Portfolio Optimization in a Big Data Context

Thierry Bazier-Matte

Summer 2015

Notation. In the following, $\boldsymbol{A}$ (capital boldface) are assumed to represent a real subset of any dimension, A (capital case) represent random variables (or distribution) and a (lower case) represent deterministic variables or realization. $\mathscr{R}$ represents the real set.

Let $M = (X, R)$ the *market* be an unknown distribution with support $\boldsymbol{M} = \boldsymbol{X} \times \boldsymbol{R} \subseteq \mathscr{R}^{p+1}$, ie. numerically qualifiable, with $(x, r) = m \sim M$ a *market observation*, consisting in one part *state* $x \in \mathscr{R}^p$ and another part *outcome* $r \in \mathscr{R}$. Typically x is a vector of observations from various variable of interests, such as financial or economical news, etc. Scalar r in this article shall represent the return from a financial asset of interest. Finally, let $M_n = \{M, \dots, M\}$ be a *random set* of n (unrealized) observations (with support $\boldsymbol{M}^n$). Therefore $\mu_n \sim M_n$ represents an iid sample of n market observations.

This article shall study *linear investment decisions* $q^T x$, with $q \in \boldsymbol{Q} \subseteq \mathscr{R}^p$.

Definition. Let $\ell : \boldsymbol{M} \to \mathscr{R}$ be a *loss function* defined by

$$\ell(m, q) = \ell(x, r, q) = -u(r\, q^T x + (1 - R_f) q^T x),$$

where $u(r) = \min(r, \beta r)$ and R_f the risk free rate.

Definition. The *empirical risk* $\hat{R} : \boldsymbol{M} \times \boldsymbol{Q} \to \mathscr{R}$ associated with decision q and market sample μ_n is given by

$$\hat{R}_{\mu_n}(q) = n^{-1} \sum_{i=1}^{n} \ell(m_i, q).$$

Definition. The *empirical decision algorithm* $\hat{A}_n : \boldsymbol{M}^n \to \boldsymbol{Q}$ associated with market sample μ_n is the optimal value of the problem

$$\text{minimize} \quad \hat{R}_{\mu_n}(q) + \lambda \|q\|_2^2.$$

From now on, $\hat{q}_n := \hat{A}_n(\mu_n)$ the empirical decision associated with market sample μ_n and $\hat{Q}_n := A_n(S_n)$ the random empirical decision, ie. $\hat{q}_n \sim \hat{Q}_n$.

Definition. The *true risk* $R_{\text{true}} : \boldsymbol{Q} \to \mathscr{R}$ associated with decision q is given by

$$R_{\text{true}}(q) = E_M[\ell(m, q)].$$

Definition. The *optimal decision* $q^\star$ is the optimal value of the problem

$$\text{minimize} \quad R_{\text{true}}(q) + \lambda \|q\|_2^2.$$