

Chapitre 3

Mesures de risque

Arthur Charpentier

Les agents face à une situation risquée ont besoin de comparer des positions ou des actions. Mais aujourd'hui, les établissements financiers (avec Bâle II) et les compagnies d'assurance (avec Solvabilité II) doivent surtout constituer des réserves pour faire face aux risques pris, c'est-à-dire qu'elle leur faut une quantification du risque pris. Il peut s'agir d'un risque de marché (changement de la valeur d'un titre), de crédit (risque de ne pas satisfaire ses engagements suite à un défaut), opérationnel (défaillance d'un processus interne, ou externe), voire de modèle (supposer les rendements Gaussiens alors qu'ils ne le sont pas).

3.1 Introduction

3.1.1 La variance comme mesure de risque ?

L'objectif général des études inférentielles est de prendre une décision. La décision est souvent un choix parmi deux modalités accepter ou rejeter entre deux alternatives. Par exemple on se demande si une personne est malade. Pour cela, on fait un test : le test est positif (la personne a de grande chance d'être malade) ou négatif (la personne n'est a priori pas malade). On espère que l'erreur est ici la plus faible possible. De manière plus générale, on a paramètre d'intérêt θ (qui pourrait être l'état de santé du patient dans l'exemple précédent, qui donne plus d'information que la dualité malade-non malade évoquée auparavant), et une statistique $\hat{\theta}$, construite partir de plusieurs caractéristiques du patient, que l'on espère « proche » de la vraie valeur (inconnue et non observable) θ . Pour juger de la proximité entre $\hat{\theta}$ et θ , on se donne classiquement une fonction de coût, L , définie comme une fonction associant à un couple $(\theta, \hat{\theta})$ une grandeur réelle. On peut aussi voir cette fonction de coût comme une erreur.

Il est possible de relier ce coût à une espérance d'utilité. L'utilité est alors une mesure de proximité entre la l'estimation $\hat{\theta}$ et la vraie valeur θ . Comme $\hat{\theta}$ est fonction d'une variable X , on introduit alors également une fonction de coût moyen, ou de risque :

$$R(\theta, \hat{\theta}) = \mathbb{E}[L(\theta, \hat{\theta}(X))] = \int L(\theta, \hat{\theta}(x)) dF_{\theta}(x),$$

où la règle de décision est $\hat{\theta}(x)$, pour chaque résultat d'une expérience aléatoire, et où $F_{\theta}(x) = \mathbb{P}(X \leq x)$.

En économétrie, la fonction de coût usuelle est un coût quadratique, suite aux travaux de Legendre et Gauss (en particulier) vers 1800, qui suggéraient de minimiser la somme des carrés des erreurs. Rappelons que Boscovich et Laplace avaient auparavant suggéré de minimiser une somme des valeurs absolues des erreurs. La fonction de coût quadratique est définie par :

$$L(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^2.$$

Si a souvent été critiqué, elle présente l'avantage de rendre les calculs simples (le plus souvent). De plus, la convexité de cette fonction quadratique permet d'éviter certains paradoxes (nous reviendrons longuement sur l'intérêt des propriétés de convexité par la suite). En particulier, un critère usuel de mesure d'incertitude associé à un estimateur $\hat{\theta}$ est l'erreur quadratique moyenne mean square error :

$$\text{mse}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2] = \text{Var}(\hat{\theta}) + (\text{biais}(\hat{\theta}, \theta))^2.$$

Aussi, pour un estimateur sans biais, c'est la variance qui permet de quantifier l'erreur associée à cette estimation.

Cette idée a été reprise par Harry Markowitz ([31] ou [32]) dans les années 50 dans le contexte de la gestion de portefeuilles (introduite - en italien - dans [11], presque 15 ans auparavant). Il visait à prendre en compte l'effet de diversification que recherchent les investisseurs, en montrant que les investisseurs construisent de façon optimale les portefeuilles efficients en minimisant le risque, mesuré par la variance, pour un niveau de rendement espéré. Il y a généralement deux manières de justifier cette approche. Classiquement, dans la théorie de l'espérance de l'utilité, les agents cherchent à maximiser l'espérance d'utilité du rendement (aléatoire) de leur portefeuille X , $\mathbb{E}(u(X))$. Or si les variations de rendements sont faibles, on peut effectuer un développement limité, en posant $X = \mathbb{E}(X) + \epsilon$:

$$u(X) \sim u(\mathbb{E}(X)) + u'(\mathbb{E}(X))\epsilon + \frac{u''(\mathbb{E}(X))}{2}\epsilon^2,$$

soit, en prenant l'espérance :

$$\mathbb{E}(u(X)) \sim u(\mathbb{E}(X)) + \frac{u''(\mathbb{E}(X))}{2}\text{Var}(X),$$

car, par construction, ϵ est centré, et $\text{Var}(X) = \text{Var}(\epsilon) = \mathbb{E}(\epsilon^2)$. Si l'agent est averse au risque, son utilité sera concave, u'' sera négatif, et il pénalisera les investissements risqués, au sens où leur variance sera trop importante. La seconde justification consiste à supposer les rendements gaussiens, et que les agents ont une aversion absolue pour le risque constante (CARA), c'est-à-dire une utilité exponentielle, $u(x) = -\exp(-\theta x)$. Dans ce cas, leur espérance d'utilité s'écrit :

$$\mathbb{E}(u(X)) = \mathbb{E}(-\exp(-\theta X)) = -\exp\left(-\theta\mathbb{E}(X) + \frac{\theta^2}{2}\text{Var}(X)\right),$$

qui tendra là aussi à pénaliser les investissements trop risqués, au sens de la variance.

Remarque 3.1 Pour des raisons de comodité, on cherchera plutôt à construire des mesures de risque monétaires, c'est-à-dire dans la même unité que X . On préférera alors généralement l'écart-type à la variance.

Enfin, cette idée que le risque peut être quantifié à l'aide de l'écart-type (qui a la même unité que les coûts) est implicitement présente dans l'idée de mutualisation de l'activité d'assurance. En effet, une des bases de l'assurance est le théorème central limite : si $S_n = X_1 + \dots + X_n$ désigne la charge totale payée sur n polices d'assurance sur une année, le théorème central limite garantit que (si les risques sont indépendants et de variance finie) :

$$\frac{S_n - n \cdot \mathbb{E}(X_i)}{\sqrt{n}\sqrt{\text{Var}(X_i)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \text{ lorsque } n \rightarrow \infty.$$

Ce qui s'interprète parfois de la manière suivante : en multipliant par 4 la taille du portefeuille, on multiplie par 4 le chiffre d'affaire (les primes étant souvent proportionnelles à $\mathbb{E}(X_i)$), mais le risque (correspondant à l'écart-type) n'est multiplié que par 2. Ce qui légitime le principe de mutualisation des risques sur des portefeuilles aussi grands que possibles (à condition que les hypothèses qui justifient l'utilisation du théorème central limite soient vérifiées).

Mais pour reprendre ce dernier exemple, on peut montrer que la variance (ou l'écart-type) ne peut pas être suffisante pour mesurer le risque, et comprendre une prise de décision. Paul Samuelson raconte une histoire intéressante à ce sujet dans [39] : considérons un jeu de pile ou face, on l'on peut gagner 200 si la pièce tombe sur pile par contre, on perd 100 si la pièce tombe sur face. On suppose que l'on a autant de chance de tomber sur pile que sur face. On décide à l'avance du nombre de lancers n que l'on fera. Si on prend une décision (nombre de lancers que l'on est prêt à faire) en calculant la variance, on note que la variance du gain X_n (ou de la perte) au bout de n lancers est ici

$$\text{Var}(X_n) = \frac{300^2}{4n} \rightarrow 0 \text{ lorsque } n \rightarrow \infty.$$

Autrement dit, on peut trouver n au delà duquel le risque pris devient aussi petit qu'on le souhaite. Pourtant Paul Samuelson raconte qu'il a trouvé des

personnes prêtes à jouer $n = 5$, voire $n = 10$ parties. Mais personne pour en jouer $n = 100$. Comme il conclut, la variance n'est probablement pas une mesure appropriée pour quantifier le risque.

Remarque 3.2 En fait, on peut noter qu'avec $n = 100$ lancers,

$$\mathbb{P}(X_{100} > 0) = \mathbb{P}(\text{au moins 34 'pile' sur 100 lancers}) \sim 99,91\%.$$

Autrement dit, avec $n = 100$ lancers, on est presque certain de ne rien perdre, ou plus formellement, le quantile (des pertes) à 99,91% est strictement positif. Pourtant, personne n'est prêt à s'engager à jouer 100 parties. La Value-at-Risk - que nous présenterons longuement dans la section 3.2.2 - n'est peut être pas, non plus, une mesure pertinente.

3.1.2 De la comparaison des risques aux mesures de risque

Dans le chapitre 2, nous avons vu comment des propriétés sur un préordre \preceq de comparaison entre risques permettaient de construire une mesure de risque \mathcal{R} , au sens où :

$$X \preceq Y \text{ si et seulement si } \mathcal{R}(X) \leq \mathcal{R}(Y).$$

En particulier, le théorème 2.1 du chapitre 2 montrait que sous des hypothèses de continuité et d'indépendance du préordre \preceq , la comparaison des risques se faisait en comparant $\mathcal{R}(X) = \mathbb{E}_{\mathbb{P}}(u(X))$, où u est une fonction dite d'utilité, selon la terminologie proposée par von Neumann et Morgenstern. Le théorème 2.2 montrait que sous l'axiomatique proposée par Savage sur le préordre, il existait une mesure de probabilité \mathbb{Q} (dite subjective) telle que $\mathcal{R}(X) = \mathbb{E}_{\mathbb{Q}}(u(X))$.

Si cette approche a été centrale en théorie de la décision (comme le rappellent [40] et [33]), l'optique a changé à la fin des années 90 (sous l'impulsion de [4]) pour proposer une axiomatique directement sur \mathcal{R} , en introduisant une notion de cohérence, puis de convexité des mesures de risque.

3.1.3 Plan du chapitre

La section 2 présentera la base axiomatique des mesures de risque, en présentant les mesures de risque usuelles, en particulier la Value-at-Risk (VaR, discutée dans la section 3.2.2), la Tail-Value-at-Risk (TVaR) et les mesures associées (CTE, CVaR, ES, discutées dans la section 3.2.3), la transformée de d'Esscher (section 3.2.4), les mesures dites de Wang (section 3.2.5) et plus généralement les mesures par distorsion (section 3.2.6). La section 3 reviendra sur les ordres de comparaison, en discutant les propriétés des ordres induits par les mesures de risque. Enfin, la section 4 fera un état de l'art succinct sur l'estimation des mesures de risque.

Remarque 3.1 Dans cette section X désigne un montant de perte. On interprétera $\mathcal{R}(X)$ comme le capital à détenir pour faire face aux pertes X . Aussi, \mathcal{R} quantifie le niveau de danger inhérent au risque X : de grandes valeurs de $\mathcal{R}(X)$ indiqueront que X est « dangereux » (dans un sens que l'on précisera). De plus, nous n'introduirons pas de dynamique ici, car la construction d'une axiomatique intégrant à la fois la composante risque et la composante temps irait au delà d'une introduction rapide sur le sujet.

3.2 Approche axiomatique des mesures de risque et mesures de risque usuelles

Une mesure de risque étant une fonction définie sur l'espace des variables aléatoires, et prenant ses valeurs dans \mathbb{R} , il convient de spécifier quelques propriétés naturelles et souhaitables de ces mesures :

- invariance en loi : si $X \stackrel{\mathcal{L}}{=} Y$, alors $\mathcal{R}(X) = \mathcal{R}(Y)$,
- croissance : si $X \geq Y$, alors $\mathcal{R}(X) \geq \mathcal{R}(Y)$,
- invariance par translation : si $k \in \mathbb{R}$, alors $\mathcal{R}(X + k) = \mathcal{R}(X) + k$,
- homogénéité : si $\lambda \in \mathbb{R}_+$, alors $\mathcal{R}(\lambda X) = \lambda \cdot \mathcal{R}(X)$,
- sous additivité : pour tous risques X et Y , $\mathcal{R}(X + Y) \leq \mathcal{R}(X) + \mathcal{R}(Y)$,
- convexité : si $\beta \in [0, 1]$, alors

$$\mathcal{R}(\beta X + [1 - \beta]Y) \leq \beta \cdot \mathcal{R}(X) + (1 - \beta) \cdot \mathcal{R}(Y).$$

Par la suite, nous ne nous intéresserons qu'aux mesures de risque invariantes en loi.

Remarque 3.3 En science actuarielle, les réflexions sur les mesures de risque sont arrivées plus tôt qu'en finance, lors des réflexions sur les principes de valorisation (« premium principles », dont [21] propose une revue de littérature, mise à jour par la suite dans [36] par exemple). La propriété d'invariance par translation est discutée dans [35], celle d'homogénéité dans [41] et celle de convexité dans [34]. Un parallèle entre les mesures de risque et les primes en sciences actuarielles a été étudié dans [28]. On notera là aussi que [11] avait initié cette axiomatique sur les mesures de risque, dans un contexte actuariel, dès 1940.

Une des conséquences de l'invariance par translation est que :

$$\mathcal{R}(X - \mathcal{R}(X)) = 0,$$

Définition 3.1 Une mesure de risque est dite :

- *monétaire* si elle est monotone et invariante par translation,
- *convexe* si elle est monétaire et convexe,
- *cohérente* si elle est monétaire, homogène et sous-additive.

Remarque 3.4 Une mesure cohérente est toujours normalisée à $\mathcal{R}(0) = 0$, par homogénéité. En revanche, si une mesure convexe est normalisée par $\mathcal{R}(0) = 0$, alors pour $\lambda \in [0, 1]$

$$\mathcal{R}(\lambda X) = \mathcal{R}(\lambda X + [1 - \lambda]0) \leq \lambda \mathcal{R}(X)$$

alors que pour $\lambda \in [1, +\infty)$,

$$\mathcal{R}(X) = \mathcal{R}\left(\frac{1}{\lambda}\lambda X\right) = \mathcal{R}\left(\frac{1}{\lambda}\lambda X + \left(1 - \frac{1}{\lambda}\right)0\right) \leq \frac{1}{\lambda}\mathcal{R}(\lambda X)$$

c'est-à-dire que $\mathcal{R}(\lambda X) \geq \lambda \mathcal{R}(X)$.

Corollaire 3.1 Si \mathcal{R} est une mesure de risque monétaire, homogène et normalisée à 0, alors la convexité et la sous-additivité sont des notions équivalentes.

Définition 3.2 Si \mathcal{R} est une mesure de risque, on définit la région de risques acceptables pour la mesure \mathcal{R} comme

$$\mathcal{A} = \{X, \mathcal{R}(X) \leq 0\}.$$

Réciproquement, si \mathcal{A} est une région de risques acceptables, la mesure de risque induite \mathcal{R} est

$$\mathcal{R}(X) = \inf\{m, X - m \in \mathcal{A}\}.$$

Proposition 3.1 Si \mathcal{R} est une mesure de risque monétaire alors

- \mathcal{R} est convexe si et seulement si \mathcal{A} est convexe,
- \mathcal{R} est positivement homogène si et seulement si \mathcal{A} est un cône.

Preuve 3.1 Pour le premier point, si \mathcal{R} est convexe, alors \mathcal{A} est convexe. Réciproquement, si \mathcal{A} est convexe, soient X_1, X_2, m_1 et m_2 tels que $X_i - m_i \in \mathcal{A}$, pour $i = 1, 2$. Par hypothèse, pour tout $\lambda, \lambda[X_1 - m_1] + (1 - \lambda)[X_2 - m_2] \in \mathcal{A}$, c'est-à-dire que $\mathcal{R}(\lambda[X_1 - m_1] + (1 - \lambda)[X_2 - m_2]) \leq 0$, soit, par la propriété d'invariance par translation, et par monotonie,

$$\mathcal{R}(\lambda X_1 + (1 - \lambda)X_2) \leq \lambda m_1 + (1 - \lambda)m_2,$$

et ceci pour tout m_1 et m_2 . Il suffit de le faire pour $m_i = \mathcal{R}(X_i)$. Pour le second point, si \mathcal{A} est un cône, alors pour tout $X - m \in \mathcal{A}$, $\lambda(X - m) \in \mathcal{A}$, pour tout $\lambda > 0$. Donc $\mathcal{R}(\lambda X - \lambda m) \leq 0$, d'où $\mathcal{R}(\lambda X) \leq \lambda m$. Si $m = \mathcal{R}(X)$, on en déduit que $\mathcal{R}(\lambda X) \leq \lambda \mathcal{R}(X)$. Et si $X - m \notin \mathcal{A}$, alors $\lambda(X - m) \notin \mathcal{A}$, et $\mathcal{R}(\lambda X) > \lambda m$. On fait alors tendre m vers $\mathcal{R}(X)$ pour avoir le résultat souhaité.

On dispose du théorème de représentation suivant

Théorème 3.1 \mathcal{R} est une mesure de risque monétaire convexe si et seulement si pour tout X bornée ($X \in L^\infty$) :

$$\mathcal{R}(X) = \max_{\mathbb{Q} \in \mathcal{M}} \{\mathbb{E}_{\mathbb{Q}} - \alpha(\mathbb{Q})\},$$

où \mathcal{M} est l'ensemble des mesures additives et normalisées à 1, et

$$\alpha(\mathbb{Q}) = \sup_{X \in \mathcal{A}} \{\mathbb{E}_{\mathbb{Q}}(X)\},$$

où \mathcal{A} est l'ensemble d'acceptation associé à \mathcal{R} .

On notera que \mathcal{M} contient plus que des mesures de probabilité.

Remarque 3.5 α est la conjugué de Legendre-Fenchel de \mathcal{R} : si f est une fonction convexe, on pose

$$f^*(x) = -\inf_{\lambda} \{f(x) - \lambda x\} = \sup_{\lambda} \{\lambda x - f(x)\}$$

telle que $(f^*)^* = f$. Alors

$$\mathcal{R}(X) = \sup_{\mathbb{Q} \in \mathcal{B}} \{\mathbb{E}_{\mathbb{Q}} - \mathcal{R}^*(\mathbb{Q})\}$$

où \mathcal{B} est l'ensemble des mesures bornées (et \mathcal{M} étant la boule unité de \mathcal{B}).

Remarque 3.6 Les mesures cohérentes peuvent s'écrire

$$\mathcal{R}(X) = \max_{\mathbb{Q} \in \mathcal{Q}} \{\mathbb{E}_{\mathbb{Q}}(X)\}$$

où

$$\mathcal{Q} = \{\mathbb{Q} \in \mathcal{M}, \alpha(\mathbb{Q}) = 0\}.$$

Ces résultats ont été établis par [4], [13], [12], [17] et [18]. Notons que des ré-écritures de la propriété de sous-additivité ont été proposées dans la littérature (en particulier [38] ou [16]).

Proposition 3.2 Si \mathcal{R} est une mesure de risque monétaire alors elle vérifie la propriété

- d'additivité pour les risques comonotones si, pour tout X et Y comonotones, $\mathcal{R}(X + Y) = \mathcal{R}(X) + \mathcal{R}(Y)$,
- de corrélation maximale (par rapport à une mesure μ) si pour tout X ,

$$\mathcal{R}(X) = \sup \{\mathbb{E}(X \cdot U) \text{ où } U \sim \mu\}$$

- de cohérence forte si pour tout X et Y , $\sup\{\mathcal{R}(\tilde{X} + \tilde{Y})\} = \mathcal{R}(X) + \mathcal{R}(Y)$,
pour $\tilde{X} \stackrel{\mathcal{L}}{=} X$ et $\tilde{Y} \stackrel{\mathcal{L}}{=} Y$.

Remarque 3.7 La VaR et la TVaR (qui seront détaillées respectivement dans les sections 3.2.2 et 3.2.3) sont additives pour les risques comonotones.

Proposition 3.3 Si \mathcal{R} est une mesure de risque convexe, les trois propriétés suivantes sont équivalentes

- \mathcal{R} est fortement cohérente,
- \mathcal{R} est additive pour des risques comonotones,
- \mathcal{R} est une mesure de corrélation maximale.

Enfin, [27] a montré le résultat suivant dès lors que \mathcal{R} est une mesure invariante en loi (ce que nous supposons dans tout ce chapitre).

Proposition 3.4 Une mesure de risque cohérente \mathcal{R} est additive pour des risques comonotones si et seulement s'il existe une fonction décroissante positive g sur $[0, 1]$ telle que

$$\mathcal{R}(X) = \int_0^1 g(t) F_X^{-1}(1-t) dt$$

où $F_X(x) = \mathbb{F}(X \leq x)$.

3.2.1 L'équivalent certain

Définition 3.3 Soit u une fonction d'utilité concave, strictement croissante, alors l'équivalent certain $\mathcal{R}(X)$ associé à une perte X vérifie

$$u(\mathcal{R}(X)) = \mathbb{E}[u(X)] \text{ soit } \mathcal{R}(X) = u^{-1}(\mathbb{E}[u(X)]).$$

Exemple 3.1 Considérons une fonction d'utilité exponentielle, $u(x) = 1 - \exp[-\theta x]$, caractérisant une aversion absolue pour le risque contante, au sens où $-u''(x)/u'(x) = \theta$. Alors l'équivalent certain associé à une perte X est

$$\mathcal{R}(X) = \frac{1}{\theta} \log(\mathbb{E}_{\mathbb{P}}[e^{-\theta X}]).$$

Cette mesure est généralement appelée mesure de risque entropique ([19] ou encore [30] ont popularisé son utilisation en actuariat et en finance). Comme l'avait noté [10], cette mesure peut également s'écrire

$$\mathcal{R}(X) = \sup_{\mathbb{Q} \ll \mathbb{P}} \left\{ \mathbb{E}_{\mathbb{Q}}[-X] - \frac{1}{\theta} H(\mathbb{Q}|\mathbb{P}) \right\},$$

où $H(\mathbb{Q}|\mathbb{P})$ désigne l'entropie relative (ou distance de Kullback-Leibler), $H(\mathbb{Q}|\mathbb{P}) = \mathbb{E}_{\mathbb{Q}} \left(\log \frac{d\mathbb{Q}}{d\mathbb{P}} \right)$.

D'où le nom de cette mesure de risque. L'idée est qu'un agent a une mesure de référence \mathbb{P} , mais qu'il considère comme une approximation de ce que doit

être la vraie mesure. Il peut faire face à un aléa X , et choisit de considérer un ensemble d'autres mesures \mathbb{Q} . Il va alors pénaliser en tenant compte de la distance à la mesure de référence \mathbb{P} . On notera que pour $\theta > 0$, $\partial\mathcal{R}(X)$ est :

$$\partial\mathcal{R}(X) = \frac{\exp[-\theta X]}{\mathbb{E}(\exp[-\theta X])},$$

que l'on verra réapparaître en introduisant la transformée d'Esscher, dans la section 3.2.4.

3.2.2 La Value-at-Risk

La Value-at-Risk est apparue (sous ce nom) dans les années 90, en réponse à de nombreux désastres qui ont touché les marchés de capitaux à cette période ([25] propose un historique intéressant de cette naissance). En fait, on peut remonter beaucoup plus tôt pour voir apparaître cette mesure de risque pour la première fois. Par exemple, lors à des débats sur l'inoculation, avant l'invention de la vaccination (par Edward Jenner en 1796). Étant donné qu'alors c'était la maladie elle-même qui était inoculée (et non une forme atténuée, comme ce fut le cas avec la vaccine), les résultats étaient assez inégaux, parfois pires que la maladie elle-même. D'Alembert disait clairement qu'il ne faut pas seulement prendre en compte le gain « en moyenne » (du au fait qu'un grand nombre de personnes développeront des anticorps) mais également le risque « Ce n'est donc ni la longueur de la vie moyenne, ni la petitesse du risque qui doit déterminer à admettre l'inoculation; c'est uniquement le rapport entre le risque d'une part, et de l'autre l'augmentation de la vie moyenne » (repris dans [7]). Il rappelle encore plus précisément que le risque est la probabilité de mourir des suites d'une inoculation ou, plus généralement, la probabilité de survenance d'un événement désagréable. On retrouve cette idée dans un cadre financier dès 1952, évoqué par Arthur Roy, l'année où Harry Markowitz propose au contraire d'utiliser la variance comme mesure de risque, avec la notion de « safety rule », dans [37].

Définition 3.4 *On appelle Value-at-Risk de niveau $\alpha \in (0, 1)$ le quantile de niveau α ,*

$$\mathcal{R}_\alpha(X) = \text{VaR}(X; \alpha) = x_\alpha \text{ où } \mathbb{P}(X \leq x_\alpha) = \alpha,$$

ou encore

$$\text{VaR}(X; \alpha) = \inf\{x, \mathbb{P}(X \leq x) \geq \alpha\} = F_X^{-1}(\alpha) = Q_X(\alpha).$$

Remarque 3.8 Avec cette notation, on notera que $\mathcal{R}_\alpha(X)$ est une fonction croissante en α , alors que certains articles et ouvrages notent $\mathcal{R}_\alpha(X)$ le quantile d'ordre $1 - \alpha$.

Commençons par rappeler que la VaR vérifie une propriété de stabilité par transformation monotone :

Lemme 3.1 *Pour tout $\alpha \in (0, 1)$, si g est un fonction strictement croissante et continue à gauche :*

$$\text{VaR}(g(X), \alpha) = F_{g(X)}^{-1}(\alpha) = g(F_X^{-1}(\alpha)) = g(\text{VaR}(X; \alpha)),$$

alors que si g est un fonction strictement décroissante, continue à droite, et si F_X est bijective :

$$\text{VaR}(g(X), \alpha) = F_{g(X)}^{-1}(\alpha) = g(F_X^{-1}(1 - \alpha)) = g(\text{VaR}(X; 1 - \alpha)).$$

Preuve 3.2 Nous ne démontrerons que le cas croissant (le raisonnement étant analogue dans le cas décroissant). Si g est strictement croissante et continue à gauche, alors, pour tout $0 < \alpha < 1$,

$$F_{g(X)}^{-1}(\alpha) \leq x \text{ si et seulement si } \alpha \leq F_{g(X)}(x).$$

Puisque g est continue à gauche,

$$g(z) \leq x \text{ si et seulement si } z \leq \sup \{y \in \mathbb{R} | g(y) \leq x\},$$

pour tout x, z . Ainsi

$$p \leq F_{g(X)}(x) \text{ si et seulement si } p \leq F_X(\sup \{y \in \mathbb{R} | g(y) \leq x\}).$$

Si $\sup \{y \in \mathbb{R} | g(y) \leq x\}$ est fini, on obtient l'équivalence souhaitée, puisque

$$p \leq F_X(\sup \{y \in \mathbb{R} | g(y) \leq x\}) \text{ si et seulement si } F_{g(X)}^{-1}(p) \leq g(F_X^{-1}(p)),$$

en utilisant le fait que $p \leq F_X(z)$ est équivalent à $F_X^{-1}(z) \leq p$.

Si $\sup \{y \in \mathbb{R} | g(y) \leq x\}$ est infini, l'équivalence ci-dessus ne peut être utilisée, mais le résultat reste valable. En effet, si $\sup \{y \in \mathbb{R} | g(y) \leq x\} = +\infty$, l'équivalence devient

$$p \leq 1 \text{ si et seulement si } F_X^{-1}(p) \leq +\infty.$$

La stricte croissante de g et la continuité à droite permettent d'obtenir

$$F_X^{-1}(p) \leq \sup \{y \in \mathbb{R} | g(y) \leq x\} \text{ si et seulement si } g(F_X^{-1}(p)) \leq x,$$

et en combinant toutes les inégalités, on peut écrire

$$F_{g(X)}^{-1}(p) \leq x \text{ si et seulement si } g(F_X^{-1}(p)) \leq x$$

pour tout x , ce qui implique $F_{g(X)}^{-1}(p) = g(F_X^{-1}(p))$ pour tout p .

La fonction quantile est utile dans les méthodes de simulation de part la propriété suivante (on parle souvent de « méthode d'inversion de la fonction de répartition ») :

Proposition 3.1 Soit $U \sim \mathcal{U}(0, 1)$ et soit X une variable aléatoire quelconque. La variable aléatoire $F_X^{-1}(U)$ a même loi que X .

Preuve 3.3 Quel que soit $x \in \mathbb{R}$, avec le Lemme 3.1, on a :

$$\mathbb{P}[F_X^{-1}(U) \leq x] = \mathbb{P}[F_X(x) \geq U] = F_X(x).$$

En prenant $g(x) = x + c$ et $g(x) = cx$, on déduit immédiatement de cette dernière propriété que la VaR est invariante par translation et homogène.

Toutefois, la VaR n'est pas sous-additive.

Exemple 3.1 Considérons les risques indépendants suivant des lois de Pareto, $X \sim \mathcal{P}ar(1, 1)$ et $Y \sim \mathcal{P}ar(1, 1)$:

$$\mathbb{P}(X > t) = \mathbb{P}(Y > t) = \frac{1}{1+t}, \quad t > 0.$$

Nous avons alors :

$$\text{VaR}(X; \alpha) = \text{VaR}(Y; \alpha) = \frac{1}{1-\alpha} - 1.$$

De plus, on peut vérifier que :

$$\mathbb{P}[X + Y \leq t] = 1 - \frac{2}{2+t} + 2 \frac{\log(1+t)}{(2+t)^2}, \quad t > 0.$$

Comme nous avons :

$$\mathbb{P}[X + Y \leq 2\text{VaR}[X; \alpha]] = \alpha - \frac{(1-\alpha)^2}{2} \log\left(\frac{1+\alpha}{1-\alpha}\right) < \alpha$$

l'inégalité :

$$\text{VaR}(X; \alpha) + \text{VaR}(Y; \alpha) < \text{VaR}(X + Y; \alpha)$$

est vraie quel que soit α , de sorte que la VaR ne peut pas être sous-additive dans ce cas.

3.2.3 La Tail Value-at-Risk

Si la Value-at-Risk s'intéressait aux probabilités d'évènements rares et extrêmes, avec la Tail-Value-at-Risk, nous nous intéresserons à ce qui se passe (en moyenne) lorsque ces évènements extrêmes surviennent.

Définition 3.1 La Tail Value-at-Risk au niveau α , notée $\text{TVaR}(X; \alpha)$ est définie par :

$$\text{TVaR}(X; \alpha) = \frac{1}{1-\alpha} \int_{\alpha}^1 \text{VaR}(X; t) dt.$$

Autrement dit, la TVaR apparaît comme la moyenne des VaR de niveau supérieur à α .

Remarque 3.1 Il est intéressant de noter qu'il existe une fonction de répartition \tilde{F}_X , appelée transformée de Hardy-Littlewood de F_X (introduite dans [22]), telle que pour tout α :

$$\tilde{F}_X^{-1}(\alpha) = \text{TVaR}(X; \alpha).$$

Si on note \tilde{X} une variable aléatoire de fonction de répartition \tilde{F}_X , on a :

$$\text{TVaR}(X; \alpha) = \text{VaR}(\tilde{X}; \alpha).$$

La TVaR d'un risque X est donc la VaR de la transformée de Hardy-Littlewood de X .

Notons que, $\text{TVaR}[X; 0] = \mathbb{E}[X]$. Et comme

$$\text{TVaR}[X; \alpha] = \frac{1}{1-\alpha} \left\{ \mathbb{E}[X] - \int_0^\alpha \text{VaR}[X; \xi] d\xi \right\}. \quad (3.1)$$

on en déduit que la Tail Value-at-Risk est une fonction croissante du niveau α . En effet :

$$\frac{d}{d\alpha} \text{TVaR}(X; \alpha) = \frac{\text{TVaR}(X; \alpha)}{1-\alpha} - \frac{\text{VaR}(X; \alpha)}{1-\alpha}.$$

Et comme $\alpha \mapsto \text{VaR}[X; \alpha]$ est une fonction croissante :

$$\text{TVaR}[X; \alpha] = \frac{1}{1-\alpha} \int_\alpha^1 \underbrace{\text{VaR}[X; t]}_{\geq \text{VaR}[X; \alpha]} dt \geq \text{VaR}[X; \alpha],$$

on en déduit que $\frac{d}{d\alpha} \text{TVaR}[X; \alpha] \geq 0$, et :

$$\text{TVaR}[X; \alpha] \geq \text{TVaR}[X; 0] = \mathbb{E}[X].$$

Ainsi, la TVaR contient toujours un chargement de sécurité.

On notera que la TVaR a été déclinée sous des formes relativement proches dans la littérature.

Définition 3.2 La *Conditional Tail Expectation* au niveau de probabilité α , notée $\text{CTE}[X; \alpha]$, représente la perte attendue sachant que la VaR au niveau α est dépassée :

$$\text{CTE}[X; \alpha] = \mathbb{E} \left[X \mid X > \text{VaR}[X; \alpha] \right].$$

Il s'agit donc de la version mathématique du concept intuitif de « perte moyenne dans les pires $1 - \alpha\%$ des cas ».

Définition 3.3 La *Conditional-VaR* au niveau de probabilité α , notée $\text{CVaR}[X; \alpha]$, est la valeur moyenne des pertes qui excèdent la VaR , correspondant à l'excédent moyen au-delà de la VaR ou encore :

$$\begin{aligned}\text{CVaR}[X; \alpha] &= \mathbb{E}\left[X - \text{VaR}[X; \alpha] \mid X > \text{VaR}[X; \alpha]\right] \\ &= e_X(\text{VaR}[X; \alpha]) \\ &= \text{CTE}[X; \alpha] - \text{VaR}[X; \alpha].\end{aligned}$$

où $e_X(s) = \mathbb{E}[X \mid X > s]$.

Définition 3.4 L'*Expected shortfall* au niveau de probabilité α , notée $\text{ES}[X; \alpha]$, n'est autre que la prime stop-loss dont la rétention est fixée à $\text{VaR}[X; \alpha]$:

$$\text{ES}[X; \alpha] = \mathbb{E}[(X - \text{VaR}[X; \alpha])_+].$$

Le résultat suivant détaille les liens qui peuvent exister entre ces différentes mesures de risque.

Proposition 3.1 Quel que soit le niveau de probabilité $\alpha \in (0, 1)$, les identités suivantes sont vérifiées :

$$\text{TVaR}[X; \alpha] = \text{VaR}[X; \alpha] + \frac{1}{1 - \alpha} \text{ES}[X; \alpha], \quad (3.2)$$

$$\text{CTE}[X; \alpha] = \text{VaR}[X; \alpha] + \frac{1}{\overline{F}_X(\text{VaR}[X; \alpha])} \text{ES}[X; \alpha]. \quad (3.3)$$

Preuve 3.4 L'expression (3.2) provient de

$$\begin{aligned}\text{ES}[X; \alpha] &= \int_0^1 (\text{VaR}[X; \xi] - \text{VaR}[X; \alpha])_+ d\xi \\ &= \int_\alpha^1 \text{VaR}[X; \xi] d\xi - \text{VaR}[X; \alpha](1 - \alpha).\end{aligned}$$

L'expression (3.3) se déduit de

$$\text{ES}[X; \alpha] = \mathbb{E}\left[X - \text{VaR}[X; \alpha] \mid X > \text{VaR}[X; \alpha]\right] \overline{F}_X(\text{VaR}[X; \alpha]). \quad (3.4)$$

La Propriété 3.1 permet d'établir le résultat suivant (qui fournit l'interprétation usuelle de la TVaR).

Proposition 3.2 La CTE et la TVaR coïncident pour des risques dont la fonction de répartition est continue :

$$\text{CTE}[X; \alpha] = \text{TVaR}[X; \alpha], \quad \text{pour } \alpha \in (0, 1). \quad (3.5)$$

Preuve 3.5 Il suffit de noter que si F_X est continue alors

$$F_X(\text{VaR}[X; \alpha]) = \alpha \text{ et } \bar{F}_X(\text{VaR}[X; \alpha]) = 1 - \alpha$$

de sorte que (3.2)-(3.3) donnent le résultat annoncé.

La TVaR est invariante par translation et homogène. En effet, comme la VaR est invariante par translation,

$$\begin{aligned} \text{TVaR}[X + c; \alpha] &= \frac{1}{1 - \alpha} \int_{\alpha}^1 \text{VaR}[X + c; \xi] d\xi \\ &= \frac{1}{1 - \alpha} \int_{\alpha}^1 (\text{VaR}[X; \xi] + c) d\xi \\ &= \text{TVaR}[X; \alpha] + c. \end{aligned}$$

De la même manière, l'homogénéité de la VaR garantit l'homogénéité de la TVaR.

Proposition 3.3 Soient le risque X et le niveau de perte x tels que $\bar{F}_X(x) > 0$. Quel que soit l'événement aléatoire E tel que $\mathbb{P}[E] = \bar{F}_X(x)$, on a $\mathbb{E}[X|E] \leq \mathbb{E}[X|X > x]$.

Preuve 3.6 Notons que l'on peut écrire

$$\mathbb{P}[E|X > x] = \frac{\mathbb{P}[X > x|E]\mathbb{P}[E]}{\mathbb{P}[X > x]} = \mathbb{P}[X > x|E].$$

Aussi,

$$\begin{aligned} \mathbb{E}[X|X > x] &= x + \mathbb{E}[X - x|X > x, E]\mathbb{P}[E|X > x] \\ &\quad + \mathbb{E}[X - x|X > x, \bar{E}]\mathbb{P}[\bar{E}|X > x] \\ &\geq x + \mathbb{E}[X - x|X > x, E]\mathbb{P}[E|X > x] \\ &= x + \mathbb{E}[X - x|X > x, E]\mathbb{P}[X > x|E] \\ &\geq x + \mathbb{E}[X - x|X > x, E]\mathbb{P}[X > x|E] \\ &\quad + \mathbb{E}[X - x|X \leq x, E]\mathbb{P}[X \leq x|E] \\ &= \mathbb{E}[X|E] \end{aligned}$$

ce qui achève la preuve.

Ce résultat est intuitif : $\mathbb{E}[X|E]$ représente le montant moyen de perte lorsqu'on sait que l'événement E est réalisé. La Propriété 3.3 nous apprend que la pire information qui soit est que X soit grande, correspondant à $X > x$.

Cette proposition 3.3 permet d'affirmer que la TVaR est sous-additive lorsque les risques sont continus. En effet, dans ce cas, la TVaR et la CTE coïncident

et

$$\begin{aligned}
\text{TVaR}[X + Y; \alpha] &= \mathbb{E}[X | X + Y > \text{VaR}[X + Y; \alpha]] \\
&\quad + \mathbb{E}[Y | X + Y > \text{VaR}[X + Y; \alpha]] \\
&\leq \mathbb{E}[X | X > \text{VaR}[X; \alpha]] \\
&\quad + \mathbb{E}[Y | Y > \text{VaR}[Y; \alpha]] \\
&= \text{TVaR}[X; \alpha] + \text{TVaR}[Y; \alpha].
\end{aligned}$$

De la même manière, la TVaR est monotone, puisque lorsque $\mathbb{P}[X \leq Y] = 1$

$$\begin{aligned}
\text{TVaR}[Y; \alpha] &= \mathbb{E}[Y | Y > \text{VaR}[Y; \alpha]] \\
&\geq \mathbb{E}[Y | X > \text{VaR}[X; \alpha]] \\
&\geq \mathbb{E}[X | X > \text{VaR}[X; \alpha]] \\
&= \text{TVaR}[X; \alpha].
\end{aligned}$$

Proposition 3.5 *La TVaR est cohérente pour les risques continus, et coïncide alors avec la CTE.*

On peut montrer que la TVaR est la plus petite mesure de risque majorant la VaR qui soit cohérente (ce résultat se retrouve dans [27] et [2], il sera montré dans la Proposition 3.6 parmi les sous classes des mesures cohérentes définies par une distorsion de probabilités).

3.2.4 La transformée d'Esscher

La mesure de risque d'Esscher consiste à prendre la prime pure de la transformée d'Esscher du risque initial, ce qui se formalise de la manière suivante :

Définition 3.5 *La mesure de risque d'Esscher de paramètre $h > 0$ du risque X , notée $\text{Es}[X; h]$, est donnée par :*

$$\text{Es}[X; h] = \frac{\mathbb{E}[X \exp(hX)]}{\mathbb{E}[\exp(hX)]} = \frac{d}{dh} \log M_X(h),$$

où $M_X(h)$ est la fonction génératrice des moments.

Remarque 3.2 En fait, $\text{Es}[X; h]$ n'est autre que la valeur espérée de la transformée d'Esscher X_h de X , dont la fonction de répartition est donnée par :

$$dF_{X,h}(t) = \frac{\exp(ht)dF_X(t)}{M_X(h)} \quad (3.6)$$

où F_X est la fonction de répartition de X , et

$$F_{X,h}(t) = \int_{\xi=-\infty}^t \exp(h\xi) \frac{dF(\xi)}{M_X(h)}. \quad (3.7)$$

c'est-à-dire :

$$\text{Es}[X; h] = \mathbb{E}[X_h] = \int_{\xi \in \mathbb{R}^+} \xi dF_{X,h}(\xi).$$

On remplace donc le risque initial X par un risque moins favorable X_h avant de calculer la prime pure.

Proposition 3.2 $\text{Es}[X; h]$ est une fonction croissante de h .

Preuve 3.7 Le résultat annoncé découle directement de

$$\begin{aligned} \frac{d}{dh} \mathbb{E}[X_h] &= \int_{\xi \in \mathbb{R}^+} x^2 dF_{X,h}(\xi) - \left(\int_{\xi \in \mathbb{R}^+} x dF_{X,h}(\xi) \right)^2 \\ &= \text{Var}[X_h] \geq 0. \end{aligned} \quad (3.8)$$

Cette dernière propriété garantit que quel que soit $h > 0$,

$$\text{Es}[X; h] \geq \text{Es}[X; 0] = \mathbb{E}[X],$$

de sorte que la mesure de risque d'Esscher contient un chargement de sécurité.

La mesure de risque d'Esscher n'est pas homogène (sauf dans le cas trivial $h = 0$). Elle est invariante par translation mais pas monotone.

Exemple 3.2 Considérons les risques X et Y tels que

$$\mathbb{P}[X = 0, Y = 0] = \mathbb{P}[X = 0, Y = 3] = \mathbb{P}[X = 6, Y = 6] = \frac{1}{3}.$$

Dans ce cas, $\mathbb{P}[X \leq Y] = 1$, mais

$$\text{Es}[X; 1/2] = 5,4567 > \text{Es}[Y; 1/2] = 5,2395.$$

Exemple 3.2 En finance de marché, le théorème fondamental de valorisation d'actifs (énoncé dans [3], repris par la suite en temps continu par [24] et [14] pour une version plus générale) nous dit qu'il n'y a pas d'opportunité d'arbitrage s'il existe une mesure risque neutre équivalente à la mesure originale \mathbb{P} telle que le prix d'un actif est l'espérance (sous cette probabilité) du payoff actualisé. Autrement dit, le prix d'une option européenne d'achat, par exemple est

$$e^{-rT} \mathbb{E}_{\mathbb{Q}}((S_T - K)_+).$$

[20] ont montré que ce prix pouvait être écrit comme une mesure de risque d'Esscher sous certaines hypothèses sur la diffusion du sous-jacent (S_t) .

3.2.5 Les mesures de risque de Wang

Nous appellerons désormais fonction de distorsion toute fonction croissante $g : [0, 1] \rightarrow [0, 1]$ telle que $g(0) = 0$ et $g(1) = 1$. Nous allons présenter dans cette section les mesures par distorsion (qui seront étudiées plus en détails dans la section suivante) dans le cas particulier proposé par [48], à savoir le cas des variables aléatoires (strictement) positives, soit $F(0) = \mathbb{P}(X \leq 0) = 0$.

Définition 3.6 *La mesure de risque de Wang associée à la fonction de distorsion g , notée \mathcal{R}_g , est définie par :*

$$\mathcal{R}_g(X) = \int_0^\infty g(1 - F_X(x)) dx = \int_0^\infty g(\bar{F}_X(x)) dx. \quad (3.9)$$

Remarque 3.3 On notera que la fonction de distorsion $g(q) = q$ (fonction identité) correspond à l'espérance mathématique $\mathbb{E}[X]$. De plus, si $g(q) \geq q$ quel que soit $q \in [0, 1]$, on a alors $\mathcal{R}_g[X] \geq \mathbb{E}[X]$, de sorte que les mesures de risque de Wang associées à de telles fonctions de distorsion contiennent un chargement de sécurité. Enfin, il est intéressant de remarquer que lorsque $g_1(q) \leq g_2(q)$ quel que soit $q \in [0, 1]$ nous avons alors $\mathcal{R}_{g_1}[X] \leq \mathcal{R}_{g_2}[X]$.

En substituant $\int_0^{\bar{F}_X(x)} dg(\alpha)$ à $g(\bar{F}_X(x))$ dans (3.15) et en permutant les intégrales, on obtient le résultat suivant.

Proposition 3.3 *Quel que soit le risque X , la mesure de risque de Wang associée à la fonction de distorsion g peut s'écrire :*

$$\mathcal{R}_g[X] = \int_0^1 \text{VaR}[X; 1 - \alpha] dg(\alpha). \quad (3.10)$$

Ainsi, les mesures de risque de Wang sont des moyennes pondérées de VaR.

Exemple 3.3 En particulier, si nous considérons la fonction de distorsion $g : [0, 1] \rightarrow [0, 1]$ définie par :

$$g_\alpha(x) = \mathbf{1}[x \geq 1 - \alpha]$$

pour une valeur $\alpha \in (0, 1)$ fixée, il vient alors par (3.10) :

$$\mathcal{R}_{g_\alpha}[X] = \text{VaR}[X; \alpha]$$

qui montre que la VaR au niveau de probabilité α est une mesure de Wang particulière correspondant à une fonction de distorsion passant de 0 à 1 en $1 - \alpha$. Dans ce cas, g_α est une fonction de répartition, associée à une masse de Dirac en $1 - \alpha$.

Exemple 3.4 De la même manière, en repartant de (3.10) avec la fonction de distorsion :

$$g_\alpha(x) = \min \left\{ \frac{x}{1-\alpha}, 1 \right\},$$

pour $\alpha \in (0, 1)$ fixé on obtient :

$$\mathcal{R}_{g_\alpha}[X] = \frac{1}{1-\alpha} \int_0^{1-\alpha} \text{VaR}[X; 1-\xi] d\xi = \text{TVaR}[X; \alpha].$$

Dans ce cas également, g_α est une fonction de répartition, correspondant à une loi uniforme sur $[0, 1-\alpha]$.

En revanche, ce n'est généralement pas le cas de l'expected shortfall (ES). Supposons par l'absurde que $\text{ES}[X; \alpha]$ puisse effectivement être représentée sous la forme d'une mesure de risque de Wang, associée à une fonction de distorsion g_α , $\text{ES}[X; \alpha] = \mathcal{R}_{g_\alpha}[X]$ quel que soit le risque X . Prenons $X \sim \mathcal{U}(0, 1)$. Dans ce cas :

$$\begin{aligned} \text{ES}[X; \alpha] &= \int_\alpha^1 (1-x) dx = \frac{1}{2}(1-\alpha)^2 \\ &= \mathcal{R}_{g_\alpha}[X] = \int_0^1 g_\alpha(1-x) dx. \end{aligned} \quad (3.11)$$

Considérons à présent $Y \sim \mathcal{B}(q)$ pour $0 < q \leq 1-\alpha$. On obtient facilement

$$\text{ES}[Y; \alpha] = q = \mathcal{R}_{g_\alpha}[Y] = g_\alpha(q)$$

qui implique $g_\alpha(q) = q$ pour $0 < q \leq 1-\alpha$. En insérant cela dans (3.11) on obtient

$$\begin{aligned} \frac{1}{2}(1-\alpha)^2 &= \int_0^\alpha g_\alpha(1-x) dx + \int_\alpha^1 (1-x) dx \\ &\geq \frac{1}{2}(1-\alpha)^2 + \alpha(1-\alpha), \end{aligned}$$

ce qui nous mène à une contradiction puisque $0 < \alpha < 1$.

Nous pouvons formellement établir par l'absurde que la CTE n'est pas une mesure de risque de Wang. Il suffit en fait de procéder comme pour l'ES. Par l'absurde, supposons qu'il existe une fonction de distorsion g_α telle que $\text{CTE}[X; \alpha] = \mathcal{R}_{g_\alpha}[X]$ quel que soit le risque X . Pour $X \sim \mathcal{U}(0, 1)$ il vient à partir de (3.3)

$$\text{CTE}[X; \alpha] = \alpha + \frac{1}{2}(1-\alpha) = \int_0^1 g_\alpha(1-x) dx,$$

qui donne après simplification

$$\int_0^1 g_\alpha(x) dx = \frac{1}{2}(1+\alpha). \quad (3.12)$$

En passant à $Y \sim \mathcal{B}(q)$ et invoquant à nouveau (3.3), on obtient $\text{CTE}[Y; \alpha] = 1$ quels que soient q et α . Dès lors,

$$g_\alpha(q) = \mathcal{R}_{g_\alpha}[Y] = \text{CTE}[Y; \alpha] = 1$$

d'où l'on tire $g_\alpha(\cdot) \equiv 1$ sur $(0, 1]$, ce qui contredit l'équation (3.12) et achève de prouver que la CTE n'est pas une mesure de risque de Wang.

Exemple 3.5 En prenant

$$g(x) = 1 - (1 - x)^\xi, \quad \xi \geq 1,$$

nous obtenons

$$\mathcal{R}_g[X] = \int_{x \geq 0} (1 - \{F_X(x)\}^\xi) dx.$$

Si ξ est entier, $\mathcal{R}_g[X]$ peut être interprétée comme la valeur attendue du maximum $M_\xi = \max\{X_1, \dots, X_\xi\}$ d'un ensemble de ξ variables aléatoires indépendantes et de même loi que X . En effet, la fonction de queue de M_ξ est donnée par

$$\mathbb{P}[M_\xi > x] = 1 - \mathbb{P}[X_1 \leq x, \dots, X_\xi \leq x] = 1 - \{F_X(x)\}^\xi,$$

de sorte que $\mathcal{R}_g[X] = \mathbb{E}[M_\xi]$.

Exemple 3.6 Considérons la fonction de distorsion

$$g(x) = x^{1/\xi}, \quad \xi \geq 1.$$

La mesure de risque PH, introduite dans [47] et [46] est donnée par

$$\text{PH}_\xi[X] = \mathcal{R}_g[X] = \int_{x \geq 0} \{\bar{F}_X(x)\}^{1/\xi} dx.$$

Notons que pour $\xi = 1$, $\text{PH}_1[X] = \mathbb{E}[X]$. Nous reviendrons par la suite sur cette mesure de risques, liée à des propriétés de hasard proportionnel dans les modèles de durées.

Les principales propriétés des mesures de risque de Wang sont les suivantes :

Proposition 3.4 *Les mesures de risque de Wang sont*

- homogènes,
- invariantes par translation,
- monotones.

Preuve 3.8 En travaillant à partir de la Propriété 3.3, et en tenant compte du fait que la VaR est invariante par translation, on obtient :

$$\begin{aligned} \mathcal{R}_g[X + c] &= \int_0^1 \text{VaR}[X + c; 1 - \alpha] dg(\alpha) \\ &= \int_0^1 (\text{VaR}[X; 1 - \alpha] + c) dg(\alpha) \\ &= \mathcal{R}_g[X] + c(g(1) - g(0)) = \mathcal{R}_g[X]. \end{aligned}$$

Table 3.1 – Principales mesures de risque par distorsion (où p est une constante comprise entre 0 et 1).

mesure de risque \mathcal{R}	fonction de distorsion g
VaR	$g(x) = \mathbf{1}[x \geq p]$
TVaR	$g(x) = \min\{x/p, 1\}$
PH	$g(x) = x^p$
Dual Power	$g(x) = 1 - (1 - x)^{1/p}$
Gini	$g(x) = (1 + p)x - px^2$
Transformation exponentielle	$g(x) = (1 - p^x) / (1 - p)$

Ceci établit l'homogénéité. De la même manière, on déduit l'homogénéité et la monotonie des mesures de Wang des propriétés correspondantes de la VaR.

Si la fonction de distorsion g est concave, la fonction $x \mapsto g(\overline{F}_X(x))$ est continue à droite et est donc la fonction de queue d'une certaine variable aléatoire. En pareille situation $\mathcal{R}_g[X]$ est effectivement une espérance mathématique (non pas de X mais d'une variable Y dont la fonction de queue est $y \mapsto g(\overline{F}_X(y))$).

Proposition 3.5 *Lorsque la fonction de distorsion est concave, la mesure de risque correspondante est sous-additive.*

Corollaire 3.1 *En combinant les Propositions 3.4 et 3.5, on constate que les mesures de risque de Wang correspondant à des fonctions de distorsion concaves sont cohérentes.*

Le résultat suivant montre que la TVaR est la plus petite mesure de risque de Wang associée à une fonction de distorsion concave et excédant la VaR de même niveau de probabilité. En d'autres termes, la TVaR est la plus petite mesure de risque cohérente excédant la VaR.

Proposition 3.6 *Quel que soit le niveau de probabilité $0 < \alpha < 1$ et le risque X , nous avons :*

$$\text{TVaR}[X; \alpha] = \min \left\{ \mathcal{R}_g[X] \mid g \text{ est concave et } \mathcal{R}_g[X] \geq \text{VaR}[X; \alpha] \right\}. \quad (3.13)$$

Preuve 3.9 La TVaR est la mesure de risque de Wang associée à la fonction de distorsion concave $\min\left(\frac{x}{1-\alpha}, 1\right)$. De plus, $\text{TVaR}[X; \alpha] \geq \text{VaR}[X; \alpha]$. Ceci implique que

$$\text{TVaR}[X; \alpha] \geq \inf \left\{ \mathcal{R}_g[X] \mid g \text{ est concave et } \mathcal{R}_g[X] \geq \text{VaR}[X; \alpha] \right\}.$$

Afin d'établir l'inégalité contraire, considérons une fonction de distorsion concave g telle que

$$\mathcal{R}_g[Y] \geq \text{VaR}[Y; \alpha],$$

soit valable quelle que soit le risque Y . En prenant alors q tel que $1 - \alpha < q < 1$, et $Y_q \sim \mathcal{B}(q)$, il est facile de vérifier que $\text{VaR}[Y_q; \alpha] = 1$, et aussi $\mathcal{R}_g[Y_q] = g(q)$. Comme $g(x) \leq 1$, nous voyons que la condition $\mathcal{R}_g[Y_q] \geq \text{VaR}[Y_q; \alpha]$ peut se réécrire $g(q) = 1$. Cette condition doit être satisfaite pour tout q tel que $1 - \alpha < q < 1$. Ceci signifie que g vaut 1 sur l'intervalle $(1 - \alpha, 1]$. Comme g est concave, ceci entraîne

$$g(x) \geq \min\left(\frac{x}{1 - \alpha}, 1\right), \text{ pour } 0 < x < 1,$$

d'où l'on déduit

$$\mathcal{R}_g[X] \geq \text{TVaR}[X; \alpha].$$

Cette dernière inégalité est satisfaite quelle que soit la fonction de distorsion concave g telle que $\mathcal{R}_g[X] \geq \text{VaR}[X; \alpha]$. Ceci implique

$$\text{TVaR}[X; \alpha] \leq \inf \left\{ \mathcal{R}_g[X] \mid g \text{ est concave et } \mathcal{R}_g[X] \geq \text{VaR}[X; \alpha] \right\} \quad (3.14)$$

et achève la démonstration.

Définition 3.7 *La mesure de risque de Wang associée à la fonction de distorsion g , notée $\mathcal{R}_g[\cdot]$, est définie par :*

$$\mathcal{R}_g[X] = \int_0^\infty g(1 - F_X(x)) dx. \quad (3.15)$$

3.2.6 Les mesures de risque par distorsion

Les mesures de Wang sont en fait un cas très particulier des mesures de risque par distorsion, définies sur les variables aléatoires qui ne sont pas nécessairement positives. L'écriture étant toutefois plus simple sur la classe des mesures de Wang (et donc les démonstrations plus courtes) nous admettrons ici la plupart des propriétés énoncées.

Définition 3.5 *On appelle mesure de risque par distorsion la quantité :*

$$\mathcal{R}(X) = \int_0^1 F_X^{-1}(1 - u) dg(u)$$

où g est une fonction de répartition sur $[0, 1]$, appelée fonction de distorsion.

De même que pour les mesures de Wang, il est possible de réécrire cette expression sous la forme

Proposition 3.6 $\mathcal{R}(X)$ peut se réécrire :

$$\mathcal{R}(X) = \int_0^{+\infty} g(1 - F_X(x))dx - \int_{-\infty}^0 [1 - g(1 - F_X(x))]dx.$$

Preuve 3.10

$$\begin{aligned} \mathcal{R}(X) &= \int_0^1 F_X^{-1}(1 - u)dg(u) \\ &= \int_0^{F_X(0)} F_X^{-1}(1 - u)dg(u) + \int_{F_X(0)}^1 F_X^{-1}(1 - u)dg(u), \end{aligned}$$

ce qui donne, en faisant une intégration par parties, et un changement de variable, $u = F_X(x)$, une réécriture la mesure de risque sous la forme

$$\mathcal{R}(X) = \int_0^{+\infty} g(1 - F_X(x))dx - \int_{-\infty}^0 [1 - g(1 - F_X(x))]dx.$$

Remarque 3.9 Dans le cas particulier où g est la fonction de répartition de la loi uniforme sur $[0, 1]$, de telle sorte que g est la fonction identité sur $[0, 1]$:

$$\mathcal{R}(X) = \int_0^{+\infty} [1 - F_X(x)]dx - \int_{-\infty}^0 F_X(x)dx = \mathbb{E}(X).$$

Remarque 3.10 Soit \mathbb{Q} la mesure induite par la transformation g sur \mathbb{P} :

$$\mathbb{Q}([a, b]) = G \circ \mathbb{P}([a, b]).$$

La croissance sur g sur $[0, 1]$ permet de construire une capacité (telles qu'elles sont été introduite dans la partie 3 du chapitre précédent.

Exemple 3.7 Considérons la fonction de répartition $g(x) = x^k$. On appellera mesure de risque à hasard proportionnel la mesure induite par cette transformation,

$$\mathcal{R}_k(X) = \int_0^1 F_X^{-1}(1 - u)ku^{k-1}du.$$

Lorsque $k < 1$, la fonction g est concave.

Remarque 3.11 On peut montrer (voir [45]) que les mesures de distorsion de risque avec g concave vérifient la propriété de sous-additivité.

Remarque 3.12 Il existe dans la littérature ([1]) un autre type de mesures de risque, appelées mesures spectrales. Une fonction spectrale (parfois aussi appelée fonction d'aversion pour le risque) est $\phi : [0, 1] \rightarrow \mathbb{R}_+$, décroissante, telle que $\int_0^1 \phi(t)dt = 1$. La mesure de risque spectrale induite est

$$\mathcal{R}(X) = \int_0^1 F_X^{-1}(t)\phi(t)dt.$$

Ces mesures de risque sont cohérentes. En fait, les mesures de distorsion de fonction de distorsion g concave sont des mesures spectrales, avec $\phi = g'$.

3.3 Comparaison entre risques et mesures de risque

Comme l'a rappelé le chapitre précédant, il existe un lien fondamental entre la comparaison des risques, et les mesures de risque (nous reprendrons ici des résultats présentés dans [15]).

3.3.1 Ordre induit par la Value-at-Risk

Définition 3.8 *Etant donnés deux risques X et Y , X sera considéré moins dangereux que Y sur base de la comparaison des VaR, ce qui se notera $X \preceq_{\text{VaR}} Y$, lorsque*

$$\text{VaR}[X; \alpha] \leq \text{VaR}[Y; \alpha] \text{ pour tout } \alpha \in (0, 1).$$

La relation \preceq_{VaR} que nous venons de définir constitue un ordre partiel sur l'ensemble des lois de probabilité (on vérifie aisément qu'elle est réflexive, antisymétrique et transitive). Par abus de notation (comme cela se fait classiquement pour les ordres stochastiques), on la définit ici sur des variables aléatoire, bien qu'elle ne soit pas antisymétriques : $X \preceq_{\text{VaR}} Y$ et $Y \preceq_{\text{VaR}} X$ n'entraînent pas $X = Y$, mais seulement $X \stackrel{L}{=} Y$.

La relation \preceq_{VaR} a été étudiée depuis fort longtemps en probabilité et en statistique, depuis son introduction par [29]. Toutefois, elle est davantage connue sous le nom de « dominance stochastique à l'ordre 1 ». D'autres notations usuelles pour \preceq_{VaR} sont \preceq_{st} , \preceq_1 ou \preceq_{FSD} (étudiés en détails dans [42] ou [33]).

Nous établissons ci-dessous plusieurs conditions équivalentes pour comparer deux risques X et Y .

Proposition 3.7 *Etant données deux variables aléatoires X et Y :*

$$\begin{aligned} X \preceq_{\text{VaR}} Y & \quad \text{si et seulement si} \quad F_X(t) \geq F_Y(t) \text{ pour tout } t \in \mathbb{R}, \\ & \quad \text{si et seulement si} \quad \overline{F}_X(t) \leq \overline{F}_Y(t) \text{ pour tout } t \in \mathbb{R}. \end{aligned}$$

L'inégalité stochastique $X \preceq_{\text{VaR}} Y$ permet de déduire bon nombre d'inégalités intéressantes, comme en témoigne le résultat suivant :

Proposition 3.8 *Etant données deux variables aléatoires X et Y , $X \preceq_{\text{VaR}} Y$ si et seulement si $\mathbb{E}[g(X)] \leq \mathbb{E}[g(Y)]$ pour toute fonction croissante g , pour autant que les espérances existent.*

Preuve 3.11 Il suffit de se rappeler que si $U \sim \mathcal{U}(0, 1)$ alors $X \stackrel{\mathcal{L}}{=} \text{VaR}[X; U]$ en vertu de la Proposition 3.1. Fort de ce résultat, nous disposons de la représentation suivante, valable quelle que soit la fonction g :

$$\mathbb{E}[g(X)] = \mathbb{E}\left[g(\text{VaR}[X; U])\right] = \int_0^1 g(\text{VaR}[X; u]) du.$$

Le résultat annoncé s'obtient alors simplement en écrivant

$$\begin{aligned} \mathbb{E}[g(X)] &= \int_0^1 g(\text{VaR}[X; u]) du \\ &\leq \int_0^1 g(\text{VaR}[Y; u]) du = \mathbb{E}[g(Y)]. \end{aligned}$$

En fait, on peut montrer que l'on peut avoir un résultat similaire en se restreignant à aux fonctions dérivables et croissantes

Proposition 3.9 *Etant données deux variables aléatoires X et Y , $X \preceq_{\text{VaR}} Y$ si et seulement si $\mathbb{E}[g(X)] \leq \mathbb{E}[g(Y)]$ pour toute fonction g telle que $g' \geq 0$, pour autant que les espérances existent.*

La Propriété 3.7 nous apprend que $X \preceq_{\text{VaR}} Y$ lorsque les fonctions de survie de X et de Y se dominent mutuellement. Le résultat suivant fournit une condition suffisante pour que cela ait lieu. Il suffit en fait que les densités de probabilité ne se croisent qu'une seule fois.

Proposition 3.10 *Quels que soient les risques X et Y , s'il existe $c \in \mathbb{R}$ tel que $f_X(t) \geq f_Y(t)$ pour $t < c$ et $f_X(t) \leq f_Y(t)$ pour $t > c$ alors $X \preceq_{\text{VaR}} Y$.*

Preuve 3.12 Il suffit de remarquer que si $x < c$,

$$F_X(x) = \int_0^x f_X(t) dt \geq \int_0^x f_Y(t) dt = F_Y(x),$$

et si $x > c$

$$\bar{F}_X(x) = \int_x^{+\infty} f_X(t) dt \geq \int_x^{+\infty} f_Y(t) dt = \bar{F}_Y(x).$$

Nous avons donc $X \preceq_{\text{VaR}} Y$ sous les conditions énoncées plus haut, ce qui achève la démonstration.

Nous pourrions nous intéresser à la comparaison des risques sachant qu'ils excèdent un certain niveau t . Ainsi, on pourrait exiger que

$$[X|X > t] \preceq_{\text{VaR}} [Y|Y > t]$$

quel que soit le niveau t . L'exemple suivant montre que cela n'est pas forcément vrai lorsque $X \preceq_{\text{VaR}} Y$.

Exemple 3.3 Montrons que

$$X \preceq_{\text{VaR}} Y \not\Rightarrow [X|X > t] \preceq_{\text{VaR}} [Y|Y > t] \text{ pour tout } t.$$

Considérons par exemple le cas où $X \sim \mathcal{U}(0, 3)$ et Y possède la densité

$$f_Y(x) = \frac{1}{6}\mathbf{1}_{]0,1[}(x) + \frac{1}{2}\mathbf{1}_{]1,2[}(x) + \frac{1}{3}\mathbf{1}_{]2,3[}(x).$$

Alors $X \preceq_{\text{VaR}} Y$, mais $[X|X > 1] \sim \mathcal{U}(1, 3)$ et $[Y|Y > 1]$ possède la densité

$$f_Y^*(x) = \frac{3}{5}\mathbf{1}_{]1,2[}(x) + \frac{2}{5}\mathbf{1}_{]2,3[}(x)$$

de telle sorte que

$$[Y|Y > 1] \preceq_{\text{VaR}} [X|X > 1].$$

On peut prouver le résultat suivant.

Proposition 3.11 *Etant donnés deux risques X et Y , $[X|X > t] \preceq_{\text{VaR}} [Y|Y > t]$ quel que soit $t \in \mathbb{R}$*

si et seulement si $t \mapsto \frac{\overline{F}_Y(t)}{\overline{F}_X(t)}$ est croissante

si et seulement si $\overline{F}_X(u)\overline{F}_Y(v) \geq \overline{F}_X(v)\overline{F}_Y(u)$ quels que soient $u \leq v$.

On peut rapprocher cette comparaison des taux de hasard, comme le montre le résultat suivant.

Proposition 3.12 *Etant donnés deux risques X et Y , $[X|X > t] \preceq_{\text{VaR}} [Y|Y > t]$ quel que soit $t \in \mathbb{R}$ si, et seulement si $r_X(t) \geq r_Y(t)$ quel que soit t , où $r(t) = -(\log \overline{F}(t))'$.*

Preuve 3.13 Le rapport $\overline{F}_Y/\overline{F}_X$ est non-décroissant si, et seulement si, $\log \overline{F}_Y/\overline{F}_X$ l'est également. Ceci revient encore à imposer que $\log \overline{F}_Y - \log \overline{F}_X$ soit non-décroissant, ce qui revient à $r_X - r_Y > 0$ puisque

$$\frac{d}{dt} \left\{ \log \overline{F}_Y(t) - \log \overline{F}_X(t) \right\} = r_X(t) - r_Y(t).$$

Remarque 3.4 La méthode de comparaison de lois de probabilité dont il est question dans les Propositions 3.11-3.12 est souvent appelée l'ordre du taux de hasard (hazard rate order) et notée \preceq_{hr} .

Supposons qu'on déflate le taux de hasard d'un facteur ξ , c'est à dire que l'on passe d'un taux r_X à un taux r_{X_ξ} donné par

$$r_{X_\xi}(t) = \frac{r_X(t)}{\xi} \leq r_X(t), \text{ pour } \xi \geq 1,$$

nous obtenons une fonction de queue

$$\bar{F}_{X_\xi}(t) = \exp\left(-\int_0^t \frac{r_X(s)}{\xi} ds\right) = \{\bar{F}_X(t)\}^{1/\xi}.$$

Dès lors, $\text{PH}_\xi[X] = \mathbb{E}[X_\xi]$. Le mesure de risque PH consiste donc à remplacer le risque initial X par un risque transformé X^* dont le taux de hasard a été déflaté, et à calculer ensuite l'espérance associée à X^* . Nous avons

$$[X|X > t] \preceq_{\text{VaR}} [X^*|X^* > t] \text{ quel que soit } t > 0.$$

3.3.2 Rapport de vraisemblance et principe d'Esscher

Nous pourrions encore songer à imposer

$$[X|a \leq X \leq a+h] \preceq_{\text{VaR}} [Y|a \leq Y \leq a+h]$$

quel que soit le niveau a et l'accroissement $h > 0$. Ceci correspond à la situation d'un réassureur qui aurait à couvrir la tranche $(a, a+h]$ d'un risque X , c'est à dire qu'il s'exposerait à une perte de

$$X_{(a, a+h]} = \begin{cases} 0 & \text{si } X < a \\ X - a & \text{si } a \leq X < a+h \\ h & \text{si } a+h \leq X, \end{cases}$$

où a est la rétention et h la portée. On peut établir le résultat suivant.

Proposition 3.4 *Considérons les variables aléatoires X et Y , toutes deux continues ou discrètes, possédant les fonctions de densité f_X et f_Y . Si*

$$\frac{f_X(t)}{f_Y(t)} \text{ décroît sur l'union des supports de } X \text{ et de } Y \quad (3.16)$$

(en prenant par convention $a/0$ égal à $+\infty$ lorsque $a > 0$), ou, de manière équivalente, si

$$f_X(u) \cdot f_Y(v) \geq f_X(v) \cdot f_Y(u) \text{ pour tout } u \leq v. \quad (3.17)$$

Alors, $[X|a \leq X \leq a+h] \preceq_{\text{VaR}} [Y|a \leq Y \leq a+h]$ quel que soit le niveau a et l'accroissement $h > 0$

Preuve 3.14 Considérons $a < b$. L'inégalité stochastique $[X|a \leq X \leq b] \preceq_{\text{VaR}} [Y|a \leq Y \leq b]$ garantit que

$$\frac{\mathbb{P}[u \leq X \leq b]}{\mathbb{P}[a \leq X \leq b]} \leq \frac{\mathbb{P}[u \leq Y \leq b]}{\mathbb{P}[a \leq Y \leq b]} \text{ lorsque } u \in [a, b].$$

Il suit alors

$$\frac{\mathbb{P}[a \leq X < u]}{\mathbb{P}[u \leq X \leq b]} \geq \frac{\mathbb{P}[a \leq Y < u]}{\mathbb{P}[u \leq Y \leq b]} \text{ lorsque } u \in [a, b].$$

C'est-à-dire

$$\frac{\mathbb{P}[a \leq X < u]}{\mathbb{P}[a \leq Y < u]} \geq \frac{\mathbb{P}[u \leq X \leq b]}{\mathbb{P}[u \leq Y \leq b]} \text{ lorsque } u \in [a, b].$$

En particulier, pour $u < b \leq v$,

$$\frac{\mathbb{P}[u \leq X < b]}{\mathbb{P}[u \leq Y < b]} \geq \frac{\mathbb{P}[b \leq X \leq v]}{\mathbb{P}[b \leq Y \leq v]}.$$

Dès lors, lorsque X et Y sont continues,

$$\frac{\mathbb{P}[a \leq X < u]}{\mathbb{P}[a \leq Y < u]} \geq \frac{\mathbb{P}[b \leq X \leq v]}{\mathbb{P}[b \leq Y \leq v]} \text{ lorsque } a < u \leq b \leq v.$$

Si nous passons à la limite pour $a \rightarrow u$ et $b \rightarrow v$ nous obtenons (3.17). La preuve dans le cas discret est similaire.

Les conditions (3.16) et (3.17), apparemment techniques et peu intuitives, sont généralement faciles à établir dans les modèles paramétriques.

Remarque 3.5 La méthode de comparaison de lois de probabilité dont il est question dans la Proposition 3.4 est encore appelée ordre du rapport de vraisemblance, et notée \preceq_{lr} .

Notons X_h la transformée d'Esscher de X . Le rapport des densités de probabilités associées à X et X_h est proportionnel à $\exp(-hx)$, qui est clairement décroissant en x . Ceci indique que

$$[X|a \leq X \leq b] \preceq_{\text{VaR}} [X_h|a \leq X_h \leq b],$$

quels que soient $a < b$.

Le résultat suivant montre sous quelles conditions les mesures de risque d'Esscher relatives à deux risques X et Y sont uniformément ordonnées.

Proposition 3.13 *Si $[X|a \leq X \leq b] \preceq_{\text{VaR}} [Y|a \leq Y \leq b]$ quels que soient $a < b$ alors $\text{Es}[X; h] \leq \text{Es}[Y; h]$ pour tout $h > 0$.*

Preuve 3.15 Nous savons de part la Proposition 3.4 que l'inégalité

$$f_X(u)f_Y(v) \geq f_X(v)f_Y(u)$$

est satisfaite pour tout $u \leq v$. En multipliant les deux membres de cette inégalité par

$$\frac{\exp(hu)}{M_X(h)} \frac{\exp(hv)}{M_Y(h)}$$

on obtient la même inégalité pour les fonctions de densité de X_h et Y_h , d'où on tire que $[X_h|a \leq X_h \leq b] \preceq_{\text{VaR}} [Y_h|a \leq Y_h \leq b]$, ce qui donne le résultat annoncé.

3.3.3 Comparaison uniforme des TVaR

Nous introduisons ici une nouvelle méthode de comparaison des risques, basée sur les TVaR.

Définition 3.9 *Quelles que soient les variables aléatoires X et Y de moyenne finie, X sera considéré moins dangereux que Y sur base de la comparaison des TVaR, ce qui se notera dorénavant $X \preceq_{\text{TVaR}} Y$, lorsque*

$$\text{TVaR}[X; \alpha] \leq \text{TVaR}[Y; \alpha] \text{ pour tout } \alpha \in [0, 1].$$

La restriction à des risques de moyenne finie garantit l'existence des TVaR, et donc la correction de la définition de \preceq_{TVaR} . Dorénavant, nous ne comparons donc que des risques dont la prime pure est finie (il est bon de noter que de nombreux résultats que nous établirons dans la suite de cette section dépendent de cette hypothèse). Cette restriction distingue également \preceq_{VaR} et \preceq_{TVaR} : en effet, \preceq_{VaR} est définie quelles que soient les variables en présence, ce qui n'est pas le cas pour \preceq_{TVaR} .

La relation \preceq_{TVaR} est très ancienne (elle est l'héritière de la relation dite de majorization entre vecteurs numériques, étudiée dans les années 1930, qui sera abordée dans le chapitre 6). Les actuaires l'appellent encore ordre stop-loss (noté \preceq_{sl}), notion intimement liée à la dominance stochastique du deuxième ordre des économistes (souvent notée \preceq_2 ou \preceq_{SSD}). La relation \preceq_{TVaR} est mieux connue des probabilistes sous le nom d'ordre convexe croissant (noté \preceq_{icx}).

Nous utiliserons encore la relation notée \preceq_{CX} , qui restreint \preceq_{TVaR} aux couples de variables aléatoires de même moyenne.

Définition 3.10 *Quelles que soient les variables aléatoires X et Y ,*

$$X \preceq_{CX} Y \quad \begin{array}{l} \text{si et seulement si} \\ \text{si et seulement si} \end{array} \quad \begin{cases} \mathbb{E}[X] = \mathbb{E}[Y], \\ X \preceq_{\text{TVaR}} Y, \\ \text{TVaR}[X; 0] = \text{TVaR}[Y; 0], \\ \text{TVaR}[X; \alpha] \leq \text{TVaR}[Y; \alpha] \text{ pour tout } \alpha \in (0, 1). \end{cases}$$

La relation \preceq_{CX} est connue sous le nom d'ordre convexe parmi les probabilistes (noté généralement \preceq_{cx}). Elle est intimement liée à d'autres relations, comme l'ordre de Lorenz.

Remarque 3.6 Notons que \preceq_{VaR} ne permettait pas de comparer des variables aléatoires de même moyenne. En effet,

$$\left. \begin{array}{l} X \preceq_{\text{VaR}} Y \\ \mathbb{E}[X] = \mathbb{E}[Y] \end{array} \right\} \Rightarrow X \stackrel{\mathcal{L}}{=} Y.$$

Afin de prouver ce résultat, il suffit d'examiner l'identité

$$\mathbb{E}[Y] - \mathbb{E}[X] = \int_{x=0}^{+\infty} \underbrace{\{\mathbb{P}[Y > x] - \mathbb{P}[X > x]\}}_{\geq 0 \text{ pour tout } x} dx = 0,$$

qui entraîne $\mathbb{P}[Y > x] = \mathbb{P}[X > x]$ pour tout x et achève la vérification.

TVaR et primes stop-loss

Les relations \preceq_{TVaR} et \preceq_{CX} introduites ci-dessus peuvent encore être interprétées en termes de primes stop-loss, comme le montrent les résultats suivants.

Proposition 3.5 *Soient deux variables aléatoires X et Y de même moyenne finie. Alors, chacune des deux inégalités suivantes est équivalente à $\mathbb{E}[(X - t)_+] \leq \mathbb{E}[(Y - t)_+]$ pour tout $t \in \mathbb{R}$:*

$$\int_0^p F_X^{-1}(u) du \geq \int_0^p F_Y^{-1}(u) du \text{ pour tout } p \in [0, 1]; \quad (3.18)$$

et

$$\int_p^1 F_X^{-1}(u) du \leq \int_p^1 F_Y^{-1}(u) du \text{ pour tout } p \in [0, 1]. \quad (3.19)$$

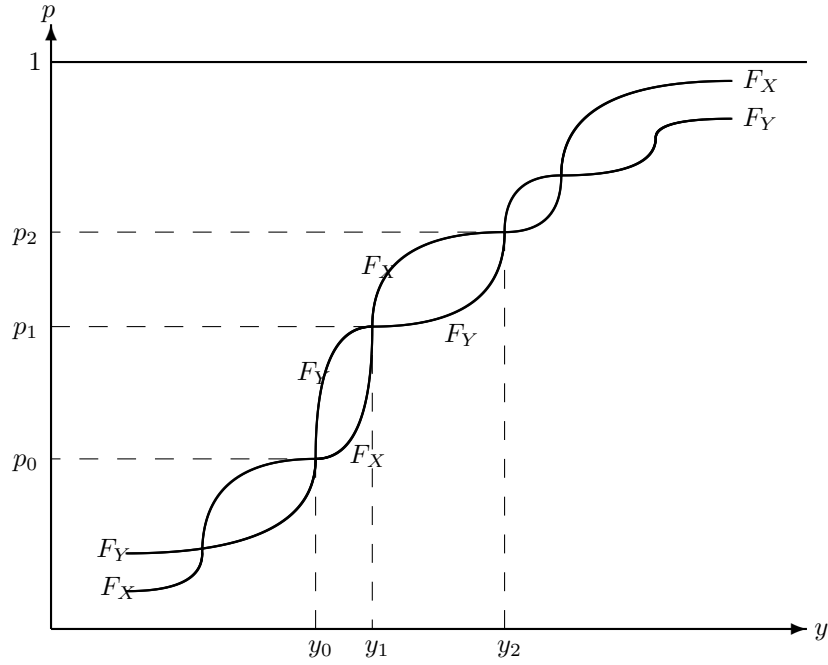
Preuve 3.16 De l'égalité des moyennes et du fait que

$$\mathbb{E}[X] = \int_0^1 F_X^{-1}(u) du \text{ et } \mathbb{E}[Y] = \int_0^1 F_Y^{-1}(u) du$$

on déduit que pour tout $p \in [0, 1]$, les inégalités (3.18) et (3.19) sont équivalentes. Ainsi, il suffit de montrer que $\mathbb{E}[(X - t)_+] \leq \mathbb{E}[(Y - t)_+]$ pour tout $t \in \mathbb{R}$ si, et seulement si, (3.18) est satisfaite.

Nous démontrons le résultat pour des fonctions de répartition continues qui se croisent à un nombre fini de reprises ; le raisonnement est similaire dans le cas général, mais techniquement beaucoup plus lourd. De l'égalité des moyennes, on déduit que F_X et F_Y doivent se croiser au moins une fois. Que les primes stop-loss soient uniformément ordonnées ou que (3.18) soit vérifiée, F_X doit être inférieure à F_Y avant ce premier croisement et, de la même manière, F_X doit dominer F_Y après le dernier croisement.

Notons (y_0, p_0) , (y_1, p_1) , et (y_2, p_2) trois points de croisement consécutifs tels que représentés à la Figure 3.1. Il est bon de garder à l'esprit que (y_0, p_0)

Figure 3.1 – Fonctions de répartition F_X et F_Y

peut être $(-\infty, 0)$ (avec la convention $0 \cdot (-\infty) \equiv 0$), et que (y_2, p_2) peut être $(\infty, 1)$ (avec la convention $0 \cdot \infty \equiv 0$). L'hypothèse de continuité garantit que

$$p_i = F_X(y_i) = F_Y(y_i), \text{ pour } i = 0, 1, 2.$$

Supposons que $\mathbb{E}[(X - t)_+] \leq \mathbb{E}[(Y - t)_+]$ pour tout $t \in \mathbb{R}$. Alors,

$$\int_{y_2}^{\infty} \bar{F}_X(x) dx \leq \int_{y_2}^{\infty} \bar{F}_Y(x) dx. \quad (3.20)$$

Donc

$$\begin{aligned} \int_{p_2}^1 F_X^{-1}(u) du &= y_2(1 - p_2) + \int_{y_2}^{\infty} \bar{F}_X(x) dx \\ &\leq y_2(1 - p_2) + \int_{y_2}^{\infty} \bar{F}_Y(x) dx \text{ par (3.20)} \\ &= \int_{p_2}^1 F_Y^{-1}(u) du. \end{aligned} \quad (3.21)$$

A présent, nous avons pour $u \in [p_1, p_2]$ que $F_X^{-1}(u) - F_Y^{-1}(u) \leq 0$. Ainsi, $\int_p^1 (F_X^{-1}(u) - F_Y^{-1}(u)) du$ est croissante en $p \in [p_1, p_2]$. De ce fait, nous déduisons

de (3.21) que

$$\int_p^1 F_X^{-1}(u) du \leq \int_p^1 F_Y^{-1}(u) du \text{ pour } p \in [p_1, p_2]. \quad (3.22)$$

De l'inégalité des primes stop-loss, nous déduisons également

$$\int_{-\infty}^{y_0} F_X(x) dx \leq \int_{-\infty}^{y_0} F_Y(x) dx. \quad (3.23)$$

De ce fait,

$$\begin{aligned} \int_0^{p_0} F_X^{-1}(u) du &= y_0 p_0 - \int_{-\infty}^{y_0} F_X(x) dx \\ &\geq y_0 p_0 - \int_{-\infty}^{y_0} F_Y(x) dx \text{ par (3.23)} \\ &= \int_0^{p_0} F_Y^{-1}(u) du. \end{aligned} \quad (3.24)$$

Comme pour $u \in [p_0, p_1]$ nous avons $F_X^{-1}(u) - F_Y^{-1}(u) \geq 0$, il vient que $\int_0^p (F_X^{-1}(u) - F_Y^{-1}(u)) du$ est croissante en $p \in [p_0, p_1]$. L'équation (3.24) donne alors

$$\int_0^p F_X^{-1}(u) du \geq \int_0^p F_Y^{-1}(u) du \text{ pour } p \in [p_0, p_1]. \quad (3.25)$$

Nous voyons donc grâce à (3.22) et (3.25) que pour chaque $p \in [0, 1]$ l'une des relations (3.18) ou (3.19) doit être vérifiée.

Réciproquement, supposons à présent que (3.18) (ou de manière équivalente (3.19)) est vraie. Alors,

$$\int_{p_2}^1 F_X^{-1}(u) du \leq \int_{p_2}^1 F_Y^{-1}(u) du. \quad (3.26)$$

De même,

$$\begin{aligned} \int_{y_2}^{\infty} \bar{F}_X(x) dx &= \int_{p_2}^1 F_X^{-1}(u) du - y_2(1 - p_2) \\ &\leq \int_{p_2}^1 F_Y^{-1}(u) du - y_2(1 - p_2) \text{ par (3.26)} \\ &= \int_{y_2}^{\infty} \bar{F}_Y(x) dx. \end{aligned} \quad (3.27)$$

Pour $x \in [y_1, y_2]$ nous avons $\bar{F}_X(x) - \bar{F}_Y(x) \leq 0$. De ce fait, $\int_y^{\infty} (\bar{F}_X(x) - \bar{F}_Y(x)) dx$ est croissante en $y \in [y_1, y_2]$. Dès lors, nous déduisons de (3.27) que

$$\int_y^{\infty} \bar{F}_X(x) dx \leq \int_y^{\infty} \bar{F}_Y(x) dx \text{ pour } y \in [y_1, y_2]. \quad (3.28)$$

On déduit encore de (3.18) que

$$\int_0^{p_0} F_X^{-1}(u) du \geq \int_0^{p_0} F_Y^{-1}(u) du. \quad (3.29)$$

Donc,

$$\begin{aligned} \int_{-\infty}^{y_0} F_X(x) dx &= y_0 p_0 - \int_0^{p_0} F_X^{-1}(u) du \\ &\leq y_0 p_0 - \int_0^{p_0} F_Y^{-1}(u) du \quad \text{par (3.29)} \\ &= \int_{-\infty}^{y_0} F_Y(x) dx. \end{aligned} \quad (3.30)$$

Pour $x \in [y_0, y_1]$ nous avons $F_X(x) - F_Y(x) \leq 0$, et donc $\int_{-\infty}^y (F_X(x) - F_Y(x)) dx$ est décroissant en $y \in [y_0, y_1]$. On déduit alors de (3.30) que

$$\int_{-\infty}^y F_X(x) dx \leq \int_{-\infty}^y F_Y(x) dx \text{ pour } y \in [y_0, y_1]. \quad (3.31)$$

Nous pouvons alors conclure car nous voyons à partir de (3.28) et (3.31) que les primes stop-loss doivent être uniformément ordonnées.

Corollaire 3.2 *Soient X et Y deux risques de même moyenne μ finie. Alors,*

$$X \preceq_{CX} Y \text{ si et seulement si } \mathbb{E}[(X - t)_+] \leq \mathbb{E}[(Y - t)_+] \text{ pour tout } t \in \mathbb{R}.$$

On peut montrer qu'un résultat du même type vaut pour \preceq_{TVaR} , à savoir

$$X \preceq_{TVaR} Y \text{ si et seulement si } \mathbb{E}[(X - t)_+] \leq \mathbb{E}[(Y - t)_+] \text{ pour tout } t \in \mathbb{R}.$$

Nous savons à présent que les relations \preceq_{CX} et \preceq_{TVaR} peuvent s'interpréter à l'aide des primes de réassurance relatives à un traité stop-loss. Une comparaison au sens \preceq_{TVaR} ou \preceq_{CX} revient à comparer les primes stop-loss de différentes rétentions. On s'en remet donc aux réassureurs lorsqu'il s'agit de choisir entre deux risques X et Y : ceux-ci sont interrogés quant à leur tarif stop-loss pur et la préférence va au risque dont les primes de réassurance sont toujours plus avantageuses, à moyenne constante.

TVaR et fonctions convexes

Tout comme \preceq_{VaR} était liée aux fonctions croissantes, les relations \preceq_{CX} et \preceq_{TVaR} sont liées aux fonctions convexes et convexes croissantes, comme le montrent les résultats suivants.

Proposition 3.14 *Etant données deux variables aléatoires X et Y de moyennes finies, on a*

1. $X \preceq_{CX} Y$ si et seulement si $\mathbb{E}[g(X)] \leq \mathbb{E}[g(Y)]$ pour toute fonction convexe g , pour autant que les espérances existent.
2. $X \preceq_{CX} Y$ si et seulement si $\mathbb{E}[g(X)] \leq \mathbb{E}[g(Y)]$ pour toute fonction g telle que $g'' \geq 0$, pour autant que les espérances existent.

Preuve 3.17 Prouvons (uniquement) 1. L'implication \Rightarrow est évidente puisque la fonction $x \mapsto (x - t)_+$ est convexe quel que soit $t \in \mathbb{R}^+$. Si on veut prouver la réciproque \Leftarrow , notons que toute fonction convexe g est la limite uniforme d'une suite $\{g_n, n = 1, 2, \dots\}$ de fonctions convexes définies par

$$g_n(x) = \alpha_1^{(n)} + \alpha_2^{(n)}x + \sum_{j=0}^n \beta_j^{(n)}(x - t_j^{(n)})_+,$$

où $\alpha_1^{(n)}, \alpha_2^{(n)} \in \mathbb{R}$, $\beta_j^{(n)} \geq 0$ pour $j = 0, 1, \dots, n$ et $0 < t_0^{(n)} < t_1^{(n)} < \dots < t_n^{(n)}$. Dès lors, comme l'inégalité

$$\begin{aligned} \mathbb{E}[g_n(X)] &= \alpha_1^{(n)} + \alpha_2^{(n)}\mathbb{E}[X] + \sum_{j=0}^n \beta_j^{(n)}\mathbb{E}[(X - t_j^{(n)})_+] \\ &\leq \alpha_1^{(n)} + \alpha_2^{(n)}\mathbb{E}[X] + \sum_{j=0}^n \beta_j^{(n)}\mathbb{E}[(Y - t_j^{(n)})_+] = \mathbb{E}[g_n(Y)] \end{aligned}$$

est satisfaite quel que soit n , il vient $\mathbb{E}[g(X)] \leq \mathbb{E}[g(Y)]$ en passant à la limite, ce qui achève la démonstration de (1).

En procédant de manière analogue, on peut encore établir le résultat suivant.

Proposition 3.15 *Etant données deux variables aléatoires X et Y de moyennes finies, on a*

1. $X \preceq_{TVaR} Y$ si et seulement si $\mathbb{E}[g(X)] \leq \mathbb{E}[g(Y)]$ pour toute fonction convexe croissante g , pour autant que les espérances existent.
2. $X \preceq_{TVaR} Y$ si et seulement si $\mathbb{E}[g(X)] \leq \mathbb{E}[g(Y)]$ pour toute fonction g telle que $g' \geq 0$ et $g'' \geq 0$, pour autant que les espérances existent.

Remarque 3.7 De manière heuristique, les fonctions convexes sont celles qui prennent leurs valeurs les plus élevées dans des régions de la forme $]-\infty, a[\cup]b, +\infty[$, avec $a < b$. Pour cette raison, si $X \preceq_{CX} Y$ ou $X \preceq_{TVaR} Y$, Y prend plus souvent des valeurs extrêmes que X , c'est-à-dire que Y est plus variable que X . Ceci peut encore se justifier par le fait que si $X \preceq_{CX} Y$, alors nécessairement

$$\text{VaR}[X] \leq \text{VaR}[Y],$$

car la fonction $x \mapsto g(x) = x^2$ est convexe.

La caractérisation suivante a été obtenue par Strassen en 1965 ([44]). Elle montre que les TVaR sont diminuées par passage à l'espérance conditionnelle.

Proposition 3.6 *Etant données deux risques X et Y , $X \preceq_{CX} Y$ si et seulement si, on peut trouver deux variables aléatoires \tilde{X} et \tilde{Y} (définies sur le même espace probabilisé) telles que $\tilde{X} \stackrel{L}{=} X$, $\tilde{Y} \stackrel{L}{=} Y$ et $\mathbb{E}[\tilde{Y}|\tilde{X}] = \tilde{X}$.*

Preuve 3.18 La construction des variables \tilde{X} et \tilde{Y} (correspondant à la preuve de la nécessité de la condition de Strassen) est difficile ; nous renvoyons le lecteur intéressé à [33] pour une démonstration intéressante. La suffisance est quant à elle facile à établir. En effet, si de telles variables \tilde{X} et \tilde{Y} existent alors quelle que soit la fonction convexe g ,

$$\begin{aligned} \mathbb{E}[g(X)] &= \mathbb{E}[g(\tilde{X})] = \mathbb{E}\left[g\left(\mathbb{E}[\tilde{Y}|\tilde{X}]\right)\right] \\ &\leq \mathbb{E}\left[\left(\mathbb{E}[g(\tilde{Y})|\tilde{X}]\right)\right] = \mathbb{E}[g(\tilde{Y})] = \mathbb{E}[g(Y)] \end{aligned}$$

grâce à l'inégalité de Jensen.

Corollaire 3.3 *En particulier, la superposition d'un aléa supplémentaire centré ε à un risque X rend la situation plus risquée (car plus variable, en fait). Formellement, quel que soit le risque X et la variable aléatoire ε indépendante de X telle que $\mathbb{E}[\varepsilon] = 0$,*

$$X \preceq_{CX} X + \varepsilon.$$

Nous donnons ici des conditions suffisantes pour \preceq_{CX} et \preceq_{TVaR} , basées sur le nombre de croisement des fonctions de répartition.

Proposition 3.16 *Soient X et Y deux variables aléatoires telles que $\mathbb{E}[X] = \mathbb{E}[Y]$. S'il existe une constante c telle que*

$$F_Y(x) \geq F_X(x) \text{ pour tout } x < c,$$

et

$$F_Y(x) \leq F_X(x) \text{ pour tout } x > c,$$

alors $X \preceq_{CX} Y$.

Preuve 3.19 Pour $t > c$, on a

$$\begin{aligned} \mathbb{E}[(X - t)_+] &= \int_{x=t}^{+\infty} \mathbb{P}[X > x] dx \\ &\leq \int_{x=t}^{+\infty} \mathbb{P}[Y > x] dx = \mathbb{E}[(Y - t)_+]. \end{aligned}$$

Pour $t \leq c$, il suffit de remarquer que

$$\begin{aligned}\mathbb{E}[(X - t)_+] &= \mathbb{E}[X] - \int_{x=0}^t \mathbb{P}[X > x] dx \\ &= \mathbb{E}[X] - \int_{x=0}^t \mathbb{P}[Y > x] dx,\end{aligned}$$

d'où le résultat puisque $\mathbb{E}[X] = \mathbb{E}[Y]$.

De la même manière, on peut prouver le résultat suivant.

Proposition 3.17 *Soient X et Y deux variables aléatoires telles que $\mathbb{E}[X] \leq \mathbb{E}[Y]$. S'il existe une constante c telle que*

$$F_Y(x) \geq F_X(x) \text{ pour tout } x < c,$$

et

$$F_Y(x) \leq F_X(x) \text{ pour tout } x > c,$$

alors $X \preceq_{TVaR} Y$.

3.4 Estimation de la Value-at-Risk

La Value-at-Risk jouant un rôle central dans la construction des mesures de risque, il est intéressant d'avoir de bons estimateurs de cette quantité. Nous verrons alors ensuite comment en déduire des estimateurs d'autres mesures de risque. Dans cette section, nous présenterons les méthodes paramétriques usuelles (en insistant sur les quantiles gaussiens, qui servent toujours de base dans la modélisation financière), puis nonparamétriques, et nous évoquerons également les estimateurs construits à l'aide de la théorie des valeurs extrêmes (mais qui seront détaillés dans les chapitres suivants).

3.4.1 Estimation paramétrique

En finance de marché, une modélisation classique des rendements consiste à faire une hypothèse de normalité. Dans ce cas, si $X \sim \mathcal{N}(\mu, \sigma)$, alors rappelons que le quantile de niveau α est donné par

$$Q(X; \alpha) = \mu + \Phi^{-1}(\alpha) \cdot \sigma,$$

où Φ est la fonction de répartition de la loi normale centrée réduite, et où $\Phi^{-1}(\alpha)$ est tabulé, avec 1,64 si $\alpha = 90\%$, 1,96 si $\alpha = 95\%$, ou encore 2,57 si $\alpha = 99.5\%$.

Exemple 3.8 Cette approche gaussienne est particulièrement intéressante pour modéliser des risques agrégés, en particulier si l'on se base sur des modèles à facteurs. Supposons que X_1, \dots, X_d soient d risques et que

$$X_i = a_i + b_i Z + u_i,$$

où les bruits u_i sont supposés indépendants entre eux, et de Z , avec $u_i \sim \mathcal{N}(0, \sigma_i^2)$, avec $Z \sim \mathcal{N}(0, 1)$, alors $X = X_1 + \dots + X_d$ est gaussien. De plus, les VaR pour chacun des risques sont

$$\text{VaR}_i(\alpha) = a_i + \Phi^{-1}(\alpha) \sqrt{b_i^2 + \sigma_i^2}$$

mais surtout

$$\text{VaR}(\alpha) = a + \Phi^{-1}(\alpha) \sqrt{b^2 + \sigma^2}$$

où $a = a_1 + \dots + a_d$, $b = b_1 + \dots + b_d$, et $\sigma^2 = \sigma_1^2 + \dots + \sigma_d^2$. On notera que

$$\text{VaR}(\alpha) - \sum_{i=1}^d \text{VaR}_i(\alpha) = \Phi^{-1}(\alpha) \left(\sqrt{b^2 + \sigma^2} - \sum_{i=1}^d \sqrt{b_i^2 + \sigma_i^2} \right)$$

où le terme entre parenthèses est négatif.

Exemple 3.9 Le modèle Gaussien peut aussi être utilisé pour dériver des bornes supérieures de quantiles, dès lors que les lois sont de variance finie. En effet, pour des lois Z de variance finie, l'inégalité de Bienaymé-Tchebychev permet d'écrire

$$\mathbb{P}(|Z - \mathbb{E}(Z)| > z) \leq \frac{\text{Var}(Z)}{z^2}$$

Aussi, si X suit une loi centrée de variance σ^2 ,

$$\mathbb{P}(X > z\sigma) \leq \frac{1}{2z^2}$$

donc, à α donné, en choisissant z de telle sorte que $z = (2(1-\alpha))^{-\frac{1}{2}}$, on obtient une borne supérieure pour $\text{VaR}(X; \alpha)$: en écrivant

$$\mathbb{P}\left(X > \frac{\sigma}{\sqrt{2(1-\alpha)}}\right) \leq \alpha,$$

on en déduit que $\text{VaR}^+(X, \alpha) = (2(1-\alpha))^{-\frac{1}{2}} \cdot \sigma$ est une borne supérieure pour $\text{VaR}(X, \alpha)$, si X est de variance finie. Or pour une loi $\mathcal{N}(0, \sigma^2)$, on connaît précisément le lien entre le quantile et σ . Aussi, en utilisant les valeurs tabulées de la loi normale centrée réduite,

$$\text{VaR}^+(X, 99\%) = 3.04 \cdot \sigma \cdot \text{VaR}(Z, 99\%) \text{ où } Z \sim \mathcal{N}(0, 1),$$

et

$$\text{VaR}^+(X, 95\%) = 1.92 \cdot \sigma \cdot \text{VaR}(Z, 95\%) \text{ où } Z \sim \mathcal{N}(0, 1),$$

ce qui permet de comprendre pourquoi, dans la littérature réglementaire, il n'est pas rare de voir apparaître des facteurs multiplicatif de l'ordre de 2 ou 3 devant les quantiles obtenus dans les modèles Gaussiens.

Définition 3.6 *Etant donné un échantillon $\{X_1, \dots, X_n\}$, l'estimateur gaussien du quantile de niveau α est*

$$\hat{q}_n(\alpha) = \hat{\mu} + \Phi^{-1}(\alpha)\hat{\sigma}, \text{ où } \hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i \text{ et } \hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{\mu})^2}.$$

Toutefois, comme le modèle gaussien est rarement pertinent, il peut être intéressant d'utiliser des approximations un peu plus fine, en particulier celle proposée par [9], appelée approximation de Cornish-Fisher :

$$Q(X; \alpha) \sim \mathbb{E}(X) + z_\alpha \sqrt{\text{Var}(X)}, \quad (3.32)$$

où

$$z_\alpha = \Phi^{-1}(\alpha) + \frac{\zeta_1}{6} [\Phi^{-1}(\alpha)^2 - 1] + \frac{\zeta_2}{24} [\Phi^{-1}(\alpha)^3 - 3\Phi^{-1}(\alpha)] - \frac{\zeta_1^2}{36} [2\Phi^{-1}(\alpha)^3 - 5\Phi^{-1}(\alpha)],$$

où ζ_1 désigne la skewness de X , et ζ_2 la kurtosis en excès,

$$\zeta_1 = \frac{\mathbb{E}([X - \mathbb{E}(X)]^3)}{\mathbb{E}([X - \mathbb{E}(X)]^2)^{3/2}} \text{ et } \zeta_2 = \frac{\mathbb{E}([X - \mathbb{E}(X)]^4)}{\mathbb{E}([X - \mathbb{E}(X)]^2)^2} - 3. \quad (3.33)$$

Définition 3.7 *Etant donné un échantillon $\{X_1, \dots, X_n\}$, l'estimation de Cornish-Fisher du quantile de niveau α est :*

$$\hat{Q}_n(\alpha) = \hat{\mu} + \hat{z}_\alpha \cdot \hat{\sigma}, \text{ où } \hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i \text{ et } \hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{\mu})^2},$$

où

$$\hat{z}_\alpha = \Phi^{-1}(\alpha) + \frac{\hat{\zeta}_1}{6} [\Phi^{-1}(\alpha)^2 - 1] + \frac{\hat{\zeta}_2}{24} [\Phi^{-1}(\alpha)^3 - 3\Phi^{-1}(\alpha)] - \frac{\hat{\zeta}_1^2}{36} [2\Phi^{-1}(\alpha)^3 - 5\Phi^{-1}(\alpha)],$$

avec $\hat{\zeta}_1$ l'estimateur usuel de la skewness, et $\hat{\zeta}_2$ l'estimateur usuel de la kurtosis

$$\hat{\zeta}_1 = \frac{\sqrt{n(n-1)}}{n-2} \frac{\sqrt{n} \sum_{i=1}^n (X_i - \hat{\mu})^3}{(\sum_{i=1}^n (X_i - \hat{\mu})^2)^{3/2}} \text{ et } \hat{\zeta}_2 = \frac{n-1}{(n-2)(n-3)} \left((n+1)\hat{\zeta}_2' + 6 \right)$$

où

$$\hat{\zeta}_2' = \frac{n \sum_{i=1}^n (X_i - \hat{\mu})^4}{(\sum_{i=1}^n (X_i - \hat{\mu})^2)^2} - 3.$$

En fait, de manière plus générale, si on suppose que la loi sous-jacente appartient à une famille paramétrique $F_X \in \mathcal{F} = \{F_\theta, \theta \in \Theta\}$ (que l'on supposera continue), $Q_X(\alpha) = F_\theta^{-1}(\alpha)$, et donc un estimateur naturel du quantile de niveau α est

$$\widehat{Q}_X(\alpha) = F_{\widehat{\theta}}^{-1}(\alpha), \quad (3.34)$$

où $\widehat{\theta}$ est un estimateur de θ (par maximum de vraisemblance, par la méthode des moments, etc).

Exemple 3.10 Pour une loi de Pareto

$$F_X(x) = \mathbb{P}(X \leq x) = 1 - \left(\frac{1}{x}\right)^\theta, \text{ pour } x \geq 1 \text{ et } \theta \geq 0.$$

La vraisemblance s'écrit

$$\log \mathcal{L}(\theta) = \sum_{i=1}^n \log \left(\frac{\theta}{X_i^{\theta+1}} \right) = n \log \theta - (\theta + 1) \sum_{i=1}^n \log X_i,$$

dont le maximum est obtenu pour

$$\widehat{\theta}_1 = \left(\frac{1}{n} \sum_{i=1}^n \log(X_i) \right)^{-1}$$

En notant que $\mathbb{E}(X) = \frac{\theta}{\theta - 1}$ pour $\theta > 1$, on en déduit l'estimateur de la méthode des moments, $\widehat{\theta}_2 = \frac{\widehat{\mu}}{\widehat{\mu} - 1}$. Notons enfin que

$$\log[1 - F(x)] = \log \overline{F}(x) = 1 - \alpha \log x,$$

autrement dit, l'estimateur (par moindres carrés) de la pente de la droite de régression peut être utilisé,

$$\widehat{\alpha}_3 = - \frac{-n \sum_{i=1}^n \log X_{i:n} \cdot \log \frac{n-i}{n} + \sum_{i=1}^n \log X_{i:n} \cdot \sum_{i=1}^n \log \frac{n-i}{n}}{n \sum_{i=1}^n [\log X_i]^2 - [\sum_{i=1}^n \log X_i]^2},$$

où $X_{i:n}$ désigne la statistique d'ordre, au sens où $X_{1:n} \leq \dots \leq X_{i:n} \leq \dots \leq X_{n:n}$. Ces trois estimateurs permettent de construire trois estimateurs de quantiles :

$$\widehat{Q}_k(\alpha) = (1 - \alpha)^{-1/\widehat{\theta}_k}.$$

Mais comme le montre la Figure 3.2, les modèles paramétriques induisent inévitablement des erreurs de modèles. La Figure de gauche montre l'estimation d'un quantile sous hypothèse de loi Gamma et de loi lognormale, et celle de droite l'estimation d'un quantile sous hypothèse de normalité et de loi de Student.

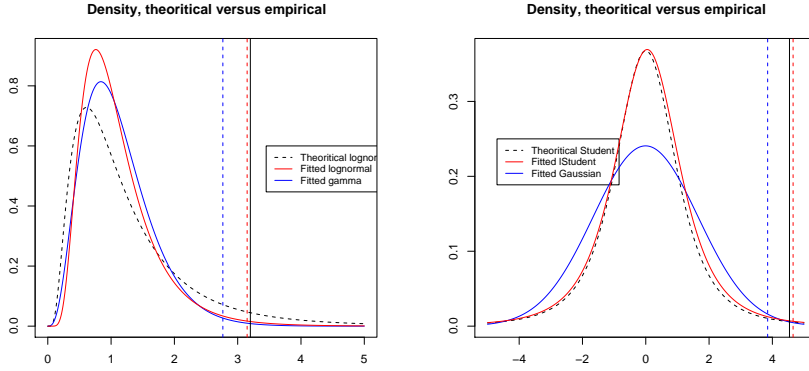


Figure 3.2 – Estimation de Value-at-Risk et erreur de modèle.

3.4.2 Estimation non-paramétrique

Afin d'éviter les erreurs dans le choix de la famille paramétrique \mathcal{F} , il est intéressant d'envisager des estimateurs nonparamétriques.

Définition 3.11 La fonction de répartition empirique F_n , construite à partir de $\{X_1, \dots, X_n\}$ est $F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X_i \leq x)$.

Rappelons que, à x fixé, $nF_n(x)$ suit une loi binomiale, centrée sur $nF(x)$ et de variance $nF(x)[1 - F(x)]$. Par la loi (forte) des grands nombres, $F_n(x)$ converge presque sûrement vers $F(x)$ quand $n \rightarrow \infty$, et le théorème de Glivenko-Cantelli assure que la convergence est uniforme, autrement dit $\sup\{|F_n(x) - F(x)|\}$ tend presque sûrement vers 0 lorsque $n \rightarrow \infty$. De plus, on a une normalité asymptotique à l'aide du théorème central limite,

$$\sqrt{n}[F_n(x) - F(x)] \xrightarrow{\mathcal{L}} \mathcal{N}(0, F(x)[1 - F(x)]),$$

lorsque $n \rightarrow \infty$. Et plus généralement, si l'on considère une convergence au sens des distributions finidimensionnelles, $\sqrt{n}[F_n(\mathbf{x}) - F(\mathbf{x})]$ converge vers un processus Gaussien $G(\mathbf{x})$ centré, de fonction de covariance

$$\text{Cov}(G(x_1), G(x_2)) = F(x_1)[1 - F(x_2)] \text{ pour } x_1 < x_2.$$

Ce processus admet la même loi que $(B_{F(x)})$ où $(B_t)_{t \in [0,1]}$ est un pont Brownien. On en déduit alors la loi de Kolmogorov Smirnov, qui permet d'assurer que $\sqrt{n} \sup\{|F_n(x) - F(x)|\}$ converge (au sens des distributions finidimensionnelles) vers $\sup\{B_t\}$.

Définition 3.8 La fonction quantile empirique, construite à partir de $\{X_1, \dots, X_n\}$ est $Q_n(\alpha) = \inf\{x \in \mathbb{R} : F_n(x) \geq \alpha\} = F_n^{-1}(\alpha)$.

Comme tenu du fait que F_n est une fonction en escalier continue à gauche, on notera que

$$Q_n(\alpha) = X_{k:n} \text{ où } \frac{k-1}{n} < \alpha \leq \frac{k}{n}.$$

Remarque 3.13 De même que la moyenne peut être obtenu par minimisation de la norme L^2 , les quantiles peuvent être obtenus par minimisation. En fait

$$Q_n(\alpha) = \operatorname{argmin}_{x \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^n H_\alpha(X_i - x)$$

où $H_\alpha(x) = x[\alpha - \mathbf{1}(x < 0)]$.

Le théorème de Glivenko Cantelli garantie la convergence (forte) de $Q_n(\alpha)$ vers $Q(\alpha)$.

Proposition 3.7 Si X_1, \dots, X_n est un échantillon *i.i.d.* de loi absolument continue de densité f , et que $f(Q(\alpha)) > 0$, alors

$$\sqrt{n}[Q_n(\alpha) - Q(\alpha)] \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\alpha[1-\alpha]}{f^2(Q(\alpha))}\right),$$

lorsque $n \rightarrow \infty$.

De manière plus générale, comme auparavant, on notera que $\sqrt{n}[Q_n(\alpha) - Q(\alpha)]$ converge vers un processus Gaussien $G(\alpha)$ centré, de fonction de covariance

$$\operatorname{Cov}(G(\alpha_1), G(\alpha_2)) = \frac{\alpha_1[1-\alpha_2]}{f(Q(\alpha_1))f(Q(\alpha_2))} \text{ pour } \alpha_1 < \alpha_2.$$

Pour terminer avec les propriétés du quantile empirique $Q_n(\alpha)$ rappelons le résultat suivant, correspondant à l'expansion de Bahadur-Kiefer

Proposition 3.8 Si X_1, \dots, X_n est un échantillon *i.i.d.* de loi absolument continue admettant une densité f telle que $f(Q(\alpha)) > 0$, alors

$$Q_n(\alpha) = Q(\alpha) + \frac{\alpha - F_n(Q(\alpha))}{f(Q(\alpha))} + Z_n$$

où, presque sûrement, $Z_n = O(n^{-3/4}[\log n]^{1/2}[\log(\log n)]^{1/4})$ selon ([6]), amélioré sous la forme $Z_n = O(n^{-3/4}[\log(\log n)]^{3/4})$ par [26].

Définition 3.9 La fonction de répartition empirique lissée, construite à partir de $\{X_1, \dots, X_n\}$ est

$$\begin{aligned}\widehat{F}_n(x) &= \int_{-\infty}^{+\infty} K\left(\frac{t-x}{h}\right) dF_n(t) \\ &= \frac{1}{nh} \sum_{i=1}^n \int_{-\infty}^x k\left(\frac{X_i-t}{h}\right) dt = \frac{1}{n} \sum_{i=1}^n K\left(\frac{X_i-x}{h}\right)\end{aligned}$$

où $K(x) = \int_{-\infty}^x k(t)dt$, k étant un noyau et $h > 0$ une fenêtre de lissage.

Classiquement, deux techniques assez différentes ont été considérées dans la littérature.

La première idée est de considérer une combinaison linéaire de statistiques d'ordre. Le quantile empirique est simplement

$$Q_n(p) = F_n^{-1}(p) = X_{[np]:n} \text{ où } [\cdot] \text{ désigne la partie entière.} \quad (3.35)$$

Si cet estimateur est simple à construire, on voit qu'il ne dépend que d'une unique observation. Il est alors naturel de considérer un lissage entre deux observations, si np n'est pas entier. Le quantile empirique pondéré est alors défini par

$$Q_n(p) = (1 - \gamma)X_{[np]:n} + \gamma X_{[np]+1:n} \text{ où } \gamma = np - [np].$$

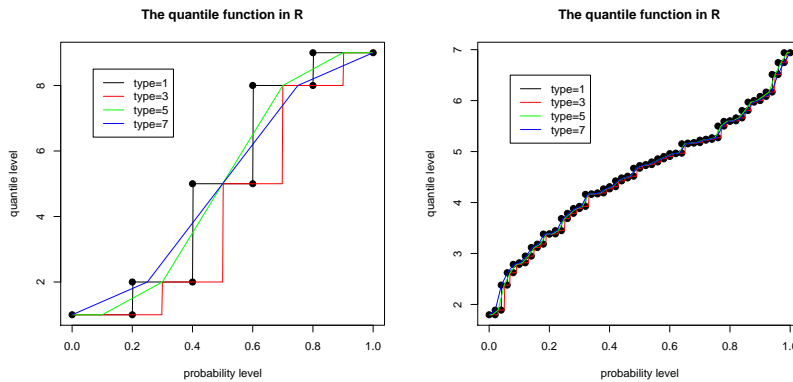


Figure 3.3 – Les quantiles empiriques usuels, sous R.

Afin d'augmenter l'efficience, il est possible, plus généralement de considérer des L statistiques,

$$Q_n(p) = \sum_{i=1}^n W_{i,n,p} X_{i:n} = \sum_{i=1}^n W_{i,n,p} F_n^{-1} \left(\frac{i}{n} \right) = \int_0^1 F_n^{-1}(t) k(p, h, t) dt \quad (3.36)$$

où F_n est la fonction de répartition empirique et où k est un noyau et h est une fenêtre. On peut alors écrire,

$$\begin{aligned} Q_n(p) &= \sum_{i=1}^n \left[\int_{\frac{i-1}{n}}^{\frac{i}{n}} k \left(\frac{t-p}{h} \right) dt \right] X_{i:n} \\ &= \sum_{i=1}^n \left[K \left(\frac{i-p}{h} \right) - K \left(\frac{i-1-p}{h} \right) \right] X_{i:n} \end{aligned} \quad (3.37)$$

où $K(x) = \int_{-\infty}^x k(t) dt$. L'idée est alors de donner davantage de poids aux statistiques d'ordre $X_{i:n}$ pour lesquelles i est proche de pn .

Par exemple, l'estimateur de Harrell-Davis (introduit dans [23]) est défini par

$$Q_n(p) = \sum_{i=1}^n \left[\int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{\Gamma(n+1)}{\Gamma((n+1)p)\Gamma((n+1)q)} y^{(n+1)p-1} (1-y)^{(n+1)q-1} \right] X_{i:n}.$$

Enfin, une approche alternative repose sur l'écriture $F \circ Q(\alpha) = \alpha$. Si \widehat{F}_n est un estimateur continu de F , alors un estimateur naturel de $Q(\alpha)$ est $\widehat{Q}_n(\alpha)$ tel que $\widehat{F}_n \circ \widehat{Q}_n(\alpha) = \alpha$, obtenu par l'algorithme de Gauss-Newton. [5], [43] ou [8] ont suggéré cette approche.

3.4.3 Estimation semiparamétrique

Une approche alternative est d'utiliser la théorie des valeurs extrêmes. Une approche peut être d'utiliser le théorème de Pickands-Balkema-de Haan : si u est suffisant grand la loi conditionnelle de $Y - u$ sachant $Y > u$ suit une de Pareto généralisée, de paramètres ξ et β (ces derniers pouvant être estimés par maximum de vraisemblance). Aussi, en posant $u = Y_{n-k:n}$, avec k suffisamment grand, notons $\widehat{\beta}_k$ et $\widehat{\xi}_k$ les estimateurs du maximum de vraisemblance de la loi de Pareto généralisé sur l'échantillon $\{Y_{n-k+1:n} - Y_{n-k:n}, \dots, Y_{n:n} - Y_{n-k:n}\}$,

$$\widehat{Q}(Y, \alpha) = Y_{n-k:n} + \frac{\widehat{\beta}_k}{\widehat{\xi}_k} \left(\left(\frac{n}{k} (1 - \alpha) \right)^{-\widehat{\xi}_k} - 1 \right) \quad (3.38)$$

Une alternative est l'estimateur de Hill, si $\xi > 0$,

$$\widehat{Q}(Y; \alpha) = Y_{n-k:n} \left(\frac{n}{k} (1 - \alpha) \right)^{-\widehat{\xi}_k}, \quad (3.39)$$

où $\widehat{\xi}_k = \frac{1}{k} \sum_{i=1}^k \log Y_{n+1-i:n} - \log Y_{n-k:n}$. Ces méthodes seront détaillées dans le Chapitre 4.

3.5 Estimation des mesures de risque

Considérons une mesure de risque par distorsion, c'est-à-dire de la forme $\mathcal{R}(X) = \int_0^1 F_X^{-1}(1-u)dg(u)$, où F_X est supposée strictement croissante, et continue. L'estimateur naturelle de cette mesure de risque est alors

$$\widehat{\mathcal{R}}(\{X_1, \dots, X_n\}) = \int_0^1 F_n^{-1}(1-u)dg(u).$$

Etant données que F_n^{-1} est une fonction en escalier, alors

$$\widehat{\mathcal{R}}(\{X_1, \dots, X_n\}) = \sum_{i=1}^n \left(g\left(\frac{i}{n}\right) - g\left(\frac{i-1}{n}\right) \right) X_{n-i+1:n},$$

qui est simplement un L estimateur. On peut alors, tout naturellement, espérer retrouver facilement des propriétés comme celles obtenues pour l'estimation nonparamétrique de la VaR. Par exemple

Proposition 3.9 *Si X_1, \dots, X_n est un échantillon i.i.d.,*

$$\sqrt{n} \left(\mathcal{R}(X) - \widehat{\mathcal{R}}(\{X_1, \dots, X_n\}) \right) \rightarrow \int_0^1 \frac{B_{1-t}}{f(F^{-1}(1-t))} dg(t),$$

où $(B_t)_{t \in [0,1]}$ est un pont brownien. La loi limite est centrée, de variance asymptotique

$$\int_0^1 \int_0^1 \frac{\min\{t, u\} - ut}{f(F^{-1}(1-t))f(F^{-1}(1-u))} dg(t)dg(u).$$

Bibliographie

- [1] C. Acerbi. Spectral measures of risk : a coherent representation of subjective risk aversion. *Journal of Banking and Finance*, 2(7) :1505–1518, 2002.
- [2] C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking and Finance*, 2(7) :1487–1503, 2002.
- [3] K.J. Arrow and G. Debreu. Existence of an equilibrium for a competitive economy. *Econometrica*, 22(3) :pp. 265–290, 1954.
- [4] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical finance*, 9(3) :203–228, 1999.
- [5] A. Azzalini. A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, 68(1) :326–328, 1981.
- [6] R. R. Bahadur. A note on quantiles in large samples. *The Annals of Mathematical Statistics*, 37(3) :pp. 577–580, 1966.
- [7] A. Charpentier. Value at risk et probabilité de ruine, entre vaccination et banque d'affaires. *Risques*, 76 :103–106, 2008.
- [8] S.X. Chen and C.Y. Tang. Nonparametric Inference of Value-at-Risk for Dependent Financial Returns. *Journal of Financial Econometrics*, 3(2) :227–255, Spring 2005.
- [9] E.A. Cornish and R.A. Fisher. Moments and cumulants in the specification of distributions. *Revue de l'Institut International de Statistique*, 5(4) :307–320, 1937.
- [10] I. Csiszár. I -divergence geometry of probability distributions and minimization problems. *Annals of Probability*, 3 :146–158, 1975.
- [11] B. de Finetti. Il problema dei « pieni ». *Giornale dell'Istituto Italiano degli Attuari*, 11(1) :1–8, 1940.
- [12] F. Delbaen. *Coherent Risk Measures*. Cattedra Galileiana. Scuola Normale Superiore di Pisa, 2000.
- [13] F. Delbaen. Coherent risk measures on general probability spaces. In *Essays in Honour of Dieter Sondermann*, pages 1–37. Springer Verlag, 2000.

- [14] F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, 300 :463–520, 1994. 10.1007/BF01450498.
- [15] M. Denuit and A. Charpentier. *Mathématiques de l'assurance non-vie : principes fondamentaux de théorie du risque. Tome 1*. Economica, 2004.
- [16] I. Ekeland, A. Galichon, and M. Henry. Comonotonic measures of multivariate risks. *Mathematical finance*, (22) :109–132, 2012.
- [17] H. Föllmer. Financial uncertainty, risk measures and robust preferences. In Marc Yor, editor, *Aspects of Mathematical Finance*, pages 3–13. Springer Berlin Heidelberg, 2008.
- [18] H. Föllmer and A. Schied. *Stochastic finance*, volume 27 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 2002.
- [19] H.U. Gerber. *An introduction to mathematical risk theory*, volume 8 of *S.S. Heubner Foundation Monograph Series*. University of Pennsylvania Wharton School S.S. Huebner Foundation for Insurance Education, Philadelphia, Pa., 1979. With a foreword by James C. Hickman.
- [20] H.U. Gerber and E.S.W. Shiu. Option pricing by esscher transforms (with discussions). *Transactions of the Society of Actuaries*, 46, 1994.
- [21] M.J. Goovaerts, F. De Vylder, and J. Haezendonck. *Insurance Premiums*. North-Holland, 1984.
- [22] G. H. Hardy and J. E. Littlewood. A maximal theorem with function-theoretic applications. *Acta Math.*, 54(1) :81–116, 1930.
- [23] F.E. Harrell and C.E. Davis. A new distribution-free quantile estimator. *Biometrika*, 69(3) :635–640, 1982.
- [24] J.M. Harrison and S.R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11(3) :215 – 260, 1981.
- [25] P. Jorion. *Value at Risk : The New Benchmark for Managing Financial Risk*. McGraw-Hill, 2006.
- [26] J. Kiefer. On Bahadur's representation of sample quantiles. *Ann. Math. Statist.*, 38 :1323–1342, 1967.
- [27] S. Kusuoka. On law invariant coherent risk measures. In *Advances in mathematical economics, Vol. 3*, volume 3 of *Adv. Math. Econ.*, pages 83–95. Springer, Tokyo, 2001.
- [28] Z. Landsman and M. Sherris. Risk measures and insurance premium principles. *Insurance : Mathematics and Economics*, 29(1) :103 – 115, 2001.
- [29] E. L. Lehmann. Ordered families of distributions. *Ann. Math. Statist.*, 26 :399–419, 1955.
- [30] M. Mania and M. Schweizer. Dynamic exponential utility indifference valuation. *Ann. Appl. Probab.*, 15(3) :2113–2143, 2005.

- [31] H.M. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1) :pp. 77–91, 1952.
- [32] H.M. Markowitz. *Mean-variance analysis in portfolio choice and capital markets*. Basil Blackwell, Oxford, 1987.
- [33] A. Müller and D. Stoyan. *Comparison methods for stochastic models and risks*. Wiley series in Probability and Statistics, 2002.
- [34] D. Olivier and H.U. Gerber. On convex principles of premium calculation. *Insurance : Mathematics and Economics*, 4(3) :179 – 189, 1985.
- [35] A. Reich. Premium principles and translation invariance. *Insurance : Mathematics and Economics*, 3(1) :57 – 66, 1984.
- [36] A. Reich. Properties of premium calculation principles. *Insurance : Mathematics and Economics*, 5(1) :97 – 101, 1986.
- [37] A. D. Roy. Safety first and the holding of assets. *Econometrica*, 20(3) :pp. 431–449, 1952.
- [38] L. Rüschendorf. Law invariant convex risk measures for portfolio vectors. *Statistics & Decisions*, 24 :97–108, 2006.
- [39] P.A. Samuelson. The “fallacy” of maximizing the geometric mean in long sequences of investing or gambling. *Proc. Nat. Acad. Sci. U.S.A.*, 68 :2493–2496, 1971.
- [40] D. Schmeidler. Integral representation without additivity. *Proc. Amer. Math. Soc.*, 97(2) :255–261, 1986.
- [41] K.S. Schmidt. Positive homogeneity and multiplicativity of premium principles on positive risks. *Insurance : Mathematics and Economics*, 8(4) :315 – 319, 1989.
- [42] M. Shaked and J.G. Shanthikumar. *Stochastic Orders*. Springer, 2006.
- [43] S.J. Sheather and J.S. Marron. Kernel quantile estimators. *J. Amer. Statist. Assoc.*, 85(410) :410–416, 1990.
- [44] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36 :423–439, 1965.
- [45] S. Wang and J.A.G. Dhaene. Comonotonicity, correlation order and premium principles. *Insurance Math. Econom.*, 22(3) :235–242, 1998.
- [46] S.S. Wang. Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance Math. Econom.*, 17(1) :43–54, 1995.
- [47] S.S. Wang. Ordering of risks under PH-transforms. *Insurance Math. Econom.*, 18(2) :109–114, 1996.
- [48] S.S. Wang and V.R. Young. Risk-adjusted credibility premiums using distorted probabilities. *Scand. Actuar. J.*, (2) :143–165, 1998.