Institut du Risque & de l'Assurance
Le Mans Université

LMM
Laboratoire Manceau de Mathématiques
Le Mans Université

Stochastic algorithms for systemic risk measures

Anis Matoussi

The author's research is part of
ANR project DREAMeS (ANR-21-CE46-0002) &

Chair Risques émergents ou atypiques en Assurance
(MMA-COVEA & Fondation du Risque de l'Institut Louis Bachelier)

Joint works with Zakaria Bensaid, Roxana Dumitrescu, Sarah Kaakai, Wissal Sabbagh and Achraf Tamtalini

FDD-FIME seminar, January 26, 2024, IHP-Paris

Introduction

- The theory of risk measure is widely used in the finance and insurance industry.
- P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risks.*Mathematical Finance*, Vol. 9, No. 3 (July 1999), 203-228.
- S. Peng : Nonlinear Expectations, Nonlinear Evaluations and Risk Measures, LNM 1856, M. Frittelli and W. Runggaldier (Eds.), pp. 165-253, 2004, Springer-Verlag Berlin Heidelberg 2004.
- P. Barrieu and N. El Karoui. Pricing, hedging and designing derivatives with risk measures. In R. Carmona, editor, Indifference Pricing, Theory and Applications, pages 77-146. Princeton Univ. Press, 2009.
- Many researchers took interest in risk measures as questions about capital requirement are arising.

Introduction

- Systemic risk measures were introduced to capture the global risk and the corresponding contagion effects that is generated by an interconnected system of financial institutions. Two approaches were suggested.:
  1. In the first one, systemic risk measures can be interpreted as the minimal amount of cash needed to secure a system after aggregating individual risks.
  2. In the second approach, systemic risk measures can be interpreted as the minimal amount of cash that secures a system by allocating capital to each single institution before aggregating individual risks. The latter is also known as Multivariate Sytemic Risk Measures (MSRM).
- We use stochastic algorithms schemes in estimating MSRM and we test numerically the performance of these algorithms on several examples [KMT22].
- Extension of numerical tests with real insurance data using ADAM (Adaptative Moment estimation) and SGDA (Stochastic Gradient Descent Ascent) [BDMS24].
- S. Kaakai, A.M., A. Tamtalini. Estimation of Systemic Shortfall Risk Measure using Stochastic Algorithms. hal-038711246 (2022), in revision ([KMT22]).
- Z. Bensaid, A.M., R. Dumitrescu, W. Sabbagh. Dynamic Multivariate Systemic Risk Measure and Deep learning algorithms. Work in progress ([BDMS24]).

Outline

1. About Risk Measures
2. Theory of Systemic Risk Measures
3. A bit of Stochastic Algorithms
4. Application to Systemic Risk Measure
5. Numerical experiments
6. From static multivariate SRM to dynamic multivariate SRM

Univariate Risk Measure

Definition

A monetary risk measure $\eta : \mathcal{L}^0(\mathbb{R}) \to \mathbb{R}$ is a map that **represents the minimal extra capital to secure a loss position X**, i.e. the minimal amount m that needed to be taken from X in order to make the resulting payoff acceptable at T :

$$\eta(X) := \inf\{m \in \mathbb{R} \mid X - m \in \mathbb{A}\}$$

Example of acceptance set :

$$\mathbb{A} := \{Z \in \mathcal{L}^0(\mathbb{R}) \mid \mathbb{E}(Z) \leq B\}, \ B \in \mathbb{R}$$

Example of Value at Risk and Some properties

- The Value at Risk at a level $\lambda \in]0,1[$ corresponds to the acceptance set:

$$\mathbb{A}^{\lambda} := \{Z \in \mathcal{L}^0(\mathbb{R}) \mid \mathbb{P}(Z < 0) \geq \lambda\}$$

- So $\mathrm{VaR}^{\lambda}(X) = \inf\{m \in \mathbb{R} \mid \mathbb{P}(X - m < 0) \geq \lambda\}$

Some characterizing features

- **Cash Additivity :**

$$\eta(X+m) = \eta(X)+m, \text{ for all } m \in \mathbb{R}$$

- **Convexity :** Reflects the effect of diversification:

$$\eta(\lambda X + (1-\lambda)Y) \leq \lambda\eta(X) + (1-\lambda)\eta(Y)$$

Outline

From one-dimensional to d-dimensional risk profiles

Consider a system of d interacting and heterogeneous portfolios and a vector of losses $\mathbf{X} = (X_1, ..., X_d)$ at T.

How to measure the risk carried by this system of portfolios ?
Or how to choose $\rho(\mathbf{X})$?

Simplest way: Sum up the risk measures of each contract : $\sum \eta_i(X_i)$

$\Rightarrow$ We ignore the dependence structure of our portfolios so we might overvalue the risk. $\Rightarrow$ No ranking of portfolios in terms of global riskiness.

First Aggregate, then inject cash

- We are interested in systemic risk of the form :

$$\rho(\mathbf{X}) = \eta(\mathbf{\Lambda}(\mathbf{X})) = \inf\{m \in \mathbb{R} \mid \Lambda(\mathbf{X}) - m \in \mathbb{A}\}$$

$\Lambda : \mathbb{R}^d \to \mathbb{R}$ is an aggregation rule that aggregate the risk factors into one risk factor $\Lambda(\mathbf{X})$.

- Example of Λ: $\Lambda(\mathbf{x}) = \sum_{i=1}^{d} x_i$, $\Lambda(\mathbf{x}) = \sum_{i=1}^{d} (x_i^+)^2$,...
- No individual risk contributions in terms of their systemic riskiness. $\Lambda : \mathbb{R}^d \to \mathbb{R}$ is an aggregation rule that aggregate the risk factors into one risk factor $\Lambda(\mathbf{X})$.
- Example: The VaR of a portfolio of assets could be computed as following:

$$\mathrm{VaR}(X) = \inf\{m \in \mathbb{R} \mid \mathbb{P}(\sum_{i=1}^{d} X_i - m < 0) \geq \lambda\}$$

First inject cash, then aggregate

- Inspiration from univariate shortfall risk (SR) measure.
- Y. Armenti, S. Crépey, S. Drapeau, and A. Papapantoleon, Multivariate Shortfall Risk Allocation and Systemic Risk. *SIAM J. Financial Math.*, Vol. 9, No. 1, pp. 90-126.

Definition

$$R(\mathbf{X}) := \inf\{\sum_{n=1}^{d} m_n \in \mathbb{R} \mid \mathbf{m} = (m_1, ..., m_d) \in \mathbb{R}^d;\ \mathbf{X} - \mathbf{m} \in \mathbb{A}\}$$

$$= \inf\{\sum_{n=1}^{d} m_n \in \mathbb{R} \mid \mathbb{E}[l(\mathbf{X} - \mathbf{m})] \le 0\},\ l \text{ a loss function.}$$

- $R(\mathbf{X})$ delivers at the same time a measure of total systemic risk and potential **ranking**$(m_1, \ldots, m_d)$ of portfolios in terms of global riskiness. 

Loss functions

Definition

A function $l : \mathbb{R}^d \mapsto (-\infty, \infty]$ is called a loss function if:

1. l is increasing, that is $l(x) \geq l(y)$ if $x \geq y$;
2. l is convex and lower-semicontinuous with $\inf l < 0$;
3. $l(x) \geq \sum x_k - c$ for some constant c.

Furthermore, a loss function l is said to be permutation invariant if $l(x) = l(\pi(x))$ for every permutation π of its components.

- The property (1) expresses the normative fact about the risk, that is, the more losses we have, the riskier is our system. As for (2), it expresses the desired property of diversification. Finally, (3) says that the loss function put more weight on high losses than a risk neutral evaluation.
- Permutation invariance : the considered risk components are often of the same type-banks, members of a clearing house, or trading desks within a trading floor. In that case, the loss function should not discriminate a particular component against another.

Example of loss functions: Let $h : \mathbb{R} \mapsto \mathbb{R}$ be one dimensional loss function satisfying condition (A1), (A2) and (A3). We could build a multivariate loss function using this one dimensional loss function in the following way:

1. $l(x) = h(\sum x_k)$;
2. $l(x) = \sum h(x_k)$;
3. $l(x) = \alpha h(\sum x_k) + (1-\alpha)\sum h(x_k)$ for $0 \leq \alpha \leq 1$.

- More specifically, in (1), we are agregating losses before evaluating the risk, whereas in (2), we evaluate individual risks before aggregating. The loss function in (3) is a convex combination of those in (1) and (2).
- One of the main examples we will be studying in this paper are the two following ones:
  - $l(x) = \frac{1}{1+\alpha}(\sum e^{\beta x_i} + \alpha e^{\beta \sum x_i})$, $\alpha > 0$.
  - $l(x) = \sum_i x_i + \frac{1}{2}\sum_i (x_i^+)^2 + \alpha \sum_{i<j} x_i^+ x_j^+$ where α is the systemic weight and β is a risk aversion coefficient.
  - Our notion of a loss function coincides with the one of aggregation function in the sense that it aggregates several loss profiles into a univariate random variable for which it can be decided whether or not it is acceptable.

- We introduce the Multivariate Orlicz Heart:

$$M^\theta := \{X \in L^0 : \ \mathbb{E}[\theta(\lambda X)] < \infty, \ \forall \lambda > 0\}, \ \theta(x) := l(|x|)$$

.

- We define $A(\mathbf{X}) := \{\mathbf{m} \in \mathbb{R}^d \mid \mathbb{E}[l(\mathbf{X} - \mathbf{m})] \le 0\}$ the corresponding set of acceptable monetary allocations.
- $X \mapsto A(X)$ defines a monetary set valued risk measure, that is, a set valued map A from M^θ into the set of monotone, closed, and convex subsets of $\mathbb{R}^d$ ($A(X)$ is different from the empty set and $\mathbb{R}^d$) .

- The following theorem from Armenti et al shows that the multivariate shortfall risk measure has the desired properties and admits a dual representation as in the case of univariate shortfall risk measure. We introduce Q^{θ^*} the set of measure densities in L^{θ^*}, the dual space of M^{θ}:

$$Q^{\theta^*} := \left\{ \frac{dQ}{d\mathbb{P}} := (Z_1, ..., Z_d),\ Z \in L^{\theta^*},\ Z_k \geq 0 \text{ and } \mathbb{E}[Z_k] = 1 \text{ for every k} \right\}$$

Theorem

The function

$$R(X) := \inf \left\{ \sum m_k :\ m \in \mathcal{A}(X) \right\}$$

is real-valued, convex, monotone and translation invariant. Moreover, it admits the dual representation:

$$R(X) = \max_{Q \in Q^{\theta^*}} \{ \mathbb{E}_Q[X] - \alpha(Q) \},\quad X \in M^{\theta}$$

where the penalty function is given by

$$\alpha(Q) = \inf_{\lambda > 0} \mathbb{E}\left[\lambda l^* \left(\frac{dQ}{\lambda d\mathbb{P}} \right) \right],\quad Q \in Q^{\theta^*}.$$

- Now, we address the question of existence and uniqueness of a risk allocation which are not straightforward in the multivariate case. Armenti et al showed that if the loss function is permutation invariant, then a risk allocations exist and they are characterized by Kuhn-Tucker conditions.

Definition

A **risk allocation** is a vector $\mathbf{m} = (\mathbf{m}_1, ..., \mathbf{m}_d)$ such that $\mathbb{E}[l(\mathbf{X} - \mathbf{m})] \leq 0$ and $R(\mathbf{X}) = \sum_{k=1}^{d} \mathbf{m}_k$.

Existence and Uniqueness ? 🤔

- When d $=$ 1, the above definition corresponds exactly to the univariate shortfall risk measure in Föllmer et Schied (2002).

Main Theorem

Assume l is a loss function verifying the following conditions:

1. $\forall \mathrm{m}_0$, $m \mapsto l(X-m)$ is differentiable at m_0;
2. l is permutation invariant.

Theorem

Under the conditions above, for every $\mathbf{X} \in M^\theta$, a **risk allocation** exists and it is characterized by the following first order conditions:

$$1 = \lambda^* \mathbb{E}[\nabla l(X - m^*)], \quad \mathbb{E}[l(X-m^*)] = 0,$$

where $\lambda^* \geq 0$ is a Lagrange coefficient. If moreover l is strictly convex along zero sum allocations for every x such that $l(x) \geq 0$, the risk allocation is unique.

Some comments

- Let $f_0(m) = \sum_{i=1}^{d} m_i$ and $f_1(m) := \mathbb{E}[l(X-m)]$, for $m \in \mathbb{R}^d$ and $X \in M^\theta$.
- The above assumption together with the convexity of the function $m \mapsto l(X-m)$, we have that, f_1 is differentiable at every $m \in \mathbb{R}^d$ and that,

$$\nabla f_1(m) = -\mathbb{E}[\nabla l(X-m)], \; m \in \mathbb{R}^d$$

- Therefore, the first order conditions given in the above theorem are equivalent to :

$$\begin{cases} \nabla f_0(m^*) + \lambda^* \nabla f_1(m^*) = 0 \\ \lambda^* f_1(m^*) = 0 \end{cases}$$

- We also know, thanks to Theorem 28.3 in Rockafellar book, that the above conditions are equivalent to saying that (m^*, λ^*) is a saddle point of the Lagrangian associated to the problem i.e.,

$$L(m,\lambda) := f_0(m) + \lambda f_1(m) = \sum_{i=1}^{d} m_i + \lambda \mathbb{E}[l(X-m)].$$

Some comments

$$R(\mathbf{X}) = \inf\left\{\sum_{k=1}^{d} m_k : \mathbb{E}[l(\mathbf{X} - \mathbf{m})] \leq 0\right\}$$

- $z^* := (m^*, \lambda^*)$ is zero of the function $h(z) := \mathbb{E}[H(X, z)]$ where:

$$H(X, z) := \begin{pmatrix} \lambda \nabla_m l(X - m) - 1 \\ l(X - m) \end{pmatrix}, \quad X \in M^\theta.$$

- In general, no closed formula for h.
- In order to find the unique risk allocation m^*, we can look for the zeros of the function h.
- We suggest here to use stochastic algorithms as they present the advantage of being incremental, less sensitive to dimension, and offer a flexible framework that can be conveniently combined with features such as importance sampling and model uncertainty.

Motivation

$$R(X) = \inf\left\{\sum_{k=1}^{d} m_k : \ \mathbb{E}[l(X-m)] \le 0\right\}$$

- The Armenti et al approach consists : estimating $\mathbb{E}[l(X-m)]$ with Monte Carlo for each m and make use of Python algorithm to find the minimum.
- Drawbacks: No result of convergence.
- The minimizing algorithm can be sensistive to the starting point of the algorithm, and there are no theoritical results on the convergence of this procedure.
- Therefore we cannot have any control over the error of the approximation
- Depends heavily on the starting point of the search algorithm.
- No control of the error of the estimation.

Outline

Some background

- Suppose we want to find the zeros of a certain function $z \mapsto h(z)$.
- If h is known, under some conditions: $Z_{n+1} = Z_n \pm \gamma_n h(Z_n)$, Z_0 given.
- If we do not have access to $h(Z_n)$ but only to some noisy estimated Y_n that are close to h on average: $Z_{n+1} = Z_n \pm \gamma_n Y_n$.
- This is typically the case when $h(z) = \mathbb{E}[H(X,z)]$, X some RV.
- An estimate of h at each step n would be $Y_n = H(X_{n+1}, Z_n)$:

$$Z_{n+1} = Z_n \pm \gamma_n H(X_{n+1}, Z_n).$$

- (X_n) is an i.i.d sequence $\sim X$.
- (γ_n) the time step sequence is taken $\to 0$ as $n \to \infty$.

Some developments

- Let $\mathcal{F}_n = \sigma(Z_0, X_1 ..., X_n)$.
- $Z_{n+1} = Z_n + \gamma_n Y_n = Z_n + \gamma_n H(X_{n+1}, Z_n) = Z_n + \gamma_n h(Z_n) + \underbrace{\gamma_n \delta M_n}_{"Noise"}$.
- $\delta M_n := H(X_{n+1}, Z_n) - h(Z_n)$. Observe that, $\mathbb{E}[\delta M_n | \mathcal{F}_n] = 0$.
- (δM_n) is therefore a martingale difference sequence.
- This is the classical version of Stochastic Algorithms introduced by Robbins-Monro (RM).
- Convergence of (Z_n) ?

Link with ODE

- First note that, since $\gamma_n \to 0$, Z_n for large n will change slowly.
- For small Δ, define m_n^Δ s.t. : $\displaystyle\sum_{i=n}^{n+m_n^\Delta-1} \gamma_i \approx \Delta$.
- $Z_{n+m_n^\Delta} - Z_n \approx \underbrace{\Delta h(Z_n)}_{\text{mean term}} + \underbrace{\displaystyle\sum_{i=n}^{n+m_n^\Delta-1} \gamma_i \delta M_i}_{\text{error}}$.
- Since δM_n is a martingale difference sequence $\mathbb{E}[\delta M_i \delta M_j] = 0$ and the variance of the error is given:

$$\mathbb{E}\left[\sum_{i=n}^{n+m_n^\Delta-1} \gamma_i \delta M_i\right]^2 = \sum_{i=n}^{n+m_n^\Delta-1} \mathbb{E}[\gamma_i^2 \delta M_i^2] = \sum_{i=n}^{n+m_n^\Delta-1} O(\gamma_i^2) = O(\Delta)\gamma_n.$$

- Thus, the main change in Z_n is due to the "mean term".

Link with ODE

- $Z_{n+m_n^\Delta} - Z_n \approx \Delta h(Z_n) + \text{"error"}$, with very small error.
- Formally, for large n, the behaviour of the algorithm can be approximated by the asymptotic behaviour of the ODE:

$$\dot{z} = h(z).$$

- Z_n will converge to limit points of the ODE.
- This requires some control over the growth of h to avoid explosion (sublinearity of h).
- One way to bypass this, is to use projection over some compact K.
- The new algorithm is $Z_{n+1} = \Pi_K(Z_n + \gamma_n Y_n)$, with $Y_n = H(X_{n+1}, Z_n)$.

Link with projected ODE

$$Z_{n+1} = \Pi_K(Z_n + \gamma_n Y_n) = Z_n + \gamma_n Y_n + \gamma_n C_n.$$

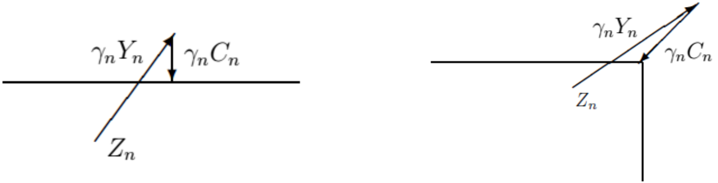


Figure: Projection when constraints are violated

- The new ODE is: $\dot{z} = h(z) + C(z)$, where $C(z) \in \mathcal{C}(z(t))$, $C(z)$ is the minimum force needed to bring back z to K.

Link with ODE

- We assume that K is hyperrectangle such that z^* is in the interior of K: $K = \{m \in \mathbb{R}^d, a_i \le m_i \le b_i\} \times [0, A]$.
- $(X_n)_{n\ge1}$ is an i.i.d sequence of random variables with the same distribution as X, independent of Z_0 and $(\gamma_n)_{n\ge1}$ is a deterministic step sequence decreasing to zero and satisfying:

$$\sum_{n\ge1} \gamma_n = +\infty \text{ and } \sum_{n\ge1} \gamma_n^2 < +\infty.$$

- Now, since z^* is interior to K and $h(z^*) = 0$, z^* is an equilibrium point for the projected ODE. In other words, this means that once $z(t)$ is equal to z^* it remains equal to z^* for all future times.
- In order to study the asymptotic stability of the equilibrium z^*, one needs to find some convenient Lyapunov function V.
- A natural and classical choice for this type of problems is $V(z) = ||z - z^*||^2$.
- It is obvious that V is positive definite. The following proposition shows that its derivative along any state trajectory is negative semi-definite on K.

Proposition

The function $V(z) = ||z - z^*||^2$ is such that $z \to \dot{V}(z) = \langle \nabla V(z), h(z) + C \rangle$ is negative semi-definite on K with the respect to the ODE.

Link with ODE

Remark

We cannot conclude that $\dot{V}$ is negative definite on K, because $z \neq z^*$ does not imply that $m \neq m^*$. Besides, if $z = (m^*, \lambda)$ such that $\lambda \neq \lambda^*$, we have $\dot{V}(z) = 0$ and $z \neq z^*$.

Proposition

The equilibrium point z^* of the ODE is asymptotically stable.

Outline

Framework

$$R(X) = \inf\left\{\sum_{k=1}^{d} m_k : \ \mathbb{E}[l(\mathbf{X} - \mathbf{m})] \leq 0\right\}$$

- The unique risk allocation m^* is characterized through the following:

$$1 = \lambda^* \mathbb{E}[\nabla l(X - m^*)], \quad \mathbb{E}[l(X - m^*)] = 0,$$

- $z^* = (m^*, \lambda^*)$ is the zero of:

$$h(z) := \mathbb{E}\begin{pmatrix} \lambda \nabla l(X - m) - 1 \\ l(X - m) \end{pmatrix} = \mathbb{E}(H(X, z))$$

- The RM Stochastic algorithms: $Z_{n+1} = \Pi_K(Z_n + \gamma_n H(X_{n+1}, Z_n))$.

More precisely

- Let γ_n be such that $\sum_{n\geq 1} \gamma_n = +\infty$ and $\sum_{n\geq 1} \gamma_n^2 < \infty$.
- Typically, $\gamma_n = \frac{c}{n^\gamma}$, $\gamma \in (1/2, 1]$.
- K is chosen such that $z^* \in K$.
- We introduce the following:

$$\begin{cases} \sigma^2(z) = \mathbb{E}[\|H(X,z) - h(z)\|^2]; \\ m^{2+p}(z) = \mathbb{E}[\|H(X,z) - h(z)\|^{2+p}]; \\ \Sigma(z) = \mathbb{E}[(H(X,z) - h(z))(H(X,z) - h(z))^\top]. \end{cases}$$

Convergence Results

$$(\mathbf{A_{a.s.}}) : \begin{cases} 1)\ h \text{ is continuous on } K; \\ 2)\ \sup_{z \in K} \sigma^2(z) < \infty. \end{cases}$$

The following result can be found in Theorem 5.2.2.1 in H.J. Kushner and G.G Yin, Stochastic Approximation and Recursive Algorithms and Applications.

Theorem

Under $(\mathbf{A_{a.s.}})$, Z_n converge to a limit point of the projected ODE: $\dot{z} = h(z) + C$

We proved that the only limiting point of the projected ODE is z^*. We conclude then using the theorem that $Z_n \to z^*$.

$$(\mathbf{A_{a.n.}}) : \begin{cases} 1) m \mapsto \mathbb{E}[\nabla l(X-m)] \text{ is a.s. } \mathcal{C}^1 \text{ Let } A := Dh(z^*); \\ 2) \text{ For some } p>0, \rho>0, \sup_{|z-z^*|\le\rho} m^{2+p}(z) < \infty, (Y_n \mathbf{1}_{|Z_n - z^*|\le\rho}) \text{ is u.i.;} \\ 3) \Sigma(\cdot) \text{ is continuous at } z^* \text{ and } \Sigma^* := \Sigma(z^*). \end{cases}$$

Theorem - Asymptotic Normality

Assume that $\gamma \in (\frac{1}{2}, 1)$ and that $(\mathbf{A_{a.s.}})$ and $(\mathbf{A_{a.n.}})$ hold. Then,

$$\sqrt{n^\gamma}(Z_n - z^*) \to \mathcal{N}\left(0, c^2 \int_0^\infty e^{cAt}\ \Sigma^*\ e^{cA^\top t} dt\right).$$

If furthermore, $cA + \frac{I}{2}$ is a stable matrix and $cI - P$ is a positive matrix where P is solution to the Lyapunov's equation $A^\top P + PA = -I$, then,

$$\sqrt{n}(Z_n - z^*) \to \mathcal{N}\left(0, c^2 \int_0^\infty e^{(cA+\frac{I}{2})t}\ \Sigma^*\ e^{(cA^\top+\frac{I}{2})t} dt\right)$$

Some comments

- A rate of convergence with $\sqrt{n}$ requires that $cA + \frac{I}{2}$ is a stable matrix $\Rightarrow c$ large enough.
- However, choosing c very large may lead to slower convergence: $Z_{n+1} = \Pi_K \left(Z_n + \frac{c}{n^\gamma} \times (h(Z_n) + "Noise") \right)$
- We could have chosen a step sequence $\gamma_n = \Gamma / n^\gamma$, where Γ is a preconditioning matrix. This will lead to the following CLT:

$$\sqrt{n}(Z_n - z^*) \to \mathcal{N}\left(0, \int_0^\infty e^{(\Gamma A + \frac{I}{2})t} \; \Gamma \Sigma^* \Gamma^\top \; e^{(A^\top \Gamma^\top + \frac{I}{2})t} dt \right).$$

- Choose Γ such that the trace of the asymptotic covariance matrix is minimised $\Rightarrow \Gamma = -A^{-1}$. But $A = Dh(z^*)$. ☹
- The corresponding optimal asymptotic covariance matrix is $V_{\text{opt}} := A^{-1} \Sigma^* (A^{-1})^\top$.

But there is way...

- Choosing the constant c is a burning issue.
- One way to bypass this is to use averaging introduced by Polyak and Ruppert (PR).
- For any arbitrary $t > 0$, we define, $\bar{Z}_n = \frac{1}{tn^\gamma} \sum_{i=n}^{n+tn^\gamma-1} Z_i$ (PR).

Theorem

Assume $\gamma \in (\frac{1}{2}, 1)$ and that $\mathbf{(A_{a.s.})}$ and $\mathbf{(A_{a.n.})}$ hold. If the matrix Σ^* is positive definite, then,

$$\sqrt{tn^\gamma}\left(\bar{Z}_n - z^*\right) \to \mathcal{N}\left(0, V_{\text{opt}} + O\left(\frac{1}{t}\right)\right)$$

where $V_{\text{opt}} := A^{-1}\Sigma^*(A^{-1})^\top$.

TCL and estimation of asymptotic covariance matrix

In practice, to derive confidence interval for the averaging procedure, one need to estimate the matrix $V_{\text{opt}} = A^{-1}\Sigma^*(A^{-1})^\top$.

Proposition

Assume **($\mathbf{A}_{\text{a.s.}}$)** and **($\mathbf{A}_{\text{a.n.}}$)** hold and that $z \to \mathbb{E}[\|H(X,z)\|^4]$ is bounded around z^*. Then,

$$\Sigma_n := \frac{1}{n}\sum_{i=1}^{n} H(X_i, Z_{i-1})^\top H(X_i, Z_{i-1}) \to \Sigma^* \; a.s. \tag{1}$$

Let $A_n^\epsilon(i,j)$ for $i,j \in \{1,\dots,d+1\}$ are defined as follows:

$$A_n^\epsilon(i,j) := \frac{1}{\epsilon n}\sum_{k=1}^{n} \left(H_i(X_k, Z_{k-1} + \epsilon e_j) - H_i(X_k, Z_{k-1})\right),$$

$$\lim_{\epsilon\to 0}\lim_{n\to\infty} A_n^\epsilon = A \; a.s. \tag{2}$$

Confidence intervals

- Instead of averaging on all observations, one could modify the estimators above and average only on recent ones.
- If we denote $V_n := A_{n,\epsilon}^{-1} \Sigma_n (A_{n,\epsilon}^{-1})^\top$, the confidence intervals of PR estimator with a confidence of $1 - \alpha$ in the following form:

$$\left[\bar{Z}_{j,n} - \sqrt{\frac{V_{jj,n}}{tn^\gamma}} q_\alpha, \bar{Z}_{j,n} + \sqrt{\frac{V_{jj,n}}{tn^\gamma}} q_\alpha \right], j \in \{1 \dots d\}, \ \gamma \in (0,1), \tag{3}$$

- q_α is the $1 - \dfrac{\alpha}{2}$ quantile of a standard normal random variable.

Outline

A toy example with $d = 2$

- $l(x_1, x_2) = \frac{1}{1+\alpha}\left[e^{\beta x_1} + e^{\beta x_2} + \alpha e^{\beta(x_1+x_2)}\right] - \frac{\alpha+2}{\alpha+1}, \alpha > 0, \beta > 0.$
- $X = (X_1, X_2) \sim \mathcal{N}(0, M)$ with $M = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2. \end{pmatrix}$
- Closed formulas for $m_i^* = \begin{cases} \frac{\beta\sigma_i^2}{2}, & \text{if } \alpha = 0, \\ \frac{\beta\sigma_i^2}{2} + \frac{1}{\beta} SRC(\rho, \sigma_1, \sigma_2, \alpha, \beta), & \text{if } \alpha > 0. \end{cases}$
- $SRC(\rho, \sigma_1, \sigma_2, \alpha, \beta) = \ln\left(\frac{\alpha e^{\rho\beta^2\sigma_1\sigma_2}}{-1 + \sqrt{1 + \alpha(\alpha+2)e^{\rho\beta^2\sigma_1\sigma_2}}}\right).$

A toy example

- we fix $\alpha = 1$, $\beta = 1$, $\sigma_1 = \sigma_2 = 1$ and $\rho \in \{-0.5, 0, 0.5\}$.

ρ	$m_1^* = m_2^*$
-0.5	0.3868
0	0.5
0.5	0.6364

Table: Exact optimal risk allocations.

- For RM/PR estimators, we take $n = 100000$, $K = [0,2]^3$ and $t = 10$.
- Z_0 was taken uniformly on K.

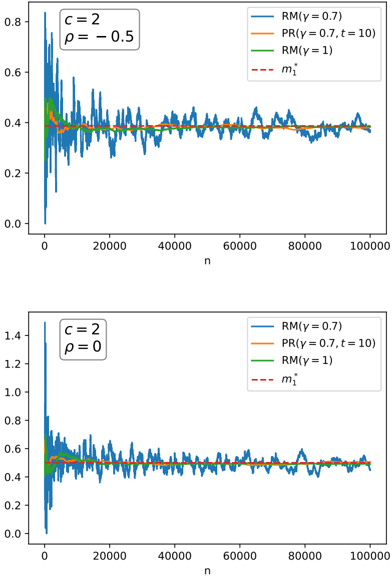
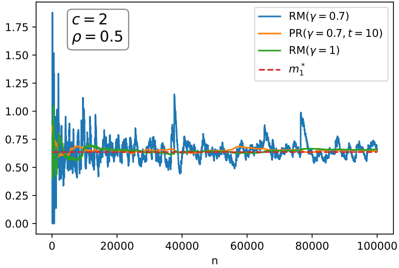


Figure: Consistency of RM/PR estimators.



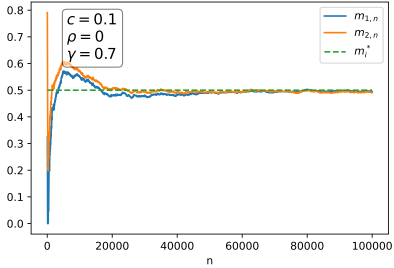
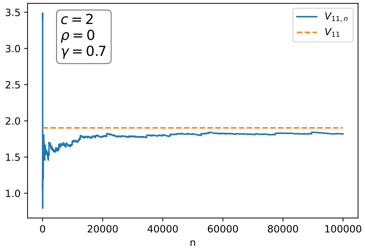


Figure: Effect of smaller value of c when $\gamma = 0.7$.

Estimator of V_{opt} and CI for PR estimator



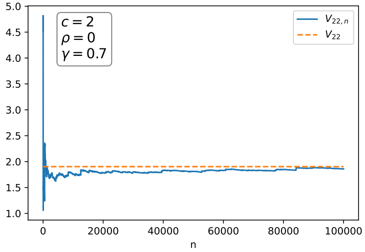


Figure: Convergence of the estimator V_n

ρ	CI for m_1^*	CI for m_2^*
-0.5	$[0.37724, 0.40478]$	$[0.36790, 0.39496]$
0	$[0.49622, 0.52594]$	$[0.49129, 0.52134]$
0.5	$[0.61944, 0.66297]$	$[0.62034, 0.66658]$

Table: Confidence intervals for PR estimators.

Second Example: $d = 10$

- $l(x_1, ..., x_d) = \sum_{i=1}^{d} x_i + \frac{1}{2}\sum_{i=1}^{d}(x_i^+)^2 + \alpha \sum_{j<k} x_i^+ x_j^+.$
- $X_i = \sum_{k=1}^{N_i^T} Z_k^i$, a Compound Poisson process. (Z_k^i) i.i.d $\sim \mathcal{N}(\mu_i, \sigma_i^2)$.

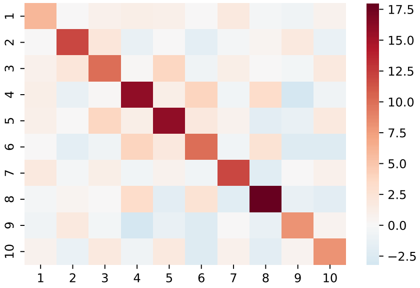


Figure: Correlation matrix of $\mathbf{X}$

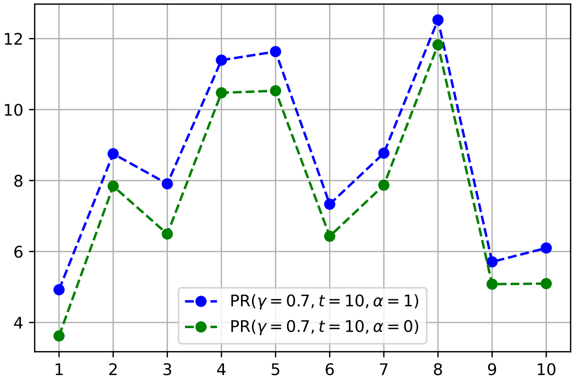


Figure: Optimal Allocations.

Work in Progress

An alternative stochastic algorithm: ADAM

- Diederik P. Kingma and Jimmy Ba, "Adam: A Method for Stochastic Optimization", 2017.
- Alexandre Défossez, Leon Bottou, Francis Bach and Nicolas Usunier, "A Simple Convergence Proof of Adam and Adagrad", 2022.
- We will use the dual formulation of the problem to directly numerically compute the solution.

The dual problem is represented by the Lagrangian $\mathcal{L}$ associated with the problem. For every $\lambda \geq 0$,

$$\mathcal{L}(m, \lambda; X) := \sum_k m_k + \lambda \mathbb{E}[l(X - m)]$$

Slater's condition holds in this case, i.e., there exists $m \in \mathbb{R}^d$ such that $\mathbb{E}[l(X - m)] < 0$. Hence,

$$\rho(X) = \sup_{\lambda \geq 0} \inf_{m \in \mathbb{R}^d} \mathcal{L}(m, \lambda; X)$$

SGDA Method Adapted to the Dual Problem

Stochastic Gradient Descent being an incremental method, we can combine the gradient descent to find the optimal allocation m and the gradient ascent to find the optimal Lagrange multiplier λ. Furthermore, we use an adaptative algorithm to regularize and accelerate the convergence. We also use mini-batches to reduce the variance. (we omit to write the mean of the batches in the sequences for simplicity)

Descent:

$$m_{n+1}^j = m_n^j - \frac{\gamma_n}{\sqrt{\hat{v}_n^j + \varepsilon}} \hat{\mu}_n^j, \quad m_0^j \in \mathbb{R}, \quad j = 1, \cdots, d,$$

where

$$\mu_{n+1}^j = \beta_1 \mu_n^j + (1 - \beta_1) \partial_{m_j} \mathcal{L}(m_n, \lambda_n; X_{n+1}), \quad j = 1, \cdots, d, \text{ (biased first moment)}$$

$$v_{n+1}^j = \beta_2 v_n^j + (1 - \beta_2) (\partial_{m_j} \mathcal{L}(m_n, \lambda_n; X_{n+1}))^2, \quad j = 1, \cdots, d, \text{ (biased second moment)}$$

$$\hat{\mu}_n^j = \frac{\mu_n^j}{1 - \beta_1^n}, \text{ (bias correction)}$$

$$\hat{v}_n^j = \frac{v_n^j}{1 - \beta_2^n}, \text{ (bias correction)}$$

Ascent:

$$\lambda_{n+1} = \lambda_n + \frac{\gamma_n}{\sqrt{\hat{\nu}_n} + \varepsilon}\hat{u}_n \quad \lambda_0 \in \mathbb{R}^+,$$

where

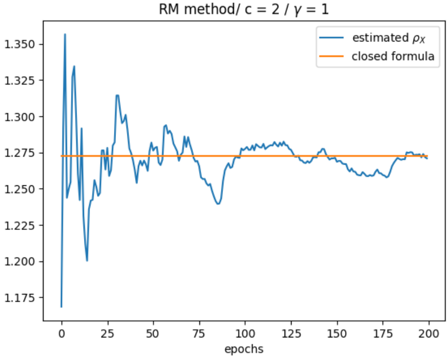
$$\begin{aligned} u_{n+1} &= \beta_1 u_n + (1-\beta_1)\partial_\lambda \mathcal{L}(m_n, \lambda_n; X_{n+1}), \\ \nu_{n+1} &= \beta_2 \nu_n + (1-\beta_2)(\partial_\lambda \mathcal{L}(m_n, \lambda_n; X_{n+1}))^2, \\ \hat{u}_n &= \frac{u_n}{1-\beta_1^n}, \\ \hat{\nu}_n &= \frac{\nu_n}{1-\beta_2^n}, \end{aligned}$$

The values of the hyperparameters can be found in the vast empirical literature around this topic: $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-7}$.

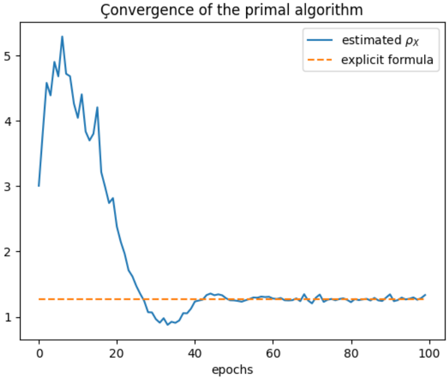
Advatanges:

- The additional moments use the historic changes to penalize (resp. increase) the learning rate when the saddle point is close (resp. far). Hence, more speed and precision.
- In practice, we don't need to project the sequences (m_n) and (λ_n) over a calibrated compact set since the moments will quickly guide the sequence towards a large compact containing the optimum, especially since the moments of the gradients will penalize them from exiting it.
- The use of the available deep learning machinery in **Tensorflow** or **Pytorch**, notably automatic differenciation. This is very important when calibrating the loss function.

Comparison results with ADAM in the toy example



RM algorithm
We take $n = 200000$, $c = 2$ and $\gamma = 1$.



Adam algorithm
We take $n = 100000$, a batch size $b = 64$ and a learning rate $\alpha = 3 \times 10^{-4}$.

Outline

Multivariate SRM with random allocations

- Biagini, F., Fouque, J. P., Frittelli, M., and Meyer-Brandis, T. (2020). On fairness of systemic risk measures. Finance and Stochastics, 24(2), 513-564.
- Alessandro Doldi and Marco Frittelli, 2021. "Conditional Systemic Risk Measures,"

Definition

The extended multivariate shortfall risk of a loss vector $X \in M^{\theta}$ is defined as:

$$\rho(X) := \inf_{Y \in \mathcal{C} \subset \mathcal{C}_{\mathbb{R}}} \left\{ \sum_{k=1}^{d} Y_k : \mathbb{E}[l(X - Y)] \leq 0 \right\} \tag{4}$$

where Y is a random vector in a subset of the class of feasible random allocations $\mathcal{C} \subset \mathcal{C}_{\mathbb{R}} := \left\{ Y \in M^{\theta};\ \text{there exists } m \in \mathbb{R} \text{ s.t. } \sum_{k=1}^{d} Y_k = m,\ \mathbb{P} - a.s \right\}$.

Characterization of the solution

Theorem (Biagini et al. 2020)

[Well-posedness of the problem] The extended shortfall risk measure $\rho(X)$ *is real-valued, convex, monotone and translation invariant. In particular, it is continuous and subdifferentiable. Moreover, it admits the dual representation:*

$$\rho(X) = \max_{\mathbb{Q} \in Q_{\theta^*} \cap dom(\alpha)} \{\mathbb{E}_{\mathbb{Q}}[X] - \alpha(\mathbb{Q})\}, \quad X \in M_\theta \tag{5}$$

where the penalty function is given by the following explicit form
$\alpha(\mathbb{Q}) = \inf_{\lambda > 0} \mathbb{E}\left[\lambda l^*(\frac{d\mathbb{Q}}{\lambda d\mathbb{P}})\right]$, $\mathbb{Q} \in Q_{\theta^*} \cap dom(\alpha)$.

Important remarks

- Interdependence structure through the law of the allocations.
- If Y is an optimal allocation, then $X + Y$ is $\sigma(X_1 + \cdots + X_d)$-measurable.
- The loss function impact on the interdependence structure of the model is less important in contrast with the deterministic multivariate Systemic Risk Measures.

Numerical method

Inspired by the SGDA method in the static case, we follow the same steps using a feed forward neural network rather than a trainable variable to approximate the optimal random allocations. The dual problem is represented by the Lagrangian $\mathcal{L}$ associated with the problem. For every $\lambda \geq 0$ and $\mu \in \mathbb{R}$,

$$\mathcal{L}(Y, \lambda, \mu; X) := \sum_{k=1}^{d} Y_k + \lambda \mathbb{E}\left[l(X - Y)\right] + \mu Var\left(\sum_{k=1}^{d} Y_k\right).$$

Hence,

$$\rho(X) = \sup_{\lambda, \mu \in \mathbb{R}^+ \times \mathbb{R}} \inf_{Y \in \mathcal{C}} \mathcal{L}(Y, \lambda, \mu; X).$$

A Deep Learning Algorithm for random allocations

- Another deep learning method : A. Doldi, Y. Feng, J.-P. Fouque, and M. Frittelli: Multivariate Systemic Risk Measures and Deep Learning Algorithms, to appear in Quantitative Finance, 2023

We introduce here a deep learning algorithm that uses advantage of the duality of the problem.

$$\rho(X) = \inf_{Y \in \mathcal{C}_{\mathbb{R}}} \left\{ \sum_{k=1}^{d} Y_k : \mathbb{E}[l(X - Y)] \leq 0 \text{ and } Var\left(\sum_{k=1}^{d} Y_k \right) = 0 \right\}, \tag{6}$$

According to the optimization problem (4), the random allocation Y can be written as a function of X in the following way $Y = \mathcal{Y}^{\xi^\star} = \Phi^{\xi^\star}(X)$.

We seek to solve the following optimization problem.

$$\varphi_X(\xi,\lambda,\mu) := \mathbb{E}\left[\sum_{k=1}^{d} \mathcal{Y}_k^{\xi}\right] + \lambda\mathbb{E}[l(X-\mathcal{Y}^{\xi})] + \mu Var\left(\sum_{k=1}^{d} \mathcal{Y}_k^{\xi}\right), \tag{7}$$

and the equivalent optimization problem

$$\sup_{(\lambda,\mu)} \inf_{\xi} \varphi_X(\xi,\lambda,\mu). \tag{8}$$

Numerical results: An explicit example

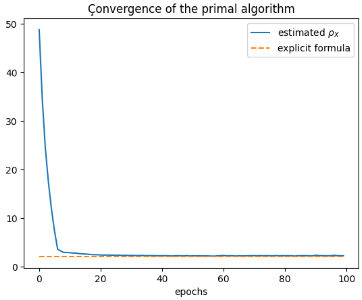
- In A. Doldi et al., this example is provided in terms of utility functions with

$$U(x) = -l(-x) \text{ and } B = 0.$$

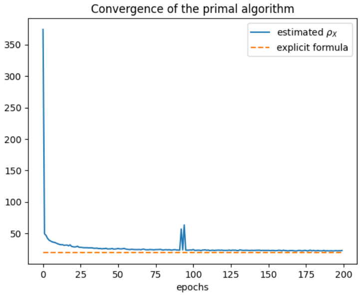
In this example, we consider a multi-dimensional vector X, representing risk factors of d positively correlated financial institutions. The loss function is defined as follows:

$$l(x) = \frac{1}{2}((\sum_{k=1}^{d} e^{\alpha_k x_k})^2 - d^2),$$

where $\{\alpha_k, k = 1, \ldots, d\}$ are the loss parameters for all financial institutions. We use again the ADAM optimizer in this setting.



(a) Convergence of the DL algorithm with d = 10



(b) Convergence of the DL algorithm with d = 100

Figure: Convergence of the deep learning algorithm

Hyper-parameters:

- (d = 10) $nbNeurons = 64$, $nbLayers = 2$, $batchSize = 10^4$, $lRate = 10^{-5}$ and $n = 10^5$
- (d = 100) $nbNeurons = 128$, $nbLayers = 3$, $batchSize = 10^4$, $lRate = 3.10^{-5}$ and $n = 2.10^5$

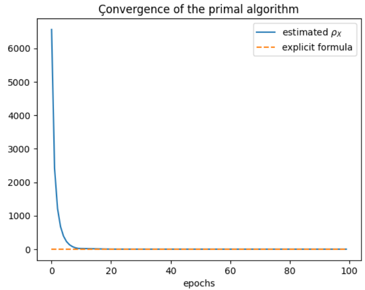


Figure: Scenario dependent allocations where X follows a multivariate compound Poisson distribution.

$$X_i = \sum_{k=1}^{N_i} G_k^i, \quad i = 1, \cdots, d,$$

where $N_i \sim \mathcal{P}(\lambda_i)$ and (G_k^i) a sequence of i.i.d normal random variables with the following parameters:

- $\lambda = (0.1, 0.2, 1., 0.5, 0.12, 0.5, 0.3, 0.22, 0.1, 1.)$, $\mu_J = 0$ and $\sigma_J = 1$.

Conditional Systemic Risk Measures

Assume that $\mathcal{G} \subseteq \mathcal{F}$ is a sub-σ-algebra of $\mathcal{F}$. Conditional Systemic Risk Measure (CSRM) were introduced in A. Doldi, M. Frittelli (2021) Conditional Systemic Risk Measures as a map

$$\rho_{\mathcal{G}} : L^0(\Omega, \mathcal{F}, \mathbb{P})^d \longrightarrow L^0(\Omega, \mathcal{G}, \mathbb{P})$$

that verifies the following axioms:

1. *Monotonicity*, that is

$$X \leq Y \text{ componentwise } \quad \mathbb{P}-a.s. \;\Rightarrow\; \rho_{\mathcal{G}}(X) \leq \rho_{\mathcal{G}}(Y) \quad \mathbb{P}-a.s.$$

2. *Conditional Convexity*, that is

$$\rho_{\mathcal{G}}(\lambda X + (1-\lambda)Y) \leq \lambda \rho_{\mathcal{G}}(X) + (1-\lambda)\rho_{\mathcal{G}}(Y) \quad \mathbb{P}-a.s. \quad \lambda \in L^{\infty}(\mathcal{G};[0,1])$$

3. *Conditional $\mathcal{G}$-Additivity*

$$\rho_{\mathcal{G}}(X+Y) = \rho_{\mathcal{G}}(X) + \sum_{j=1}^{N} Y^j \quad \mathbb{P}-a.s. \quad \text{if } Y \in (L^{\infty}(\mathcal{G}))^d \cap L_{\mathcal{F}}$$

Definition of Maps and Sets

$$\mathcal{D}_{\mathcal{G}} := \left\{ Y \in (L^0(\mathcal{F}))^d \middle| \sum_{k=1}^{d} Y_k \in L^0(\Omega, \mathcal{G}, \mathbb{P}) \right\}.$$

The set of admissible allocations $\mathcal{B}_{\mathcal{G}}$ is defined as follows:

$$\mathcal{B}_{\mathcal{G}} \subseteq \mathcal{D}_{\mathcal{G}},$$

and

$$\mathcal{C}_{\mathcal{G}} := \mathcal{B}_{\mathcal{G}} \cap \mathcal{B}_{\mathcal{G}}^{(\mathbf{I}),\infty} \cap (L^1(\mathcal{F}))^d,$$

where

$$\mathcal{B}_{\mathcal{G}}^{(\mathbf{I}),\infty} := \left\{ Y \in (L^0(\mathcal{F}))^d \middle| \exists d = [d_1, \cdots, d_h] \in (L^\infty(\mathcal{G}))^h \middle| \sum_{i \in I_m} Y_i = d_m \text{ for } m \leq h \right\}.$$

Let us now define the following maps and sets that will be used in the main theorem in the static framework:

$$\rho_{\mathcal{G}}^{\infty}(X) := \operatorname*{ess\,inf}_{Y \in \mathcal{C}_{\mathcal{G}} \cap (L^{\infty}(\mathcal{F}))^d} \left\{ \sum_{k=1}^{d} Y_k : \mathbb{E}[l(X - Y)|\mathcal{G}] \le 0 \right\};$$

$$\rho_{\mathcal{G}}(X) := \operatorname*{ess\,inf}_{Y \in \mathcal{C}_{\mathcal{G}}} \left\{ \sum_{k=1}^{d} Y_k : \mathbb{E}[l(X - Y)|\mathcal{G}] \le 0 \right\};$$

$$\alpha^1(\mathbb{Q}) := \operatorname*{ess\,sup}_{X \in (L^{\infty}_{\mathcal{F}})^d, \mathbb{E}[l(X)|\mathcal{G}] \le 0} \sum_{j=1}^{d} \mathbb{E}_{\mathbb{Q}_j}[X_j|\mathcal{G}], \quad \mathbb{Q} \in \mathcal{Q}_{\mathcal{G}};$$

$$\mathcal{Q}_{\mathcal{G}} := \left\{ \mathbb{Q} = (\mathbb{Q}_1, \cdots, \mathbb{Q}_d) \ll \mathbb{P} \,\middle|\, \frac{d\mathbb{Q}}{d\mathbb{P}} := (Z_1, \cdots, Z_d) \in L^1(\mathcal{F}), \forall k, \quad \mathbb{E}[Z_k|\mathcal{G}] = 1 \right\};$$

$$\mathcal{Q}_{\mathcal{G}}^1 := \left\{ \mathbb{Q} \in \mathcal{Q}_{\mathcal{G}} | \alpha^1(\mathbb{Q}) \in L^1(\mathcal{G}) \text{ and } \sum_{k=1}^{d} \mathbb{E}_{\mathbb{Q}^k}[Y_k|\mathcal{G}] \le \sum_{k=1}^{d} Y_k, \ \forall Y \in \mathcal{C}_{\mathcal{G}} \cap (L^{\infty}(\mathcal{F}))^d \right\}.$$

Main theorem

Theorem (Fritelli et al. 2021)

Consider the maps $\rho_{\mathcal{G}}^{\infty}$ and $\rho_{\mathcal{G}}$ defined above and under some assumptions that we will not discuss here.

- $\rho_{\mathcal{G}}^{\infty}(X) \in L^{\infty}(\mathcal{G})$ *for all* $X \in (L^{\infty}(\mathcal{F}))^d$ *and* $\rho_{\mathcal{G}}^{\infty}$ *is a Conditional Systemic Risk Measure as* $\rho_{\mathcal{G}}^{\infty}$ *is monotone, conditionally convex, and conditionally monetary. It is also continuous from above and from below.*
- *For every* $X \in (L^{\infty}(\mathcal{F}))^d$, $\rho_{\mathcal{G}}^{\infty}(X) = \rho_{\mathcal{G}}(X)$ *and the essential infimum is attained.*
- *The CSRM* $\rho_{\mathcal{G}}^{\infty}$ *admits the following dual representation:*

$$\rho_{\mathcal{G}}^{\infty}(X) = \operatorname*{ess\,sup}_{\mathbb{Q}\in\mathcal{Q}_{\mathcal{G}}^{1}} \left(\sum_{j=1}^{d} \mathbb{E}_{\mathbb{Q}_j}\left[X_j|\mathcal{G}\right] - \alpha^{1}(\mathbb{Q}) \right) \quad \forall X \in (L^{\infty}(\mathcal{F}))^d \tag{9}$$

Exponential case

Let's take $l_k(x) = e^{\alpha_k x} - 1$, for $k = 1, \ldots, d$, with $\alpha_1, \ldots, \alpha_d > 0$. Finally our loss function, takes the following form:

$$l(x) := \sum_{k=1}^{d} l_k(x_k) = \sum_{k=1}^{d} e^{\alpha_k x_k} - d.$$

We consider only the case $\mathcal{B}_{\mathcal{G}} = \mathcal{D}_{\mathcal{G}}$, which corresponds to the case of full sharing among all agents in the system (i.e., the extreme case of one single group).

Theorem (Fritelli et al. 2021)

Consider the map $\rho_{\mathcal{G}}$ defined above and a general sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, $X \in (L^\infty(\mathcal{F}))^d$. Then,

$$\rho_{\mathcal{G}}(X) = \gamma \log\left(\mathbb{E}_{\mathbb{P}}\left[\exp\left(\frac{\sum_{k=1}^{d} X_k}{\gamma}\right)\Big|\mathcal{G}\right]\right) - \sum_{k=1}^{d} \frac{1}{\alpha_k} \log\left(\frac{1}{\alpha_k}\right),$$

where $\gamma = \sum_{k=1}^{d} \frac{1}{\alpha_k}$ *and* $\widehat{\mathbf{Y}} = [\widehat{Y}_1, \cdots, \widehat{Y}_d] \in (L^\infty(\mathcal{F}))^d$ *is an optimal allocation for* $\rho_{\mathcal{G}}(X)$ *and* $\widehat{\mathbb{Q}} = [\widehat{\mathbb{Q}}_1, \cdots, \widehat{\mathbb{Q}}_d]$ *is an optimum for the dual representation of* $\rho_{\mathcal{G}}(X)$*, where for* $k = 1, \cdots, d$

$$\widehat{Y}_k := X_k + \frac{1}{\gamma \alpha_k}\left(-\sum_{k=1}^{d} X_k + \rho_{\mathcal{G}}(X) + \sum_{k=1}^{d} \frac{1}{\alpha_k} \log\left(\frac{1}{\alpha_k}\right)\right) - \frac{1}{\alpha_k} \log\left(\frac{1}{\alpha_k}\right), \tag{10}$$

Dynamic MSRM

Fix a time horizon $T < \infty$ and let $(\Omega, \mathbb{F}, \mathbb{P})$ be a complete probability space.

- Let W be a d-dimensional Brownian motion and $\mathcal{J}(dt, de)$ an independent Poisson random measure with compensator $\nu(de)dt$ such that $\nu(de)$ is a σ-finite measure on $\mathbb{R}^d \setminus \{0\}$, equipped with its Borel field $\mathcal{B}(\mathbb{R}^d \setminus \{0\})$.
- Let $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ be the (completed) filtration associated with W and $\mathcal{J}$.
- In the dynamic setting, we condition with respect to the sub-σ-algebra $\mathcal{F}_t$ for all $t \in [0, T]$.

a QBSDEJ

- Nicole El Karoui, A.M., Armand Ngoupeyou, "Quadratic Exponential Semimartingales and Application to BSDEs with jumps", 2016.
- Sarah Kaakai, A.M. , Achraf Tamtalini, "Utility Maximization Problem with Uncertainty and a Jump Setting", 2022.

Theorem

If $X_T \in L^\infty(\mathcal{F}_T)$, and for all $t \in [0,T]$ the map $\rho_{\mathcal{F}_t}$ is a CSRM with the following loss function $l : x \to \sum_{k=1}^{d} e^{\alpha_k x_k} - d$, then there exists (Z,U) in $\mathcal{H}^2 \times \mathcal{H}^2_\nu$ such that :

$$\begin{cases} -d\rho_{\mathcal{F}_t}(X_T) = \left(\frac{1}{2\gamma}|Z_t|^2 + \int_{\mathbb{R}^d \setminus \{0\}} \left\{ \gamma \exp\left(\frac{U_t(e)}{\gamma} \right) - U_t(e) - \gamma \right\} \nu(de) \right) dt - Z_t dW_t - \int_{\mathbb{R}^d \setminus \{0\}} U_t(e) \tilde{\mathcal{J}}(dt, de), \\ \rho_{\mathcal{F}_T}(X_T) = \sum_{k=1}^{d} X_T^k - \frac{1}{\alpha_k} \log\left(\frac{1}{\alpha_k} \right). \end{cases}$$

where $\gamma = \sum_{k=1}^{d} \frac{1}{\alpha_k}$.

Elements of proof

- Show (using the dual approach)

$$\rho_{\mathcal{F}_t}(X_T) = \gamma \log\left(\mathbb{E}_{\mathbb{P}}\left[\exp\left(\frac{\sum_{k=1}^{d} X_T^k}{\gamma}\right)\Big|\mathcal{F}_t\right]\right) - \sum_{k=1}^{d} \frac{1}{\alpha_k} \log\left(\frac{1}{\alpha_k}\right)$$

- Use the multiplicative martingale representation theorem and the relationship between stochastic and ordinary exponentials.

Numerical method

- Alasseur, Bensaid, Dumitrescu, Warin, "Deep learning algorithms for FBSDEs with jumps", 2024.

Discrete time approximation :

$$\begin{cases} X^\pi_{i+1} = X^\pi_i + b(t_i, X^\pi_i)\Delta t_i + \sigma(t_i, X^\pi_i)\Delta W_i + \sum_{l=1}^{\Delta N_i} \beta(t_i, X_i, \Delta J^i_l), \\ Y^\pi_i \cong Y^\pi_{i+1} + (\frac{1}{\gamma}|Z^\pi_i|^2 \ + \Gamma^\pi_i - \gamma)\Delta t_i - Z^\pi_i \Delta W_i - \left(u(t_i, X^\pi_i + \sum_{l=1}^{\Delta N_i} \beta(t_i, X^\pi_i, \Delta J^i_l)) - u(t_i, X^\pi_i) \right) \\ + \mathbb{E}\left[u(t_i, X^\pi_i + \sum_{l=1}^{\Delta N_i} \beta(t_i, X^\pi_i, \Delta J^i_l)) - u(t_i, X^\pi_i) \Big| \mathcal{F}_{t_i} \right], \\ Z^\pi_i = \mathbb{E}\left[Y^\pi_{i+1} \frac{\Delta W_i}{\Delta t_i} | \mathcal{F}_{t_i} \right], \\ U^\pi_i(e) = u(t_i, X^\pi_i + \beta(t_i, X^\pi_i, e)) - u(t_i, X^\pi_i), \\ \Gamma^\pi_i = \int_{\mathbb{R}^d \setminus \{0\}} \left\{ \gamma \exp\left(\frac{U^\pi_i(e)}{\gamma} \right) - U^\pi_i(e) \right\} \nu(de), \\ X^\pi_0 = \xi, \quad Y^\pi_M = g(X^\pi_M), \\ i = 0, \cdots, M-1. \end{cases}$$

Neural Networks

Global solver. We use two feed forward neural networks: $\mathcal{Z}^\theta$ to approximate the control Z and $\mathcal{W}^\theta$ to approximate the jump part. The Deep merged BSDE method consists in training the neural networks by solving in a forward way the backward representation of the solution, i.e. instead of solving the BSDE starting from the terminal condition, one estimates Y_0 with a trainable parameter θ_0 and solves the forward optimization problem with the aim of minimizing $\mathbb{E}\left[|Y_T - g(X_T)|^2\right]$.

$$Y_{i+1}^{\pi,\theta} = Y_i^{\pi,\theta} - (\frac{1}{\gamma}|\mathcal{Z}^\theta(t_i, X_i^\pi)|^2 \ + \lambda\Phi^\theta(t_i, X_i^\pi) - \gamma)\Delta t_i + \mathcal{Z}^\theta(t_i, X_i^\pi)\Delta W_i$$

$$+\mathcal{W}^\theta(t_i, X_i^\pi, \sum_{l=1}^{\Delta N_i} \tilde{\beta}_i(\Delta J_l^i)) - \Theta^\theta(t_i, X_i^\pi)$$

where $\tilde{\beta}_i(\Delta J_l^i) = \beta(t_i, X_i^\pi, \Delta J_l^i)$.

Note that we have the following result to compute the conditional expectation by means of Monte Carlo on each trajectory of the batch:

$$\Theta^\theta(t_i, X_i^\pi) = \mathbb{E}\left[\mathcal{W}^\theta(t_i, X_i^\pi, \sum_{l=1}^{\Delta N_i} \tilde{\beta}_i(\Delta J_l^i)) \Big| \mathcal{F}_{t_i}\right],$$

and, $$\Theta^\theta(t, x) = \mathbb{E}\left[\mathcal{W}^\theta(t, x, \sum_{l=1}^{\Delta N_i} \beta(t, x, \Delta J_l^i))\right],$$

$$\Phi^\theta(t_i, X_i) = \frac{1}{\lambda}\int_{\mathbb{R}^d \setminus \{0\}} \left\{\gamma \exp\left(\frac{U_i^\pi(e)}{\gamma}\right) - U_i^\pi(e)\right\} \nu(de),$$

and, $$\Phi^\theta(t, x) = \mathbb{E}\left[\gamma \exp\left(\frac{\mathcal{W}^\theta(t, x, \beta(t, x, \Delta J))}{\gamma}\right) - \mathcal{W}^\theta(t, x, \beta(t, x, \Delta J))\right]$$

Numerical example

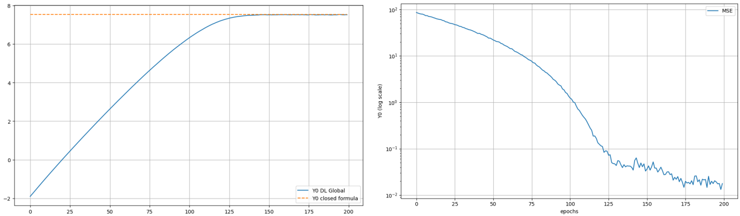
We consider the following dynamics for the portfolio:

$$\begin{cases} dX_t = X_{t^-}(rdt + \sigma dW_t + \int_{\mathbb{R}^\star} (e^e - 1)\tilde{\mathcal{J}}(dt, de)), \\ X_0 = x_0; \end{cases} \quad t \in [0, T], \tag{11}$$

where $\tilde{\mathcal{J}}(dt, de)$ is the compensated jump measure associated with the compound Poisson process $\sum_{i=1}^{N_t} Y_i$ with intensity measure $\nu(de)$, where ν is given by :

$$\nu(de) = \frac{\lambda}{\xi\sqrt{2\pi}} \exp\left(-\frac{(e-\alpha)^2}{2\xi^2}\right) de.$$

We set $d = 5$, $T = 1$, $M = 50$ steps, the interest rate $r = 0.1$, the diffusion volatility $\sigma = 0.3$, the jumps intensity $\lambda = 3$, the parameters of the jumps distribution $\alpha = 0$ and $\xi = 0.2$, the initial condition $X_0 = 1$, we take the same values for all the components of the vectors $r, \sigma, \alpha, \xi, x_0$ in $\mathbb{R}^d$, for simplicity.



(a) Convergence of Y0 (b) Convergence of the MSE error

Figure: Convergence of the deep BSDE solver

Hyper-parameters:

$nbNeurons = 25$, $nbLayers = 2$, $batchSize = 10$, $lRate = 10^{-3}$ and $n = 20000$.

A larger class

In this part, we consider a new class of multivariate SRM inspired from the remarks discussed above. As explained we noticed that if we find a class of loss functions written in a specific way, we can write the risk measure in the dual problem in the following way

$$\rho_\tau(X_T) = \mathbb{P} - ess \sup_{\eta \in \mathcal{D}^c(\tau)} \left\{ \mathbb{E}_{\mathbb{Q}^\eta} \left[g(X_T) - \int_\tau^T h_s(\eta_s) ds | \mathcal{F}_\tau \right] \right\} \tag{12}$$

where $h : [0.T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ is an $\mathbb{F}$-progressively measurable map with

$$\mathbb{E}_{\mathbb{Q}} \left[\int_0^T |h_s(\eta_s)| ds \right] < \infty, \quad \forall \mathbb{Q} \in \mathcal{Q}^1_{\mathcal{F}_t}, \quad \forall t \in [0, T],$$

and $g : \mathbb{R}^d \to \mathbb{R}$ a lipschitz continuous function. Thus, to go back to the notations of the static setting, we chose the following penalty function :

$$\alpha^1_{\mathcal{F}_\tau}(\mathbb{Q}^\eta) = \mathbb{E}_{\mathbb{Q}_\eta} \left[\int_\tau^T h_s(\eta_s) ds | \mathcal{F}_\tau \right].$$

Furthermore, we introduce the Hamiltonian H defined on $[0, T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$ by

$$H_t(x) := \sup_{y \in \mathbb{R}^d} \{ x \cdot y - h_t(y) \}$$

This collection of dynamic risk measures with a continuous filtration is expected to be a solution of the following BSDE:

$$\begin{cases} -dY_t = H_t(Z_t)dt - Z_t dW_t, \\ Y_T = g(X_T). \end{cases} \tag{13}$$

Perspectives

- Theoretical convergence of the ADAM/Deep learning algorithms in the deterministic and random allocations setting.
- Dynamic representation for a general class of Multivariate Systemic Risk Measures.
- Modelling the dependence structure through mean-field interactions.


thank
you!