, on présente la méthode des moments généralisée ainsi que son utilisation dans le cadre de l'estimation sous contraintes. On présente ensuite certains tests d'hypothèse paramétriques.

Au chapitre 5, on présente la méthode des équations d'estimation optimales ainsi qu'une version légèrement modifiée qui diminue la quantité de calcul nécessaire.

Au chapitre 6, on applique les méthodes d'estimation des deux chapitres précédents à la distribution de Laplace asymétrique généralisée. On présente en premier lieu une méthode qui permet d'obtenir un point de départ pour l'algorithme d'optimisation. Puis, on détaille les résultats obtenus.

Au chapitre 7, on présente les tests de normalité de Shapiro-Wilk et d'Epps-Pulley, puis ceux d'adéquation du  $\chi^2$  de Pearson, celui de Kolmogorov-Smirnov ainsi qu'un autre basé sur la fonction génératrice des moments.

Au chapitre 8, on présente différentes méthodes pour l'évaluation d'options. On présente d'abord quelques notions liées aux produits dérivés, puis on détaille trois méthodes pouvant être utilisées avec la distribution de Laplace asymétrique généralisée. Enfin, on présente quelques particularités liées à certains titres financiers.

Au chapitre 9, on présente un exemple d'application des outils développés aux chapitres 3 à 8 avec un ensemble de données.

Enfin, on présente en annexe certaines notions de théorie des probabilités et de statistique.

# Chapitre 1

## Les modèles de rendements financiers

### 1.1 L'utilisation de modèles en finance

On doit considérer les implications de l'utilisation de modèles en finance avant d'entreprendre leur étude. On doit aussi prendre connaissance des différents types ainsi que les risques liés à chacun d'entre eux. Pour ce faire, on se réfère à la note «Model Risk» publiée par Derman (1996).

Durant les dernières décennies, plusieurs modèles sont apparus afin de fournir une approche fondamentale aux concepts de tarification, d'offre et de demande et d'arbitrage aux intervenants des milieux financiers. Au cours des années 1970, on se préoccupe particulièrement des fluctuations des taux d'intérêt, un phénomène qui marque cette époque. Les notions de durée et de convexité font alors leurs débuts. Sur les marchés de capitaux propres, on s'intéresse à la discordance entre le prix négocié des contrats à terme et le prix raisonnable calculé selon une perspective théorique.

Puis, la confiance développée envers le modèle de tarification d'options de Black et Scholes (1973) et ses extensions a favorisé la croissance du marché des produits dérivés. La puissance de calcul croissante des ordinateurs a aussi permis l'élaboration et l'utilisation de modèles de plus en plus sophistiqués. La dépendance qui peut se développer envers ceux-ci apporte son lot de considérations. On doit donc se rappeler l'utilisation désirée par les auteurs de ceux-ci et le risque associé à leur usage à grande échelle.

#### 1.1.1 Différents types de modèles

Toujours selon Derman, un modèle financier peut être classé parmi au moins trois catégories :

1. Le **modèle fondamental**, basé sur un système de postulats et de données, entre lesquels on peut établir différentes relations. Le modèle de Black-Scholes en est un exemple.

2. Le **modèle phénoménologique**, qui présente une description ou une analogie, afin d'illustrer quelque chose qui ne peut être directement observé. C'est un modèle moins fondamental, basé aussi sur des liens de cause à effet. Un modèle qui chercherait à expliquer l'impact du retrait du porteur de parts majoritaire d'une entreprise sur la valeur des actions de celle-ci serait phénoménologique.
3. Le **modèle statistique**, basé sur une régression ou un réglage optimal entre différents ensembles de données. On ne cherche pas ici à expliquer une dynamique, mais à décrire une tendance ou une corrélation. Le modèle d'évaluation des actifs financiers et celui des trois facteurs de Fama et French (1993) en sont des exemples.

Un modèle financier est en partie basé sur des variables qui représentent des opinions et des anticipations, et non seulement des quantités mesurables. Ces variables peuvent être, entre autres, le rendement et la volatilité future espérés. Cette considération sera importante notamment lorsque l'on voudra déterminer le prix raisonnable d'un produit dérivé. En effet, un modèle de tarification est essentiellement un moyen de refléter l'intuition des acteurs du marché à propos de ces variables sous la forme d'un prix exprimé dans une unité monétaire. Un bon modèle doit faciliter l'extrapolation de ce prix sous certaines conditions de marché.

Contrairement à la physique classique, un principe fondamental en finance est l'incertitude. On ne peut anticiper la valeur d'un titre à un moment donné dans le futur avec la même précision qu'on peut prévoir la position d'un objet à cet instant. Les outils mathématiques principalement utilisés seront alors les processus stochastiques, les statistiques et les distributions de probabilités, en plus du calcul différentiel et intégral.

### 1.1.2 Le risque de modélisation

Plusieurs risques inhérents à la modélisation en finance existent. Quelques-uns d'entre eux seront décrits dans cette section.

La modélisation peut tout simplement ne pas être applicable à la situation étudiée. L'exemple le plus probant serait de tenter de prévoir les mouvements du prix d'un titre financier à court terme.

Un modèle peut être incorrect pour plusieurs raisons. Entre autres, il peut ignorer certains facteurs ou poser une hypothèse déterministe inappropriée sur ceux-ci. Il peut aussi considérer une dynamique incorrecte pour un des facteurs ou encore une relation inappropriée entre ceux-ci. Enfin, il peut n'être applicable que sous certaines conditions bien précises ou encore que son utilisation soit limitée à court terme, notamment lorsqu'il nécessite un temps de calibration pour être statistiquement valable. Il peut aussi être inutilisable par une mauvaise estimation des paramètres.

Un modèle peut aussi être correct, mais avoir une solution erronée. Cela se produit notamment lorsqu'on tente de dériver une solution analytique ou que l'on doit utiliser des méthodes numériques pour obtenir celle-ci. On se doit, dans ce cas, de connaître l'erreur maximale possible de la méthode utilisée. Un modèle correct peut aussi être utilisé dans le mauvais contexte. Par exemple, on pourrait avoir recours à des paramètres inadéquats de simulation, ou encore réutiliser le modèle dans une autre situation sans tenir compte des conditions de validité de celui-ci.

Son utilisation peut générer des prix déraisonnables ; on parle alors d'arbitrage de modèle. Par exemple, si un titre est évalué à l'aide du modèle d'évaluation des actifs financiers, son prix sera différent de celui qui serait obtenu avec la régression à trois facteurs de Fama et French. Un investisseur peut alors faire du profit en achetant le titre à celui qui demande le prix le plus faible pour le revendre à celui qui offre le plus élevé.

L'utilisation de données instables peut produire des résultats différents selon la période étudiée. La possibilité qu'une estimation basée sur des données historiques soit erronée doit être considérée.

Enfin, comme la plupart des modèles financiers sont implémentés sous forme de logiciels, différents bogues informatiques peuvent se retrouver dans le code source. On considère entre autres des erreurs d'arrondissement, de logique et de clarté du code, ainsi que des particularités du matériel qui n'auraient pas été prises en compte par le programmeur. Ces erreurs peuvent être difficiles à détecter, c'est pourquoi un grand nombre de tests devrait être effectués avant de publier un logiciel de modélisation financière.

## 1.2 Les rendements financiers

Le **rendement** est défini comme étant le gain ou la perte de valeur d'un actif sur une période donnée. Il est constitué des revenus occasionnés et des gains en capitaux d'un investissement et est habituellement représenté sous la forme d'un pourcentage. Ces derniers peuvent prendre la forme de coupons pour les titres à revenus fixes et de dividendes pour les actions échangées sur les marchés boursiers. On ne considèrera, dans ce texte, que les titres boursiers sans dividende, dont le rendement est lié uniquement aux gains en capitaux.

### 1.2.1 Définitions et notations

On définit le prix  $S(t) > 0$  d'un titre financier observé au temps  $t$ . Implicitement, le prix considéré est celui à la fermeture. On définit aussi le taux de rendement effectif  $R(t)$  sur une période comprise dans l'intervalle de temps  $[t - 1, t]$ . C'est le taux composé continu, aussi appelé force d'intérêt, qui aurait occasionné les mêmes gains ou pertes sur un montant déposé

en banque au cours de la période concernée. Le taux de rendement est la variable d'intérêt dans le contexte de la modélisation financière.

On associe le taux de rendement effectif à la différence entre le logarithme du prix initial et final. Dans la situation où le taux de rendement est déterministe et non aléatoire, on obtient l'équation différentielle suivante :

$$\frac{dS(t)}{dt} = R(t) \cdot S(t).$$

On peut interpréter cette équation en affirmant que la variation du prix  $dS(t)$  sur un intervalle de temps infiniment petit  $dt$  est proportionnelle à la valeur actuelle  $S(t)$ . Cette équation différentielle a pour solution générale :

$$S(t) = S(0)e^{R(t) \cdot t}. \quad (1.2.1)$$

Afin de définir les propriétés de l'échantillon sélectionné, on pose comme hypothèse :

**Hypothèse 1.1.** *Le rendement  $R(t)$  est constant durant la période définie par l'intervalle de temps  $[t-1, t]$ , mais il est différent d'une à l'autre :  $R(s) \neq R(t), s \neq t$ .*

On peut alors représenter le rendement  $R(t)$  comme étant la différence entre les logarithmes des prix observés au temps  $t$  et  $t-1$ , ou encore le logarithme du quotient de ces mêmes prix :

$$\begin{aligned} R(t) &= \ln(S(t)) - \ln(S(t-1)) \\ &= \ln\left(\frac{S(t)}{S(t-1)}\right). \end{aligned} \quad (1.2.2)$$

On définit aussi le **rendement cumulé**  $L(t)$ . Il correspond à la somme des rendements effectifs observés sur l'intervalle  $[0, t]$  :

$$\begin{aligned} L(t) &= \sum_{i=1}^t R(i) \\ &= \sum_{i=1}^t [\ln(S(i)) - \ln(S(i-1))] \\ &= \ln(S(t)) - \ln(S(0)) \\ &= \ln\left(\frac{S(t)}{S(0)}\right). \end{aligned} \quad (1.2.3)$$

Cette représentation permet d'exprimer le prix actuel  $S(t)$  en fonction de la valeur initiale  $S(0)$  sous une forme similaire à la solution (1.2.1), mais tenant compte de l'hypothèse émise



précédemment :

$$\begin{aligned} e^{L(t)} &= \frac{S(t)}{S(0)} \\ S(t) &= S(0) \cdot e^{L(t)} \\ &= S(0) \cdot \exp\left(\sum_{i=1}^t R(i)\right). \end{aligned} \tag{1.2.4}$$

### 1.2.2 Rendements cumulés

On pose l'hypothèse suivante :

**Hypothèse 1.2.** *Les rendements  $R(i), i \in 1, \dots, t$  sont indépendants, mais pas nécessairement identiquement distribués.*

On peut alors obtenir la distribution du rendement cumulé  $L(t)$  en utilisant le produit de convolution (A.2.12). Considérons  $\phi_{R(i)}(\xi)$  la fonction caractéristique d'un rendement  $R(i)$  et  $\phi_{L(t)}(\xi)$  celle du cumulé  $L(t)$ . On obtient alors que cette dernière est égale au produit des fonctions caractéristiques des rendements effectifs sur chacune des périodes de l'intervalle  $[0, t]$  :

$$\phi_{L(t)}(\xi) = \prod_{i=1}^t \phi_{R(i)}(\xi). \tag{1.2.5}$$

On considère la situation où l'on posera plutôt l'hypothèse suivante :

**Hypothèse 1.3.** *Les rendements  $R(i), i \in 1, \dots, t$  sont à la fois indépendants et identiquement distribués.*

Alors, la fonction caractéristique des rendements est égale pour chaque période :

$$\phi_R(\xi) = \phi_{R(1)}(\xi) = \dots = \phi_{R(t)}(\xi). \tag{1.2.6}$$

On peut donc simplifier l'expression (1.2.5) pour obtenir la fonction caractéristique :

$$\phi_{L(t)}(\xi) = [\phi_R(\xi)]^t. \tag{1.2.7}$$

Considérer une distribution qui est fermée sous la convolution pour modéliser les rendements sur une période  $R(i)$  peut alors être intéressant. Le rendement cumulé  $L(t)$  pourra aussi être modélisé à l'aide de la même distribution. Pour ce faire, on modifie un paramètre d'échelle en fonction de la longueur  $t$  de l'intervalle de temps considéré.

### 1.2.3 Données disponibles

Les données disponibles auprès des fournisseurs d'informations financières prennent habituellement la forme de séries chronologiques discontinues. Celles-ci incluent les prix à l'ouverture, le plus bas et le plus élevé au courant de la journée ainsi qu'à la fermeture, pour chaque jour où les marchés financiers sont en activité. Afin de mesurer le rendement quotidien d'un titre, seuls les prix à la fermeture seront considérés.

## 1.3 Les premiers modèles

### 1.3.1 Le modèle de Bachelier

Un des premiers modèles proposés afin de représenter les rendements financiers a été celui de Bachelier (1900).

Le prix d'un titre peut varier, durant une période, de n'importe quelle valeur comprise dans l'intervalle  $[-S(t), \infty]$ . Il propose donc que cet intervalle soit remplacé par l'ensemble du domaine réel  $\mathbb{R}$ . La probabilité que le titre atteigne une valeur nulle ou négative ou que celle-ci double devrait donc être négligeable. Il ajoute aussi que la variation est indépendante du prix actuel du titre  $S(t)$  et que la distribution de probabilités de celle-ci est symétrique et centrée en ce point.

Il utilise le principe selon lequel la probabilité que deux événements indépendants consécutifs aient lieu est le produit de celles que chacun d'entre eux se réalise, pour établir la distribution des variations du prix. Par exemple, la variation du prix sur une première période prend la valeur  $x$  et celle sur une seconde,  $z - x$ , comme illustré à la figure 1.1.

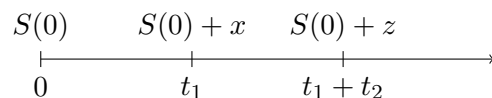


FIGURE 1.1: Modèle de Bachelier : probabilité composée

On définit  $f(x, t)$  la fonction de densité de la variation du prix  $S(t)$  par rapport au niveau initial  $S(0)$ . Alors, selon le principe précédent, on obtient l'expression

$$f(z, t_1 + t_2) = \int_{-\infty}^{\infty} f(x, t_1) \cdot f(z - x, t_2) \cdot dx. \quad (1.3.1)$$

La solution proposée est que la densité de probabilité soit de la forme

$$f(x, t) = A \cdot \exp \{-B^2 x^2\}.$$

Afin que la fonction  $f(x, t)$  soit une densité de probabilité, la condition suivante doit être respectée :

$$\int_{-\infty}^{\infty} A \cdot \exp \{-B^2 x^2\} dx = 1. \quad (1.3.2)$$

Ceci implique que

$$B = A\sqrt{\pi}.$$

En posant  $x = 0$ , on a  $A = f(0, t)$  et l'on en déduit :

$$f(x, t) = f(0, t) \cdot \exp \{-\pi \cdot f(0, t)^2 \cdot x^2\}. \quad (1.3.3)$$

En reprenant l'intégrale (1.3.1), on obtient que la densité de probabilité  $f(z, t_1 + t_2)$  soit aussi de la forme (1.3.2) :

$$f(z, t_1 + t_2) = \frac{f(x, 1)f(z - x, 2)}{\sqrt{f(x, 1)^2 + f(z - x, 2)^2}} \exp \left\{ -\pi \frac{f(x, 1)f(z - x, 2)}{f(x, 1)^2 + f(z - x, 2)^2} z^2 \right\}. \quad (1.3.4)$$

On reconnaitra que cette densité est, à un changement de variable près, une loi normale. La démarche suggère qu'il recherchait une distribution qui était fermée sous la convolution, une propriété souhaitable pour un modèle cohérent des rendements financiers.

Ce modèle implique un processus de Wiener-Bachelier selon lequel les incréments, ou les changements de prix, suivent une distribution normale :

$$S(T) - S(t) \sim N(0, \sigma^2 (T - t)). \quad (1.3.5)$$

On doit noter que ce modèle implique que la variance des fluctuations n'est pas proportionnelle au prix initial. Une première correction sera apportée au modèle afin de considérer le logarithme du prix. Ce changement permettra d'obtenir un modèle où elle est désormais proportionnelle au prix initial. Le processus du prix suivra alors un mouvement brownien géométrique :

$$S(T) - S(t) \sim LN(0, \sigma(T - t)). \quad (1.3.6)$$

Le logarithme du prix suivra alors un processus de Wiener-Bachelier :

$$\ln(S(T)) - \ln(S(t)) \sim N(0, \sigma(T - t)). \quad (1.3.7)$$

Un des principaux avantages du processus de Bachelier modifié est que le rendement cumulé  $L(t)$  est aussi une variable aléatoire gaussienne. Cette propriété est appelée L-stabilité ou

invariance sous l'addition. La distribution gaussienne est la seule ayant cette propriété où le second moment est fini. Le sujet des distributions L stables sera aussi abordé à la section 1.3.2.

Quelques années après sa publication, ce modèle est l'objet de critiques de la part d'économistes et de financiers. En se référant à Mitchell (1916), on observe que, sur une base annuelle, les variations négatives par rapport à la moyenne (149) sont plus fréquentes que celles qui sont positives (126), pour un ensemble de 40 titres boursiers, entre 1890 et 1915 (figure 1.2). Une asymétrie négative des rendements sera alors présente.

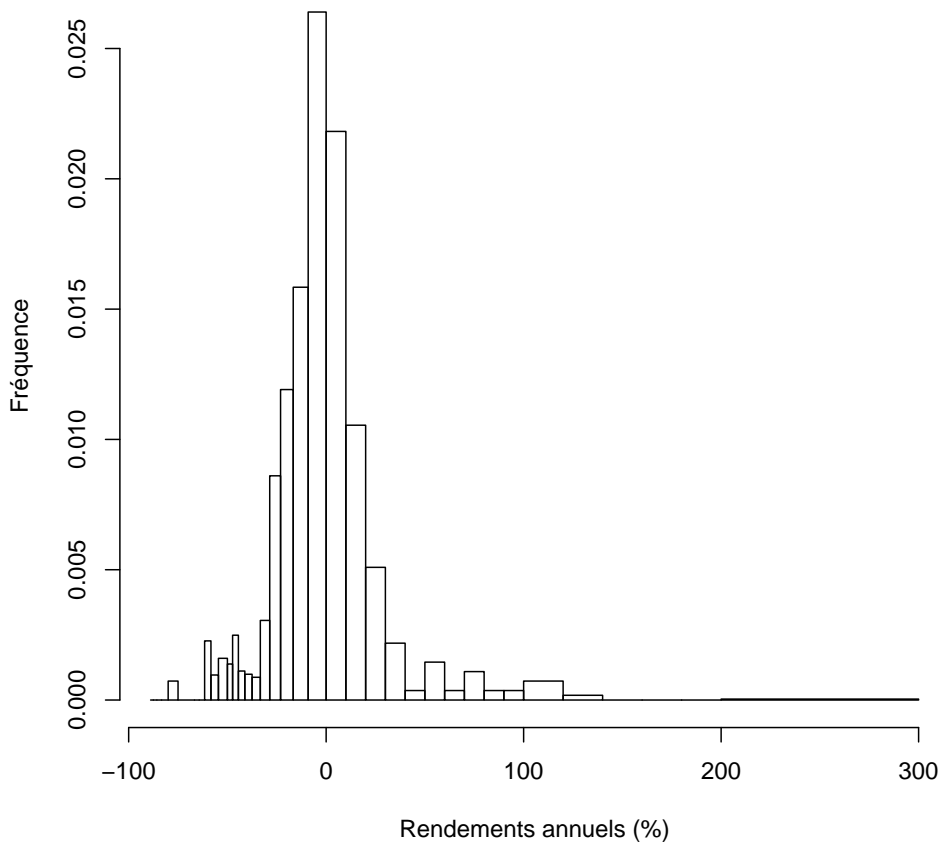


FIGURE 1.2: Distribution des rendements annuels de 40 titres boursiers, de 1890 à 1915, Table XVIII de Mitchell (1916)

De plus, les variations extrêmes sont plus fréquentes que ne pourrait le prédire un modèle basé sur un mouvement brownien. La distribution des rendements aurait donc des queues plus épaisses<sup>1</sup> que la normale. On doit trouver un modèle qui permet de tenir compte de ces

1. traduction de l'anglais heavy tailed

particularités.

### 1.3.2 Proposition de Mandelbrot

Mandelbrot (1963) propose un modèle qui vise à combler les lacunes du processus brownien géométrique (1.3.6). Il explique que les distributions empiriques des changements de prix sont habituellement trop *pointues* pour être considérées comme des échantillons d'une population gaussienne.

Il identifie différentes caractéristiques qu'un bon modèle des rendements financiers devrait posséder :

1. Il doit tenir compte de la fréquence des grands changements de prix. Il doit donc être basé sur une distribution leptocurtique, plus pointue au centre que la normale.
2. Il doit permettre des changements instantanés et imprévisibles de toute amplitude.
3. Il doit admettre une probabilité non nulle que plusieurs changements consécutifs semblent corrélés.
4. Il doit admettre un processus de prix non stationnaire, car la variance échantillonnale prend différentes valeurs à travers le temps.

La famille de distributions L stable semble être celle qui répond le mieux à l'ensemble de ces conditions (Walter, 1995). L'équation suivante définit la propriété de L-stabilité de la distribution de la variable aléatoire des rendements sur une période  $R$  :

$$(a_1 R_1 + b_1) + (a_2 R_2 + b_2) \stackrel{d}{=} aR + b \quad (1.3.8)$$

$$\forall a_1, a_2 > 0, \forall b_1, b_2. \quad (1.3.9)$$

La solution générale de cette équation a été découverte par Lévy en 1925. Le logarithme de la fonction caractéristique de celle-ci prend la forme suivante :

$$\ln(\phi_R(\xi)) = i\delta\xi - \gamma|\xi|^\alpha \left[ 1 + \frac{i\beta\xi}{|\xi|} \tan \frac{\alpha\pi}{2} \right]. \quad (1.3.10)$$

Le domaine et le rôle des paramètres de la distribution L stable est décrit à la table 1.1. La flexibilité apportée par les quatre paramètres permet de remplir les quatre conditions établies au début de cette section. De plus, l'absence, dans la majorité des cas, de moments finis d'ordre supérieur à l'espérance, permet de tenir compte du mouvement erratique des prix et ainsi produire de larges discontinuités de son processus. Elle permet aussi d'expliquer l'apparence de corrélation sérielle, en considérant une probabilité non négligeable que cette caractéristique soit présente. Cependant, ce modèle est difficile à appliquer à l'évaluation de

Paramètre	Domaine	Rôle	Observations
$\alpha$	$]0, 2]$	Aplatissement	Plus sa valeur est petite, plus la distribution est leptocurtique. $\alpha = 2$ correspond à la distribution normale.
$\beta$	$] -1, 1]$	Asymétrie	Défini seulement lorsque $\alpha \neq 1$ . Lorsque $\alpha = 1$ et $\beta = 0$ , on obtient la distribution de Cauchy.
$\gamma = s^\alpha$	$\mathbb{R} \setminus \{0\}$	Échelle	On doit prendre la racine $\alpha$ pour obtenir un paramètre d'échelle $s$ tel que défini par Pearson.
$\delta$	$\mathbb{R}$	Localisation	

TABLE 1.1: Domaine et rôle des paramètres de la distribution L stable de Mandelbrot

produits dérivés pour cette raison, étant donné que l'on devra être en mesure de quantifier la volatilité.

L'approche classique, selon Mandelbrot, pour expliquer les grands changements de prix a été de considérer un mélange de deux distributions normales, dont une pour les fluctuations régulières et une qui a une variance plus importante, pour les discontinuités. Il remarque que pour expliquer adéquatement le comportement des données empiriques, on doit introduire un mélange de plusieurs distributions normales, ce qui rendrait le modèle plus complexe. Par contre, on retrouve une approche intéressante avec le modèle présenté à la section suivante.

### 1.3.3 Le modèle de Press

Press (1967) propose un modèle statistique basé sur un processus de Poisson composé auquel on ajoute un mouvement brownien  $W(t)$ . C'est donc d'un processus ayant des incréments stationnaires et indépendants. Il présente donc les caractéristiques d'un processus de Lévy. Press utilise aussi la transformation logarithmique (1.3.6) afin que la variation soit proportionnelle au prix. Il remarque aussi que le modèle logarithmique de Bachelier est inadéquat, car il ne tient pas compte des queues de la distribution empirique des rendements qui sont plus épaisses que celles de la normale. Il ajoute que le modèle proposé par Mandelbrot est discutable, car il ne trouve aucune évidence, à partir des données observées, que la distribution de la population aurait une variance infinie.

Le processus de Poisson  $\{N(t)\}$  de paramètre  $\lambda t$  est un processus de comptage qui détermine les occurrences des sauts  $Y_k, k = 1, \dots, N(t)$ . Ces sauts surviennent généralement lorsqu'une information importante est rendue publique par rapport à un titre. Ceux-ci sont aussi de distribution normale, mais leur espérance n'est pas nulle et leur variance est différente de celle du processus  $W(t)$ . Cette composante que l'on ajoute au modèle de Bachelier modifié permet d'expliquer les variations plus importantes et moins fréquentes observées empiriquement.

Le processus du logarithme du prix  $\{s(t)\} \equiv \{\ln(S(t))\}$  est donc représenté par l'équation suivante :

$$s(t) = s(0) + \sum_{k=1}^{N(t)} Y_k + W(t). \quad (1.3.11)$$

On définit les différentes variables aléatoires composant le processus comme suit :

$$\begin{aligned} Y_k &\sim N(\theta, \sigma_2^2) \\ W(t) &\sim N(0, \sigma_1^2 t) \\ N(t) &\sim \text{Poisson}(\lambda t) \end{aligned}$$

Comme pour la plupart des processus de Lévy, on ne peut obtenir une forme explicite pour la fonction de densité, car celle-ci se présente sous la forme d'une série infinie. On représente alors ces processus par leur fonction caractéristique, formée par le produit de celles de leurs différentes composantes.

La distribution du logarithme du prix  $s(t)$  est définie par la fonction caractéristique  $\phi_{s(t)}(\xi)$ , qui est le produit de celle de la constante et celles des processus de Wiener et de Poisson composé :

$$\begin{aligned} \phi_{s(t)}(\xi) &= E \left[ e^{i\xi s(t)} \right] \\ &= \exp \{ i\xi \cdot s(0) \} \times \exp \left\{ -\frac{t\sigma_1^2 \xi^2}{2} \right\} \times \exp \left\{ \lambda t \left[ e^{i\theta\xi - (\sigma_2^2 \xi^2 / 2)} - 1 \right] \right\} \\ &= \exp \left\{ i\xi \cdot s(0) - \frac{t}{2} \sigma_1^2 \xi^2 + \lambda t \left[ e^{i\theta\xi - (\sigma_2^2 \xi^2 / 2)} - 1 \right] \right\}. \end{aligned} \quad (1.3.12)$$

Afin d'estimer le modèle, on s'intéressera plutôt à la distribution d'un incrément  $\Delta s(t) = s(t) - s(t-1)$  de ce processus. La fonction caractéristique  $\phi_{\Delta s(t)}(\xi)$  de cette variable aléatoire peut être facilement identifiée à partir de celle du processus (1.3.12). Essentiellement, on pose  $s(0) = 0$  et  $t = 1$ , pour obtenir :

$$\begin{aligned} \phi_{\Delta s(t)}(\xi) &= E \left[ e^{i\xi \Delta s(t)} \right] \\ &= \exp \left\{ -\frac{\sigma_1^2 \xi^2}{2} + \lambda \left[ e^{i\xi\theta - (\sigma_2^2 \xi^2 / 2)} - 1 \right] \right\}. \end{aligned} \quad (1.3.13)$$

Pour estimer les paramètres du modèle, on privilégie la méthode des cumulants, qui est similaire à la méthode des moments. Considérons les quatre premiers cumulants de la distribution

de l'incrément  $\Delta s(t)$  :

$$K_1 = \lambda\theta \quad (1.3.14a)$$

$$K_2 = \sigma_1^2 + \lambda(\theta^2 + \sigma_2^2) \quad (1.3.14b)$$

$$K_3 = \lambda\theta(\theta^2 + 3\sigma_2^2) \quad (1.3.14c)$$

$$K_4 = \lambda(\theta^4 + 6\theta^2\sigma_2^2 + 3\sigma_2^4). \quad (1.3.14d)$$

En utilisant les quatre premiers cumulants empiriques (A.2.18), on obtient les équations suivantes :

$$0 = \hat{\theta}^4 - \frac{\bar{K}_3}{\bar{K}_1}\hat{\theta}^2 + \frac{3\bar{K}_4}{2\bar{K}_1}\hat{\theta} - \frac{\bar{K}_3^2}{2\bar{K}_1^2} \quad (1.3.15a)$$

$$\hat{\lambda} = \frac{\bar{K}_1}{\hat{\theta}} \quad (1.3.15b)$$

$$\hat{\sigma}_2^2 = \frac{\bar{K}_3 - \hat{\theta}^2\bar{K}_1}{3\bar{K}_1} \quad (1.3.15c)$$

$$\hat{\sigma}_1^2 = \bar{K}_2 - \frac{\bar{K}_1}{\hat{\theta}} \left( \hat{\theta}^2 + \frac{\bar{K}_3 - \bar{K}_3\theta^2}{3\bar{K}_1} \right). \quad (1.3.15d)$$

En résolvant numériquement l'équation (1.3.15a) pour le  $\hat{\theta}$ , puis par substitutions successives dans les équations (1.3.15), on obtient des estimateurs convergents pour les quatre paramètres du modèle.

Un modèle similaire a aussi été présenté par Merton (1976), cependant, il inclut un paramètre de dérive  $\alpha$ , et considère que les sauts  $Y$ , qui sont des facteurs multiplicatifs, peuvent suivre une autre distribution que la normale. Il présente le modèle sous la forme d'une équation différentielle stochastique :

$$\frac{dS}{S} = (\alpha - \lambda k)dt + \sigma dW + dq. \quad (1.3.16)$$

La constante  $k$  représente l'espérance de la variation relative si un saut se produit et  $q$ , le processus de Poisson composé. La solution de cette équation est, selon le lemme d'Itô :

$$S(t) = \tilde{S}(0) \exp \left\{ \left( \alpha - \frac{1}{2}\sigma^2 - \lambda k \right) t + \sigma W(t) \right\} \quad (1.3.17)$$

où

$$\tilde{S}(0) = \begin{cases} S(0) & \text{si } N(t) = 0 \\ S(0) \sum_{k=1}^{N(t)} Y_k & \text{si } N(t) \geq 1. \end{cases}$$

En spécifiant un paramètre de dérive  $\delta = \alpha - \frac{1}{2}\sigma^2 - \lambda k$  et en considérant que les sauts  $Y$  sont de distribution lognormale, on peut réécrire la fonction caractéristique d'un incrément (1.3.13)



du modèle de Press :

$$\begin{aligned}\phi_{\Delta s(t)}(\xi) &= E \left[ e^{i\xi\Delta s(t)} \right] \\ &= \exp \left\{ i\delta\xi - \frac{\sigma^2\xi^2}{2} + \lambda \left[ e^{i\xi\theta - (\sigma^2\xi^2/2)} - 1 \right] \right\}.\end{aligned}\quad (1.3.18)$$

L'utilisation de ce modèle présente deux désavantages. L'estimation du modèle est difficile lorsque la moyenne s'approche de 0, car le quotient (1.3.15b) tend alors vers une indétermination. De plus, contrairement à d'autres modèles, il est difficile d'identifier le rôle des paramètres par rapport à un moment en particulier (classification de Pearson), contrairement à ce qu'on pourra observer avec la distribution de Laplace asymétrique généralisée.

### 1.3.4 Le modèle de Praetz

Praetz (1972) propose un modèle inspiré par la physique des particules. Il pose comme hypothèse que deux intervalles qui ne se chevauchent pas forment une marche aléatoire, et que les éléments qui composent la séquence des rendements financiers  $\{R(t)\}$  sont mutuellement indépendants. Il considère qu'un état stable existe où les rendements suivent une loi normale de paramètres  $\mu$  et  $\sigma^2$ .

Cependant, cet état stable n'est jamais réellement atteint, et la fonction de densité empirique généralement observée suppose une distribution symétrique concave, pointue au centre et ayant des queues épaisses. Il fait une analogie entre la température d'un gaz et le niveau d'activité sur les marchés, où la variance du mouvement brownien est proportionnelle à ces deux quantités. Il propose que le paramètre de variance de la normale  $\sigma^2$  suive une distribution  $g(\sigma^2)$  ayant un support positif. La distribution conditionnelle est normale lorsque ce paramètre est connu.

$$h_{R(t)}(r) = \int_0^\infty f_{R(t)}(r|\sigma^2)g(\sigma^2)d\sigma^2 \quad (1.3.19)$$

$$f_{R(t)}(r|\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(r-\mu)^2}{2\sigma^2} \right\} \quad (1.3.20)$$

Il propose comme solution acceptable pour la densité  $g(\sigma^2)$ , la distribution gamma inverse de paramètres  $m$  et  $s^2$  :

$$g(\sigma^2) = \frac{s^{2m}(m-1)^m e^{-(m-1)\frac{s^2}{\sigma^2}}}{\sigma^{2(m-1)}\Gamma(m)}. \quad (1.3.21)$$

Cette distribution a pour moyenne  $s^2$  et variance  $\frac{s^2}{m-2}$ . La distribution non conditionnelle des rendements  $h_{R(t)}$  est approximativement une Student avec  $2m$  degrés de liberté à un facteur

d'échelle de  $\left(\frac{m}{m-1}\right)^{1/2}$  près :

$$h_{R(t)}(r) = \frac{\Gamma(m) [2(m-1)\pi]^{1/2} s}{\left[1 + \frac{(y-\mu)^2}{s^2(2m-2)}\right]^{m+1/2}}. \quad (1.3.22)$$

D'autres distributions pourraient être utilisées au lieu de la gamma inverse. En utilisant la loi gamma, on obtient la distribution de Laplace asymétrique généralisée, qui sera l'objet d'une étude approfondie aux chapitres suivants. Il propose enfin d'utiliser aussi la distribution a priori gamma inverse pour le paramètre  $\mu$ . Par contre, il remarque qu'il obtient aussi une distribution similaire à celle de Student. Cette généralisation n'est donc pas nécessaire.

## 1.4 Conditions essentielles de Madan et Seneta

Inspirés par les travaux de Mandelbrot, Press et Praetz, Madan et Seneta (1990) présentent un ensemble de conditions considérées essentielles dans l'élaboration d'un modèle de rendements financiers. Ils se baseront sur celles-ci pour proposer le modèle Variance Gamma :

1. La distribution des rendements  $R$  doit avoir une queue épaisse. Ainsi, la probabilité que cette variable aléatoire ait une valeur supérieure à  $r + t$  avec un  $t$  petit, sachant qu'elle est supérieure à  $r$ , doit tendre vers 1, ce qui signifie que la fonction de survie converge lorsque cette quantité est grande.

$$\lim_{r \rightarrow \infty} P[R > r + t | R > r] = 1 \quad (1.4.1)$$

$$\bar{F}(r + t) \sim \bar{F}(r), \quad r \rightarrow \infty$$

2. La distribution doit posséder des moments finis pour les  $n$  premières puissances des rendements  $R$ . Étant donné que l'on cherche à modéliser la queue de la distribution, on fixe  $n = 4$ .

$$E[R^k] < \infty, \quad k \in \{1, 2, 3, 4\} \quad (1.4.2)$$

3. a) Le modèle doit proposer un processus de temps continu ayant des accroissements stationnaires et indépendants.
- b) Les distributions des accroissements doivent appartenir à la même famille, quelle que soit leur longueur. Cette condition est essentielle afin de permettre l'échantillonnage et l'analyse des séries chronologiques.
4. Le modèle doit permettre une extension multivariée avec une distribution elliptique afin de conserver la validité du modèle d'évaluation des actifs financiers.

Chacun des modèles présentés précédemment respecte la majorité ou toutes ces conditions. Les résultats se retrouvent à la table 1.2.

Modèles	Conditions			
	1	2	3	4
Mouvement brownien de Bachelier		*	*	*
Distribution stable symétrique de Mandelbrot	*			*
Processus de Poisson composé de Press	*	*	*	*
Mélange gaussien/inverse gamma de Praetz	*	*		*
Modèle Variance Gamma de Madan et Seneta	*	*	*	*

TABLE 1.2: Respect des conditions émises par Madan et Seneta pour les différents modèles présentés

On remarque que le modèle de Press remplit toutes les conditions émises par Madan et Seneta. Cependant, ils remarqueront que ce n'est pas un processus de sauts, car il contient aussi une composante de diffusion (Section A.4.1), ce qui va à l'encontre de l'intuition derrière la continuité de la trajectoire du prix. C'est cette dernière observation qui les incitera à proposer le modèle Variance Gamma, qui est un processus de sauts. Ce modèle, aussi étudié sous le nom de distribution de Laplace asymétrique généralisée par Kotz et al. (2001), a acquis beaucoup de notoriété dans le domaine de la finance mathématique. De plus, avec le développement de l'informatique et des méthodes numériques, on peut maintenant utiliser de manière efficace la fonction caractéristique dans le cadre de la calibration, des tests statistiques et de la tarification d'options. C'est pourquoi un intérêt particulier est apporté à cette distribution dans ce texte.



## Chapitre 2

# La distribution de Laplace asymétrique généralisée

Dans ce chapitre, on présente, en premier lieu, le processus de Laplace ainsi que les deux processus sous-jacents à sa construction, le processus gamma et le processus de Wiener. Ensuite, on présente la distribution de Laplace asymétrique généralisée et ses principales propriétés qui seront utilisées pour modéliser les rendements de titres financiers. Puis, on présente quelques cas particuliers. Dans le chapitre suivant, on présentera différentes méthodes pour obtenir une approximation de la fonction de densité et la fonction de répartition.

La distribution de Laplace asymétrique généralisée a été principalement étudiée par Kozubowski et Podgórski (1999). Cependant, elle a été introduite près d'une décennie auparavant par Madan et Seneta (1990), sous le nom de distribution Variance Gamma. La différence entre les approches des deux auteurs est majeure. Madan et Seneta (1990) développent un modèle financier à partir du mouvement brownien géométrique, qu'ils généralisent en proposant que la variance suive une distribution gamma. Kozubowski et Podgórski (1999) généralisent la distribution de Laplace asymétrique. Leur approche est plus générale, car ils ne cherchent pas à développer un modèle financier, mais une nouvelle classe de distributions utilisable dans divers domaines scientifiques. Étant donné leur approche plus détaillée et plus intuitive, c'est leur formulation du modèle qui sera développée. On rappellera enfin que les deux modèles sont équivalents même si leurs paramétrisations sont différentes.

### 2.1 Le processus de Laplace

Le processus de Laplace est défini comme étant un processus de Wiener subordonné par un processus gamma. En d'autres termes, c'est un processus de Wiener évalué à des temps aléatoires déterminés par un processus gamma. Selon Kotz et al. (2001), ce dernier est à la distribution de Laplace ce que le mouvement brownien est à la loi normale. Il est aussi un

cas particulier des processus de Lévy, et en conserve donc la principale propriété, celle d'être infiniment divisible.

Il a certains points en commun avec le mouvement brownien dont des moments finis pour tout ordre et des incréments indépendants et stationnaires. Cependant, la plupart des caractéristiques diffèrent :

- Discontinuité des trajectoires (processus de sauts) ;
- Distribution asymétrique des accroissements ;
- Paramètre d'échelle et de temps entièrement dissociés.

Enfin, il possède une représentation alternative qui n'implique aucun processus de Wiener. Il peut en fait être représenté comme la différence de deux processus gamma indépendants. On peut le représenter en utilisant la forme générale des processus de Lévy.

### 2.1.1 Le processus gamma

Le processus gamma, noté  $\{G(t; \tau, \beta)\}$ , est un processus de sauts purs (donc aucune composante de dérive ni de diffusion) dont les incréments  $G(t+1; \tau, \beta) - G(t; \tau, \beta)$  suivent une distribution gamma de paramètres de forme  $\tau$  et d'échelle  $\beta$ , définie par les fonctions de densité  $f_{\tau, \beta}(x)$  et caractéristique  $\phi_{\tau, \beta}(\xi)$  :

$$f_{\tau, \beta}(x) = \frac{\beta^\tau}{\Gamma(\tau)} x^{\tau-1} e^{-\beta x} 1_{\{x \geq 0\}} \quad (2.1.1)$$

$$\begin{aligned} \phi_{\tau, \beta}(\xi) &= E \left[ e^{i\xi X} \right] \\ &= \int_0^\infty e^{i\xi x} f_{\tau, \beta}(x) dx \\ &= \frac{1}{\left(1 - \frac{i\xi}{\beta}\right)^\tau}. \end{aligned} \quad (2.1.2)$$

On s'intéresse à la situation où le paramètre d'échelle est de valeur unitaire ( $\beta = 1$ ). Le processus gamma agit alors à titre de compteur et sa valeur  $G(t; \tau, \beta = 1)$  au temps  $t$  correspondra au nombre de sauts depuis  $t = 0$ . La fonction de densité  $G(t+1; \tau, \beta = 1) - G(t; \tau, \beta = 1)$  sera alors :

$$f_{\tau, \beta=1}(x) = \frac{1}{\Gamma(\tau)} x^{\tau-1} e^{-x} 1_{\{x \geq 0\}}. \quad (2.1.3)$$

Le paramètre  $\tau$ , qui définit la forme de la distribution, déterminera la fréquence moyenne des sauts du processus gamma  $\Gamma(t; \tau, \beta = 1)$ , étant donné l'espérance  $E[G(t)] = \tau \cdot t$ . La fonction caractéristique de ce processus sera donc  $\phi(\xi, t; \tau, \beta = 1)$ , en utilisant la propriété de convolution (A.2.13) (même si le temps  $t$  n'est pas entier, car la distribution est infiniment

divisible) :

$$\begin{aligned}\phi(\xi, t; \tau, \beta = 1) &= \left[ \frac{1}{\left(1 - \frac{i\xi}{1}\right)^\tau} \right]^t \\ &= \frac{1}{(1 - i\xi)^{\tau \cdot t}}.\end{aligned}\tag{2.1.4}$$

On peut réécrire la fonction caractéristique d'un incrément de ce processus  $\phi(\xi; t = 1; \tau, \beta = 1)$  sous la représentation de Lévy-Khintchine (A.4.1), avec l'exposant caractéristique  $\Xi(\zeta)$  :

$$\Xi(\zeta; t = 1; \tau, \beta = 1) = \tau \ln(1 - i\zeta)\tag{2.1.5}$$

$$\begin{aligned}&= \tau (e^0 - e^{-\infty}) \ln(1 - i\zeta) \\ &= \tau \int_0^\infty \frac{e^{-x} - e^{-(1-i\zeta)x}}{x} dx \\ &\quad \text{(intégrale de Frullani (Spiegel et Liu, 1999), p.115)} \\ &= \tau \int_0^\infty \left(1 - e^{i\zeta x}\right) \frac{1}{x} e^{-x} dx.\end{aligned}\tag{2.1.6}$$

On a donc, par cette représentation, la démonstration que le processus gamma est un processus de sauts purs. Il pourra donc être utilisé comme subordonnant dans la construction d'un processus subordonné (A.4.5).

### 2.1.2 Le processus de Wiener

Le processus de Wiener  $\{W(t; \mu, \sigma^2)\}$  est un processus de diffusion avec dérive. Il n'a donc pas de composante de saut. Ses incréments suivent une distribution normale :

$$W(t+1; \mu, \sigma^2) - W(t; \mu, \sigma^2) \sim N(\mu, \sigma^2).\tag{2.1.7}$$

Cette distribution est définie par la fonction de densité  $f_{\mu, \sigma}(x)$  et la fonction caractéristique  $\phi_{\mu, \sigma}(\xi)$  :

$$f_{\mu, \sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right\}\tag{2.1.8}$$

$$\phi_{\mu, \sigma}(\xi) = \exp\left\{i\mu\xi - \frac{\sigma^2\xi^2}{2}\right\}.\tag{2.1.9}$$

Notons que la variance d'un incrément est proportionnelle à la longueur de celui-ci. Soit deux incréments indépendants d'un même processus :  $I_1 = W(t+q; \mu, \sigma^2) - W(t; \mu, \sigma^2) \sim N(q\mu, q\sigma^2)$  et  $I_2 = W(t+q+s; \mu, \sigma^2) - W(t+q; \mu, \sigma^2) \sim N(s\mu, s\sigma^2)$ . La somme de ces

incrément suit une distribution normale dont la moyenne et la variance sont respectivement la somme de celles des deux incrément :

$$I_1 + I_2 \sim N((q + s)\mu, (q + s)\sigma^2). \quad (2.1.10)$$

Comme la distribution normale est aussi infiniment divisible, on peut obtenir la fonction caractéristique du processus  $\phi(\xi; t; \mu, \sigma^2)$  en utilisant la propriété de convolution (A.2.13) :

$$\phi(\xi; t; \mu, \sigma^2) = \exp \left\{ i\mu t\xi - \frac{\sigma^2 t\xi^2}{2} \right\}. \quad (2.1.11)$$

On déduit donc facilement l'exposant caractéristique  $\Lambda(\xi; t = 1; \mu, \sigma^2)$  d'un incrément de ce processus, sous la représentation de Lévy-Khintchine :

$$\Lambda(\xi; t = 1; \mu, \sigma^2) = -(i\mu\xi - \frac{\sigma^2\xi^2}{2}). \quad (2.1.12)$$

Ceci démontre que le processus de Wiener est un processus avec dérive et diffusion, mais sans composante de saut. Il pourra donc être utilisé pour construire un processus subordonné (A.4.5).

### 2.1.3 Le processus de Laplace est un processus subordonné

On considère un processus gamma  $G(t; \tau, \beta = 1)$  et un processus de Wiener  $W(t; \mu, \sigma^2)$ . On se rappelle que la variance d'un incrément (2.1.7) de ce dernier est proportionnelle à la longueur de l'intervalle de temps. En utilisant une propriété appelée la subordination, on peut modifier l'échelle de temps du processus de Wiener de sorte que la variance soit aléatoire pour tout intervalle. Tout processus de Lévy peut être utilisé comme subordonnant pour définir cette échelle de temps. Si on utilise le processus gamma, on obtiendra le processus de Laplace sans dérive  $\{Y(t; \sigma, \mu, \tau)\}$  défini comme suit :

$$\{Y(t; \sigma, \mu, \tau)\} \equiv \{W(G(t; \tau, \beta = 1); \mu, \sigma^2)\}. \quad (2.1.13)$$

On obtient l'exposant caractéristique  $\Psi(\xi, t = 1; \sigma, \mu, \tau)$  d'un incrément  $Y(t + 1; \sigma, \mu, \tau) - Y(t; \sigma, \mu, \tau)$  en utilisant la propriété de subordination définie par l'équation (A.4.6), où  $\Xi(\zeta, t = 1; \tau, \beta = 1)$  est l'exposant caractéristique du processus gamma et  $\Lambda(\xi, t = 1; \mu, \sigma^2)$  celui du processus de Wiener :

$$\begin{aligned} \Psi(\xi, t = 1; \sigma, \mu, \tau) &= \Xi(i\Lambda(\xi, t = 1; \mu, \sigma^2), t = 1; \tau, \beta = 1) \\ &= \tau \ln(1 - i(i\Lambda(\xi))) \\ &= \tau \ln \left( 1 + \left( \frac{\sigma^2\xi^2}{2} - i\mu\xi \right) \right). \end{aligned} \quad (2.1.14)$$



Le processus de Laplace sans dérive est donc, par définition, un processus de Lévy et par conséquent infiniment divisible. En utilisant l'exposant caractéristique (2.1.14) et la définition (A.4.7), on obtient sa fonction caractéristique :

$$\begin{aligned}
 \phi_{Y(t;\sigma,\mu,\tau)}(\xi) &= \exp \{-t \cdot \Psi(\xi, t = 1; \sigma, \mu, \tau)\} \\
 &= \exp \left\{ -t \cdot \left( \tau \ln \left( 1 + \left( \frac{\sigma^2 \xi^2}{2} - i\mu\xi \right) \right) \right) \right\} \\
 &= \left( 1 + \frac{\sigma^2 \xi^2}{2} - i\mu\xi \right)^{-\tau \cdot t}.
 \end{aligned} \tag{2.1.15}$$

Une manière simple pour expliquer le mécanisme derrière le processus subordonné est d'en construire une trajectoire à l'aide de la simulation.

On simule un temps d'arrivée  $T_1$ , de distribution gamma, puis une hauteur de saut  $X_1$ , de distribution normale. On obtient ainsi le premier incrément de la trajectoire, tel qu'illustré à la figure 2.1.

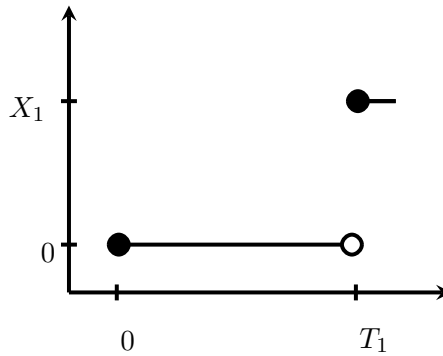


FIGURE 2.1: Premier incrément d'un processus subordonné

Une réalisation d'une trajectoire de ce processus par simulation se trouve à la figure 2.2.

Le processus gamma  $G(t; \tau, \beta = 1)$ , en tant que subordonnant dans ce cas-ci, définit une échelle de temps économique, selon laquelle on situe l'arrivée d'évènements pouvant influencer le prix d'un titre financier. Cette dernière ne peut être mesurée, elle est donc abstraite. L'échelle de temps où sont effectuées les observations du processus correspond au temps calendrier. C'est la seule qui puisse être mesurée. Étant donné que ces deux échelles sont indépendantes, plusieurs sauts entre deux observations sont possibles. L'échelle de temps économique est donc soit étirée, soit compressée, en comparaison au temps calendrier. Autrement dit, si l'on définit une journée économique comme étant l'intervalle de temps entre deux sauts, on peut en avoir plusieurs au cours d'une seule journée de calendrier ( $G(t + 1; \tau, \beta = 1) - G(t; \tau, \beta = 1) > 1$ ). À l'opposé, une d'entre elles peut chevaucher plusieurs journées calendrier ( $G(t + 1; \tau, \beta = 1) - G(t; \tau, \beta = 1) \leq 1$ ).

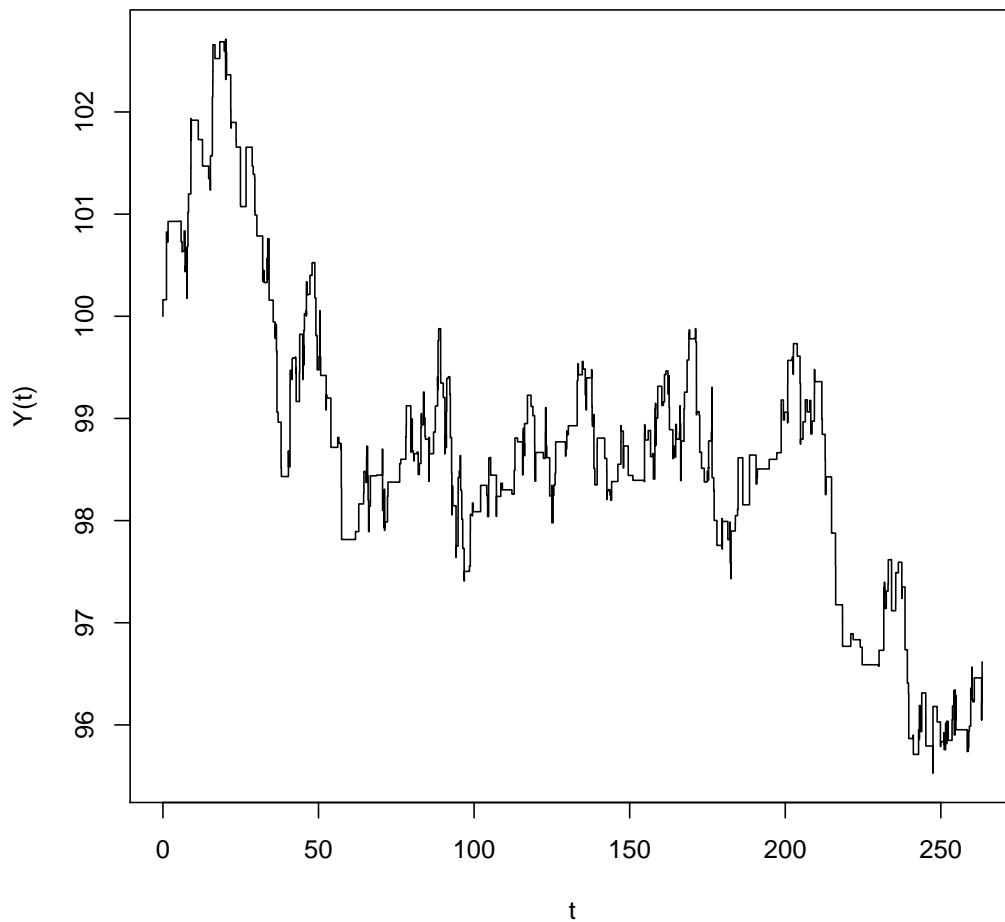


FIGURE 2.2: Simulation d'un processus de Wiener subordonné par un processus gamma

Un processus stochastique qui représente le comportement du prix d'un titre financier doit inclure une composante de dérive. Celle-ci exprime le rendement moyen réalisé et est indépendante du processus de sauts. Pour cette raison, on ajoute un coefficient de dérive  $\theta$  au processus de Laplace sans dérive, pour obtenir sa forme générale. Comme ce coefficient est constant, on peut multiplier la fonction caractéristique (2.1.15) par la transformée de Fourier inverse du produit de celui-ci et de la longueur de l'intervalle de temps  $t$ ,  $\mathcal{F}^{-1}(\theta \cdot t) = e^{i\xi\theta \cdot t}$ ,

pour obtenir celle du processus de Laplace :

$$\begin{aligned}\phi_{Y(t;\theta,\sigma,\mu,\tau)}(\xi) &= \frac{e^{i\xi\theta \cdot t}}{\left(1 + \frac{\sigma^2\xi^2}{2} - i\mu\xi\right)^{\tau \cdot t}} \\ &= \left(\frac{e^{i\xi\theta}}{\left(1 + \frac{\sigma^2\xi^2}{2} - i\mu\xi\right)^\tau}\right)^t.\end{aligned}\tag{2.1.16}$$

Le processus  $\{Y(t; \theta, \sigma, \mu, \tau)\}$  définit, dans le contexte financier, l'évolution du logarithme du prix, tel que présenté par Kotz et al. (2001). Pour des fins de simplification, on fixe le prix initial à 1 :  $Y(0; \theta, \sigma, \mu, \tau) = 0$

La fonction caractéristique (2.1.16) constituera la principale représentation du processus de Laplace pour la suite de ce texte. La construction du modèle «Variance Gamma» de Madan et Seneta (1990) est similaire, à l'exception que la paramétrisation et le processus gamma utilisés sont différents, ce qui rend leur approche moins intuitive, bien que le résultat soit équivalent.

## 2.2 Distribution de Laplace asymétrique généralisée

La distribution de Laplace asymétrique généralisée caractérise un intervalle du processus de Laplace avec dérive. Aussi appelée distribution de Bessel, elle a été introduite par Karl Pearson en 1929, en lien avec la covariance d'un échantillon tiré d'une population normale à deux variables. C'est aussi une généralisation de la distribution de Laplace asymétrique qui sera présentée à la section 2.3.1.

### 2.2.1 Fonction caractéristique

On définit cette distribution principalement par sa fonction caractéristique. Celle-ci s'obtient facilement à partir de la fonction caractéristique du processus de Laplace avec dérive (2.1.16), en considérant un incrément de longueur  $t = 1$ . À partir de la définition de ce dernier (2.1.13), on déduit qu'elle est en fait un mélange de la loi normale dont le paramètre de variance suit une distribution gamma.

$Y$  est une variable aléatoire définie comme étant la somme :

- d'un paramètre de translation  $\theta$ ,
- du produit :
  - d'une variable aléatoire  $W$  issue d'une distribution gamma (2.1.3)
  - et d'un paramètre  $\mu$
- et du produit :
  - d'un paramètre  $\sigma$ ,

- de la racine carrée de la variable aléatoire  $W$
- et d'une variable aléatoire  $Z$  issue d'une distribution normale centrée réduite :

$$Y = \theta + \mu W + \sigma\sqrt{W}Z \quad (2.2.1)$$

où

$$Z \sim N(0, 1) \text{ et } W \sim \Gamma(\tau, \beta = 1).$$

Alors, la variable aléatoire  $Y$ , sachant que  $W = w$ , suit une distribution normale de moyenne  $w\mu$  et de variance  $w\sigma^2$  :

$$(Y|W = w) \sim N(w\mu, w\sigma^2). \quad (2.2.2)$$

La fonction caractéristique de la variable aléatoire  $Y$  peut donc être obtenue en utilisant celle de la loi normale  $\phi_{\mu w, w\sigma^2}^N(t)$  et la formule de l'espérance conditionnelle :

$$\begin{aligned} \phi_Y(t; \theta, \sigma, \mu, \tau) &= E [E [e^{itY} | W]] \\ &= \int_0^\infty E [e^{it(\theta + \mu w + \sigma\sqrt{w}Z)}] g(w) dw \quad (\text{en utilisant (2.2.1)}) \\ &= e^{i\theta t} \int_0^\infty \phi_{\mu w, w\sigma^2}^N(t) g(w) dw \\ &= e^{i\theta t} \int_0^\infty e^{iw\mu t - \frac{w\sigma^2 t^2}{2}} \times \frac{1}{\Gamma(\tau)} w^{\tau-1} e^{-w} dw \\ &= e^{i\theta t} \int_0^\infty \frac{1}{\Gamma(\tau)} w^{\tau-1} e^{-w(1 + \frac{1}{2}\sigma^2 t^2 - i\mu t)} dw. \end{aligned}$$

En complétant l'intérieur de l'intégrale de façon à retrouver la densité de la loi gamma de paramètres  $\alpha = \tau$  et  $\beta = (1 + \frac{1}{2}\sigma^2 t^2 - i\mu t)^\tau$ , on obtient la fonction caractéristique de la variable aléatoire  $Y$  de distribution Laplace asymétrique généralisée :

$$\begin{aligned} \phi_Y(t; \theta, \sigma, \mu, \tau) &= \frac{e^{i\theta t}}{(1 + \frac{1}{2}\sigma^2 t^2 - i\mu t)^\tau} \int_0^\infty \frac{(1 + \frac{1}{2}\sigma^2 t^2 - i\mu t)^\tau}{\Gamma(\tau)} w^{\tau-1} e^{-w(1 + \frac{1}{2}\sigma^2 t^2 - i\mu t)} dw \\ &= \frac{e^{i\theta t}}{(1 + \frac{1}{2}\sigma^2 t^2 - i\mu t)^\tau}. \end{aligned} \quad (2.2.3)$$

### 2.2.2 Invariance d'échelle

Une seconde paramétrisation pour la famille de distributions de Laplace introduit la propriété d'invariance d'échelle. Cette propriété permet d'appliquer un changement d'échelle à une variable aléatoire en modifiant un seul paramètre sans que la valeur des autres ne soit affectée. Si l'on revient à la définition de la variable aléatoire conditionnelle  $Y|W$  (2.2.2), on remarque que

les paramètres  $\mu$  et  $\sigma^2$  sont influencés par la valeur de  $W$ . Ils sont donc corrélés. En introduisant un paramètre d'invariance d'échelle  $\kappa$  en remplacement de  $\mu$ , on élimine cette corrélation. Le paramètre  $\kappa$  est obtenu à l'aide de la transformation suivante :

$$\kappa = \frac{\sqrt{2\sigma^2 + \mu^2} - \mu}{\sqrt{2}\sigma}. \quad (2.2.4)$$

À l'inverse, on peut retrouver le paramètre  $\mu$  en l'isolant dans l'équation précédente. On obtient donc :

$$\mu = \frac{\sigma}{\sqrt{2}} \left( \frac{1}{\kappa} - \kappa \right). \quad (2.2.5)$$

On définit la fonction vectorielle  $T_{\mu \rightarrow \kappa}$  comme étant la transformation qui permet le passage de la forme en  $\mu$  à celle en  $\kappa$  :

$$T_{\mu \rightarrow \kappa}(\theta, \sigma, \mu, \tau) = \begin{bmatrix} \theta \\ \sigma \\ \frac{\sqrt{2\sigma^2 + \mu^2} - \mu}{\sqrt{2}\sigma} \\ \tau \end{bmatrix}. \quad (2.2.6)$$

On définit aussi la transformation inverse  $T_{\kappa \rightarrow \mu}$  :

$$T_{\kappa \rightarrow \mu}(\theta, \sigma, \kappa, \tau) = \begin{bmatrix} \theta \\ \sigma \\ \frac{\sigma}{\sqrt{2}} \left( \frac{1}{\kappa} - \kappa \right) \\ \tau \end{bmatrix}. \quad (2.2.7)$$

Cette notation permet d'utiliser un vecteur de paramètres. On peut obtenir la matrice de variance-covariance d'une forme paramétrique en connaissant celle de l'autre et en utilisant le gradient de ces transformations.

Pour la première transformation, on a :

$$\nabla T_{\mu \rightarrow \kappa} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \frac{\mu\sqrt{4\sigma^2 + \mu^2} - \mu^2}{2\sigma^2\sqrt{4\sigma^2 + \mu^2}} & -\frac{\sqrt{4\sigma^2 + \mu^2} - \mu}{2\sigma\sqrt{4\sigma^2 + \mu^2}} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (2.2.8)$$

Pour la seconde transformation, on a :

$$\nabla T_{\kappa \rightarrow \mu} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -\frac{\kappa^2 - 1}{\sqrt{2}\kappa} & -\frac{(\kappa^2 + 1)\sigma}{\sqrt{2}\kappa^2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (2.2.9)$$

La forme en  $\mu$  sera privilégiée pour l'estimation, car elle est plus compacte. Cependant, certaines propriétés de la distribution font appel à la forme utilisant le paramètre  $\kappa$ .

### 2.2.3 Fonctions génératrices

En utilisant la relation (A.2.16), on obtient la fonction génératrice des moments à partir de la fonction caractéristique :

$$\begin{aligned} M_Y(\xi) &= \phi_Y(-i\xi) \\ &= \frac{e^{\theta\xi}}{\left(1 - \frac{1}{2}\sigma^2\xi^2 - \mu\xi\right)^\tau}, \end{aligned} \quad (2.2.10)$$

$$\text{où } 1 - \frac{1}{2}\sigma^2\xi^2 - \mu\xi > 0. \quad (2.2.11)$$

La condition 2.2.11 permet de s'assurer que le dénominateur prend une valeur réelle strictement positive. Cette condition impose aussi une restriction à l'espace des paramètres  $\Omega$ . La fonction génératrice des moments permet d'obtenir tous les moments  $E[Y^r]$  d'une variable aléatoire en la dérivant successivement par rapport à la variable de transformation et en égalant cette dernière à 0 :

$$E[Y^r] = \left[ \frac{d^r M_Y(\xi)}{d\xi^r} \right]_{\xi=0}. \quad (2.2.12)$$

La fonction génératrice des cumulants  $K_Y(\xi)$  est aussi intéressante à utiliser. On l'obtient à partir du logarithme de la fonction génératrice des moments (A.2.15) :

$$\begin{aligned} K_Y(\xi) &= \ln(M_Y(\xi)) \\ &= \ln\left(\frac{e^{\theta\xi}}{\left(1 - \frac{1}{2}\sigma^2\xi^2 - \mu\xi\right)^\tau}\right), \quad 1 - \frac{1}{2}\sigma^2\xi^2 - \mu\xi > 0. \end{aligned} \quad (2.2.13)$$

Elle a une utilité similaire à la fonction génératrice des moments, sauf qu'elle permet d'obtenir les cumulants  $K_r$  de la distribution en la dérivant successivement par rapport à la variable de transformation et en égalant celle-ci à 0 :

$$K_r = \left[ \frac{d^r \ln(M_Y(\xi))}{d\xi^r} \right]_{\xi=0}. \quad (2.2.14)$$

### 2.2.4 Moments et rôle des paramètres

En mathématiques, les moments sont des quantités décrivant la forme d'un ensemble de points. En statistique, les moments décrivent certaines caractéristiques d'une population ou d'un échantillon. Ces caractéristiques sont utilisées dans la sélection d'une distribution de probabilité appropriée pour représenter la population à partir de l'échantillon, ce qu'on appelle

l'inférence statistique. Les moments bruts et centraux sont évalués par rapport à 0 et à la moyenne respectivement.

On obtient les premiers moments bruts de cette distribution à l'aide de la relation décrite précédemment (2.2.12).

$$\begin{aligned}
E[Y] &= \theta + \tau \mu \\
E[Y^2] &= \theta^2 + 2 \mu \tau \theta + \tau \sigma^2 + \mu^2 \tau^2 + \mu^2 \tau \\
E[Y^3] &= \theta^3 + 3 \mu \tau \theta^2 + (3 \tau \sigma^2 + 3 \mu^2 \tau^2 + 3 \mu^2 \tau) \theta \\
&\quad + (3 \mu \tau^2 + 3 \mu \tau) \sigma^2 + \mu^3 \tau^3 + 3 \mu^3 \tau^2 + 2 \mu^3 \tau \\
E[Y^4] &= \theta^4 + 4 \mu \tau \theta^3 + (6 \tau \sigma^2 + 6 \mu^2 \tau^2 + 6 \mu^2 \tau) \theta^2 \\
&\quad + ((12 \mu \tau^2 + 12 \mu \tau) \sigma^2 + 4 \mu^3 \tau^3 + 12 \mu^3 \tau^2 + 8 \mu^3 \tau) \theta \\
&\quad + (3 \tau^2 + 3 \tau) \sigma^4 + (6 \mu^2 \tau^3 + 18 \mu^2 \tau^2 + 12 \mu^2 \tau) \sigma^2 \\
&\quad + \mu^4 \tau^4 + 6 \mu^4 \tau^3 + 11 \mu^4 \tau^2 + 6 \mu^4 \tau.
\end{aligned}$$

À partir de ceux-ci, on obtient aussi les quatre premiers moments centraux :

$$m_1 = E[Y] = \theta + \tau \mu \quad (2.2.15a)$$

$$m_2 = E[(Y - m_1)^2] = \tau \sigma^2 + \tau \mu^2 \quad (2.2.15b)$$

$$m_3 = E[(Y - m_1)^3] = 3 \tau \sigma^2 \mu + 2 \tau \mu^3 \quad (2.2.15c)$$

$$m_4 = E[(Y - m_1)^4] = 3 \tau^2 \sigma^4 + 3 \tau^2 \mu^4 + 3 \tau \sigma^4 + 6 \tau \mu^4 + 6 \tau^2 \mu^2 \sigma^2 + 12 \tau \sigma^2 \mu^2. \quad (2.2.15d)$$

On résume le domaine et le rôle des paramètres à la table 2.1.

On retrouve quelques exemples de courbes de la fonction de densité avec différents paramètres d'asymétrie et d'aplatissement à la figure 2.3.

Afin de comparer la distribution de Laplace asymétrique généralisée avec la normale, que l'on cherche à remplacer dans le contexte des rendements financiers, le comportement des coefficients d'asymétrie  $\gamma_1$  et d'aplatissement  $\gamma_2$  de celle-ci peut être intéressant à observer.

On obtient ces derniers à partir des moments centraux (2.2.15) :

$$\gamma_1(Y) = \frac{m_3}{(m_2)^{3/2}} = \frac{2 \mu^3 + 3 \sigma^2 \mu}{(\mu^2 + \sigma^2)^{3/2} \sqrt{\tau}} \quad (2.2.16a)$$

$$\gamma_2(Y) = \frac{m_4}{(m_2)^2} - 3 = \frac{(3 \mu^4 + 6 \sigma^2 \mu^2 + 3 \sigma^4) \tau^2 + (6 \mu^4 + 12 \sigma^2 \mu^2 + 3 \sigma^4) \tau}{(\mu^2 + \sigma^2)^2 \tau^2} - 3. \quad (2.2.16b)$$

Le coefficient d'asymétrie de la normale vaut  $\gamma_1^N(Y) = 0$  et celui d'excès d'aplatissement,  $\gamma_2^N(Y) = 0$ , puisqu'il a été défini à partir de cette distribution. Si l'on fixe tous les paramètres

Paramètre	Domaine	Rôle	Observations
$\theta$	$\mathbb{R}$	Localisation	N'influence que la moyenne. Équivaut au mode lorsque $\mu = 0$ .
$\sigma$	$\mathbb{R}^+ \setminus \{0\}$	Échelle	Vrai paramètre d'échelle lorsque $\kappa$ est utilisé.
$\mu$	$\mathbb{R}$	Asymétrie	Distribution asymétrique à gauche lorsque négatif et à droite lorsque positif. Déplace la moyenne dans la même direction. Corrélation positive avec la variance et le coefficient d'aplatissement.
$\kappa$	$\mathbb{R}^+ \setminus \{0\}$	Asymétrie	Valeur dans l'intervalle $[0, 1[$ lorsque $\mu < 0$ , dans $[1, \infty]$ lorsque $\mu \geq 0$ .
$\tau$	$\mathbb{R}^+ \setminus \{0\}$	Aplatissement	Négativement corrélé avec le coefficient d'aplatissement. Une petite valeur donne une distribution pointue.

TABLE 2.1: Domaine et rôle des paramètres de la distribution de Laplace asymétrique généralisée

sauf  $\mu$ , la valeur minimale du coefficient d'excès d'aplatissement  $\gamma_2(Y)$  est atteinte lorsque  $\mu$  est de 0 :

$$\begin{aligned} \min_{\mu} \gamma_2(Y) &= [\gamma_2(Y)]_{\mu=0} \\ &= \frac{3}{\tau}. \end{aligned}$$

Étant donné que le paramètre  $\tau$  est strictement positif, la valeur minimale que peut prendre le coefficient d'excès d'aplatissement  $\gamma_2(Y)$  est plus grande que 0. Par conséquent, l'aplatissement de la distribution de Laplace asymétrique généralisée sera toujours supérieur à celui de la normale ; c'est donc une distribution leptocurtique, ce qui est une propriété désirable selon les observations de Madan et Seneta (1990).

### 2.2.5 Changement d'échelle et de localisation

Parfois, on doit modifier un ensemble de données afin d'effectuer des comparaisons ou pour bénéficier des avantages d'une méthode numérique. Les principales transformations utilisées sont :

1. un changement de localisation, où l'on additionne une constante à l'ensemble des données
2. un changement d'échelle, où l'on multiplie l'ensemble des données par un facteur
3. une combinaison des deux transformations précédentes.

Soit deux constantes  $a, b \neq 0$  et une variable aléatoire  $X$  qui suit une distribution Laplace asymétrique généralisée :  $X \sim GAL(\theta, \sigma, \kappa, \tau)$ .  $Y$  correspond à la somme :



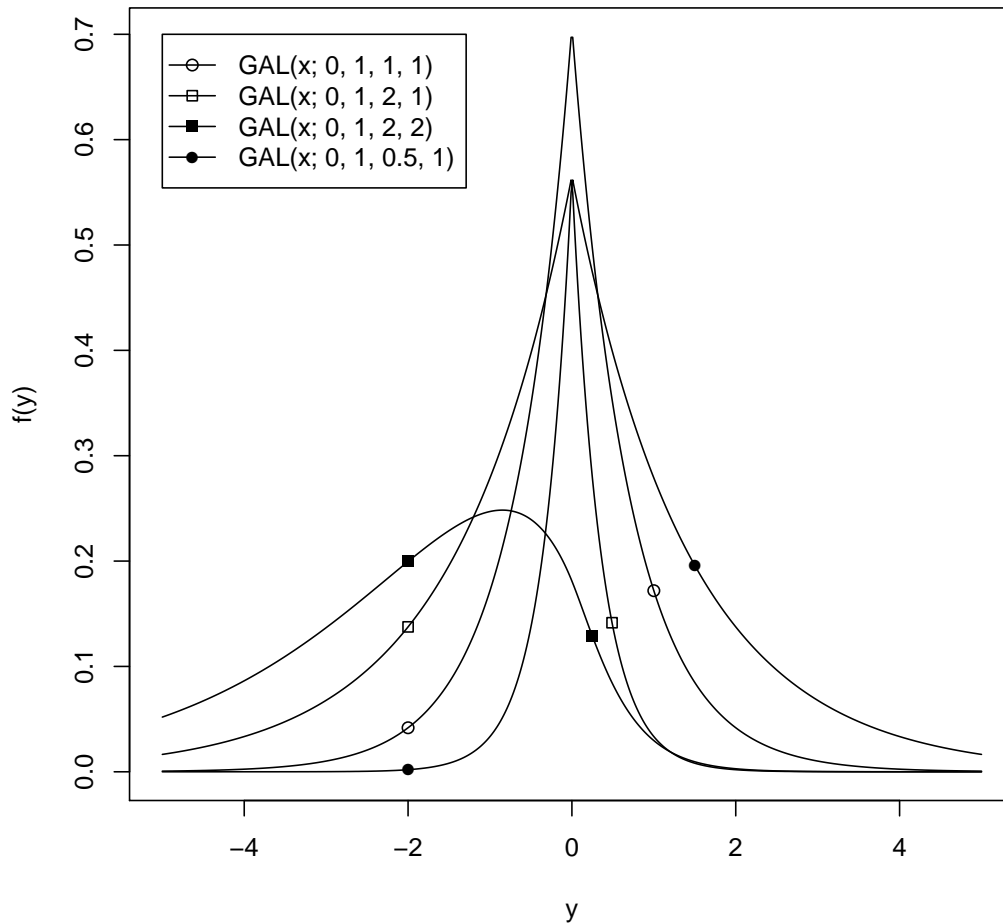


FIGURE 2.3: Fonction de densité de la distribution Laplace asymétrique généralisée avec différents paramètres :  $GAL(y; \theta, \sigma, \kappa, \tau)$

- de la constante  $b$  et
- du produit :
  - de la constante  $a$  et
  - de la variable aléatoire  $X$ .

Selon Kotz et al. (2001), en utilisant le paramètre  $\kappa$ , on peut effectuer un changement d'échelle et de localisation et obtenir une variable aléatoire qui suit encore cette distribution :

- le paramètre de localisation  $\theta$  subit la même transformation que la variable aléatoire  $X$  ;
- le paramètre d'échelle  $\sigma$  est multiplié par la constante  $a$  ;
- le paramètre  $\kappa$  est inversé si cette constante est négative. Ce dernier cas est en fait une réflexion de la distribution autour du mode.

On a donc :

$$Y = aX + b \sim GAL(a\theta + b, a\sigma, \kappa^{\text{sgn } a^{-1}}, \tau). \quad (2.2.17)$$

On estime les paramètres  $\hat{\theta}, \hat{\sigma}, \hat{\mu}$  et  $\hat{\tau}$  sur un ensemble de données centrées et réduites  $X_t$  à partir d'un échantillon original  $Y_t$ .  $\hat{m}$  et  $\hat{s} > 0$  sont respectivement la moyenne et l'écart-type de l'échantillon  $Y_t$ . En utilisant l'équation de transformation (2.2.17) avec les paramètres estimés précédemment, on peut alors obtenir ceux correspondants pour l'échantillon :

$$Y_t = \hat{s}X_t + \hat{m} \sim GAL(\hat{s}\theta + \hat{m}, \hat{s}\sigma, \kappa, \tau). \quad (2.2.18)$$

Le paramètre  $\kappa$  n'est pas modifié puisque l'écart-type  $\hat{s}$  est strictement positif. Cette propriété permettra de pratiquer l'estimation sur des données centrées et réduites, ce qui diminue le risque d'erreurs numériques sans nuire à sa précision, puisqu'aucune information contenue dans l'échantillon n'est perdue.

## 2.2.6 Représentation alternative et simulation

Le processus de Laplace  $Y(t; \theta, \sigma, \kappa, \tau)$  peut être représenté sous la forme d'une différence de deux processus gamma  $G_1(t; \tau) - G_2(t; \tau)$  à laquelle on additionne une composante de dérive. Le premier processus compte les gains, alors que le second, les pertes. Les deux ont des incréments qui suivent une distribution gamma  $\Gamma(\alpha = \tau, \beta = 1)$ , c'est-à-dire la même distribution que le processus subordonnant utilisé à la section 2.1.3 :

$$G_i(\tau) \sim \Gamma(\tau, \beta = 1), \quad i \in \{1, 2\}. \quad (2.2.19)$$

Il peut être représenté sous la forme du processus composé suivant :

$$Y(t) \stackrel{d}{=} \theta + \frac{\sigma\sqrt{2}}{2} \left( \frac{1}{\kappa} G_1(t) - \kappa G_2(t) \right). \quad (2.2.20)$$

La distribution de Laplace asymétrique généralisée peut alors être représentée sous la forme d'un incrément de ce processus. Les variables aléatoires  $G_1$  et  $G_2$  sont respectivement des réalisations indépendantes de distribution gamma avec densité (2.1.3) :

$$Y \stackrel{d}{=} \theta + \frac{\sigma\sqrt{2}}{2} \left( \frac{1}{\kappa} G_1 - \kappa G_2 \right). \quad (2.2.21)$$

Simuler une réalisation de chacune d'entre elles suffira pour obtenir une réalisation de la distribution gamma. On peut aussi réécrire la définition précédente de la variable aléatoire  $Y$  (2.2.21) sous la forme simplifiée suivante :

$$Y \stackrel{d}{=} \theta + (G_3 - G_4). \quad (2.2.22)$$

Les deux variables aléatoires sont alors définies comme suit :

$$\begin{aligned} G_3 &\sim \Gamma\left(\tau, \beta = \frac{\sqrt{2}}{\kappa\sigma}\right) \\ G_4 &\sim \Gamma\left(\tau, \beta = \frac{\kappa\sigma}{\sqrt{2}}\right). \end{aligned}$$

On peut facilement démontrer l'équivalence en distribution (2.2.22). En effectuant le produit des fonctions caractéristiques respectives  $\phi_Y(\xi)$ ,  $\phi_{G_3}(\xi)$  et  $\phi_{G_4}(\xi)$  des variables aléatoires  $Y$ ,  $G_3$  et  $G_4$ , on retrouve celle de la forme en  $\kappa$  :

$$\begin{aligned} \phi_Y(\xi) &= e^{i\xi\theta} \cdot \phi_{G_3}(\xi) \cdot \phi_{G_4}(\xi) \\ &= e^{i\xi\theta} \cdot \left(\frac{1}{1 + i\xi \cdot \frac{\kappa\sigma}{\sqrt{2}}}\right)^\tau \cdot \left(\frac{1}{1 - i\xi \cdot \frac{\kappa\sigma}{\sqrt{2}}}\right)^\tau \\ &= \frac{e^{i\xi\theta}}{\left(1 - i\xi \frac{\sigma}{\sqrt{2}} \left(\frac{1}{\kappa} - \kappa\right) + \frac{\sigma^2 \xi^2}{2}\right)^\tau}. \end{aligned} \tag{2.2.23}$$

La figure 2.4 présente un histogramme et un estimateur de densité par noyau d'un échantillon de 2500 réalisations de la variable aléatoire  $Y$ , suivant une distribution de Laplace asymétrique généralisée de paramètres  $\theta = 0$ ,  $\sigma = 1$ ,  $\kappa = 2$  et  $\tau = 1$ . On remarquera que le paramètre  $\kappa$  produit une distribution asymétrique à gauche puisqu'il prend une valeur supérieure à 1.

### 2.2.7 Fonction de Bessel et densité

La densité de la distribution Laplace asymétrique généralisée peut être exprimée à l'aide de la fonction de Bessel modifiée du troisième type (Abramowitz et Stegun, 1965) :

$$K_\lambda(u) = \frac{(u/2)^\lambda \Gamma(1/2)}{\Gamma(\lambda + 1/2)} \int_1^\infty e^{-ut} (t^2 - 1)^{\lambda-1/2} dt, \quad \lambda \geq -1/2. \tag{2.2.24}$$

On définit les paramètres  $\delta = \mu\sigma^{-1}$  et  $\eta = \sqrt{1 + \delta^2}$  afin de simplifier la notation. La démonstration qui permet d'obtenir la fonction de densité est fondée sur la définition de la distribution utilisant la différence de deux processus gamma (2.2.21). Elle est développée par Kozubowski et Podgórski (1999) :

$$\begin{aligned} f(x) &= \frac{e^{-\kappa \operatorname{sgn}(x)|x|}}{\Gamma(\tau)} \int_0^\infty y^{\tau-1} (|x| + y)^{\tau-1} e^{-\eta y} dy \\ &= \frac{1}{\Gamma(\tau)\sqrt{\pi}} \left(\frac{|x|}{\eta}\right)^{\tau-1/2} e^{\delta x/2} K_{\tau-1/2}\left(\frac{\eta|x|}{2}\right). \end{aligned} \tag{2.2.25}$$

Afin d'incorporer le paramètre de localisation  $\theta$ , on remplace  $x$  par la valeur centrée  $x - \theta$ , et les paramètres  $\eta$  et  $\delta$  par ceux d'origine, tels que définis précédemment, pour obtenir la forme

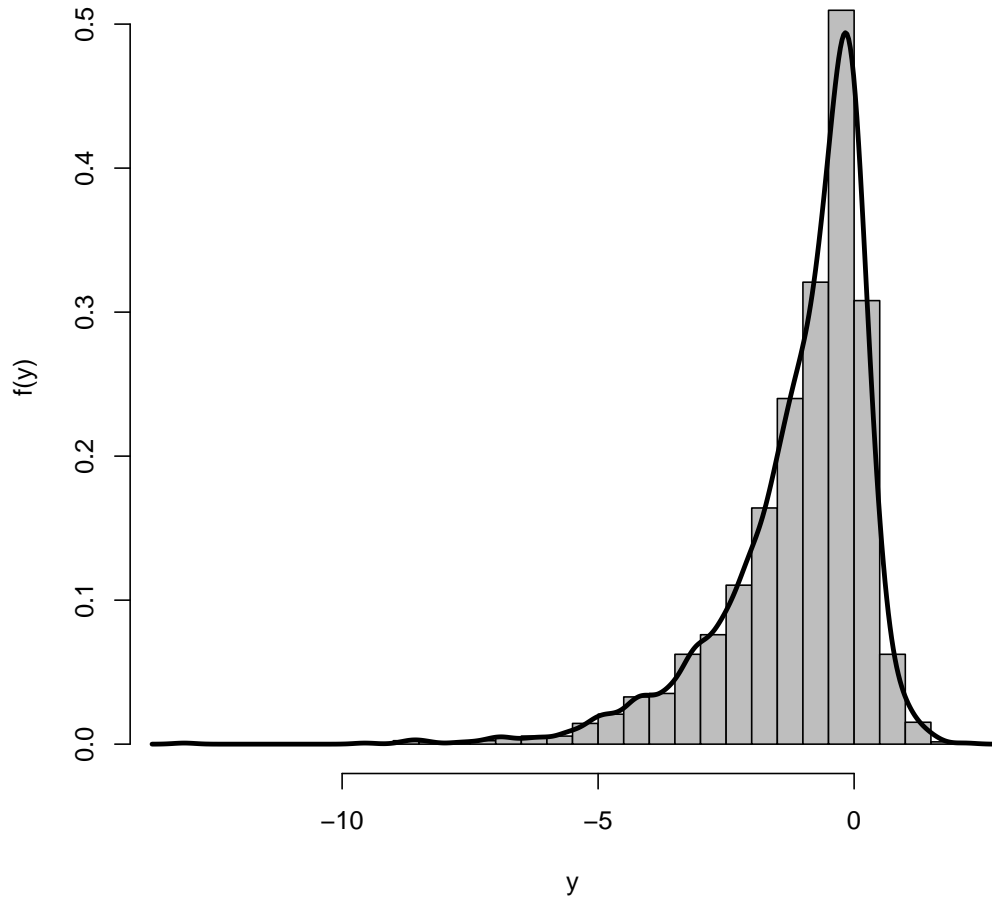


FIGURE 2.4: Histogramme et estimateur de densité par noyau de 2500 réalisations de la variable aléatoire  $Y \sim GAL(\theta = 0, \sigma = 1, \kappa = 2, \tau = 1)$

suivante (Kotz et al., 2001) :

$$f(x) = \frac{\sqrt{2}e^{\frac{\sqrt{2}}{2\sigma}(1/\kappa-\kappa)(x-\theta)}}{\sqrt{\pi}\sigma^{\tau+1/2}\Gamma(\tau)} \left(\frac{\sqrt{2}|x-\theta|}{\kappa+1/\kappa}\right)^{\tau-1/2} K_{\tau-1/2}\left(\frac{\sqrt{2}}{2\sigma}\left(\frac{1}{\kappa}+\kappa\right)|x-\theta|\right). \quad (2.2.26)$$

Cette fonction a été implémentée dans un algorithme de maximum de vraisemblance pour le logiciel GNU R par Scott et Dong (2012). Elle reste cependant sensible numériquement en plus de demander un temps de calcul important pour de grands échantillons. De plus, comme la fonction de densité n'est pas différentiable, on ne peut pas obtenir les estimateurs du maximum de vraisemblance et leurs matrices de variance-covariance. Donc, il est impossible d'effectuer de tests d'adéquation ou d'hypothèse avec ces résultats. C'est une des principales raisons qui

ont motivé la recherche d'autres méthodes d'estimation plus efficaces pour cette distribution.

## 2.3 Cas particuliers

La distribution de Laplace asymétrique généralisée peut être considérée comme une extension de plusieurs cas particuliers. Ces distributions ainsi que leur notation et leur fonction de densité sont présentées dans la table 2.2.

Cas	Distribution	Notation	Densité
$\theta = 0$ $\sigma = 0$ $\mu > 0$ $\tau = 1$	Exponentielle de moyenne $\mu > 0$	$GAL(0, 0, \mu, 1)$	$\frac{1}{\mu}e^{-x/\mu} (x > 0)$
$\theta = 0$ $\sigma = 0$ $\mu > 0$	Gamma de paramètres $\alpha = \tau, \beta = \mu > 0$	$GAL(0, 0, \mu, \tau)$	$\frac{x^{\tau-1}e^{-x/\mu}}{\mu^\tau \Gamma(\tau)}, (x > 0)$
$\sigma > 0$ $\mu = 0$ $\tau = 1$	Laplace symétrique ( $\sigma > 0$ )	$GAL(\theta, \sigma, 0, 1)$	$\frac{1}{\sqrt{2}\sigma}e^{-\sqrt{2} x-\theta /\sigma}, (x \in \mathbb{R})$
$\sigma > 0$ $\mu \neq 0$ $\tau = 1$	Laplace asymétrique ( $\sigma > 0$ )	$GAL(\theta, \sigma, \mu, 1)$	$\frac{\sqrt{2}\kappa}{\sigma(1+\kappa^2)} \begin{cases} e^{\frac{\sqrt{2}\kappa}{\sigma}(\theta-x)}, & x \geq \theta \\ e^{\frac{\sqrt{2}}{\sigma\kappa}(x-\theta)}, & x < \theta \end{cases}$
$\sigma = 0$ $\mu = 0$ $\tau = 0$	Dégénérée à $\theta$	$GAL(\theta, 0, 0, 0)$	$1_{x=\theta}$

TABLE 2.2: Cas spéciaux de la distribution de Laplace asymétrique généralisée

Les deux premiers cas spéciaux sont des distributions de probabilité classiques. L'attention sera portée sur la distribution de Laplace asymétrique et son cas spécial, celle de Laplace symétrique. La dégénérée n'a aucune application pratique, mais en mentionner l'existence est pertinent, puisque ce ne sont pas toutes les distributions qui admettent ce cas.

### 2.3.1 Distribution de Laplace asymétrique

La distribution de Laplace asymétrique a été introduite par Hinkley et Revankar (1977) et étudiée en profondeur par Kozubowski et Podgórski (1999). Cette distribution a elle aussi plusieurs caractéristiques intéressantes pour remplacer la normale, dans le contexte de la modélisation financière, comme le présentent Kozubowski et Podgórski (2001). Elle est notée  $AL(\theta, \sigma, \kappa)$ .

Cette distribution, en tant que cas particulier de la distribution de Laplace asymétrique généralisée, conserve la majorité de ses propriétés. Entre autres, elle est infiniment divisible et

possède des moments finis. Les paramètres  $\theta$ ,  $\sigma$  et  $\kappa$  conservent leurs rôles et propriétés définis à la section 2.2.4. On peut donc contrôler la localisation, l'échelle et l'asymétrie de la distribution. Cependant, aucun paramètre ne permet de contrôler le coefficient d'aplatissement.

La densité de probabilité de cette distribution  $f(x; \theta, \sigma, \kappa)$  présente une forme analytique simple qui ne requiert pas le recours à la fonction de Bessel modifiée de troisième type (2.2.24).

$$f(x; \theta, \sigma, \kappa) = \frac{\sqrt{2}}{\sigma} \frac{\kappa}{1 + \kappa^2} \begin{cases} \exp\left(\frac{-\sqrt{2}\kappa}{\sigma}(x - \theta)\right), & x \geq \theta \\ \exp\left(\frac{\sqrt{2}}{\theta\kappa}(\theta - x)\right), & x < \theta. \end{cases} \quad (2.3.1)$$

La fonction caractéristique de cette distribution prend la forme suivante :

$$\phi(t; \theta, \sigma, \kappa) = \frac{e^{i\theta t}}{1 + \frac{\sigma^2 t^2}{2} - i \frac{\sigma}{\sqrt{2}} \left(\frac{1}{\kappa} - \kappa\right) t}. \quad (2.3.2)$$

Cette dernière est utilisée principalement lorsque l'application d'une propriété nécessite une distribution avec un seul paramètre d'échelle. La seconde paramétrisation de cette distribution utilise le paramètre  $\mu$  qui n'est pas invariant d'échelle. Cette paramétrisation donne une forme beaucoup plus simple de la fonction caractéristique (2.3.2) :

$$\phi(t\theta, \sigma, \mu) = \frac{e^{i\theta t}}{1 + \frac{\sigma^2 t^2}{2} - i\mu t}. \quad (2.3.3)$$

On reconnaît la fonction caractéristique d'un incrément du processus de Laplace (2.1.16) avec les paramètres ( $\tau = 1$ ,  $t = 1$ ). Une forme standardisée de cette distribution existe aussi avec le paramètre d'échelle  $\sigma = 1$  et de dérive  $\theta = 0$ . Enfin, on peut avoir une représentation alternative de cette distribution, utilisée principalement pour la simulation, basée sur l'exponentielle :

$$Y = \theta + \frac{\sigma}{\sqrt{2}} \left( \frac{1}{\kappa} W_1 - kW_2 \right) \quad (2.3.4)$$

où

$$W_1 \sim \text{Exponentielle} \left( \frac{\sigma}{\sqrt{2}\kappa} \right) \text{ et } W_2 \sim \text{Exponentielle} \left( \frac{\sigma\kappa}{\sqrt{2}} \right)$$

Cette méthode n'est cependant pas la plus efficace pour simuler une réalisation  $Y$  de la distribution de Laplace asymétrique, car on peut la faire à partir de deux réalisations indépendantes  $U_1$  et  $U_2$  d'une variable aléatoire uniforme définie sur l'intervalle  $[0, 1]$  :

$$Y = \theta + \frac{\sigma}{\sqrt{2}} \ln \frac{U_1^\kappa}{U_2^{1/\kappa}}. \quad (2.3.5)$$

Lorsque le paramètre  $\tau$  de la distribution de Laplace asymétrique généralisée est un entier, on peut générer une réalisation  $X$  de celle-ci en sommant un total de  $\tau$  réalisations indépendantes de la variable aléatoire  $Y$  :

$$X = \sum_{i=1}^{\tau} Y_i$$

$$Y_i \sim AL(\theta, \sigma, \kappa).$$

## 2.4 Relation avec le modèle de Madan et Seneta (1990)

Le modèle Variance Gamma de Madan et Seneta (1990) utilise une paramétrisation différente de celle de Kotz et al. (2001). On doit, afin de conserver une compatibilité des développements des prochains chapitres, détailler les changements de paramètres qui permettent de passer d'un modèle à l'autre. Le modèle Variance Gamma utilise les paramètres  $C, \sigma, \theta$  et  $\nu$ . Afin d'éviter la confusion, les paramètres seront indicés de 1 à 3 selon le modèle d'où ils proviennent. Le modèle 1 sera celui de Madan et Seneta (1990), les modèles 2 et 3 seront respectivement les paramétrisations en  $\mu$  et en  $\kappa$  de Kotz et al. (2001). Les différentes équivalences se trouvent à la table 2.3.

Vers le modèle		À partir du modèle		À partir du modèle	
<b>1</b>		<b>2</b>		<b>3</b>	
	$C$	=	$\theta_2$	=	$\theta_3$
	$\sigma_1$	=	$\sqrt{\tau_2}\sigma_2$	=	$\sqrt{\tau_3}\sigma_3$
	$\theta_1$	=	$\mu\tau_2$	=	$\tau_3\sigma_3(1/\kappa - \kappa)/\sqrt{2}$
	$\nu$	=	$1/\tau_2$	=	$1/\tau_3$
<b>2</b>		<b>1</b>		<b>3</b>	
	$\theta_2$	=	$C$	=	$\theta_3$
	$\sigma_2$	=	$\sigma_1\sqrt{\nu}$	=	$\sigma_3$
	$\mu$	=	$\theta_1 * \nu$	=	$\sigma_3(1/\kappa - \kappa)/\sqrt{2}$
	$\tau_2$	=	$1/\nu$	=	$\tau_3$
<b>3</b>		<b>1</b>		<b>2</b>	
	$\theta_3$	=	$C$	=	$\theta_2$
	$\sigma_3$	=	$\sigma_1\sqrt{\nu}$	=	$\sigma_2$
	$\kappa$	=	$\frac{\sqrt{2(\sigma_1\sqrt{\nu})^2 + (\theta_1*\nu)^2} \sim \theta_1*\nu}{\sigma_1\sqrt{\nu}\sqrt{2}}$	=	$\frac{\sqrt{2(\sigma_2)^2 + \mu^2} - \mu}{\sigma_2\sqrt{2}}$
	$\tau_3$	=	$1/\nu$	=	$\tau_2$

TABLE 2.3: Changements de paramétrisation





## Chapitre 3

# Approximation de la densité et de la fonction de répartition

Il existe plusieurs situations où il n'est pas possible d'obtenir une forme analytique des fonctions de densité et de répartition de la distribution d'une variable aléatoire. C'est particulièrement le cas pour la majorité des distributions issues d'un processus de Lévy, des sommes de longueur finie ou aléatoire de variables aléatoires et des mélanges de distributions. Cependant, dans la majorité de ces cas, il est facile d'obtenir la fonction caractéristique ou encore la fonction génératrice des moments. D'ailleurs, il pourrait être intéressant d'utiliser l'intégration numérique afin d'obtenir les transformées de Fourier ou de Laplace de ces fonctions. Par contre, la convergence de ces méthodes d'intégration est souvent lente, c'est-à-dire qu'il faut un pas de discrétisation très fin afin d'obtenir une approximation satisfaisante (Lugannani et Rice, 1980). Il est aussi possible d'utiliser l'algorithme de la transformée de Fourier rapide, mais celui-ci a comme désavantage d'être très peu précis dans l'évaluation des probabilités aux extrémités du support de la densité (figure 3.1).

C'est dans cette optique qu'ont été développées les approximations par la méthode du point de selle. Puisque celle-ci requiert l'utilisation de l'approximation de Laplace pour la fonction de densité et de l'approximation de Temme pour la fonction de répartition, on les introduit avant de développer la méthode à proprement parler. Il est à noter que la majorité de ces démonstrations sont tirées de Butler (2007) qui présente une monographie assez complète sur la méthode du point de selle.

### 3.1 L'approximation de Laplace

L'approximation de Laplace sera utilisée afin de développer la méthode du point de selle pour la fonction de densité. Soit  $g(x)$ , une fonction régulière sur l'intervalle  $[c, d]$  ayant un minimum global au point  $\hat{x} \in ]c, d]$ . L'approximation de Laplace de premier ordre prend la

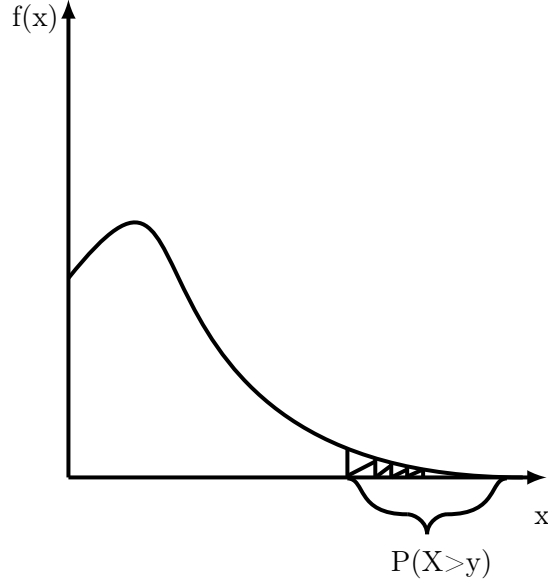


FIGURE 3.1: Probabilité à l'extrémité du support

forme suivante :

$$\int_c^d e^{-g(x)} dx \approx \frac{\sqrt{2\pi} e^{-g(\hat{x})}}{\sqrt{g''(\hat{x})}}. \quad (3.1.1)$$

La démonstration se fait facilement en utilisant le développement de Taylor de  $g(x)$  autour du point  $\hat{x}$  :

$$g(x) = g(\hat{x}) + g'(\hat{x})(x - \hat{x}) + \frac{1}{2}g''(\hat{x})(x - \hat{x})^2 + R. \quad (3.1.2)$$

La valeur  $R$  représente le reste du développement et est une quantité relativement petite. Comme  $\hat{x}$  est un minimum local, alors la dérivée première  $g'(\hat{x})$  vaut 0 et la dérivée seconde  $g''(\hat{x})$  est supérieure à 0. On obtient donc l'intégrale suivante, qui ressemble à celle de la fonction de répartition de la distribution normale :

$$e^{-g(\hat{x})} \int_c^d \exp \left\{ -\frac{1}{2}g''(\hat{x})(x - \hat{x})^2 \right\} dx \approx e^{-g(\hat{x})} \sqrt{\frac{2\pi}{g''(\hat{x})}}. \quad (3.1.3)$$

En utilisant un terme de plus du développement de Taylor (3.1.2), on obtient l'approximation de Laplace de second ordre, qui a une plus grande précision :

$$\int_c^d e^{-g(x)} dx \approx \frac{\sqrt{2\pi} e^{-g(\hat{x})}}{\sqrt{g''(\hat{x})}} \left\{ \frac{5}{24} \hat{\kappa}_3^2 - \frac{1}{8} \hat{\kappa}_4 \right\} \quad (3.1.4)$$

où

$$\hat{\kappa}_i = \frac{g^{(i)}(\hat{x})}{\{g''(\hat{x})\}^{i/2}}, \quad i \geq 3. \quad (3.1.5)$$

## 3.2 L'approximation de Temme

L'approximation de Temme sera utilisée afin de dériver la méthode du point de selle pour la fonction de répartition. On considère une variable aléatoire  $Z_{w_0}$  de distribution normale standard tronquée de telle sorte qu'elle ne puisse prendre des valeurs supérieures au point  $w_0$ .

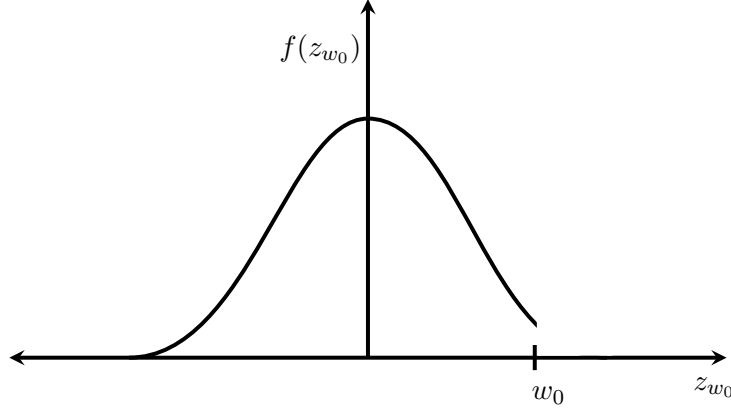


FIGURE 3.2: Distribution normale tronquée à  $w_0$

On définit l'espérance d'une fonction  $h$  de cette variable aléatoire, à partir des fonctions de densité  $\phi(w)$  et de répartition  $\Phi(w)$  de la loi normale, comme suit :

$$E[h(Z_{w_0})] = \frac{\int_{-\infty}^{w_0} h(w)\phi(w)dw}{\Phi(w_0)}. \quad (3.2.1)$$

Une approximation du numérateur est donnée par le développement suivant :

$$\begin{aligned} \int_{-\infty}^{w_0} h(w)\phi(w)dw &= h(0)\Phi(w_0) + \int_{-\infty}^{w_0} \frac{h(w) - h(0)}{w} w\phi(w)dw \\ &= h(0)\Phi(w_0) - \int_{-\infty}^{w_0} \frac{h(w) - h(0)}{w} d\phi(w) \quad (\text{car } w\phi(w)dw = -d\phi(w)) \\ &\approx h(0)\Phi(w_0) - \left[ \frac{h(w) - h(0)}{w} \phi(w) \right]_{w=-\infty}^{w=w_0}. \end{aligned}$$

On ignore l'intégrale de l'intégration par parties, puisqu'elle prend une petite valeur, pour obtenir l'approximation de Temme :

$$\int_{-\infty}^{w_0} h(w)\phi(w)dw \approx h(0)\Phi(w_0) + \phi(w_0) \left[ \frac{h(w_0) - h(0)}{w_0} \right]. \quad (3.2.2)$$

### 3.3 La méthode du point de selle

L'approximation de la distribution d'une variable aléatoire par la méthode du point de selle a été introduite par Daniels (1954). Il cherchait une façon d'estimer la distribution de la moyenne de variables aléatoires.

#### 3.3.1 Approximation de la densité

Soit une variable aléatoire  $Y$  dont on cherche à évaluer la fonction de densité  $f_Y(y)$ . Les fonctions génératrices des moments  $M_Y(s)$  et des cumulants  $K_Y(s)$  sont définies de manière équivalente par les intégrales suivantes, où l'on cherche une forme qui permettra d'utiliser l'approximation de Laplace :

$$M_Y(s) = e^{K_Y(s)} = \int_{-\infty}^{\infty} e^{sy} f_Y(y) dy \quad (3.3.1)$$

$$= \int_{-\infty}^{\infty} e^{sy + \ln f_Y(y)} dy \quad (3.3.2)$$

$$= \int_{-\infty}^{\infty} e^{-g(s,y)} dy. \quad (3.3.3)$$

Cette intégrale doit converger pour toute valeur  $s$  comprise dans un intervalle  $]-c_1, c_2[$  contenant l'origine et où la somme des bornes est positive :  $c_1 + c_2 > 0$ . Cet intervalle doit être choisi de sorte qu'il soit le plus grand possible. On considère pour l'instant que la fonction  $g(s, y) = -sy - \ln f_Y(y)$  répond aux conditions de l'approximation de Laplace énoncées au début de la section 3.1. On obtient alors l'approximation suivante pour une valeur de  $s$  fixée :

$$e^{K(s)} \approx \sqrt{\frac{2\pi}{g''(s, \hat{y}_s)}} e^{s\hat{y}_s} f_Y(\hat{y}_s). \quad (3.3.4)$$

La valeur  $\hat{y}_s$  est celle qui minimise la fonction  $g(s, y)$  pour un point  $s$  fixé. C'est donc la solution de la condition de premier ordre :

$$\begin{aligned} 0 &= -g'(s, \hat{y}_s) \\ &= s + \frac{\partial \ln(f_Y(\hat{y}_s))}{\partial \hat{y}_s} \\ s &= -\frac{\partial \ln(f_Y(\hat{y}_s))}{\partial \hat{y}_s}. \end{aligned} \quad (3.3.5)$$

L'hypothèse posée afin de faire l'approximation est que la fonction  $g(s)$  est convexe, ce qui implique que la dérivée seconde de celle-ci est positive ( $\frac{\partial^2 g(s,y)}{\partial y^2} > 0$ ) par définition. On examine la relation entre le point de selle  $s$  et le point  $\hat{y}_s$  déterminé par l'équation (3.3.5). En dérivant celle-ci par rapport au point  $\hat{y}_s$ , on constate qu'il existe une relation croissante et monotone

entre les deux :

$$\frac{ds}{d\hat{y}_s} = -\frac{\partial^2 \ln(f_Y(\hat{y}_s))}{\partial \hat{y}_s^2} > 0. \quad (3.3.6)$$

On doit maintenant trouver une solution de l'équation de premier ordre (3.3.5). Pour ce faire, on isole le terme  $\ln f_Y(y)$  dans l'équation (3.3.4) :

$$\ln f_Y(y) \approx K(s) - s\hat{y}_s - \frac{1}{2} \ln \left( \frac{2\pi}{-\left(\frac{\partial^2 \ln(f_Y(\hat{y}_s))}{\partial \hat{y}_s^2}\right)} \right). \quad (3.3.7)$$

Si l'on considère que le dernier terme de la soustraction est relativement constant par rapport au point  $\hat{y}_s$ , alors sa dérivée sera pratiquement nulle. Ce terme peut donc être négligé dans la prochaine étape, où on dérive (3.3.7) par rapport à  $y_s$  :

$$\frac{\partial \ln f_Y(\hat{y}_s)}{\partial \hat{y}_s} \approx (K'(s) - \hat{y}_s) \frac{\partial s}{\partial \hat{y}_s} - s. \quad (3.3.8)$$

En combinant l'approximation (3.3.8) et la condition de premier ordre (3.3.5), il en résulte que le terme de gauche de la partie de droite est approximativement égal à 0 :

$$(K'(s) - \hat{y}_s) \frac{\partial s}{\partial \hat{y}_s} \approx 0. \quad (3.3.9)$$

Cela conduit donc à l'équation qui met en relation le point de selle  $s$  et le point  $\hat{y}_s$  :

$$\hat{y}_s = K'(s). \quad (3.3.10)$$

Enfin, la dernière relation consiste à exprimer  $g''(s, \hat{y}_s)$  seulement en fonction du point de selle  $s$  en utilisant la condition de second ordre (3.3.6) et en dérivant l'équation précédente (3.3.10) en  $s$  :

$$\begin{aligned} g''(s, \hat{y}_s) &= -\frac{\partial^2 \ln(f_Y(\hat{y}_s))}{\partial \hat{y}_s^2} \\ &= \frac{ds}{d\hat{y}_s} \\ &= \left( \frac{d\hat{y}_s}{ds} \right)^{-1} \\ &= (K''(s))^{-1}. \end{aligned} \quad (3.3.11)$$

On peut donc réécrire l'approximation de Laplace (3.3.4) en utilisant les relations (3.3.10) et (3.3.11) :

$$\begin{aligned} e^{K(s)} &\approx \sqrt{2\pi K''(s)} \exp\{s\hat{y}_s\} f_Y(\hat{y}_s) \\ \hat{f}_Y(y_s) &\approx \frac{1}{\sqrt{2\pi K''(\hat{s})}} \exp\{K(\hat{s}) - \hat{s}y\} = \hat{f}_Y(y). \end{aligned} \quad (3.3.12)$$

L'équation (3.3.12) est une approximation par la méthode du point de selle de la densité de la variable aléatoire  $Y$  au point  $y_s$ .

Par convention, on considère que l'approximation  $\hat{f}_Y(y)$  est une fonction de densité, mais, elle n'en est pas exactement une puisqu'elle n'intègre pas à 1. Cependant, il est possible de la normaliser en calculant la valeur de l'intégrale de celle-ci sur le domaine  $\chi$  de la distribution, puis en divisant l'approximation  $\hat{f}_Y(y)$  par celle-ci :

$$c = \int_{\chi} \hat{f}_Y(y) dy \neq 1 \quad (3.3.13a)$$

$$\bar{f}_Y(y) = c^{-1} \hat{f}_Y(y), \quad \forall y \in \chi. \quad (3.3.13b)$$

### 3.3.2 Unicité du point de selle

Daniels (1954) démontre que l'équation du point de selle (3.3.10) a une solution unique dans l'intervalle  $[-c_1, c_2]$  sélectionné précédemment lorsque  $y_0$  est situé dans un intervalle  $[a, b]$ .

Pour ce faire, il présente une condition essentielle. Les limites inférieure et supérieure de la dérivée première de la fonction génératrice des cumulants doivent correspondre aux points  $a$  et  $b$  respectivement (3.3.14) :

$$\lim_{s \rightarrow c_2} K'(s) = b \quad (3.3.14)$$

$$\lim_{s \rightarrow -c_1} K'(s) = a. \quad (3.3.15)$$

On définit la fonction génératrice des moments de la différence entre la variable aléatoire  $Y$  et le point  $y_0$  par  $M(s, y)$ .

Quand  $a < y_0 < b$ , la dérivée première prend la forme suivante :

$$M'(s, y_0) = \int_a^b (y - y_0) e^{s(y - y_0)} dF(y). \quad (3.3.16)$$

Les limites inférieure et supérieure de la fonction  $M'(s, y_0)$  sont respectivement  $M'(-\infty, y_0) = -\infty$  et  $M'(\infty, y_0) = \infty$ . De plus, la dérivée de celle-ci est toujours positive  $M''(s, y_0) > 0$ . Donc, pour chaque valeur  $a < y_0 < b$ , une racine unique  $\hat{s}$  de l'équation  $M'(s, y_0) = 0$  existe nécessairement. Par conséquent, l'équation du point de selle (3.3.10), qui lui est équivalente, a aussi une solution réelle unique.

### 3.3.3 Approximation de la fonction de répartition

La fonction de répartition de la variable aléatoire  $Y$  au point  $y$  est définie comme étant la probabilité  $P(y \leq Y)$  et s'obtient en intégrant la densité sur l'intervalle  $[-\infty, y]$ . On considère

le changement de variable  $y \mapsto \hat{w}$  défini comme suit :

$$\frac{\hat{w}^2}{2} = \hat{s}y - K(\hat{s}). \quad (3.3.17)$$

La valeur  $\hat{w}_y$  est obtenue en résolvant l'équation du point de selle pour la racine  $\hat{s}$  et en remplaçant celle-ci dans l'équation (3.3.17). Ce changement de variable est une transformation monotone, continue et croissante dont la dérivée est donnée par l'équation suivante :

$$\frac{dy}{d\hat{w}} = \begin{cases} \hat{w}/\hat{s}, & \hat{s} \neq 0 \\ \sqrt{K''(0)}, & \hat{s} = 0. \end{cases} \quad (3.3.18)$$

L'approximation de premier ordre de la fonction de répartition s'obtient en intégrant celle de premier ordre de la densité (3.3.12) sur ce même intervalle à l'aide du changement de variable (3.3.17) :

$$\begin{aligned} F_Y(y) &\approx \int_{-\infty}^y \frac{1}{\sqrt{2\pi K''(s)}} \exp\{K(s) - sx\} dx \\ &= \int_{-\infty}^y \frac{1}{\sqrt{K''(s)}} \phi(\hat{w}) dx \\ &= \int_{-\infty}^{\hat{w}_y} \frac{\hat{w}}{\hat{s}\sqrt{K''(s)}} \phi(\hat{w}) d\hat{w}. \end{aligned} \quad (3.3.19)$$

On applique l'approximation de Temme à l'intégrale précédente avec l'équation :

$$h(\hat{w}) = \frac{\hat{w}}{\hat{s}\sqrt{K''(s)}}.$$

On obtient ainsi l'approximation de premier ordre de la fonction de répartition :

$$\begin{aligned} F_Y(y) &\approx \Phi(\hat{w}_y) + \phi(\hat{w}_y) \frac{1 - h(\hat{w}_y)}{\hat{w}_y} \\ &\approx \Phi(\hat{w}_y) + \phi(\hat{w}_y) \left( \frac{1}{\hat{w}_y} - \frac{1}{\hat{u}_y} \right). \end{aligned} \quad (3.3.20)$$

Les deux constantes  $\hat{w}_y$  et  $\hat{u}_y$  sont respectivement définies comme suit :

$$\hat{w}_y = \text{sgn}(\hat{s}) \sqrt{2\{\hat{s}y - K(\hat{s})\}} \quad (3.3.21a)$$

$$\hat{u}_y = \hat{s} \sqrt{K''(\hat{s})}. \quad (3.3.21b)$$

### 3.3.4 Quelques propriétés des approximations

Les approximations par la méthode du point de selle ont pour principale propriété de respecter le principe d'invariance. Cette propriété permet notamment d'obtenir une approximation

de la densité ainsi que la fonction de répartition en utilisant des paramètres estimés à partir d'un échantillon de données centrées et réduites. Puis, elle permet d'appliquer la transformation linéaire  $Y = \sigma X + \mu$ , telle qu'utilisée à la section 2.2.5, tout en conservant la même approximation.

Une autre propriété de ces approximations est qu'elle respecte la symétrie de la distribution. Ainsi, lorsque la distribution est asymétrique, on doit s'attendre à ce que l'approximation le soit aussi.

## 3.4 Application de la méthode du point de selle

### 3.4.1 Approximation de la densité

On applique la méthode du point de selle à la distribution de Laplace asymétrique généralisée, afin d'obtenir une approximation de la densité (2.2.25) et aussi pouvoir représenter la fonction de répartition, car la première n'est pas intégrable analytiquement. Cependant, pour des fins de simplification, la paramétrisation  $\mu$  sera utilisée. On rappelle qu'on peut passer d'une forme à l'autre à l'aide des équations (2.2.4) et (2.2.5).

On évalue d'abord la dérivée première de la fonction génératrice des cumulants (2.2.13) par rapport à la variable de transformation  $s$  :

$$\begin{aligned} K'_Y(s) &= \frac{\partial}{\partial s} \ln \left( \frac{e^{\theta s}}{\left(1 - \frac{1}{2}\sigma^2 s^2 - \mu s\right)^\tau} \right) \\ &= \frac{\sigma^2 \theta s^2 + (2\mu\theta - 2\sigma^2 \tau) s - 2\theta - 2\mu\tau}{\sigma^2 s^2 + 2\mu s - 2}, \end{aligned} \quad (3.4.1)$$

$$\text{où } 1 - \frac{1}{2}\sigma^2 s^2 - \mu s > 0. \quad (3.4.2)$$

On détermine ensuite le point de selle  $\hat{s}_y$  correspondant à la valeur  $y$  pour laquelle on veut estimer la fonction de densité ou de répartition, à l'aide de l'équation (3.3.10) et de la première dérivée de la fonction génératrice des cumulants (3.4.1) :

$$\begin{aligned} y &= K'_Y(\hat{s}_y) \\ &= \frac{\sigma^2 \theta \hat{s}_y^2 + (2\mu\theta - 2\sigma^2 \tau) \hat{s}_y - 2\theta - 2\mu\tau}{\sigma^2 \hat{s}_y^2 + 2\mu \hat{s}_y - 2}. \end{aligned}$$

L'équation précédente a deux solutions dont une seule correspond à un minimum. Pour déterminer quelle solution est appropriée, on doit évaluer la dérivée seconde en ces deux points :

$$\hat{s}_y = \frac{\pm \sqrt{2\sigma^2 y^2 + \mu^2 y^2 - 4\sigma^2 \theta y - 2\mu^2 \theta y + 2\sigma^2 \theta^2 + \mu^2 \theta^2 + \sigma^4 \tau^2}}{\sigma^2 (y - \theta)} - \frac{\mu}{\sigma^2} - \frac{\tau}{y - \theta}. \quad (3.4.3)$$



On évalue la dérivée seconde de la fonction génératrice des cumulants :

$$K_Y''(s) = \frac{2(s^2\sigma^4 + 2\mu s\sigma^2 + 2\sigma^2 + 2\mu^2)\tau}{(s^2\sigma^2 + 2\mu s - 2)^2} > 0. \quad (3.4.4)$$

On peut maintenant déterminer quelle valeur  $\hat{s}_y$  utiliser, car la fonction  $K_Y''(s)$  sera positive pour celle-ci. De manière équivalente, on utilise la condition 2.2.11, qui permet l'existence de la fonction génératrice des moments, afin d'identifier plus aisément la solution unique.

On peut maintenant évaluer (3.3.12), pour obtenir l'approximation de premier ordre de la fonction de densité de la distribution de Laplace asymétrique généralisée.

### 3.4.2 Approximation de la fonction de répartition

En évaluant les valeurs  $\hat{w}_y$  et  $\hat{u}_y$  (3.3.21) à l'aide de celles du point de selle (3.4.3), de la fonction génératrice des cumulants (2.2.13) et de la dérivée seconde de cette dernière (3.4.4), on obtient l'approximation de premier ordre de la fonction de répartition.

Tout au long de ce chapitre, on a supposé que les paramètres de la distribution étaient connus. Cependant, ceux-ci doivent être estimés à partir des données d'un échantillon représentatif de la population étudiée. C'est d'ailleurs ce qui sera fait dans les chapitres suivants.



# Chapitre 4

## Méthode des moments généralisée

Si l'on pose l'hypothèse selon laquelle un échantillon issu d'une population a une certaine distribution de probabilité, on doit ensuite déterminer quels sont les paramètres de celle-ci. Ces paramètres peuvent être estimés à l'aide de différentes méthodes statistiques. La méthode des moments généralisée est une de celles-ci.

### 4.1 Introduction

On considère un échantillon de taille  $T$  formé de plusieurs réalisations  $\mathbf{y} = y(1), \dots, y(T)$  indépendantes entre elles et identiquement distribuées. Selon le modèle paramétrique, cet échantillon est formé de réalisations d'une variable aléatoire  $Y$  suivant une distribution particulière. Le vecteur de paramètres de celle-ci,  $\theta$ , de longueur  $a$ , appartenant à l'espace  $\Omega \subset \mathbb{R}^a$ , doit être estimé à partir de l'échantillon. Une première approche consiste à maximiser la fonction de vraisemblance  $LL(\theta; \mathbf{y})$ , qui équivaut au produit de la densité  $f(y; \theta)$  évaluée à chacune des réalisations  $y(t)$  :

$$LL(\theta; \mathbf{y}) = \prod_{t=1}^T f(y(t); \theta), \quad \theta \in \Omega. \quad (4.1.1)$$

L'estimateur  $\hat{\theta}$  est, dans ce cas, la valeur pour laquelle l'échantillon a la plus grande probabilité d'être observé. Cependant, puisque cela apporte plusieurs simplifications intéressantes, on cherchera à maximiser la fonction de log-vraisemblance  $l(\theta; \mathbf{y}) = \ln(L(\theta; \mathbf{y}))$  :

$$\hat{\theta} = \arg \max_{\theta} l(\theta; \mathbf{y}). \quad (4.1.2)$$

Le maximum est obtenu en résolvant la condition de premier ordre, c'est-à-dire en égalant le gradient de la fonction de log-vraisemblance à 0 :

$$\frac{\partial l(\theta; \mathbf{y})}{\partial \theta} = 0. \quad (4.1.3)$$

Cette méthode permet d'obtenir un estimateur convergent de variance minimale, car elle utilise l'ensemble de l'information contenue dans l'échantillon. Cependant, la fonction de vraisemblance doit pouvoir être spécifiée sous une forme analytique, et de plus, celle-ci doit être différentiable afin de pouvoir utiliser cette méthode, ce qui n'est pas le cas avec la distribution de Laplace asymétrique généralisée. Lorsque cette situation se présente, une méthode alternative est préférée, celle des moments généralisée, comme proposée par Hansen (1982). Celle-ci est décrite par Hamilton (1994) dans le contexte de l'étude des séries chronologiques. Elle a pour avantage de nécessiter seulement la spécification de certaines conditions de moment. Par contre, elle n'utilise pas toute l'information fournie par l'échantillon, ce qui ne permettra pas d'obtenir un estimateur de variance minimale.

#### 4.1.1 Méthode classique des moments

Pour certaines distributions, on ne peut directement estimer les paramètres. On cherchera alors des fonctions des paramètres qui sont facilement estimables de manière convergente. L'ensemble de fonctions le plus commun qui répond à cette condition est celui des moments, d'où le nom de la méthode.

La méthode classique des moments a été introduite par Pearson en 1894. On considère un échantillon de taille  $T$  dont les observations seront notées  $y(t)$ . On veut estimer le vecteur de paramètres  $\theta$ , de longueur  $a$  de la distribution. On définit les moments  $\mathbf{m} = m_1, \dots, m_a$  de la population totale, représentée par la variable aléatoire  $Y$ , comme étant l'espérance des puissances de celle-ci, et donc une fonction des paramètres de la distribution. On considèrera le même nombre de moments que de paramètres à estimer.

$$m_i(\theta) = E[Y^i], \quad i = 1, \dots, a \quad (4.1.4)$$

Cette méthode consiste à résoudre un système d'équations (4.1.5) où l'on égale les moments empiriques  $m_i(\theta)$  à ceux de la distribution  $\hat{m}(\mathbf{y})$ .

$$\left\{ \begin{array}{l} m_1(\theta) = \frac{1}{T} \sum_{t=1}^T y_t \\ m_2(\theta) = \frac{1}{T} \sum_{t=1}^T y_t^2 \\ \vdots \\ m_a(\theta) = \frac{1}{T} \sum_{t=1}^T y_t^a \end{array} \right\} \quad (4.1.5)$$

L'estimateur des moments  $\hat{\theta}_T$  est celui qui résout ce système.

## 4.2 Méthode des moments généralisée

La méthode classique des moments utilise le même nombre d'équations d'estimation que de paramètres ( $r = a$ ). De plus, le système d'équations formé par celles-ci doit admettre une

solution réelle appartenant à l'espace des paramètres  $\Omega$ , ce qui n'est pas toujours le cas. Lorsque ces deux conditions ne sont pas réunies, on doit choisir un vecteur de paramètres  $\theta$  pour lequel les moments de la population  $m_i$  ont une valeur la plus près possible de ceux de l'échantillon  $\hat{m}_i$  correspondants. Cette distance est notée par le vecteur  $g(\theta; \mathbf{y})$  et correspond au cas le plus simple de la méthode des moments généralisée.

$$g(\theta; \mathbf{y}) = \begin{bmatrix} m_1 - \hat{m}_1 \\ \vdots \\ m_r - \hat{m}_r \end{bmatrix}. \quad (4.2.1)$$

Pour obtenir ces estimateurs, on cherchera plutôt à minimiser une fonction objectif notée  $Q(\theta; \mathbf{y})$ , qui correspond à une norme quadratique pondérée par une matrice définie positive  $W$  :

$$Q(\theta; \mathbf{y}) \equiv g(\theta; \mathbf{y})' W g(\theta; \mathbf{y}). \quad (4.2.2)$$

Hansen (1982) nomme cette procédure «méthode des moments généralisée». Elle est aussi nommée «méthode du  $\chi^2$  minimum» par Berkson (1980), bien qu'elle n'en soit qu'un cas particulier. On retrouve aussi le nom d'estimateur de distance minimale, par Wolfowitz (1957).

#### 4.2.1 Définition

On considère un vecteur  $\mathbf{y}_T$  de longueur  $T$  contenant les données  $y(t)$  de l'échantillon tirées d'une population représentée par la variable aléatoire  $Y$ . On considère de plus un vecteur de paramètres  $\theta \in \Omega$  de longueur  $a$  dont la vraie valeur, qui est inconnue, est représentée par la constante  $\theta_0$ . Soit une fonction vectorielle de longueur  $r$  de la variable aléatoire  $Y$  appelée condition de moment ou d'orthogonalité :

$$h(\theta, Y) : (\mathbb{R}^a \times \mathbb{R}) \longrightarrow \mathbb{R}^r. \quad (4.2.3)$$

L'espérance de cette fonction, sous l'hypothèse  $\theta = \theta_0$ , est un vecteur nul noté  $\mathbf{0}_r$ .

$$E[h(\theta_0, Y)] = \mathbf{0}_r. \quad (4.2.4)$$

On définit aussi la fonction  $g(\theta, \mathbf{y}_T) : \mathbb{R}^a \longrightarrow \mathbb{R}^r$  comme étant la moyenne empirique des conditions de moment  $h(\theta, y)$  :

$$g(\theta, \mathbf{y}_T) \equiv \frac{1}{T} \sum_{t=1}^T h(\theta, y(t)). \quad (4.2.5)$$

L'idée derrière la méthode des moments généralisée est de choisir un ensemble de paramètres  $\theta$  de sorte que la valeur de la fonction  $g(\theta, \mathbf{y}_T)$  soit aussi près que possible du vecteur nul

$\mathbf{0}_r$ . Selon la norme utilisée pour mesurer cette distance, les propriétés de l'estimateur  $\hat{\theta}_T$  vont varier. Étant donné les nombreuses propriétés bien établies dans le domaine des statistiques, on utilise la norme quadratique avec pondération, appelée aussi moindres carrés généralisés, qui prend la forme suivante :

$$Q(\theta, \mathbf{y}_T) = [g(\theta, \mathbf{y}_T)]' W_T [g(\theta, \mathbf{y}_T)]. \quad (4.2.6)$$

Cette fonction de minimisation permettra d'utiliser des tests statistiques basés sur la distribution  $\chi^2$  de Pearson. La matrice carrée  $W_T$  de dimension  $r \times r$  est définie positive et est habituellement une fonction des données de l'échantillon  $\mathbf{y}_T$  et des paramètres  $\theta$ . Une matrice de pondération optimale sera déterminée à la section 4.2.3.

## 4.2.2 Convergence

Si le nombre de paramètres  $a$  est égal à celui de conditions de moments  $r$ , alors la fonction objectif atteindra un minimum de 0 au point  $\hat{\theta}$ . On obtiendra ce dernier en résolvant l'équation suivante pour le paramètre  $\theta$  :

$$g(\hat{\theta}_T, \mathbf{y}_T) = 0. \quad (4.2.7)$$

Lorsque le nombre de conditions de moments est plus grand que celui des paramètres, on ne pourra pas obtenir une solution pour l'équation précédente (4.2.7). La proximité entre la valeur de chaque condition de moment et 0 sera déterminée par la matrice de pondération  $W_T$ . Étant donné que la fonction  $g(\hat{\theta}_T, \mathbf{y}_T)$  est la moyenne échantillonnale de la fonction aléatoire  $h(\theta_0, Y)$ , on a, par la loi des grands nombres, la relation suivante entre ces deux quantités :

$$g(\theta, \mathbf{y}_T) \xrightarrow{P} E[h(\theta, Y)]. \quad (4.2.8)$$

On considère la suite d'observations  $\{y(t)\}_{t=1}^T$  comme un processus stochastique. L'ensemble des conditions de régularité suivantes permet d'obtenir un estimateur convergent (Hansen, 1982).

1. Le processus stochastique  $\{y(t)\}_{t=1}^T$  est **stationnaire**, dont la distribution conjointe de plusieurs observations ne change pas dans le temps :

$$F_Y(y(t(1) + \tau), \dots, y(t(k) + \tau)) = F_Y(y(t(1)), \dots, y(t(k))), \quad \forall \tau \in \mathbb{R}. \quad (4.2.9)$$

Il est aussi **ergodique**, c'est-à-dire que l'on peut déduire les propriétés du processus à partir d'un échantillon (ou réalisation) suffisamment long de celui-ci. Entre autres, la moyenne et la variance d'un échantillon recueilli sur une période en particulier sont représentatives de celles de n'importe quel autre intervalle de temps de ce processus.

Dans ces deux dernières situations, on parle aussi de convergence en moyenne ( $\mathbb{L}^1$ ) et en norme quadratique ( $\mathbb{L}^2$ ) :

$$\lim_{T \rightarrow \infty} E(|Y_T - Y|^r) = 0, \quad r \in \{1, 2\}. \quad (4.2.10)$$

2. L'**espace métrique**  $(\Omega, \sigma)$ , défini par l'espace des paramètres  $\Omega$  et la norme valeur absolue est **séparable** :

$$\sigma = |Y_T - Y|^r. \quad (4.2.11)$$

Cette condition définit l'unicité du vecteur de paramètres, car si la distance est nulle, les estimateurs correspondent aux vrais paramètres :

$$\sigma = 0 \iff \hat{\theta} = \theta_0, \quad \theta_0 \in \Omega. \quad (4.2.12)$$

3. La fonction  $h(\theta, \mathbf{y})$  est **mesurable au sens de Borel** pour chaque vecteur  $\theta$ , c'est-à-dire qu'un sous-ensemble de l'espace des paramètres  $\Omega$  existe pour chaque valeur qu'elle peut prendre. De plus, la fonction est continue sur l'ensemble  $\Omega$  pour chaque échantillon  $\mathbf{y}$ .
4. L'espérance de la fonction  $h(\theta, Y)$  existe et est définie pour toute valeur  $\theta \in \Omega$ . De plus, par définition, l'espérance de la fonction pour les vrais paramètres est de 0 :

$$E[h(\theta_0, Y)] = 0.$$

5. La séquence de matrices de pondération  $\{W_T\}_{T=1}^\infty$  converge presque sûrement, élément par élément, vers une constante  $W_0$ , en utilisant la norme valeur absolue définie précédemment (4.2.11).

$\hat{\theta}_T \in \Omega$  est un estimateur convergent de  $\theta$  lorsque les conditions précédentes sont respectées.

$$\begin{aligned} \hat{\theta}_T &= \arg \min_{\theta \in \Omega} Q(\theta, \mathbf{y}_T) \\ &= \arg \min_{\theta \in \Omega} [g(\theta, \mathbf{y}_T)]' W_T [g(\theta, \mathbf{y}_T)]. \end{aligned} \quad (4.2.13)$$

### 4.2.3 Matrice de pondération optimale

On définit la variance-covariance  $\mathbf{S}(\theta; \mathbf{y})$  de la moyenne échantillonnale de la fonction  $h(\theta, \mathbf{y})$ . Cette matrice est formée par l'espérance, élément par élément, du produit extérieur de l'estimateur par sa transposée, multiplié par la taille de l'échantillon  $T$

$$\mathbf{S}(\theta; \mathbf{y}) = T \cdot E \{ [h(\theta, \mathbf{y})] [h(\theta, \mathbf{y})]' \}. \quad (4.2.14)$$

La variance-covariance asymptotique de la moyenne échantillonnale de la fonction  $h(\theta_0, y(t))$  est obtenue en évaluant la matrice (4.2.14) au point  $\theta_0$ .

$$\mathbf{S}(\theta_0; \mathbf{y}) = T \left[ \cdot \lim_{T \rightarrow \infty} E \{ [h(\theta, \mathbf{y})] [h(\theta, \mathbf{y})]' \} \right]_{\theta=\theta_0} \quad (4.2.15)$$

La valeur optimale de la matrice de pondération  $W_T$  de l'équation (4.2.13) est obtenue en inversant la variance-covariance asymptotique :

$$W_T = \mathbf{S}^{-1}(\theta_0; \mathbf{y}_T). \quad (4.2.16)$$

Cependant, comme on ne connaît pas la valeur de  $\theta_0$ , on utilisera l'estimateur convergent  $\hat{\theta}$ , qui minimise la condition suivante :

$$Q_T(\theta, \mathbf{y}_T) = [g(\theta, \mathbf{y}_T)]' \mathbf{S}_T^{-1}(\theta; \mathbf{y}_T) [g(\theta, \mathbf{y}_T)]. \quad (4.2.17)$$

Le problème d'optimisation se note alors comme suit :

$$\hat{\theta}_T = \arg \min_{\theta \in \Omega} Q_T(\theta, \mathbf{y}_T). \quad (4.2.18)$$

Comme la séquence  $\{h(\theta_0, y(t))\}_{t=-\infty}^{\infty}$  ne présente pas de corrélation sérielle, on pourrait estimer la variance-covariance  $S_T$  de manière convergente en évaluant la moyenne empirique du produit extérieur de la condition de moment (Hamilton, 1994, p.413) :

$$\mathbf{S}_T^*(\theta; \mathbf{y}_T) = \frac{1}{T} \sum_{t=1}^T [g(\theta_0, y(t))] [g(\theta_0, y(t))]' . \quad (4.2.19)$$

L'estimateur  $\hat{\mathbf{S}}_T(\hat{\theta}; \mathbf{y}_T)$  converge en probabilité vers la vraie valeur de la matrice  $\mathbf{S}(\theta_0; \mathbf{y}_T)$ . Étant donné que l'on estime la fonction  $h(\theta, \mathbf{y}_T)$  à l'aide de la fonction  $g(\hat{\theta}; \mathbf{y}_T)$ , on a aussi la convergence en probabilité :

$$\hat{\mathbf{S}}_T(\hat{\theta}; \mathbf{y}_T) = \frac{1}{T} \sum_{t=1}^T \left[ g(\hat{\theta}_T, w(t)) \right] \left[ g(\hat{\theta}_T, w(t)) \right]' \xrightarrow{P} \mathbf{S}(\theta_0; \mathbf{y}_T). \quad (4.2.20)$$

L'estimateur de la matrice de pondération optimale est alors défini comme étant l'inverse de la variance-covariance estimée :

$$\hat{W}_T = \hat{\mathbf{S}}_T^{-1}(\hat{\theta}; \mathbf{y}_T). \quad (4.2.21)$$

Comme cette matrice dépend de l'estimateur  $\hat{\theta}$ , qui est pour l'instant inconnu, elle ne pourra pas être utilisée pour une première optimisation de l'équation de minimisation (4.2.13). Cependant, elle pourra être utilisée dans une procédure itérative à la section 4.2.4. On devra considérer l'utilisation d'un point de départ alternatif pour le vecteur de paramètres dans l'algorithme de minimisation.



#### 4.2.4 Méthode des moments généralisée itérative

La méthode des moments généralisée itérative de Hall (2005) permet de contourner le problème de l'estimation de la matrice  $\hat{\mathbf{S}}_T(\hat{\theta}; \mathbf{y}_T)$ . Elle consiste, en premier lieu, à calculer un estimateur préliminaire  $\hat{\theta}^{(0)}$ , en utilisant la matrice identité  $W_T = I_r$  dans l'équation (4.2.13). On suggère d'utiliser un vecteur de valeurs initiales  $\hat{\theta}^I$  obtenues par une autre méthode d'estimation, lorsque possible, comme point de départ de l'optimisation numérique.

À l'aide de l'estimateur initial  $\hat{\theta}^{(0)}$ , on obtient une première évaluation de la matrice de pondération :

$$W_T = \left[ \hat{\mathbf{S}}_T(\hat{\theta}^{(0)}; \mathbf{y}_T) \right]^{-1}. \quad (4.2.22)$$

En utilisant la matrice  $W_T$  comme pour pondérer la fonction objectif (4.2.13), on obtiendra un nouvel estimateur  $\hat{\theta}_T^{(1)}$ .

Par la suite, on répète cette procédure jusqu'à ce qu'on obtienne deux estimateurs consécutifs ( $\hat{\theta}_T^{(j)}$  et  $\hat{\theta}_T^{(j+1)}$ ) qui ne sont pas significativement différents, selon un critère d'arrêt  $\epsilon_i < \epsilon$ . Ce critère d'arrêt prendra ici la forme suivante, qui correspond à la norme euclidienne de la différence entre les deux derniers estimateurs obtenus :

$$\epsilon_i = \sqrt{\left[ \hat{\theta}_T^{(i)} - \hat{\theta}_T^{(i-1)} \right]' \left[ \hat{\theta}_T^{(i)} - \hat{\theta}_T^{(i-1)} \right]}. \quad (4.2.23)$$

On peut aussi fixer un nombre maximal d'itérations  $i_{max}$  si l'on ne parvient pas à obtenir le niveau de précision voulu. Par contre, dans cette situation, on préfère utiliser un autre point de départ.

Chaque estimateur  $\hat{\theta}_T^{(i)}$  a la même distribution asymptotique. La méthode itérative a pour avantage, en pratique, que les estimateurs produits sont invariants d'échelle et ne dépendent pas de la matrice  $W_T$  initiale.

#### 4.2.5 Distribution asymptotique des estimateurs

On considère  $\hat{\theta}_T$ , la valeur qui minimise la fonction objectif  $Q(\theta, \mathbf{y}_T)$  (4.2.17). Cette minimisation équivaut à résoudre le système non linéaire où l'on égale la dérivée de l'équation d'optimisation à 0 :

$$\begin{aligned} \frac{d}{d\theta'} Q(\theta, \mathbf{y}_T) &= \left[ \frac{d}{d\theta'} g(\theta, \mathbf{y}_T) \Big|_{\theta=\hat{\theta}} \right]' \cdot \hat{W}_T \cdot g(\theta, \mathbf{y}_T) \\ &= 0. \end{aligned} \quad (4.2.24)$$

où le gradient  $D'(\theta, \mathbf{y}_T)$  est

$$\begin{aligned} D'(\theta, \mathbf{y}_T) &= \left[ \frac{d}{d\theta'} g(\theta, \mathbf{y}_T) \right]' \\ &= \begin{bmatrix} \frac{d}{d\theta_1} g_1(\theta, \mathbf{y}_T) & \cdots & \frac{d}{d\theta_1} g_a(\theta, \mathbf{y}_T) \\ \vdots & \ddots & \vdots \\ \frac{d}{d\theta_k} g_1(\theta, \mathbf{y}_T) & \cdots & \frac{d}{d\theta_k} g_a(\theta, \mathbf{y}_T) \end{bmatrix}. \end{aligned} \quad (4.2.25)$$

En appliquant le théorème central limite multivarié (B.2.5) à l'estimateur des conditions de moments  $g(\theta, \mathbf{y}_T)$ , on obtient que, lorsque la taille  $T$  est suffisamment grande, celui-ci converge en loi vers un vecteur aléatoire de distribution normale multivariée de moyenne  $\mathbf{0}_a$  et de variance-covariance  $T^{-1}\mathbf{S}(\theta; \mathbf{y}_T)$  :

$$\sqrt{T}g(\theta, \mathbf{y}_T) \xrightarrow{L} N(\mathbf{0}_a, \mathbf{S}(\theta; \mathbf{y}_T)). \quad (4.2.26)$$

Cette relation permet de conclure que l'estimateur  $\hat{\theta}$  est gaussien. On n'a donc qu'à calculer sa variance-covariance asymptotique.

Soit  $\{\hat{\mathbf{S}}_T\}_{T=1}^{\infty}$ , une séquence de matrices  $(r \times r)$  définies positives qui convergent en probabilité vers la variance-covariance asymptotique :

$$\hat{\mathbf{S}}_T \xrightarrow{P} \mathbf{S}.$$

On ajoute que la fonction  $g(\theta, \mathbf{y}_T)$  doit être différentiable par rapport au vecteur  $\theta$  pour tout échantillon  $\mathbf{y}_T$ .

On doit préalablement poser un ensemble de conditions supplémentaires de régularité :

1. L'estimateur  $\hat{\theta}_T$  converge en probabilité vers la vraie valeur des paramètres  $\theta_0$  :

$$\hat{\theta}_T \xrightarrow{P} \theta_0. \quad (4.2.27)$$

2. Le théorème central limite s'applique pour la fonction  $g(\theta, \mathbf{y}_T)$  (4.2.26).
3. Pour toute séquence d'estimateurs  $\{\theta_T^*\}_{T=1}^{\infty}$  convergents en probabilité  $\theta_T^* \xrightarrow{P} \theta_0$ , on peut évaluer le gradient  $D'(\theta, \mathbf{y}_T)$  de l'équation  $g(\theta, \mathbf{y}_T)$  (4.2.25) à l'aide de limites :

$$\begin{aligned} D'_T(\theta, \mathbf{y}_T) &\equiv \text{plim} \left\{ \frac{dg(\theta, \mathbf{y}_T)}{d\theta'} \Big|_{\theta=\theta_T^*} \right\} \\ &= \text{plim} \left\{ \frac{dg(\theta, \mathbf{y}_T)}{d\theta'} \Big|_{\theta=\theta_0} \right\}. \end{aligned} \quad (4.2.28)$$

On note que les colonnes de la matrice  $D$  sont linéairement indépendantes.

Pour obtenir la distribution asymptotique de l'estimateur  $\hat{\theta}$ , on utilise le premier ordre du développement de Taylor de la fonction  $g(\hat{\theta}, \mathbf{y}_T)$  autour de la valeur du vrai paramètre  $\theta_0$ , tel qu'avancé par Gourieroux et Monfort (1989) :

$$g(\hat{\theta}, \mathbf{y}_T) = g(\theta_0, \mathbf{y}_T) + D'_T(\theta, \mathbf{y}_T) (\hat{\theta} - \theta_0). \quad (4.2.29)$$

On multiplie de part et d'autre par la matrice  $\left( \frac{dg(\theta, \mathbf{y}_T)}{d\theta'} \Big|_{\theta=\theta_T} \times W_T \right)$  de dimension  $(a \times r)$  :

$$\begin{aligned} \left( \frac{dg(\theta, \mathbf{y}_T)}{d\theta'} \Big|_{\theta=\theta_T} \times W_T \right) \times g(\hat{\theta}, \mathbf{y}_T) &= \left( \frac{dg(\theta, \mathbf{y}_T)}{d\theta'} \Big|_{\theta=\theta_T} \times W_T \right) \times g(\theta_0, \mathbf{y}_T) \\ &+ \left( \frac{dg(\theta, \mathbf{y}_T)}{d\theta'} \Big|_{\theta=\theta_T} \times W_T \right) \times D'_T(\theta, \mathbf{y}_T) (\hat{\theta} - \theta_0). \end{aligned} \quad (4.2.30)$$

L'équation de premier ordre (4.2.24) nous indique que le côté gauche de l'égalité précédente (4.2.30) vaut 0. On retrouve alors une expression de la distance entre l'estimateur et la vraie valeur des paramètres, qui dépend de la matrice de pondération, de la fonction  $g(\theta_0, \mathbf{y}_T)$  et du gradient  $D'_T(\theta, \mathbf{y}_T)$  :

$$\begin{aligned} (\hat{\theta} - \theta_0) &= - \left[ \frac{dg(\theta, \mathbf{y}_T)}{d\theta'} \Big|_{\theta=\theta_T} \times W_T \times D'_T(\theta, \mathbf{y}_T) \right]^{-1} \\ &\times \frac{dg(\theta, \mathbf{y}_T)}{d\theta'} \Big|_{\theta=\theta_T} \times W_T \times g(\theta_0, \mathbf{y}_T). \end{aligned} \quad (4.2.31)$$

La condition de régularité (4.2.28) permet la convergence de chaque rangée de l'estimateur  $D'_T(\theta, \mathbf{y}_T)$  vers celles du gradient  $D'(\theta_0, \mathbf{y}_T)$ . De plus, l'équation (4.2.31) implique la relation de convergence suivante :

$$\begin{aligned} \sqrt{T} (\hat{\theta} - \theta_0) &\xrightarrow{P} - \{D(\theta, \mathbf{y}_T) W_T D'(\theta, \mathbf{y}_T)\}^{-1} \\ &\times \{D(\theta, \mathbf{y}_T) W_T \sqrt{T} \cdot g(\theta_0, \mathbf{y}_T)\}. \end{aligned} \quad (4.2.32)$$

Afin de simplifier la notation, on définit la constante  $C(\theta, \mathbf{y}_T)$  :

$$C(\theta, \mathbf{y}_T) = - \{D(\theta, \mathbf{y}_T) W_T D'(\theta, \mathbf{y}_T)\}^{-1} \times D(\theta, \mathbf{y}_T) W_T.$$

L'équation (4.2.32) devient

$$\sqrt{T} (\hat{\theta} - \theta_0) \xrightarrow{P} C(\theta, \mathbf{y}_T) \sqrt{T} \cdot g(\theta_0, \mathbf{y}_T). \quad (4.2.33)$$

En combinant la relation (4.2.26), où l'on applique le théorème central limite à la fonction  $g(\theta_0, \mathbf{y}_T)$ , avec la méthode delta multivariée (B.3.4), on retrouve la forme suivante, avec une convergence en loi cependant, puisque celle-ci est moins forte que celle en probabilité :

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{L} N(0, \mathcal{J}_0^{-1}) \quad (4.2.34)$$

où

$$\begin{aligned} \mathcal{J}_0^{-1} &= C(\theta, \mathbf{y}_T) \{W_T\}^{-1} C(\theta, \mathbf{y}_T) \\ &= \{D(\theta, \mathbf{y}_T) W_T D'(\theta, \mathbf{y}_T)\}^{-1} D(\theta, \mathbf{y}_T) W_T \{W_T\}^{-1} \\ &\quad \times W_T D'(\theta, \mathbf{y}_T) \{D(\theta, \mathbf{y}_T) W_T D'(\theta, \mathbf{y}_T)\}^{-1} \\ &= \{D(\theta, \mathbf{y}_T) W_T D'(\theta, \mathbf{y}_T)\}^{-1}. \end{aligned} \quad (4.2.35)$$

### 4.3 Estimation sous contraintes

La méthode des moments généralisée suppose que le vrai vecteur de paramètres  $\theta_0$  appartient à l'ensemble  $\Omega$ . En pratique, les paramètres sont souvent soumis à certaines contraintes à l'égalité.

On définit un ensemble de  $q$  contraintes linéaires implicites appliquées au vecteur de paramètres  $\theta$  de longueur  $a$  :

$$\left\{ \begin{array}{l} r_{(1,1)}\theta_1 + \dots + r_{(1,a)}\theta_a = r_{(1,0)} \\ \dots \\ r_{(q,1)}\theta_1 + \dots + r_{(q,a)}\theta_a = r_{(q,0)} \end{array} \right\}. \quad (4.3.1)$$

On peut les présenter sous la forme d'un système matriciel :

$$\underbrace{\begin{bmatrix} r_{(1,1)} & \dots & r_{(1,a)} \\ \vdots & \ddots & \vdots \\ r_{(q,1)} & \dots & r_{(q,a)} \end{bmatrix}}_{\mathbf{R}} \times \underbrace{\begin{bmatrix} \theta_1 \\ \vdots \\ \theta_a \end{bmatrix}}_{\boldsymbol{\theta}} = \underbrace{\begin{bmatrix} r_{(1,0)} \\ \vdots \\ r_{(q,0)} \end{bmatrix}}_{\mathbf{r}}. \quad (4.3.2)$$

Afin de les inclure dans un problème de minimisation, on préférera utiliser la notation  $a(\theta) = R\theta - r$ . On notera au passage que le gradient du vecteur de contraintes équivaut à la matrice de coefficients :

$$\frac{\partial}{\partial \theta'} a(\theta) = R. \quad (4.3.3)$$

Ainsi, on peut estimer les paramètres de la distribution contrainte à l'aide de la méthode des moments généralisée, de manière analogue à la distribution non contrainte, comme il a été

présenté à la section précédente. On utilisera la technique du multiplicateur de Lagrange afin d'inclure la contrainte  $a(\theta)$  dans l'équation de minimisation (4.2.18). Le vecteur  $\gamma$  associe un multiplicateur à chaque contrainte linéaire. On définit le lagrangien  $\mathcal{L}(\tilde{\theta})$  mettant en relation la fonction objectif  $Q_T(\theta)$  (4.2.17) et les contraintes  $a(\theta)$  :

$$\mathcal{L}(\theta) = -Q_T(\theta) - a(\theta)' \gamma. \quad (4.3.4)$$

L'estimateur contraint  $\tilde{\theta}$  est obtenu en maximisant ce lagrangien :

$$\tilde{\theta} = \arg \max_{\theta \in \Omega} \mathcal{L}(\theta). \quad (4.3.5)$$

La solution optimale s'obtient en résolvant les conditions de premier ordre par rapport au vecteur de paramètres  $\theta$  et celui des multiplicateurs de Lagrange  $\gamma$  :

$$\frac{\partial}{\partial \theta} Q_T(\tilde{\theta}) - \frac{\partial}{\partial \theta'} a(\tilde{\theta})' \gamma_T = 0 \quad (4.3.6)$$

$$a(\tilde{\theta}) = 0. \quad (4.3.7)$$

On s'intéresse aussi à la distribution asymptotique de cet estimateur contraint. Pour ce faire, on doit développer les conditions de premier ordre comme il a été fait à la section 4.2.5 pour l'estimateur non contraint.

### 4.3.1 Distribution asymptotique des estimateurs contraints

Supposons que les conditions de premier ordre sont deux fois continûment dérivables par rapport au vecteur  $\theta$ . On développe les équations (4.3.6) et (4.3.7) autour de la vraie valeur du paramètre contraint  $\theta_0$ . Puis, on les multiplie par les facteurs  $\frac{1}{\sqrt{T}}$  et  $\sqrt{T}$  respectivement. Notons que la fonction  $a(\theta)$  vaut 0 au point  $\theta_0$ , ce qui permettra de simplifier la seconde équation :

$$\frac{1}{\sqrt{T}} \frac{\partial}{\partial \theta} Q_T(\theta_0) + \frac{1}{T} \frac{\partial^2}{\partial \theta \partial \theta'} Q_T(\theta_0) \sqrt{T}(\tilde{\theta} - \theta_0) - \frac{\partial}{\partial \theta'} a(\tilde{\theta})' \frac{\gamma_T}{\sqrt{T}} \approx 0 \quad (4.3.8a)$$

$$\frac{\partial}{\partial \theta'} a(\theta_0) \sqrt{T}(\tilde{\theta} - \theta_0) \approx 0. \quad (4.3.8b)$$

On définit la matrice d'information de Fisher comme étant la limite de l'espérance de la valeur de la dérivée seconde de la fonction objectif. Au point  $\theta_0$ , on identifie l'estimateur de cette matrice par  $\mathcal{J}_0$ , la variance-covariance de l'estimateur non contraint (4.2.35) :

$$\mathcal{J}_0 = \lim_{T \rightarrow \infty} -\frac{1}{T} \frac{\partial^2}{\partial \theta \partial \theta'} Q_T(\theta_0). \quad (4.3.9)$$

On reprend l'équivalent asymptotique de l'équation (4.3.8a) pour l'estimateur non contraint :

$$\frac{1}{\sqrt{T}} \frac{\partial}{\partial \theta} Q_T(\theta_0) - \mathcal{J}_0 \sqrt{T}(\hat{\theta} - \theta_0) \approx 0. \quad (4.3.10)$$

En combinant ces deux dernières expressions, on peut formuler les conditions de premier ordre (4.3.8) comme étant asymptotiquement des fonctions linéaires de l'estimateur non contraint  $\sqrt{T}(\hat{\theta} - \theta_0)$  :

$$\mathcal{J}_0 \sqrt{T}(\hat{\theta} - \theta_0) - \mathcal{J}_0 \sqrt{T}(\tilde{\theta} - \theta_0) - \frac{\partial}{\partial \theta} a(\theta_0)' \frac{\gamma_T}{\sqrt{T}} \approx 0 \quad (4.3.11)$$

$$\frac{\partial}{\partial \theta'} a(\theta_0) \sqrt{T}(\tilde{\theta} - \theta_0) \approx 0. \quad (4.3.12)$$

En réorganisant la première équation (4.3.11), on obtient :

$$\sqrt{T}(\tilde{\theta} - \theta_0) \approx \sqrt{T}(\hat{\theta} - \theta_0) - \mathcal{J}_0^{-1} \frac{\partial}{\partial \theta} a(\theta_0)' \frac{\gamma_T}{\sqrt{T}}. \quad (4.3.13)$$

En la reportant dans la seconde équation (4.3.12), on obtient :

$$\frac{\partial}{\partial \theta'} a(\theta_0) \sqrt{T}(\hat{\theta} - \theta_0) - \frac{\partial}{\partial \theta'} a(\theta_0) \mathcal{J}_0^{-1} \frac{\partial}{\partial \theta} a(\theta_0)' \frac{\gamma_T}{\sqrt{T}} + \frac{\partial}{\partial \theta'} a(\theta_0) \sqrt{T}(\tilde{\theta} - \theta_0) \approx 0. \quad (4.3.14)$$

Comme le rang de la matrice  $\frac{\partial}{\partial \theta'} a(\theta_0)$  est égal au nombre de contraintes  $r$ , alors  $\frac{\partial}{\partial \theta'} a(\theta_0) \mathcal{J}_0^{-1} \frac{\partial}{\partial \theta} a(\theta_0)'$  est inversible, et l'on peut donc isoler le multiplicateur de Lagrange  $\frac{\gamma_T}{\sqrt{T}}$  en fonction des estimateurs contraints  $\sqrt{T}(\tilde{\theta} - \theta_0)$  et non contraints  $\sqrt{T}(\hat{\theta} - \theta_0)$  :

$$\frac{\gamma_T}{\sqrt{T}} \approx \left( \frac{\partial}{\partial \theta'} a(\theta_0) \mathcal{J}_0^{-1} \frac{\partial}{\partial \theta} a(\theta_0)' \right)^{-1} \frac{\partial}{\partial \theta'} a(\theta_0) \sqrt{T}(\hat{\theta} - \theta_0). \quad (4.3.15)$$

On définit l'estimateur contraint en fonction de l'estimateur non contraint en utilisant le lagrangien (4.3.15) dans la condition (4.3.13) :

$$\begin{aligned} \sqrt{T}(\tilde{\theta} - \theta_0) &\approx \left( I - \underbrace{\mathcal{J}_0^{-1} \frac{\partial}{\partial \theta} a(\theta_0)' \left( \frac{\partial}{\partial \theta'} a(\theta_0) \mathcal{J}_0^{-1} \frac{\partial}{\partial \theta} a(\theta_0)' \right)^{-1} \frac{\partial}{\partial \theta'} a(\theta_0)}_P \right) \mathcal{J}_0^{-1} \sqrt{T}(\hat{\theta} - \theta_0) \\ &\approx (I - P) \mathcal{J}_0^{-1} \sqrt{T}(\hat{\theta} - \theta_0). \end{aligned} \quad (4.3.16)$$

La variance asymptotique de l'estimateur contraint est donc, à partir du résultat précédent (4.3.16) et de la définition (4.3.10) :

$$V \left[ \sqrt{T}(\tilde{\theta} - \theta_0) \right] = (I - P) \mathcal{J}_0^{-1}. \quad (4.3.17)$$

L'estimateur  $\tilde{\theta}$  suit donc asymptotiquement une distribution normale multivariée de moyenne  $\theta_0$  et de variance  $T(I - P)\mathcal{J}_0^{-1}$  :

$$\tilde{\theta} \sim \mathcal{N}(\theta_0, T(I - P)\mathcal{J}_0^{-1}). \quad (4.3.18)$$

## 4.4 Tests d'hypothèses paramétriques

Les tests d'hypothèses paramétriques sont utilisés afin d'évaluer une hypothèse concernant les paramètres d'une distribution, en fonction d'un échantillon de données. Afin d'effectuer ces tests, on présume que la différence entre l'estimateur  $\hat{\theta}$  et la vraie valeur des paramètres  $\theta_0$  suit une distribution normale multivariée. Les hypothèses sont habituellement formulées sous la forme de contraintes linéaires, ainsi, les statistiques de test sont obtenues à partir du calcul matriciel. Les trois tests les plus couramment utilisés dans le cadre de l'estimation par maximum de vraisemblance peuvent être adaptés à la méthode des moments généralisée (Newey et McFadden, 1994). Pour l'ensemble de ces tests, l'hypothèse nulle correspond à la contrainte linéaire suivante :

$$H_0 : a(\theta) = R\theta - r = 0. \quad (4.4.1)$$

### 4.4.1 Test de Wald

Le test de Wald permet de vérifier si la différence entre l'estimateur non contraint  $\hat{\theta}$  et l'estimateur contraint  $\tilde{\theta}$  est significative. Les contraintes linéaires posées ne seront pas applicables lorsque le résultat est positif. Pour ce faire, on doit connaître la distribution asymptotique de celle-ci. Comme la distribution asymptotique des deux estimateurs est normale, alors celle de cette différence l'est aussi :

$$(\hat{\theta} - \tilde{\theta}) \sim \mathcal{N}(0, TP\mathcal{J}_0^{-1}). \quad (4.4.2)$$

On obtient l'espérance et la variance de la statistique  $\sqrt{T}(\hat{\theta} - \tilde{\theta})$  en utilisant le fait que la somme de deux variables aléatoires normales l'est aussi :

$$\begin{aligned} E[\sqrt{T}(\hat{\theta} - \tilde{\theta})] &= E[\sqrt{T}(\hat{\theta} - \theta_0)] - E[\sqrt{T}(\tilde{\theta} - \theta_0)] \\ &= \theta_0 - \theta_0 \\ &= 0 \end{aligned} \quad (4.4.3)$$

$$\begin{aligned} V[\sqrt{T}(\hat{\theta} - \tilde{\theta})] &= V[\sqrt{T}(\hat{\theta} - \theta_0)] - V[\sqrt{T}(\tilde{\theta} - \theta_0)] \\ &= (I - (I - P))\mathcal{J}_0^{-1} \\ &= P\mathcal{J}_0^{-1}. \end{aligned} \quad (4.4.4)$$

On définit la statistique  $\chi^{WALD,1}$ , qui a asymptotiquement une distribution  $\chi^2$  avec  $q$  degrés de liberté :

$$\begin{aligned}\chi^{WALD,1} &= T \left( \hat{\theta} - \tilde{\theta} \right)' P \left( \hat{\theta} - \tilde{\theta} \right) \\ &= T \left( \hat{\theta} - \tilde{\theta} \right)' \mathcal{J}_0^{-1} \frac{\partial}{\partial \theta} a(\theta_0)' \left( \frac{\partial}{\partial \theta'} a(\theta_0) \mathcal{J}_0^{-1} \frac{\partial}{\partial \theta} a(\theta_0)' \right)^{-1} \frac{\partial}{\partial \theta'} a(\theta_0) \left( \hat{\theta} - \tilde{\theta} \right).\end{aligned}\tag{4.4.5}$$

On rejettera l'hypothèse nulle (4.4.1) lorsque la valeur de la statistique  $\chi^{WALD,1}$  sera supérieure à un seuil critique  $\chi_{q,1-\alpha}^2$  au niveau de confiance  $1 - \alpha$ .

Une version asymptotiquement équivalente de ce test qui ne requiert pas de calculer la valeur de l'estimateur contraint existe. Ce test est équivalent lorsque les contraintes définissent certains paramètres comme des constantes. On vérifie si un cas particulier d'une distribution s'applique, par exemple avec celle de Laplace asymétrique généralisée.

On définit alors la statistique  $\chi^{WALD,2}$ , qui a aussi asymptotiquement une distribution  $\chi^2$  avec  $q$  degrés de liberté. Par contre, ici, on teste si la valeur de la contrainte linéaire est significativement différente de 0 :

$$\chi^{WALD,2} = T a'(\hat{\theta}) \left( \frac{\partial}{\partial \theta'} a(\theta_0) \mathcal{J}_0^{-1} \frac{\partial}{\partial \theta} a'(\theta_0) \right)^{-1} a(\hat{\theta}).\tag{4.4.6}$$

#### 4.4.2 Test du multiplicateur de Lagrange

Le test du multiplicateur de Lagrange, ou du score, introduit par Newey et West (1987), est basé uniquement sur l'estimateur contraint et est équivalent asymptotiquement au test de Wald présenté à la section précédente. Il vérifie l'application de la contrainte (4.3.2) à l'estimateur  $\tilde{\theta}$ . Selon la définition du lagrangien (4.3.4), si la contrainte est vérifiée, alors la restriction  $a(\tilde{\theta})$  vaudra 0. Selon la condition de premier ordre (4.3.7), la dérivée de la fonction objectif de l'estimateur non contraint  $\frac{\partial}{\partial \theta} Q_T(\theta)$ , le score,  $\tilde{\theta}$  devrait aussi être égale à 0. On cherchera donc à tester si cette valeur est significativement différente de 0.

On définit la statistique du multiplicateur de Lagrange  $\chi^{LM,1}$ , qui a asymptotiquement une distribution  $\chi_q^2$  avec  $q$  degrés de liberté :

$$\chi^{LM,1} = T \frac{\partial}{\partial \theta} Q_T(\tilde{\theta}) P \frac{\partial}{\partial \theta} Q_T(\tilde{\theta}).\tag{4.4.7}$$

On rejettera l'hypothèse nulle (4.4.1) lorsque la valeur de la statistique  $\chi^{LM,1}$  sera supérieure à un seuil critique  $\chi_{q,1-\alpha}^2$  au niveau de confiance  $1 - \alpha$ .

On peut construire un test équivalent, basé sur la valeur du multiplicateur, dont la matrice de variance-covariance est l'inverse de celle de la contrainte. On définit alors la statistique  $\chi^{LM,2}$



suivante :

$$\chi^{LM,2} = T\gamma' \frac{\partial}{\partial\theta'} a(\theta_0) \mathcal{J}_0^{-1} \frac{\partial}{\partial\theta} a'(\theta_0) \gamma. \quad (4.4.8)$$

#### 4.4.3 Test basé sur la statistique de métrique de distance

La statistique de métrique de distance est basée sur la différence entre les valeurs minimales de la fonction objectif  $Q_T(\theta)$  obtenues lors de l'optimisation avec contraintes (4.3.4) et sans contraintes (4.2.13).

On définit la statistique  $\chi^{DM}$  :

$$\chi^{DM} = -T \left[ Q_T(\tilde{\theta}) - Q_T(\hat{\theta}) \right]. \quad (4.4.9)$$

Cette statistique a asymptotiquement une distribution  $\chi_q^2$  avec  $q$  degrés de liberté. Elle est l'analogie de la statistique du ratio de vraisemblance dans le cadre de l'estimation par la méthode du maximum de vraisemblance.

Le test basé sur la métrique de distance vérifie que la contrainte (4.3.2) posée lors de l'estimation du vecteur  $\tilde{\theta}$  est valide. Un des désavantages de ce test est qu'il requiert deux optimisations. Par contre, on peut facilement récupérer les valeurs de  $Q_T(\theta)$  lors de l'estimation.

On rejettera l'hypothèse nulle (4.4.1) lorsque la valeur de la statistique  $\chi^{DM}$  sera supérieure à un seuil critique  $\chi_{q,1-\alpha}^2$  au niveau de confiance  $1 - \alpha$ .

#### 4.4.4 En résumé

On rassemble les différentes statistiques permettant d'effectuer un test d'hypothèse paramétrique à la table 4.1.

Statistique	Valeur
$\chi^{WALD,1}$	$T \left( \hat{\theta} - \tilde{\theta} \right)' \mathcal{J}_0^{-1} \frac{\partial}{\partial\theta} a(\theta_0)' \left( \frac{\partial}{\partial\theta'} a(\theta_0) \mathcal{J}_0^{-1} \frac{\partial}{\partial\theta} a(\theta_0)' \right)^{-1} \frac{\partial}{\partial\theta'} a(\theta_0) \left( \hat{\theta} - \tilde{\theta} \right)$
$\chi^{WALD,2}$	$T a'(\hat{\theta}) \left( \frac{\partial}{\partial\theta'} a(\theta_0) \mathcal{J}_0^{-1} \frac{\partial}{\partial\theta} a'(\theta_0) \right)^{-1} a(\hat{\theta})$
$\chi^{LM,1}$	$T \frac{\partial}{\partial\theta} Q_T(\tilde{\theta}) P \frac{\partial}{\partial\theta} Q_T(\tilde{\theta})$
$\chi^{LM,2}$	$T\gamma' \frac{\partial}{\partial\theta'} a(\theta_0) \mathcal{J}_0^{-1} \frac{\partial}{\partial\theta} a'(\theta_0) \gamma$
$\chi^{DM}$	$-T \left[ Q_T(\tilde{\theta}) - Q_T(\hat{\theta}) \right]$

TABLE 4.1: Tests d'hypothèse paramétriques pour la méthode des moments généralisée



## Chapitre 5

# Méthode de l'équation d'estimation optimale

Une équation d'estimation est une fonction des données de l'échantillon et des paramètres d'un modèle qui spécifie de quelle manière on doit procéder pour estimer ces derniers, lorsque la distribution de la population est inconnue. Cette approche a un avantage sur les méthodes de vraisemblance, car elle ne requiert pas l'utilisation de la fonction de densité ou de répartition, mais seulement les moments de la distribution. (Everitt et Skrondal, 2006)

Crowder (1986) définit l'équation d'estimation sous une forme générale, composée d'une matrice d'échelle  $S(\theta; Y)$ , un vecteur de fonctions aléatoires  $u(\theta; Y)$  et un vecteur de pondération  $\mathbf{w}(\theta; Y)$  :

$$g(\theta; Y) = S^{-1}(\theta; Y) \sum_{t=1}^T u(\theta; Y) \mathbf{w}(\theta; Y). \quad (5.0.1)$$

En faisant abstraction du vecteur de pondération  $\mathbf{w}(\theta; Y)$ , on retrouve une forme qui rappelle la méthode des moments généralisée. La fonction d'estimation  $u(\theta; Y)$  définit, dans ce cas, des conditions de moment. Toutefois, en faisant abstraction de la matrice  $S(\theta; Y)$ , on obtient une classe d'équations d'estimation  $g(\theta; Y)$  qui généralise plusieurs méthodes connues, comme développé par Crowder (1987) :

$$g(\theta; Y) = \sum_{t=1}^T u(\theta; Y) \mathbf{w}(\theta; Y). \quad (5.0.2)$$

La propriété fondamentale de toute équation d'estimation est que son espérance est nulle. L'équation d'estimation (5.0.1), et donc (5.0.2), produit des estimateurs sans biais :

$$E [g(\theta; Y)] = 0. \quad (5.0.3)$$

Par analogie avec la méthode du maximum de vraisemblance, on pourra alors la considérer comme une équation de quasi-score. On peut en effet représenter l'équation précédente par celle du score (5.0.4) donnée par l'espérance de la dérivée de la fonction de log-vraisemblance par rapport au vecteur de paramètres  $\theta$  :

$$E \left[ g(\theta; Y)^{EMV} \right] = E \left[ \frac{d \ln L(\theta; Y)}{d\theta} \right] = 0. \quad (5.0.4)$$

On désigne la moyenne et la variance de la distribution par les notations  $\mu(\theta)$  et  $\sigma^2(\theta)$  respectivement. On définit aussi les dérivées premières de la moyenne et de l'écart-type par rapport au vecteur de paramètres :

$$\mu'(\theta) = \frac{d\mu(\theta)}{d\theta} \quad (5.0.5a)$$

$$\sigma'(\theta) = \frac{d\sqrt{\sigma^2(\theta)}}{d\theta}. \quad (5.0.5b)$$

Dans ce chapitre, on considère une équation d'estimation quadratique de la forme suivante :

$$g(\theta; Y) = \sum_{t=1}^n [\mathbf{a}(\theta; y_t)(y_t - \mu(\theta)) + \mathbf{b}(\theta; y_t) ((y_t - \mu(\theta))^2 - \sigma^2(\theta))]. \quad (5.0.6)$$

Le vecteur de pondération  $\mathbf{w}(\theta; Y)$  de l'équation (5.0.2) est composé de deux fonctions déterministes :  $\mathbf{a}(\theta; Y)$  et  $\mathbf{b}(\theta; Y)$  :

$$\mathbf{w}(\theta; Y) = \begin{bmatrix} \mathbf{a}(\theta; Y) \\ \mathbf{b}(\theta; Y) \end{bmatrix}. \quad (5.0.7)$$

On démontre facilement que c'est une équation d'estimation en calculant son espérance :

$$\begin{aligned} E[g(\theta; Y)] &= E \left[ \sum_{t=1}^n [\mathbf{a}(\theta; y_t)(y_t - \mu(\theta)) + \mathbf{b}(\theta; y_t) ((y_t - \mu(\theta))^2 - \sigma^2(\theta))] \right] \\ &= \sum_{t=1}^n (\mathbf{a}(\theta; y_t) \cdot E[y_t - \mu(\theta)] + \mathbf{b}(\theta; y_t) \cdot E[(y_t - \mu(\theta))^2 - \sigma^2(\theta)]) \\ &= 0. \end{aligned} \quad (5.0.8)$$

La forme quadratique regroupe certaines méthodes d'estimation bien établies, recensées à la table 5.1.

La méthode d'estimation gaussienne de Whittle présente un avantage par rapport aux modèles linéaires conventionnels (moindres carrés pondérés et maximum de quasi-vraisemblance) du fait qu'elle prend en considération la variance comme critère d'optimisation et non seulement comme pondération.

Méthode	$\mathbf{a}(\theta; y_t)$	$\mathbf{b}(\theta; y_t)$
Moindres carrés non pondérés	$\mu'(\theta)$	0
Maximum de quasi-vraisemblance	$\frac{\mu'(\theta)}{\sigma^2(\theta)}$	0
Estimation gaussienne de Whittle (Fox et Taqqu, 1986)	$\frac{\mu'(\theta)}{\sigma^2(\theta)}$	$\frac{\sigma'(\theta)}{\sigma^2(\theta)^{3/2}}$

TABLE 5.1: Méthodes d'estimation représentables par la forme quadratique (5.0.6)

Crowder (1987) développe, à partir de cette dernière, une équation d'estimation optimale qui vise à remplacer la méthode de quasi-vraisemblance basée sur la moyenne et la variance jusqu'alors utilisée.

## 5.1 Équation d'estimation optimale

Soit une équation d'estimation de la forme générale (5.0.2). On a, par définition, que l'espérance de la composante stochastique  $u(\theta; Y_t)$  de celles-ci sous les vrais paramètres  $\theta_0$  est nulle. De plus, si l'on considère que les données ne sont pas corrélées, l'espérance du produit de deux de ces fonctions est nulle, sauf lorsqu'elle est évaluée au même point  $Y_t$  :

$$E[u(\theta_0; Y_t)] = 0 \quad (5.1.1)$$

$$E[u(\theta_0; Y_s)u(\theta_0; Y_t)] = \delta_{s,t} \text{Var}[u(\theta_0; Y_t)]. \quad (5.1.2)$$

Les matrices non singulières  $M(\theta_0)$  et  $V(\theta_0)$  représentent respectivement, pour l'équation d'estimation, l'espérance de sa dérivée par rapport à  $\theta$  et l'inverse de sa variance-covariance, évaluée à la vraie valeur des paramètres  $\theta_0$  :

$$M(\theta_0; Y) = \sum_{t=1}^T \mathbf{w}(\theta_0) E[u'(\theta_0; Y_t)] \quad (5.1.3a)$$

$$V(\theta_0; Y) = \sum_{t=1}^T \mathbf{w}(\theta_0) \text{Var}[u(\theta_0; Y_t)] \mathbf{w}^T(\theta_0). \quad (5.1.3b)$$

La variance asymptotique de l'estimateur  $\hat{\theta}$  prend alors la forme suivante, en utilisant la méthode delta multivariée :

$$\text{Var}[\hat{\theta}] = M^{-1}(\theta_0) V(\theta_0) [M^{-1}(\theta_0)]^T. \quad (5.1.4)$$

On désire obtenir des équations d'estimation optimales de telle sorte que la variance soit minimale. Le *théorème 4.1* de Crowder (1986) permet d'établir quel vecteur de pondération  $\mathbf{w}^*(\theta)$  utiliser.

Selon ce théorème, lorsqu'on utilise le vecteur de pondération optimal (5.1.5), on obtient que la dérivée de la fonction d'estimation  $M^*(\theta)$  soit égale à l'inverse de la matrice de variance-covariance  $V^*(\theta)$  :

$$\mathbf{w}^*(\theta) = \{E [u'(\theta_0; Y_t)]\}^T \{Var [u(\theta; Y_t)]\}^{-1}. \quad (5.1.5)$$

La variance asymptotique (5.1.4) du vecteur de paramètres optimal  $\hat{\theta}^*$  qui résout l'équation d'estimation  $g^*(\theta)$  se simplifie alors sous la forme :  $[V^*(\theta_0; y_t)]^{-1}$

$$\begin{aligned} M^*(\theta_0) &= \sum_{t=1}^T \{E [u'(\theta_0; Y_t)]\}^T \{Var [u(\theta_0; Y_t)]\}^{-1} E [u'(\theta_0; Y_t)] \quad (5.1.6) \\ V^*(\theta_0) &= \sum_{t=1}^T \{E [u'(\theta_0; Y_t)]\}^T \{Var [u(\theta_0; Y_t)]\}^{-1} Var [u(\theta_0; Y_t)] \\ &\quad \times \left[ \{E [u'(\theta_0; Y_t)]\}^T \{Var [u(\theta_0; Y_t)]\}^{-1} \right]^T \\ &= \sum_{t=1}^T \{E [u'(\theta_0; Y_t)]\}^T \{Var [u(\theta_0; Y_t)]\}^{-1} E [u'(\theta_0; Y_t)] \\ &\Leftrightarrow M^*(\theta_0) = V^*(\theta_0) \\ &\Leftrightarrow Var(\hat{\theta}^*) = [M^*(\theta_0)]^{-1} V^*(\theta_0) \left[ [M^*(\theta_0)]^{-1} \right]^T \\ &= [V^*(\theta_0)]^{-1}. \quad (5.1.7) \end{aligned}$$

On considère la fonction  $u(\theta; Y_t)$  de l'équation quadratique (5.0.6) , dont on cherche le vecteur de pondération optimal correspondant selon la proposition (5.1.5) :

$$u(\theta; Y_t) = \begin{bmatrix} Y_t - \mu(\theta) \\ (Y_t - \mu(\theta))^2 - \sigma^2(\theta) \end{bmatrix}. \quad (5.1.8)$$

On évalue l'espérance, sous le vecteur de vrais paramètres  $\theta_0$ , de la dérivée par rapport à  $\theta$  de la fonction  $u(\theta; Y)$  :

$$\begin{aligned} E [u'(\theta; Y_t)] &= \begin{bmatrix} E [-\mu'(\theta)] \\ E \left[ -2(Y - \mu(\theta))\mu'(\theta) - 2\sqrt{\sigma^2(\theta)}\sigma'(\theta) \right] \end{bmatrix} \\ &= \begin{bmatrix} -\mu'(\theta) \\ -2\sqrt{\sigma^2(\theta)}\sigma'(\theta) \end{bmatrix}. \quad (5.1.9) \end{aligned}$$

Puis, on évalue la variance-covariance du vecteur  $u(\theta)$ . Pour ce faire, on devra évaluer séparément chaque élément composant cette matrice :

$$Var [u(\theta; Y_t)] = \begin{bmatrix} Var [Y - \mu(\theta)] & Cov [Y - \mu(\theta), (Y - \mu(\theta))^2 - \sigma^2(\theta)] \\ Cov [Y - \mu(\theta), (Y - \mu(\theta))^2 - \sigma^2(\theta)] & Var [(Y - \mu(\theta))^2 - \sigma^2(\theta)] \end{bmatrix}. \quad (5.1.10)$$

On définit les coefficients d'asymétrie et d'aplatissement normalisés, ainsi qu'une constante  $\gamma_3(\theta)$  qui en découle, afin de simplifier les expressions qui seront obtenues conséquemment :

$$\begin{aligned}\gamma_1(\theta) &= \frac{m_3(\theta)}{m_2(\theta)^{3/2}} \\ &= \frac{E[(Y - \mu(\theta))^3]}{\sigma^2(\theta)^{3/2}}\end{aligned}\quad (5.1.11a)$$

$$\begin{aligned}\gamma_2(\theta) &= \frac{m_4(\theta)}{m_2(\theta)^2} - 3 \\ &= \frac{E[(Y - \mu(\theta))^4]}{\sigma^2(\theta)^2} - 3\end{aligned}\quad (5.1.11b)$$

$$\gamma_3(\theta) = \gamma_2(\theta) + 2 - \gamma_1(\theta)^2. \quad (5.1.11c)$$

On développe les différentes composantes de la matrice : (5.1.10)

$$\begin{aligned}Var[Y - \mu(\theta)] &= E[(Y - \mu(\theta))^2] - E[Y - \mu(\theta)]^2 \\ &= \sigma^2(\theta)\end{aligned}\quad (5.1.12)$$

$$\begin{aligned}Cov[Y - \mu(\theta), (Y - \mu(\theta))^2 - \sigma^2(\theta)] &= E[(Y - \mu(\theta))((Y - \mu(\theta))^2 - \sigma^2(\theta))] \\ &\quad - E[Y - \mu(\theta)]E[(Y - \mu(\theta))^2 - \sigma^2(\theta)] \\ &= E[(Y - \mu(\theta))^3 - \sigma^2(\theta)(Y - \mu(\theta))] \\ &= \sigma^2(\theta)^{3/2}\gamma_1(\theta)\end{aligned}\quad (5.1.13)$$

$$\begin{aligned}Var[(Y - \mu(\theta))^2 - \sigma^2(\theta)] &= E[((Y - \mu(\theta))^2 - \sigma^2(\theta))((Y - \mu(\theta))^2 - \sigma^2(\theta))] \\ &= E[(Y - \mu(\theta))^4 - 2\sigma^2(\theta)(Y - \mu(\theta))^2 + \sigma^2(\theta)^2] \\ &= \sigma^2(\theta)^2(\gamma_2(\theta) + 3) - 2\sigma^2(\theta)^2 + \sigma^2(\theta)^2 \\ &= \sigma^2(\theta)^2(\gamma_2(\theta) + 2).\end{aligned}\quad (5.1.14)$$

On obtient alors la matrice de variance-covariance de la fonction  $u(\theta)$ , dont on évalue par la suite l'inverse :

$$Var[u(\theta; Y_t)] = \begin{bmatrix} \sigma^2(\theta) & \sigma^2(\theta)^{3/2}\gamma_1(\theta) \\ \sigma^2(\theta)^{3/2}\gamma_1(\theta) & \sigma^2(\theta)^2(\gamma_2(\theta) + 2) \end{bmatrix} \quad (5.1.15)$$

$$Var^{-1}[u(\theta; Y_t)] = \frac{1}{\sigma^2(\theta)^3\gamma_3(\theta)} \begin{bmatrix} \sigma^2(\theta)^2(\gamma_2(\theta) + 2) & -\sigma^2(\theta)^{3/2}\gamma_1(\theta) \\ -\sigma^2(\theta)^{3/2}\gamma_1(\theta) & \sigma^2(\theta) \end{bmatrix}. \quad (5.1.16)$$

On peut enfin évaluer l'expression (5.1.5) à l'aide des résultats précédents :

$$\mathbf{w}^*(\theta) = \begin{bmatrix} -\mu'(\theta) & -2\sqrt{\sigma^2(\theta)}\sigma'(\theta) \end{bmatrix} \times \frac{1}{\sigma^2(\theta)^3\gamma_3(\theta)} \begin{bmatrix} \sigma^2(\theta)^2(\gamma_2(\theta) + 2) & -\sigma^2(\theta)^{3/2}\gamma_1(\theta) \\ -\sigma^2(\theta)^{3/2}\gamma_1(\theta) & \sigma^2(\theta) \end{bmatrix}.$$

On obtient alors le vecteur de pondération optimal (5.0.7) formé des fonctions  $\mathbf{a}^*(\theta)$  et  $\mathbf{b}^*(\theta)$  :

$$\mathbf{a}^*(\theta) = \frac{\{-(\gamma_2(\theta) + 2)\mu'(\theta) + 2\gamma_1(\theta)\sigma'(\theta)\}}{\sigma^2(\theta)\gamma_3(\theta)} \quad (5.1.17a)$$

$$\mathbf{b}^*(\theta) = \frac{\gamma_1(\theta)\mu'(\theta) - 2\sigma'(\theta)}{\sigma^2(\theta)^{3/2}\gamma_3(\theta)}. \quad (5.1.17b)$$

On obtient aisément les matrices  $M^*(\theta)$  et  $V^*(\theta)$  à partir de leur définition (5.1.3) :

$$\begin{aligned} M^*(\theta) &= V^*(\theta) \\ &= T \cdot \sigma(\theta)^{-2} \left\{ (\mu'(\theta))(\mu'(\theta))^T + \gamma_3(\theta)^{-1} (\gamma_1(\theta)\mu'(\theta) - 2\sigma'(\theta))(\gamma_1(\theta)\mu'(\theta) - 2\sigma'(\theta))^T \right\}. \end{aligned} \quad (5.1.18)$$

Afin d'obtenir la solution optimale  $\theta^*$ , on définit la fonction objectif  $\Lambda(\theta; \mathbf{y}_T)$  :

$$\Lambda(\theta; \mathbf{y}_T) = g(\theta; \mathbf{y}_T) W(\theta; \mathbf{y}_T) g(\theta; \mathbf{y}_T)^T. \quad (5.1.19)$$

$W(\theta; \mathbf{y}_T)$  est une matrice définie positive. En premier lieu, on suggère d'utiliser la matrice identité. Puis, on peut raffiner l'estimation en lui substituant la matrice  $V^*(\hat{\theta})$  de manière itérative, de la même manière qu'avec la méthode des moments généralisée itérative, à la section 4.2.4.

## 5.2 Équation d'estimation optimale modifiée

Il est possible que, pour certaines distributions, les expressions formant le vecteur de pondération optimal  $\mathbf{w}^*(\theta)$  soient particulièrement complexes. On pourra toujours les utiliser, cependant, il peut être intéressant de développer une approximation de celles-ci. Pour ce faire, on substitue les valeurs théoriques des coefficients d'asymétrie et d'aplatissement par une valeur estimée à partir de l'échantillon :

$$\mathbf{a}_{mod}^*(\theta) = \frac{\{-(\hat{\gamma}_2(\mathbf{y}_T) + 2)\mu'(\theta) + 2\hat{\gamma}_1(\mathbf{y}_T)\sigma'(\theta)\}}{\sigma^2(\theta)\hat{\gamma}_3(\mathbf{y}_T)} \quad (5.2.1a)$$

$$\mathbf{b}_{mod}^*(\theta) = \frac{\hat{\gamma}_1(\mathbf{y}_T)\mu'(\theta) - 2\sigma'(\theta)}{\sigma^2(\theta)^{3/2}\hat{\gamma}_3(\mathbf{y}_T)} \quad (5.2.1b)$$

où

$$\hat{\gamma}_1(\mathbf{y}_T) = \frac{\sum_{t=1}^T (y_t - \bar{y})^3}{\left(\sum_{t=1}^T (y_t - \bar{y})^2\right)^{3/2}} \quad (5.2.1c)$$

$$\hat{\gamma}_2(\mathbf{y}_T) = \frac{\sum_{t=1}^T (y_t - \bar{y})^4}{\left(\sum_{t=1}^T (y_t - \bar{y})^2\right)^2} \quad (5.2.1d)$$

$$\hat{\gamma}_3(\mathbf{y}_T) = \hat{\gamma}_2(\mathbf{y}_T) + 2 - \hat{\gamma}_1(\mathbf{y}_T)^2. \quad (5.2.1e)$$



On pourra ainsi réduire considérablement la taille des expressions à évaluer tout en conservant les propriétés asymptotiques des estimateurs, puisque l'espérance de chacune des statistiques utilisées est égale à la valeur que l'on a remplacée.

Au prochain chapitre, nous appliquerons la méthode des moments généralisée et la méthode de l'équation d'estimation optimale à la distribution de Laplace asymétrique généralisée.



## Chapitre 6

# Estimation des paramètres de la distribution de Laplace asymétrique généralisée

Ce chapitre présente les différents éléments qui permettent d'obtenir des estimateurs convergents à l'aide de la méthode des moments généralisée et de l'équation d'estimation optimale. Premièrement, une méthode pour obtenir un point de départ efficace pour l'algorithme d'optimisation sera détaillée. Enfin, les résultats découlant des deux chapitres précédents seront présentés.

### 6.1 Vecteur de paramètres initiaux

Comme chacune des deux méthodes d'estimation étudiées requiert l'utilisation d'un algorithme d'optimisation numérique, on doit être en mesure de fournir un vecteur de paramètres initiaux qui favorisera la convergence de celui-ci. Ce vecteur doit faire partie de l'espace des paramètres  $\Omega$ , tel que défini à la table 2.1. Ce vecteur est plus facile à obtenir de manière empirique, par exemple, avec la méthode des moments, lorsque c'est possible. Cependant, pour la distribution de Laplace asymétrique généralisée, le système d'équations formé à partir des quatre premiers moments centraux empiriques et théoriques (2.2.15) n'a pas de solution analytique.

Seneta (2004) propose une technique pour obtenir un vecteur de paramètres initiaux basé sur la méthode des moments. En utilisant la même démarche que ce dernier, on peut obtenir un vecteur pour la forme en  $\mu$  de la fonction caractéristique (2.2.3). On pose comme hypothèse dans les équations des quatre premiers moments (2.2.15a), (2.2.15b), (2.2.16a) et (2.2.16b), que le paramètre  $\mu$  est significativement nul, donc que la distribution est presque symétrique. On impose donc que les puissances de ce paramètre plus grandes ou égales à deux soient

nulles :

$$\mu^k = 0, \quad k \in \{2, 3, 4\}. \quad (6.1.1)$$

En posant l'égalité entre les moments théoriques et leur estimateur correspondant, on obtient le système d'équations suivant :

$$\begin{aligned} \hat{m}_1 &= \theta + \tau \mu \\ \hat{m}_2 &= \sigma^2 \tau \\ \hat{\gamma}_1 &= \frac{3\mu}{|\sigma| \sqrt{\tau}} \\ \hat{\gamma}_2 &= \frac{3}{\tau}. \end{aligned}$$

Ce système possède deux solutions analytiques, dont une respecte la condition selon laquelle le paramètre  $\sigma$  doit être positif. On note que ce résultat ne constitue qu'un point de départ pour l'algorithme d'optimisation numérique et qu'on ne peut, en aucun cas, utiliser celui-ci à des fins statistiques. C'est un estimateur biaisé et non convergent étant donné la condition (6.1.1) :

$$\hat{\theta} = -\frac{\hat{\gamma}_1 \sqrt{\hat{m}_2} - \hat{\gamma}_2 \hat{m}_1}{\hat{\gamma}_2} \quad (6.1.2a)$$

$$\hat{\sigma} = \frac{\sqrt{\hat{\gamma}_2 \hat{m}_2}}{\sqrt{3}} > 0 \quad (6.1.2b)$$

$$\hat{\mu} = \frac{\hat{\gamma}_1 \sqrt{\hat{m}_2}}{3} \quad (6.1.2c)$$

$$\hat{\tau} = \frac{3}{\hat{\gamma}_2}. \quad (6.1.2d)$$

## 6.2 Méthode des moments généralisée

En premier lieu, on utilise la méthode des moments généralisée afin d'estimer les paramètres de la distribution de Laplace asymétrique généralisée. Aux fins d'illustration, seulement deux conditions de moments seront utilisées. Cependant, pour obtenir des résultats optimaux en termes de convergence asymptotique, on doit utiliser au moins autant de conditions de moment que de paramètres. Ainsi, en considérant la différence entre les moyennes et variances empirique et théorique, on obtient les conditions de moment suivantes, qui ont une espérance nulle sous les vrais paramètres  $\theta_0$  :

$$h(\theta; Y) = \begin{bmatrix} Y - m_1 \\ (Y - m_1)^2 - m_2 \end{bmatrix} \quad (6.2.1)$$

$$= \begin{bmatrix} Y - (\theta + \mu\tau) \\ (Y - (\theta + \mu\tau))^2 - \tau(\sigma^2 + \mu^2) \end{bmatrix}. \quad (6.2.2)$$

On définit la contrepartie empirique de ces conditions de moment comme suit :

$$\begin{aligned} g(\theta; \mathbf{y}_T) &= \frac{1}{T} \begin{bmatrix} g_1(\theta; \mathbf{y}_T) \\ g_2(\theta; \mathbf{y}_T) \end{bmatrix} \\ &= \frac{1}{T} \begin{bmatrix} \sum_{t=1}^T y_t - m_1 \\ \sum_{t=1}^T (y_t - m_1)^2 - m_2 \end{bmatrix} \end{aligned} \quad (6.2.3)$$

$$= \frac{1}{T} \begin{bmatrix} \sum_{t=1}^T y_t - (\theta + \mu\tau) \\ \sum_{t=1}^T (y_t - (\theta + \mu\tau))^2 - \tau(\sigma^2 + \mu^2) \end{bmatrix}. \quad (6.2.4)$$

On doit définir la matrice de pondération optimale.

### 6.2.1 Matrice de pondération optimale

On s'intéresse à la forme analytique de la matrice de pondération optimale (4.2.14). Celle-ci est obtenue en prenant l'espérance du produit extérieur du vecteur des conditions de moment théoriques sur elles-mêmes. Pour des fins de simplification, on définit le produit des conditions  $i$  et  $j$  par la notation  $H_{i,j}(\theta; Y) = h_i(\theta; Y)h_j(\theta; Y)$ .

$$\begin{aligned} S(\theta; Y) &= E \begin{bmatrix} H_{(1,1)}(\theta; Y) & H_{(1,2)}(\theta; Y) \\ H_{(2,1)}(\theta; Y) & H_{(2,2)}(\theta; Y) \end{bmatrix} \\ &= \begin{bmatrix} E[H_{(1,1)}(\theta; Y)] & E[H_{(1,2)}(\theta; Y)] \\ E[H_{(2,1)}(\theta; Y)] & E[H_{(2,2)}(\theta; Y)] \end{bmatrix} \end{aligned} \quad (6.2.5)$$

où

$$\begin{aligned} H_{(1,1)}(\theta; Y) &= (Y - \theta - \mu\tau)^2 \\ H_{(1,2)}(\theta; Y) &= (Y^2 - 2\theta Y - 2\mu\tau Y + \theta^2 + 2\mu\tau\theta + \mu^2\tau^2 - \sigma^2\tau - \mu^2\tau)^2 \\ H_{(2,2)}(\theta; Y) &= (Y - \theta - \mu\tau) (Y^2 - 2\theta Y - 2\mu\tau Y + \theta^2 + 2\mu\tau\theta + \mu^2\tau^2 - \sigma^2\tau - \mu^2\tau). \end{aligned}$$

On évalue ensuite l'espérance de chacun des éléments de la matrice, où l'on définit de manière analogue celle du produit des conditions de moments,  $W_{(i,j)}(\theta; Y) = E[H_{(i,j)}(\theta; Y)]$ .

$$W(\theta; Y) = \begin{bmatrix} W_{(1,1)}(\theta; Y) & W_{(1,2)}(\theta; Y) \\ W_{(2,1)}(\theta; Y) & W_{(2,2)}(\theta; Y) \end{bmatrix} \quad (6.2.6)$$

où

$$\begin{aligned}
W_{(1,1)} &= \mu^2 \tau^2 - 2\mu^2 \nu \tau + \nu \sigma^2 + \mu^2 \nu^2 + \mu^2 \nu \\
W_{(2,2)} &= \mu^4 \tau^4 + (-2\mu^2 \sigma^2 - 4\mu^4 \nu - 2\mu^4) \tau^3 + (\sigma^4 + (10\mu^2 \nu + 2\mu^2) \sigma^2 + 6\mu^4 \nu^2 + 10\mu^4 \nu + \mu^4) \tau^2 \\
&\quad + (-2\nu \sigma^4 + (-14\mu^2 \nu^2 - 16\mu^2 \nu) \sigma^2 - 4\mu^4 \nu^3 - 14\mu^4 \nu^2 - 10\mu^4 \nu) \tau + (3\nu^2 + 3\nu) \sigma^4 \\
&\quad + (6\mu^2 \nu^3 + 18\mu^2 \nu^2 + 12\mu^2 \nu) \sigma^2 + \mu^4 \nu^4 + 6\mu^4 \nu^3 + 11\mu^4 \nu^2 + 6\mu^4 \nu \\
W_{(1,2)} &= W_{(2,1)} = 3\theta^4 + (6\mu\nu - 3) \theta^3 + (3\nu\sigma^2 + 3\mu^2\nu^2 + (3\mu^2 - 3\mu)\nu) \theta^2 - \mu^3\tau^3 \\
&\quad + (\mu\sigma^2 + 3\mu^3\nu + \mu^3) \tau^2 + (-4\mu\nu\sigma^2 - 3\mu^3\nu^2 - 4\mu^3\nu) \tau + (3\mu\nu^2 + 3\mu\nu) \sigma^2 \\
&\quad + \mu^3\nu^3 + 3\mu^3\nu^2 + 2\mu^3\nu.
\end{aligned}$$

On évalue quelle serait la valeur de la variance-covariance sous des paramètres optimaux. Ce résultat permet par la suite d'évaluer la distribution asymptotique des estimateurs.

On obtient l'estimateur de la matrice optimale (4.2.21) en effectuant le produit extérieur du vecteur de conditions de moment empirique (6.2.2), puis en évaluant la moyenne des matrices résultantes de l'équation (4.2.19) :

$$\hat{S}(\theta; \mathbf{y}_T) = \frac{1}{T} \sum_{t=1}^T \begin{bmatrix} G_{(1,1)}(\theta; y_t) & G_{(1,2)}(\theta; y_t) \\ G_{(2,1)}(\theta; y_t) & G_{(2,2)}(\theta; y_t) \end{bmatrix} \quad (6.2.7)$$

où

$$\begin{aligned}
G_{(1,1)}(\theta; y_t) &= (y_t - \theta - \mu\tau)^2 \\
G_{(1,2)}(\theta; y_t) &= (y_t^2 - 2\theta y_t - 2\mu\tau y_t + \theta^2 + 2\mu\tau\theta + \mu^2\tau^2 - \sigma^2\tau - \mu^2\tau)^2 \\
G_{(2,2)}(\theta; y_t) &= (y_t - \theta - \mu\tau) (y_t^2 - 2\theta y_t - 2\mu\tau y_t + \theta^2 + 2\mu\tau\theta + \mu^2\tau^2 - \sigma^2\tau - \mu^2\tau).
\end{aligned}$$

### 6.2.2 Variance-covariance des paramètres

On obtient la variance-covariance asymptotique des paramètres à partir de la variance-covariance associée aux conditions de moment, en utilisant la méthode delta multivariée. Pour ce faire, on évalue d'abord la valeur théorique du gradient  $D(\theta)$ .

$$\begin{aligned}
D(\theta) &= E \begin{bmatrix} -1 & -2(Y - \theta - \mu\tau) \\ 0 & -2\sigma\tau \\ -\tau & -2\tau(Y - \theta - \mu\tau) - 2\mu\tau \\ -\nu & -2\mu(Y - \theta - \mu\tau) - \sigma^2 - \mu^2 \end{bmatrix} \\
&= \begin{bmatrix} -1 & 0 \\ 0 & -2\sigma\tau \\ -\tau & -2\mu\tau \\ -\nu & -\sigma^2 - \mu^2 \end{bmatrix}. \quad (6.2.8)
\end{aligned}$$

Ensuite, on utilise le résultat (4.2.35) afin d'obtenir la valeur de la variance-covariance  $\mathcal{J}_0^{-1}$ .

De la même manière, on évalue la variance-covariance des paramètres optimaux en utilisant les valeurs empiriques de la variance-covariance des conditions de moment (6.2.7) et du gradient  $\hat{D}(\theta, \mathbf{y}_T)$ , défini par la matrice (4.2.25). On a essentiellement à calculer la moyenne empirique des gradients obtenus en chaque point  $y_t$ , puis à utiliser le résultat (4.2.35) avec les estimateurs  $\hat{W}_T = \hat{S}^{-1}(\theta, \mathbf{y}_T)$  et  $\hat{D}(\theta, \mathbf{y}_T)$  :

$$\hat{D}(\theta, \mathbf{y}_T) = \frac{1}{T} \sum_{t=1}^T \begin{bmatrix} -1 & -2(y_t - \theta - \mu\tau) \\ 0 & -2\sigma\tau \\ -\tau & -2\tau(y_t - \theta - \mu\tau) - 2\mu\tau \\ -\nu & -2\mu(y_t - \theta - \mu\tau) - \sigma^2 - \mu^2 \end{bmatrix}. \quad (6.2.9)$$

### 6.2.3 Contraintes linéaires

Une fois l'estimation des paramètres effectuée, on peut tester si ces derniers pourraient en fait correspondre à un cas particulier de la distribution de Laplace asymétrique généralisée ayant le même support, parmi celles figurant à la table 2.2. Pour ce faire, on doit déterminer les paramètres de la distribution sous un ensemble de contraintes linéaires, sous la forme (4.3.2), correspondant au cas particulier, tel qu'énuméré dans la table 6.1.

Distribution	Paramètres	
	$R$	$r$
Laplace symétrique	$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$
Laplace asymétrique	$\begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 \end{bmatrix}$
Dégénérée à $\theta$	$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

TABLE 6.1: Contraintes linéaires pour les cas particuliers de la distribution de Laplace asymétrique généralisée  $GAL(\theta, \sigma, \mu, \tau)$

On doit ensuite utiliser un algorithme d'optimisation numérique afin de maximiser le lagrangien (4.3.4) défini par la fonction objectif  $Q_T(\theta)$  (4.2.17) et l'ensemble de contraintes  $a(\theta) = R\theta - r$  sélectionné précédemment. Enfin, on peut effectuer les tests statistiques de la section 4.4.1 afin de vérifier si les paramètres de la distribution contrainte sont significativement différents de ceux de la distribution non contrainte. Dans cette situation, l'hypothèse selon laquelle le cas particulier serait approprié pour décrire la distribution des données de l'échantillon serait rejetée.

Notons qu'on peut aussi estimer les paramètres de chacun des cas particuliers directement par la méthode des moments généralisée en utilisant des conditions de moment basées sur la moyenne  $m_1$  et la variance  $m_2$  de la forme (6.2.1). Par contre, dans ces cas, on devra utiliser des tests d'adéquation non paramétriques afin de décider laquelle représente mieux la distribution de l'échantillon de données. Ces tests seront présentés au chapitre suivant, à la section 7.2.

### 6.3 Méthode de l'équation d'estimation optimale

On utilise maintenant la méthode de l'équation d'estimation optimale afin d'estimer les paramètres de la distribution de Laplace asymétrique généralisée. On considère une équation d'estimation de la forme quadratique (5.0.6), basée sur les conditions de moments (6.2.2) utilisées à la section précédente.

On évalue d'abord les expressions des dérivées premières de la moyenne et de l'écart-type par rapport au vecteur de paramètres (5.0.5) :

$$\begin{aligned}\mu'(\theta) &= \frac{d}{d\theta} (\theta + \tau \mu) \\ &= \begin{bmatrix} 1 \\ 0 \\ \tau \\ \mu \end{bmatrix}^T\end{aligned}\tag{6.3.1a}$$

$$\begin{aligned}\sigma'(\theta) &= \frac{d}{d\theta} \sqrt{\tau \sigma^2 + \tau \mu^2} \\ &= \frac{1}{\sqrt{\tau \sigma^2 + \tau \mu^2}} \begin{bmatrix} 0 \\ \tau \sigma \\ \tau \mu \\ \frac{\sigma^2 + \mu^2}{2} \end{bmatrix}^T.\end{aligned}\tag{6.3.1b}$$

On peut alors obtenir une forme analytique en utilisant les expressions déterminées précédemment pour  $\mathbf{a}(\theta)$  et  $\mathbf{b}(\theta)$  (5.1.17), les moments de la distribution (2.2.15) ainsi que les coefficients d'asymétrie et d'aplatissement (2.2.16). Étant donné la longueur des expressions obtenues, elles ne seront pas présentées dans ce texte. Cependant, on peut facilement les évaluer à l'aide d'un logiciel de calcul symbolique. On combine ensuite celles-ci pour obtenir l'équation d'estimation optimale  $g(\theta; \mathbf{y}_T)$  de forme quadratique.

On obtient cependant des expressions plus faciles à manipuler pour  $\mathbf{a}(\theta)$  et  $\mathbf{b}(\theta)$  en utilisant



la version modifiée du vecteur de pondération optimal (5.2.1) :

$$\mathbf{a}_{mod}^*(\theta; \mathbf{y}_T) = \begin{bmatrix} \frac{-\hat{\gamma}_2(\mathbf{y}_T) - 2}{(\sigma^2 + \mu^2) \tau \hat{\gamma}_3(\mathbf{y}_T)} \\ \frac{2 \sigma \hat{\gamma}_1(\mathbf{y}_T)}{(\sigma^2 + \mu^2)^{3/2} \sqrt{\tau} \hat{\gamma}_3(\mathbf{y}_T)} \\ \frac{2 \mu \sqrt{\tau} \hat{\gamma}_1(\mathbf{y}_T) + \tau (-\hat{\gamma}_2(\mathbf{y}_T) - 2)}{\sqrt{\sigma^2 + \mu^2}} \\ \frac{(\sigma^2 + \mu^2) \tau \hat{\gamma}_3(\mathbf{y}_T)}{\sqrt{\sigma^2 + \mu^2} \hat{\gamma}_1(\mathbf{y}_T) + \mu (-\hat{\gamma}_2(\mathbf{y}_T) - 2)} \\ \frac{(\sigma^2 + \mu^2) \tau \hat{\gamma}_3(\mathbf{y}_T)}{(\sigma^2 + \mu^2) \tau \hat{\gamma}_3(\mathbf{y}_T)} \end{bmatrix} \quad (6.3.2a)$$

$$\mathbf{b}_{mod}^*(\theta; \mathbf{y}_T) = \begin{bmatrix} \frac{\hat{\gamma}_1(\mathbf{y}_T)}{(\sigma^2 + \mu^2)^{3/2} \tau^{3/2} \hat{\gamma}_3(\mathbf{y}_T)} \\ \frac{2 \sigma}{(\sigma^2 + \mu^2)^2 \tau \hat{\gamma}_3(\mathbf{y}_T)} \\ \frac{\tau \hat{\gamma}_1(\mathbf{y}_T) - \frac{2 \mu \sqrt{\tau}}{\sqrt{\sigma^2 + \mu^2}}}{(\sigma^2 + \mu^2)^{3/2} \tau^{3/2} \hat{\gamma}_3(\mathbf{y}_T)} \\ \frac{\mu \hat{\gamma}_1(\mathbf{y}_T) - \frac{\sqrt{\sigma^2 + \mu^2}}{\sqrt{\tau}}}{(\sigma^2 + \mu^2)^{3/2} \tau^{3/2} \hat{\gamma}_3(\mathbf{y}_T)} \end{bmatrix} \cdot \quad (6.3.2b)$$

On insère ensuite ces dernières dans la forme générale (5.0.6) pour obtenir l'équation d'estimation optimale modifiée  $g_{mod}^*(\theta; Y)$  :

$$g_{mod}^*(\theta; Y) = \sum_{t=1}^n \left[ \mathbf{a}_{mod}^*(\theta; \mathbf{y}_T) (y_t - \mu(\theta)) + \mathbf{b}_{mod}^*(\theta; \mathbf{y}_T) \left( (y_t - \mu(\theta))^2 - \sigma^2(\theta) \right) \right]. \quad (6.3.3)$$

On note qu'afin d'éviter des irrégularités numériques, on suggère, dans les deux cas, de faire l'estimation sur des données centrées et réduites :

$$X_t = \frac{Y_t - \hat{m}_1}{\sqrt{\hat{m}_2}}. \quad (6.3.4)$$

Comme précédemment, on pourra utiliser la propriété d'invariance (2.2.17) de la paramétrisation en  $\kappa$  afin de retrouver les paramètres de la distribution de la variable aléatoire  $Y_t$  une fois l'estimation effectuée :

$$Y_t = \hat{\sigma}_t X_t + \hat{\mu}_t \sim GAL(\hat{\sigma}_t \theta + \hat{\mu}_t, \hat{\sigma}_t \sigma, \kappa, \tau). \quad (6.3.5)$$

Afin d'obtenir le vecteur de paramètres estimés  $\hat{\theta}$ , on minimise la fonction objectif  $\Lambda(\theta; \mathbf{y}_T)$  (5.1.19). Une fois les paramètres obtenus, on peut ensuite évaluer la variance-covariance de ces estimateurs, à partir du résultat (5.1.18).



# Chapitre 7

## Tests statistiques

### 7.1 Test de normalité

On utilise un test de normalité afin de vérifier si l'on peut rejeter l'hypothèse de normalité de la distribution des rendements. Comme présenté au premier chapitre, c'est cette observation qui est à la base de la recherche de modèles alternatifs, étant donné que cette hypothèse est rejetée pour la plupart des séries financières. Il existe plusieurs tests disponibles pour ce faire, dont ceux de Shapiro-Wilk et d'Epps-Pulley.

#### 7.1.1 Test de Shapiro-Wilk

Le test de Shapiro-Wilk (Shapiro et Wilk, 1965) est basé sur les statistiques d'ordre de la distribution normale. Ce test est particulièrement efficace même pour de petits échantillons ( $T < 20$ ), selon les auteurs.

On considère l'échantillon ordonné  $y_{(1)}, y_{(2)}, \dots, y_{(T)}$ . On évalue la variance  $S^2$  de l'échantillon puis la statistique  $b$ , où les valeurs  $a_{T-i+1}$  proviennent de la *Table 5* de Shapiro et Wilk (1965) :

1. Si la taille de l'échantillon  $T$  est paire, alors  $k = \frac{T}{2}$  et

$$b = \sum_{i=1}^k a_{T-i+1} (y_{(T-i+1)} - y_{(i)}). \quad (7.1.1)$$

2. Si la taille de l'échantillon  $T$  est impaire, alors  $k = \frac{T-1}{2}$  et

$$b = a_T (y_{(T)} - y_{(1)}) + \dots + a_{k+2} (y_{(k+2)} - y_{(k)}). \quad (7.1.2)$$

La statistique de Shapiro-Wilk  $W = \frac{b}{S^2}$  suit une distribution particulière dont on retrouve différents quantiles à la *Table 6* de Shapiro et Wilk (1965). On rejette l'hypothèse de la normalité lorsque la statistique  $W$  est inférieure au seuil critique déterminé à partir de cette dernière table.

### 7.1.2 Test d'Epps-Pulley

Le test d'Epps-Pulley (Epps et Pulley, 1983) est basé sur le carré de la différence entre les fonctions caractéristiques empirique et théorique de la distribution normale. Ce test est considéré comme un des plus puissants pour un seuil de tolérance donné, pour de larges échantillons. La statistique de test prend la forme suivante, où  $\bar{X}$  et  $S$  sont respectivement la moyenne et l'écart-type échantillonnal :

$$EP_T = T \int_{-\infty}^{\infty} \left| \frac{1}{T} \sum_{j=1}^T \exp \left[ it \frac{(X_j - \bar{X})}{S} \right] - \exp \left[ -\frac{t^2}{2} \right] \right|^2 w(t) dt. \quad (7.1.3)$$

Henze (1990) propose une procédure simple à implémenter pour effectuer un test basé sur la statistique d'Epps-Pulley  $EP_T$ , où la pondération de la différence quadratique  $w(\cdot)$  est remplacée par la densité de la loi normale centrée réduite, ce qui permet d'accorder davantage d'importance aux observations près de l'origine et aussi d'obtenir une forme intégrable :

$$EP_T = \frac{2}{T} \sum_{1 \leq j < k \leq T} \exp \left[ -\frac{(X_j - X_k)^2}{2S^2} \right] - \sqrt{2} \sum_{j=1}^T \exp \left[ -\frac{(X_j - \bar{X})^2}{4S^2} \right] + \frac{T}{\sqrt{3}} + 1. \quad (7.1.4)$$

Lorsque  $T > 10$ , on calcule une version modifiée de la statistique :

$$EP_T^* = \left( EP_T - \frac{0.365}{T} + \frac{1.34}{T^2} \right) \left( 1 + \frac{1.3}{T} \right). \quad (7.1.5)$$

On obtient ensuite une approximation d'un quantile de la loi normale centrée réduite :

$$Z_T = \gamma + \delta \ln \left( (EP_T^* - \xi) (\xi + \lambda - EP_T^*) \right). \quad (7.1.6)$$

Les constantes sont dérivées à l'équation 4.1 de Henze (1990) :

$$\begin{aligned} \gamma &\approx 3.55295 \\ \delta &\approx 1.23062 \\ \lambda &\approx 2.26664 \\ \xi &\approx -0.020682. \end{aligned}$$

Puis, on compare cette statistique au quantile  $Z_{1-\alpha}$  correspondant au seuil de tolérance  $\alpha$ . Si  $Z_n > Z_{1-\alpha}$  alors on rejette la normalité des données.

## 7.2 Tests d'adéquation

Les tests d'adéquation vérifient l'ajustement à l'échantillon de la fonction de répartition estimée soit :

- à partir de l'échantillon ayant servi à l'estimation du vecteur de paramètres
- à partir d'un échantillon séparé, afin de vérifier, par exemple, si un modèle estimé à partir d'anciennes données s'applique toujours à de nouvelles informations.

On cherche à rejeter ou non une hypothèse concernant l'échantillon  $Y$  et le vecteur de paramètres estimés :

- Si on considère le vecteur de paramètres comme étant estimé à partir du même échantillon sur lequel on effectue le test, alors on pose l'hypothèse composée suivante :

$$H_0 : Y \sim F_Y(\hat{\theta}). \quad (7.2.1)$$

- Si on considère le vecteur de paramètres comme étant connu ou estimé à partir d'un autre échantillon, alors on pose l'hypothèse simple suivante :

$$H_0 : Y \sim F_Y(\theta_0). \quad (7.2.2)$$

La différence entre ces deux hypothèses est le nombre de degrés de liberté de la statistique de test. Les deux premiers tests présentés sont des classiques, largement utilisés, mais dont l'application est basée sur la fonction de répartition. Étant donné qu'on ne dispose que d'une approximation de celle-ci, on utilisera alors un test basé sur la fonction génératrice des moments, qui a une forme analytique.

### 7.2.1 Test $\chi^2$ de Pearson

Le test  $\chi^2$  de Pearson (Hogg et Craig, 1978, ch. 8) est basé sur une approximation multinomiale de la fonction de répartition  $F_X(x; \theta)$  pour laquelle on veut vérifier l'ajustement des données échantillonnelles  $X_1, \dots, X_n$ . Pour ce faire, on divise le domaine de la variable aléatoire  $X$  en  $k$  intervalles, appelés classes. On associe la probabilité  $p_i(\theta)$  que  $X$  prend une valeur dans l'intervalle  $[c_{i-1}, c_i]$ . Cette probabilité est évaluée à l'aide de la fonction de répartition  $F_X(x; \theta)$  de la variable aléatoire  $X$  :

$$p_i(\theta) = F_X(c_i; \theta) - F_X(c_{i-1}; \theta), \quad i = 1, \dots, k. \quad (7.2.3)$$

Soit  $N_i$ , le nombre de données de l'échantillon observées dans l'intervalle défini précédemment. On définit alors la statistique  $Q$  comme étant la somme, pour chaque intervalle, du carré de la différence entre le nombre d'observations obtenues et espérées, pondérée par cette dernière valeur. Cette quantité a approximativement une distribution asymptotique  $\chi^2(k-1)$  avec  $k-1$  degrés de liberté :

$$Q_{k-1} = \sum_{i=1}^k \frac{(N_i - np_i(\theta_0))^2}{np_i(\theta_0)} \sim \chi^2(k-1). \quad (7.2.4)$$

La statistique  $Q_{k-1}$  peut être utilisée pour vérifier une hypothèse simple, c'est-à-dire lorsque les paramètres sont connus :

$$\mathcal{H}_0 : \theta = \theta_0. \quad (7.2.5)$$

Cependant, comme les  $q$  paramètres de la distribution  $f_X$  doivent être estimés pour évaluer les probabilités  $p_i(\theta)$ , on retrouve alors une hypothèse composée :

$$\mathcal{H}_0 : F \in \{F_\theta\}, \theta \in \Omega. \quad (7.2.6)$$

On doit retrancher le même nombre de degrés de liberté à la distribution asymptotique approximative de la statistique  $Q$  qui devient :

$$Q_{k-q-1} = \sum_{i=1}^k \frac{(N_i - np_i(\hat{\theta}))^2}{np_i(\hat{\theta})} \sim \chi^2(k - q - 1). \quad (7.2.7)$$

Il est important de noter que ce test est approximatif, car pour obtenir asymptotiquement la distribution  $\chi^2$ , les paramètres doivent être estimés en minimisant la statistique  $Q_{k-q-1}$  (7.2.7), qui devient alors une fonction objectif.

On rejette l'hypothèse composée  $\mathcal{H}_0$  lorsque la condition  $Q_{k-q-1} > \chi_{1-\alpha}^2(k - q - 1)$  est respectée. On effectue le même test pour l'hypothèse simple, en posant  $q = 0$ . Le point critique de la distribution  $\chi^2(k - q - 1)$  est sélectionné selon le critère

$$\Pr(Q_{k-q-1} \leq \chi_{1-\alpha}^2(k - q - 1)) = 1 - \alpha. \quad (7.2.8)$$

## 7.2.2 Test de Kolmogorov-Smirnov

Le test de Kolmogorov-Smirnov vérifie l'hypothèse simple (7.2.2). On définit la statistique  $D_n$  comme étant la valeur maximale de la distance entre les fonctions de répartition empiriques  $F_n(x)$  et celle  $F_X(x)$  de la distribution exacte spécifiée par l'hypothèse nulle  $H_0$  :

$$D_n = \sup_x |F_n(x) - F_X(x)| \quad (7.2.9)$$

où

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{X_i \leq x}. \quad (7.2.10)$$

La statistique  $\sqrt{n}D_n$  suit asymptotiquement une distribution de Kolmogorov, selon Jingbo Wang et Marsaglia (2003), dont la fonction de répartition est définie comme suit :

$$\Pr(\sqrt{n}D_n \leq x) = 1 - 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2 x^2} \quad (7.2.11)$$

$$= \frac{\sqrt{2\pi}}{x} \sum_{k=1}^{\infty} e^{-(2k-1)^2 \pi^2 / (8x^2)}. \quad (7.2.12)$$

On rejette l'hypothèse nulle  $H_0$  lorsque la condition  $\sqrt{n}D_n > K_{1-\alpha}$  est respectée. On sélectionne le quantile de cette distribution selon le critère suivant :

$$\Pr(\sqrt{n}D_n \leq K_{1-\alpha}) = 1 - \alpha. \quad (7.2.13)$$

### 7.2.3 Test de distance minimale basé sur la fonction génératrice des moments

Luong et Thompson (1987) développent un ensemble de tests d'ajustement pour l'hypothèse simple (7.2.2) basée sur une transformation de la fonction de densité ou de répartition. Parmi cette classe de tests, on retrouve celui d'Epps-Pulley, développé pour vérifier l'hypothèse de normalité des données, tel que présenté à la section 7.1.2. On retrouve aussi le test K-L de Feuerverger et McDunnough (1981) basé sur les parties réelles et imaginaires de la fonction caractéristique. Cependant, pour la distribution de Laplace asymétrique généralisée, on ne peut effectuer cette séparation. On préférera donc utiliser la fonction génératrice des moments  $M_Y(\xi)$  afin de construire la statistique de test  $D(F_T; f_{\theta_0})$ . Comme on a déjà estimé les paramètres, on veut donc vérifier une hypothèse simple. On se réfère à Koutrouvelis (1980), qui a développé le test K-L pour cette situation. Cependant, on remplacera la fonction caractéristique par la fonction génératrice des moments.

On considère un ensemble de fonctions  $h_j(y), j = 1, \dots, K$ . La transformée de la fonction de répartition  $F_Y(y)$  est donnée par le vecteur  $\mathbf{z}(F) = [z_1(F), \dots, z_K(F)]$  :

$$\begin{aligned} z_j(F) &= \int_{-\infty}^{\infty} h_j(y) dF_Y(y) \\ &= E[h_j(y)] \\ &= M_Y(t_j), \quad j = 1, \dots, K. \end{aligned} \quad (7.2.14)$$

En posant  $h_j(y) = e^{t_j y}$ , on définit  $z_j(F)$  comme étant la fonction génératrice des moments. On définit aussi la transformée de la fonction de répartition empirique  $F_T(y)$  par le vecteur  $\mathbf{z}(F_T) = [z_1(F_T), \dots, z_K(F_T)]$ , où

$$z_j(F_T) = \frac{1}{T} \sum_{t=1}^T e^{t_j y_t}. \quad (7.2.15)$$

La statistique de distance quadratique prend alors la forme suivante :

$$\begin{aligned} Td(F_n, F_{\theta}) &= T \{ \mathbf{z}(F_n) - \mathbf{z}(F_{\theta}) \}' \mathbf{Q}(F_{\theta}) \{ \mathbf{z}(F_n) - \mathbf{z}(F_{\theta}) \} \\ &= \mathbf{v}_n' \mathbf{Q}(F_{\theta}) \mathbf{v}_n. \end{aligned} \quad (7.2.16)$$

On doit maintenant sélectionner une matrice  $\mathbf{Q}(F_{\theta})$  de sorte que la distribution de la statistique  $Td(F_n, F_{\theta})$  soit  $\chi^2$ . Comme  $\mathbf{v}_n = \sqrt{T} \{ \mathbf{z}(F_n) - \mathbf{z}(F_{\theta}) \}$  suit asymptotiquement une distribution

normale multivariée de moyenne 0 et de variance-covariance  $\Sigma$ , on a, si la matrice  $\Sigma$  est inversible et que l'on pose  $\mathbf{Q} = \Sigma^{-1}$ , que  $\mathbf{v}'_n \mathbf{Q}(F_\theta) \mathbf{v}_n$  suit une distribution  $\chi^2$  avec  $K$  degrés de liberté, dans le cadre d'une hypothèse simple. Afin d'évaluer la variance-covariance  $\Sigma$ , on définit l'espérance du produit des deux fonctions  $h_j(y)$  et  $h_k(y)$  :  $z_{j,k}(F) = E[h_j(y)h_k(y)]$ .

On décrit la matrice de variance-covariance  $\Sigma$  comme suit :

$$\Sigma = \begin{bmatrix} \sigma_{1,1} & \cdots & \sigma_{1,k} \\ \vdots & \ddots & \vdots \\ \sigma_{j,1} & \cdots & \sigma_{j,k} \end{bmatrix} \quad (7.2.17)$$

où

$$\begin{aligned} \sigma_{j,k} &= z_{j,k}(F) - z_j(F)z_k(F) \\ &= M_Y(t_j + t_k) - M_Y(t_j)M_Y(t_k), \quad j, k = 1, \dots, K. \end{aligned} \quad (7.2.18)$$

On peut maintenant évaluer la statistique  $Td(F_n, F_\theta)$  et effectuer le test. On rejette l'hypothèse  $H_0$  lorsque la valeur de celle-ci est supérieure au quantile correspondant au seuil  $\alpha$  de la distribution  $\chi^2$  avec  $K$  degrés de liberté :  $Td(F_n, F_\theta) > \chi^2_{1-\alpha}(K)$ .



## Chapitre 8

# Évaluation d'options

Un des principaux intérêts de connaître la distribution des rendements d'un actif est de pouvoir évaluer la valeur de différents produits dérivés.

### 8.1 Définitions

En se référant à Bingham et Kiesel (2004), on définit :

- Un **produit dérivé** est un contrat financier dont la valeur à la date d'échéance  $T$  est déterminée par le prix de l'actif sous-jacent au temps  $T$  ou par celles prises au cours de l'intervalle  $[0, T]$ .
- Une **option** est un instrument financier qui donne le droit (et non l'obligation) d'effectuer une transaction avant ou à une date et pour un prix spécifiés.
- Une **option d'achat (de vente) européenne** donne le **droit** d'acheter (de vendre) un actif au **prix d'exercice**  $K$  au temps  $T$ .
- Lorsque la valeur actuelle du titre est, par rapport au prix d'exercice :
  - supérieure ( $S(t) > K$ ), l'option est dite **dans le cours** ;
  - égale ( $S(t) = K$ ), l'option est dite **au cours** ;
  - inférieure ( $S(t) < K$ ), l'option est dite **hors du cours**.

Le prix à l'échéance d'une option d'achat européenne  $C(t)$  est défini comme suit :

$$C(T) = \begin{cases} S(T) - K & , S(T) > K \\ 0 & S(T) \leq K. \end{cases} \quad (8.1.1)$$

La valeur du prix à l'échéance est appelée une **créance éventuelle**. L'évaluation d'un produit dérivé équivaut à calculer la valeur espérée actualisée de la réclamation contingente définie par

le contrat. Pour l'option d'achat européenne, on obtient la formule suivante :

$$\begin{aligned} C(t) &= B(t, T)E[C(T)] \\ &= B(t, T) \int_K^\infty (S(T) - K)d\hat{F}_{S(T)}. \end{aligned} \quad (8.1.2)$$

$B(t, T)$  est la valeur au temps  $t$  d'une obligation zéro-coupon au taux sans risque  $r$  d'échéance  $T$ .  $\hat{F}_t(S(T))$  est la fonction de répartition de la distribution neutre au risque de  $S(T)$ .

Les déterminants de la valeur d'une option sont :

- le prix courant de l'actif  $S(t)$
- le prix d'exercice  $K$
- la volatilité de l'actif  $\sigma$
- le temps d'ici l'échéance  $T - t$
- le taux d'intérêt sans risque  $r$ .

On peut aussi récrire (8.1.2) avec le logarithme du prix  $k = \ln K$ . On utilise donc  $s(t) = \ln S(T)$ , dont la fonction de répartition est définie par  $F_{s(T)}(y) = Pr[\ln S(T) < y]$  :

$$C(t) = B(t, T) \int_k^\infty (e^{s(T)} - e^k)dF_{s(T)}. \quad (8.1.3)$$

Enfin, une relation fondamentale, appelée parité vente-achat, permet, dans un scénario sans arbitrage, de calculer le prix d'une option de vente  $P(t)$  (d'achat  $C(t)$ ) en connaissant la valeur :

- de l'autre lorsqu'elles sont de mêmes échéance et prix d'exercice
- de l'obligation  $B(t, T)$
- initiale du titre sous-jacent  $S(t)$  commun aux deux options.

$$P(t) + S(t) = C(t) + B(t, T)K. \quad (8.1.4)$$

Cette relation permettra d'évaluer les deux types d'option à l'aide d'un seul calcul.

### 8.1.1 Équation martingale

L'évaluation d'options se fait traditionnellement dans un univers ou un espace de probabilités neutre au risque, noté  $\mathbb{Q}$ , c'est-à-dire un point de vue selon lequel les investisseurs n'exigent pas une prime de risque. Cette approche a été introduite par Black et Scholes (1973). Dans cette perspective, les rendements espérés futurs sont escomptés au taux sans risque. La distribution des rendements cumulés (1.2.3) répond à la propriété martingale, selon laquelle la

valeur actuelle d'un titre financier reflète l'ensemble de l'information connue sur ce dernier. Cette propriété permet de fixer un seul prix pour les options. **L'équation martingale** établit l'égalité entre la valeur espérée du titre au temps  $t$  et celle d'un investissement de même valeur dans un compte en banque créditant le taux sans risque :

$$E[\exp(L_t)] = e^{rt}. \quad (8.1.5)$$

### 8.1.2 Paramètres neutres au risque

Afin de pouvoir utiliser les résultats de l'estimation paramétrique des chapitres précédents pour évaluer le prix de produits dérivés, on doit tout d'abord identifier les paramètres neutres au risque de la distribution de Laplace asymétrique généralisée associant le rendements cumulé  $L_t$  au taux d'intérêt sans risque  $r$ . Pour ce faire, on utilise l'équation martingale (8.1.5) ainsi que la fonction génératrice des moments (2.2.10). On obtient alors une expression pour le paramètre de dérive neutre au risque  $\theta_{RN}$  :

$$\begin{aligned} e^{rt} &= E[\exp(L_t)] = E[\exp(R_1 + \dots + R_t)] \\ &= M_{R_1 + \dots + R_t}(1) = (M_R(1))^t \\ &= \left( \frac{e^{\theta_{RN}}}{(1 - \mu - \sigma^2/2)^\tau} \right)^t \\ r &= \ln \left( \frac{e^{\theta_{RN}}}{(1 - \mu - \sigma^2/2)^\tau} \right) \\ &= \theta_{RN} - \tau \ln(1 - \mu - \frac{\sigma^2}{2}) \\ \theta_{RN} &= r + \tau \ln(1 - \mu - \frac{\sigma^2}{2}). \end{aligned} \quad (8.1.6)$$

On obtient ainsi le paramètre de correction de la dérive  $\omega$  à partir de l'équation martingale (8.1.6) :

$$\omega = \tau \ln(1 - \mu - \frac{\sigma^2}{2}). \quad (8.1.7)$$

Cette égalité impose une contrainte aux paramètres  $\mu$  et  $\sigma$  qui doit respecter la condition

$$\mu + \frac{\sigma^2}{2} < 1. \quad (8.1.8)$$

On obtient ensuite la fonction caractéristique de la mesure neutre au risque  $\mathbb{Q}$  pour  $L_T$  en remplaçant  $\theta$  par  $\theta_{RN}$  :

$$\phi(\xi, L_T) = \frac{\exp(i\xi * (\ln(S_t) + (T-t)(r + \omega)))}{(1 - i\mu\xi + (\sigma^2\xi^2)/2)^{\tau(T-t)}}. \quad (8.1.9)$$

La distribution neutre au risque des rendements est donc  $R_t \sim GAL(r + \tau \ln(1 - \mu - \sigma^2/2), \sigma, \mu, \tau)$ .

## 8.2 Aperçu du modèle de Black-Scholes

Étant donné l'importance du modèle de Black-Scholes en finance, on se doit de le présenter comme outil de référence lorsque l'on veut le remplacer. Ce modèle est basé sur les quatre hypothèses suivantes :

1. Les rendements ont une distribution normale de moyenne  $\mu$  et variance  $\sigma^2$ , ou, de manière équivalente, le processus de prix  $\{S(t)\}$  est un mouvement brownien géométrique de dérive  $\mu$  et de volatilité  $\sigma$ .
2. Une obligation zéro-coupon au taux d'intérêt sans risque  $r$  existe sur le marché.
3. Le marché est complet et sans friction. Une position prise sur ce marché peut être couverte de manière continue.
4. La vente à découvert est autorisée, de plus tous les actifs sont infiniment divisibles.

En utilisant l'équation martingale (8.1.5), on obtient la distribution neutre au risque des rendements, qui est normale avec une moyenne  $r - \frac{\sigma^2}{2}$  et une variance  $\sigma^2$ .

À partir de cette information, on peut dériver la formule de Black-Scholes pour le prix d'une option d'achat ou de vente (8.1.4) européenne :

$$C(S(t), K, T) = \Phi(\Delta_1) S(t) - \Phi(\Delta_2) KB(t, T) \quad (8.2.1)$$

$$\Delta_1 = \frac{\ln\left(\frac{S(t)}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}$$

$$\Delta_2 = \Delta_1 - \sigma\sqrt{T-t}$$

$$P(S(t), K, T) = \Phi(-\delta_2)KB(t, T) - \Phi(-\delta_1)S(t). \quad (8.2.2)$$

L'avantage de ce modèle est que toutes les données requises pour évaluer le prix d'options sur des actifs cotés sur les marchés boursiers, à l'exception de la volatilité, sont accessibles auprès de fournisseurs d'informations financières. Comme les options sont négociées sur les marchés financiers, on a aussi accès à leur prix, ce qui permet d'extrapoler la valeur de la volatilité implicite. C'est pour cette raison que le modèle, bien qu'il soit basé sur des hypothèses très restrictives, est toujours utilisé comme point de référence.

## 8.3 Méthodes d'évaluation pour options européennes

Le modèle de Black-Scholes est un des seuls qui présentent une forme analytique simple pour le prix des options. Bien qu'on puisse en dériver une pour plusieurs modèles en utilisant les équations différentielles stochastiques et le calcul d'Îto, les résultats sont souvent difficiles à utiliser. On préférera alors utiliser des méthodes numériques pour calculer le prix des options.

### 8.3.1 Méthode de Heston

L'approche utilisée par Heston (1993) est de calculer directement l'espérance de la créance éventuelle associée à l'option de vente européenne  $P(S(t), K, T - t)$  sous la mesure neutre au risque  $\mathbb{Q}$  :

$$P(S(t), K, T) = B(t, T) \int_0^\infty \max(K - S(T), 0) \cdot dF_{S(T)} \quad (8.3.1)$$

$$\begin{aligned} &= B(t, T)K \int_0^K dF_{S(T)} - B(t, T) \int_0^K S(T) \cdot dF_{S(T)} \\ &= B(t, T)KF_{S(t)}(K) - S(t)G_{S(t)}(K). \end{aligned} \quad (8.3.2)$$

On définit la fonction de répartition  $F_{S(t)}(K)$  de la variable aléatoire  $S(t) \leq K$  sous la mesure  $\mathbb{Q}$ . On définit aussi la fonction de répartition  $G_{S(t)}(K)$  sous une autre mesure équivalant à  $\mathbb{Q}$ . Soit  $\phi_F(\xi)$  la fonction caractéristique de  $S(T)$  sous  $\mathbb{Q}$ , celle de cette mesure est alors définie comme suit :

$$\phi_G(\xi) = \frac{\phi_F(\xi - i)}{\phi_F(-i)}. \quad (8.3.3)$$

C'est une transformée d'Esscher de paramètre  $h = 1$  de la mesure  $\mathbb{Q}$ . On remarquera la forme de (8.3.2) qui est très similaire à la formule du prix de l'option de vente du modèle de Black-Scholes (8.2.2).

Les fonctions de répartition  $F_{S(t)}(K)$  et  $G_{S(t)}(K)$  peuvent être évaluées à partir de (A.2.11) ou de la méthode du point de selle.

Carr et Madan (1999) notent qu'on ne peut pas inverser l'approximation (A.2.11) en utilisant la méthode de la transformée de Fourier rapide. Cependant, l'approche de la méthode du point de selle n'a pas été considérée, bien qu'elle puisse fournir une solution analytique dans plusieurs situations, en particulier pour la distribution de Laplace asymétrique généralisée (3.4.3).

### 8.3.2 Méthode de Carr et Madan

La méthode de Heston nécessite l'évaluation de deux probabilités, donc deux inversions de fonctions caractéristiques. Dilip B Madan et Chang (1998) ont développé une méthode qui ne nécessite qu'une seule inversion, décrite par Epps (2007). Cette méthode nécessite cependant la sélection d'un paramètre d'amortissement  $\theta_D$  par tâtonnement afin d'obtenir de bons résultats. Elle utilise l'expression de la créance éventuelle *amortie*<sup>1</sup> de l'option d'achat, construite à partir de l'expression (8.1.3) :

$$C_{\theta_D}(S(t), k, T) = e^{\theta_D k} C(S(t), k, T) \quad (8.3.4)$$

$$= e^{\theta_D k} B(t, T) \int_K^\infty (e^{s(T)} - e^k) dF_s(T). \quad (8.3.5)$$

---

1. traduction de l'anglais damped

On applique la transformée de Fourier :

$$\mathcal{F}C_{\theta_D}(\nu) = B(t, T) \frac{\phi(\nu - i(1 + \theta_D))}{(\theta_D + i\nu)(1 + \theta_D + i\nu)}. \quad (8.3.6)$$

Ce résultat est inversé et *amplifié*<sup>2</sup> pour retrouver le prix de l'option :

$$\begin{aligned} C(S(t), k, T) &= e^{-\theta_D k} \mathcal{F}^{-1}(\mathcal{F}C_{\theta_D}(\nu)) \\ &= \frac{e^{-\theta_D k}}{2\pi} \int_{-\infty}^{\infty} e^{-i\nu k} (\mathcal{F}C_{\theta_D})(\nu) \cdot d\nu \\ &= \frac{e^{-\theta_D k}}{\pi} \int_0^{\infty} e^{-i\nu k} (\mathcal{F}C_{\theta_D})(\nu) \cdot d\nu \quad (\text{par symétrie}). \end{aligned} \quad (8.3.7)$$

On peut utiliser la méthode de la transformée de Fourier rapide (section A.3) pour retrouver les prix d'options en se référant à Carr et Madan (1999).

On doit donc discrétiser (8.3.7) et effectuer des changements de variables pour obtenir la forme (A.3.1). On fixe le nombre de points  $N$  (idéalement une puissance de deux) et  $\eta$  le pas de discrétisation. On définit l'incrément  $\lambda$  pour le vecteur  $k_u$ ,  $u = 1, \dots, N$  des logarithmes des prix d'exercice :

$$\lambda = \frac{2\pi}{N\eta}. \quad (8.3.8)$$

La valeur maximale  $b$  de ce prix d'exercice sera :

$$b = \frac{N\lambda}{2}. \quad (8.3.9)$$

Enfin, on obtient la suite des points d'évaluation de la fonction caractéristique  $\nu_j = \eta(j - 1)$ . L'ensemble de ces substitutions permet d'obtenir une expression qui est de la forme requise pour appliquer la méthode de la transformée de Fourier rapide.

$$\begin{aligned} C(S(t), k_u, T) &\approx \frac{e^{-\theta_D k}}{\pi} \sum_{j=1}^N e^{-i\nu_j k} (\mathcal{F}C_{\theta_D})(\nu_j) \\ &\approx \frac{e^{-\theta_D k}}{\pi} \sum_{j=1}^N e^{-i\lambda\eta(j-1)(u-1)} e^{ib\nu_j} (\mathcal{F}C_{\theta_D})(\nu_j)\eta. \end{aligned} \quad (8.3.10)$$

C'est essentiellement une intégration par la méthode du trapèze. On suggère d'implémenter la correction de Simpson, qui permet une intégration plus précise tout en conservant le même nombre de points de discrétisation. On obtiendra alors la formule suivante :

$$C(S(t), k_u, T) \approx \frac{e^{-\theta_D k}}{\pi} \sum_{j=1}^N e^{-i\lambda\eta(j-1)(u-1)} e^{ib\nu_j} (\mathcal{F}C_{\theta_D})(\nu_j) \frac{\eta}{3} (3 + (-1)^j + \delta_{j-1}). \quad (8.3.11)$$

---

2. traduction de l'anglais amplified

On peut par la suite calculer un prix d'option pour chaque prix d'exercice en utilisant une méthode d'interpolation (splines cubiques par exemple).

### 8.3.3 Prix d'exercice hors du cours

L'intégration numérique de (8.3.7) pose problème quand l'échéance  $T - t$  est petite, ou encore le prix d'exercice  $K$  est hors du cours ( $k > \ln(S(0))$ ). Dans ce contexte particulier, Carr et Madan (1999) développent une formule alternative à l'équation (8.3.7) pour évaluer le prix de l'option d'achat :

$$C(S(t), k, T) = \frac{1}{\pi \sinh(\theta_D^* k)} \int_0^\infty e^{-i\nu k} (\mathcal{F}C_{\theta_D^*}^*)(\nu) \cdot d\nu. \quad (8.3.12)$$

La transformée de Fourier de l'expression du prix de l'option d'achat européenne est exprimée sous la forme

$$\begin{aligned} \mathcal{F}C_{\theta_D^*}^* &= \frac{\zeta_T(\nu - i\theta_D^*) - \zeta_T(\nu + i\theta_D^*)}{2} \\ \zeta_T(\nu) &= e^{-r(T-t)} \left[ \frac{1}{1 + i\nu} - \frac{e^{r(T-t)}}{i\nu} - \frac{\phi(\nu - i)}{\nu(\nu - i)} \right]. \end{aligned} \quad (8.3.13)$$

On note que ce paramètre d'amortissement  $\theta_D^*$  peut être différent du paramètre  $\theta_D$  qui a été utilisé précédemment.

L'expression à utiliser pour la méthode de la transformée de Fourier rapide devient

$$C^*(S(t), k_u, T) \approx \frac{1}{\pi \sinh(\theta_D^* k)} \sum_{j=1}^N e^{-i\lambda\eta(j-1)(u-1)} e^{i\nu_j} (\mathcal{F}C_{\theta_D^*}^*)(\nu_j) \frac{\eta}{3} (3 + (-1)^j + \delta_{j-1}). \quad (8.3.14)$$

Les paramètres  $N, \eta, \lambda$  et  $\nu_j$  prennent les valeurs utilisées pour évaluer le prix dans la situation où le titre est dans le cours (8.3.11).

### 8.3.4 Critique de la méthode de Carr-Madan

Un paramètre d'amortissement  $\theta_D$  inférieur à la valeur optimale (lorsque comparé avec la méthode de Heston ou d'Epps) aura tendance à surestimer le prix des options d'achat européennes. Itkin (2005) démontre que pour certaines régions de l'espace des paramètres, l'intégrale (8.3.7) a un comportement très irrégulier. De plus, il décrit plusieurs restrictions pour  $\theta_D$  qui, dans plusieurs cas, ne permettent pas d'estimation convergente.

### 8.3.5 Méthode d'Epps

Epps (2007) propose une méthode qui, contrairement à celle de Carr et Madan, ne nécessite pas de paramètre d'amortissement et qui ne requiert aussi qu'une seule inversion de la fonction

caractéristique. Pour ce faire, on se base sur l'expression de l'option de vente développée précédemment (8.3.1), qui est ensuite exprimée sous la forme du logarithme, à la manière de (8.1.3). Après l'utilisation de la formule d'inversion (A.2.8), on obtient le résultat suivant :

$$\begin{aligned} E[\max(K - S_T; 0)] &= \int_{-\infty}^k F_{S(T)}(s) e^s \cdot ds \\ &= \frac{K}{2} - \frac{1}{2} \int_{-\infty}^k \lim_{c \rightarrow \infty} \underbrace{\int_{-c}^c \frac{e^{-i\nu s}}{\pi i \nu} \phi(\nu) d\nu}_{a(c,s)} e^s ds. \end{aligned} \quad (8.3.15)$$

La formule d'inversion implique que lorsque  $|\lim_{c \rightarrow \infty} a(c, s)| = |1 - 2F_{S(t)}(s)| \leq 1$ , pour toute valeur  $\epsilon > 0$ ,  $c_\epsilon$  existe telle que l'on a le résultat suivant :

$$|\sup_s [1 - 2F_{S(t)}(s) - a(c_\epsilon, s)]| < \epsilon. \quad (8.3.16)$$

Dans cette situation,  $|a(c, s)e^s| \leq e^s(1 + \epsilon)$  lorsque la constante  $c$  est suffisamment grande. Comme la fonction  $e^s$  est intégrable sur le support  $(-\infty, k)$ , le théorème de convergence dominée de Lebesgue (section A.5.1) implique que l'on peut poser l'égalité suivante :

$$\begin{aligned} \int_{-\infty}^k \lim_{c \rightarrow \infty} a(c, s) e^s ds &= \lim_{c \rightarrow \infty} \int_{-\infty}^k a(c, s) e^s ds \\ &= \lim_{c \rightarrow \infty} \int_{-\infty}^k \int_{-c}^c \frac{e^{(1-i\nu)s}}{\pi i \nu} \phi(\nu) d\nu ds. \end{aligned}$$

De plus, comme la limite de la double intégrale précédente est égale à  $K - 2E[\max(K - S_T; 0)]$  qui appartient à l'intervalle  $[-K, K]$ , le théorème de Fubini (section A.5.2) permet d'inverser l'ordre d'intégration :

$$\begin{aligned} \lim_{c \rightarrow \infty} \int_{-\infty}^k \int_{-c}^c \frac{e^{-i\nu s}}{\pi i \nu} \phi(\nu) d\nu ds &= \lim_{c \rightarrow \infty} \int_{-c}^c \int_{-\infty}^k e^{(1-i\nu)s} ds \frac{\phi(\nu)}{\pi i \nu} d\nu \\ &= \frac{K}{\pi} \lim_{c \rightarrow \infty} \int_{-c}^c \frac{K^{-i\nu}}{i\nu + \nu^2} \phi(\nu) d\nu. \end{aligned} \quad (8.3.17)$$

En remplaçant le résultat (8.3.17) dans l'équation de départ (8.3.15), on obtient ainsi une expression particulièrement simple pour le prix de l'option d'achat :

$$P(S(t), K, T) = B(t, T) K \left[ \frac{1}{2} - \frac{1}{2\pi} \int_{-c}^c K^{-i\nu} \frac{\phi(\nu)}{\nu(i + \nu)} d\nu \right]. \quad (8.3.18)$$

Cependant, puisque la forme ne se prête pas à l'utilisation de l'algorithme de la transformée de Fourier rapide, on a recours à une procédure d'intégration numérique.



## 8.4 Particularités

### 8.4.1 Option sur actions avec dividendes

Lorsqu'une option a, pour titre sous-jacent, une action qui verse des dividendes, on doit en tenir compte dans l'évaluation de son prix. Si l'on considère que la valeur au marché de l'action  $S^*(t)$  a été évaluée avec la méthode de l'actualisation des flux financiers futurs, on doit tenir compte de la valeur actualisée des dividendes qui seront versés avant l'échéance de l'option et la soustraire de ce prix.

Parfois, les dividendes ne sont pas fixes, mais proportionnels à la valeur de l'action à la date ex-dividende avec un taux  $\delta$ . Soit  $n(T)$  le nombre de dates ex-dividende dans l'intervalle  $]t, T]$ , le prix initial  $S(t)$  considéré pour le calcul de la valeur de l'option est alors défini comme suit :

$$S(t) = S^*(t)(1 - \delta)^{n(T)}. \quad (8.4.1)$$

Lorsque l'on considère un indice composé de plusieurs titres, on peut prendre un dividende versé de manière continue à un taux  $q$ . Dans ce cas, le prix initial  $S(t)$  sera

$$S(t) = S^*(t)e^{-q(T-t)}. \quad (8.4.2)$$

### 8.4.2 Options sur contrats à terme et taux de change

Black (1976) (p.177) démontre que le prix d'une option sur un contrat à terme a le même prix que sur une action dont le taux de dividende équivaut au taux sans risque. On peut donc utiliser les résultats (8.4.1) et (8.4.2) en posant  $q = r$ .

De même, on peut évaluer le prix d'une option sur une monnaie étrangère en considérant le taux sans risque étranger  $r_f$  comme un taux de dividende  $q = r_f$ . Selon la Banque des Règlements Internationaux, c'est le type d'options le plus transigé sur les marchés non réglementés en date de 2005, même s'il est moins étudié que les autres.



# Chapitre 9

## Exemple d'application

On considère un échantillon  $S_1$  formé de l'ensemble des prix  $S_1(t)$  à la fermeture du titre Abbey National entre le 31 juillet et le 8 octobre 1991. La table C.1 présente l'ensemble des 50 observations. Cet échantillon a notamment été étudié précédemment par Buckle (1995).

### 9.1 Description des données

On évalue tout d'abord les rendements quotidiens  $R_1$  à l'aide de l'équation (1.2.2). On obtient alors 49 observations du processus des rendements  $R(t)$ , que l'on présente à la figure 9.1 sous forme de série chronologique.

On énumère d'abord quelques propriétés de cet échantillon qui pourront compléter l'analyse. À la table 9.1, on présente quelques statistiques d'ordre. De plus, à la table 9.2, on retrouve quelques valeurs relatives aux premiers moments.

Statistique d'ordre	Valeur
Minimum	-0.027500
1er quartile	-0.009790
Médiane	-0.003260
3e Quartile	0.006620
Maximum	0.043400

TABLE 9.1: Statistiques d'ordre de l'échantillon  $R_1$

On présente maintenant la distribution des rendements sous la forme d'une courbe de densité à la figure 9.2.

À l'aide du test de normalité d'Epps-Pulley (Section 7.1.2), on vérifie si la distribution des rendements  $R_1$  est significativement différente de la normale. On fixe le seuil de tolérance  $\alpha = 5\%$ .

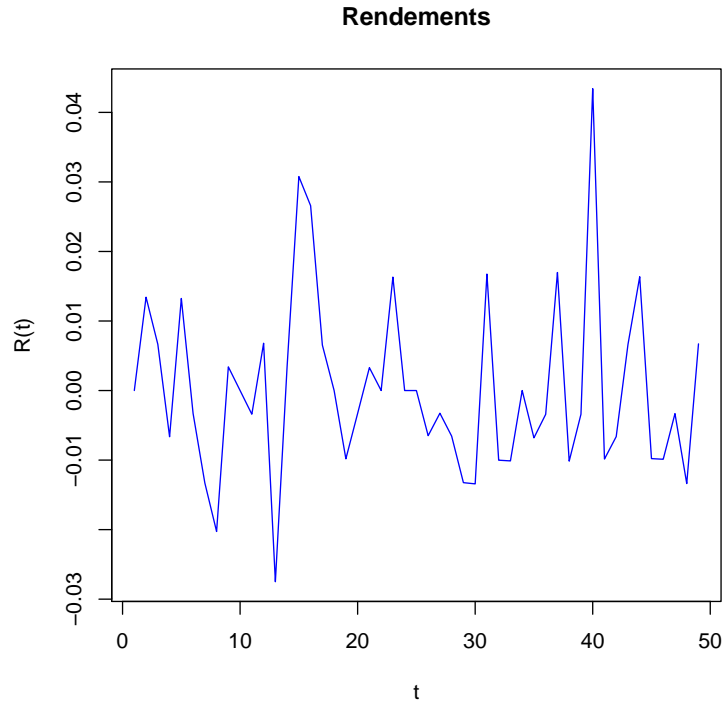


FIGURE 9.1: Représentation en série chronologique de l'échantillon  $R_1$

Statistique	Valeur
Moyenne	0.000206
Variance	0.000169
Coefficient d'asymétrie	0.977563
Coefficient d'aplatissement	4.597114

TABLE 9.2: Valeurs relatives aux premiers moments de l'échantillon  $R_1$

On évalue d'abord la statistique  $EP_T$  (7.1.3), puis la statistique modifiée  $EP_T^*$  (7.1.5) étant donné que  $T > 10$ . La table 9.3 présente les résultats.

Statistique	Valeur
$EP_T$	0.626033
$EP_T^*$	0.635568
$Z_T$	2.44824
$p$	0.007178

TABLE 9.3: Test de normalité d'Epps-Pulley pour  $R_1$

Étant donné que la valeur  $p$  est inférieure au seuil  $\alpha$ , on rejette l'hypothèse de normalité de l'échantillon  $R_1$ . On peut vérifier cette affirmation à l'aide du graphique de comparaison des

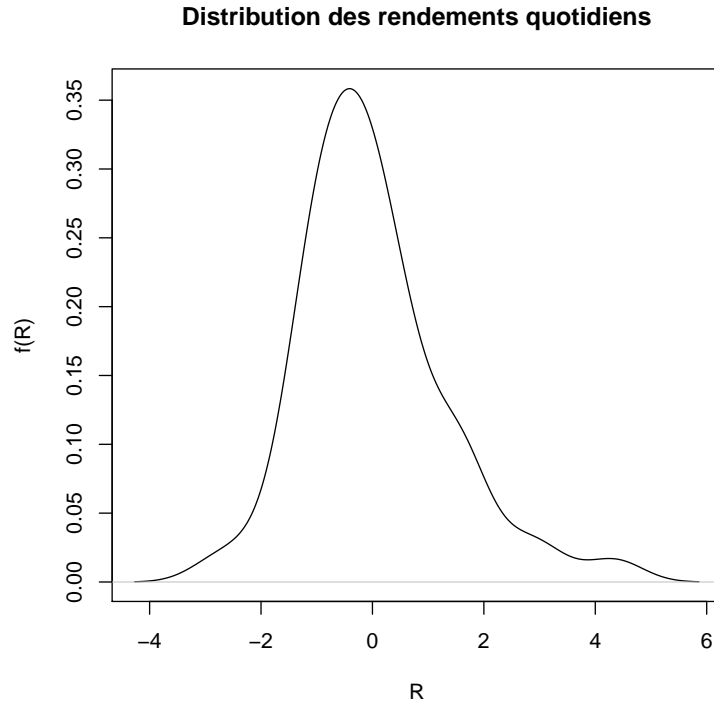


FIGURE 9.2: Distribution de la variable aléatoire  $R_1$

quantiles empiriques avec ceux de la loi normale présenté à la figure 9.3.

## 9.2 Estimation

Dans cette section, on estime les paramètres de la distribution de Laplace asymétrique généralisée pour l'échantillon de données centrées et réduites  $R_1^*$ , à l'aide des méthodes :

- des moments généralisée itérative (Section 4.2.4)
- d'estimation gaussienne de Whittle (table 5.1)
- de l'équation d'estimation optimale de Crowder (Section 5.1) et
- de l'équation d'estimation optimale modifiée (Section 5.2).

On évalue d'abord le vecteur de paramètres initiaux  $\theta_0$  à l'aide des équations (6.1.2). On effectue ensuite une première optimisation à l'aide de chacune des trois méthodes afin de pouvoir évaluer la matrice de pondération qui servira à la seconde. La table 9.4 présente les paramètres obtenus  $\theta_1$ .

On obtient la matrice de variance-covariance des paramètres pour les méthodes basées sur une équation d'estimation (5.1.4) en évaluant l'inverse de la variance-covariance des éléments composant celle-ci (5.1.3b) et le gradient (5.1.3a) :

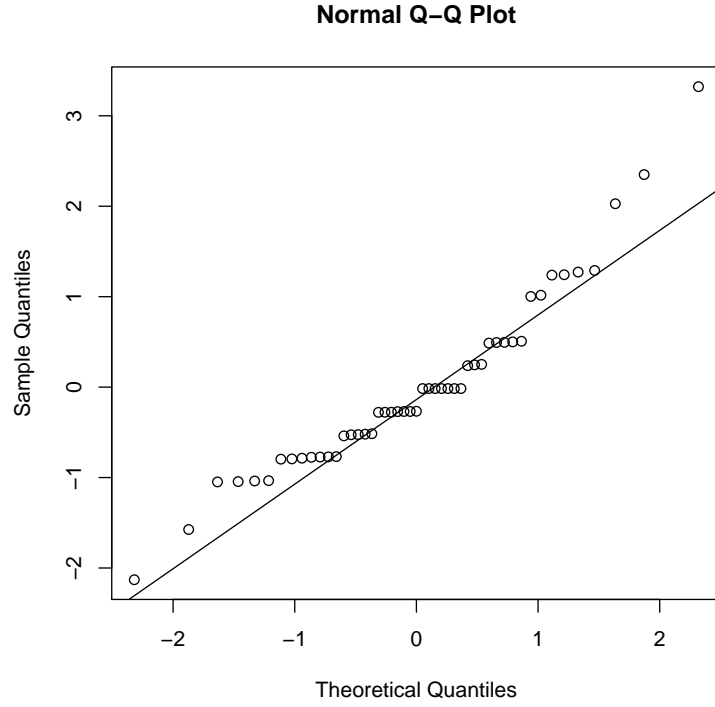


FIGURE 9.3: Graphique Quantile-Quantile

Méthode	$\theta$	$\sigma$	$\mu$	$\tau$
Paramètres initiaux	-0.612081	0.515932	0.325854	1.878388
Moments généralisée	-0.641646	0.625908	0.326366	1.965995
Estimation gaussienne	-0.776204	0.581154	0.385120	2.015437
Équation d'estimation optimale	-0.660300	0.645678	0.376654	1.753069
Équation d'estimation optimale modifiée	-0.711439	0.606642	0.362932	1.960299

TABLE 9.4: Paramètres  $\theta_1$  de la première optimisation

— Pour la méthode d'estimation gaussienne :

$$\mathbf{S}_T^*(\hat{\theta}_1; R_1^*) = \begin{bmatrix} 1.13756e-03 & -0.00207546 & 0.000917316 & 7.46064e-06 \\ -2.07546e-03 & 0.00984423 & 0.002340623 & 1.24328e-03 \\ 9.17316e-04 & 0.00234062 & 0.003399880 & 8.38932e-04 \\ 7.46064e-06 & 0.00124328 & 0.000838932 & 2.60840e-04 \end{bmatrix} \quad (9.2.1)$$

— Pour la méthode de l'équation d'estimation optimale :

$$\mathbf{S}_T^*(\hat{\theta}_1; R_1^*) = \begin{bmatrix} 1.47029e-03 & -0.00212833 & 0.00133597 & 2.84688e-05 \\ -2.12833e-03 & 0.01115599 & 0.00277669 & 1.95192e-03 \\ 1.33597e-03 & 0.00277669 & 0.00396182 & 1.18855e-03 \\ 2.84688e-05 & 0.00195192 & 0.00118855 & 4.92502e-04 \end{bmatrix} \quad (9.2.2)$$

— Pour la méthode de l'équation d'estimation optimale modifiée :

$$\mathbf{S}_T^*(\hat{\theta}_1; R_1^*) = \begin{bmatrix} 1.03165e-03 & -0.00146603 & 0.001145269 & 6.63861e-05 \\ -1.46603e-03 & 0.00812868 & 0.001989247 & 1.17588e-03 \\ 1.14527e-03 & 0.00198925 & 0.003435163 & 8.33622e-04 \\ 6.63861e-05 & 0.00117588 & 0.000833622 & 2.71162e-04 \end{bmatrix} \quad (9.2.3)$$

On peut ainsi construire des intervalles de confiance pour les paramètres estimés. On utilise un seuil de tolérance de  $\alpha = 5\%$  :

— Pour la méthode d'estimation gaussienne :

	<b>Borne inférieure</b>	<b>Valeur estimée</b>	<b>Borne supérieure</b>
$\theta$	-0.8328	-0.7762	-0.7197
$\sigma$	0.4148	0.5812	0.7475
$\mu$	0.2874	0.3851	0.4829
$\tau$	1.9884	2.0154	2.0425

— Pour la méthode de l'équation d'estimation optimale :

	<b>Borne inférieure</b>	<b>Valeur estimée</b>	<b>Borne supérieure</b>
$\theta$	-0.7246	-0.6603	-0.5960
$\sigma$	0.4686	0.6457	0.8228
$\mu$	0.2711	0.3767	0.4822
$\tau$	1.7159	1.7531	1.7903

— Pour la méthode de l'équation d'estimation optimale modifiée :

	<b>Borne inférieure</b>	<b>Valeur estimée</b>	<b>Borne supérieure</b>
$\theta$	-0.7653	-0.7114	-0.6576
$\sigma$	0.4555	0.6066	0.7578
$\mu$	0.2647	0.3629	0.4612
$\tau$	1.9327	1.9603	1.9879

À l'aide de l'inverse de la variance-covariance des conditions de moments et des équations d'estimation, on peut effectuer une seconde optimisation afin d'obtenir des estimateurs convergents. Pour la méthode des moments généralisée, on utilisera plutôt une procédure itérative. Les paramètres présentés à la table 9.5 sont donc ceux obtenus avec la convergence de l'algorithme itératif de la section 4.2.4 avec un critère d'arrêt  $\epsilon = 10^{-7}$  (4.2.23).

Pour la méthode des moments généralisée, puisque l'on utilise une procédure itérative, présenter la première matrice de variance-covariance des paramètres n'est pas pertinent. On obtient celle à la convergence de l'algorithme (4.2.35) en évaluant l'inverse de la variance-covariance

Méthode	$\theta$	$\sigma$	$\mu$	$\tau$
Moments généralisée	-0.640067	0.625431	0.324311	1.973623
Estimation gaussienne	-0.775071	0.581413	0.384342	2.016619
Équation d'estimation optimale	-0.658697	0.646516	0.376251	1.750685
Équation d'estimation optimale modifiée	-0.712450	0.606193	0.363196	1.961614

TABLE 9.5: Paramètres  $\theta_1$  de la première optimisation

des conditions de moments (4.2.19) et le gradient (4.2.25). On obtient donc :

$$\mathbf{S}_T^*(\hat{\theta}_{OPT}; R_1^*) = \begin{bmatrix} 0.00203708 & 0.00438553 & 0.00174636 & 0.00154237 \\ 0.00438553 & 0.01207044 & 0.00239640 & 0.00384905 \\ 0.00174636 & 0.00239640 & 0.00220402 & 0.00104816 \\ 0.00154237 & 0.00384905 & 0.00104816 & 0.00127406 \end{bmatrix}. \quad (9.2.4)$$

Pour les méthodes basées sur une équation d'estimation, on obtient, pour la seconde optimisation, les variances-covariances suivantes :

— Pour la méthode d'estimation gaussienne :

$$\mathbf{S}_T^*(\hat{\theta}_2; R_1^*) = \begin{bmatrix} 1.13314e-03 & -0.00206602 & 0.000919367 & 7.53773e-06 \\ -2.06602e-03 & 0.00983078 & 0.002332240 & 1.24238e-03 \\ 9.19367e-04 & 0.00233224 & 0.003395736 & 8.36472e-04 \\ 7.53773e-06 & 0.00124238 & 0.000836472 & 2.60254e-04 \end{bmatrix} \quad (9.2.5)$$

— Pour la méthode de l'équation d'estimation optimale :

$$\mathbf{S}_T^*(\hat{\theta}_2; R_1^*) = \begin{bmatrix} 1.47327e-03 & -0.00212557 & 0.00134221 & 2.89119e-05 \\ -2.12557e-03 & 0.01116770 & 0.00277803 & 1.96072e-03 \\ 1.34221e-03 & 0.00277803 & 0.00396652 & 1.19169e-03 \\ 2.89119e-05 & 0.00196072 & 0.00119169 & 4.95537e-04 \end{bmatrix} \quad (9.2.6)$$

— Pour la méthode de l'équation d'estimation optimale modifiée :

$$\mathbf{S}_T^*(\hat{\theta}_2; R_1^*) = \begin{bmatrix} 1.03121e-03 & -0.00146899 & 0.001142696 & 6.60715e-05 \\ -1.46899e-03 & 0.00812703 & 0.001987649 & 1.17298e-03 \\ 1.14270e-03 & 0.00198765 & 0.003432414 & 8.32388e-04 \\ 6.60715e-05 & 0.00117298 & 0.000832388 & 2.70299e-04 \end{bmatrix} \quad (9.2.7)$$

On peut donc construire des intervalles de confiance :

— Pour la méthode des moments généralisée :



	<b>Borne inférieure</b>	<b>Valeur estimée</b>	<b>Borne supérieure</b>
$\theta$	-0.715736	-0.640067	-0.564397
$\sigma$	0.441236	0.625431	0.809627
$\mu$	0.245602	0.324311	0.403020
$\tau$	1.913780	1.973623	2.033466

— Pour la méthode d'estimation gaussienne :

	<b>Borne inférieure</b>	<b>Valeur estimée</b>	<b>Borne supérieure</b>
$\theta$	-0.8315	-0.7751	-0.7186
$\sigma$	0.4152	0.5814	0.7476
$\mu$	0.2866	0.3843	0.4820
$\tau$	1.9896	2.0166	2.0437

— Pour la méthode de l'équation d'estimation optimale :

	<b>Borne inférieure</b>	<b>Valeur estimée</b>	<b>Borne supérieure</b>
$\theta$	-0.7230	-0.6587	-0.5943
$\sigma$	0.4693	0.6465	0.8237
$\mu$	0.2707	0.3763	0.4818
$\tau$	1.7134	1.7507	1.7880

— Pour la méthode de l'équation d'estimation optimale modifiée :

	<b>Borne inférieure</b>	<b>Valeur estimée</b>	<b>Borne supérieure</b>
$\theta$	-0.7663	-0.7124	-0.6586
$\sigma$	0.4551	0.6062	0.7573
$\mu$	0.2650	0.3632	0.4614
$\tau$	1.9341	1.9616	1.9892

En utilisant la propriété (2.2.18), on retrouve les paramètres, correspondants aux données  $R(t) = \sqrt{Var[R_1]}R^*(t) + E[R_1]$ ,  $t = 1, \dots, 50$ , présentés à la table 9.6.

<b>Méthode</b>	$\theta$	$\sigma$	$\mu$	$\tau$
Moments généralisée	-0.008119	0.008134	0.002983	1.973623
Estimation gaussienne	-0.009875	0.007562	0.003535	2.016619
Équation d'estimation optimale	-0.008361	0.008409	0.003460	1.750685
Équation d'estimation optimale modifiée	-0.009060	0.007884	0.003340	1.961614

TABLE 9.6: Paramètres des données  $R_1$

### 9.3 Approximation

On effectue l'approximation des fonctions de densité et de répartition pour un ensemble de points  $r \in \{-0.03, -0.02, -0.01, 0, 0.01, 0.02, 0.03\}$  à l'aide de la méthode du point de selle. On utilise les paramètres obtenus par la méthode des moments généralisée.

On résout d'abord l'équation du point de selle (3.3.10). On illustre graphiquement cette équation pour  $r = 0.01$  à la figure 9.4 :

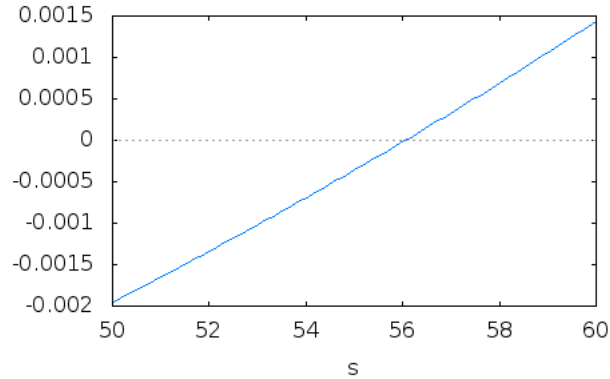


FIGURE 9.4: Équation du point de selle pour  $r = 0.01$

Le point de selle prend alors une valeur de  $\hat{s} = 56.050951$  dans cette situation. En utilisant l'approximation de premier ordre  $\hat{f}_{R,1}(r)$  (3.3.12) de la fonction de densité, et en comparant le résultat avec la fonction de densité  $f_R(r)$  (2.2.26), on obtient les résultats présentés à la table 9.7. Dans les deux cas, la fonction de densité a été normalisée à l'aide de l'intégrale (3.3.13a).

$r$	Densité		Erreur relative
	$\hat{f}_{R,1}(r)$	$f_R(r)$	$(\hat{f}_{R,1}(r) - f_R(r))/f_R(r)$
-0.03	1.561365	2.065547	-0.244091
-0.02	9.038616	10.702896	-0.155498
-0.01	31.216317	37.581108	-0.169361
0.00	33.015869	31.073988	0.062492
0.01	15.941633	12.450742	0.280376
0.02	5.988373	4.120017	0.453483
0.03	2.026838	1.246205	0.626409

TABLE 9.7: Approximation de la densité de  $R_1$

On évalue ensuite la valeur de la fonction de répartition à l'aide de l'approximation de premier ordre  $\hat{F}_{R,1}(r)$  (3.3.20). On compare celle-ci à la valeur de la fonction de répartition obtenue en intégrant numériquement la fonction de densité. Les résultats sont présentés à la table 9.8.

$r$	Fonction de répartition		Erreur relative
	$\hat{F}_{R,1}(r)$	$F_R(r)$	$(\hat{F}_{R,1}(r) - F_R(r))/F_R(r)$
-0.03	0.007402	0.011577	-0.360606
-0.02	0.048942	0.064944	-0.246402
-0.01	0.249911	0.292034	-0.144241
0.00	0.625225	0.681167	-0.082126
0.01	0.857830	0.889832	-0.035965
0.02	0.951508	0.965916	-0.014916
0.03	0.984393	0.990080	-0.005744

TABLE 9.8: Approximation de la fonction de répartition de  $R_1$

## 9.4 Graphiques

On illustre graphiquement, aux figures 9.5 et 9.6, la fonction de densité de la distribution de Laplace asymétrique généralisée avec les paramètres estimés par la méthode des moments généralisée et la méthode de l'équation d'estimation optimale. La table 9.9 décrit chacune des courbes.

Abbréviation	Description
Emp.	Données empiriques.
Norm.	Densité de la distribution normale ayant les mêmes moyenne et variance que la variable aléatoire $R$ .
Estim.	Densité de la distribution de Laplace asymétrique généralisée avec les paramètres estimés.
Pt selle o.1	Approximation de premier ordre avec la méthode du point de selle.
FFT	Transformée de Fourier rapide.

TABLE 9.9: Courbes de densité

On remarquera au passage que la méthode du point de selle nécessite une normalisation. La valeur de l'intégrale  $c$  (3.3.13a) pour les quatre méthodes se trouve à la table 9.10.

Méthode	Valeur de l'intégrale $c$
Moments généralisée	0.920148
Estimation gaussienne	0.932246
Équation d'estimation optimale	0.917326
Équation d'estimation optimale modifiée	0.925781

TABLE 9.10: Valeur de l'intégrale de l'approximation de la densité par la méthode du point de selle

### Densité de $R^*$

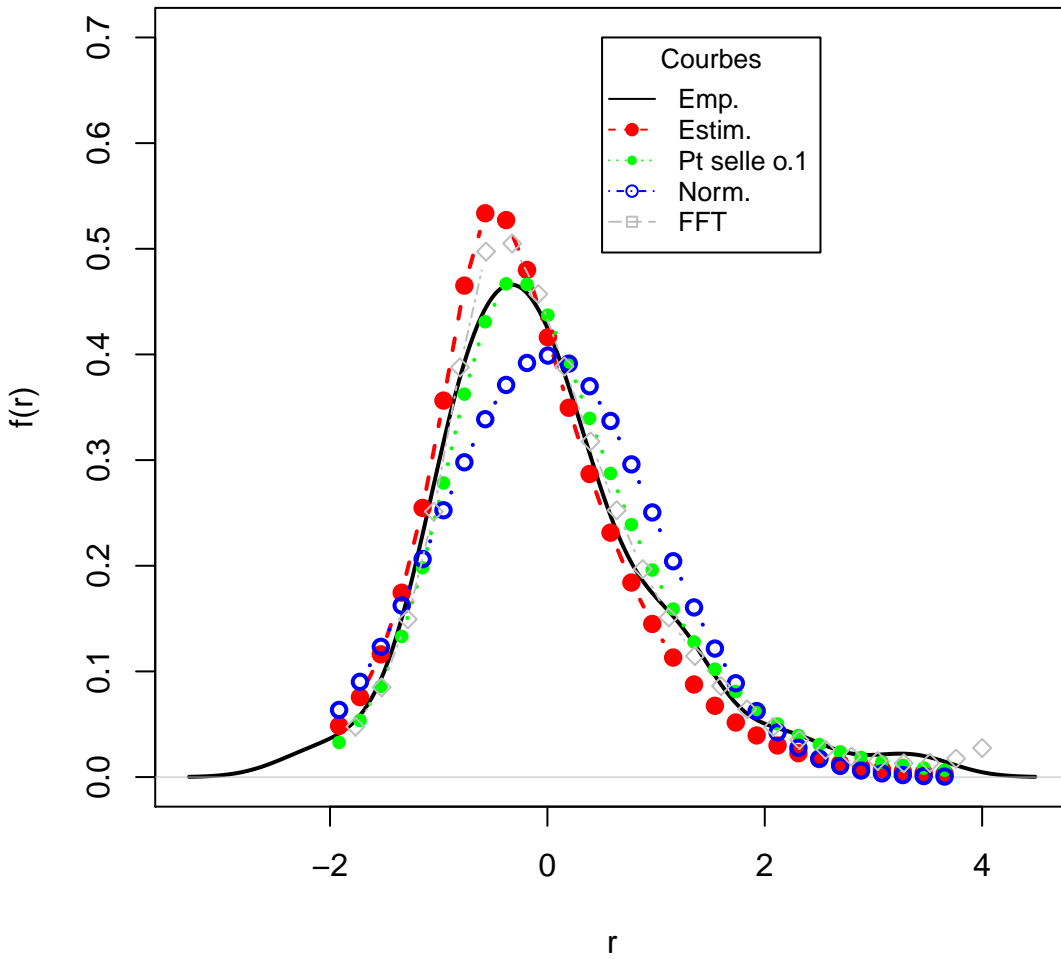


FIGURE 9.5: Densité de  $R_1^*$  selon la méthode des moments généralisée

### Densité de $R^*$

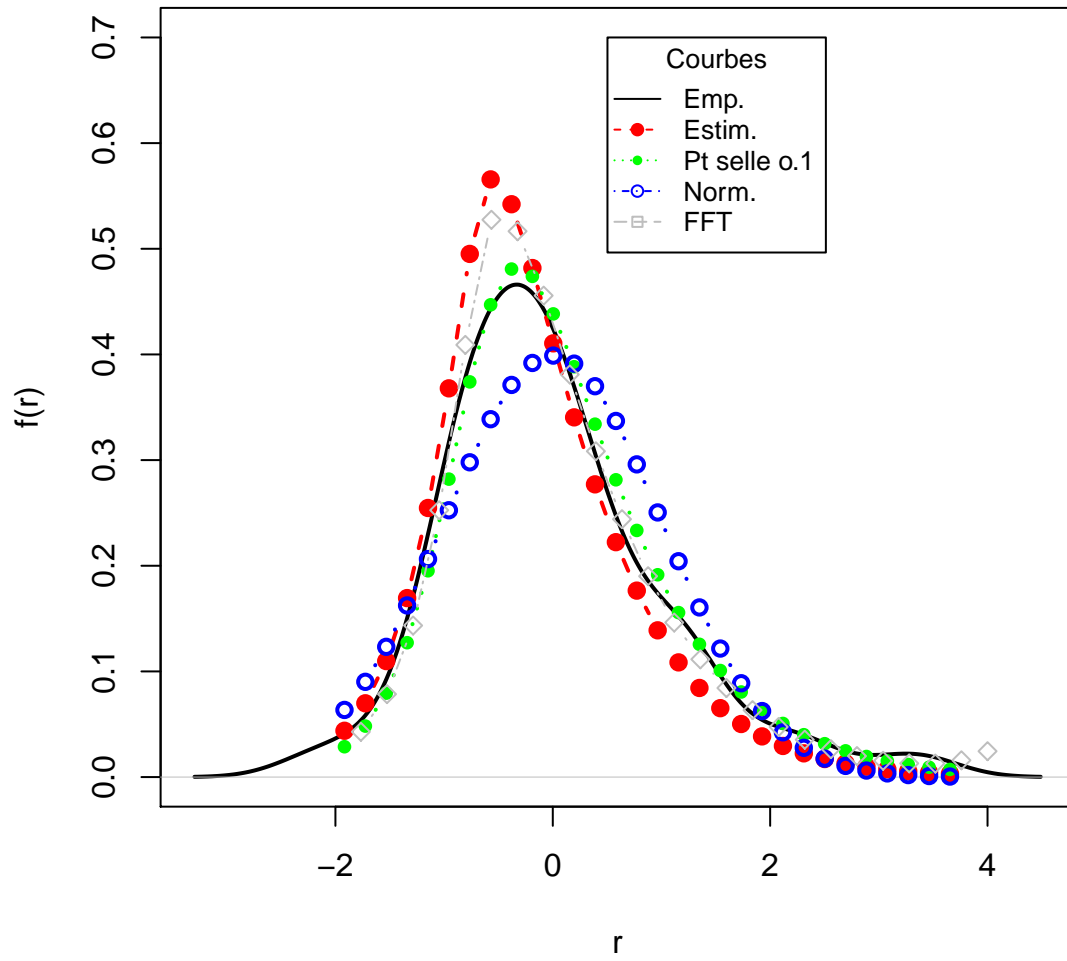


FIGURE 9.6: Densité de  $R_1^*$  selon la méthode de l'équation d'estimation optimale

## 9.5 Tests statistiques

On effectue le test du  $\chi^2$  en utilisant sept classes optimales déterminées par l'algorithme du logiciel GNU R. On obtient la fonction de répartition à partir de la fonction caractéristique en utilisant la formule d'inversion (A.2.11). On se rappelle que cet algorithme produit surtout des erreurs aux extrémités de la distribution. Par contre, ce test donne davantage d'importance aux classes ayant un plus grand nombre de données par construction. On utilisera donc l'inversion plutôt que la méthode du point de selle. On obtient la statistique  $Q_6$  présentée à la table 9.11 à partir de la définition (7.2.4). On utilise un seuil de tolérance  $\alpha = 5\%$ .

Méthode	Valeur de la statistique $Q_6$
Moments généralisée	0.484919
Estimation gaussienne	0.473527
Équation d'estimation optimale	0.531888
Équation d'estimation optimale modifiée	0.494769

TABLE 9.11: Test du  $\chi^2$

Ce test ne permet pas de rejeter l'hypothèse de la distribution de Laplace asymétrique généralisée avec les paramètres obtenus pour aucune des méthodes d'estimation.

On effectue maintenant le test de Kolmogorov-Smirnov. Comme celui-ci est basé sur la fonction de répartition, on utilisera la même méthode que précédemment, impliquant l'inversion de la fonction caractéristique. On évalue la fonction de répartition en chaque point de l'échantillon  $R_1^*$  et l'on obtient la statistique  $D_{49}$  présentée à la table 9.12 à l'aide de la définition (7.2.9). En utilisant un seuil de tolérance de  $\alpha = 5\%$ , on obtient une valeur critique de 0.194286.

Méthode	Valeur de la statistique $D_{49}$
Moments généralisée	0.0684329
Estimation gaussienne	0.0784436
Équation d'estimation optimale	0.0758317
Équation d'estimation optimale modifiée	0.074668

TABLE 9.12: Test de Kolmogorov-Smirnov

Encore une fois, on ne peut pas rejeter l'hypothèse de la distribution de Laplace asymétrique généralisée avec les paramètres estimés pour chacune des méthodes.

Ces deux tests sont approximatifs, puisque les paramètres n'ont pas été estimés en minimisant la statistique utilisée. Cependant, on obtient un test asymptotiquement exact en utilisant une statistique de distance minimale basée sur la fonction génératrice des moments, tel que développé à la section 7.2.3. Avec un seuil de tolérance de  $\alpha = 5\%$ , on obtient une valeur critique de 67.5048. Les statistiques obtenues sont présentées à la table 9.13.

Méthode	Valeur de la statistique $Td(F_{49}, F_\theta)$
Moments généralisée	0.226891
Estimation gaussienne	0.110573
Équation d'estimation optimale	0.074020
Équation d'estimation optimale modifiée	0.108625

TABLE 9.13: Test de distance minimale basé sur la fonction génératrice des moments

On ne peut pas rejeter l'hypothèse de la distribution de Laplace asymétrique généralisée avec les paramètres estimés pour chacune des méthodes avec ce test.

## 9.6 Évaluation d'options

À titre d'exemple, on évaluera une option européenne dont les différentes caractéristiques figurent à la table 9.14.

Caractéristique	Valeur
Type	Option de vente
Échéance ( $T$ )	30 jours
Valeur actuelle du titre ( $S(0)$ )	299
Prix d'exercice ( $K$ )	entre 95% et 105% de la valeur actuelle
Taux sans risque annuel ( $r_f$ )	5%

TABLE 9.14: Caractéristiques de l'option

À partir des paramètres de la table 9.6, en utilisant l'équation martingale appliquée à la distribution de Laplace asymétrique généralisée (8.1.6), on obtient l'ensemble correspondant pour la mesure neutre au risque. On rappelle que celle-ci n'est pas unique étant donné que le processus de Laplace est un processus de sauts. La table 9.15 présente ces paramètres.

	$\theta$	$\sigma$	$\mu$	$\tau$
Moments généralisée	-0.005824	0.008134	0.002983	1.973623
Estimation gaussienne	-0.007062	0.007562	0.003535	2.016619
Équation d'estimation optimale	-0.005993	0.008409	0.003460	1.750685
Équation d'estimation optimale modifiée	-0.006487	0.007884	0.003340	1.961614

TABLE 9.15: Paramètres neutres au risque

On présente les graphiques de la valeur du prix de l'option de vente pour les méthodes des moments généralisée et de l'équation d'estimation optimale aux figures 9.7 et 9.8. On peut facilement remarquer le manque de précision de l'approche de Carr-Madan, qui s'approche de la courbe de Black-Scholes lorsque le titre est dans la monnaie et qui se met à osciller dès que

le titre est hors de la monnaie. Les méthodes de Epps et de Heston donnent des résultats très similaires, et l'approximation du point de selle d'ordre 1 est très précise dans ce contexte.

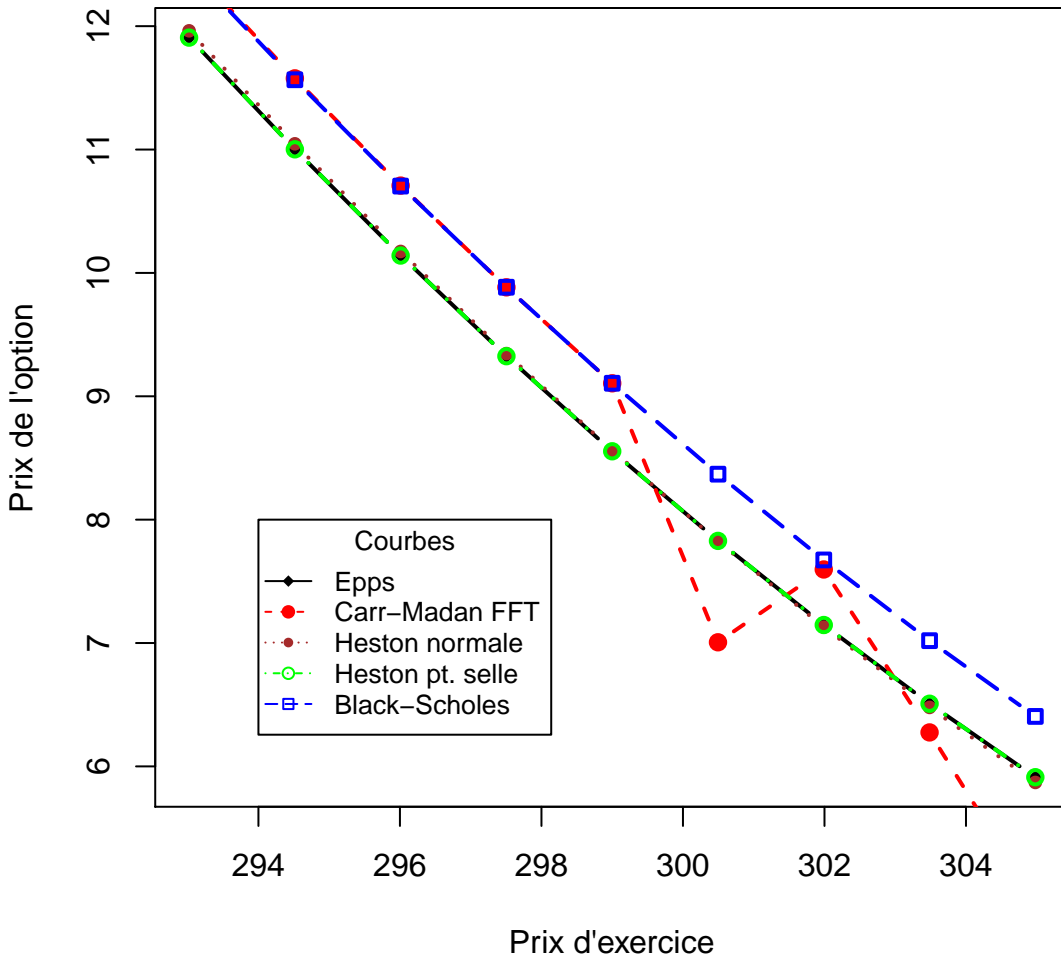


FIGURE 9.7: Prix de l'option selon les paramètres estimés avec la méthode des moments généralisée



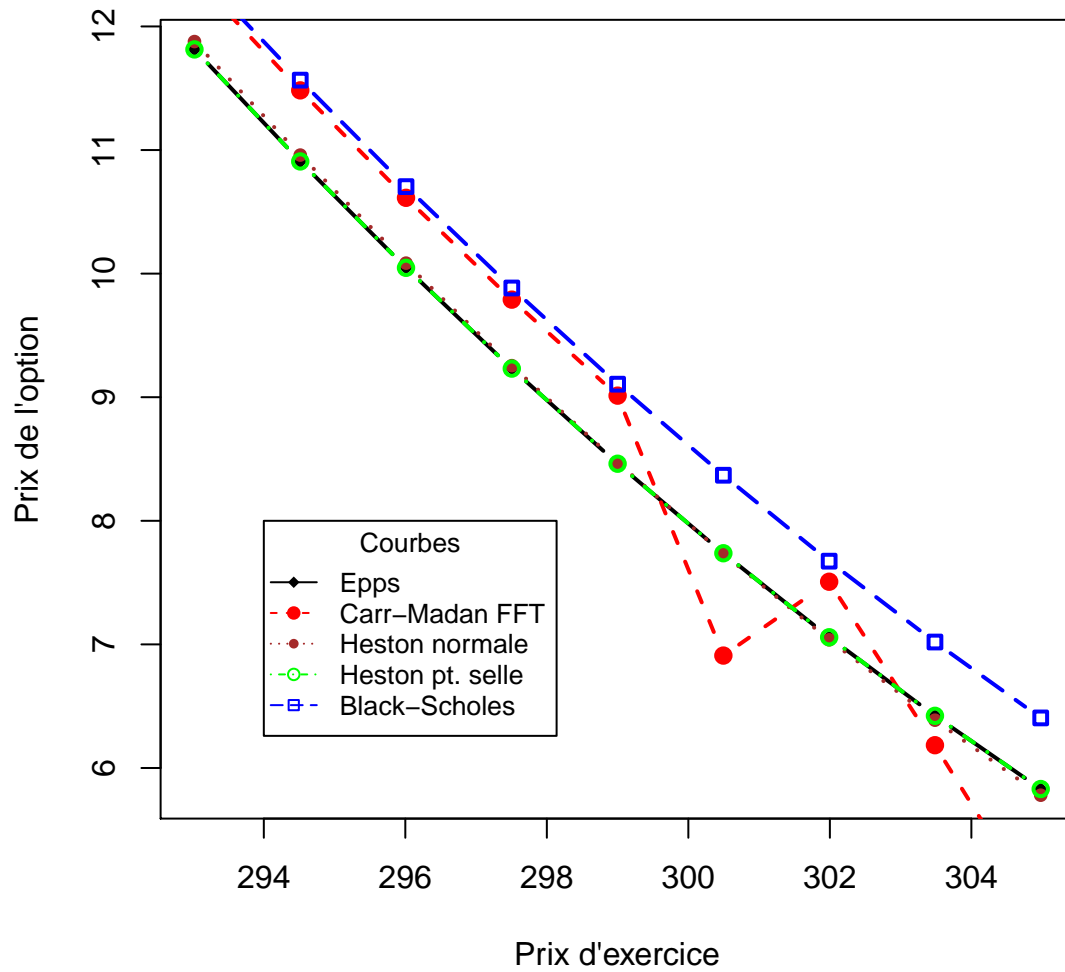


FIGURE 9.8: Prix de l'option selon les paramètres estimés avec la méthode de l'équation d'estimation optimale



# Conclusion

En guise de conclusion, j'aimerais tout d'abord effectuer un retour sur les différents éléments introduits au début du premier chapitre, concernant le risque de modélisation. Le modèle présenté n'en est pas exempt, bien au contraire. Cependant, il nécessite moins d'hypothèses restrictives que les autres modèles présentés pour être valide, bien qu'il exige toujours l'indépendance des observations. Il tient compte de la possibilité de sauts tout en conservant une composante de mouvement aléatoire, ce qui décrit adéquatement les observations empiriques à ce jour. Comme tout modèle paramétrique, il reste dépendant du nombre et de la qualité des données disponibles. Étant donné que l'utilisation d'algorithmes d'optimisation numérique est inévitable, il subsiste un risque important autour de l'estimation des paramètres et de l'approximation de la distribution. De plus, étant donné qu'il n'existe pas de mesure neutre au risque unique, l'arbitrage de modèle reste possible et doit être considéré. L'utilisation d'un échantillon de données instables à travers le temps peut produire des résultats inattendus, surtout au niveau de la distribution de la volatilité historique, un aspect qui pourra être approfondi ultérieurement. Enfin, subsiste toujours le risque d'erreurs de nature informatique qui pourraient produire de faux résultats.

Ce retour permet de constater qu'il y a toujours place à l'amélioration des outils développés. Entre autres, il pourrait être pertinent d'étudier les différences entre le comportement à court et à long terme du modèle. En se basant sur la théorie de l'utilité, on pourrait développer une meilleure approche pour déterminer les paramètres de la distribution neutre au risque. Il pourrait aussi être intéressant de développer des mesures de risque cohérentes pour les processus de Lévy, notamment avec les avancées de celles basées sur l'entropie. L'extension multivariée de ce modèle n'a toujours pas été développée dans la littérature, alors il pourrait être pertinent de s'y attarder, entre autres pour étudier les titres indicels et optimiser la composition de portefeuilles. Enfin, il pourrait être intéressant d'aborder le problème inverse de l'estimation des paramètres à partir des prix des produits dérivés observés sur les marchés financiers.



# Annexe A

## Éléments de théorie des probabilités

### A.1 Définitions de base

Une **variable aléatoire** est une variable dont la valeur est déterminée par une expérience aléatoire. On associe une probabilité à chacune des valeurs possibles. On appelle **évènement** tout ensemble de réalisations possible de l'expérience aléatoire. L'entièreté de tous les évènements se nomme l'**ensemble fondamental** et est notée  $\Omega$ .

Une **probabilité** associée à un évènement  $A$  doit satisfaire un ensemble de trois axiomes (Dodge, 2004).

1. Une probabilité est une valeur comprise entre 0 et 1 :

$$0 \leq P(A) \leq 1. \quad (\text{A.1.1})$$

2. La probabilité de l'évènement correspondant à l'ensemble fondamental est 1 :

$$P(\Omega) = 1. \quad (\text{A.1.2})$$

3. Pour chaque séquence d'évènements mutuellement exclusifs  $A_1, A_2, \dots$  :

$$P \left[ \bigcup_{i=1}^{\infty} A_i \right] = \sum_{i=1}^{\infty} P(A_i). \quad (\text{A.1.3})$$

Le respect de ces conditions devra notamment être vérifié lors de l'utilisation de méthodes numériques.

Toujours selon Dodge (2004), on définit :

1. La **densité**, notée  $f_X(x)$ , permet de déterminer la probabilité qu'une variable aléatoire  $X$  prenne une valeur dans un intervalle fixé  $[a, b]$ . En intégrant cette fonction sur l'intervalle  $[a, b]$ , on obtient cette probabilité :

$$P(a \leq X \leq b) = \int_a^b f_X(x) dx. \quad (\text{A.1.4})$$

2. La **fonction de répartition**, notée  $F_X(x)$ , d'une variable aléatoire est définie comme la probabilité que celle-ci prenne une valeur inférieure ou égale à un certain nombre  $b \in \mathbb{R}$  :

$$F_X(b) = P(X \leq b) = \int_{-\infty}^b f_X(x) dx. \quad (\text{A.1.5})$$

## A.2 Transformées d'une variable aléatoire

Les transformées d'une variable aléatoire résultent de l'application d'un opérateur intégral, ou à noyau, sur la fonction de densité.

Certaines d'entre elles permettent de déterminer entièrement leur distribution. Parmi celles-ci, on retrouve la fonction caractéristique et les fonctions génératrice des moments et des cumulants, qui sont les plus couramment utilisées.

Certaines transformées permettent de modifier la distribution d'une variable aléatoire. Parmi celles-ci, la transformée d'Esscher, qui est employée en actuariat et en finance, est aussi étroitement liée au concept d'utilité en sciences économiques.

### A.2.1 La fonction caractéristique

#### Transformée de Fourier

La transformée de Fourier est une opération qui transforme une fonction  $g(\xi)$  en une autre  $f(x)$  par l'intégration, sur son domaine, du produit  $e^{-ix\xi} g(\xi)$  :

$$\mathcal{F}(g) : x \mapsto f(x) = \int_{-\infty}^{\infty} e^{-ix\xi} g(\xi) d\xi. \quad (\text{A.2.1})$$

On définit aussi la transformée de Fourier inverse, qui permet de retrouver la fonction initiale  $g(\xi)$  à partir de la transformée  $f(x)$  :

$$g(\xi) = \mathcal{F}^{-1}(f(x)) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ix\xi} f(x) dx. \quad (\text{A.2.2})$$

#### Définition

La fonction caractéristique est définie comme étant la transformée de Fourier inverse de la densité, dans le cas continu, ou de masse de probabilité, dans le cas discret. C'est donc une application directe de (A.2.2) qui s'exprime par l'intégrale de Riemann-Stieltjes :

$$\phi_X(s) = \int_{-\infty}^{\infty} e^{isx} dF_X(x). \quad (\text{A.2.3})$$

Cette intégrale est toujours convergente, comme le démontre [Stuart et Ord \(1987\)](#).

## Les moments

On définit le moment d'ordre  $r$  d'une distribution de probabilités comme étant la quantité représentée par l'espérance de la puissance  $r$  d'une variable aléatoire :

$$E[X^r] = \int_{-\infty}^{\infty} x^r dF_X(x). \quad (\text{A.2.4})$$

La fonction caractéristique, lorsqu'elle est différentiable, permettra de générer les moments de la distribution en utilisant la propriété suivante (Lukacs, 1960) :

$$\frac{d^r \phi_X(s)}{ds^r} = i^r \int_{-\infty}^{\infty} e^{isx} x^r dF_X(x).$$

En posant  $s = 0$  par la suite, on obtient, pour les différentes dérivées, l'équation suivante :

$$\left[ \frac{d^r \phi_X(s)}{ds^r} \right]_{s=0} = i^r E[X^r].$$

Ce qui nous permet de définir les différents moments de la distribution à partir de la fonction caractéristique :

$$E[X^r] = (-i)^r \left[ \frac{d^r \phi_X(s)}{ds^r} \right]_{s=0}. \quad (\text{A.2.5})$$

### A.2.2 Inversion de la fonction caractéristique

#### La densité

On obtient la densité de la variable aléatoire  $X$  en calculant la transformée de Fourier (A.2.1) de la fonction caractéristique suivante :

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isx} \phi_X(s) ds. \quad (\text{A.2.6})$$

On peut utiliser l'intégration numérique si l'intégrale n'a pas de solution analytique, la méthode de la transformée de Fourier rapide (section A.3) ou encore la méthode du point de selle.

#### La fonction de répartition

On peut obtenir directement l'expression de la fonction de répartition, sans passer par l'intégration de la densité de probabilité, en utilisant le théorème de Gil-Pelaez (1951) :

$$F_X(x) = \frac{1}{2} + \frac{1}{2\pi} \int_0^{\infty} \frac{e^{isx} \phi_X(-s) - e^{-isx} \phi_X(s)}{is} ds. \quad (\text{A.2.7})$$

On peut exprimer cette intégrale sous la forme suivante lorsque la densité  $f(x)$  est strictement continue (Epps, 2007, p.66) :

$$F_X(x) = \frac{1}{2} - \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-isx}}{is} \phi(s) ds. \quad (\text{A.2.8})$$

Cette forme est moins appropriée pour l'intégration numérique, mais sera utile dans plusieurs calculs. On préférera la ramener à la forme suivante avec le théorème 2 de Wendel (1961) :

$$F_X(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\text{Im} [e^{-isx} \phi_X(s)]}{s} ds. \quad (\text{A.2.9})$$

Cette fonction peut être difficile à intégrer de manière efficace numériquement, surtout lorsqu'on l'évalue en des points situés aux extrémités de la distribution. On peut alors privilégier l'égalité donnée par le théorème 4 de Shephard (1991), qui utilise le noyau de Fejér afin de réduire l'erreur d'intégration, pour obtenir le résultat suivant :

$$F_X(x) = \frac{1}{2} - \frac{1}{2\pi} \lim_{n \rightarrow \infty} \int_0^n \underbrace{\left[1 - \frac{s}{n}\right]}_{\text{Noyau de Fejér}} \left[ \frac{e^{isx} \phi_X(-s) - e^{-isx} \phi_X(s)}{is} \right] ds. \quad (\text{A.2.10})$$

De même, en utilisant le raisonnement qui permet de passer de la définition de la transformée (A.2.7) à celle de son inverse (A.2.9), on obtient, pour le résultat (A.2.10), l'expression suivante :

$$F_X(x) = \frac{1}{2} - \frac{1}{\pi} \lim_{n \rightarrow \infty} \int_0^n \left[1 - \frac{s}{n}\right] \frac{\text{Im} [e^{-isx} \phi_X(s)]}{s} ds. \quad (\text{A.2.11})$$

On peut alors obtenir une approximation en fixant la borne d'intégration supérieure  $n$  qui définit la précision désirée. Certains logiciels d'intégration numérique prennent en charge les bornes infinies.

La fonction caractéristique permet d'identifier la distribution d'une somme de variables aléatoires indépendantes  $Z = X_1, \dots, X_n$ , appelée produit de convolution et noté  $Z = X_1 * \dots * X_n$  :

$$\phi_Z(s) = \phi_{X_1 + \dots + X_n}(s) = \prod_{i=1}^n \phi_{X_i}(s). \quad (\text{A.2.12})$$

Lorsque les variables aléatoires sommées sont aussi identiquement distribuées, la fonction caractéristique de  $Z$  est la  $n^e$  puissance de celle de  $X$  :

$$\phi_Z(s) = \phi_{X_1 + \dots + X_n}(s) = [\phi_{X_i}(s)]^n. \quad (\text{A.2.13})$$

Cette fonction est donc une solution de rechange intéressante à utiliser lorsque aucune forme analytique pour la densité ou la fonction de répartition pour une distribution donnée n'existe.

### A.2.3 La fonction génératrice des moments

La fonction génératrice des moments  $M_X(s)$  est définie comme étant la transformée de Laplace inverse de la densité :

$$M_X(s) = \int_{-\infty}^{\infty} e^{sx} dF_X(x). \quad (\text{A.2.14})$$



Cette intégrale, contrairement à (A.2.3), ne converge pas toujours, ce qui signifie que certaines distributions n'ont pas de fonction génératrice des moments. Comme son nom l'indique, cette fonction permet de générer les moments de la distribution de la variable  $X$ , ce qui se fait essentiellement de la même manière qu'avec la définition (A.2.5). Par contre, on n'a pas à éliminer le terme complexe, ce qui a pour avantage de simplifier les calculs :

$$E[X^r] = \left[ \frac{d^r M_X(s)}{ds^r} \right]_{s=0}. \quad (\text{A.2.15})$$

La fonction génératrice des moments peut aussi être obtenue à partir de la fonction caractéristique :

$$M_X(s) = \phi(-is). \quad (\text{A.2.16})$$

Ceci permet d'en déduire qu'elles possèdent des caractéristiques communes, notamment le produit de convolution.

#### A.2.4 La fonction génératrice des cumulants

La fonction génératrice des cumulants  $K_X(s)$  est définie comme le logarithme de la fonction génératrice des moments :

$$K_X(s) = \ln M_X(s). \quad (\text{A.2.17})$$

Cette fonction est utilisée pour générer les cumulants, des quantités étroitement liées aux moments, qui peuvent à leur tour être utilisés dans le cadre de méthodes d'estimation paramétrique. Parmi celles-ci, on retrouve la méthode des cumulants, dont l'objectif est de former un système d'équations où les valeurs empiriques et théoriques de ces quantités sont égales. Les valeurs empiriques peuvent être obtenues à partir des moments échantillonnaires (Stuart et Ord, 1987) :

$$\hat{K}_1 = \hat{m}_1 \quad (\text{A.2.18a})$$

$$\hat{K}_2 = \hat{m}'_2 - \hat{m}_1^2 \quad (\text{A.2.18b})$$

$$\hat{K}_3 = \hat{m}'_3 - 3\hat{m}_1\hat{m}'_2 + 2\hat{m}_1^3 \quad (\text{A.2.18c})$$

$$\hat{K}_4 = \hat{m}'_4 - 3\hat{m}_1^2\hat{m}'_2 - 4\hat{m}_1\hat{m}'_3 + 12\hat{m}_1^2\hat{m}'_2 - 6\hat{m}_1^4. \quad (\text{A.2.18d})$$

On remarquera aussi la forme de la première dérivée de cette dernière, qui sera utilisée pour estimer la densité et la fonction de répartition avec la méthode du point de selle :

$$\frac{dK_X(s)}{ds} = K'_X(s) = \frac{M'_X(s)}{M_X(s)}. \quad (\text{A.2.19})$$

### A.2.5 La transformée d'Esscher

La transformée d'Esscher transforme la densité en une autre en utilisant un coefficient exponentiel :

$$f(x; h) = \frac{e^{hx} f(x)}{\int_{-\infty}^{\infty} e^{hx} f(x) dx}. \quad (\text{A.2.20})$$

Lorsque l'on ne dispose pas d'une forme analytique de la densité, on peut exprimer la transformée d'Esscher sous la forme de sa fonction génératrice des moments :

$$M_X(s; h) = \frac{M_X(s + h)}{M_X(h)}. \quad (\text{A.2.21})$$

## A.3 La transformée de Fourier rapide

Étant donné l'importance de la transformée de Fourier et de son inverse en sciences physiques, plusieurs algorithmes ont été développés afin d'en effectuer l'intégration numérique. Ces algorithmes se regroupent sous l'appellation de transformée de Fourier rapide.

L'existence de ces algorithmes favorise l'utilisation de la fonction caractéristique et de ses propriétés, notamment pour l'agrégation de risques en actuariat, ou encore le calcul de probabilités pour des distributions n'ayant pas de forme explicite pour la fonction de répartition.

Ces algorithmes numériques sont utilisés respectivement pour calculer de manière optimisée des sommes de la forme

$$\hat{f}(x_u) = \sum_{j=1}^N \phi(\zeta_j) e^{-\frac{2\pi i}{N}(j-1)(u-1)} \quad u = 1, \dots, n. \quad (\text{A.3.1})$$

$$\hat{\phi}(\zeta_j) = \sum_{u=1}^N f(x_u) e^{\frac{2\pi i}{N}(j-1)(u-1)} \quad j = 1, \dots, n. \quad (\text{A.3.2})$$

Par exemple, pour évaluer l'intégrale (A.2.6) afin d'obtenir la valeur de la densité sur un certain domaine  $[a, b]$ , on devra discrétiser celle-ci sur nombre  $N$  de points de discrétisation. On définit alors les différents paramètres de ces deux algorithmes :

- $\eta = \frac{b-a}{N}$ , le pas de discrétisation pour la variable aléatoire  $u$
- $\zeta = \eta(j-1)$ , la variable de transformation
- $\lambda = \frac{2\pi}{N\eta}$ , le pas de discrétisation pour la variable de transformation  $\zeta$
- $c = \frac{N\lambda}{2}$ , la borne supérieure d'intégration

On obtient alors une sommation de la forme (A.3.1) :

$$f_{FFT}(x_u) \approx \frac{1}{2\pi} \sum_{j=1}^N e^{-i\lambda\zeta_j(u-1)} e^{-ic\zeta_j} \phi(-c + \lambda(j-1))\eta, \quad (\text{A.3.3a})$$

$$\approx \frac{1}{2\pi} \sum_{j=1}^N e^{-i\frac{2\pi}{N}(j-1)(u-1)} e^{-ic\zeta_j} \phi(-c + \lambda(j-1))\eta. \quad (\text{A.3.3b})$$

Plusieurs logiciels d'analyse numérique possèdent une implémentation de la méthode de la transformée de Fourier rapide. Essentiellement, c'est une fonction  $f_{FFT}(X_u)$  qui prend comme argument un vecteur de valeurs, appelé signal, et qui en retourne un autre, de même longueur, le spectre de fréquences. En statistique, la fonction caractéristique empirique constitue le signal et la densité, le spectre de fréquences. Le signal  $X_u$  est défini comme suit, à partir des constantes définies précédemment :

$$X_u = e^{-i\lambda\zeta_j(u-1)} \phi(-c + (j-1)\lambda) \eta. \quad (\text{A.3.4})$$

La somme (A.3.3a) devient alors

$$f_{FFT}(X_u) \approx \frac{1}{2\pi} \sum_{j=1}^N X_u e^{-ic\zeta_j}. \quad (\text{A.3.5})$$

On recouvre la densité aux points  $x_u = a + (j-1)\eta, j = 1, \dots, N$  en appliquant la transformation suivante au vecteur  $f_{FFT}(X_u)$  :

$$f(x_u) = \frac{1}{N} e^{icx_u} f_{FFT}(X_u). \quad (\text{A.3.6})$$

## A.4 Processus de Lévy

### A.4.1 Définition et propriétés

Le processus de Lévy, tel que présenté par O.E.E. Barndorff-Nielsen et Resnick (2001), est une classe générale de processus stochastiques regroupant entre autres les processus de Poisson composés et les processus de Wiener. Il est défini comme continu à droite, limité à gauche (càdlàg), ayant pour point de départ l'origine et ayant des incréments indépendants et homogènes. Il est infiniment divisible, tout processus de Lévy peut ainsi être considéré comme une convolution de plusieurs autres. Cette propriété est très intéressante dans un contexte de rendements financiers, car la même distribution pourra être utilisée avec n'importe quel intervalle d'échantillonnage des prix.

## Représentation de Lévy-Khintchine

Le processus de Lévy est généralement représenté par sa fonction caractéristique sous la forme de Lévy-Khintchine :

$$\mathbb{E}\left[e^{i\theta X(t)}\right] = \exp\left(\underbrace{ait\theta}_{\text{composante de dérive}} - \underbrace{\frac{1}{2}\sigma^2 t\theta^2}_{\text{composante de diffusion}} + t \underbrace{\int_{\mathbb{R}\setminus\{0\}} (e^{i\theta x} - 1 - i\theta x \mathbf{I}_{|x|<1}) \nu(dx)}_{\text{composante de saut}}\right) \quad (\text{A.4.1})$$

$$= \exp\left(-t\Psi(\theta)\right). \quad (\text{A.4.2})$$

$\Psi(\theta)$  est appelé l'exposant caractéristique de l'incrément de longueur 1 ( $X(t+1) - X(t)$ ). Un processus de Lévy est souvent décrit par son triplet générateur  $(a, \sigma^2, \nu)$ . Cette description permettra de classer certains processus de Lévy selon deux catégories :

- Si  $\sigma^2 = 0$ ,  $\{X(t)\}$  est un processus de sauts
- Si  $a = 0$  et  $\sigma^2 = 0$ , alors le processus  $\{X(t)\}$  est un processus de sauts purs

L'élément  $\nu(dx)$  de la composante de saut est appelé la mesure de Lévy. Un exemple de processus de Lévy représenté sous la forme de Lévy-Khintchine qui ne fait pas partie de ces deux catégories est défini par le modèle de Press (1.3.12).

## Représentation de Lévy-Itô

La décomposition de Lévy-Itô est une représentation alternative décrivant la trajectoire d'une réalisation du processus. Cette dernière a une interprétation intéressante en finance décrite par Applebaum (2004) :

$$\begin{aligned} X(t) &= \underbrace{at + B_{\sigma^2}(t)}_{\text{mouvement brownien}} \\ &+ \underbrace{\int_{|x|<1} xN(t, dx) - t\nu(dx)}_{\text{martingale de sauts purs de carré intégrable}} \\ &+ \underbrace{\int_{|x|>1} xN(t, dx)}_{\text{processus de Poisson composé}} \end{aligned} \quad (\text{A.4.3})$$

$$= X_1(t) + X_2(t) + X_3(t). \quad (\text{A.4.4})$$

La portion mouvement brownien  $(X_1(t))$  décrit le comportement général du processus, en spécifiant le rendement espéré  $a$  et la volatilité du titre  $\sigma^2$ . La première intégrale  $(X_2(t))$  décrit le bruit causé par les transactions financières quotidiennes qui font varier un peu le prix, alors que la seconde  $(X_3(t))$  introduit les sauts occasionnés par des événements plus rares, comme les catastrophes naturelles et les crises politiques. Dans les deux cas,  $N(t, dx)$  est une mesure aléatoire de Poisson.

#### A.4.2 Processus subordonné

On considère les processus de Lévy  $\{X(t)\}$  et  $\{Z(t)\}$ . Celui qui suit est défini comme étant un processus subordonné et aussi un processus de Lévy, comme le démontre Sato (1999) et Schoutens (2003) :

$$\{Y(t)\} = \{X(Z(t))\}. \quad (\text{A.4.5})$$

Les processus de Lévy les plus couramment utilisés en finance sont des mouvements browniens subordonnés. Ils sont plus faciles à manipuler et permettent néanmoins de représenter les phénomènes de queues longues présents dans les distributions de rendements. Une bonne introduction à ce sujet est faite par Kyprianou (2007).

Si  $\Lambda(\theta)$  est l'exposant caractéristique du processus  $\{X(t)\}$  et  $\Xi(\theta)$ , celui du processus subordonné  $\{Z(t)\}$ , alors celui du processus  $\{Y(t)\}$  prend la forme suivante :

$$\begin{aligned} \Psi(\theta) &= \Xi(\theta) \circ i\Lambda(\theta) \\ &= \Xi(i\Lambda(\theta)). \end{aligned} \quad (\text{A.4.6})$$

La fonction caractéristique du processus  $\{Y(t)\}$  est alors

$$\phi_{Y(t)}(\xi) = e^{-t\Xi(i\Lambda(\xi))}. \quad (\text{A.4.7})$$

La densité de la variable aléatoire  $Y(t)$  s'obtient à l'aide de la formule de l'espérance conditionnelle :

$$\begin{aligned} f_{Y(t)}(y) &= E [f_{X(t)}(y|Z(t) = z)] \\ &= \int_{-\infty}^{\infty} f_{X(t)}(y|Z(t) = z) \cdot f_{Z(t)}(z) dz. \end{aligned} \quad (\text{A.4.8})$$

### A.5 Théorèmes d'intégration

Ces deux théorèmes sont présentés sans démonstration afin de compléter la démonstration de la méthode d'Épps de la section 8.3.5. La démonstration de ceux-ci nécessite des connaissances en théorie de l'intégration. On peut retrouver davantage d'explications sur ceux-ci dans Teschl (2004).

### A.5.1 Théorème de convergence dominée de Lebesgue

Le théorème de convergence dominée de Lebesgue est un des principaux éléments de la théorie de l'intégration. Il sert à démontrer qu'une fonction est intégrable, sachant qu'une autre l'est aussi et qu'elle répond à certaines conditions.

Soit  $(f_n)_{n \in \mathbb{N}}$ , une suite de fonctions réelles ou complexes intégrables dans un intervalle  $I$  qui convergent vers une fonction  $f$  :

$$f(x) = \lim_{n \rightarrow \infty} f_n(x). \quad (\text{A.5.1})$$

On suppose qu'une fonction  $g$  intégrable dans l'intervalle  $I$  existe telle que pour toute valeur de  $n$  et pour tout point  $x \in I$  où elle est définie, la valeur absolue de  $f_n(x)$  est inférieure ou égale à  $g(x)$  :

$$\forall n \in \mathbb{N}, \forall x \in I, |f_n(x)| \leq g(x). \quad (\text{A.5.2})$$

Alors, la fonction  $f$  est intégrable.

### A.5.2 Théorème de Fubini

Le théorème de Fubini permet, entre autres, d'inverser l'ordre d'intégration lorsque certaines conditions sont remplies.

Soit une fonction  $f$  continue sur le rectangle  $R$  suivant :

$$R = \{(x, y) | a \leq x \leq b, c \leq y \leq d\}. \quad (\text{A.5.3})$$

on peut inverser l'ordre d'intégration :

$$\int_{X \times Y} f(x, y) d(\mu \times \nu)(x, y) = \int_X \left[ \int_Y f(x, y) d\nu(y) \right] d\mu(x) = \int_Y \left[ \int_X f(x, y) d\mu(x) \right] d\nu(y). \quad (\text{A.5.4})$$

Il est possible de généraliser cet énoncé en utilisant la théorie de la mesure.

## Annexe B

# Éléments de statistique mathématique

### B.1 Loi faible des grands nombres

La **loi faible des grands nombres** est un résultat important en probabilité, car il permet de définir la notion d'estimateur convergent. Soit une suite de variables aléatoires indépendantes et identiquement distribuées  $\{X_T\}_{T=1}^{\infty}$  ayant une espérance  $E[X]$  et une variance  $V[X]$  finies. Selon la loi faible des grands nombres, pour tout nombre réel strictement positif  $\varepsilon$ , la probabilité que la différence entre la moyenne empirique  $Y_T = \frac{X_1 + X_2 + \dots + X_T}{T}$  et l'espérance  $E[X]$  soit supérieure à la valeur  $\varepsilon$  tend vers 0 lorsque  $T$  tend vers l'infini.

$$\lim_{T \rightarrow +\infty} \mathbb{P}(|Y_T - E[X]| \geq \varepsilon) = 0, \quad \forall \varepsilon > 0 \quad (\text{B.1.1})$$

On dit alors que la suite d'estimateurs  $\{Y_T\}_{T=1}^{\infty}$  converge en probabilité vers l'espérance  $E[X]$ . L'estimateur de l'espérance  $Y_T$  est alors **convergent**.

### B.2 Théorème central limite

Le **théorème central limite** est un résultat fondamental en probabilité qui énonce le rôle de la distribution normale. Il démontre que toute somme de variables aléatoires indépendantes et identiquement distribuées suit approximativement une loi normale. Ce résultat permet entre autres d'identifier la distribution limite d'un estimateur convergent.

### B.2.1 Cas univarié

Soit une suite d'observations  $X_1, \dots, X_T$  d'un échantillon aléatoire d'une distribution de moyenne  $\mu$  et de variance  $\sigma^2$  :

$$\begin{aligned} Y_T &= \frac{1}{\sqrt{T}\sigma} \left( \sum_{t=1}^T X_t - T\mu \right) \\ &= \frac{\sqrt{T}}{\sigma} (\bar{X}_T - \mu). \end{aligned} \quad (\text{B.2.1})$$

Alors, cette variable aléatoire converge en distribution vers une variable aléatoire normale centrée réduite :

$$Y_T \xrightarrow{L} \mathcal{N}(0, 1). \quad (\text{B.2.2})$$

### B.2.2 Cas multivarié

On peut aussi généraliser ce théorème pour des observations multivariées. On considère alors une série d'observations multivariées  $\mathbf{X}_1, \dots, \mathbf{X}_T$  où

$$\mathbf{X}_t = \begin{bmatrix} X_{t(1)} \\ \vdots \\ X_{t(k)} \end{bmatrix}, \quad t = 1, \dots, T. \quad (\text{B.2.3})$$

On définit maintenant la variable aléatoire correspondante :

$$\mathbf{Y}_T = \frac{1}{T} \begin{bmatrix} \sum_{t=1}^T [X_{t(1)}] \\ \vdots \\ \sum_{t=1}^T [X_{t(k)}] \end{bmatrix}. \quad (\text{B.2.4})$$

Dans cette situation, la variable aléatoire converge en distribution vers une variable aléatoire de distribution normale multivariée centrée de matrice de variance-covariance  $\Sigma$  :

$$\sqrt{T} (\mathbf{Y}_T - \mu) \xrightarrow{L} \mathcal{N}_k(0, \Sigma) \quad (\text{B.2.5})$$

où

$$\Sigma = \begin{bmatrix} \omega_{(1,1)} & \cdots & \omega_{(1,k)} \\ \vdots & \ddots & \vdots \\ \omega_{(k,1)} & \cdots & \omega_{(k,k)} \end{bmatrix}, \quad \text{avec } \omega_{(j,k)} = \begin{cases} \text{Var} [X_{1(j)}], & (j = k) \\ \text{Cov} [X_{1(j)}, X_{1(k)}], & (j \neq k). \end{cases}$$



### B.3 Méthode delta multivariée

Dans le cas univarié, on utilise la méthode delta pour évaluer la distribution d'une fonction d'un estimateur, en supposant que la distribution de cet estimateur est asymptotiquement normale de variance connue.

Dans le cas multivarié, on estime la distribution d'une fonction d'un vecteur d'estimateurs, dont la distribution asymptotique est normale multivariée, avec une matrice de variance-covariance  $\Sigma$ .

On a donc, pour un estimateur convergent  $\hat{\theta}_T$ , en appliquant le théorème central limite, le résultat suivant :

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{L} \mathcal{N}(0, \Sigma). \quad (\text{B.3.1})$$

On cherche la distribution d'une fonction  $h(\hat{\theta}_T)$ . On développe cette fonction sous la forme d'une série de Taylor et en conservant seulement les deux premiers termes :

$$h(\hat{\theta}_T) \approx h(\theta_0) + \nabla [h(\theta_0)]' \cdot (\hat{\theta}_T - \theta_0). \quad (\text{B.3.2})$$

On a donc, après quelques manipulations, que la variance de la fonction  $h(\hat{\theta}_T)$  est approximativement

$$\text{Var} [h(\hat{\theta}_T)] \approx \nabla [h(\theta_0)]' \left( \frac{\Sigma}{T} \right) \nabla [h(\theta_0)]. \quad (\text{B.3.3})$$

La distribution de la fonction  $h(\hat{\theta}_T)$  est alors asymptotiquement

$$h(\hat{\theta}_T) \xrightarrow{L} N \left( h(\theta_0), \nabla [h(\theta_0)]' \left( \frac{\Sigma}{T} \right) \nabla [h(\theta_0)] \right). \quad (\text{B.3.4})$$



# Annexe C

## Données

Ce tableau présente l'échantillon de données utilisées au chapitre 9.

TABLE C.1: Prix du titre Abbey National (penny sterling) du 31 juillet au 8 octobre 1991

<b>Date</b>	<b>Prix</b>	<b>Date</b>	<b>Prix</b>	<b>Date</b>	<b>Prix</b>	<b>Date</b>	<b>Prix</b>	<b>Date</b>	<b>Prix</b>
31/07	296	14/08	294	28/08	303	11/09	296	25/09	306
01/08	296	15/08	293	29/08	304	12/09	301	26/09	303
02/08	300	16/08	295	30/08	304	13/09	298	27/09	301
05/08	302	19/08	287	02/09	309	16/09	295	30/09	303
06/08	300	20/08	288	03/09	309	17/09	295	01/10	308
07/08	304	21/08	297	04/09	309	18/09	293	02/10	305
08/08	303	22/08	305	05/09	307	19/09	292	03/10	302
09/08	299	23/08	307	06/09	306	20/09	297	04/10	301
12/08	293	26/08	307	09/09	304	23/09	294	07/10	297
13/08	294	27/08	304	10/09	300	24/09	293	08/10	299



# Bibliographie

- Milton Abramowitz et Irene A. Stegun. *Handbook Of Mathematical Functions : With Formulas, Graphs, And Mathematical Tables*, volume 55. Dover Publications, 1965.
- David Applebaum. Lévy processes : From probability to finance and quantum groups. *Notices Of The American Mathematical Society*, 51(11) :1336–1347, 2004.
- Louis Bachelier. *Théorie De La Spéculation*. Gauthier-Villars, 1900.
- Joseph Berkson. Minimum chi-square, not maximum likelihood! *The Annals of Statistics*, pages 457–487, 1980.
- Nicholas H. Bingham et Rüdiger Kiesel. *Risk-Neutral Valuation : Pricing And Hedging Of Financial Derivatives*. Springer, 2004.
- Fischer Black. The pricing of commodity contracts. *Journal Of Financial Economics*, 3(1) : 167–179, 1976.
- Fischer Black et Myron Scholes. The pricing of options and corporate liabilities. *The Journal Of Political Economy*, pages 637–654, 1973.
- D. J. Buckle. Bayesian inference for stable distributions. *Journal of the American Statistical Association*, 90(430) :605–613, 1995.
- Ronald W. Butler. *Saddlepoint Approximations With Applications*, volume 22. Cambridge University Press, 2007.
- Peter Carr et Dilip Madan. Option valuation using the fast fourier transform. *Journal Of Computational Finance*, 2(4) :61–73, 1999.
- Martin Crowder. On consistency and inconsistency of estimating equations. *Econometric Theory*, pages 305–330, 1986.
- Martin Crowder. On linear and quadratic estimating functions. *Biometrika*, 74(3) :591–597, 1987.

- Henry E Daniels. Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics*, pages 631–650, 1954.
- Emanuel Derman. Model risk. Technical report, Goldman Sachs, 1996.
- Peter P Carr Dilip B Madan et Eric C Chang. The variance gamma process and option pricing. *European Finance Review*, 2(1) :79–105, 1998.
- Yadolah Dodge. *Statistique : Dictionnaire Encyclopédique*. Springer Verlag France, 2004.
- Thomas W Epps. *Pricing Derivative Securities*. World Scientific Publishing Company Incorporated, 2007.
- Thomas W Epps et Lawrence B Pulley. A test for normality based on the empirical characteristic function. *Biometrika*, 70(3) :723–726, 1983.
- Brian Everitt et Anders Skrondal. *The Cambridge Dictionary Of Statistics*, volume 4. Cambridge University Press Cambridge, 2006.
- Eugene F Fama et Kenneth R French. Common risk factors in the returns on stocks and bonds. *Journal Of Financial Economics*, 33(1) :3–56, 1993.
- Andrey Feuerverger et Philip McDunnough. On the efficiency of empirical characteristic function procedures. *Journal Of The Royal Statistical Society. Series B (methodological)*, pages 20–27, 1981.
- Robert Fox et Murad S Taqqu. Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series. *The Annals of Statistics*, 14(2) :517–532, 1986.
- J Gil-Pelaez. Note on the inversion theorem. *Biometrika*, 38(3-4) :481–482, 1951.
- Christian Gourieroux et Alain Monfort. *Statistique Et Modèles Économétriques : Notions Générales, Estimation, Préviation, Algorithmes*, volume 1. Economica, 1989.
- Alastair R Hall. *Generalized Method Of Moments*. Oxford University Press Oxford, 2005.
- James Douglas Hamilton. *Time Series Analysis*, volume 2. Cambridge University Press, 1994.
- Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica : Journal Of The Econometric Society*, pages 1029–1054, 1982.
- N Henze. An approximation to the limit distribution of the epps-pulley test statistic for normality. *Metrika*, 37(1) :7–18, 1990.
- Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review Of Financial Studies*, 6(2) :327–343, 1993.

- David V Hinkley et Nagesh S Revankar. Estimation of the pareto law from underreported data : A further analysis. *Journal Of Econometrics*, 5(1) :1–11, 1977.
- R.V. Hogg et A.T. Craig. *Introduction To Mathematical Statistics*. Macmillan, 1978. ISBN 9780029789902.
- Andrey Itkin. Pricing options with vg model using fft. *arXiv preprint physics/0503137*, 2005.
- Wai Wan Tsang Jingbo Wang et George Marsaglia. Evaluating kolmogorov’s distribution. *Journal of Statistical Software*, 8(18) :1–4, 2003.
- S. Kotz, T.J. Kozubowski, et K. Podgórski. *The Laplace Distribution And Generalizations : A Revisit With Applications To Communications, Exonomics, Engineering, And Finance*. Progress in Mathematics Series. Birkhäuser, 2001. ISBN 9780817641665.
- I. A. Koutrouvelis. A Goodness-Of-Fit Test Of Simple Hypotheses Based On The Empirical Characteristic Function. *Biometrika*, 67(1) :238–240, 1980. doi : 10.1093/biomet/67.1.238.
- Tomasz J Kozubowski et Krzysztof Podgórski. A class of asymmetric distributions. *Actuarial Research Clearing House*, 1 :113–134, 1999.
- Tomasz J Kozubowski et Krzysztof Podgórski. Asymmetric laplace laws and modeling financial data. *Mathematical And Computer Modeling*, 34(9) :1003–1021, 2001.
- Andreas E Kyprianou. *Introductory Lectures On Fluctuations Of LéVy Processes With Applications*. Springer, 2007.
- Robert Lugannani et Stephen Rice. Saddle point approximation for the distribution of the sum of independent random variables. *Advances In Applied Probability*, pages 475–490, 1980.
- Eugene Lukacs. *Characteristic Functions*, volume 4. Griffin London, 1960.
- Andrew Luong et Mary E. Thompson. Minimum-distance methods based on quadratic distances for transforms. *Canadian Journal of Statistics*, 15(3) :239–251, 1987.
- Dilip B Madan et Eugene Seneta. The variance gamma model for share market returns. *Journal Of Business*, pages 511–524, 1990.
- Benoit Mandelbrot. The variation of certain speculative prices. *Journal Of Business*, pages 394–419, 1963.
- Robert C Merton. Option pricing when underlying stock returns are discontinuous. *Journal Of Financial Economics*, 3(1) :125–144, 1976.
- Wesley C Mitchell. A critique of index numbers of the prices of stocks. *The Journal Of Political Economy*, pages 625–693, 1916.

- Whitney K Newey et Daniel McFadden. Large sample estimation and hypothesis testing. *Handbook Of Econometrics*, 4 :2111–2245, 1994.
- Whitney K Newey et Kenneth D West. Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 28(3) :777–787, 1987.
- T.E. Mikosch O.E.E. Barndorff-Nielsen et S.I.E. Resnick. *Lévy Processes : Theory and Applications*. Birkhäuser, 2001. ISBN 9780817641672.
- Peter D Praetz. The distribution of share price changes. *Journal Of Business*, pages 49–55, 1972.
- S James Press. A compound events model for security prices. *Journal Of Business*, pages 317–335, 1967.
- K. Sato. *Lévy Processes And Infinitely Divisible Distributions*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1999. ISBN 9780521553025.
- Wim Schoutens. *Lévy Processes In Finance*. Wiley, 2003.
- David Scott et Christine Yang Dong. *Variancegamma : The Variance Gamma Distribution*, 2012. URL <http://CRAN.R-project.org/package=VarianceGamma>. R package version 0.3-1.
- Eugene Seneta. Fitting the variance-gamma model to financial data. *Journal Of Applied Probability*, pages 177–187, 2004.
- Samuel Sanford Shapiro et Martin B Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4) :591–611, 1965.
- Neil G Shephard. From characteristic function to distribution function : A simple framework for the theory. *Econometric Theory*, 7(04) :519–529, 1991.
- Murray R Spiegel et John Liu. *Schaum's Mathematical Handbook Of Formulas And Tables*, volume 1000. McGraw-Hill, 1999.
- Alan Stuart et J Keith Ord. *Kendall's Advanced Theory Of Statistics, Vol. 1*. Oxford University Press, New York, 1987.
- Gerald Teschl. Topics in real and functional analysis. 2004.
- Christian Walter. "levy-stability under addition and fractal structure of markets : Implications for the actuaries and emphasized examination of matif national contract". In *Proceedings of the 5th AFIR colloquium*, 1995.



JG Wendel. The non-absolute convergence of gil-pelaez' inversion integral. *The Annals of Mathematical Statistics*, 32(1) :338–339, 1961.

Jacob Wolfowitz. The minimum distance method. *The Annals Of Mathematical Statistics*, pages 75–88, 1957.

Jeffrey M Wooldridge. *Econometric Analysis Of Cross Section And Panel Data*. MIT press, 2001.



# Contrat de partage

Cette création est mise à disposition selon le contrat *Attribution - Partage dans les Mêmes Conditions 4.0 International* de Creative Commons disponible à l'adresse <http://creativecommons.org/licenses/by-sa/4.0/deed.fr>

En vertu de ce contrat, vous êtes autorisé à :

- **Partager** — copier, distribuer et communiquer le matériel par tous moyens et sous tous formats
- **Adapter** — remixer, transformer et créer à partir du matériel pour toute utilisation, y compris commerciale.

L'Offrant ne peut retirer les autorisations concédées par la licence tant que vous appliquez les termes de cette licence.

Selon les conditions suivantes :



**Attribution** — Vous devez créditer l'Oeuvre, intégrer un lien vers la licence et indiquer si des modifications ont été effectuées à l'Oeuvre. Vous devez indiquer ces informations par tous les moyens possibles mais vous ne pouvez pas suggérer que l'Offrant vous soutient ou soutient la façon dont vous avez utilisé son Oeuvre.



**Partage à l'identique** — Dans le cas où vous effectuez un remix, que vous transformez, ou créez à partir du matériel composant l'Oeuvre originale, vous devez diffuser l'Oeuvre modifiée dans les mêmes conditions, c'est-à-dire avec la même licence avec laquelle l'Oeuvre originale a été diffusée.

**Aucune autre restriction** — Vous n'êtes pas autorisé à appliquer des conditions légales ou des mesures techniques qui restreindraient légalement autrui à utiliser l'Oeuvre dans les conditions décrites par la licence.

