

Introduction

Thierry Bazier-Matte

February 3, 2017

1 Introduction

1.1 Avant propos

[**Todo:** Discuter du rôle croissant que jouent l’informatique et les statistiques dans la construction de portefeuille. Contraster avec les math. stochastiques. Citer Simons et cet article de Quandl selon lequel data is the new shit.]

1.2 Exposition du problème et hypothèses

Ce mémoire vise à établir clairement et rigoureusement comment un investisseur *averse au risque* disposant d’*information complémentaire* au *marché* peut utiliser cette information pour accroître son *utilité espérée* ou, de façon équivalente, son *rendement équivalent certain*.

Modélisation du marché Nous entendons ici par *marché* n’importe quel type d’actif financier ou spéculatif dans lequel on peut investir une partie de sa fortune dans l’espoir de la voir fructifier au cours d’une période de temps arbitraire. Ainsi, tout au long de l’exposé théorique qui suivra, il peut être pertinent d’avoir en tête les rendements quotidiens issus des grands indices boursiers (par exemple les 500 plus grandes capitalisations américaines). Cependant, le traitement qui sera développé pourrait tout aussi bien s’appliquer à une action cotée en bourse dont on considère les rendements mensuels.^[Nécessaire?] Mathématiquement, l’idée de marché peut ainsi être réduite à celle d’une variable aléatoire $R(t)$ décrivant l’évolution du rendement de l’actif en question.

Relativement à l’idée de marché, nous ferons également l’hypothèse que l’univers a une influence sur ces rendements. Il serait par exemple raisonnable de croire que le prix du pétrole a une influence sur l’évolution du rendement du marché américain. De la même façon, l’annonce d’un scandale aura à son tour des répercussions sur la valeur du titre de la compagnie dont il est l’objet. En outre, il a été montré par Fama et French que le rendement d’une action pouvait s’expliquer comme une combinaison de quelques facteurs fondamentaux (la taille de l’entreprise, le risque de marché et le ratio cours/valeur).

On peut alors considérer un vecteur d'information $\vec{X}(t) = (X_1(t), X_2(t), \dots)$ dont chaque composante représente une information particulière, par exemple l'absence ou la présence d'un certain type de scandale, un ratio comptable, le prix d'un certain actif financier.^[Rephrase] D'un point de vue probabiliste, on dira donc qu'il existe une forme de dépendance entre $R(t)$ et $\{\vec{X}(\tau) \mid \tau < t\}$ l'ensemble des événements antérieurs à t . Le processus joint de ces deux événements sera désormais défini comme *la distribution totale de marché*, ou simplement le marché.

Stationarité Bien qu'un tel modèle permette de représenter de façon très générale l'évolution d'un marché, nous formulerons l'hypothèse supplémentaire selon laquelle le marché est un processus *stationnaire*. Ceci permet notamment d'évacuer la notion temporelle afin de ne représenter qu'une distribution de causes (l'information X) et d'effet (l'observation des rendements R). Cette hypothèse est assez contraignante. Elle suppose d'une part que les réalisations passées n'ont aucun effet sur les réalisations futures (indépendance) et d'autre part que la distribution de marché est figée dans le temps, ce qui implique notamment l'absence de probabilité de faillite. Elle implique aussi que le marché ne peut être vu comme un environnement adversarial qui réagirait par exemple aux décisions d'un investisseur. Ceci vient notamment mettre en cause la théorie des marchés efficients selon laquelle une brèche dans l'absence d'arbitrage serait immédiatement colmatée par des spéculateurs (effet d'autorégulation). Nous aurons toutefois l'occasion de revenir plus en détail sur les liens à faire entre cet exposé et l'efficacité des marchés.

Approche mathématique et statistique Dans ce qui suit, nous noterons par M la distribution de marché. Le vecteur aléatoire d'information sera par ailleurs formé de m composantes; pour l'instant, aucune hypothèse par rapport à la dépendance des composantes de X ne sera formulée. À ce point-ci, on a donc le modèle de marché suivant:

$$M = (R, X_1, \dots, X_m). \quad (1)$$

On fera également l'hypothèse qu'on possède un ensemble de n éléments échantillonnés à partir de M , de sorte que:

$$\{r_i, x_{i1}, \dots, x_{im}\}_{i=1}^n \sim M \quad (2)$$

représente notre ensemble d'échantillonnage (aussi appelé ensemble d'entraînement). Le domaine des rendements possibles de R sera noté $\mathbf{R} \subseteq \mathcal{R}$ et celui du vecteur d'information X sera noté $\mathbf{X} \subseteq \mathcal{R}^m$. Le vecteur d'observations de rendement sera noté $r \in \mathcal{R}^n$ et la matrice d'information par $X \in \mathcal{R}^{n \times m}$.

Modélisation de la préférence Indépendamment de la notion de marché, on a d'autre part l'aspect d'aversion au risque qui est modélisé par une fonction d'utilité $u : \mathbf{R} \rightarrow \mathbf{U}$, où $\mathbf{R} \subseteq \mathcal{R}$ est le domaine (fermé ou non) des rendements considérés et $\mathbf{U} \subseteq \mathcal{R}$ celui des *utilités*.

Bien qu'en pratique il soit plus facile de travailler sur des fonctions possédant des valeurs dans U , en pratique cet espace est adimensionnel^[Citation needed], de sorte que nos résultats seront présentés dans l'espace des rendements R .

Fonction de décision Donnés ces éléments de base, le but de ce mémoire sera alors de déterminer une fonction de décision d'investissement $q : X \rightarrow P \subseteq \mathcal{R}$ maximisant l'utilité espérée de l'investissement.

Mathématiquement on a donc le problème fondamental suivant:

$$\underset{q \in \mathcal{Q}}{\text{maximiser}} \quad Eu(R \cdot q(X)), \quad (3)$$

où l'optimisation a lieu dans un espace de fonctions \mathcal{Q} à préciser.

Cependant, comme la distribution $(X, R) = M$ est inconnue, il est impossible de déterminer la fonction q^* minimisant cet objectif. On dispose toutefois d'un échantillon de M dont on peut se servir pour approximer le problème (SAA, voir Shapiro^[Citation needed]):

$$\underset{q \in \mathcal{Q}}{\text{maximiser}} \quad n^{-1} \sum_{i=1}^n u(r_i q(x_i)), \quad (4)$$

mais encore ici le problème est mal spécifié, puisqu'aucune contrainte n'a été posée sur l'espace \mathcal{Q} . Par exemple, il suffirait de prendre pour q un dictionnaire associant à x_i la valeur αr_i , où $\alpha > 0$, et à toute autre valeur de x une valeur nulle pour avoir une valeur d'utilité arbitrairement grande à mesure que $\alpha \rightarrow \infty$.

Risque in-échantillon et hors échantillon une telle fonction q est qu'elle se généralise très mal. En effet pour toute observation x qui ne figurerait dans l'ensemble d'entraînement, q prescrirait alors un investissement nul. Il y a alors une énorme différence entre l'utilité observée au sein de notre échantillon et l'utilité hors échantillon.

Donnée une fonction de décision $q \in \mathcal{Q}$ et un échantillon de M , on définit le *risque in-échantillon* ou *risque empirique* par

$$\hat{R}(q) = n^{-1} \sum_{i=1}^n \ell(r_i q(x_i)), \quad (5)$$

où $\ell = -u$. De la même façon, on définit le *risque hors-échantillon* ou *erreur de généralisation* par

$$R(q) = Eu(R \cdot q(X)). \quad (6)$$

On peut souhaiter d'une bonne fonction de décision qu'elle performe bien hors échantillon, aussi la quantité $R(q) - \hat{R}(q)$ sera-t-elle primordiale et beaucoup d'attention lui sera consacrée dans les prochaines sections. Notons que le risque hors-échantillon étant théoriquement impossible à calculer, en pratique on segmentera l'ensemble d'échantillonnage en deux parties, l'une dédiée à l'apprentissage, l'autre à évaluer la performance hors échantillon.

Régularisation Afin de contrecarrer le risque hors échantillon, la solution est en fait de pénaliser la complexité de la fonction de décision q (rasoir d’Occam). Ainsi, on étudiera en profondeur le choix d’une fonctionnelle $R : \mathcal{Q} \rightarrow \mathcal{R}$ permettant de quantifier la complexité de q . L’objectif serait alors

$$\underset{q \in \mathcal{Q}}{\text{maximiser}} \quad n^{-1} \sum_{i=1}^n u(r_i q(x_i)) - R(q). \quad (7)$$

Par exemple, comme les mesures sur x peuvent comporter de l’incertitude ou du bruit, il serait souhaitable que la décision $q(x_1)$ soit proche de $q(x_2)$, si x_1 et x_2 sont eux même proches dans l’espace \mathbf{X} . Si R encodait une telle préférence, ne fonction discontinue comme le dictionnaire présenté plus haut sera alors hautement défavorisée, et une fonction plus lisse y serait préférée.

[**Todo:** Introduire la validation croisée ainsi que le paramètre λ dans l’objectif.]

Espaces de décision En pratique, ce mémoire ne considérera que des espaces de Hilbert pour \mathcal{Q} . Un des avantages des espaces de Hilbert, c’est qu’ils induisent naturellement une notion de norme $\|\cdot\|_H$, qu’on peut intuitivement relier au concept de complexité. Nous nous intéresserons donc aux propriétés induites par $R(q) = \|q\|_H^2 = \langle q, q \rangle$. Il y a aussi moyen, sous des conditions assez techniques (théorème de la représentation) de généraliser la norme L_2 de q à une norme L_p général. En particulier, nous verrons qu’une régularisation donnée par norme L_1 induit certaines propriétés d’éparsité dans la solution.

Décisions linéaires De façon générale, la forme de décision la plus simple est celle qui combine linéairement les p observations de $x \in \mathbf{X} \subseteq \mathcal{R}^p$; autrement dit lorsque qu’on contraint $\mathcal{Q} = \mathbf{X}^*$, *i.e.*, à l’espace dual de \mathbf{X} . En langage plus clair, à toute fonction $q \in \mathcal{Q}$ il existe un vecteur de dimension p tel que la décision dérivée de l’observation x sera donnée par $q(x) = \langle q, x \rangle = q^T x$.

La régularisation L_2 de q devient alors tout simplement $R(q) = q^T q = \|q\|^2$ et la fonction optimale de décision q^* sera alors déterminée en résolvant le problème d’optimisation suivant:

$$\underset{q \in \mathcal{R}^p}{\text{maximiser}} \quad n^{-1} \sum_{i=1}^n u(r_i q^T x_i) - \lambda \|q\|^2. \quad (8)$$

1.3 Dimensionnalité de l’information

[**Todo:** Discussion du phénomène big data, de l’importance de p]

1.4 Risque et garanties statistiques sur la décision

[**Todo:** Discussion sur les méthodes de risques hors échantillon, complexité de l'échantillonnage, mesure Rademacher, distance par rapport à la “meilleure” décision]

1.5 Interprétations

Interprétation géométrique dans l'espace X

Interprétation statistique (avec matrix covariance)

Autre?

1.6 Objectifs