



HAL
open science

Simulation of (nested/extreme) risks in finance: regression Monte-Carlo, MCMC, stochastic algorithms

Emmanuel Gobet

► To cite this version:

Emmanuel Gobet. Simulation of (nested/extreme) risks in finance: regression Monte-Carlo, MCMC, stochastic algorithms. Doctoral. France. 2018. cel-01692008

HAL Id: cel-01692008

<https://cel.hal.science/cel-01692008v1>

Submitted on 24 Jan 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Lectures at the 17th Winter school on Mathematical Finance, Netherlands

Simulation of (nested/extreme) risks in finance: regression Monte-Carlo, MCMC, stochastic algorithms

emmanuel.gobet@polytechnique.edu

Centre de Mathématiques Appliquées,
Ecole Polytechnique



These slides correspond with lectures given by the author at the 17th Winter school on Mathematical Finance, Netherlands, in January 2018.

The author would like to thank Professors Peter Spreij and Michel Vellekoop for the invitation to give these lectures, and all the participants for their feedbacks.

AGENDA OF THE COURSE

1. Financial context: risk measure calculations and stress tests
2. Overview on standard (and sometimes inefficient) methods for simulating nested expectations OR extreme risks (not both at the same time)
3. Markov chain techniques on path space (MCMC, IPS) (explicit rare event)
4. Nested rare event statistics
5. Nested extreme risks using SA (implicit rare event)

And if enough time, I will talk also on

6. Uncertainty propagation in $\mathbb{V}aR$, $\mathbb{C}VaR$

General references on stochastic simulation:

- ✓ L. Devroye: “Nonuniform random variate generation” (1986)
(<http://luc.devroye.org/rnbookindex.html>)
- ✓ S. Asmussen and P.W. Glynn: "Stochastic simulation" (2007)
- ✓ G. Rubino and B. Tuffin: "Rare events simulation" (2009)
- ✓ P. Glasserman: “Monte Carlo methods in financial engineering” (2003)
- ✓ A. Schwartz and A. Weiss: "Large Deviations for Performance Analysis" (1995)
- ✓ J.A. Bucklew: "Large deviation techniques in decision, simulation, and estimation" (1990)


1 Financial context: risk measure calculations and stress tests

1.1 Few definitions and notations

Definition (VaR, CVaR).

- ✓ Loss V (scalar random variable), assumed to be integrable
- ✓ VaR (**Value-at-risk**) is the *left quantile at level α* of V , i.e.

$$\text{VaR}^\alpha(V) = \inf\{x \in \mathbb{R} : \mathbb{P}(V \leq x) \geq \alpha\}$$

- ▶ “Loss that, with a probability α , will not be exceeded.”
 - ▶ Typically $\alpha = 99\%$
 - ▶  Does not take into account the losses beyond that threshold.
 - ▶ Ref: [Artzner, Delbaen, Eber, Heath, Math Fin 2002; Follmer et Schied, *Stochastic finance*, 2002]
- ✓ CVaR (**Conditional Value-at-risk**) accounts for *losses exceeding VaR*.
Depends on the full distribution in the right tail.

Several similar definitions:

- ▶ Averaged VaR:

$$\mathbb{A}\text{VaR}^\alpha(\mathbf{V}) = \frac{1}{(1-\alpha)} \int_\alpha^1 \text{VaR}^\beta(\mathbf{V}) d\beta.$$

- ▶ Convex optimization point-of-view [**Rockafellar-Uryasev 00**]:

$$\mathbb{A}\text{VaR}^\alpha(V) = \inf_x \left(x + \frac{1}{1-\alpha} \mathbb{E} [(V - x)_+] \right)$$

and the inf is attained for $x = \text{VaR}^\alpha(V)$.

[\Rightarrow $\mathbb{A}\text{VaR}^\alpha$ is sub-additivity, homogeneous, convex...]

Proof. Set $Q(\beta) := \text{VaR}^\beta(V)$ and $F(x) := \mathbb{P}(V \leq x)$. Then

$$\begin{aligned} \mathbb{E}((V - s)_+) + (1 - \alpha)s &= \mathbb{E}(Q(U) - s)_+ + (1 - \alpha)s \\ &= \int_0^1 (Q(r) - s) \mathbf{1}_{Q(r) > s} dr + (1 - \alpha)s \\ &= \int_0^1 (Q(r) - s) \mathbf{1}_{r > F(s)} dr + (1 - \alpha)s \\ &= \int_\alpha^1 Q(r) dr + \int_{F(s)}^\alpha Q(r) dr - s \int_{F(s)}^1 dr + s \int_\alpha^1 dr \\ &= \int_\alpha^1 Q(r) dr + \underbrace{\int_{F(s)}^\alpha (Q(r) - s) dr}_I. \end{aligned}$$

- * **Case $\alpha > F(s)$.** Then $r > F(s) \Leftrightarrow Q(r) > s \rightsquigarrow I > 0$.
- * **Case $\alpha \leq F(s)$.** Then $r \leq F(s) \Leftrightarrow Q(r) \leq s \rightsquigarrow I \geq 0$.
- * **Case $s = Q(\alpha)$.** Then $F(s) = F(Q(\alpha)) \geq \alpha$ and $Q(r) = s$ for $r \in [\alpha, F(s)]$
 $\rightsquigarrow I = 0$.

□

- **Lemma (The inf may be attained for several x).** Assume a continuous c.d.f. for V : the arginf solve

$$x : \mathbb{P}(V \leq x) = \alpha,$$

and any such solution leads to the same value for

$$x + \frac{1}{1 - \alpha} \mathbb{E} [(V - x)_+].$$

- ▮► Selecting any α -quantile provides the same $\mathbb{A}\text{VaR}$ (useful remark for later use of Stochastic Approximation algorithm).

Proof. a) Already proved. Can be re-obtained by writing first-order optimality condition.

b) Take x and x' , with $x < x'$ s.t. $\mathbb{P}[V \leq x] = \mathbb{P}[V \leq x'] = \alpha$. Then,

$$\begin{aligned}x + \frac{1}{1-\alpha} \mathbb{E}[(V-x)_+] &= x + \frac{1}{1-\alpha} (\mathbb{E}[V \mathbf{1}_{V \geq x}] - x \mathbb{P}[V \geq x]) \\ &= \frac{1}{1-\alpha} \mathbb{E}[V \mathbf{1}_{V \geq x}].\end{aligned}$$

Since by assumption $\mathbb{P}[x < V \leq x'] = 0$, we have

$$x' + \frac{1}{1-\alpha} \mathbb{E}[(V-x')_+] = \frac{1}{1-\alpha} \mathbb{E}[V \mathbf{1}_{V \geq x'}].$$

□

► Expected Shortfall:

$$\mathbb{E}\mathbb{S}^\alpha(\mathbf{V}) = \mathbb{E}[\mathbf{V} \mid \mathbf{V} \geq \mathbb{V}\text{aR}^\alpha(\mathbf{V})].$$

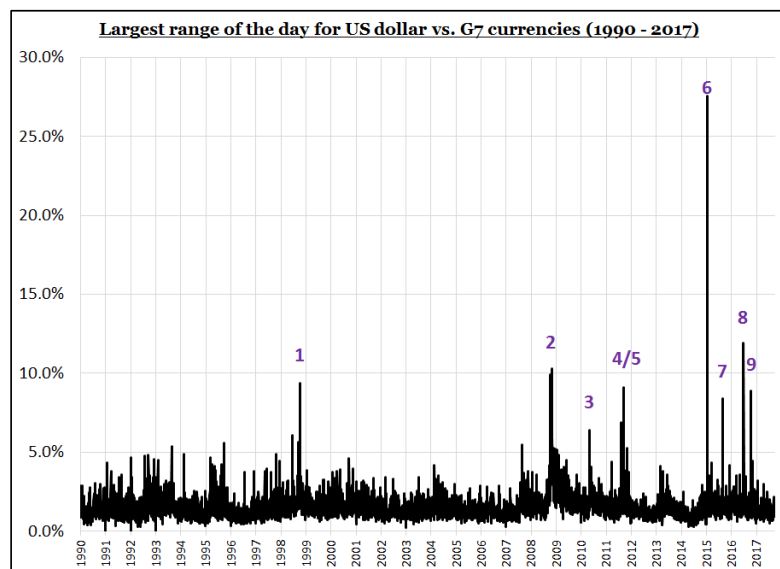
► Worst case expectation:

$$\mathbb{W}\text{C}\mathbb{E}^\alpha(\mathbf{V}) := \sup_{\mathbf{A}: \mathbb{P}(\mathbf{A}) \geq 1-\alpha} \mathbb{E}[\mathbf{V} \mid \mathbf{A}].$$

- ✓ In full generality on V , $\mathbb{A}\mathbb{V}\text{aR}^\alpha(\mathbf{V}) \geq \mathbb{W}\text{C}\mathbb{E}^\alpha(\mathbf{V}) \geq \mathbb{E}\mathbb{S}^\alpha(\mathbf{V})$.
- ✓ When the c.d.f. of V is continuous at $\mathbb{V}\text{aR}^\alpha(V)$, then these three quantities coincide.
- ✓ In the following, we simply write $\mathbb{C}\mathbb{V}\text{aR}^\alpha(V)$ and assume implicitly no atoms in the tails (continuous c.d.f.).
- ✓ We take $\alpha \in (0, 1)$ to avoid boundary problems.

1.2 Risk calculations in finance

Risks/volatilities get higher and higher...



USD versus G7 moves larger than 6%.

Source: HSBC FX, 2017.

1. Asia Crisis, LTCM unwind, October 8, 1998
2. Global Fin. Crisis, October 2008
3. US stock market flash crash, May 6, 2010
4. US downgraded by Standard & Poors, August 9, 2011
5. SNB puts in the floor, Sep 6, 2011
6. SNB removes the floor, January 15, 2015
7. China reval, August 24, 2015
8. UK votes to leave EU, June 24, 2016
9. Flash crash, October 7, 2016

Usually VaR – CVaR depend on time horizon, risk level, and losses under consideration ...

1.2.1 Usual VaR – CVaR

The main financial risk measures in practice are VaR and CVaR (as an expected shortfall) of level $\alpha \in (0, 1)$

Insurance: Solvency capital requirement (SCR) determined as the *99.5%-value at risk* for Solvency II, and the *99%-expected shortfall* for the Swiss Solvency Test, of the *one-year loss and profit* of the company

Banking: Basel II Pillar II defines economic capital as the *99% value-at-risk* of the *one-year depletion of core equity tier I capital* (CET1), the regulatory metric that represents the wealth of the shareholders within the bank. Moreover, the FRTB (2019?) requires a shift from 99% value-at-risk to 97.5% expected shortfall as the reference risk measure in capital calculations.

Market risk: risk computations over 10 trading days.

1.2.2 Initial-Margin Variation

Ref: *Basel Committee on Banking Supervision. Margin requirements for non-centrally-cleared derivatives*, 2015. <http://www.bis.org/bcbs/publ/d317.pdf>

Usual *linear risk-neutral pricing of financial contracts*: change of paradigm in last few years, influenced by the regulators.

- ✓ Nowadays, banks and financial institutions have to post collateral to a central counterparty (CCP = clearing house) in order to secure their positions.
- ✓ Variation margin= collateral to cover a new contract at inception and daily changes
- ✓ Initial margin [**Basel 2015, p.11 3(d)**]: IM protects the transacting parties from the potential future exposure that could arise from future changes in the mark-to-market value of the contract during the time it takes to close out and replace the position in the event that one or more counterparties default. The amount of initial margin reflects the size of the potential future exposure. It depends on a variety of factors, [...] and can change over time [...].

Example (Toy model).

- ✓ IM deposits = proportional to $\mathbb{C}\text{VaR}$ of the contract price over a future period of length $\Delta = 1$ week or 10 days.
- ✓ constant interest rate (money account)
- ✓ one tradable asset, denoted S :

$$dS_t = \mu S_t dt + \sigma S_t dW_t,$$

- ✓ π_t amount invested in S
- ✓ Objective: hedging of an option with maturity $T > 0$ and payoff $\Phi(S_T)$.
- ✓ Usual self-financing replicating portfolio: its value evolves as

$$\begin{cases} dV_t &= r(V_t - \pi_t) dt + \pi_t \frac{dS_t}{S_t}, \quad t \in [0, T], \\ V_T &= \Phi(S_T). \end{cases}$$

⇒ Linear BSDE (Backward Stochastic Differential Equation) [EPQ97] and linear pricing rule.

✓ Accounting for IM requirement:

$$dV_t = \left(r(V_t - \pi_t) - R \times \text{CVaR}_{\mathcal{F}_t}^\alpha(V_t - V_{t+\Delta}) \right) dt + \pi_t \frac{dS_t}{S_t},$$

where R is the net funding rate for the deposit (funding rate minus the interest rate paid by the CCP for the deposit, e.g. $R \approx 3\%$).

⇒ McKean Anticipative BSDE [CESS17, ADG⁺18]

$$V_t = \Phi(S_T) + \int_t^T \left(-r(V_s - \pi_s) - \mu\pi_s + R \text{CVaR}_{\mathcal{F}_s}^\alpha(V_s - V_{(s+\Delta) \wedge T}) \right) ds - \int_t^T \pi_s \sigma dW_s, \quad t \in [0, T].$$

⇒ Non linear nested expectations.

1.2.3 Credit risk

Aim: measuring the statistics of losses when some counter-parties of the bank make defaults.

Example (Credit Value Adjustment (CVA)). Simplified computations with independence between default and losses:

$$\text{CVA}_0 = \text{LGD} \sum_i \text{PD}(t_{i-1}, t_i) \mathbb{E} [(\mathbb{E} [\Phi_{t_i, T} | \mathcal{F}_{t_i}])_+]$$

where $\Phi_{t_i, T}$ is the value cashflow occurring between t_i and T .

▣ Simple nested computations.

1.2.4 XVA

Ref: [Armenti, Crépey 2017; Abbas-Turki, Crépey, Diallo 2017]

Cost of funding initial margin:

$$\text{MVA}_0 = \lambda \mathbb{E} \left[\int_0^T \text{IM}_t^\alpha dt \right].$$

Cost of capital (for remunerating shareholders for their capital at risk):

$$\text{KVA}_0 = \mathbb{E} \left[\int_0^T e^{-ht} \text{CVaR}_t^\alpha dt \right].$$

⇒ Double nested computations.

1.3 Stress-test

Ref: European Banking Authority (EBA), *Draft Guidelines on the revised common procedures and methodologies for the supervisory review and evaluation process (SREP) and supervisory stress testing*, EBA/CP/2017/18, October 2017.

New recommendations about how to handle stress tests for the banks.

Objective: to define *good practices* for all banks.

Three major axes:

1. Possibility of identifying unlikely scenarii for the whole system, but quite adverse for the bank, by applying chocks on the bank portfolio. *Reverse Stress Test*.
2. Analyzing model risk when using a given model for stress testing.
3. Accounting for structures of dependency between risks: economic, crash of industrial sectors, political risks, natural and ecological risks, operational risks, reputation risks, cyber-risks...

New challenges for modelling and simulation of dependent and extreme risks ...

- ▣ In all these examples, the VaR, CVaR are usually not explicit, conditional expectations for valuation are not available in closed form
- ▣ Monte-Carlo methods are a *must-use*

1.4 Two problems with naive simulations

1.4.1 Extreme/rare events

Definition (informal). An event A is extreme when $\mathbb{P}(A) \leq 10^{-2}$, and rare when $\mathbb{P}(A) \leq 10^{-4}$.

Property (Probability vs Frequency). When i.i.d. events, the first occurrence time of A is $\stackrel{d}{=} \mathcal{G}(\mathbb{P}(A)) \rightsquigarrow \mathbb{E}(\tau_A) = \frac{1}{\mathbb{P}(A)}$ time periods.

✓ In trading: VaR at 99% over 10 days: $\mathbb{E}(\tau_A) = 100 * 10j \approx 3$ years.

✓ In aeronautics: probability= 10^{-9} over 1 hour: $\mathbb{E}(\tau_A) = 10^9h \approx 114\ 000$ years!

Examples (other). Traffic congestion in telecommunications (Internet, Twitter,...) , natural disaster (floods, earthquakes), accident involving planes/nuclear plants, financial crisis...

✓ Naive Monte-Carlo sampling scheme give poor statistics.

Example (Confidence intervals from Monte Carlo methods).

Computation of $p = \mathbb{P}(A) = \mathbb{E}[\mathbf{1}_A]$.

Sample $(X_i = \mathbf{1}_{A_i} : i \geq 1)$ having Bernoulli distribution with parameter p .

- ▶ Variance $\sigma^2 = p(1 - p) \approx p$.
- ▶ Accuracy in absolute value (95% CI): $\approx 1.96\sqrt{\frac{p}{n}}$
- ▶ Accuracy in relative value (95% CI): $\approx 1.96\sqrt{\frac{1}{pn}}$.

* If relative accuracy of 10% is expected, then the required number of

simulations is $n_{10\%} \approx 400/p$:

p	10^{-4}	10^{-7}
$n_{10\%}$	4 000 000	4 000 000 000

* If relative accuracy of 1% is expected, then

$n_{1\%} \approx 40000/p$:

p	10^{-4}	10^{-7}
$n_{1\%}$	4 000 000 000	4 000 000 000 000

- ✓ The rarer, the more inefficient.
- ✓ Tail statistics depend much on model mis-specification:

Example. Let $A_x = \{X > x\}$ and $X \stackrel{d}{=} \mathcal{E}(\lambda)$ under \mathbb{P}_λ . Then

$$\mathbb{P}_\lambda(A_x) = \exp(-\lambda x) \quad \text{and} \quad \lim_{x \rightarrow +\infty} \frac{\mathbb{P}_{\lambda+\varepsilon}(A_x)}{\mathbb{P}_\lambda(A_x)} = \begin{cases} 0 & \text{if } \varepsilon > 0 \\ +\infty & \text{if } \varepsilon < 0 \end{cases} .$$

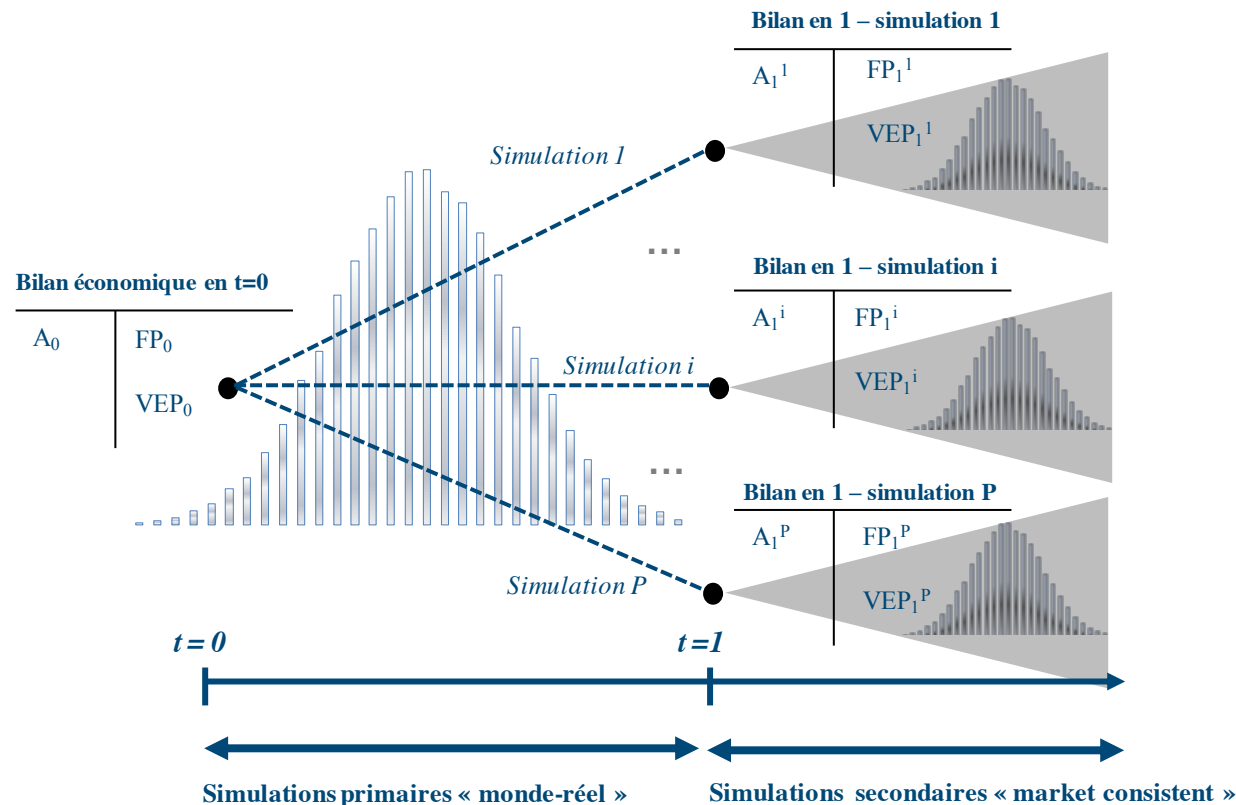
- ✓ Need for sampling adverse events (stress tests) for testing the resilience of the system...

Some solutions will be provided in this course.

1.4.2 Nested expectations

Crude MC for $\mathbb{E}[f(O, \mathbb{E}[R | O])]$: *Simulations in Simulations algorithm*

- ✓ Outer (primary) Monte-Carlo
- ✓ Inner (secondary) Monte-Carlo



PICTURE CREDIT:
from [DL09]

Crude Monte-Carlo algorithm:

- ✓ Sample M i.i.d. of O
- ✓ For each $O^{(m)}$, sample N i.i.d. samples of $R | O^{(m)}$:

$$\mathbb{E} \left[R | O^{(m)} \right] \approx \frac{1}{N} \sum_{k=1}^N R^{(m,k)}.$$

- ✓ Simple scheme, but presumably not very efficient for both the outer and the inner stages

A priori, the cost is $M \times N$.

- ✓ What is the accuracy? $O(M^{-1/2} + N^{-1/2})$?
- ✓ Can we do better?
- ✓ What about if the cost of sampling O is big (rare event)?

Some solutions later in this course.

2 Overview on standard (and sometimes inefficient) methods for simulating nested expectations OR extreme risks (not both at the same time)

2.1 Acceptance/rejection method

How to sample a random variable Z conditionally to an event A ?


Theorem. For a r.v. Z and an event A with positive probability, let $(Z_n, A_n)_{n \geq 1}$ be a sequence of i.i.d. elements distributed as (Z, A) . Define the first time when A is realized:

$$\nu = \inf\{n \geq 1 : A_n \text{ is realized}\}.$$

Then, the r.v. Z_ν has the conditional law of Z given A .

✓ Random number of loops = r.v. with geometric distribution $\mathcal{G}(1/\mathbb{P}(A))$

$$\implies \mathbb{E}[\nu] = \frac{1}{\mathbb{P}(A)}$$

✓  When A is rare, inefficient procedure

Proof. For any Borel set B , we have

$$\begin{aligned}\mathbb{P}(Z_\nu \in B) &= \sum_{n \geq 1} \mathbb{P}(Z_n \in B; A_1^c; \cdots; A_{n-1}^c; A_n) \\ &= \sum_{n \geq 1} (1 - \mathbb{P}(A))^{n-1} \mathbb{P}(Z_n \in B; A_n) \\ &= \frac{\mathbb{P}(Z \in B; A)}{\mathbb{P}(A)} = \mathbb{P}(Z \in B | A).\end{aligned}$$

□

Example (How to sample X conditionally to $\{X \in B\} = A$?).

Repeat

sample X_1, X_2, \dots independently

Until $\nu = n$ such that $X_n \in B$

▮ X_ν has the right distribution

2.2 Importance sampling

1. Modify the sampling distribution
2. Quite common in probability (Girsanov transform) and in statistics (likelihood, parametric model $(\mathbb{P}^\theta, \theta \in \Theta), \dots$)
3. Can be efficient for rare events (in some cases...)

2.2.1 Computing expectation

Definition. A probability measure \mathbb{Q} on (Ω, \mathcal{F}) defines a change of probability measure (w.r.t. \mathbb{P}), if there exists a **positive r.v.** L (the likelihood of \mathbb{Q} w.r.t. \mathbb{P}) satisfying $\mathbb{E}_{\mathbb{P}}(\mathbf{L}) = \mathbf{1}$, such that

$$\mathbb{Q}(\mathbf{A}) = \mathbb{E}[\mathbf{L}\mathbf{1}_{\mathbf{A}}] \quad \text{if} \quad \mathbf{A} \in \mathcal{F}.$$

In general we write $\mathbb{Q} = \mathbf{L} \cdot \mathbb{P}$, or $d\mathbb{Q} = \mathbf{L} \cdot d\mathbb{P}$, or $\frac{d\mathbb{Q}}{d\mathbb{P}} = \mathbf{L}$.

- ✓ Probability measures \mathbb{P} and \mathbb{Q} equivalent if $\mathbb{P}(L > 0) = 1$.
- ✓ If $\mathbb{P} \sim \mathbb{Q}$, the null-sets under \mathbb{P} and \mathbb{Q} are the same.
- ✓ If $\mathbb{P} \sim \mathbb{Q}$, $L^{-1} =$ likelihood of \mathbb{P} w.r.t. \mathbb{Q} .

2.2.2 Computing expectation

For bounded (and then integrable) $f(Y)$, we have

$$\mathbb{E}_{\mathbb{Q}}(\mathbf{f}(\mathbf{Y})) = \mathbb{E}_{\mathbb{P}}(\mathbf{L}\mathbf{f}(\mathbf{Y})), \quad \mathbb{E}_{\mathbb{P}}(\mathbf{f}(\mathbf{Y})) = \mathbb{E}_{\mathbb{Q}}(\mathbf{L}^{-1}\mathbf{f}(\mathbf{Y})).$$

Example (with density). Let Y have a probability density $p(\cdot)$ on \mathbb{R}^d – w.r.t. Lebesgue measure – and let $q(\cdot)$ be another p.d. with

$$\{x \in \mathbb{R}^d : q(x) = 0\} = \{x \in \mathbb{R}^d : p(x) = 0\}.$$

Then $\mathbf{L} = \frac{q}{p}(\mathbf{Y})$ defines a new probability measure \mathbb{Q} and the **\mathbb{Q} -distribution of Y has density q .**



Changes of probability measure can modify

- ✓ the expectation,
- ✓ the variance,
- ✓ the independence between components...

▣ Very flexible tool, but ...

2.2.3 Applications to Monte Carlo methods

Proposition. Assume $\mathbb{E}_{\mathbb{P}}|X| < +\infty$. Let $\mathbb{Q} \sim \mathbb{P}$ with likelihood L . Assume that we can sample (L, X) under \mathbb{Q} and that $(L_m, X_m)_{m \geq 1}$ are i.i.d. distributed as (L, X) under \mathbb{Q} , then

$$I_{\mathbb{Q}, M}^{\text{Imp. Samp}} = \frac{1}{M} \sum_{m=1}^M L_m^{-1} X_m \xrightarrow[M \rightarrow +\infty]{a.s.} \mathbb{E}_{\mathbb{P}}(\mathbf{X}) = \mathbb{E}[\mathbf{X}].$$

Convergence rates given by CLT.  The variance can be larger, or smaller.

Property (Optimal \mathbb{Q}). In the case $X = \mathbf{1}_A$, $L = \frac{\mathbf{1}_A}{\mathbb{P}(A)}$ gives a method with zero variance.

Untractable in practice! But $\mathbb{Q}^*(A) = 1$, showing that good IS should sample more often the extreme event.

2.2.4 Probability changes with affine transform

Proposition. Let Y be a random variable on \mathbb{R}^d , with (explicit) probability density $p(\cdot)$, strictly positive on \mathbb{R}^d .

✓ **(Change of mean)** For any $\theta_\mu \in \mathbb{R}^d$,

$$\mathbf{L} = \frac{\mathbf{p}(\mathbf{Y} - \theta_\mu)}{\mathbf{p}(\mathbf{Y})}$$

defines a new probability \mathbb{Q} under which Y has **the same distribution as $Y + \theta_\mu$ under \mathbb{P} .**

✓ **(Changes of mean and covariance)** More generally, for any $\theta_\mu \in \mathbb{R}^d$ and any square invertible matrix θ_σ ,

$$\mathbf{L} = \frac{\frac{1}{|\det.(\theta_\sigma)|} \mathbf{p}(\theta_\sigma^{-1}(\mathbf{Y} - \theta_\mu))}{\mathbf{p}(\mathbf{Y})}$$

defines a new probability \mathbb{Q} under which Y has **the same distribution as $\theta_\sigma Y + \theta_\mu$ under \mathbb{P} .**

Corollary (Gaussian distribution in dimension 1). Let $Y \stackrel{d}{=}_{\mathbb{P}} \mathcal{N}(0, 1)$.

✓ **(Change of mean)** For any $\theta_{\mu} \in \mathbb{R}$, the likelihood

$$L = \exp\left(\theta_{\mu}Y - \frac{1}{2}\theta_{\mu}^2\right)$$

leads to $Y \stackrel{d}{=}_{\mathbb{Q}} \mathcal{N}(\theta_{\mu}, 1)$.

✓ **(Changes of mean and variance)** For any $\theta_{\mu} \in \mathbb{R}$ and $\theta_{\sigma} > 0$, the likelihood

$$L = \frac{1}{\theta_{\sigma}} \exp\left(\frac{1}{2}\left(1 - \frac{1}{\theta_{\sigma}^2}\right)Y^2 + \frac{Y\theta_{\mu}}{\theta_{\sigma}^2} - \frac{\theta_{\mu}^2}{2\theta_{\sigma}^2}\right)$$

leads to $Y \stackrel{d}{=}_{\mathbb{Q}} \mathcal{N}(\theta_{\mu}, \theta_{\sigma}^2)$.

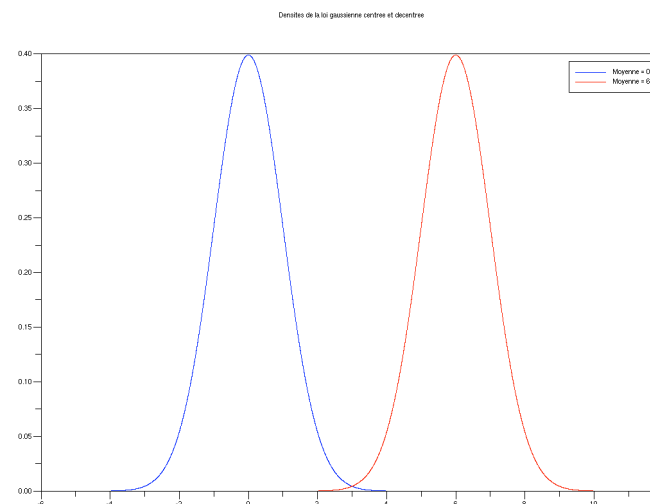
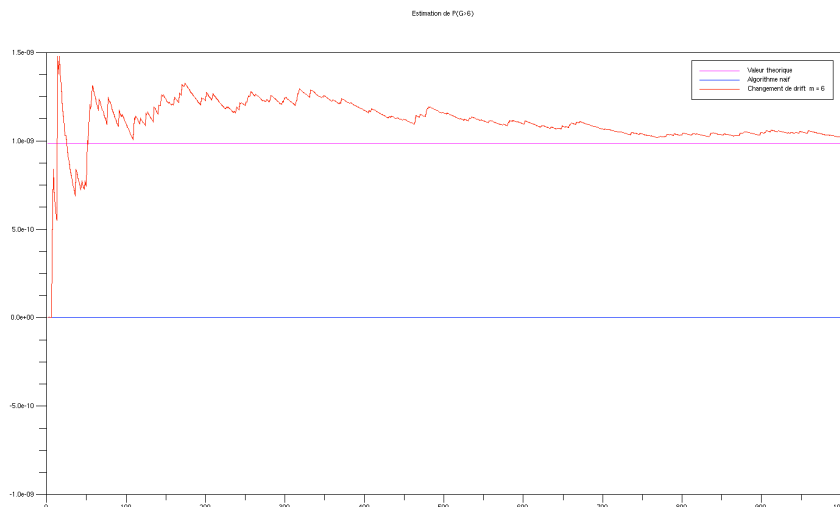
▷ Similar formulas in several dimensions

▷ In [GHS00], Delta-Gamma expansions for VaR of $V = u(t, S_t) - u(0, S_0)$:

$$\mathbf{V} \approx \partial_t \mathbf{u}(\mathbf{0}, \mathbf{S}_0) \Delta_t + \partial_x \mathbf{u}(\mathbf{0}, \mathbf{S}_0) \cdot \mathbf{S}_0 \sigma_0 \Delta \mathbf{W}_t + \frac{1}{2} \mathbf{S}_0 \sigma_0 \Delta \mathbf{W}_t \cdot \partial_{xx}^2 \mathbf{u}(\mathbf{0}, \mathbf{S}_0) \mathbf{S}_0 \sigma_0 \Delta \mathbf{W}_t$$

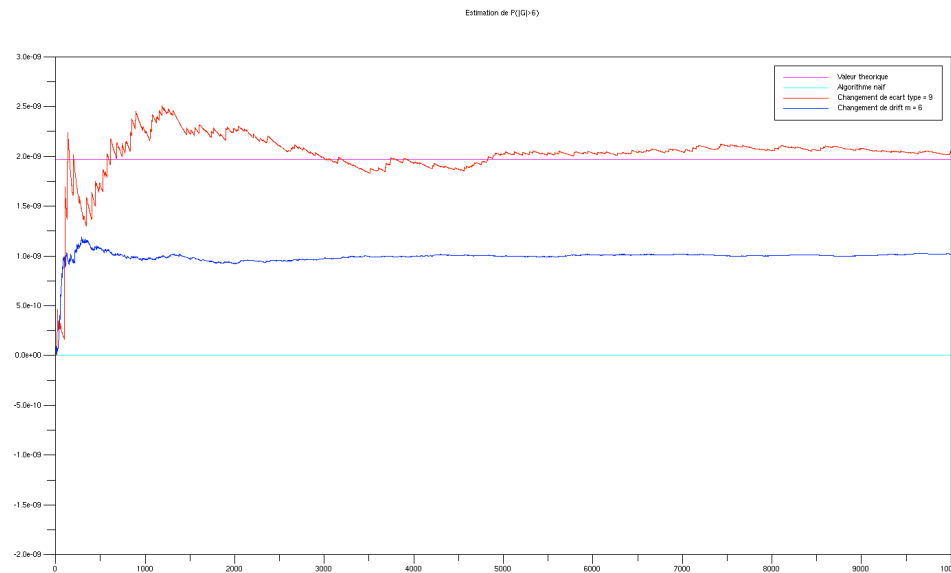
in small time ($t = 10$ days). Then, use of Gaussian IS (with optimal new drift and new covariance) as a proxy for good IS on V .

Example. $\mathbb{E}(Y)$ for $Y = \mathbf{1}_{G \geq x}$ and $G \stackrel{d}{=} \mathcal{N}(0, 1)$. Graphs for $x = 6$, with and without change of mean.



x	Exact value of $\mathbb{E}(Y) = \mathcal{N}(-x)$	Empirical mean (simple MC)	Half width of Conf. Int. (at 95%)	Empirical mean (Imp. Samp.)	Half width of Conf. Int. (at 95%)	Ratio of variances
1	1,59E-001	1,48E-001	6,96E-003	1,58E-001	3,74E-003	3,46
2	2,28E-002	2,12E-002	2,82E-003	2,24E-002	6,77E-004	17,4
3	1,35E-003	1,70E-003	8,07E-004	1,34E-003	4,84E-005	279
4	3,17E-005	0,00E+000	0,00E+000	3,24E-005	1,34E-006	∞
5	2,87E-007	0,00E+000	0,00E+000	2,90E-007	1,35E-008	∞

Numerical results for $\mathbb{E}(Y)$ with $Y = 1_{|G| \geq x}$, with or without change of variance, with or without change of variance.



Change of mean: does not converge to the right value.

Change of variance: much better.

How to know in advance which to use?

 Be careful with blind choice of measure changes, it can produce worst results.

In finance, for realistic problems (see later), we don't know if we have to change the mean, the variance, or...

Conclusion. These IS methods can be very efficient when one has a certain knowledge on the extreme event. Without this knowledge? look for data-driven method or simulation-based method?

2.3 Nested expectations

Aim: compute $\mathcal{I} = \mathbb{E}[f(\mathbf{O}, \mathbb{E}(\mathbf{R}|\mathbf{O}))]$.

 The inner and outer expectations could be under different measures

- ✓ outer \Leftarrow historical
- ✓ inner \Leftarrow risk-neutral

Four approaches:

- ✓ Simulations in the simulations (a.k.a. *nested simulations*)
- ✓ Direct (and unbiased) simulation in the case of f polynomials
- ✓ Multi-level approximations
- ✓ Empirical (linear or non-linear) regression for $\mathbb{E}(R|O)$

Sometimes for certain f , derivation of non-nested lower and upper bounds is possible.

2.3.1 Preliminaries about the propagation of bias for computing $f(\mathbb{E}(X))$

Set $\bar{X}_M = \frac{1}{M} \sum_{m=1}^M X_m$. **Bias: in general $\mathbb{E}(f(\bar{X}_M)) \neq f(\mathbb{E}[X])$**

Proposition (bias in the substitution method).

- i) If f is convex (resp. concave), the substitution method gives over-estimation (under-estimation) of the desired quantity.
- ii) If f is \mathcal{C}_b^2 and if X has finite polynomial moments (up to some order), then

$$\mathbb{E}(f(\bar{X}_M)) - f(\mathbb{E}[X]) = \frac{c_1}{M} + o(M^{-1})$$

for some constant c_1 .

▮ One can expect that in a nested MC scheme, the bias of the inner MC could be still neglected compared to the variance. Would it imply that less MC effort is required in the inner stage?

When f is related to the maximum: split the sample

Proposition (Lower and upper biased estimators). Let

$\bar{X}_{1,M} = \frac{2}{M} \sum_{m=1}^{M/2} X_i$ and $\bar{X}_{2,M} = \frac{2}{M} \sum_{m=M/2+1}^M X_i$. Set

$$\bar{f}_M = (\bar{X}_M - a)_+, \quad \underline{f}_M = \mathbf{1}_{\bar{X}_{1,M} \geq a} (\bar{X}_{2,M} - a).$$

Then \bar{f}_M and \underline{f}_M converge *a.s.* to $(\mathbb{E}[X] - a)_+$ as $M \rightarrow +\infty$, with

$$\mathbb{E}(\underline{\mathbf{f}}_M) \leq (\mathbb{E}[\mathbf{X}] - \mathbf{a})_+ \leq \mathbb{E}(\bar{\mathbf{f}}_M).$$

- ✓ Allow to sandwich the unknown quantity by two computable estimators.
- ✓ Similar techniques used in pricing Bermuda options [**Broadie-Glasserman JEDC 1997**].

2.3.2 Simulations in the simulations (a.k.a. nested simulations or Monte-Carlo²) for $\mathcal{I} = \mathbb{E}[f(O, \mathbb{E}(R|O))]$

2 stages:

- ✓ *Outer expectation*: M i.i.d. simulations (O_1, \dots, O_M)
- ✓ *Inner conditional expectation* $\mathbb{E}(R|O)$: for each O_m , N independent simulations $(R_{m,1}, \dots, R_{m,N})$ de $R | O_m$:

$$\mathbb{E}(\mathbf{R} | \mathbf{O}_m) \approx \frac{1}{N} \sum_{k=1}^N \mathbf{R}_{m,k}.$$

$$\checkmark \mathbb{E}[f(\mathbf{O}, \mathbb{E}(\mathbf{R}|\mathbf{O}))] \approx \mathbf{I}_{M,N} := \frac{1}{M} \sum_{m=1}^M f \left(\mathbf{O}_m, \frac{1}{N} \sum_{k=1}^N \mathbf{R}_{m,k} \right).$$

Theorem. Assume that the conditional distribution of $R \mid O$ has finite moments of order 4, uniformly in O :

$$\sup_x \mathbb{E} (|R - \mathbb{E}(R \mid O = x)|^4) < +\infty.$$

i) If $f : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}$ is bounded Lipschitz in the second variable, then

$$(\mathbb{E} [|I_{M,N} - \mathcal{I}|^2])^{1/2} \leq O \left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{M}} \right).$$

ii) If moreover, f is C^1 in the second variable, with a derivative bounded Lipschitz, then

$$(\mathbb{E} [|I_{M,N} - \mathcal{I}|^2])^{1/2} \leq O \left(\frac{1}{N} + \frac{1}{\sqrt{M}} \right).$$

Corollary (complexity analysis). Computational cost: $\mathcal{C} = O(NM)$. Denote by TOL the error tolerance.

a) If f is Lipschitz (as in i)): then $\mathbf{M} = \mathbf{N} = \text{TOL}^{-2}$ and $\mathcal{C} = \mathbf{O}(\text{TOL}^{-4})$.

b) If f is smoother (as in ii)): then $\mathbf{M} = \mathbf{N}^2 = \text{TOL}^{-2}$ and $\mathcal{C} = \mathbf{O}(\text{TOL}^{-3})$.

2.3.3 Direct (unbiased) simulation when f is a polynomial

Assume that $\mathbf{f}(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^n \mathbf{a}_k(\mathbf{x}) \mathbf{y}^k$ and consider the problem of finding a non-nested unbiased Monte-Carlo estimator of $\mathbb{E}[\mathbf{f}(\mathbf{O}, \mathbb{E}(\mathbf{R}|\mathbf{O}))]$.


- ✓ Draw a M -sample i.i.d. of O
- ✓ To each O_m , sample n i.i.d. copies $(R_{m,1}, \dots, R_{m,n})$ of $R | O_m$

Theorem. Assume $\mathbb{E}[|R|] < +\infty$ and $\sup_{1 \leq k \leq n} \mathbb{E}[|a_k(O)| \mathbb{E}[|R| | O]] < +\infty$.

- ✓ The estimator $\mathbf{I}^{(M)} := \frac{1}{M} \sum_{m=1}^M \sum_{k=1}^n \mathbf{a}_k(\mathbf{O}_m) \prod_{i=1}^k \mathbf{R}_{m,i}$ is an unbiased estimator, in L_1 , *a.s.* convergent to $\mathbb{E}[\mathbf{f}(\mathbf{O}, \mathbb{E}(\mathbf{R}|\mathbf{O}))]$.
- ✓ Confidence intervals available under additional integrability conditions.

Corollary (complexity analysis). Computational cost: $\mathcal{C} = O(M)$. If TOL is the error tolerance, then $\mathbf{M} = \mathbf{TOL}^{-2}$ and $\mathcal{C} = \mathbf{O}(\mathbf{TOL}^{-2})$

▣ Better convergence order compared to the case of general f .

Useable for a general f (like $x \mapsto x_+$)?  Not a good idea (in general) to approximate an arbitrary f by a polynomial and use the polynomial-type estimator (large variance).

2.3.4 Lower and upper non-nested estimator of $\mathbb{E}[(\mathbb{E}[R | O])_+]$

Takes the form of CVA computations.

1. **Non-nested lower estimator:** for any φ , we have

$$\mathbb{E}[(\mathbb{E}[\mathbf{R} | \mathbf{O}])_+] \geq \mathbb{E}[\mathbf{R}\mathbf{1}_{\varphi(\mathbf{O}) \geq 0}]$$

and the equality holds for $\varphi(O) = \mathbb{E}[R | O]$.

Proof. Use $x_+ = x\mathbf{1}_{x \geq 0}$.

$$\text{LHS-RHS} = \mathbb{E}[\mathbb{E}[R | O] (\mathbf{1}_{\mathbb{E}[R|O] \geq 0} - \mathbf{1}_{\varphi(O) \geq 0})] \geq 0.$$

2. **Non-nested upper estimator:** for any ε with $\mathbb{E}[\varepsilon | O] = 0$, we have

$$\mathbb{E}[(\mathbb{E}[\mathbf{R} | \mathbf{O}])_+] \leq \mathbb{E}[(\mathbf{R} - \varepsilon)_+]$$

and the equality holds for $\varepsilon = R - \mathbb{E}[R | O]$.

Proof. Jensen inequality.

2.3.5 Multi-level Monte-Carlo estimations

How to retrieve the $\mathbf{M} = \text{TOL}^{-2}$ et $\mathbf{C} = \mathbf{O}(\text{TOL}^{-2})$ for quite general f (under mild conditions on f)?

Multi-Level MC (smart variance reduction): [Heinrich 01, Giles '08, Giles '15].

- ✓ Several level $\ell = 0, \dots, L$
- ✓ At level ℓ , sample i.i.d. $(O_{\ell,m})_{1 \leq m \leq M_\ell}$ and for each $O_{\ell,m}$, sample i.i.d. $(R_j^{\ell,m})_{1 \leq j \leq n_\ell}$ according to $R | O = O_{\ell,m}$
- ✓ Independent sampling between levels

Definition. The multi-level estimator is defined by

$$\hat{I}_{M,n}^{ML} = \frac{1}{M_0} \sum_{m=1}^{M_0} f \left(O_{0,m}, \frac{1}{n_0} \sum_{j=1}^{n_0} R_{0,m,j} \right) + \sum_{\ell=1}^L \frac{1}{M_\ell} \sum_{m=1}^{M_\ell} \left[f \left(O_{\ell,m}, \frac{1}{n_\ell} \sum_{j=1}^{n_\ell} R_{\ell,m,j} \right) - f \left(O_{\ell,m}, \frac{1}{n_{\ell-1}} \sum_{j=1}^{n_{\ell-1}} R_{\ell,m,j} \right) \right].$$

Error analysis. Set $\mathbb{E} \left[f \left(O, \frac{1}{n} \sum_{j=1}^n R_j \right) \right]$ and assume

- ✓ $n_\ell = n_0 2^\ell$
- ✓ bias: $\mathbb{E} \left[f \left(O, \frac{1}{n} \sum_{j=1}^n R_j \right) \right] = I + O(1/n)$ with i.i.d. $R_j \stackrel{d}{=} R \mid O$.
- ✓ f is Lipschitz

Theorem. For any choice of M_ℓ , we have

- i) $\mathbb{E}[\hat{I}_{M,n}^{ML}] = I + O(1/n_L) \rightsquigarrow$ bias-accuracy at the finer level
- ii) The computational cost is $\sum_{\ell=0}^L M_\ell n_\ell$.
- iii) The quadratic error is bounded (up to constant) as

$$\mathbb{E}[(\hat{I}_{M,n}^{ML} - I)^2] \leq C \left(\frac{1}{n_L^2} + \sum_{\ell=0}^L \frac{1}{M_\ell n_\ell} \right).$$

Corollary (Asymptotic optimization). Given an error tolerance $\text{TOL} \rightarrow 0^+$, take

$$\checkmark L = \lceil \frac{-\log(n_0 \text{TOL})}{\log(2)} \rceil$$

$$\checkmark M_\ell n_\ell \text{ independent on } \ell \implies M_\ell = M_0 2^{-\ell}$$

$$\checkmark \text{Computational cost} = (L + 1)M_0 n_0 = O\left(\left|\frac{\log(\text{TOL})}{\text{TOL}}\right|^2\right).$$

😊 We achieve (up to the log factor) the optimal 1/2-order of MC methods.

2.3.6 Empirical regression

Definition (terminology from Machine Learning). Let O and R two random variables

- ✓ $R \in \mathbb{R}$, the **response** (square integrable);
- ✓ $O \in \mathbb{R}^d$, the **observation**;
- ✓ **regression function** $\mathfrak{M}(\cdot)$: $\mathbb{E}(\mathbf{R}|\mathbf{O}) = \mathfrak{M}(\mathbf{O})$;
- ✓ **learning sample** of \mathfrak{M} : $(R^{(m)}, O^{(m)})_{1 \leq m \leq M}$ with size M ;
- ✓ approximation by $\widetilde{\mathfrak{M}}_M(\cdot)$ on a **linear vector space Φ with finite dimension (size K)**:

$$\Phi = \text{Vect.}(\phi_1, \dots, \phi_K) = \left\{ \varphi(\cdot) := \sum_{\mathbf{k}=1}^K \alpha_{\mathbf{k}} \phi_{\mathbf{k}}(\cdot) := \alpha \cdot \phi(\cdot), \quad \alpha \in \mathbb{R}^K \right\}.$$

Later, we will consider other approximations in non-linear space (like neural networks).

✓ $\mathfrak{M}(\mathbf{O}) = \operatorname{argmin}_{\mathfrak{M}_O} \mathbb{E}(\mathbf{R} - \mathfrak{M}_O)^2$

since $\mathbb{E}(R - \mathfrak{M}_O)^2 = \mathbb{E}(R - \mathbb{E}(R|O))^2 + \mathbb{E}(\mathbb{E}(R|O) - \mathfrak{M}_O)^2$;

✓ **empirical coefficients:** $\alpha^M := \operatorname{argmin}_{\alpha \in \mathbb{R}^K} \frac{1}{M} \sum_{m=1}^M (\mathbf{R}^{(m)} - \alpha \cdot \phi(\mathbf{O}^{(m)}))^2$.

⇒ Ordinary Least-Squares problem (linear regression)



There are maybe several minimizing coefficients (colinearity in the basis function along the learning sample): solutions are in a hyperplane.

⇒ **The good solution** (numerically and mathematically) is to take the SVD minimizer (minimizing the distance of the hyperplan to the origin = coefficients with minimal norm).

✓ **empirical regression function:**

$$\tilde{\mathfrak{M}}_M(\cdot) = \alpha^M \cdot \phi(\cdot) = \sum_{k=1}^K \alpha_k^M \phi_k(\cdot).$$



Function with random coefficients.

- ✓ Final estimator: $\mathbb{E}[\mathbf{f}(\mathbf{O}, \mathbb{E}(\mathbf{R}|\mathbf{O}))] \approx \frac{1}{M} \sum_{m=1}^M \mathbf{f}(\mathbf{O}^{(m)}, \widetilde{\mathfrak{M}}_M(\mathbf{O}^{(m)}))$.
- 😊 We obtain an evaluation of $x \mapsto \mathbb{E}(R|O = x)$ everywhere

FURTHER REMARKS

- ✓ We do not assume that the basis functions is orthormalized with respect to the distribution of O \implies allows great flexibility in the choice of basis functions and in applicability to any model for O (including jump processes ...)
 \implies data-driven method
- ✓ However, if $A_{i,j} = \mathbb{E}(\phi_i(O)\phi_j(O)) = \delta_{i,j}$, then one could set

$$\alpha_k^M = \frac{1}{M} \sum_{m=1}^M R^{(m)} \phi_k(O^{(m)}).$$


Called **quasi-regression method** \neq regression method

Applicable only in few cases (Hermite polynomials and Gaussian model, etc).

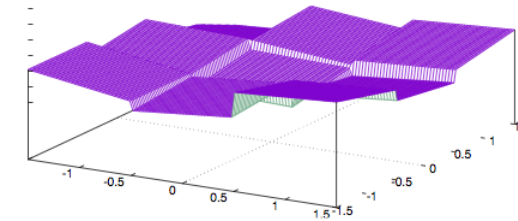
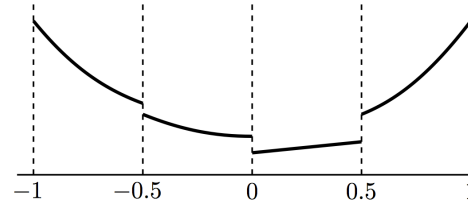
See books [**Funaro '92 ; Canuto-Hussaini-Quarteroni-Zang '06; Schoutens book'00**].

EXAMPLES OF APPROXIMATION SPACE

1st example: global polynomials

- ✓ **Popular in finance** (ex. with Bermuda options and Longstaff-Schwarz '01 algorithm).  Polynomials of LogNormal r.v. are not dense in L_2 .
- ✓ **Global approximation**, suitable for smooth functions (is low degree sufficient?)
- ✓ If \mathfrak{M} has singularities or discontinuities, slow convergence
- ✓ Mind **large values/coefficients** when O is unbounded
- ✓ **Error analysis**: very few results. See [**Funaro '92 ; Canuto-Hussaini-Quarteroni-Zang '06**].
- ✓ In full generality, **curse of dimensionality** if we seek the perfect accuracy. Sometimes, few polynomials are enough.
- ✓ **Extra available tools/results**: Cross-validation techniques, model selection, parcimony penalization...

2nd exemple: local polynomials. Local approximations on different hypercubes.



Definition.

- ✓ **Hypercubes** $C_{i_1, \dots, i_d} = [i_1 \Delta, (i_1 + 1) \Delta] \times \dots \times [i_d \Delta, (i_d + 1) \Delta]$, on a large set $[-H, H]^d$
- ✓ On each hypercube, polynomials of degree $< k$

Properties.

- ✓ Very fast regression: equivalent to independent regression on each hypercube, with low dimension
- ✓ **Approximation error:** if \mathfrak{M} of class $C^k(\dots)$ and O (with distribution μ) has exponential moments (take $H \sim c \log(1/\Delta)$ with c large enough), then

$$\inf_{\varphi \in \Phi} |\varphi - \mathfrak{M}|_{L_2(\mu)} \leq c \Delta^k.$$

ERROR ANALYSIS WITHOUT (MANY) MODEL ASSUMPTIONS



Error has to be performed on the function and not on its coefficients.

Theorem (key). Define

- i) L_2 empirical norm: $|\varphi|_{L_2(\mu^M)}^2 = \frac{1}{M} \sum_{m=1}^M \varphi^2(O^{(m)})$;
- ii) empirical regression function:


$$\widetilde{\mathfrak{M}}_M(\cdot) := \arg \min_{\varphi \in \Phi} \frac{1}{M} \sum_{m=1}^M |R^{(m)} - \varphi(O^{(m)})|^2.$$

[HYP.] Conditional variance of R is bounded: $\sigma^2 = \sup_{o \in \mathbb{R}^d} \text{Var}(R|O = o) < +\infty$.

Then, the *mean quadratic error* can be decomposed regarding bias and variance:

$$\mathbb{E} \left(|\widetilde{\mathfrak{M}}_M - \mathfrak{M}|_{L_2(\mu^M)}^2 \right) \leq \min_{\varphi \in \Phi} |\varphi - \mathfrak{M}|_{L_2(\mu)}^2 + \sigma^2 \frac{K}{M}.$$

FURTHER REMARKS

- ✓  Trade-off between K et M ? approximation error vs statistical error.
- ✓ Can be optimized.
- ✓ At the end, using local polynomials,

$$\left[\mathbb{E} \left(\left| \widetilde{\mathfrak{M}}_M - \mathfrak{M} \right|_{L_2(\mu^M)}^2 \right) \right]^{\frac{1}{2}} \underset{M \rightarrow +\infty}{\leq} C M^{-\frac{k}{2k+d}}$$

with a computational cost $O(M)$.

Take-home message: the order of convergence as a function of computational

cost is $\frac{1}{2 + \frac{\text{dimension}}{\text{smoothness}}}$.

- ✓ If $\frac{\text{dimension}}{\text{smoothness}} \ll 1$, we retrieve the MC rate!
- ✓ Otherwise, empirical regression not appropriate

BACK TO THE INITIAL PROBLEM $\mathcal{I} = \mathbb{E}(f(O, \mathbb{E}(R | O)))$

Theorem. Assume

- (i) $f : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}$ is globally Lipschitz in the second variable;
- (ii) $f(O, \mathfrak{M}(O))$ is square integrable.

Then

$$\left(\mathbb{E} \left[|\hat{\mathcal{I}}_M - \mathcal{I}|^2 \right] \right)^{1/2} \leq C \left(\left[\mathbb{E} \left(|\tilde{\mathfrak{M}}_M - \mathfrak{M}|_{\mathbf{L}_2(\mu^M)}^2 \right) \right]^{\frac{1}{2}} + \sqrt{\frac{1}{M}} \right)$$

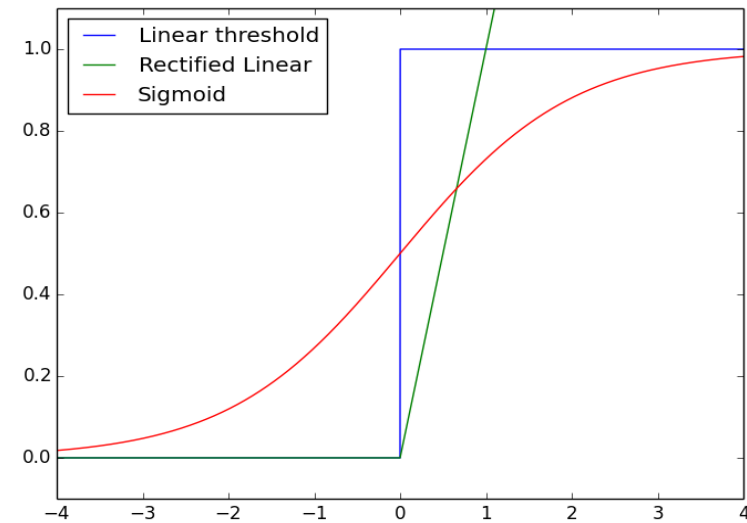
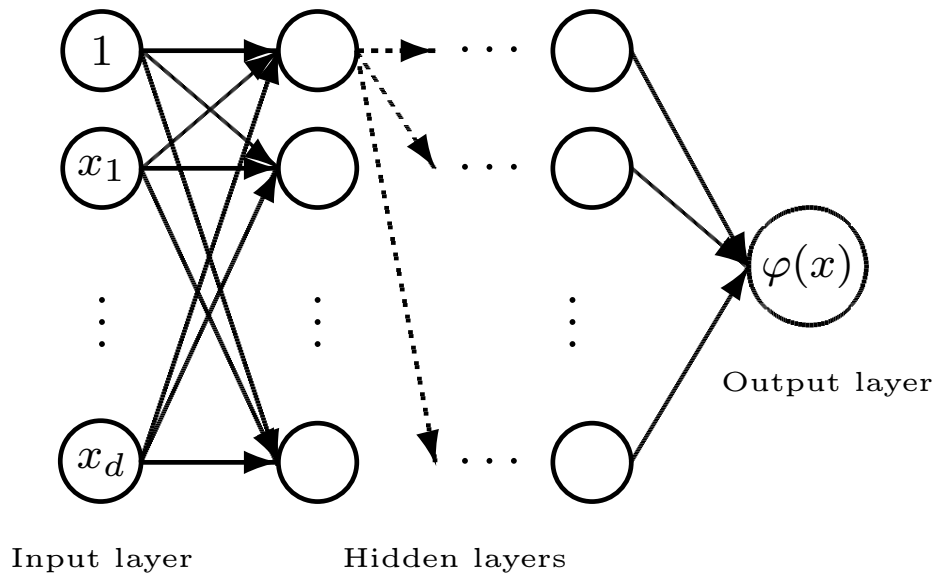
with $\hat{\mathcal{I}}_M := \frac{1}{M} \sum_{m=1}^M f(O^{(m)}, \tilde{\mathfrak{M}}_M(O^{(m)}))$.

To summarize:

- ✓ If $\frac{\text{dimension}}{\text{smoothness}} \ll 1$, nested regression is as efficient as the MultiLevel MC rate
- ✓ However, the Nested MC or MLMC do not give any information on $\mathbb{E}(R|O)$ (or its distribution)

NON-LINEAR REGRESSION

Example (Feed-forward Neural Network).



Representation of neural networks

Different activation functions: $\mathbf{1}_{y \geq 0}$, $\max(y, 0)$, $1/(1 + e^{-y})$.

[Anthony, Bartlett, *Neural Network Learning*, 1999.]

Theorem. Let (O, R) be a random vector in $\mathbb{R}^d \times \mathbb{R}$, with R bounded, and let Φ be a dictionary of functions $\mathbb{R}^d \mapsto \mathbb{R}$. Assume that the dictionary is countably generated (like for NN). Consider a learning sample $(O^{(m)}, R^{(m)})_{m=1, \dots, M}$ and define

$$\tilde{\mathfrak{M}}_M := \arg \min_{\varphi \in \Phi} \frac{1}{M} \sum_{m=1}^M |R^{(m)} - \varphi(O^{(m)})|^2.$$

Then

$$\mathbb{E} \left[\int_{\mathbb{R}^d} \left| [\tilde{\mathfrak{M}}_M]_{|\mathbf{R}|_\infty}(\mathbf{x}) - \mathfrak{M}(\mathbf{x}) \right|^2 \mu(d\mathbf{x}) \right] \leq_c \text{VC}(\Phi) \frac{(1 + \log M)}{M} + \inf_{\varphi \in \Phi} \mathbb{E} [|\varphi(\mathbf{O}) - \mathbb{E}[\mathbf{R}|\mathbf{O}]|^2]$$

where $\text{VC}(\Phi)$ is the Vapnik-Chervonenkis dimension of Φ .

- ✓ Essentially, same control as for vector space case, assuming additionally R bounded
- ✓ In the vector space case, $\text{VC}(\Phi^{\text{VS}}) = O(K)$.
- ✓ For NN and RELU activation, $\text{VC}(\Phi^{\text{NN}}) = O(\#\{\text{Weights}\} \times \#\{\text{Layers}\}^2)$
- ✓ Computing $\tilde{\mathfrak{M}}_M$ may require a specific optimization algorithm (for non convex problems)

3 Sampling extremes using Markov chains on path space

3.1 Monte-Carlo Markov Chain

3.1.1 Brief overview of important principles

- ✓ Markov chain with transition kernel $P(x, \cdot)$:

$$X_{k+1} \stackrel{d}{=} P(X_k, \cdot)$$

- ✓ Long-term behavior? convergence to a stationary distribution π ?

$$x = X_0 \xrightarrow{P} X_1 \xrightarrow{P} \dots \xrightarrow{P} X_n \overset{\text{dist.}}{\approx} \pi \xrightarrow{P} \dots \xrightarrow{P} X_{2n} \overset{\text{dist.}}{\approx} \pi \dots$$

- ✓ Characterization of π using reversibility: if P is π -reversible, that is

$$(X_0, X_1) \stackrel{d}{=} (X_1, X_0)$$

then **the distribution π of X_0 is stationary for the transition $P(\cdot, \cdot)$.**

✓ Chain with rejection in \mathcal{A} : transition Q defined on \mathcal{A} by

$$\int \varphi(y)Q(x, dy) := \int_{\mathcal{A}} \varphi(y)P(x, dy) + \varphi(x)P(x, \mathcal{A}^c).$$

If P is π -reversible, then Q is $\pi|_{\mathcal{A}}$ -reversible.

Proof.

$$\begin{aligned} \int_{\mathcal{A}} \int_{\mathcal{A}} \gamma(x)\varphi(y)Q(x, dy)\pi(dx) &= \int_{\mathcal{A}} \int_{\mathcal{A}} \gamma(x)\varphi(y)P(x, dy)\pi(dx) + \int_{\mathcal{A}} \gamma(x)\varphi(x)P(x, \mathcal{A}^c)\pi(dx) \\ &= \int_{\mathcal{A}} \int_{\mathcal{A}} \gamma(y)\varphi(x)P(x, dy)\pi(dx) + \int_{\mathcal{A}} \gamma(x)\varphi(x)P(x, \mathcal{A}^c)\pi(dx) \\ &= \int_{\mathcal{A}} \int_{\mathcal{A}} \gamma(y)\varphi(x)Q(x, dy)\pi(dx). \end{aligned}$$

3.1.2 Convergence in finite dimension

Ready-to-use statement.

Theorem ([FM03, Proposition 2],[MT09, Theorem 15.0.1]). Let Q be a phi-irreducible aperiodic transition kernel on $\mathcal{A} \subset \mathbb{R}^d$ and let $\mathcal{C} \subset \mathcal{A}$ be an accessible petite set. Assume there exist $\delta \in (0, 1)$, $b \in \mathbb{R}$ and a measurable $V : \mathcal{A} \rightarrow [1, +\infty)$, bounded on \mathcal{C} , such that

$$QV(\mathbf{x}) \leq \delta V(\mathbf{x}) + b\mathbf{1}_{\mathbf{x} \in \mathcal{C}}, \quad \mathbf{x} \in \mathcal{A}.$$

Let $p \geq 2$, assume that $\mathbb{E}[V(X_0)] < +\infty$. Then for some constant $0 < r < 1$, for any g bounded in $V^{1/p}$ -norm, we have

$$\left(\mathbb{E} \left[\left| \frac{1}{n} \sum_{i=1}^n g(\mathbf{X}_i) - \pi(g) \right|^p \right] \right)^{1/p} \leq \frac{C}{\sqrt{n}}, \quad |\mathbb{E}[g(\mathbf{X}_n)] - \pi(g)| \leq Cr^n,$$

for some finite constant $C > 0$.

3.1.3 Some (important) examples according to π

In the sequel, for sampling efficiently extreme events, we will use reversible Markov transition, with small transformation: we call them **shakers**.

We write the transition $P(x, \cdot)$ under its sampling form:

$$\mathbf{K}(\mathbf{x}, \mathbf{Y}) \stackrel{d}{=} \mathbf{P}(\mathbf{x}, \cdot)$$

for some independent random variable Y .

On the top on this, there are some probabilistic models that are functions of X and that describe the extreme risks;

$$\mathbf{Z} = \phi_{\mathbf{Z}}(\mathbf{X}).$$

Low-dimensional parameterization of reversible shakers

✓ If π is a standard Gaussian distribution,

$$\mathbf{K}_\rho(\mathbf{x}, \mathbf{y}) = \rho\mathbf{x} + \sqrt{1 - \rho^2}\mathbf{y}$$

with $Y \stackrel{d}{=} \mathcal{N}(0, 1)$, $-1 \leq \rho \leq 1$.

If $X \stackrel{d}{=} \mathcal{N}(0, 1) \perp Y$, then $(\mathbf{X}, \mathbf{K}_\rho(\mathbf{X}, \mathbf{Y})) \stackrel{d}{=} (\mathbf{K}_\rho(\mathbf{X}, \mathbf{Y}), \mathbf{X})$.

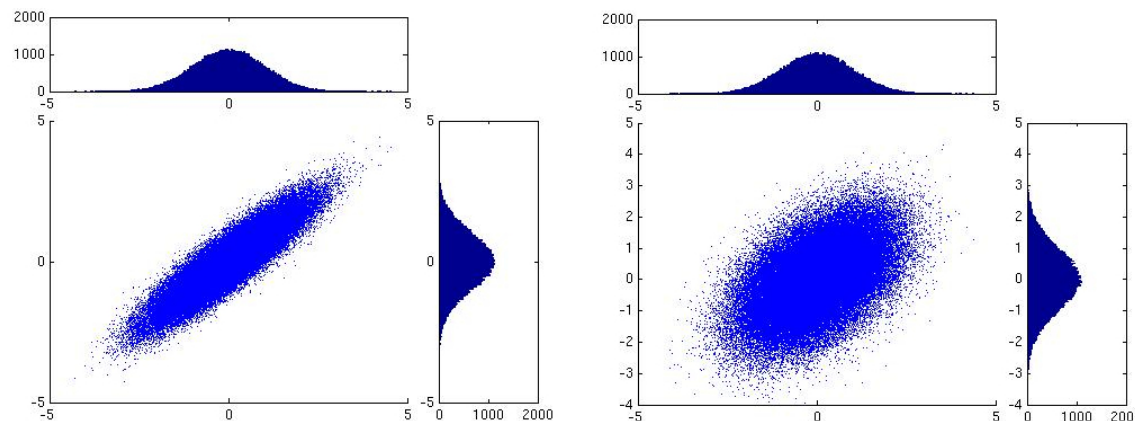


Figure 1: Shaking $\mathcal{N}(0, 1)$ with $\rho = 0.9$ and $\rho = 0.5$

✓ If X is a standard Brownian motion,

$$K(X, Y) = \left(\int_0^t \rho_s dX_s + \int_0^t \sqrt{1 - \rho_s^2} dY_s \right)_{0 \leq t \leq T}$$

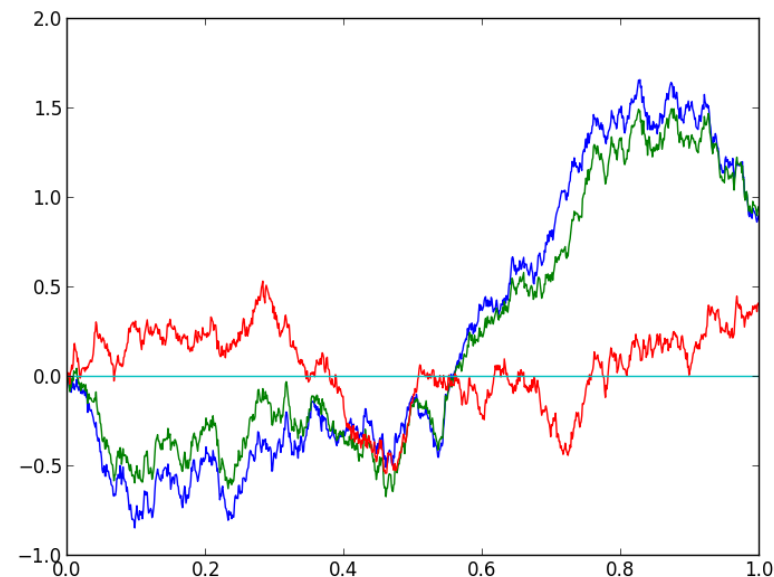
with Y is an independant BM and $\rho \in [-1, 1]$ deterministic function.

Shaker means $\rho \approx 1$.

Shaking OU model

$$dZ_t = \lambda(\mu - Z_t)dt + \sigma dX_t.$$

Path of Z , shaken Z with $\rho = 0.9$
and $\rho = 0.1$



Can be extended to any Gaussian Hilbert space (Fractional Brownian Motion, ...). [Nualart 2006]

✓ For a Gamma distribution $Ga \sim \text{Gamma}(\alpha, \beta)$, i.e

$$\mathbb{P}(Ga \in dx) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx, x > 0$$

we can use a transformation based on the Beta-Gamma algebra

$$\mathbf{K}(\mathbf{Ga}) = \mathbf{Ga} * \text{Beta}(\alpha(\mathbf{1} - \mathbf{p}), \alpha\mathbf{p}) + \text{Gamma}(\alpha\mathbf{p}, \beta),$$

(includes exponential distribution).

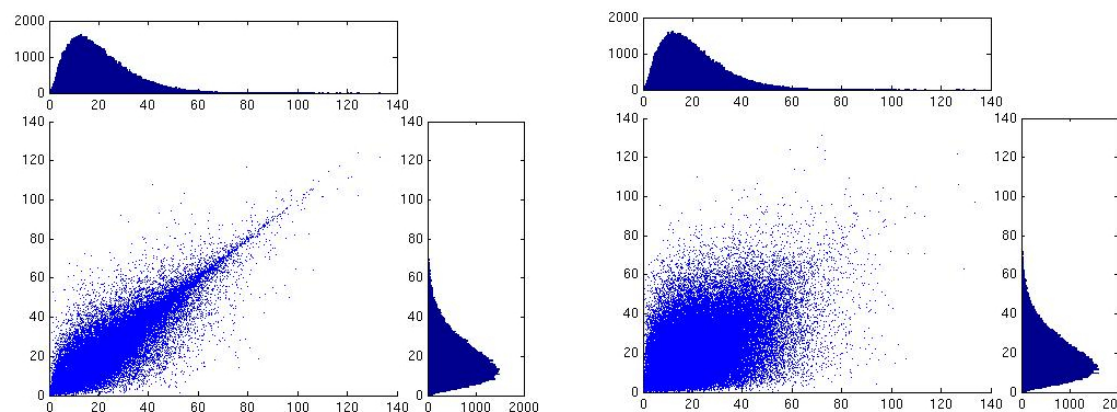


Figure 2: Shaking $\text{Gamma}(2.5, 0.12)$ with $p = 0.1$ and $p = 0.5$

✓ **Compound Poisson process.** Let $X_t = \sum_{k=1}^{N_t} J_k$ be a $\text{CPP}(\lambda, \mu)$.

$$\text{CPP decomposition: } X_t = X_t^a + X_t^b$$

where $X^a \stackrel{d}{=} \text{CPP}((1-p)\lambda, \mu)$ and $X^b \stackrel{d}{=} \text{CPP}(p\lambda, \mu)$

$$\mathbf{K}(\mathbf{X}, \mathbf{Y}) = (\mathbf{X}_t^a + \mathbf{Y}_t)_{0 \leq t \leq T} \quad \text{with} \quad \mathbf{Y}_t \stackrel{d}{=} \text{CPP}(p\lambda, \mu)$$

▮ remove each jump with probability p and add extra independent ones.

Useful when extreme risks are described by jump processes.

3.2 Back to rare-event statistics

Here, extreme event is given explicitly (and not implicitly by a VaR).

Aim: efficient numerical evaluation of

$$\mathbb{P}(\mathbf{Z} \in \mathbf{A}^{\mathbf{Z}}) \quad \text{or} \quad \mathbb{E}(\varphi^{\mathbf{Z}}(\mathbf{Z}) \mathbf{1}_{\mathbf{Z} \in \mathbf{A}^{\mathbf{Z}}}),$$

$$\mathbb{E}(\varphi^{\mathbf{Z}}(\mathbf{Z}) \mid \mathbf{Z} \in \mathbf{A}^{\mathbf{Z}}),$$

$$\partial_{\theta} \mathbb{E}_{\theta}(\varphi^{\mathbf{Z}}(\mathbf{Z}) \mid \mathbf{Z} \in \mathbf{A}^{\mathbf{Z}}),$$

✓ Z = economic and financial variables of interest with

$$\mathbf{Z} := \phi_{\mathbf{Z}}(\mathbf{X})$$

✓ X = stochastic variable generating the risk factors

✓ $A^{\mathbf{Z}}$ = set of critical scenarios (defined by user/regulator)

✓ $A = \phi_{\mathbf{Z}}^{-1}(A)$, $\varphi^{\mathbf{Z}}(\phi_{\mathbf{Z}}(x)) = \varphi(x)$, so that $\mathbb{E}(\varphi^{\mathbf{Z}}(\mathbf{Z}) \mid \mathbf{Z} \in \mathbf{A}^{\mathbf{Z}}) = \mathbb{E}(\varphi(\mathbf{X}) \mid \mathbf{X} \in \mathbf{A})$

✓ The Markov chain will be performed at the level of X (usually a stochastic "path" in finance)

3.2.1 Setting and Shaker

- ✓ Given probability space: $(\Omega, \mathcal{F}, \mathbb{P})$
- ✓ **State space \mathbb{S} for stochastic element X .**
 - ▶ $\mathbb{S} = \mathbb{R}^d$, i.e. $X =$ finite-dimensional random variable
 - ▶ $\mathbb{S} = \mathbb{C}([0, 1], \mathbb{R}^d)$, e.g. $X =$ Brownian motion or SDE or Itô process
 - ▶ $\mathbb{S} = \mathbb{D}([0, 1], \mathbb{R}^d)$, e.g. $X =$ compound Poisson process, jump processes
 - ▶ $\mathbb{S} = \mathbb{R}^{\mathbb{N}}$, i.e. $X =$ countably infinite number of random variables

If a path space, may be non-Markovian.

Definition (of reversible shaking transformation). Similar to before.

- ✓ Define $\mathcal{K}(\cdot) = \mathbf{K}(\cdot, \mathbf{Y})$ for some $K(\cdot)$ and Y independent of everything else
- ✓ $\mathcal{K}(\cdot)$ is a **reversible shaker** for X if:

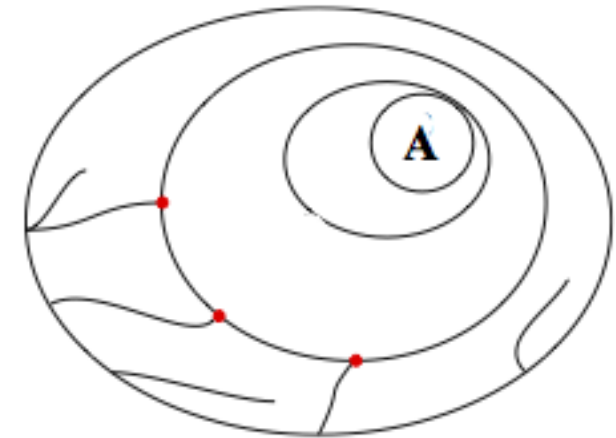
$$(\mathbf{X}, \mathbf{K}(\mathbf{X}, \mathbf{Y})) \stackrel{d}{=} (\mathbf{K}(\mathbf{X}, \mathbf{Y}), \mathbf{X}).$$

3.2.2 Splitting

Define a sequence of n nested subsets of \mathbb{S}

$$\mathbb{S} := A_0 \supset \cdots \supset A_k \supset \cdots \supset A_n := A,$$

$$\mathbb{P}(\mathbf{X} \in \mathbf{A}) = \prod_{k=1}^n \mathbb{P}(\mathbf{X} \in \mathbf{A}_k | \mathbf{X} \in \mathbf{A}_{k-1}).$$



For instance: Scenarios of Portfolio loss of 1 G€ is seen as a nested sequence of larger and larger losses: 10M€, 100M€, 500M€, 800M€, 900M€

3.2.3 Shaker with rejection

$$\mathcal{M}_k^{\mathcal{K}}(\mathbf{X}) := \begin{cases} \mathcal{K}(X) & \text{if } \mathcal{K}(X) \in A_k \\ X & \text{if } \mathcal{K}(X) \notin A_k. \end{cases}$$



We know it preserves the conditional distribution of $X \mid X \in A_k$.

3.2.4 1st method: the Parallel One-Path (POP) approach

😊 The Markov transformation $\mathcal{M}_k^{\mathcal{K}}$ leaves invariant the distribution of $X|X \in A_k$.

- ✓ **Birkhoff's ergodic theorem:** for ergodic Markov chains $(Z_i)_i$ with a unique invariant distribution π , we have

$$\frac{1}{N} \sum_{i=1}^N f(Z_i) \xrightarrow[N \rightarrow +\infty]{} \int f d\pi \quad a.s.$$

- ✓ **Proposition (Markov chain at level k).** Given an initial position

$$\mathbf{X}_{k,0} \in \mathbf{A}_k$$

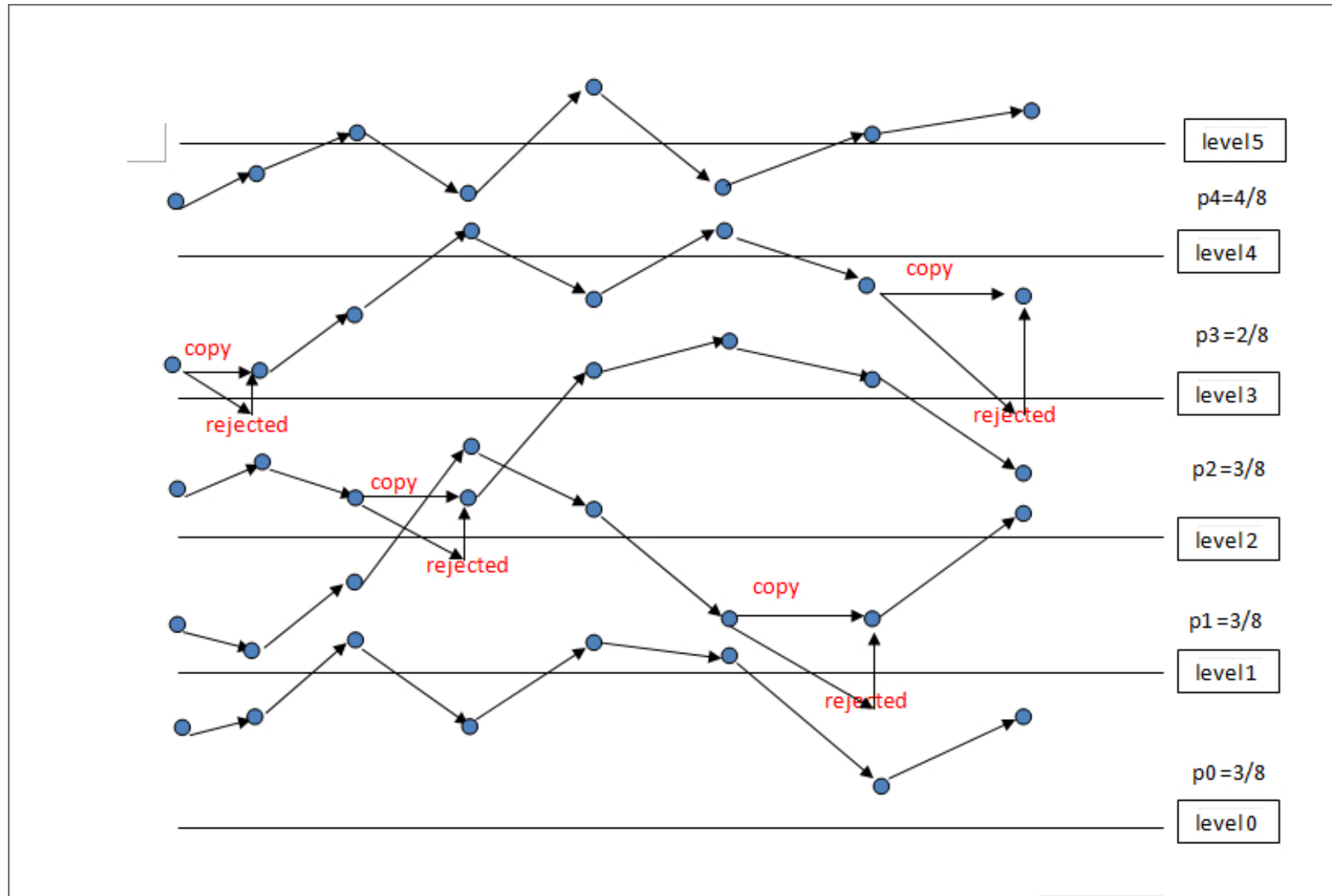
define

$$\mathbf{X}_{k,i} := \mathcal{M}_k^{\mathcal{K}}(\mathbf{X}_{k,i-1}), \quad i \geq 1.$$

Then, as $N \rightarrow +\infty$ (average over time),

$$\mathbb{P}(\mathbf{X} \in \mathbf{A}_{k+1} | \mathbf{X} \in \mathbf{A}_k) \approx \frac{1}{N} \sum_{i=0}^{N-1} \mathbf{1}_{\mathbf{A}_{k+1}}(\mathbf{X}_{k,i}).$$

POP SAMPLING SCHEME



Theorem. Assume the conditions of Theorem of slide 54. If N is the length of simulations, L_2 -convergence at rate \sqrt{N} .

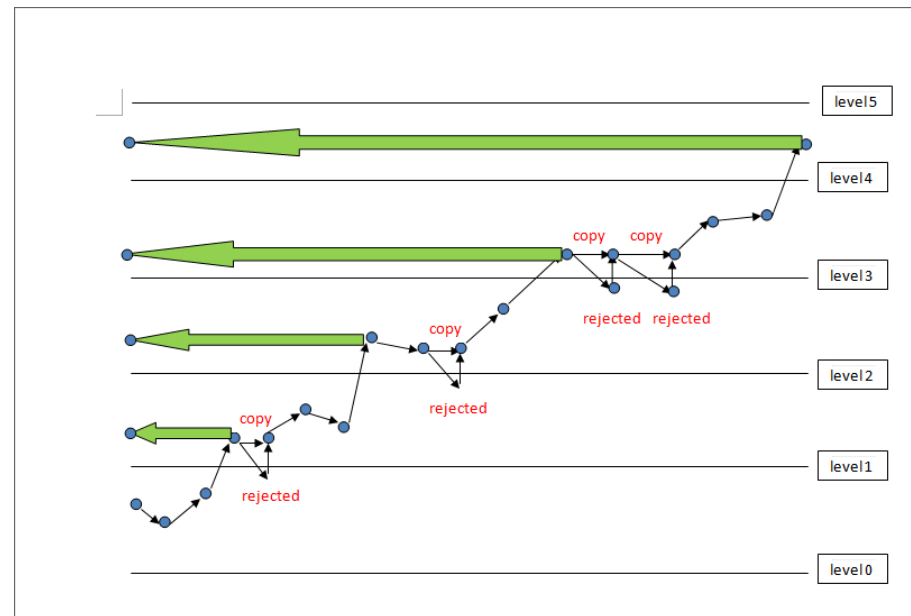
In the case of Gaussian shaker, conditions are fulfilled.

Property. Independent computations

- ✓ in the computing sense (parallel-processing \implies fast computations)
- ✓ in the probabilistic sense (independent Markov chains \implies low variance)

How to get initial positions?

- ✓ $X_{0,0}$ is simulated as an independent copy of X , then apply $X_{0,i+1} = \mathcal{M}_0^K(X_{0,i})$ until getting one point in A_1 , then take it as $X_{1,0}$
- ✓ apply $X_{1,i+1} = \mathcal{M}_1^K(X_{1,i})$ until hitting A_2 , then take it as $X_{2,0}$ etc
- ✓ Extra computational cost is negligible w.r.t the global cost of the algorithm.



Adaptive levels. Possibility of an adaptive version where the levels are chosen online to ensure 10% for the conditional probabilities. See refs on the summerschool webpage.

3.2.5 2nd method: Interacting Particles System (IPS) approach

We define $(X_i)_{0 \leq i \leq n-1}$ as a Markov chain taking values in \mathbb{S} :

$$\mathbf{X}_0 \stackrel{d}{=} \mathbf{X} \quad \text{and} \quad \mathbf{X}_i := \mathcal{M}_i^{\mathcal{K}}(\mathbf{X}_{i-1}), \mathbf{1} \leq i \leq n-1.$$

Theorem (Feymann-Kac measure representation).

$$\mathbb{P}(X \in A_n) = \mathbb{E}\left(\prod_{i=0}^{n-1} \mathbf{1}_{A_{i+1}}(X_i)\right),$$

$$\mathbb{E}(\varphi(X) | X \in A_n) = \frac{\mathbb{E}(\varphi(X_{n-1}) \prod_{i=0}^{n-1} \mathbf{1}_{A_{i+1}}(X_i))}{\mathbb{E}(\prod_{i=0}^{n-1} \mathbf{1}_{A_{i+1}}(X_i))}.$$

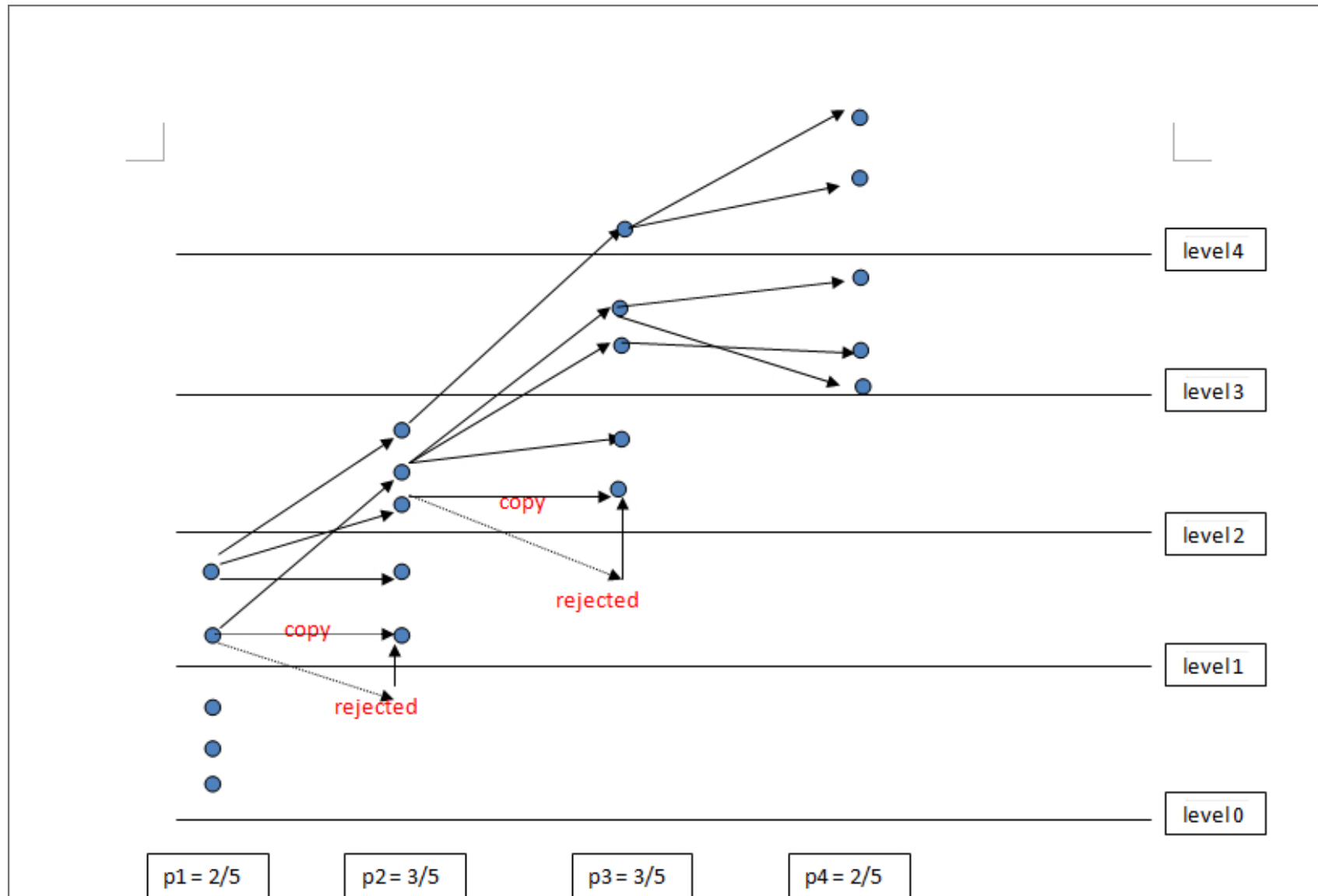
- ✓ Interacting particles system (IPS, [Del Moral 2004]) method gives estimation of $\mathbb{E}\left(f(X_0, \dots, X_n) \prod_{i=0}^{n-1} G_i(X_i)\right)$

Theorem. (Average over space) Using a system with M particles in interaction, L_2 -convergence at rate \sqrt{M} .



No parallel computation, requires to store in memory all the particles.

IPS SAMPLING SCHEME



3.3 Numerical tests

3.3.1 Maximum of Ornstein-Uhlenbeck process

Model:

$$dY_t = \lambda(\mu - Y_t)dt + \sigma dW_t, \quad Y_0 = 0 \quad \text{with } \lambda = 1, \mu = 0, \sigma = 1, T = 1.$$

We want to estimate $\mathbb{P}(\max_{0 \leq l \leq 100} \tilde{Y}_{t_l} > \mathbf{3.6})$ with $t_l = hl$ where \tilde{Y} is the Euler scheme simulation of Y with time step $h = 0.01$.

Experiments:

- ✓ 10^7 MC simulations with IS gives $\text{CI} = [0.9772, 1.0038] \times 10^{-7}$
- ✓ Intermediate levels are defined as $L_i = 3.6 \times (\frac{i}{5})^{1/2}, 1 \leq i \leq 5$.
- ✓ Output statistics computed with 50 macro MC runs.

	mean	std	std/mean
$\rho = 0.9$	1.06e-07	5.12e-08	0.48
$\rho = 0.75$	9.51e-08	2.15e-08	0.22
$\rho = 0.5$	9.32e-08	9.42e-08	1.01

IPS: $M = 10^5$

	mean	std	std/mean
$\rho = 0.9$	9.80e-08	6.74e-09	0.07
$\rho = 0.75$	1.00e-07	9.52e-09	0.10
$\rho = 0.5$	1.05e-07	2.78e-08	0.27

POP: $N = 10^5$

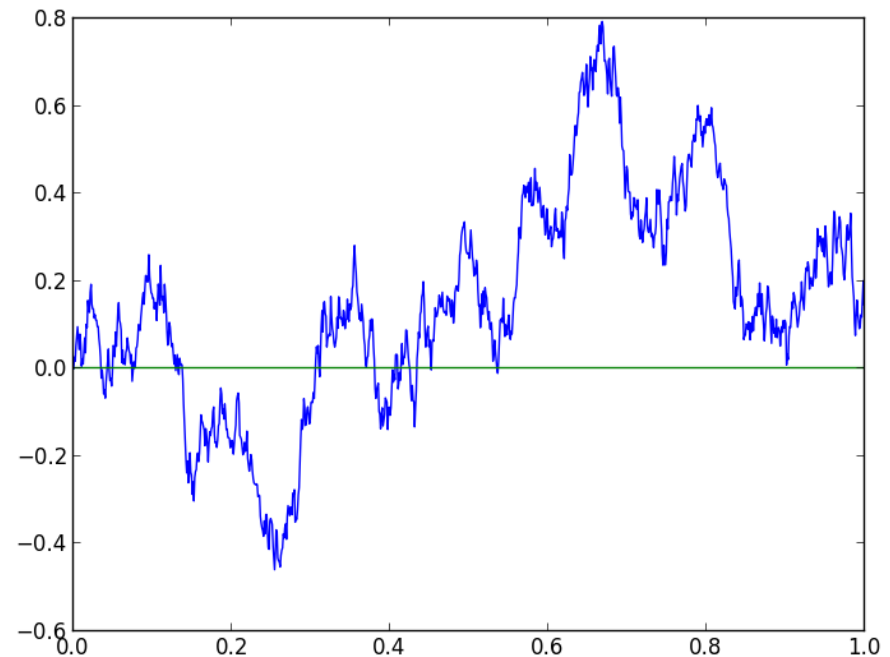
3.3.2 Oscillation of OU process

Let W be a standard Brownian motion, consider the solution to

$$dZ_t = \lambda(\mu - Z_t)dt + \sigma dW_t, \quad Z_0 = 0.$$

We wish to compute the rare event probability

$$\mathbb{P}\left(\max_{0 \leq t \leq T} Z_t > 1.6 \text{ and } \min_{0 \leq t \leq T} Z_t < -1.6\right).$$



Importance sampling techniques are difficult to apply on this example.

Aim at computing (the time-discretized quantity)

$$\mathbb{P}\left(\max_{0 \leq l \leq 100} \tilde{Y}_{t_l} > 1.6 \text{ and } \min_{0 \leq l \leq 100} \tilde{Y}_{t_l} < -1.6\right).$$

✓ 7×10^9 MC simulation gives $CI = [3.9709, 4.3691] \times 10^{-7}$

✓ Set $L_i = 1.6 \times (\frac{i}{5})^{1/2}$ and $A_i = \left\{ \max_{0 \leq l \leq 100} \tilde{Y}_{t_l} > L_i \text{ and } \min_{0 \leq l \leq 100} \tilde{Y}_{t_l} < -L_i \right\}$

✓ 100 macro MC runs

	mean	std	std/mean
$\rho = 0.9$	4.01e-07	1.23e-07	0.31
$\rho = 0.75$	4.10e-07	1.67e-07	0.41
$\rho = 0.5$	2.44e-07	4.76e-07	1.95

IPS: $M = 10^5, \alpha = 1$

	mean	std	std/mean
$\rho = 0.9$	4.14e-07	2.68e-08	0.06
$\rho = 0.75$	4.18e-07	4.60e-08	0.11
$\rho = 0.5$	4.29e-07	1.26e-07	0.29

POP: $N = 10^5$

3.3.3 Measuring the default probabilities in credit portfolios

[Model as in Carmona-Fouque-Vestal 2009]

$$\checkmark \text{ } I \text{ Firms: } \begin{cases} d\mathbf{X}_i(t) = r\mathbf{X}_i(t)dt + \sigma(t)\mathbf{X}_i(t)d\mathbf{W}_i(t), \\ d\sigma(t) = \kappa(\bar{\sigma} - \sigma(t))dt + \gamma\sqrt{\sigma(t)}d\mathbf{W}_t \end{cases}$$

with positive $\kappa, \bar{\sigma}$ and γ and $d\langle W_i, W_j \rangle = \rho dt, i \neq j, \quad d\langle W_i, W \rangle = \rho_\sigma dt.$

✓ Time of default for firm i :

$$\tau_i(\mathbf{B}_i) := \inf \{t : \mathbf{X}_i(t) \leq \mathbf{B}_i\}$$

with a fixed default boundary B_i .

✓ Probability of $L + 1$ defaults at least: $\mathbb{P}(\mathbf{L}) = \mathbb{P} \left(\sum_{i=1}^I \mathbf{1}_{\{\tau_i(\mathbf{B}_i) \leq T\}} > \mathbf{L} \right).$

✓ Path-dependent problem. As opposed to [CFV09], we do not perform IPS based on time-discretization of (X, σ) .

Good nested sets:

$$A_k = \left\{ \omega : \sum_{i=1}^I \mathbf{1}_{\left\{ \tau_i \left(\mathbf{X}_0(\mathbf{i}) - \frac{k}{n} (\mathbf{X}_0(\mathbf{i}) - \mathbf{B}_i) \right) (\omega) \leq T \right\}} > L \right\}, \quad 1 \leq k \leq n.$$

Less efficient nested sets:

$$\tilde{A}_k = \left\{ \omega : \sum_{i=1}^I \mathbf{1}_{\left\{ \tau_i(B_i)(\omega) \leq T \right\}} > \frac{k}{n} \mathbf{L} \right\}, \quad 1 \leq k \leq n.$$

Model parameters:

$X_i(0)$	r	ρ	$\sigma(0)$	κ	$\bar{\sigma}$	γ	ρ_σ
90	0.06	0.10	0.4	3.5	0.4	0.7	-0.06

Credit risk parameters: $I = 125$, threshold level $B_i = B$, $T = 1$.

Benchmark value for $L = 100$ and $B = 36$ (by Crude Monte-Carlo using 3×10^9 samples),

$$\mathbb{P}(\mathbf{L}) \in [4.92, 5.13] \times 10^{-6}$$

with 99%-confidence.

DETAILED PERFORMANCES

Non adaptive version

	IPS			POP		
	mean	std.	std./mean	mean	std.	std./mean
	$(\times 10^{-5})$	$(\times 10^{-6})$		$(\times 10^{-6})$	$(\times 10^{-7})$	
$\rho = 0.9$	1.18	5.72	0.48	5.01	8.03	0.16
$\rho = 0.7$	1.53	8.73	0.57	4.99	10.21	0.20
$\rho = 0.5$	1.40	15.40	1.1	5.02	19.38	0.39

Table 1: Estimators of default probability (mean) for $L = 100$ and $B = 36$ with empirical standard deviation (std.) for IPS and POP method based on 100 macro-runs.

Each intermediate level estimator in both the methods is based on $\mathbf{M} = \mathbf{N} = 10^4$ simulations.

Adaptive version

	Adaptive IPS			Adaptive POP		
	mean ($\times 10^{-6}$)	std. ($\times 10^{-6}$)	std./mean	mean ($\times 10^{-6}$)	std. ($\times 10^{-6}$)	std./mean
$\rho = 0.9$	4.93	1.91	0.39	5.16	0.85	0.16
$\rho = 0.7$	5.42	1.58	0.29	4.98	1.02	0.20
$\rho = 0.5$	6.40	5.00	0.78	5.35	2.05	0.38

Table 2: Estimators of default probability (mean) for $L = 100$ and $B = 36$ with empirical standard deviation (std.) for adaptive IPS and POP methods ($p = 10\%$) based on 100 algorithm macro-runs. Each intermediate level estimator in both methods is based on $M = N = 10^4$ simulations.

 No significant increase of statistical error.

Tuning the shaking parameters for the non-adaptive version

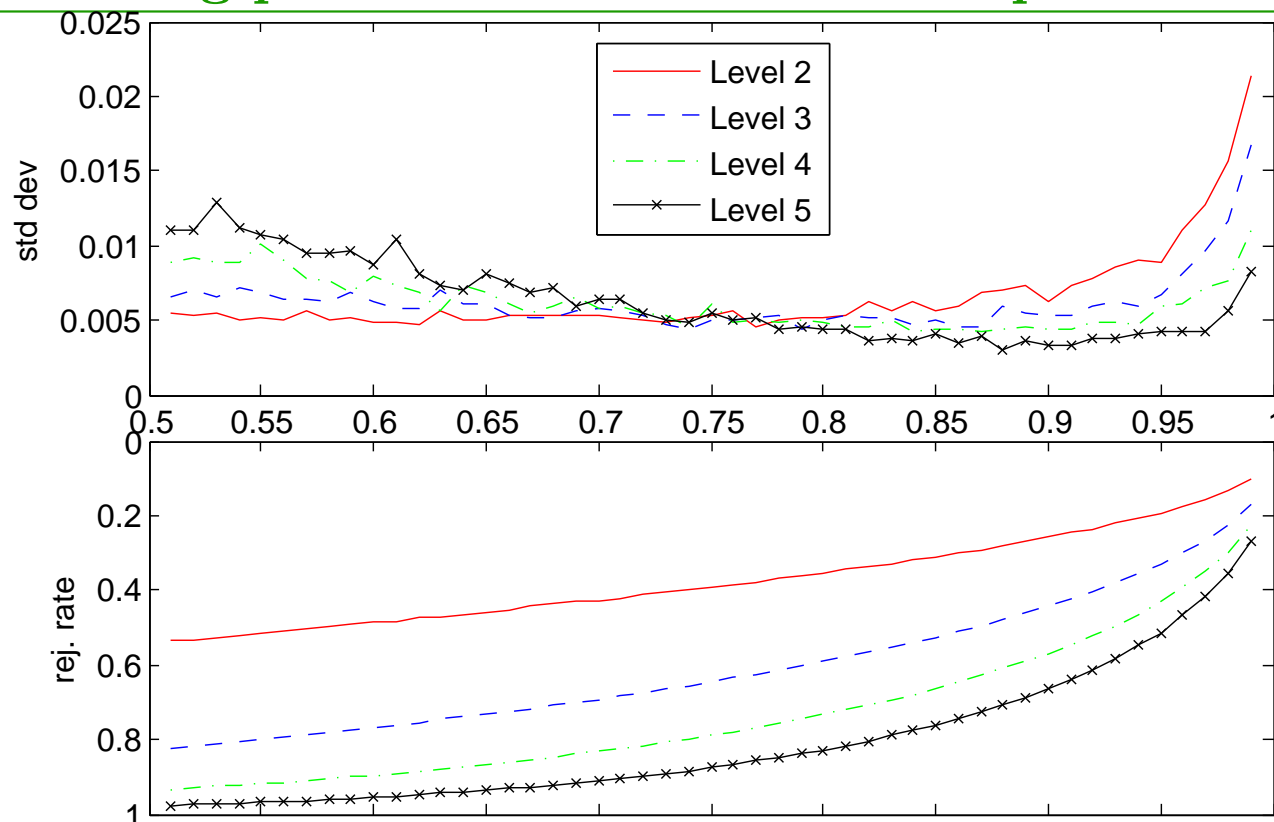


Figure 3: POP method, standard deviation (std. dev.) of each conditional probability estimator and corresponding rejection rate (rej. rate), based on 100 macro-runs, for different values of ρ .

▣▶ **Rule-of-thumb: minimal variances for POP correspond to acceptance rate $\approx 30 - 40\%$.**

Tuning the shaking parameters for the adaptive version

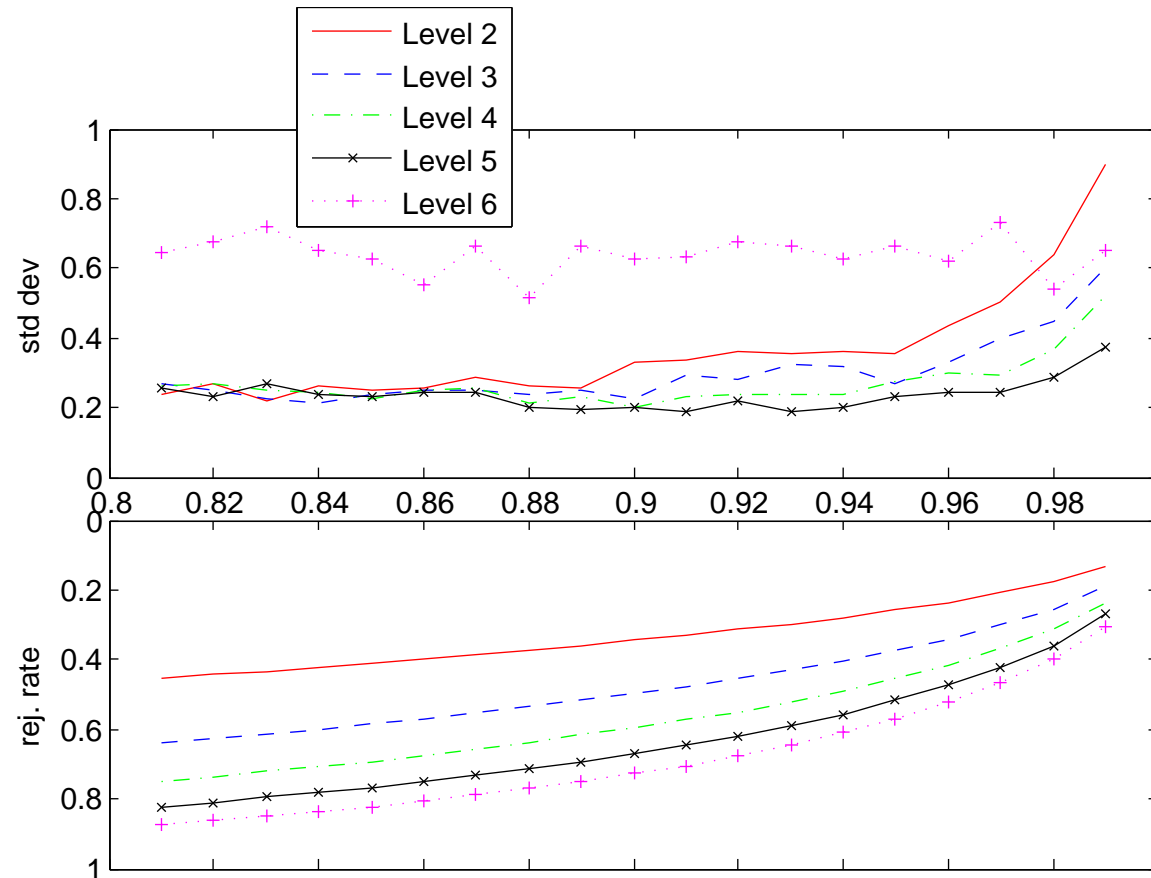


Figure 4: Adaptive POP method ($p = 10\%$). Standard deviation (std. dev.) and corresponding rejection rate (rej. rate), based on 100 macro-runs, of each quantile estimator and last level occupation measure estimator, for different values of ρ . The std. dev. of occupation measure (level 6) estimator has been scaled by 10 for easier comparison.

Extra shaking for IPS between selection steps

$\times 10^{-6}$	$J = 1$	$J = 2$	$J = 3$	$J = 4$	$J = 5$	$J = 6$	$J = 7$	$J = 8$	$J = 9$	$J = 10$	POP
$\rho = 0.9$	2.50	1.21	1.04	1.01	1.03	1.06	1.11	1.19	1.31	1.37	0.80
$\rho = 0.7$	1.74	1.40	1.29	1.25	1.20	1.29	1.28	1.33	1.37	1.43	1.02
$\rho = 0.5$	4.46	3.69	3.56	3.11	3.18	2.90	2.67	2.73	2.63	2.61	1.94

Table 3: Empirical standard deviation of IPS estimators of default probability for $L = 100$ and $B = 36$ based on 1000 algorithm macro-runs, with $M = 10^4$ and particle system size equal to $M' = \lfloor \frac{M}{J} \rfloor$. The last column is the empirical standard deviation of POP method using $n = M = 10^4$ iterations at each level.

Impact of time-discretization (curse of dimensionality of MCMC?)

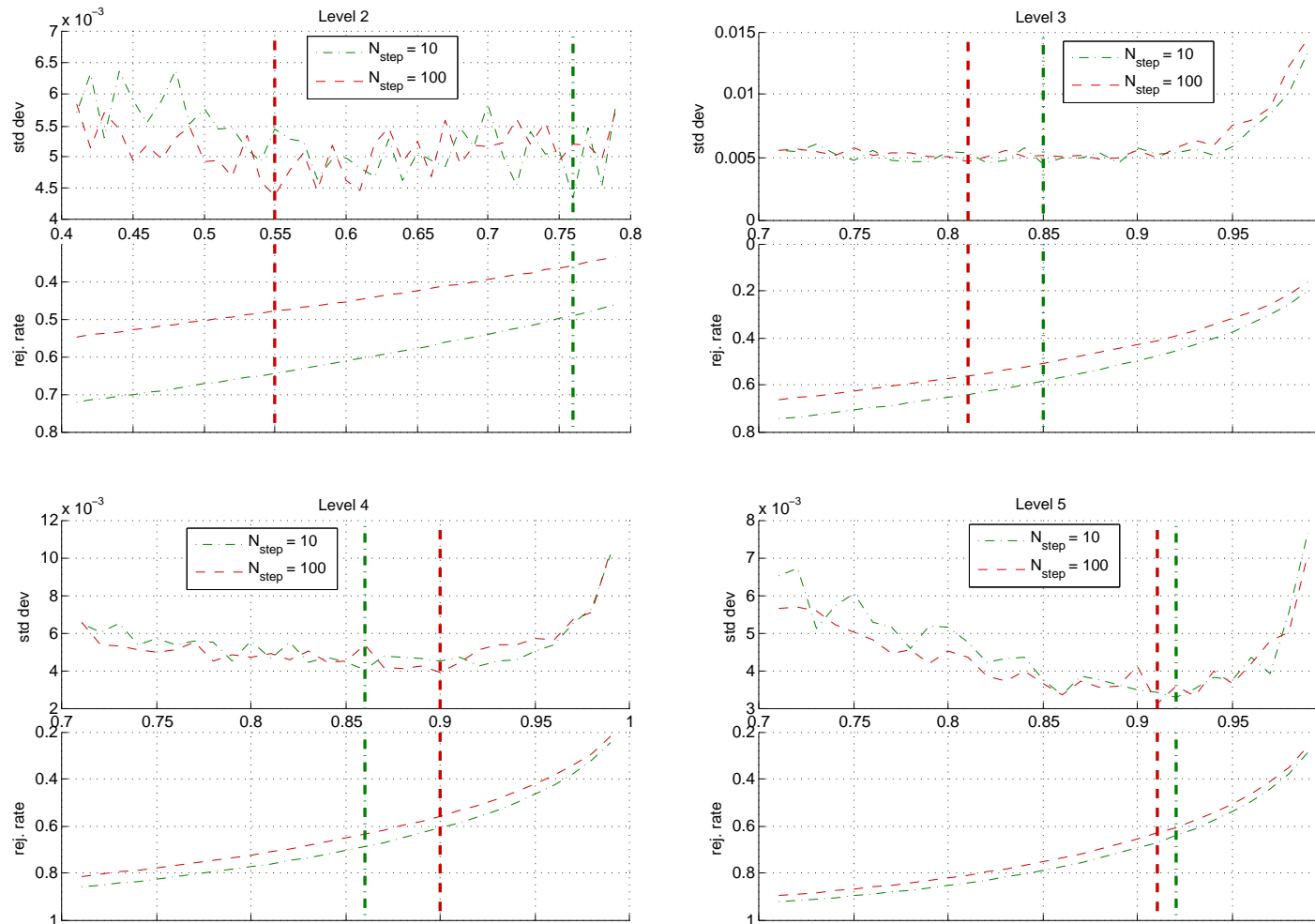
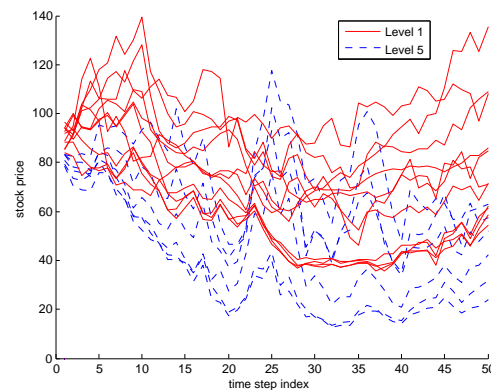
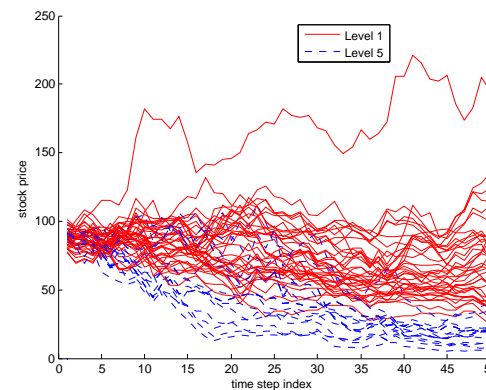


Figure 5: The impact of the number of discretization times (N_{step}) on the optimal shaking parameter (ρ) at each level of POP method.

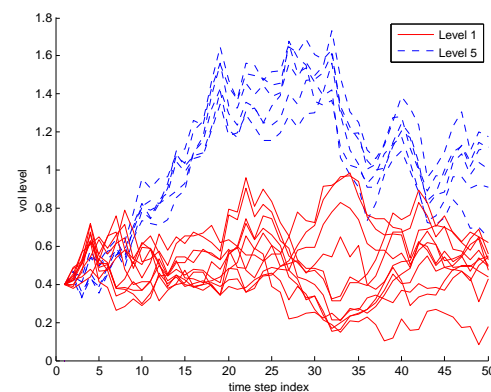
Generation of stress-tests



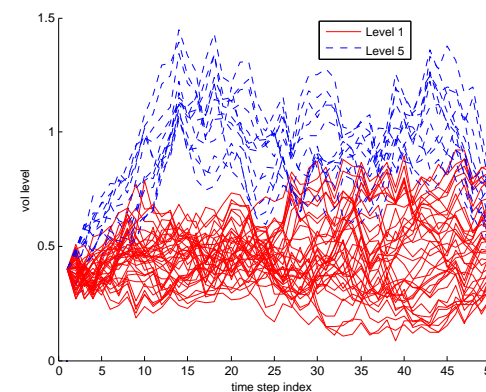
(a) $B = 18$



(b) $B = 36$



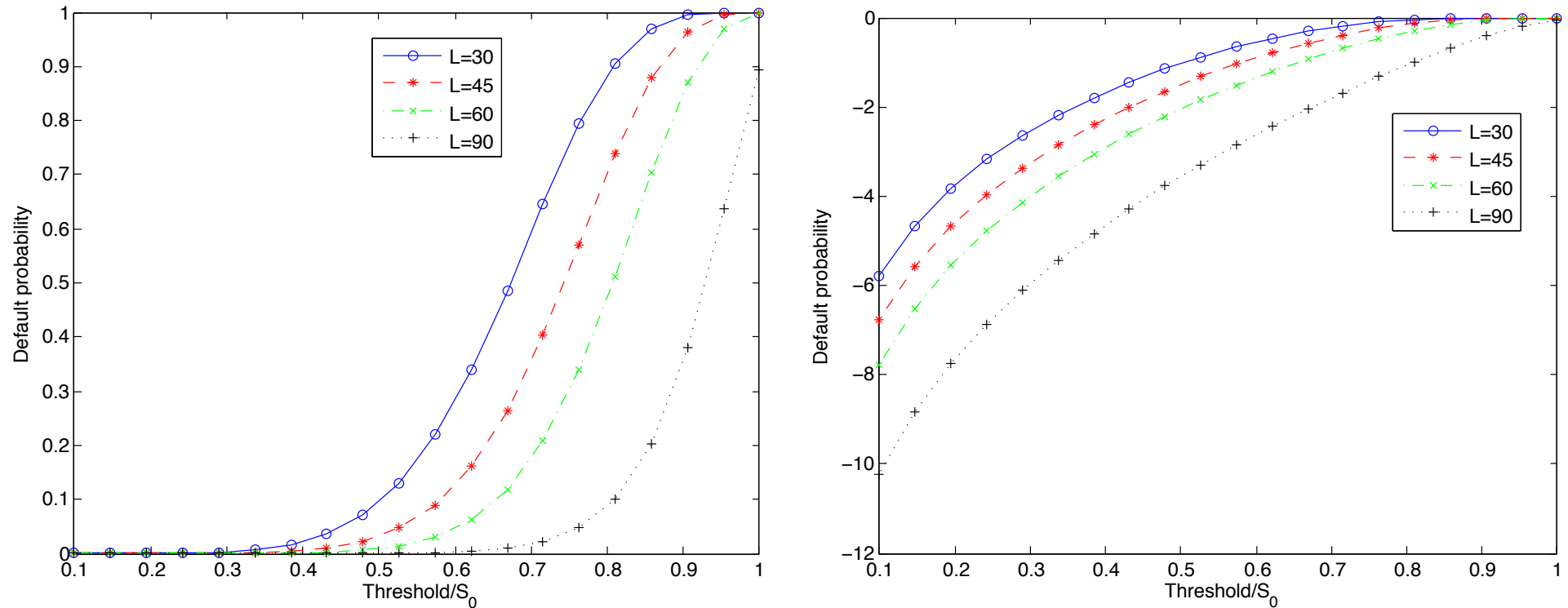
(c) $B = 18$



(d) $B = 36$

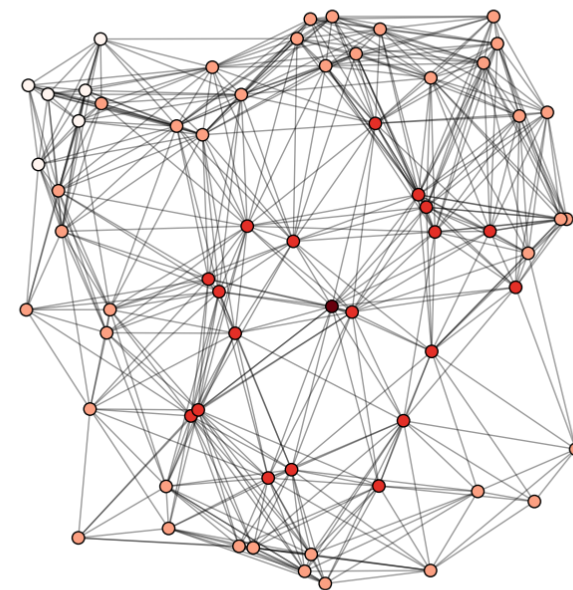
Figure 6: Sample paths for the asset price of firm 1 at Level 1 and Level 5 in the POP method and the respective volatility sample paths.

Rare event probability estimates on a large range of parameters



Plot (a) and log-plot (b) of default probabilities for varying B/X_0 .

3.3.4 Random graph



- ✓ **Erdos-Renyi model:** V vertices, vertex i is connected with vertex j with probability q .
- ✓ $\mathbb{P}(\text{Triangle}(\mathbf{X}) > \frac{V(V-1)(V-2)}{6} t^3)$ for $t > q$, with $V = 64, q = 0.35, t = 0.4$
- ✓ In [BHLN13], this probability is estimated as $2.19e - 06$ and $2.03e - 06$

- ✓ Shaker: 10% of all X_{ij} by P with $qP(1, 0) = (1 - q)P(0, 1)$
- ✓ Intermediate levels $L_i = \frac{V(V-1)(V-2)}{6} t^3 (\frac{i}{5})^{\frac{1}{5}}, i = 1, 2, 3, 4, 5$
- ✓ 50 macro MC runs

$P(1, 0)$	0.25	0.5	0.75
mean	1.79e-06	1.83e-06	1.92e-06
std	2.29e-06	1.30e-06	1.04e-06
std/mean	1.28	0.71	0.54

IPS: $M = 10^4$

$P(1, 0)$	0.25	0.5	0.75
mean	2.15e-06	2.05e-06	2.06e-06
std	5.76e-07	4.52e-07	3.23e-07
std/mean	0.27	0.22	0.16

POP: $N = 10^4$

- ✓ On-going applications to systemic risks.

3.3.5 Hawkes process (high-frequency modeling...)

- ✓ **Model:** self-exciting counting process $(N_t)_{t \geq 0}$ with stochastic intensity

$$d\lambda_t = \theta(\mu - \lambda_t)dt + dN_t.$$

Here $\theta = 2$, $\mu = 1$, the terminal time $T = 24$ and $\lambda_0 = 1$.

- ✓ Jump instants: $(\tau_j)_{j \geq 1}$. $\mathbf{H} = \max\{\tau_j - \tau_i : \tau_{k+1} - \tau_k < 0.5, i \leq k < j - 1\}$ = longest period during which all jump inter-arrivals are less than 0.5.
- ✓ Estimation of $\mathbb{P}(\mathbf{H} > 11)$?
- ✓ Using 3×10^8 crude MC simulations gives a 99%-CI = $[3.2469, 3.8064] \times 10^{-6}$.
- ✓ Intermediate sets as $\{H > L_k\}$ where $(L_k)_{k=1, \dots, 5} = [3.5, 5.5, 7.5, 9.5, 11]$

IPS	$p = 0.1$	$p = 0.3$	$p = 0.5$
mean	3.30e-06	5.19e-06	3.88e-06
std	2.84e-06	1.37e-05	1.60e-05
std/mean	0.86	2.64	4.12

IPS: $M = 10^4$

POP	$p = 0.1$	$p = 0.3$	$p = 0.5$
mean	3.33e-06	3.51e-06	2.69e-06
std	1.25e-06	2.92e-06	3.71e-06
std/mean	0.37	0.83	1.38

POP: $N = 10^4$

3.3.6 Ruins in insurance

- ✓ **Model:** Insurance company's capital reserve

$$R_t = x + ct - \sum_{k=1}^{N_t} Z_k.$$

Take $c = 1$, λ (intensity for N_t) = 0.005, $x = 100$, $T = 1$ and suppose Z_k 's are Gamma(2.5, 0.12).

- ✓ Estimation of $\mathbb{P}(\min_{0 \leq t \leq T} R_t < 0)$
- ✓ 10^5 MC simulations with IS gives CI = $[1.0419, 1.1881] \times 10^{-6}$
- ✓ Set $L_i = x(1 - (\frac{i}{5})^2)$, $i = 1, 2, 3, 4, 5$ and $A_i = \{\min_{0 \leq t \leq T} R_t < L_i\}$
- ✓ 100 macro MC runs

IPS: works when shaking only jumps (Gamma variables) or when shaking jumps and inter-arrival times.

$M = 10^4$, shaking only jumps

	$p = 0.1$	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.5$	$p = 0.6$
mean	1.25e-006	1.11e-006	1.01e-006	1.02e-006	1.15e-006	1.09e-006
std	2.82e-006	1.30e-006	6.46e-007	8.39e-007	5.15e-007	4.11e-007
std/mean	2.26	1.17	0.64	0.82	0.45	0.38

POP: works only when shaking jumps and inter-arrival times.

$N = 10^4$

	$p = 0.1$	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.5$	$p = 0.6$
mean	1.14e-06	1.11e-06	1.12e-06	1.05e-06	1.12e-06	9.29e-07
std	5.08e-07	4.44e-07	4.80e-07	6.74e-07	8.24e-07	9.52e-07
std/mean	0.45	0.40	0.43	0.64	0.74	1.02

3.3.7 Model risk and robustness

- ✓ Model with switching volatilities according to the MA [Guyon 2014]
- ✓ Two volatilities $\sigma_- < \sigma_+$
- ✓ Discrete-time dynamics (non-Markovian)

$$S_{i+1} = S_i \exp\left(-\frac{1}{2}\sigma_-^2 \Delta t + \sigma_- \Delta W_i\right), i = 0, 1, 2, 3,$$

$$S_{i+1} = \begin{cases} S_i \exp\left(-\frac{1}{2}\sigma_-^2 \Delta t + \sigma_- \Delta W_i\right) & \text{if } S_i \geq \frac{\sum_{k=i-4} S_k}{4} \\ S_i \exp\left(-\frac{1}{2}\sigma_+^2 \Delta t + \sigma_+ \Delta W_i\right) & \text{if } S_i < \frac{\sum_{k=i-4} S_k}{4} \end{cases} \quad \text{when } i \geq 4.$$

- ✓ Hedging of a Call using the BS hedge with volatility σ_- :

$$\mathbf{PL}_{\text{trader}} := \text{BS Call option price} + \sum_{j=0}^9 \delta_j (\mathbf{S}_{5(j+1)} - \mathbf{S}_{5j}) - (\mathbf{S}_T - \mathbf{K})_+.$$

- ✓ Probability of large losses due to model misspecification:

$$\mathbb{P}(\mathbf{PL}_{\text{trader}} \leq \mathbf{L})?$$

Model parameters: $S_0 = 10$, $\sigma_- = 0.2$, $\sigma_+ = 0.27$, $K = 10$ and $L = -2.4$.

MC parameters: $L_k := kL/5$, $k = 1, 2, 3, 4, 5$.

Crude MC (with 5×10^8 simulations): $\mathbb{P}(\mathbf{PL}_{\text{trader}} \leq \mathbf{L}) \in [2.93, 3.34] \times 10^{-6}$ with 99%-confidence.

	IPS			POP		
	mean ($\times 10^{-6}$)	std. ($\times 10^{-7}$)	std./mean	mean ($\times 10^{-6}$)	std. ($\times 10^{-7}$)	std./mean
$\rho = 0.9$	3.10	5.29	0.17	3.13	2.07	0.07
$\rho = 0.7$	3.23	13.3	0.41	3.11	3.98	0.13
$\rho = 0.5$	2.79	25.9	0.93	3.18	8.44	0.27

Table 4: Estimators of $\mathbb{P}(PL_{\text{trader}} \leq L)$ (mean) with empirical standard deviation (std.) for IPS and POP method based on 100 macro-runs.

Each intermediate level estimator in both the methods is based on $M = N = 10^5$ simulations.

3.3.8 Tails of implied volatility in fSABR model

Denote by $(B_t^{(H)})_{t \in \mathbb{R}}$ the fractional Brownian motion with Hurst exponent $H \in (0, 1)$: i.e. the \mathbb{R} -valued Gaussian process, centered with covariance function

$$\mathbb{E}(B_t^{(H)} B_s^{(H)}) = \frac{1}{2} (|t|^{2H} + |s|^{2H} - |t - s|^{2H}).$$

Mandelbrot and van Ness representation:

$$B_t^{(H)} = C_H \left[\int_{-\infty}^t [(t-s)^{H-\frac{1}{2}} - (-s)_+^{H-\frac{1}{2}}] dB_s, \right.$$

$$C_H = \sqrt{\frac{2H\Gamma(3/2-H)}{\Gamma(H+1/2)\Gamma(2-2H)}}.$$

Recent interests for fBM in modeling random volatility: [\[Comte-Renault 98\]](#)[\[Gatheral-Jaisson-Rosenbaum 14\]](#)...

fSABR model:
$$\begin{cases} \frac{dS_t}{S_t} = \sigma_t dZ_t \\ \sigma_t = \bar{\sigma} \exp\left(-\frac{\alpha^2}{2}t^{2H} + \alpha B_t^{(H)}\right) \end{cases}$$

with $d\langle B, Z \rangle_t = \rho_{BZ} dt$.

Parameter values: $S_0 = 40, \bar{\sigma} = 0.235, r = 0, T = 1.0$.

What is the impact of parameters H, α, ρ_{BZ} on prices? Which ones can we retrieve from calibration?

- ✓ Use information from the tails (ITM-OTM options, Implied Vol smile).
- ✓ Lee formula (left tail):

$$\beta_L := \limsup_{x \rightarrow -\infty} \frac{I^2(x)}{x/T}$$

where $I(x)$ is the BS implied volatility corresponding to the log-moneyness x and option maturity T

- ✓ If $\tilde{q} := \sup\{q \geq 0 : \mathbb{E}[S_T^{-q}] < \infty\}$, then

$$\tilde{q} = 1/2\beta_L + \beta_L/8 - 1/2.$$

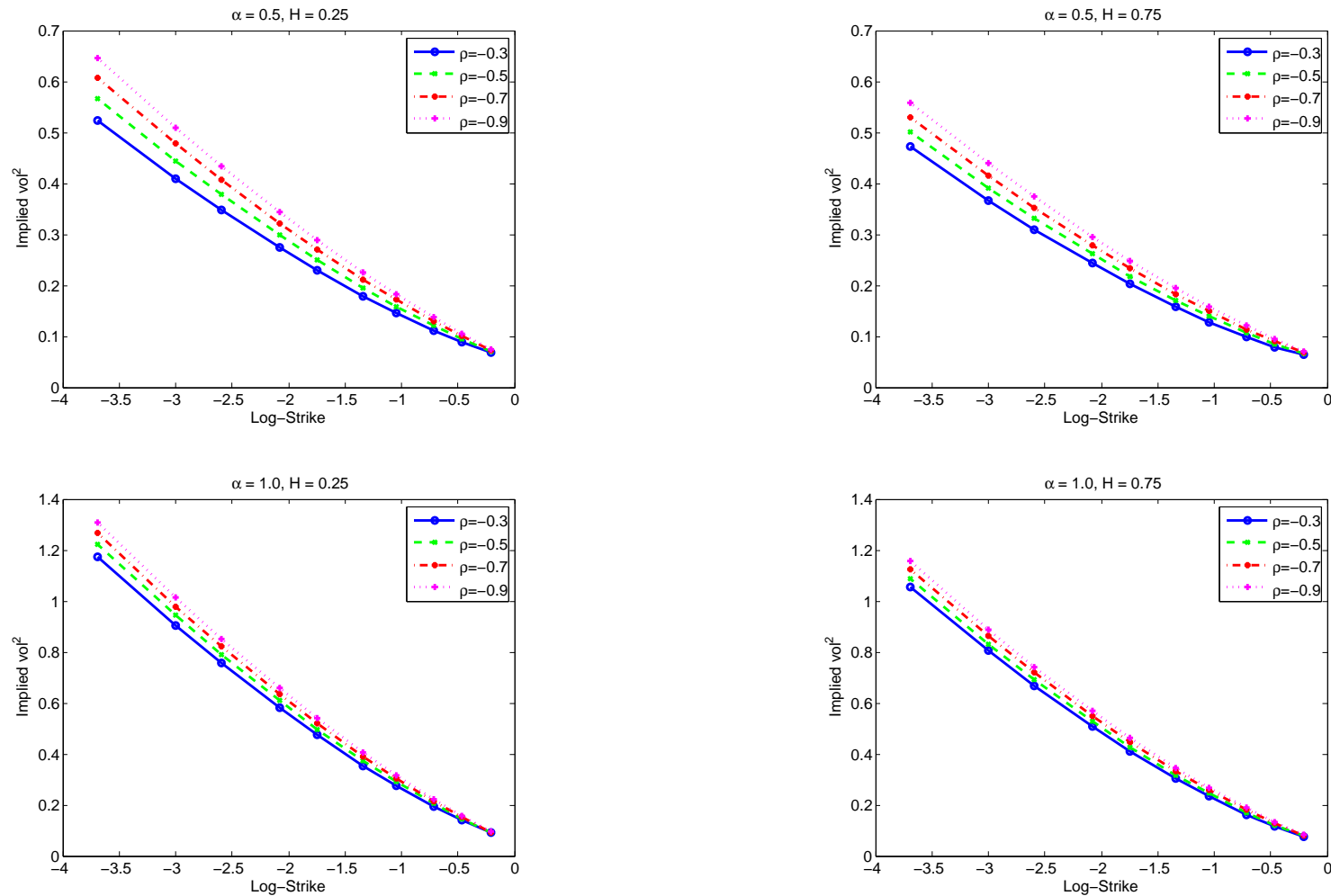


Figure 7: Plot of estimated squared implied volatility estimates w.r.t. log-strike in fSABR model, for different correlations ρ_{BZ} .

ρ_{BZ}	$\alpha = 0.5$				$\alpha = 1.0$			
	$H = 0.15$	$H = 0.25$	$H = 0.75$	$H = 0.9$	$H = 0.15$	$H = 0.25$	$H = 0.75$	$H = 0.9$
-0.3	2.6133	2.5515	2.8058	2.9753	0.8251	0.8267	0.9211	0.9632
-0.5	2.4222	2.3823	2.6733	2.8715	0.7905	0.7913	0.8950	0.9449
-0.7	2.2593	2.2042	2.5465	2.7918	0.7597	0.7591	0.8686	0.9277
-0.9	2.1235	2.0653	2.4339	2.6919	0.7325	0.7297	0.8449	0.9113

Table 5: *Estimates of maximum negative moment \tilde{q} in fSABR model using POP method.*

It suggests that

- ✓ calibration of α and ρ_{BZ} seems tractable
- ✓ calibration of H -index may be difficult from vanilla prices

3.3.9 Estimating sensitivities for OTM options

Insurance contracts for asset managers.

Model:

$$\frac{dS_t^i}{S_t^i} = \mu^i dt + \sigma^i d(LW_t)^i.$$

Aim: sensitivity of $\mathbb{E} [(K - p_1 S_T^1 - p_2 S_T^2)_+]$

Model parameters:

- ✓ $r = \mu^i = 0.01$
- ✓ $K = 100, T = 1,$
- ✓ Correlation matrix $C = LL^*$
- ✓ $\sigma^1 = 25\%, \sigma^2 = 22.5\%, C_{1,2} = 90\%$ (ruled out by [GT13, Theorem 1])
- ✓ $p_1 = 10, S_0^1 = 10, p_2 = 5, S_0^2 = 20$
- ✓ Rare-event regime: $\mathbb{P} [p_1 S_T^1 + p_2 S_T^2 \leq K] \approx 1.7 \times 10^{-3}.$

For extreme event, direct sensitivity is not relevant \rightsquigarrow better to use relative sensitivity, defined by $\frac{\partial_\theta \mathbb{E} [(K - p_1 S_T^1 - p_2 S_T^2)_+]}{\mathbb{E} [(K - p_1 S_T^1 - p_2 S_T^2)_+]}$.

We use the Malliavin calculus to get

$$\partial_\theta \mathbb{E} [(K - p_1 S_T^1 - p_2 S_T^2)_+] = \mathbb{E} [Z_\theta \mathbf{1}_{K - p_1 S_T^1 - p_2 S_T^2 \geq 0}]$$

so that, setting $A := \{K - p_1 S_T^1 - p_2 S_T^2 \geq 0\}$,

$$\frac{\partial_\theta \mathbb{E} [(K - p_1 S_T^1 - p_2 S_T^2)_+]}{\mathbb{E} [(K - p_1 S_T^1 - p_2 S_T^2)_+]} = \frac{\mathbb{E} [Z_\theta | A]}{\mathbb{E} [K - p_1 S_T^1 - p_2 S_T^2 | A]}.$$

 Can be evaluated directly by the Markov chain at the last level \rightsquigarrow very simple and efficient.

Explicit weights:

$$\checkmark \theta = p_i \rightsquigarrow Z_\theta = \frac{1}{p_i T} \Sigma^{-1} e^i \cdot W_T$$

$$\checkmark \theta = \sigma^i \rightsquigarrow Z_\theta = \Sigma^{-1} e^i \cdot W_T \left(-\sigma^i + \frac{1}{T} (LW_T)^i \right) - (L\Sigma^{-1})_{i,i}$$

$$\checkmark \theta = C_{i,j} \rightsquigarrow Z_\theta = \frac{1}{T} (L^{-1}W_T)^i (L^{-1}W_T)^j - (C^{-1})_{i,j}$$

Estimates of relative sensitivity w.r.t. different model parameters.

	Sensitivity w.r.t.		
	p_1	σ^1	$C_{1,2}$
POP method (10^6 paths) (mean/std)	-0.7155 (0.0046)	24.0078 (0.1760)	3.1058 (0.0253)
Finite difference (10^6 paths POP with CRN) (mean/std)	-0.7120 (0.0157)	23.9252 (0.4838)	3.0866 (0.1128)
Finite difference (10^9 paths) (99% conf. interval) for benchmark	(-0.7155, -0.7129)	(23.9285, 24.0108)	(3.0801, 3.0990)

3.3.10 Summary and perspectives

- ✓ Efficient framework to compute rare event probabilities and generate stress tests using reversible Markov chains (shakers) on path space
- ✓ IPS (average in space) \neq POP (average in time)
- ✓ In general, better performance of POP compared to IPS (variances are divided by about 10)
- ✓ Advantages of POP: low memory, parallel and independent computations
- ✓ Tuning the intermediate levels can be done adaptively
- ✓ In Gaussian setting, sensitivity analysis available (Malliavin calculus)
- ✓ The variance optimal ρ for POP depends on the level.
Rule of thumb: adjust ρ so that acceptance rate is $\approx 30 - 40\%$.

3.4 MCMC/Regression scheme for $\mathbb{E}[f(Y, \mathbb{E}[R | Y]) | Y \in \mathcal{A}]$

Example. R = future cashflow at time 2; $\mathbb{E}[R | Y]$ = price at time 1;

Evaluation in adverse situations \mathcal{A} .

3.4.1 Algorithm

```

1 /* Simulation of the design and the observations */
2  $O^{(0)} \sim \xi$ , where  $\xi$  is a distribution on  $\mathcal{A} \subset \mathbb{R}^d$  ;
3 for  $m = 1$  to  $M$  do
4    $O^{(m)} \sim \mathbf{P}(O^{(m-1)}, dx)$  (apply the Shaker with rejection);
5    $R^{(m)} \sim \mathbf{Q}(O^{(m)}, dr)$ ;
6 /* Least-Squares regression */
7 Choose  $\hat{\alpha}_M \in \mathbb{R}^L$  solving  $\arg \min_{\alpha \in \mathbb{R}^L} \frac{1}{M} \sum_{m=1}^M \left| R^{(m)} - \langle \alpha; \underline{\phi}(O^{(m)}) \rangle \right|^2$  and set
    $\hat{\phi}_M(x) := \langle \hat{\alpha}_M; \underline{\phi}(x) \rangle$  ;
8 /* Final estimator using ergodic average */
9 Return  $\hat{\mathcal{I}}_M := \frac{1}{M} \sum_{m=1}^M f(O^{(m)}, \hat{\phi}_M(O^{(m)}))$ .

```

3.4.2 Convergence results about regression

Notations:

- ✓ Let $\mu \, d\lambda$ be the distribution of $Y | Y \in \mathcal{A}$, λ =positive σ -finite measure
- ✓ Let $L_2(\mu)$ be the set of measurable functions $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ such that

$$|\varphi|_{L_2(\mu)} := \left(\int \varphi^2 \mu \, d\lambda \right)^{1/2} < +\infty$$
- ✓ Let $\psi_\star = \arg \inf_{\varphi \in \text{Span}(\phi_1, \dots, \phi_L)} |\mathfrak{M} - \varphi|_{L_2(\mu)}$ be the projection of \mathfrak{M} on the basis functions (linear approximation space)

Theorem (Non asymptotic error estimates on the regression function).

Assume that

- (i) the transition kernel P and the initial distribution ξ satisfy: there exists a constant C_P and a rate sequence $\{\rho(m), m \geq 1\}$ such that for any $m \geq 1$,

$$\left| \xi P^m [(\psi_\star - \mathfrak{M})^2] - \int (\psi_\star - \mathfrak{M})^2 \mu \, d\lambda \right| \leq C_P \rho(m).$$

(ii) the conditional distribution \mathbb{Q} satisfies

$$\sigma^2 := \sup_{x \in \mathcal{A}} \left\{ \int r^2 \mathbb{Q}(x, dr) - \left(\int r \mathbb{Q}(x, dr) \right)^2 \right\} < \infty.$$

Let $O^{(1:M)}$ and $\hat{\phi}_M$ be given by the previous Algorithm. Then,

$$\begin{aligned} \Delta_M &:= \mathbb{E} \left[\frac{1}{M} \sum_{m=1}^M \left(\hat{\phi}_M(\mathbf{O}^{(m)}) - \mathfrak{M}(\mathbf{O}^{(m)}) \right)^2 \right] \\ &\leq \frac{\sigma^2 L}{M} + |\psi_\star - \mathfrak{M}|_{L_2(\mu)}^2 + \frac{C_P}{M} \sum_{m=1}^M \rho(m). \end{aligned}$$

Remarks.

- ✓ Finite dimensional Gaussian shaker $\implies \rho(m) = O(r^m)$ with $r \in (0, 1)$. See slide 54.
- ✓ Usually $\sum_{m=1} \rho(m) < +\infty$
- ✓ The use of MCMC design does not impact significantly the statistical error. Compare with slide 47

Corollary (Non asymptotic error estimates on the outer expectation $\mathbb{E}[f(Y, \mathbb{E}[R | Y]) | Y \in \mathcal{A}]$). Assume

(i) $f : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}$ is globally Lipschitz in the second variable

$$|f(y, r_1) - f(y, r_2)| \leq C_f |r_1 - r_2|.$$

(ii) There exists a finite constant C such that for any M

$$\mathbb{E} \left[\left(M^{-1} \sum_{m=1}^M f \left(O^{(m)}, \mathfrak{M}(O^{(m)}) \right) - \int f(x, \mathfrak{M}(x)) \mu(x) d\lambda(x) \right)^2 \right] \leq \frac{C}{M}.$$

Then

$$\begin{aligned} & \left(\mathbb{E} \left[\left| \frac{1}{M} \sum_{m=1}^M f(\mathbf{O}^{(m)}, \hat{\phi}_M(\mathbf{O}^{(m)})) - \mathbb{E}[f(\mathbf{Y}, \mathbb{E}[\mathbf{R} | \mathbf{Y}]) | \mathbf{Y} \in \mathcal{A}] \right|^2 \right] \right)^{1/2} \\ & \leq C_f \sqrt{\Delta_M} + \sqrt{\frac{C}{M}}. \end{aligned}$$

3.4.3 Numerical examples

Goal: to approximate

$$\mathcal{I} := \mathbb{E} \left[\left(\mathbb{E} \left[(\mathbf{K} - \mathbf{h}(\mathbf{S}_{T'}))_+ \mid \mathbf{S}_T \right] - \mathbf{p}_* \right)_+ \mid \mathbf{S}_T \in \mathcal{S} \right]$$

for various choices of h , where $\{S_t, t \geq 0\}$ is a d -dimensional geometric Brownian motion, $T < T'$ and $\{S_T \in \mathcal{S}\}$ is a rare event.

▷ **A toy example in dimension 1.** Here $h(y) = y$ and $\mathcal{S} = \{s \in \mathbb{R}_+ : s \leq s_*\}$ so that

$$\mathcal{I} = \mathbb{E} \left[\left(\mathbb{E} \left[(\mathbf{K} - \mathbf{S}_{T'})_+ \mid \mathbf{S}_T \right] - \mathbf{p}_* \right)_+ \mid \mathbf{S}_T \leq \mathbf{s}_* \right].$$

T	T'	S_0	K	σ	s_*	p_*
1	2	100	100	30%	30	10

Here $\mathbb{P}(Y \in \mathcal{A}) \approx 5.6e-5$

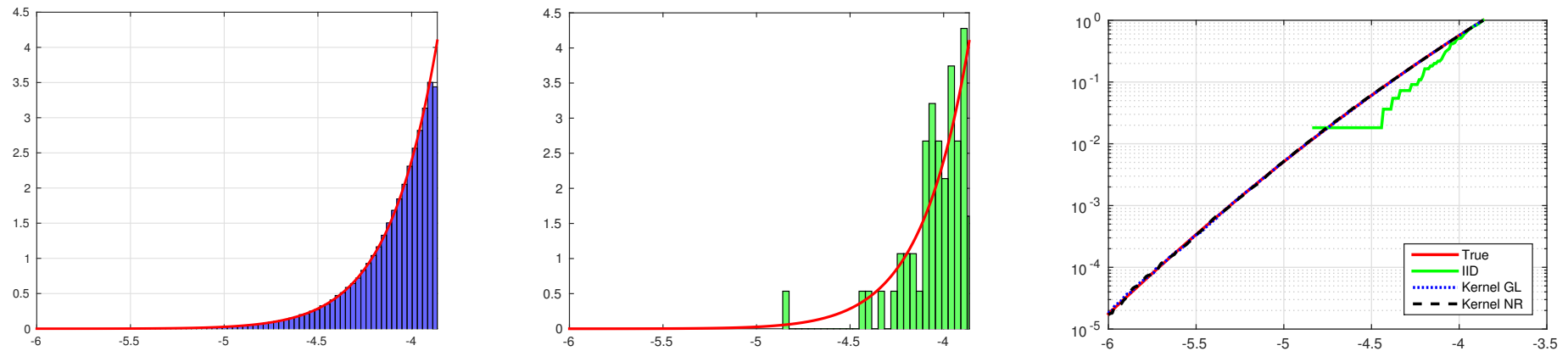


Figure 8: Normalized histograms of the $M = 1e6$ points from the Markov chain (left), from the i.i.d. sampler with rejection (middle). (right) Restricted to $[-6, y_\star]$, the cdf of Y given $\{Y \in \mathcal{A}\}$ with MCMC/crude MC estimates.

😊 MCMC sampler gives very accurate sampling of the tails

As before, choosing the shaking parameter is important!

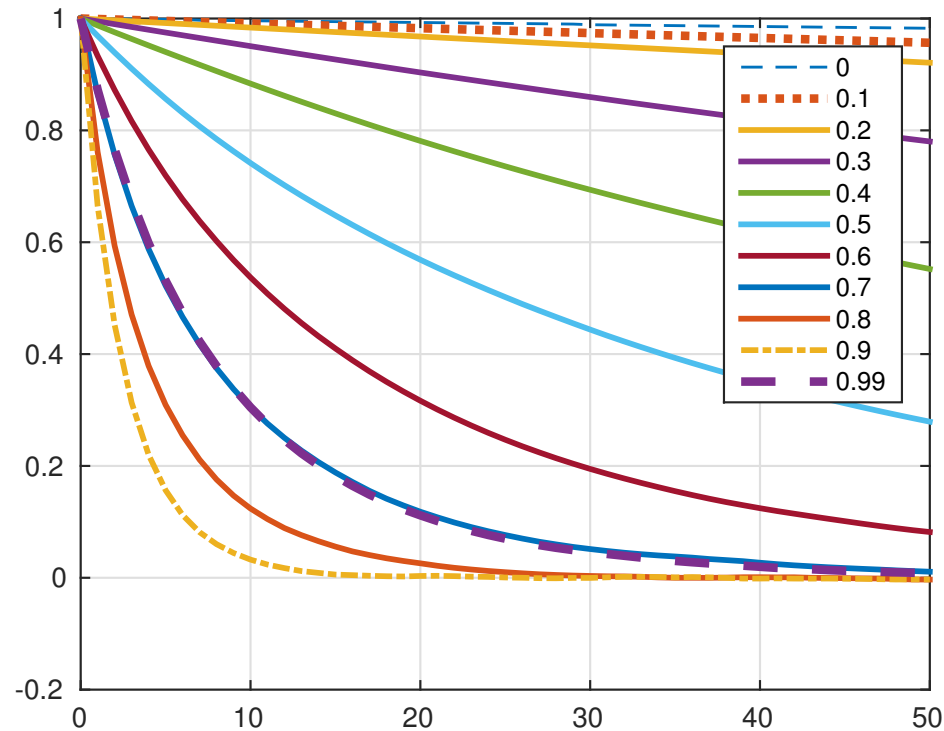


Figure 9: For different values of ρ , estimation of the autocorrelation function (over 100 independent runs) of the Markov chain. Each curve is computed using $1e6$ sampled points.

► Tune the shaking parameter according to the acceptance rate

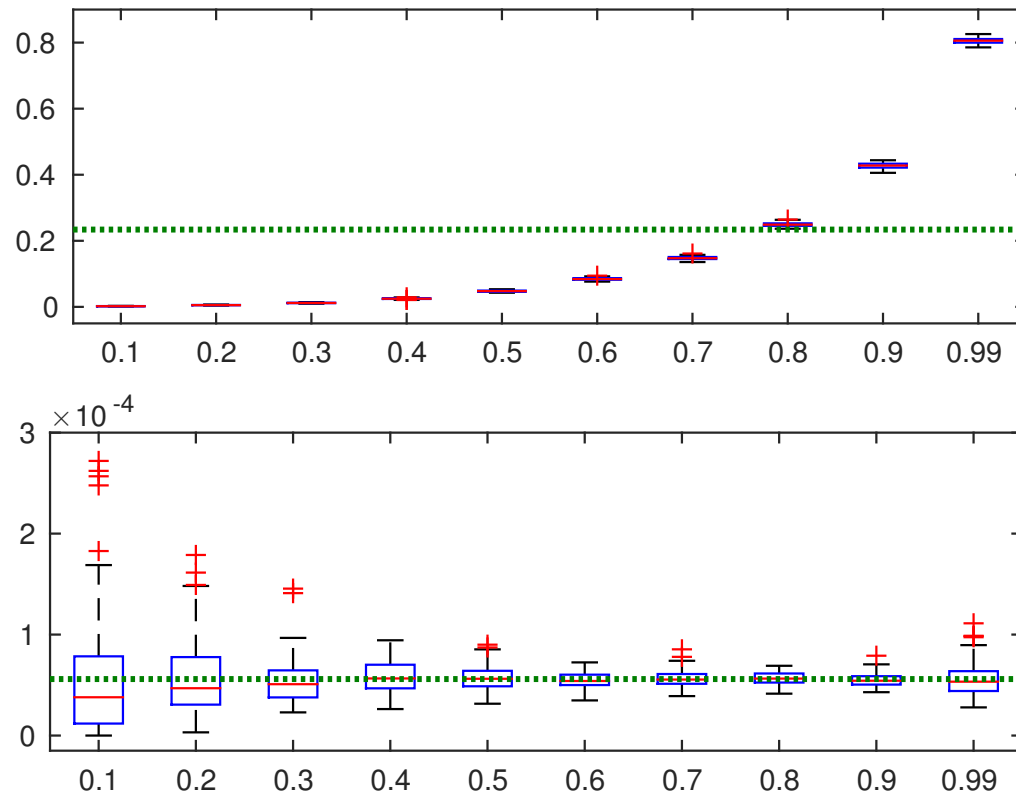


Figure 10: (top) Mean acceptance rate after M chain iterations of the chain (green = 23.4% [Rosenthal 2008]) (bottom) Estimation of $\mathbb{P}(Y \in \mathcal{A})$ by combining splitting and MCMC

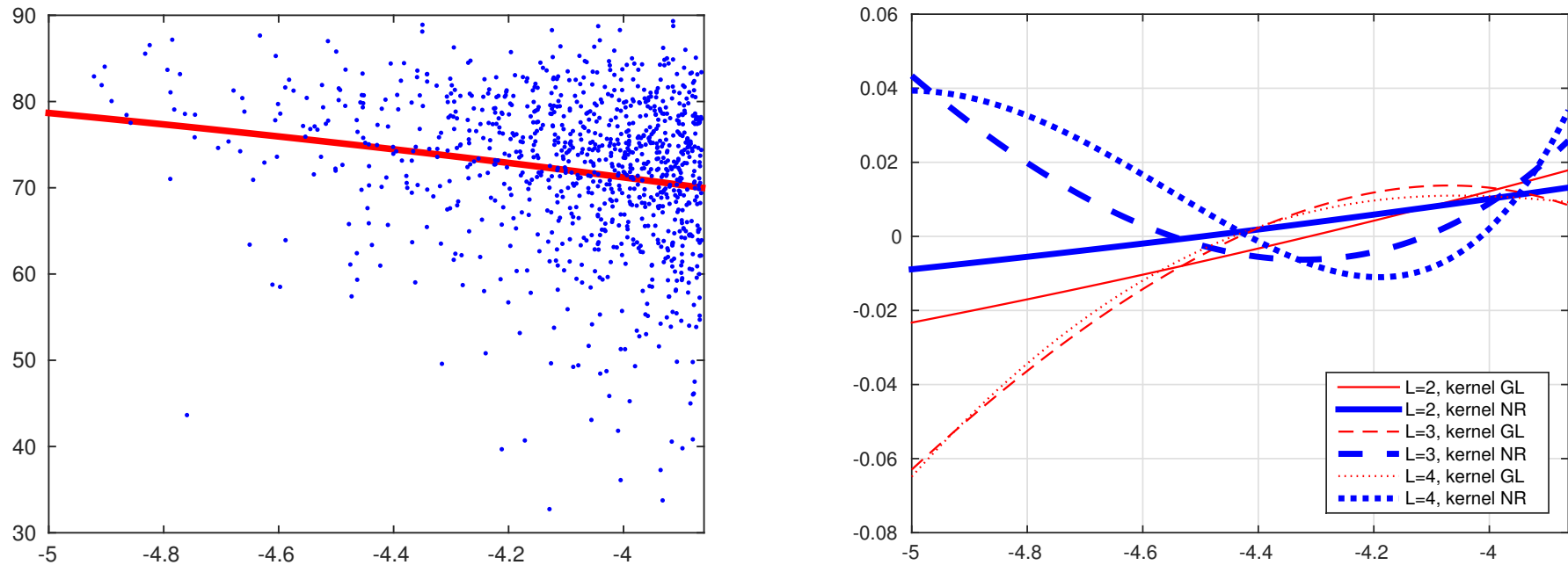
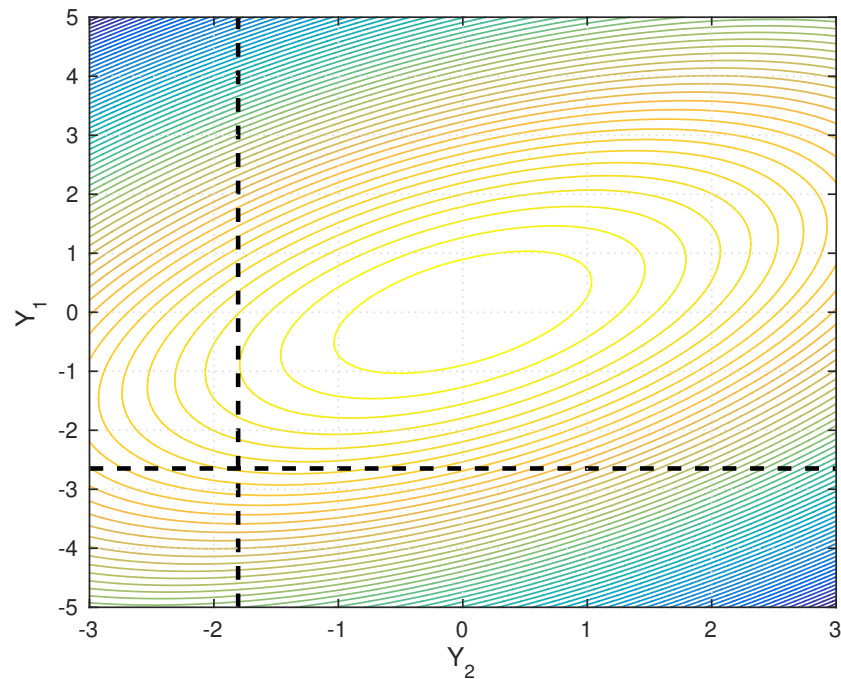


Figure 11: (left) 1000 sampled points $(O^{(m)}, R^{(m)})$ (using the MCMC sampler), together with \mathfrak{M} ; (right) A realization of the error function $x \mapsto \hat{\phi}_M(x) - \mathfrak{M}(x)$ on $[-5, y_\star]$, for different values of $L \in \{2, 3, 4\}$ and two different kernels when sampling $O^{(1:M)}$.

▷ **Correlated geometric Brownian motions in dim. 2.** We consider

$$\mathcal{I} := \mathbb{E} \left[\left(\mathbb{E} \left[\left(\mathbf{K} - \sqrt{\mathbf{S}_{1,T'} \mathbf{S}_{2,T'}} \right)_+ \mid \mathbf{S}_T \right] - \mathbf{p}_\star \right)_+ \mid \mathbf{S}_T \in \mathcal{S} \right]$$

with $\mathcal{S} = \{(s_1, s_2) \in \mathbb{R}_+ \times \mathbb{R}_+ : s_1 \leq s_\star, s_2 \leq s_\star\}$.



Level curves of the density function of $(S_{1,T}, S_{2,T})$ and the rare set in the lower left area delimited by the two hyperplanes.

For the basis functions, we take

$$\begin{aligned} \varphi_1(x) &= 1, & \varphi_2(x) &= \sqrt{x_1}, & \varphi_3(x) &= \sqrt{x_2}, \\ \varphi_4(x) &= x_1, & \varphi_5(x) &= x_2, & \varphi_6(x) &= \sqrt{x_1 x_2}. \end{aligned}$$

T	T'	$S_{0,1}$	$S_{0,2}$	K	σ_1	σ_2	ρ	s_\star	p_\star
1	2	100	100	100	25%	35%	50%	50	5

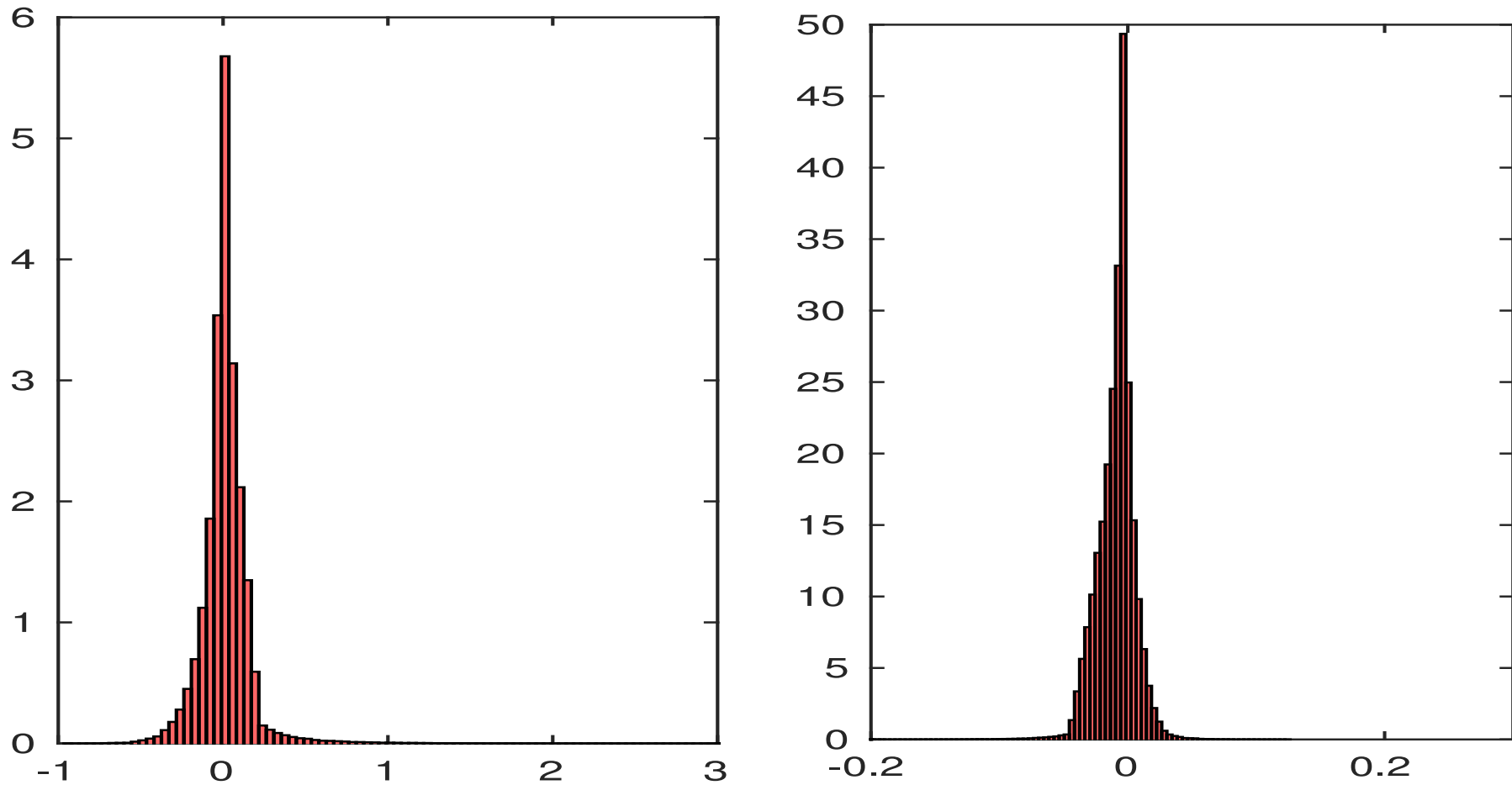


Figure 12: (left) Normalized histograms of the error $\{\hat{\phi}_M(O^{(m)}) - \mathfrak{M}(O^{(m)}), m = 1, \dots, M\}$, when $L = 3$. (right): the same case with $L = 6$.

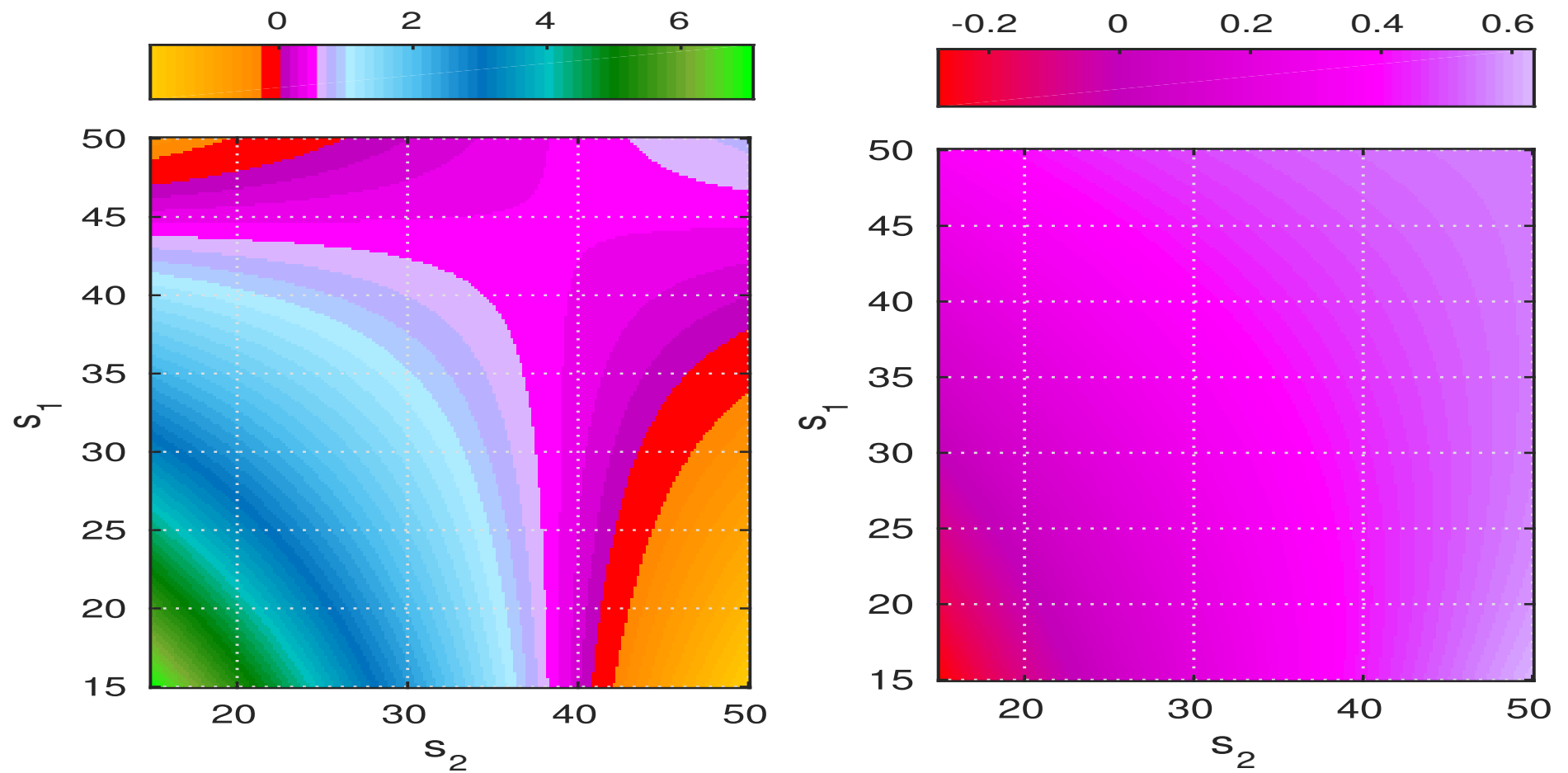


Figure 13: (left) Error function with $L = 3$. (right) The same case with $L = 6$.

3.4.4 Conclusion

- ✓ Design of a regression method suitable for design points in an extreme set \mathcal{A}
- ✓ Extreme design points are sampled using MCMC
- ✓ Error estimates available as if the set \mathcal{A} were not rare

On-going work: Rare event set \mathcal{A} depending itself on conditional expectations (on cashflow value) \Rightarrow adaptive MCMC-regression scheme

4 Stochastic algorithms, VaR and nested risks

Able to handle the case where the extreme set is not explicitly given (but given implicitly by some conditional expectations)

4.1 Usual Robbins-Monro algorithm

Seeking the zeros of a function $h : \mathbb{R}^d \mapsto \mathbb{R}^d$, i.e.

$$T^* = \{\theta \in \mathbb{R}^d : h(\theta) = 0\}$$

where h has the form

$$h(\theta) = \mathbb{E}(H(\theta, X)),$$

for $X : \Omega \rightarrow \mathbb{R}^q$, a r.v. that can be sampled, and $H : \mathbb{R}^d \times \mathbb{R}^q \rightarrow \mathbb{R}^d$.

Example ($\text{VaR}^\alpha(X)$). Assume that $X \in \mathbb{R}$ has a continuous c.d.f. Then the zeros of

$$h(q) = \mathbb{E}[\mathbf{1}_{X \leq q} - \alpha]$$

give the α -quantile of X . VaR is the inf of such zeros.

Definition (Stochastic approximation). Define

$$\theta_{n+1} = \theta_n - \gamma_{n+1} H(\theta_n, X_{n+1}),$$

for i.i.d. X_n and where $(\gamma_n)_{n \in \mathbb{N}}$ are the steps of the algorithm such that

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

Example. Take $\gamma_n = \frac{c}{n^\alpha}$ with $\alpha \in (\frac{1}{2}, 1]$.

Theorem (Convergence). Let $h \in C^0$, $T^* = h^{-1}(\{0\})$ and assume that for any $\theta^* \in T^*$ and $\theta \notin T^*$ we have

$$\langle \theta - \theta^*, h(\theta) \rangle > 0, \quad (\text{separation condition}).$$

If $\mathbb{E}(|H(\theta, X)|^2) \leq C_H(1 + |\theta|^2)$ for a certain constant C_H , there exists a r.v. $\theta_\infty : \Omega \rightarrow T^*$ such that $\theta_n \xrightarrow{a.s.} \theta_\infty$.

Furthermore, the convergence holds also in L^p with $p \in (0, 2)$.

Under some extra conditions, convergence rate is equal to $1/\sqrt{\gamma_n}$.

Further remarks

- ✓ The conditions for convergence are fulfilled for the α -quantile. We could compute the $\mathbb{C}\text{VaR}$ along the same sequence since it does not depend on the specific representant of the α -quantile. See slide 7.
- ✓ 😊 SA is a sequential algorithm that can be stopped at any time at the current result.
- ✓ The more we simulate, the more accurate the result.
- ✓ We can incorporate a IS scheme in the SA for $\mathbb{V}\text{aR}$, [BFP09].
- ✓ Alternative scheme for $\mathbb{V}\text{aR}$: sample and sort. Efficient for non-nested computations. Not clear for nested expectations.

Proof (of convergence). Let $\theta^* \in T^*$. Write

$$|\theta_{n+1} - \theta^*|^2 = |\theta_n - \theta^*|^2 + \gamma_{n+1}^2 |H(\theta_n, X_{n+1})|^2 - 2\gamma_{n+1} \langle \theta_n - \theta^*, H(\theta_n, X_{n+1}) \rangle.$$

Step 1. This shows that $\theta_n \in L^2, \forall n$.

Step 2. Set $\Delta M_n = H(\theta_n, X_{n+1}) - h(\theta_n)$ where $h(\theta_n) = \mathbb{E}(H(\theta_n, X_{n+1}) | \mathcal{F}_n)$, then

$$\begin{aligned} |\theta_{n+1} - \theta^*|^2 &= |\theta_n - \theta^*|^2 + \gamma_{n+1}^2 |H(\theta_n, X_{n+1})|^2 - 2\gamma_{n+1} \langle \theta_n - \theta^*, \Delta M_n \rangle \\ &\quad - 2\gamma_{n+1} \langle \theta_n - \theta^*, h(\theta_n) \rangle. \end{aligned}$$

Then, for any $c \geq 0$, define

$$S_n := \frac{|\theta_n - \theta^*|^2 + \sum_{k=0}^{n-1} 2\gamma_{k+1} \langle \theta_k - \theta^*, h(\theta_k) \rangle + c \sum_{k \geq n+1} \gamma_k^2}{\prod_{k=1}^n (1 + c\gamma_k^2)} \geq 0.$$

For a constant $c \geq 0$ large enough, we have

$$\mathbb{E}(S_{n+1} | \mathcal{F}_n) \leq S_n.$$

▮ S_n is surmartingale (and positive).

Recall

$$S_n := \frac{|\theta_n - \theta^*|^2 + \sum_{k=0}^{n-1} 2\gamma_{k+1} \langle \theta_k - \theta^*, h(\theta_k) \rangle + c \sum_{k \geq n+1} \gamma_k^2}{\prod_{k=1}^n (1 + c\gamma_k^2)}.$$

Step 3. Therefore, for some integrable $S_\infty \geq 0$, $S_n \xrightarrow{a.s.} S_\infty$. Thus,

- ✓ $|\theta_n - \theta^*|^2 \xrightarrow{a.s.}$ and consequently, $\sup_n |\theta_n - \theta^*|^2$ is a.s. finite
- ✓ $\sum_{k \geq 1} \gamma_{k+1} \langle \theta_k - \theta^*, h(\theta_k) \rangle < +\infty$.

Step 4. Define

$$l = \liminf_{k \rightarrow \infty} \langle \theta_k - \theta^*, h(\theta_k) \rangle \geq 0.$$

We must have $l = 0$, otherwise there is a contradiction with the convergence.

Step 5: $T^* = \{\theta^*\}$. Up to an subsequence extraction (θ_k bounded), we show that $\theta_{\sigma(k)} \xrightarrow{a.s.} \theta^*$ and finally the whole sequence $\theta_k \xrightarrow{a.s.} \theta^*$.

Step 6: T^* contains several points. Similar arguments using a dense subset of $T^* \dots$ □

4.2 SA (Robbins-Monro) algorithm with nested simulations

From [Barrera, Crépey, Diallo, Fort, G', Stazhinsky, 2018] for XVA computations.

We aim at computing the VaR of losses of the form

$$L := \Phi + D \mathbb{E}_1 [W] - \mathbb{E}_0 [W'],$$

where

- ✓ Z =observable financial/economic variables serving to project values on,
- ✓ the second/third terms of L are the future and present liability valuations,
 - ▶ $(W \mid Z_1 = z_1) \stackrel{d}{=} \Pi(z_1, \cdot)$
 - ▶ $(W' \mid Z_0 = z) \stackrel{d}{=} \Pi'(z, \cdot)$
- ✓ the first term Φ is the realized loss on the interval $[0, 1]$,
- ✓ D is the (random) discount factor.

Example. Euro Median Term Note (EMTN), issued by a bank, with a performance linked to the Euribor rate denoted by Z . Then

$$L := \Phi + D \mathbb{E}_1 [W] - \mathbb{E}_0 [W'],$$


corresponds to

$$\checkmark W' = \varphi(Z_1) \mathbf{1}_{\mathcal{D} \geq 1} + D\varphi(Z_2) \mathbf{1}_{\mathcal{D} \geq 2},$$

$$\checkmark W = \varphi(Z_2) \mathbf{1}_{\mathcal{D} \geq 2},$$

$$\checkmark \Phi = \varphi(Z_1) \mathbf{1}_{\mathcal{D} \geq 1}.$$

Algorithm principle (nested simulations in SA algorithm) Instead of sampling exactly $L := \Phi + D \mathbb{E}_1 [W] - \mathbb{E}_0 [W']$, we sample approximately using a MC average of $\mathbb{E}_1 [W]$ for each Z_1 and a MC average for $\mathbb{E}_0 [W']$.

 We may not compute anymore the VaR of L , except if the inner MC contains more and more samples. Interaction with the SA step γ_k ?

```

1 Input: A positive sequence  $\{\gamma_k, k \geq 1\}$ ,  $\mathbb{N}^*$ -valued sequences
    $\{M_k, M'_k, k \geq 1\}$ ,  $\xi_0 \in \mathbb{R}$ ,  $\chi_0 = 0$  and  $z \in \mathbb{R}^q$ .
2 for  $k = 1$  to  $K$ , do
3     /* Sampling step */
4     Sample  $(\Phi^k, D^k, Z_1^k)$  with the same distribution as  $(\Phi, D, Z_1)$ 
       conditionally to  $Z_0 = z$ , independently from the past draws ;
5     Sample  $M'_k$  independent copies  $\{W'_m, 1 \leq m \leq M'_k\}$  with the
       distribution  $\Pi'(z, \cdot)$ ;
6     Given  $Z_1^k$ , sample  $M_k$  independent copies  $\{W_{m,k}, 1 \leq m \leq M_k\}$  with the
       distribution  $\Pi(Z_1^k, \cdot)$ ;
7     Compute  $L_k := \Phi^k + D^k \frac{1}{M_k} \sum_{m=1}^{M_k} W_{m,k} - \frac{1}{M'_k} \sum_{m=1}^{M'_k} W'_m$  ;
8     /* Update the VaR estimates */
9      $\xi_k = \xi_{k-1} - \gamma_k (\mathbf{1}_{L_k \leq \xi_{k-1}} - \alpha)$  ;
10 Return(The sequences  $\{\xi_k, 1 \leq k \leq K\}$ )

```

Algorithm 1: Estimate of $\text{VaR}^\alpha(L)$

4.2.1 Convergence

Assume

- i) The distribution of L has a bounded density and D is bounded.
- ii) For some $p_\star \geq 2$, W and W' have finite polynomial moments of order p_\star .
- iii) There exists $\kappa \in (0, 1]$ such that

$$\sum_k \gamma_k = +\infty, \quad \sum_k \gamma_k^{1+\kappa} < +\infty,$$

$$\sum_{k \geq 1} \gamma_k^{1-\kappa} (M_k \wedge M'_k)^{-p_\star/(1+p_\star)} < +\infty, \quad \sum_{k \geq 1} \gamma_k (M_k \wedge M'_k)^{-1/2} < +\infty.$$

Flexibility in the trade-off between $\gamma_k \rightarrow 0$ and $M_k, M'_k \rightarrow +\infty$.

Theorem. Assume the previous assumptions. Then there exists a bounded random variable ξ_∞ such that

$$\xi_k \xrightarrow{a.s.} \xi_\infty.$$

and ξ_∞ is a α -quantile of $L = \Phi + D \mathbb{E}_1 [W] - \mathbb{E}_0 [W']$.

See [Barrera, Crépey, Diallo, Fort, G', Stazhinsky, 2018] for numerical results.

4.2.2 Conclusion, perspectives

- ✓ Inner MC into SA for computing VaR in future valuations
- ✓ Sequential procedure \neq batch procedure of *sample-sort scheme*
- ✓ Convergence rates are available
- ✓ What about boosting the convergence by IS or POP sampling?
- ✓ Other topics: model mis-specification (model risk)
 - ▶ quite important problem in the tails
 - ▶ sensitivity analysis may be not enough
 - ▶ better with Uncertainty Quantification. Some first results in [Crépey, Fort, G', Stazhinsky, 2017]
<https://hal.archives-ouvertes.fr/hal-01629952/>

References

- [AB09] M. Anthony and P.L. Bartlett. *Neural network learning: Theoretical foundations*. cambridge university press, 2009.
- [AC17] Y. Armenti and S. Crépey. XVA metrics for CCP optimisation. Technical report, Working paper available at <https://math.maths.univ-evry.fr/crepey>, 2017.
- [ACC17] C. Albanese, S. Caenazzo, and S. Crépey. Credit, funding, margin, and capital valuation adjustments for bilateral portfolios. *Probability, Uncertainty and Quantitative Risk*, 2(1):7, 2017.
- [ADEH99] P. Artzner, F. Delbaen, J.M. Eber, and D. Heath. Coherent measures of risk. *Math. Finance*, 9(3):203–228, 1999.
- [ADG⁺18] A. Agarwal, S. De Marco, E. Gobet, J. López-Salas, F. Noubiagain, and A. Zhou. Numerical approximations of McKean anticipative backward stochastic differential equations arising in Initial Margin requirements. *Hal preprint hal-01686952*, 2018.
- [ADGL15] A. Agarwal, S. De Marco, E. Gobet, and G. Liu. Rare event simulation related to financial risks: efficient estimation and sensitivity analysis. *HAL preprint, hal-01219616*, 2015.
- [ADGL18] A. Agarwal, S. De Marco, E. Gobet, and G. Liu. Study of new rare event simulation schemes and their application to extreme scenario generation.

Mathematics and Computers in Simulation, 143:89–98, 2018.

- [AG07] S. Asmussen and P.W. Glynn. *Stochastic simulation: Algorithms and analysis*. Stochastic Modelling and Applied Probability 57. New York, NY: Springer., 2007.
- [ATCD17] L. Abbas-Turki, S. Crépey, and B. Diallo. XVA principles, nested Monte Carlo, and GPU optimizations. Technical report, Working paper available at <https://math.maths.univ-evry.fr/crepey>, 2017.
- [BFP09] O. Bardou, N. Frikha, and G. Pages. Computing VaR and CVaR using stochastic approximation and adaptive unconstrained importance sampling. *Monte-Carlo Methods and Applications*, 15(3):173–210, 2009.
- [BGG⁺16] C.-E. Bréhier, M. Gazeau, L. Goudenege, T. Lelièvre, and M. Rousset. Unbiasedness of some generalized adaptive multilevel splitting algorithms. *The Annals of Applied Probability*, 26(6):3559–3601, 2016.
- [BHLN13] S. Bhamidi, J. Hannig, C.Y. Lee, and J. Nolen. The importance sampling technique for understanding rare event in Erdos-Renyi random graphs. *preprint arXiv:1302.6551*, 2013.
- [Buc90] J.A. Bucklew, editor. *Large deviation techniques in decision, simulation, and estimation*. John Wiley & Sons, 1990.
- [CBCR14] M. Chiachio, J.L. Beck, J. Chiachio, and G. Rus. Approximate bayesian computation by subset simulation. *SIAM Journal on Scientific Computing*,

- 36(3):A1339–A1358, 2014.
- [CC10] R. Carmona and S. Crépey. Particle methods for the estimation of credit portfolio loss distributions. *Int. J. Theor. Appl. Finance*, 13(4):577–602, 2010.
- [CDG11] F. Cérou, P. Del Moral, and A. Guyader. A nonasymptotic theorem for unnormalized Feynman-Kac particle models. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(3):629–649, 2011.
- [CESS17] S. Crépey, R. Élie, W. Sabbagh, and S. Song. When capital is a funding source: The XVA anticipated BSDEs. *preprint*, 2017.
- [CFG17] S. Crépey, G. Fort, E. Gobet, and U. Stazhynski. Uncertainty quantification for stochastic approximation limits using chaos expansion. *hal-01629952*, 2017.
- [CFV09] R. Carmona, J.P. Fouque, and D. Vestal. Interacting particle systems for the computation of rare credit portfolio losses. *Finance Stoch.*, 13(4):613–633, 2009.
- [CG07] F. Cérou and A. Guyader. Adaptive multilevel splitting for rare event analysis. *Stoch. Anal. Appl.*, 25(2):417–443, 2007.
- [CMFG12] F. Cérou, P. Del Moral, T. Furon, and A. Guyader. Sequential Monte-Carlo for rare event estimation. *Stat. Comput.*, 22(3):795–808, 2012.
- [Del04] P. Del Moral. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer, New-York, 2004.
- [Dev86] L. Devroye. *Nonuniform random variate generation*. Springer-Verlag, New York,

1986.

- [DG05] P. Del Moral and J. Garnier. Genealogical particle analysis of rare events. *Annals of Applied Probability*, 15:2496–2534, 2005.
- [DL09] L. Devineau and S. Loisel. Construction d’un algorithme d’accélération de la méthode des «simulations dans les simulations» pour le calcul du capital économique solvabilité ii. *Bulletin Français d’Actuariat*, 10(17):188–221, 2009.
- [EPQ97] N. El Karoui, S.G. Peng, and M.C. Quenez. Backward stochastic differential equations in finance. *Math. Finance*, 7(1):1–71, 1997.
- [FM03] G. Fort and E. Moulines. Convergence of the monte carlo expectation maximization for curved exponential families. *Annals of Statistics*, pages 1220–1259, 2003.
- [FS02] H. Föllmer and A. Schied. *Stochastic finance*, volume 27 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 2002. An introduction in discrete time.
- [GHS00] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Variance reduction techniques for estimating value-at-risk. *Management Science*, 46:1349–1364, 2000.
- [Gil08] M.B. Giles. Multilevel Monte-Carlo path simulation. *Operation Research*, 56:607–617, 2008.
- [Gil15] M.B. Giles. Multilevel monte carlo methods. *Acta Numerica*, 24:259–328, 2015.
- [GJR14] J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *preprint available*

at *SSRN 2509457*, 2014.

- [GL15] E. Gobet and G. Liu. Rare event simulation using reversible shaking transformations. *SIAM Scientific Computing*, 37(5):A2295–A2316, 2015.
- [Gla03] P. Glasserman. *Monte-Carlo methods in Financial Engineering*. Springer Verlag, New York, 2003.
- [GT13] A. Gulisashvili and P. Tankov. Tail behavior of sums and differences of log-normal random variables. *arXiv preprint arXiv:1309.3057*, 2013.
- [Guy14] J. Guyon. Path-dependent volatility. *Preprint available at <http://ssrn.com/abstract=2425048>*, september 2014.
- [Hei01] S. Heinrich. Multilevel Monte-Carlo Methods. In *LSSC '01 Proceedings of the Third International Conference on Large-Scale Scientific Computing*, volume 2179 of *Lecture Notes in Computer Science*, pages 58–67. Springer-Verlag, 2001.
- [LK10] O.P. Le Maître and O.M. Knio. *Spectral Methods for Uncertainty Quantification. With Applications to Computational Fluid Dynamics*. Scientific Computation. Springer, 2010.
- [LMN13] K. Latuszyński, B. Miasojedow, and W. Niemiro. Nonasymptotic bounds on the estimation error of MCMC algorithms. *Bernoulli*, 19(5A):2033–2066, 2013.
- [Mor13] P. Del Moral. *Mean field simulation for Monte-Carlo integration*, volume 126 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL,

2013.

- [MT09] S. Meyn and R.L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009.
- [Nua06] D. Nualart. *Malliavin calculus and related topics*. Springer Verlag, second edition, 2006. (with corrections on the webpage of the author).
- [RT09] G. Rubino and B. Tuffin. *Rare Event Simulation using Monte-Carlo Methods*. Wiley, 2009.
- [RU00] R.T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.
- [Sim16] E. Simonnet. Combinatorial analysis of the adaptive last particle method. *Statistics and Computing*, 26(1-2):211–230, 2016.
- [SPB16] D. Straub, I. Papaioannou, and W. Betz. Bayesian analysis of rare events. *Journal of Computational Physics*, 314:538–556, 2016.
- [SW95] A. Schwartz and A. Weiss, editors. *Large Deviations for Performance Analysis*. Chapman and Hall, London, 1995.