Value-at-Risk, Expected Shortfall and coherent risk measures

Theoretical and numerical aspects

Noufel Frikha

Université Paris 1 Panthéon-Sorbonne

October 2024

Contents

Value-at-Risk (VaR)

Definition

We follow the usual convention on risk measures (that originally appeared in insurance) by considering the loss variable $L = -\text{P\&L}_{t,t+h}$.

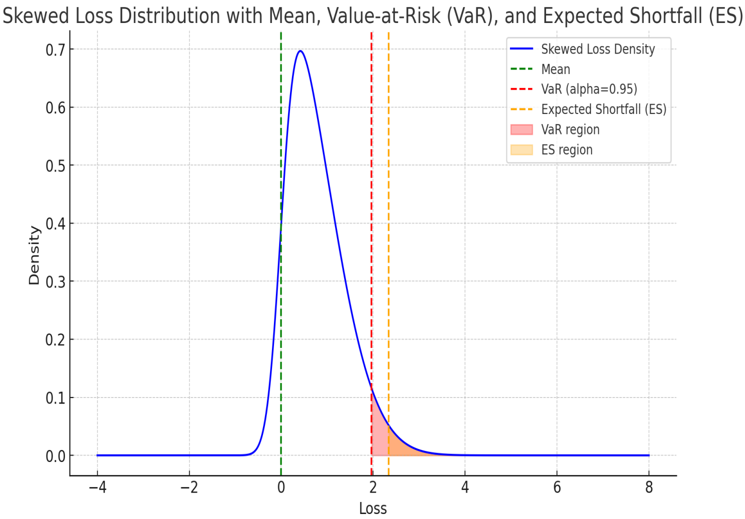
Definition

Given a loss portfolio L over a time horizon h, and with cumulative distribution function (c.d.f.) F_L, we call Value-at-Risk for a confidence level $\alpha \in (0,1)$, denoted $\text{VaR}_\alpha(L)$, the smallest value having a probability smaller than $1-\alpha$ to be lost, i.e.:

$$\text{VaR}_\alpha(L) = \inf\{\ell \in \mathbb{R} : \mathbb{P}(L > \ell) \leq 1 - \alpha\} = \inf\{\ell \in \mathbb{R} : F_L(\ell) \geq \alpha\}.$$

Remarks:

- Confidence levels α are often of the order of 95%, 99%, or 99.9% depending on the context. The horizon h is typically 1 or 10 days (market risk) or 1 year (credit risk).
- $\text{VaR}_\alpha(L)$ is increasing with α.


Skewed Loss Distribution with Mean, Value-at-Risk (VaR), and Expected Shortfall (ES)
Skewed Loss Density
Mean
VaR (alpha=0.95)
Expected Shortfall (ES)
VaR region
ES region
Density
Loss
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
−4
−2
0
2
4
6
8

Generalized Inverse and Quantile Function

Reminder: The c.d.f. of a random variable X is the function $F : \mathbb{R} \to [0,1]$, defined as $F(x) = \mathbb{P}[X \leq x]$. It is non-decreasing, right-continuous with left-limit and satisfying:

$$\lim_{x \to -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \to +\infty} F(x) = 1.$$

Definition

The quantile function of X is the generalized inverse of its c.d.f. F, defined on $(0,1)$ as:

$$F^{-1}(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\},$$

with the convention that $\inf \emptyset = +\infty$. The function F^{-1} is non-decreasing and left-continuous.

For $\alpha \in (0,1)$, the quantile of order α of F is denoted:

$$q_\alpha(F) = F^{-1}(\alpha),$$

which can also be written as $q_\alpha(X)$ when F is the c.d.f. of the random variable X. Using this notation, we can express:

$$\mathrm{VaR}_\alpha(L) = q_\alpha(F_L),$$

where F_L is the c.d.f. of the loss variable L.

Some Useful Properties of the Quantile Function

Let X be a random variable with c.d.f. F. The following properties hold for the quantile function:

(P1) For $\alpha \in (0,1)$, $F(q_\alpha(F)) \geq \alpha$.

(P2) For $\alpha \in (0,1)$, $F(x) \geq \alpha$ if and only if $x \geq q_\alpha(F)$.

(P3) If F is continuous, then $F(q_\alpha(F)) = \alpha$ for $\alpha \in (0,1)$.

(P4) Let U follow a uniform distribution on $[0,1]$. Then, $F^{-1}(U)$ has the same distribution as X, i.e., its c.d.f. is F.

Remarks:

- If F is continuous, property (P3) shows that $F^{-1}(\alpha)$ is strictly increasing.
- If F is continuous and strictly increasing, i.e., invertible, then the generalized inverse coincides with the usual inverse function.
- The generalized inverse is useful in cases where F is not invertible, such as when F is discontinuous or constant on non-empty intervals.
- Property (P4) is the basis of the inversion method for simulating random variables with c.d.f. F.

Proof of Properties (P1) and (P2)

Proof:

- (P1): By definition of $q_\alpha = F^{-1}(\alpha)$, it is clear that if $F(x) \geq \alpha$, then $x \geq q_\alpha$, and moreover, for all n, $\exists x_n \leq q_\alpha + \frac{1}{n}$ such that $F(x_n) \geq \alpha$. Since F is non-decreasing, we have $F(q_\alpha + \frac{1}{n}) \geq \alpha$, and so by right-continuity of F, we deduce that $F(q_\alpha) \geq \alpha$, which proves (P1).

- (P2): Again, since F is non-decreasing, this implies that if $x \geq q_\alpha$, then $F(x) \geq F(q_\alpha) \geq \alpha$, which proves (P2).

- (P3): From (P2), if $x < q_\alpha$, then $F(x) < \alpha$. With (P1), and if F is continuous at q_α, this proves that $F(q_\alpha) = \alpha$, i.e. (P3).

- (P4): From (P2), we have for all $x \in \mathbb{R}$,

$$\mathbb{P}[F^{-1}(U) \leq x] = \mathbb{P}[U \leq F(x)] = F(x),$$

hence, $F^{-1}(U)$ and X have the same c.d.f., and thus the same distribution.

Corollary (Quantile Function of a Transformed Variable)

If g is continuous and strictly increasing on $\mathbb{R}$ (hence invertible), then

$$q_\alpha(g(X)) = g(q_\alpha(X)).$$

Proof: Let F be the c.d.f. of X. Then, the c.d.f. of $Y = g(X)$ is $G(y) = F(g^{-1}(y))$. From (P2), for all $y \in \mathbb{R}$, we have the equivalence:

$$\begin{aligned} y \geq q_\alpha(Y) &\iff G(y) \geq \alpha \iff F(g^{-1}(y)) \geq \alpha \\ &\iff g^{-1}(y) \geq q_\alpha(X) \iff y \geq g(q_\alpha(X)), \end{aligned}$$

which shows that $q_\alpha(Y) = g(q_\alpha(X))$.

Example: Properties of Quantiles for Some Functions

For example:

- $q_\alpha(X^3) = q_\alpha(X)^3$
- $q_\alpha(e^X) = e^{q_\alpha(X)}$
- $q_\alpha(aX + b) = aq_\alpha(X) + b$, for $a > 0$

Attention: In general:

- $q_\alpha(X^2) \neq q_\alpha(X)^2$
- $q_\alpha(-X) \neq -q_\alpha(X)$

If F, the c.d.f. of X, is invertible, then $q_\alpha(-X) = -q_{1-\alpha}(X)$.

Application: Affine transformation of VaR

$$\mathrm{VaR}_\alpha(aL + b) = a\mathrm{VaR}_\alpha(L) + b, \ a > 0.$$

Interpretation: The risk measured in VaR of a shares of a portfolio is a times the risk of one share. Moreover, adding (if $b > 0$) or withdrawing (if $b < 0$) an amount b of this portfolio changes the risk by b.

Value-at-Risk

Example 1: Gaussian Loss Distribution

Assume that the loss distribution follows a Gaussian law $L \sim N(\mu, \sigma^2)$. Then the normalized loss $\overline{L} = \frac{L-\mu}{\sigma}$ follows a centered standard normal distribution, and we have:

$$\mathrm{VaR}_\alpha(L) = \mu + \sigma\Phi^{-1}(\alpha),$$

where Φ is the cumulative distribution function (c.d.f.) of the standard normal distribution $N(0,1)$, and $\Phi^{-1}(\alpha)$ is the α-quantile of Φ.

Reminder: Φ being continuous and $N(0,1)$ being symmetric around 0, we have:

$$\Phi^{-1}(\alpha) = -\Phi^{-1}(1-\alpha), \quad \alpha \in (0,1).$$

Example 2: Portfolio of Stocks

Consider a portfolio consisting of a long position in $\beta = 5$ shares of a stock with an initial price $S_0 = 100$. The intra-day log-return of the asset $\Delta_1 Y_{t+1} = \ln(S_{t+1}/S_t)$, $t = 0, 1, \ldots$ are assumed to be i.i.d. and normally distributed with mean 0 and standard deviation $\sigma = 0.1$.

(i) We denote by L_1 the portfolio loss between today and tomorrow. We have:

$$L_1 = -P\&L_1 = -\beta(S_1 - S_0) = -\beta S_0(e^{\Delta_1 Y_1} - 1) = -500(e^{\Delta_1 Y_1} - 1).$$

Then, using the properties on the VaR and the fact that $\Delta_1 Y_1 \overset{law}{=} -\Delta_1 Y_1$:

$$\begin{aligned}
\mathrm{VaR}_\alpha(L_1) &= -500\mathrm{VaR}_{1-\alpha}(e^{\Delta_1 Y_1} - 1) \\
&= -500(e^{\mathrm{VaR}_{1-\alpha}(\Delta_1 Y_1)} - 1) \\
&= 500(1 - e^{-\mathrm{VaR}_\alpha(\Delta_1 Y_1)}) \\
&= 500\left(1 - e^{-0.1\Phi^{-1}(\alpha)}\right).
\end{aligned}$$

For $\alpha = 0.99$, $\Phi^{-1}(\alpha) \approx 2.3$, giving $\mathrm{VaR}_{0.99}(L_1) \approx 100$.

(ii) We keep the long position on the portfolio for 100 days. The portfolio loss over this period is:

$$L_{100} = -\beta(S_{100} - S_0) = -500(e^{\Delta_{100} Y_{100}} - 1),$$

where

$$\Delta_{100} Y_{100} = \ln(S_{100}/S_0) = \sum_{t=0}^{99} \Delta_1 Y_{t+1} \sim N(0,1).$$

The VaR over this period is:

$$\mathsf{VaR}_\alpha(L_{100}) = 500 \left(1 - e^{-\Phi^{-1}(\alpha)}\right),$$

hence for $\alpha = 0.99$, $\mathsf{VaR}_{0.99}(L_{100}) \approx 450$.
A linear approximation of the loss gives:

$$L_{100} = -500(e^{\Delta_{100} Y_{100}} - 1) \approx \tilde{L}_{100} = -500\Delta_{100} Y_{100},$$

leading to:

$$\mathsf{VaR}_\alpha(\tilde{L}_{100}) = 500\Phi^{-1}(\alpha),$$

which for $\alpha = 0.99$ gives $\mathsf{VaR}_{0.99}(\tilde{L}_{100}) \approx 1150$, a poor approximation of $\mathsf{VaR}_{0.99}(L_{100})$.

Expected Shortfall (ES)

Although VaR is popular among practitioners, it has several limitations. In particular, it does not consider the magnitude of losses beyond the VaR level.

Definition (Definition of Expected Shortfall)

Let L be a loss variable with cumulative distribution function F_L such that $\mathbb{E}[|L|] < \infty$. The expected shortfall at confidence level $\alpha \in (0,1)$ is defined as:

$$\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L)du = \frac{1}{1-\alpha}\int_\alpha^1 q_u(F_L)du,$$

where $q_u(F_L)$ is the quantile function of F_L.

Instead of fixing a confidence level α, ES looks at the average of losses exceeding VaR at level α, i.e., in the tail of the loss distribution. ES is sometimes called Conditional VaR (CVaR), Average VaR (AVaR) or Tail VaR (TVaR).

For continuous loss distributions, we have an equivalent definition of expected shortfall:

Proposition: If $L \in L^1(\mathbb{P})$ has a continuous cdf, then

$$\mathsf{ES}_\alpha(L) = \mathbb{E}[L | L \geq \mathsf{VaR}_\alpha(L)] = \frac{1}{1-\alpha}\mathbb{E}[L\mathbf{1}_{L \geq q_\alpha(F_L)}].$$

Interpretation: This means that ES is the average of losses exceeding VaR.

Proof. Recall that if $U \sim U([0;1])$ then $F_L^{-1}(U)$ has the same distribution as L. We deduce that :

$$\begin{aligned}\mathbb{E}[L 1_{L \geq q_\alpha(F_L)}] &= \mathbb{E}[F_L^{-1}(U) 1_{F_L^{-1}(U) \geq F_L^{-1}(\alpha)}] \\ &= \mathbb{E}[F_L^{-1}(U) 1_{U \geq \alpha}] = \int_\alpha^1 F_L^{-1}(a)\, da = \int_\alpha^1 \mathrm{VaR}_a(L)\, da.\end{aligned}$$

where in the second equality we used the fact that F_L^{-1} is strictly increasing (since F_L is continuous). We conclude by noting that when F_L is continuous, $\mathbb{P}(L \geq \mathrm{VaR}_\alpha(L)) = 1 - \alpha$.

Remark: In the general case when F_L may be discontinuous, we have :

$$\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha}\Big(\mathbb{E}[L 1_{L \geq q_\alpha(F_L)}] + \mathrm{VaR}_\alpha(L)(1 - \alpha - \mathbb{P}(L \geq \mathrm{VaR}_a(L))\Big).$$

Comparison between VaR and ES

Comparison between VaR and ES

Consider a continuous loss distribution, where:

$$\mathbb{P}(L \geq \text{VaR}_\alpha(L)) = 1 - \alpha.$$

◦ For example, when $\alpha = 95\%$, $\text{VaR}_\alpha(L) = 10,000$ euros means there is a 5% probability of losing more than 10,000 euros.

◦ For ES, we have:

$$\text{ES}_\alpha(L) = \mathbb{E}[L | L \geq \text{VaR}_\alpha(L)],$$

so, for instance, $\text{ES}_{0.95}(L) = 13,000$ euros means that, on average, the "bad" losses exceeding 10,000 euros are 13,000 euros.

Basic Inequalities

- Recall that VaR_α is nondecreasing with α so that

$$\mathrm{ES}_\alpha(L) \geq \mathrm{VaR}_\alpha(L),$$

  i.e., ES is more conservative than VaR.

- When L follows a Gaussian distribution, we have remarkably:

$$\mathrm{VaR}_{99\%}(L) \approx \mathrm{ES}_{97.5\%}(L).$$

Calculations of Gaussian VaR and ES

Recall that for $L \sim N(0,1)$:

$$\mathrm{VaR}_\alpha(L) = \Phi^{-1}(\alpha), \quad \mathrm{ES}_\alpha(L) = \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}.$$

α	99.9%	99%	95%
VaR_α	3.090	2.326	1.645
ES_α	3.367	2.665	2.063

Other Examples of Distribution

The following two examples are left as an exercise.

- Laplace distribution (double exponential), i.e. density $f(\ell) = \frac{\lambda}{2} e^{-\lambda|\ell|}$, $\lambda > 0$:

$$\mathrm{VaR}_\alpha(L) = -\frac{1}{\lambda}\ln(2(1-\alpha)), \quad \mathrm{ES}_\alpha = \frac{1}{\lambda}\left[1 - \ln(2(1-\alpha))\right], \quad \alpha > \frac{1}{2}.$$

- Pareto distribution with index p, i.e. density $f(\ell) = p\ell^{-p-1}1_{\ell \geq 1}, p > 1$:

$$\mathrm{VaR}_\alpha(L) = (1-\alpha)^{-1/p}, \quad \mathrm{ES}_\alpha = \frac{p}{p-1}(1-\alpha)^{-1/p}.$$

Comparison Between VaR and ES

- VaR
  - Introduced in the early 90's by JP Morgan (RiskMetrics).
  - Standard in the financial sector.
  - Basel III based on VaR.
- ES
  - Used more and more often by fund managers and in insurance.
  - Discussion for replacing $\mathrm{VaR}(99\%)$ by $\mathrm{ES}(97.5\%)$ in Basel regulation.
- Others
  - Similar estimation method
  - ES is coherent but not VaR
  - VaR is defined for any distribution law while ES requires integrable tail distribution (e.g. ∞ for the Cauchy distribution).

Aggregation of Risks and Coherent Risk Measures

Define concepts and reasonable properties to take into account the aggregation and diversification of risks, leading to the class of coherent risk measures.

- $(\Omega, \mathcal{F})$ is a probability space, and L is the set of random variables on $(\Omega, \mathcal{F})$. An element $L \in \mathcal{C}$ represents a portfolio loss over a horizon h. We assume that $\mathcal{C}$ is convex.
- A risk measure is a function $\rho : \mathcal{C} \to \mathbb{R}$, which is law-invariant. $\rho(L)$ is interpreted as the amount of equity that must be added to the initial position for it to become acceptable to a regulator.
- A position L such that $\rho(L) \leq 0$ is acceptable without additional capital; if $\rho(L) < 0$, capital can even be withdrawn from the position.

(IT) Invariance by Translation

For all $L \in \mathcal{C}$, we have:

$$\rho(L + \ell) = \rho(L) + \ell, \quad \forall \ell \in \mathbb{R}.$$

Interpretation: Axiom (IT) formulates the requirement for capital: if $\rho(L) > 0$, adding the capital $\rho(L)$ to the initial position leads to an adjusted loss $\bar{L} = L - \rho(L)$ with $\rho(\bar{L}) = 0$, so that the position becomes acceptable. A measure ρ satisfying (IT) is called a monetary risk measure.

Remark: We have already seen that VaR and ES satisfy the axiom (IT).

(M) Monotonicity

For all $L_1, L_2 \in \mathcal{C}$, if $L_1 \leq L_2$ a.s., then:

$$\rho(L_1) \leq \rho(L_2).$$

Interpretation: A position with a higher loss in all states of the world requires more capital.

Remark: If $L_1 \leq L_2$ then $F_{L_2}(l) = P(L_2 \leq l) \leq \mathbb{P}(L_1 \leq l) = F_{L_1}(l)$. (stochastic dominance of first order), from which we deduce that: $\mathrm{VaR}_\alpha(L_1) \leq \mathrm{VaR}_\alpha(L_2)$, i.e. VaR satisfies (M). By integration, we also deduce that ES satisfies (M).

(Sub) Sub-additivity

For all $L_1, L_2 \in \mathcal{C}$, we have:

$$\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L_2).$$

Interpretation and advantages:

- The sub-additivity property encourages financial institutions to aggregate their positions to reduce risk, i.e., the capital required by the regulator.
- If $L = L_1 + \cdots + L_n$, where L_i represents the position of the internal unit i, then:

  $$\rho(L) \leq \rho(L_1) + \cdots + \rho(L_n).$$

  The estimation of partial risk $\rho(L_i)$ is generally more precise, and thus, $\sum_{i=1}^{n} \rho(L_i)$ gives a reliable estimate for the aggregated risk $\rho(L)$.

However:

- The axiom of sub-additivity is sometimes subject to controversy, particularly because it excludes, in general, the VaR (Value-at-Risk) measure, as we shall see later.
- Sub-additivity is satisfied by the Expected Shortfall (ES) risk measure.

(PH) Positive Homogeneity

For all $L \in \mathcal{C}$, we have:

$$\rho(aL) = a\rho(L), \quad \forall a \geq 0.$$

Interpretation and remarks:

- The axiom (PH) means that when one changes the currency (or numéraire), the risk is modified accordingly.
- Both Value-at-Risk (VaR) and Expected Shortfall (ES) satisfy (PH).
- A risk measure satisfying both (Sub) and (PH) is convex (Conv):

$$\rho(\lambda L_1 + (1-\lambda)L_2) \leq \lambda\rho(L_1) + (1-\lambda)\rho(L_2), \quad \forall L_1, L_2 \in L, \ \lambda \in [0,1].$$

- However, (PH) is sometimes criticized, especially in illiquid markets where the risk of n shares of a position L, for large n, might be strictly larger than n times the risk of L. This is not satisfied with (PH).
- This criticism has led to the replacement of (Sub) and (PH) by the weaker property of convexity (Conv).

Coherent Risk Measure

A risk measure $\rho : \mathcal{C} \to \mathbb{R}$ is said to be coherent if it satisfies the following four axioms:

1. (IT)
2. (M)
3. (Sub)
4. (PH)

◦ Consequences:

- (PH) + (IT) imply that $\rho(0) = 0$, and more generally $\rho(c) = c$ for any constant c. If the loss c occurs with certainty, an accounting provision of c is required.
- (M) implies that if $L \geq 0$, then $\rho(L) \geq 0$. If the loss is certain, the funds must be deposited.

VaR is not sub-additive (hence not coherent)

Example: Consider a portfolio of $d = 100$ bonds which may default with initial value 100 and nominal 105 at maturity in 1 year.

- The defaults are independent and occur with probability $p = 2\%$ for each bond.
- The loss of bond i is:

$$L_i = 100 - 105(1 - Y_i) = 105Y_i - 5$$

where Y_i is the default indicator: $Y_i = 1$ if default occurs, otherwise 0. Hence $Y_i \sim \mathcal{B}(p)$ and

$$L_i = \begin{cases} 100, & \text{with probability } p = 2\% \\ -5, & \text{with probabiity } 1 - p = 98\%. \end{cases}$$

VaR is not sub-additive (hence not coherent)

Consider two portfolios, each with initial value 10,000 euros:

- Portfolio A: 100 shares in one bond: $L_A = 100 L_1 = 10500 Y_1 - 500$
- Portfolio B: one share in each bond:
  $L_B = \sum_{i=1}^{100} L_i = 105 \sum_{i=1}^{100} Y_i - 500 = 105S - 500, \quad S \sim \mathcal{B}(100, 2\%)$.

Note that $\mathbb{P}(L_1 \leq -5) = 0,98$ and for $l < -5$, $\mathbb{P}(L_1 \leq l) = 0 < 0,95$, hence $\text{VaR}_\alpha(L_1) = -5$ and

$$\text{VaR}_{0.95}(L_A) = 100\text{VaR}_{0.95}(L_1) = -500$$

and, since $\mathbb{P}(S \leq 5) \approx 0,984 \geq 0,95$, $\mathbb{P}(S \leq 4) \approx 0,949 < 0,95$, one has $\text{VaR}_{0.95}(S) = 5$ and

$$\text{VaR}_{0.95}(L_B) = 105\text{VaR}_{0.95}(S) - 500 = 525 - 500 = 25.$$

Conclusion: Measuring risk with VaR can lead to nonsensical results!

$$\text{VaR}_{0.95}(\sum_{i=1}^{100} L_i) = 25 > -500 = \text{VaR}_{0.95}(100L_1) = \sum_{i=1}^{100} \text{VaR}_{0.95}(L_i).$$

Remarks

- In the previous example, the non-sub-additivity of VaR arises due to the fact that the i.i.d. loss variables L_i have a strongly asymmetric distribution (high skewness), typical of bond portfolios with defaults.
- There are other counter-examples of sub-additivity of VaR for distributions law with zero skewness but with fat distribution tails, like the Cauchy distribution (density $f(x) = \frac{1}{\pi(1+x^2)}$) or the Pareto distribution (density $f(x) = p/x^{p+1}\mathbf{1}_{x\geq 1}$, $p > 0$).
- On the other hand, VaR is sub-additive for Gaussian variables and, more generally, for random variables with elliptical distributions.

VaR for Gaussian variables

Let us consider a model with N sources of risk where the loss L_i over a period is given by:

$$L_i = a_i + b_i Z + \varepsilon_i, \quad i = 1, \dots, N,$$

where $Z \sim \mathcal{N}(0,1)$, and $(\varepsilon_i)_{i=1}^N$ are i.i.d. white noises with law $\mathcal{N}(0, \sigma_i^2)$, independent of Z. The parameters are a_i, b_i. The variable Z is interpreted as a common risk factor, and the ε_i are idiosyncratic risks.

The global loss is:

$$L = \sum_{i=1}^N L_i = a + bZ + \varepsilon,$$

with $a = \sum_{i=1}^N a_i$, $b = \sum_{i=1}^N b_i$, and

$$\varepsilon = \sum_{i=1}^N \varepsilon_i \sim \mathcal{N}(0, \sigma^2), \quad \sigma^2 = \sum_{i=1}^N \sigma_i^2.$$

Contagion effect

The loss L_i follows a Gaussian distribution $\mathcal{N}(a_i, b_i^2 + \sigma_i^2)$, while the global loss $L \sim \mathcal{N}(a, b^2 + \sigma^2)$. From the affine transformation property of VaR, we have:

$$\text{VaR}_\alpha(L_i) = a_i + \sqrt{b_i^2 + \sigma_i^2}\,\Phi^{-1}(\alpha), \quad i = 1, \dots, N,$$

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

The coefficient $b = \sum_{i=1}^N b_i$ depends on the correlations between the losses L_i and Z. The larger $|b|$ is, the larger $\text{VaR}_\alpha(L)$ becomes. b^2 is a measure of contagion.

In particular, if $b = 0$, we say that the risk field is protected against the common risk factor.

Diversification

It holds

$$\mathsf{VaR}_\alpha(L) - \sum_{i=1}^{N} \mathsf{VaR}_\alpha(L_i) = \left[\sqrt{b^2+\sigma^2} - \sum_{i=1}^{N}\sqrt{b_i^2+\sigma_i^2}\right]\Phi^{-1}(\alpha).$$

By writing:

$$\left(\sum_{i=1}^{N}\sqrt{b_i^2+\sigma_i^2}\right)^2 = \sum_{i=1}^{N}(b_i^2+\sigma_i^2) + \sum_{i\neq j}\sqrt{(b_i^2+\sigma_i^2)(b_j^2+\sigma_j^2)},$$

we have:

$$\sqrt{b^2+\sigma^2} - \sum_{i=1}^{N}\sqrt{b_i^2+\sigma_i^2} = \frac{b^2+\sigma^2 - \sum_{i=1}^{N}(b_i^2+\sigma_i^2) - \sum_{i\neq j}\sqrt{(b_i^2+\sigma_i^2)(b_j^2+\sigma_j^2)}}{\sqrt{b^2+\sigma^2}+\sum_{i=1}^{N}\sqrt{b_i^2+\sigma_i^2}}$$

$$= \frac{\sum_{i\neq j}(b_ib_j - \sqrt{b_i^2+\sigma_i^2}\sqrt{b_j^2+\sigma_j^2})}{\sqrt{b^2+\sigma^2}+\sum_{i=1}^{N}\sqrt{b_i^2+\sigma_i^2}} \leq 0.$$

Thus, diversification reduces the risk.

ES is coherent

Proposition: ES is a coherent risk measure.

Proof: We already know that ES satisfies the properties of (IT), (M), and (PH). Let us show that ES is also sub-additive.

For random variables L_1 and L_2 with continuous distributions, and denoting $L_3 = L_1 + L_2$, we have:

$$(1-\alpha)\left[\mathsf{ES}_\alpha(L_1) + \mathsf{ES}_\alpha(L_2) - \mathsf{ES}_\alpha(L_3)\right] = \mathbb{E}\left[L_1\left(\mathbb{I}_{L_1 \geq \mathsf{VaR}_\alpha(L_1)} - \mathbb{I}_{L_3 \geq \mathsf{VaR}_\alpha(L_3)}\right)\right]$$

$$+\mathbb{E}\left[L_2\left(\mathbb{I}_{L_2 \geq \mathsf{VaR}_\alpha(L_2)} - \mathbb{I}_{L_3 \geq \mathsf{VaR}_\alpha(L_3)}\right)\right].$$

ES is coherent (continued)

Now, for $i = 1, 2$, the terms:

$$(L_i - \mathsf{VaR}_\alpha(L_i))\left(\mathbb{I}_{L_i \geq \mathsf{VaR}_\alpha(L_i)} - \mathbb{I}_{L_3 \geq \mathsf{VaR}_\alpha(L_3)}\right) \geq 0,$$

since the two factors in parentheses have the same sign. We deduce that:

$$\begin{aligned}(1-\alpha)&\left[\mathsf{ES}_\alpha(L_1) + \mathsf{ES}_\alpha(L_2) - \mathsf{ES}_\alpha(L_3)\right] \\ &\geq \mathsf{VaR}_\alpha(L_1)\mathbb{E}\left[\mathbb{I}_{L_1 \geq \mathsf{VaR}_\alpha(L_1)} - \mathbb{I}_{L_3 \geq \mathsf{VaR}_\alpha(L_3)}\right] \\ &+ \mathsf{VaR}_\alpha(L_2)\mathbb{E}\left[\mathbb{I}_{L_2 \geq \mathsf{VaR}_\alpha(L_2)} - \mathbb{I}_{L_3 \geq \mathsf{VaR}_\alpha(L_3)}\right] = 0,\end{aligned}$$

since $\mathbb{E}[\mathbb{I}_{L_i \geq \mathsf{VaR}_\alpha(L_i)}] = \mathbb{P}[L_i \geq \mathsf{VaR}_\alpha(L_i)] = 1 - \alpha$, for $i = 1, 2, 3$. Therefore, ES is sub-additive.

The special case of Elliptical distributions

Spherical distributions

Definition. An $\mathbb{R}^d$-valued random vector X has a spherical distribution if there exists a function $\psi_X : \mathbb{R}_+ \to \mathbb{R}$ such that the characteristic function of X satisfies:

$$\varphi_X(u) := \mathbb{E}\left[\exp(iu^\top X)\right] = \psi_X(\|u\|^2), \quad u \in \mathbb{R}^d.$$

We then denote $X \sim S_d(\psi_X)$.

Lemma: Let $X : \Omega \to \mathbb{R}^d$ be a random variable and $\varphi_X : \mathbb{R}^d \to \mathbb{R}, u \mapsto \mathbb{E}(e^{i\langle u, X\rangle})$ its characteristic function. The following assertions are equivalent:

(i) For each orthogonal linear map $O : \mathbb{R}^d \to \mathbb{R}^d$, one has $OX \sim X$.

(ii) There is a function $\psi_X : \mathbb{R}^+ \to \mathbb{R}$ with $\varphi_X(u) = \psi_X(\|u\|^2)$, i.e. $X \sim S_d(\psi_X)$.

(iii) For each $a \in \mathbb{R}^d$, we have $\langle a, X\rangle \sim \|a\| X_1$, where X_1 is the first component of the vector X.

Proof

- (i) ⇒ (ii): For each orthogonal linear map O and each $u \in \mathbb{R}^d$, we have

$$\varphi_X(u) = \varphi_{OX}(u) = \mathbb{E}\left(e^{i\langle u, OX\rangle}\right) = \mathbb{E}\left(e^{i\langle O^T u, X\rangle}\right) = \varphi_X(O^T u).$$

  The characteristic function $\varphi_X(\cdot)$ is therefore invariant under orthogonal transformations, and the property (ii) follows.
- (ii) ⇒ (iii): Assume $a \in \mathbb{R}^d$. Then we get for each $t \in \mathbb{R}$,

$$\varphi_{\langle a, X\rangle}(t) = \mathbb{E}\left(e^{it\langle a, X\rangle}\right) = \mathbb{E}\left(e^{i\langle ta, X\rangle}\right) = \varphi_X(ta) = \psi_X(t^2\|a\|^2).$$

  On the other hand, we have

$$\varphi_{\|a\|X_1}(t) = \mathbb{E}\left(e^{it\|a\|X_1}\right) = \mathbb{E}\left(e^{i\langle t\|a\|e_1, X\rangle}\right) = \varphi_X(t\|a\|e_1) = \psi_X(t^2\|a\|^2),$$

  and the property (iii) follows.
- (iii) ⇒ (i): We have

$$\varphi_{OX}(u) = \mathbb{E}\left(e^{i\langle u, OX\rangle}\right) = \mathbb{E}\left(e^{i\langle O^T u, X\rangle}\right) = \varphi_{\langle O^T u, X\rangle}(1) = \varphi_{\|O^T u\|X_1}(1)$$
$$= \varphi_{\|u\|X_1}(1) = \varphi_X(u)$$

  which shows that $\varphi_X(u)$ is invariant under orthogonal transformations,

Examples

∘ Normal distribution: If $X \sim N_d(0, I_d)$ then

$$\varphi_X(u) = \exp(-\frac{1}{2}\|u\|^2) = \psi_X(\|u\|^2), \quad \psi(t) = \exp(-\frac{1}{2}t).$$

∘ Normal mixture:

The $\mathbb{R}^d$-valued random vector X is said to have a multivariate normal variance mixture distribution if:

$$X \equiv \mu + \sqrt{W} AZ \sim M_d(\mu, \Sigma, \hat{F}_W)$$

where:

- $Z \sim N_k(0, I_k)$
- $W \geq 0$ is a nonnegative random variable, independent of Z, $\hat{F}_W(\theta) := \mathbb{E}[\exp(-\theta W)]$ (Laplace-Stieljes transform).
- $A \in \mathbb{R}^{d \times k}$ and $\mu \in \mathbb{R}^d$ are constants

$$\varphi_X(u) = \mathbb{E}[\mathbb{E}[e^{iu^T X}|W]] = \exp(iu^T\mu)\hat{F}_W(\frac{1}{2}u^T\Sigma u).$$

$\rightsquigarrow$ If $\mu = 0$ and $\Sigma = AA^T = I_d$ then $X \sim S_d(\psi_X)$ with $\psi_X = \hat{F}_W(\frac{1}{2}t)$.

Elliptical Distributions

- Definition:
  An $\mathbb{R}^d$-valued random vector X has an *Elliptical distribution* if:

$$X \equiv \mu + AY,$$

  where $Y \sim S_k(\psi)$ (a spherical distribution), $A \in \mathbb{R}^{d\times k}$, and $\mu \in \mathbb{R}^d$ are constants.
- Characteristic function:
  The characteristic function of an elliptical distribution is given by:

$$\varphi_X(u) = \mathbb{E}[e^{iu^T X}] = e^{iu^T \mu}\psi(u^T \Sigma u),$$

  where $\Sigma = AA^\top$. We then denote $X \sim E_d(\mu, \Sigma, \psi)$
- Examples:
  - Multivariate normal distribution: $X \sim N_d(\mu, \Sigma)$ has an elliptical distribution.
  - Normal mixture: $X \sim M_d(\mu, \Sigma, \hat{F}_W)$ then $X \sim E_d(\mu, \Sigma, \psi)$ with $\psi(t) = \hat{F}_W(t/2)$.
  - Multivariate t-distribution: An elliptical distribution with heavier tails compared to the normal distribution.

Sub-additivity of VaR for Elliptical Distributions

Proposition: Let $X \sim E_d(\mu, \Sigma, \psi)$. Then, for any $u, w \in \mathbb{R}^d$, and $\alpha \in [0, 1]$:

$$\mathrm{VaR}_\alpha(u^\top X + w^\top X) \leq \mathrm{VaR}_\alpha(u^\top X) + \mathrm{VaR}_\alpha(w^\top X).$$

Proof: We have $X \equiv \mu + AY$ with $AA^\top = \Sigma$ and $Y \sim S_d(\psi)$. From the proposition on spherical distribution, for any $u \in \mathbb{R}^d$:

$$u^\top X \overset{d}{=} u^\top \mu + \|A^\top u\| Y_1.$$

This implies that for any $u, w \in \mathbb{R}^d$ and $\alpha \in [0, 1]$:

$$\mathrm{VaR}_\alpha(u^\top X + w^\top X) = (u + w)^\top \mu + \|A^\top (u + w)\| \mathrm{VaR}_\alpha(Y_1).$$

The triangle inequality gives

$$\|A^\top (u + w)\| \leq \|A^\top u\| + \|A^\top w\|.$$

Therefore, we have:

$$\begin{aligned} \mathrm{VaR}_\alpha(u^\top X + w^\top X) &\leq u^\top \mu + w^\top \mu + (\|A^\top u\| + \|A^\top w\|) \mathrm{VaR}_\alpha(Y_1) \\ &= \mathrm{VaR}_\alpha(u^\top X) + \mathrm{VaR}_\alpha(w^\top X). \end{aligned}$$

Other examples of risk measures

Other examples of coherent risk measures

Expected Shortfall (ES) is a coherent risk measure, defined as:

$$\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L)du$$

◦ Construction of new coherent risk measures on the basis of existing coherent risk measures.

◦ Spectral risk measures

The ES can be directly generalized to take into account individual risk aversion. Instead of averaging over all $\mathrm{VaR}_z(X)$ for $z \geq \alpha$ with a uniform weight, one can employ a more general weighting function ϕ.

Definition

Let $(A, \mathcal{A}, \mu)$ be a probability space with σ-Algebra $\mathcal{A}$ and probability measure μ. Then an integrable map $\phi : A \to \mathbb{R}$ is called a weight function, if ϕ has the following properties:

(i) $\phi(\alpha) \geq 0$ for almost every $\alpha \in A$,

(ii) $\int_A \phi(\alpha)\, d\mu(\alpha) = 1$.

Definition (Spectral Risk Measure)

Let $\phi \in L^1([0,1])$ be a weight function. The risk measure

$$M_\phi(X) = \int_0^1 \mathsf{VaR}_p(X)\phi(p)\,dp$$

is called the spectral measure of ϕ.

◦ The concept of a spectral measure allows the representation of an individual profile of risk aversion.
◦ The VaR is a limit case of spectral measures

$$\mathsf{VaR}_\alpha(X) = \int_0^1 \mathsf{VaR}_p(X)\delta_\alpha(p)\,dp,$$

where δ_α denotes the Dirac distribution.

Theorem

Let $(A, \mathcal{A}, \mu)$ be a probability space with σ-Algebra $\mathcal{A}$ and probability measure μ. Let $\{\rho_\alpha\}_{\alpha \in A}$ be a family of risk measures and M a vector space of real-valued random variables X, such that $\rho_\alpha(X)$ are μ-almost everywhere defined and μ-integrable. If all ρ_α are translation invariant, positively homogeneous, monotone, and subadditive, then the risk measure

$$\rho : M \to \mathbb{R}, \quad X \mapsto \rho(X) = \int_A \rho_\alpha(X) d\mu(\alpha)$$

also has the corresponding property.

Proof.

Let $c \in \mathbb{R}$ and X, Y be arbitrary random variables.

- Translation invariance: since μ is a probability measure, $\rho(X+c) = \int_A \rho_\alpha(X+c)d\mu(\alpha) = \int_A (\rho_\alpha(X)+c)\, d\mu(\alpha) = \rho(X)+c$,
- Positive homogeneity: For $c \geq 0$,

$$\rho(cX) = \int_A \rho_\alpha(cX)d\mu(\alpha) = \int_A c\rho_\alpha(X)d\mu(\alpha) = c\rho(X).$$

- Monotony: If $X \geq Y$ almost everywhere, then $\rho_\alpha(X) \geq \rho_\alpha(Y)$, so

$$\rho(X) = \int_A \rho_\alpha(X)d\mu(\alpha) \geq \int_A \rho_\alpha(Y)d\mu(\alpha) = \rho(Y).$$

- Subadditivity:

$$\rho(X+Y) = \int_A \rho_\alpha(X+Y)d\mu(\alpha) \leq \int_A (\rho_\alpha(X)+\rho_\alpha(Y))\, d\mu(\alpha) = \rho(X)+\rho(Y).$$

Thus, the risk measure ρ inherits all the properties of the ρ_α. □

Coherence of spectral risk measures

Theorem (Coherence of spectral risk measures)

A spectral measure M_ϕ is coherent, if the weight function ϕ is (almost everywhere) monotone increasing.

Examples of Spectral Risk Measures

- For $\phi(u) = \frac{1}{1-\alpha} 1_{[0,1-\alpha]}(u)$, we recover Expected Shortfall (ES).
- Other choices of $\phi(u)$ lead to different spectral risk measures that emphasize extreme losses.

Proof.

Since ϕ is monotone increasing, we can define a measure on $([0,1],\mathcal{B})$ by $\phi(p) := \nu([0,p])$. By Fubini's theorem, it follows that:

$$M_\phi(X) = \int_0^1 \mathsf{VaR}_p(X)\phi(p)dp = \int_0^1 \mathsf{VaR}_p(X)\left(\int_0^p d\nu(\alpha)\right)dp$$

$$= \int_0^1 \left(\int_0^1 1_{[0,p]}(\alpha)\mathsf{VaR}_p(X)d\nu(\alpha)\right)dp = \int_0^1 \left(\int_0^1 1_{[\alpha,1]}(p)\mathsf{VaR}_p(X)dp\right)d\nu(\alpha)$$

$$= \int_0^1 \left(\int_\alpha^1 \mathsf{VaR}_p(X)dp\right)d\nu(\alpha) = \int_0^1 (1-\alpha)\mathsf{ES}_\alpha(X)d\nu(\alpha)$$

where we used the identity $1_{[0,p]}(\alpha) = 1_{[\alpha,1]}(p)$ for $\alpha, p \in [0,1]$. The assertion now follows from the previous theorem with $d\mu(\alpha) = (1-\alpha)d\nu(\alpha)$, since:

$$\int_0^1 d\mu(\alpha) = \int_0^1 (1-\alpha)d\nu(\alpha) = \int_0^1 \left(\int_\alpha^1 dp\right)d\nu(\alpha)$$

$$= \int_0^1 \left(\int_0^1 1_{[\alpha,1]}(p)dp\right)d\nu(\alpha) = \int_0^1 \left(\int_0^1 1_{[0,p]}(\alpha)d\nu(\alpha)\right)dp = \int_0^1 \varphi(p)dp = 1.$$

□

Distortion Risk Measures

- Denote by Ψ the cumulative distribution function (CDF) on $[0, 1]$ with density ϕ, so that

$$M_\phi(L) = R_\Psi(L) = \int_0^1 F_L^{-1}(1-u)d\Psi(u)$$

- More generally, when Ψ is a CDF on $[0, 1]$, called a distortion function, R_Ψ is called a distortion risk measure.

- In the particular case where Ψ is the distribution function of the Dirac law in $1-\alpha$, i.e., $\Psi(x) = 1_{x \geq 1-\alpha}$, we have:

$$R_\Psi(L) = \mathsf{VaR}_\alpha(L)$$

Wang Risk Measure

- Assume for simplification that F_L is invertible, i.e., F_L is continuous and strictly increasing.
- By integration by parts and a change of variable $(u \mapsto 1-u)$, we have:

$$\begin{aligned} R_\Psi(L) &= \int_0^{1-F_L(0)} F_L^{-1}(1-u)d\Psi(u) + \int_{1-F_L(0)}^1 F_L^{-1}(1-u)d[\Psi(u)-1] \\ &= -\int_0^{1-F_L(0)} \Psi(u)dF_L^{-1}(1-u) - \int_{1-F_L(0)}^1 [\Psi(u)-1]dF_L^{-1}(1-u) \\ &= \int_{F_L(0)}^1 \Psi(1-u)dF_L^{-1}(u) + \int_0^{F_L(0)} [\Psi(1-u)-1]dF_L^{-1}(u) \end{aligned}$$

- Further, with a change of variable $u = F_L(l)$, we get the formula of Wang risk measure:

$$R_\Psi(L) = \int_0^{+\infty} \Psi(F_L^c(l))dl - \int_{-\infty}^0 [1-\Psi(F_L^c(l))]dl$$

- Here, $F_L^c(l) = 1 - F_L(l)$ represents the survival function.

◦ The interpretation is the following: the initial survival function F^c_L is replaced by a survival function $\Psi(F^c_L)$ and the integral in R_Ψ is called the Choquet integral or distorted expectation.

◦ When $\Psi(u) = u$, we recover the usual integral and expectation:

$$R_\Psi(L) = \int_0^{+\infty} \mathbb{E}[1_{L>\ell}]d\ell - \int_{-\infty}^0 \mathbb{E}[1_{L\leq\ell}]d\ell$$

$$= \mathbb{E}\left[\int_0^{+\infty} 1_{L>\ell}d\ell - \int_{-\infty}^0 1_{L\leq\ell}d\ell\right] = \mathbb{E}[L].$$

◦ When Ψ is concave, the Choquet integral gives more weight to the large values of L (extreme risks), and one shows that R_Ψ is sub-additive, which is consistent with the decreasing monotonicity of $\psi = \Psi'$ when Ψ admits a density.

Examples

- Distortion risk measure with proportional hazard rate: This corresponds to a distortion function:

$$\Psi(u) = u^p, \quad u \in [0,1], \text{ and } p > 0.$$

When $p < 1$, Ψ is concave : the extreme losses are over-weighted. The associated risk measure R_Ψ is sub-additive.

- Exponential distortion risk measure: this corresponds to a distortion function:

$$\Psi(u) = \frac{1 - e^{-pu}}{1 - e^{-p}}, \quad u \in [0,1], \text{ and } p > 0,$$

which is concave.

Coherence and Independence

- We could think that the risk of two independent risks aggregates together, i.e., $\rho(L_1 + L_2) = \rho(L_1) + \rho(L_2)$ for independent L_1 and L_2.
- It is wrong in general!
- Let L_1, L_2 be i.i.d. centered Gaussian. Then $L_1 + L_2 \sim \sqrt{2}L_1$, and thus for a risk measure satisfying (PH) (e.g., VaR and ES):

$$\rho(L_1 + L_2) = \rho(\sqrt{2}L_1) = \sqrt{2}\rho(L_1) < 2\rho(L_1) = \rho(L_1) + \rho(L_2)$$

whenever $\rho(L_1) > 0$.

Computing the VaR and ES in practice

◦ We here focus on non-parametric approaches which rely on an i.i.d. sample $X_1, \cdots, X_n$ of size n with the same law as X with cdf F.

◦ A natural idea to estimate $\text{VaR}_\alpha(X) = F^{-1}(\alpha)$ is use the order statistics

$$X_{(1)} = \min_{1\leq k\leq n} X_k \leq X_{(2)} \leq \cdots \leq X_{(n-1)} \leq X_{(n)} = \max_{1\leq k\leq n} X_k$$

defined by sorting the realizations of $X_1, \cdots, X_n$ in increasing order.

◦ We then estimate $\text{VaR}_\alpha(X)$ by $X_{(\lceil n\alpha \rceil)}$ where $\lceil x \rceil$ is the unique integer s.t. $\lceil x \rceil - 1 < x \leq \lceil x \rceil$.

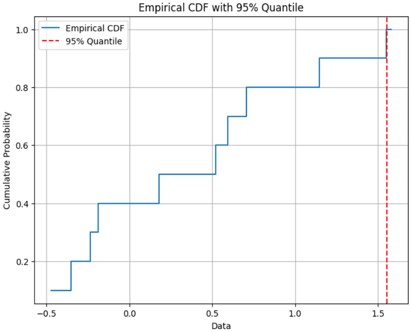
◦ Remark: One can estimate $\text{VaR}_\alpha(X)$ by

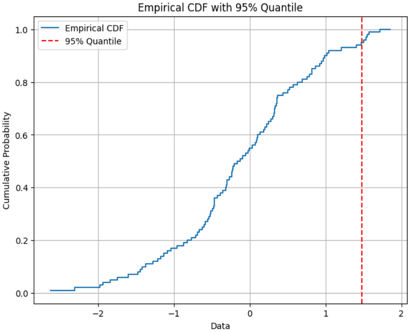
$$\begin{cases} X_{((n+1)\alpha)} & \text{if } (n+1)\alpha \text{ is an integer.} \\ \frac{1}{2}(X_{(\lfloor (n+1)\alpha \rfloor)} + X_{(\lfloor (n+1)\alpha \rfloor)+1}) & \text{otherwise.} \end{cases}$$

◦ Example: $n = 100$ and $\alpha = 95\%$ then we estimate the 95% quantile by $X_{(95)}$.

Examples

- Empirical cdf with 10 and 100 samples.





Computing $X_{(\lceil n\alpha \rceil)}$ is nothing but the α-quantile of the empirical cdf of the data

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k \le x} = \begin{cases} 0, & \text{if } x \le X_{(1)}, \\ i/n, & \text{if } X_{(i)} \le x < X_{(i+1)}, \\ 1, & \text{if } x \ge X_{(n)}. \end{cases}$$

Fix $\alpha \in (0,1)$ and select i s.t. $\frac{i-1}{n} < \alpha \le \frac{i}{n}$ so that $i - 1 < n\alpha \le i \Leftrightarrow \lceil n\alpha \rceil = i$.

Recalling that

$$F_n^{-1}(\alpha) = \inf\{x : F_n(x) \ge \alpha\}$$

we get

$$F_n^{-1}(\alpha) = X_{(i)} = X_{(\lceil n\alpha \rceil)}.$$

◦ As a direct application of the LLN and CLT, for any $x \in \mathbb{R}$

$$F_n(x) \overset{a.s.}{\to} F(x) \text{ and } \sqrt{n}(F_n(x) - F(x)) \overset{d}{\Rightarrow} \mathcal{N}(0, F(x)(1 - F(x)), \quad \text{as} \quad n \uparrow \infty.$$

◦ According to the Glivenko-Cantelli theorem, if F is continuous, then

$$\|F_n - F\|_\infty := \sup_{x \in \mathbb{R}} |(F_n - F)(x)| \overset{a.s.}{\to} 0, \quad \text{as} \quad n \uparrow \infty.$$

Theorem

Assume that F is continuous and increasing. Then, for any $\alpha \in (0,1)$,

$$F_n^{-1}(\alpha) \xrightarrow{a.s.} F^{-1}(\alpha), \quad \text{as} \quad n \uparrow \infty.$$

Proof.

Since F is invertible and F^{-1} is continuous, it suffices to prove that

$$F(F_n^{-1}(\alpha)) \xrightarrow{a.s.} F(F^{-1}(\alpha)) = \alpha, \quad \text{as} \quad n \uparrow \infty.$$

Then, we write

$$\begin{aligned} |F(F_n^{-1}(\alpha)) - F(F^{-1}(\alpha))| &\leq |F(F_n^{-1}(\alpha)) - F_n(F_n^{-1}(\alpha))| \\ &\quad + |F_n(F_n^{-1}(\alpha)) - F(F^{-1}(\alpha))| \\ &\leq \|F_n - F\|_\infty + \left|\frac{\lceil n\alpha \rceil}{n} - \alpha\right| \\ &\to 0, \quad \text{as} \quad n \uparrow \infty, \end{aligned}$$

using the Glivenko-Cantelli theorem for the first term. □

Computation of the ES

◦ Regarding the ES, a simple idea consists in writing

$$\mathsf{ES}_\alpha(X) = \frac{1}{1-\alpha}\mathbb{E}[X\mathbf{1}_{X\geq \mathsf{VaR}_\alpha(L)}] \approx \frac{1}{1-\alpha}\frac{1}{n}\sum_{i=1}^{n} X_i \mathbf{1}_{X_i \geq X_{(\lceil n\alpha \rceil)}} = \hat{\mathsf{ES}}_\alpha(X).$$

Notice that

$$\hat{\mathsf{ES}}_\alpha(X) = \frac{1}{1-\alpha}\frac{1}{n}\sum_{i=\lceil n\alpha\rceil}^{n} X_{(i)}$$

which is computed using the same sample $X_1, \cdots, X_n$ as the one used to compute $F_n^{-1}(\alpha)$.

Theorem

Assume that $X \in L^1(\mathbb{P})$ and that its cdf is continuous and increasing. Then, it holds

$$\hat{ES}_\alpha(X) \xrightarrow{a.s.} ES_\alpha(X) \quad \text{as} \quad n \to \infty.$$

Proof.

Step 1: prove the decomposition

$$\hat{\mathrm{ES}}_\alpha(X) = \mathrm{VaR}_\alpha(X) + \frac{1}{1-\alpha}\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathrm{VaR}_\alpha(X))_+$$
$$+ X_{(\lceil n\alpha \rceil)} - \mathrm{VaR}_\alpha(X) + \frac{1}{1-\alpha}\frac{1}{n}\sum_{i=1}^{n}(X_i - X_{(\lceil n\alpha \rceil)})_+ - (X_i - \mathrm{VaR}_\alpha(X))_+$$
$$+ \frac{1}{1-\alpha}X_{(\lceil n\alpha \rceil)}(\alpha - \frac{\lceil \alpha n \rceil}{n})$$

Step 2: prove that as $n \uparrow \infty$

- $A_n \xrightarrow{a.s.} \mathrm{ES}_\alpha(X)$ (LLN)
- $B_n \xrightarrow{a.s.} 0$ (Lipschitz reg x_+ + $X_{(\lceil n\alpha \rceil)} \xrightarrow{a.s.} \mathrm{VaR}_\alpha(X)$)
- $C_n \xrightarrow{a.s.} 0.$

Stochastic approximation point of view for the VaR-ES

◦ We here present another point of view to compute the couple (VaR, ES). We first remark that if the cdf of X is continuous and increasing then the VaR is the unique solution to

$$\mathbb{P}(X \leq \xi) = \alpha \Leftrightarrow \mathbb{E}[H_1(\xi, X)] = 0, \quad \text{with} \quad H_1(\xi, X) := \mathbf{1}_{X \leq \xi} - \alpha$$

A natural idea to compute the (unique) zero of $h_1(\xi) = \mathbb{E}[H_1(\xi, X)]$ is to use the (online) Robbins-Monro algorithm with dynamics

$$\xi_{k+1} = \xi_k - \gamma_{k+1} H_1(\xi_k, X_{k+1}) = \xi_k - \gamma_{k+1}(h_1(\xi_k) + \varepsilon_{k+1}),$$

where $(X_k)_{k\geq 1}$ is an i.i.d. sequence with the same law as X and ξ_0 is a real-valued random variable independent of $(X_k)_{k\geq 1}$.
Here, $(\gamma_k)_{k\geq 1}$ is a deterministic decreasing and positive sequence satisfying

$$\sum_{n\geq 1} \gamma_n = \infty \quad \text{and} \quad \sum_{n\geq 1} \gamma_n^2 < \infty.$$

◦ Example: $\gamma_n = \gamma n^{-\beta}$, with $\beta \in (1/2, 1]$ and $\gamma > 0$.

What about the ES?

- A natural idea is to proceed as before

$$\mathsf{ES}_\alpha(X) = \frac{1}{1-\alpha}\mathbb{E}[X\mathbf{1}_{X\geq \mathsf{VaR}_\alpha(L)}] \approx \frac{1}{1-\alpha}\frac{1}{n}\sum_{k=1}^{n} X_k \mathbf{1}_{X_k \geq \xi_{k-1}} = C_n, \quad n \geq 1$$

Notice that the sequence $(C_n)_{n\geq 0}$ (with $C_0 = 0$) defined above can be written in the recursive form

$$C_{k+1} = C_k - \frac{1}{k+1} H_2(\xi_k, C_k, X_{k+1}), \text{ with } H_2(\xi, C, x) := C - x\mathbf{1}_{x\geq \xi}$$

The resulting (online) stochastic algorithm reads as

$$\begin{cases} \xi_{k+1} & = \xi_k - \gamma_{k+1} H_1(\xi_k, X_{k+1}) \\ C_{k+1} & = C_k - \frac{1}{k+1} H_2(\xi_k, C_k, X_{k+1}) \end{cases}$$

A general convergence result

The convergence of the Robbins-Monro algorithm and the stochastic gradient descent algorithm can be framed as the following general result.

Theorem

Define $h(z) = \mathbb{E}[H(z, X)]$, $H : \mathbb{R}^q \times \mathbb{R}^d \to \mathbb{R}^d$. *Let* $T^\star = \{h = 0\}$. *Assume that the following mean-reverting assumption is satisfied:*

$$\forall z \in \mathbb{R}^d \backslash T^\star, \forall z^\star \in T^\star, \quad \langle z - z^\star, h(z)\rangle > 0,$$

and

$$\mathbb{E}[|H(z, X)|^2] \leq C(1 + |z|^2).$$

Then, the sequence $(z_n)_{n\geq 0}$ *defined by*

$$z_{n+1} = z_n - \gamma_{n+1} H(z_n, X_{n+1}), \quad n \geq 0$$

where $(X_n)_{n\geq 1}$ *is an i.i.d. sequence of r.v. having the same distribution as* X *and* z_0 *is a r.v. independent of* $(X_n)_{n\geq 1}$ *satisfying* $\mathbb{E}[|z_0|^2] < \infty$, *satisfies*

$$z_n \xrightarrow{a.s.} z_\infty, \qquad \text{as} \quad n \uparrow \infty,$$

where z_∞ *is a r.v. taking values in* $T^\star$.

◦ We apply the above general theorem to $h_1(\xi) = \mathbb{E}[H_1(\xi, X)] = \mathbb{P}(X \leq \xi) - \alpha$.

- $\langle h_1(\xi), \xi - \xi^\star \rangle = (\mathbb{P}(X \leq \xi) - \alpha)(\xi - \xi^\star) > 0$, for all $\xi \neq \xi^\star$.
- $|H_1(\xi, X)|^2 \leq 2(1+\alpha^2) \leq 2(1+\alpha^2)(1+|\xi|^2) \Rightarrow \mathbb{E}[|H_1(\xi, X)|^2] \leq C(1+|\xi|^2)$ with $C := 2(1+\alpha^2)$.

$\rightsquigarrow$ the sequence $(\xi_n)_{n\geq 0}$ converges $a.s.$ to $\xi^\star = \mathrm{VaR}_\alpha(X)$.

◦ To prove the $a.s.$ convergence of $(C_n)_{n\geq 0}$, we use the following decomposition:

$$
\begin{aligned}
C_n &= \frac{1}{1-\alpha}\frac{1}{n}\sum_{k=1}^{n} X_k \mathbf{1}_{X_k \geq \xi_{k-1}} \\
&= \frac{1}{1-\alpha}\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}[X\mathbf{1}_{X\geq\xi}]_{|\xi=\xi_{k-1}} + \frac{1}{1-\alpha}\frac{1}{n}\sum_{k=1}^{n}(X_k\mathbf{1}_{X_k\geq\xi_{k-1}} - \mathbb{E}[X\mathbf{1}_{X\geq\xi}]_{|\xi=\xi_{k-1}}) \\
&=: \mathrm{A}_n + \mathrm{B}_n.
\end{aligned}
$$

- Ought to Cesarò's lemma and the continuity of $\xi \mapsto \mathbb{E}[X\mathbf{1}_{X\geq\xi}]$, one gets $\mathrm{A_n} \overset{a.s.}{\to} \frac{1}{1-\alpha}\mathbb{E}[X\mathbf{1}_{X\geq\xi^\star}] = \mathrm{ES}_\alpha(X)$, as $n \uparrow \infty$.

- It thus remains to prove that $\mathrm{B}_n \xrightarrow{a.s.} 0$ as $n \uparrow \infty$.

◦ Note that $\mathrm{B}_n := \frac{1}{n}\sum_{k=1}^{n}\varepsilon_k$, with $\varepsilon_k = \frac{1}{1-\alpha}(X_k\mathbf{1}_{X_k\geq\xi_{k-1}} - \mathbb{E}[X\mathbf{1}_{X\geq\xi}]_{|\xi=\xi_{k-1}})$.

◦ We introduce the filtration $\mathcal{F} = (\mathcal{F}_n)_{n\geq 1}$, $\mathcal{F}_n = \sigma(\xi_0, X_1, \cdots, X_n)$ and the process

$$N_n = \sum_{k=1}^{n}\frac{1}{k}\varepsilon_k, \quad n \geq 1.$$

◦ Note that since $X_k \perp\!\!\!\perp \mathcal{F}_{k-1}$, one has

$$\mathbb{E}[\varepsilon_k|\mathcal{F}_{k-1}] = \frac{1}{1-\alpha}(\mathbb{E}[X\mathbf{1}_{X\geq\xi}]_{\xi=\xi_{k-1}} - \mathbb{E}[X\mathbf{1}_{X\geq\xi}]_{\xi=\xi_{k-1}}) = 0$$

so that $(N_n)_{n\geq 1}$ is an $\mathcal{F}$-martingale.

◦ Assuming that $X \in L^2(\mathbb{P})$, for some compact set $\mathcal{K}$ containing $(\xi_n)_{n\geq 1}$, one has

$$\mathbb{E}[\varepsilon_k^2|\mathcal{F}_{k-1}] = \frac{\text{var}(X\mathbf{1}_{X\geq\xi_{k-1}}|\mathcal{F}_{k-1})}{(1-\alpha)^2} \leq \frac{\mathbb{E}[X^2\mathbf{1}_{X\geq\xi}]_{|\xi=\xi_{k-1}}}{(1-\alpha)^2} \leq \frac{\sup_{\xi\in\mathcal{K}}\mathbb{E}[X^2\mathbf{1}_{X\geq\xi}]}{(1-\alpha)^2}$$

Hence,

$$\langle N\rangle_\infty = \lim_n \langle N\rangle_n = \sum_{k\geq 1}\frac{1}{k^2}\mathbb{E}[\varepsilon_k^2|\mathcal{F}_{k-1}] < \infty \quad a.s.$$

which in turn yields the $a.s.$ convergence of $(N_n)_{n\geq 1}$. Using Kronecker's lemma, we conclude

$$\mathrm{B}_n = \frac{1}{n}\sum_{k=1}^{n}\varepsilon_k \xrightarrow{a.s.} 0, \quad \text{as} \quad n\uparrow\infty.$$

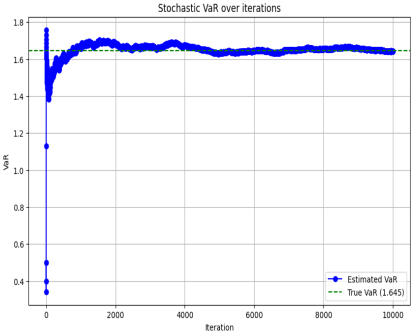
Conclusion: The (online) stochastic algorithm

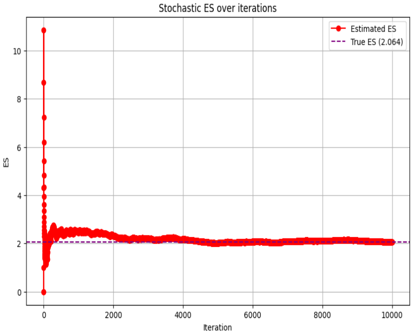
$$\begin{cases} \xi_{n+1} & = \xi_n - \gamma_{n+1}H_1(\xi_n, X_{n+1}) \\ C_{n+1} & = C_n - \frac{1}{n+1}H_2(\xi_n, C_n, X_{n+1}) \end{cases}$$

satisfies

$$(\xi_n, C_n) \xrightarrow{a.s.} (\mathsf{VaR}_\alpha(X), \mathsf{ES}_\alpha(X)), \quad \text{as} \quad n\uparrow\infty.$$

◦ Take $X \sim \mathcal{N}(0,1)$ and set $\gamma_n = 1/n^{\beta}$, $\beta = 0.8$, $\xi_0 = 0.5$, $C_0 = 1$, $\alpha = 95\%$ and $M = 10000$ iterations. $(\text{VaR}_{\alpha}(X), \text{ES}_{\alpha}(X)) = (1.645, 2.064)$.





Thank you!