
SAVE THE DATE: JULY 17 - AUGUST 25, 2017

SUMMER SCHOOL (1 WEEK)+ PROJECTS IN COMPUTATIONAL STOCHASTICS (5 WEEKS)

Website: <http://smai.emath.fr/cemracs/cemracs17/>



CEMRACS 2017

*Numerical methods for stochastic models:
control, uncertainty quantification, mean-field*

July 17 - August 25, CIRM, Marseille



[Home](#)

[News](#)

[Summer school](#)

[Research projects](#)

[Seminars](#)

[Registration](#)

[Participants](#)

CEMRACS concept

The CEMRACS is a scientific event of the SMAI (the french Society of Applied and Industrial Mathematics). The Cemracs concept was initiated in 1996 by Yvon Maday and Frédéric Coquel and takes place every year at CIRM in Luminy (Marseille, France) during 6 weeks. The goal of this event is to bring together scientists from both the academic and industrial communities and discuss these topics.

During the first week, a classical [summer school](#) is proposed. It consists of several lectures given by leading scientists and related to the topics of the research projects. The remaining 5 weeks are dedicated to working on the [research projects](#), possibly after a morning [seminar](#).

MCMC design-based non-parametric regression for rare-event. Application to nested risk computations

emmanuel.gobet@polytechnique.edu

Centre de Mathématiques Appliquées and FiME,
Ecole Polytechnique and CNRS



With the support of



Joint work with G. Fort and E. Moulines .

Preprint available at <https://hal.archives-ouvertes.fr/hal-01394833>

STATEMENT OF THE PROBLEM

Aim: to compute

$$\mathcal{J} := \mathbb{E} [f(\mathbf{Y}, \mathbb{E} [\mathbf{R} \mid \mathbf{Y}]) \mid \mathbf{Y} \in \mathcal{A}]$$

where

- ✓ R and Y are vector-valued random variables,
- ✓ \mathcal{A} is a rare subset, i.e. $\mathbb{P}(Y \in \mathcal{A})$ small.

Equivalent to

$$\mathbb{E} [f(X, \mathbb{E} [R \mid X])]$$

where X has the conditional distribution of Y given $\{Y \in \mathcal{A}\}$



Nested expectations. Computation in two stages:

- ✓ one **inner (cond.) expectation**: $\phi_{\star}(\mathbf{X}) = \mathbb{E} [\mathbf{R} \mid \mathbf{X}]$
- ✓ one **outer expectation** related to a rare-event

Applications of computing nested expectations

- ✓ **Dynamic programming equations** for stochastic control and optimal stopping problems, see [TR01, LS01, Egl05, LGW06, BKS10].
But coupling with rare-event is usually not required.
- ✓ **Financial and actuarial risk management** [MFE05]
 - ▶ risk management of portfolios written with derivative options [Gordy, Juneja, [GJ10]]
 R : aggregated cashflows of derivatives at time T'
 Y : for the underlying financial variables at time $T < T'$.
 - ▶ computation of the extreme exposure (Value at Risk, Conditional VaR) of the portfolio.
 - ▶ Essential concerns for Solvency Capital Requirement in insurance [Devineau, Loisel [DL09]].

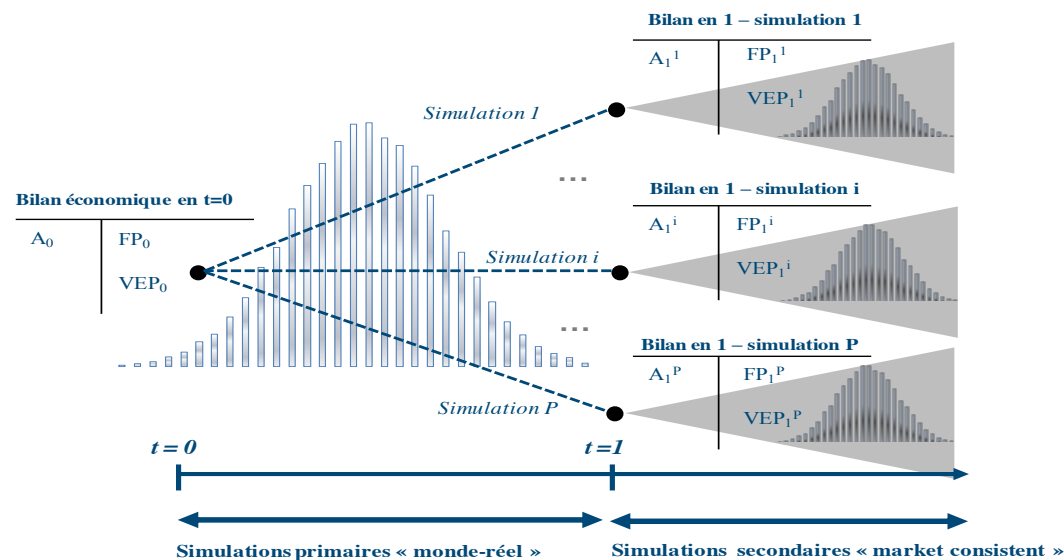
1 Several approaches for outer/inner stages

1.1 Crude MC/Crude MC for $\mathbb{E}[f(X, \mathbb{E}[R | X])]$

- ✓ Sample M i.i.d. of $X^{(m)}$ (using for instance rejection algorithms)
- ✓ For each $X^{(m)}$, sample N i.i.d. samples of $R | X^{(m)}$:

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

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



Simulations In Simulations algorithm, picture taken from [DL09]

1.2 Crude MC/Other spatial approximation of $\mathbb{E}[\cdot | X]$

- ✓ [Hong, Juneja [HJ09]] for kernel estimators
- ✓ [Liu, Staum [LS10]] for kriging techniques
- ✓ [Broadie, Du, Moallemi [BDM15]] for least-squares regression methods (with possible weighting)

Outer stage remains unsatisfactory, because Y are sampled i.i.d. and rejected when outside \mathcal{A} .

1.3 Our approach

- ✓ use a MCMC scheme for the outer stage
- ✓ design a regression scheme for the inner stage, using ϕ_1, \dots, ϕ_L basis functions

2 Efficient sampling under the conditional probability: MCMC approach

From [G', Liu [GL15]] and [Agarwal, De Marco, G', Liu[ADGL15]], known as POP scheme.

Aim: sampling of $Y \mid Y \in \mathcal{A}$.

2.1 Preliminaries

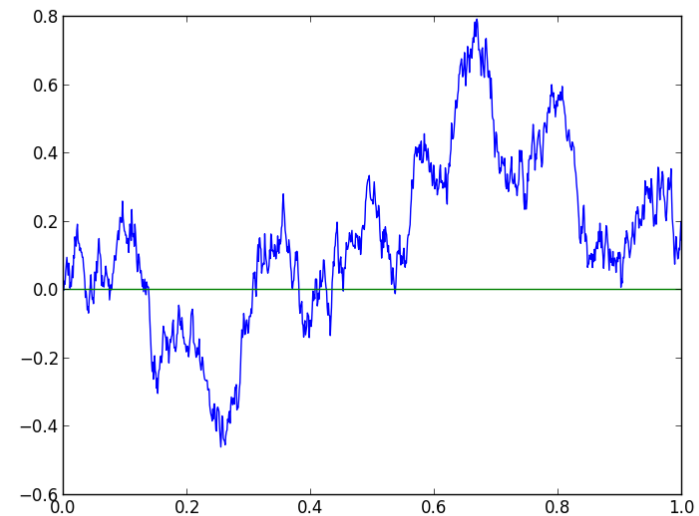
We seek a flexible approach able to overcome some constraints of importance sampling.

Example (Large oscillation of Ornstein-Uhlenbeck 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 sample on the rare event

$$\mathcal{A} := \left\{ \max_{0 \leq t \leq T} Z_t > 1.6 \text{ and } \min_{0 \leq t \leq T} Z_t < -1.6 \right\}.$$



Importance sampling techniques are difficult to apply on this example.

2.2 MCMC shaker

Definition (of reversible shaking transformation).

- ✓ Define $\mathcal{K}(\cdot) = \mathbf{K}(\cdot, \mathbf{U})$ for some $K(\cdot)$ and U independent of Y .
- ✓ $\mathcal{K}(\cdot)$ is a **reversible shaker** for Y if $(\mathbf{Y}, \mathcal{K}(\mathbf{Y})) \stackrel{d}{=} (\mathcal{K}(\mathbf{Y}), \mathbf{Y})$.

Similar to balance equation in **symmetric Metropolis-Hastings** sampler.

Example (SDE shaker). If Y is a standard Brownian motion,

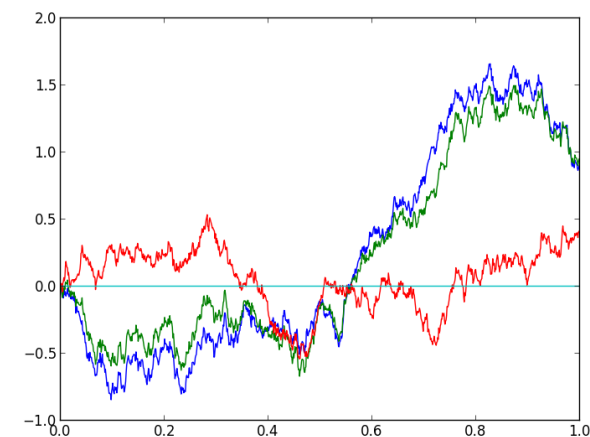
$$K(Y, U) = \left(\int_0^t \rho_s dY_s + \int_0^t \sqrt{1 - \rho_s^2} dU_s \right)_{0 \leq t \leq T}$$

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

Shaking OU

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

Path of \mathbf{Z} , shaked Z with $\rho = 0.9$ and $\rho = 0.1$



▷ **SHAKER WITH REJECTION**: $\mathcal{M}^{\mathcal{K}}(\mathbf{Y}) := \begin{cases} \mathcal{K}(Y) & \text{if } \mathcal{K}(Y) \in \mathcal{A} \\ Y & \text{if } \mathcal{K}(Y) \notin \mathcal{A}. \end{cases}$

Proposition (conditional invariance under shaking with rejection).

The distribution of Y conditionally on $\{Y \in \mathcal{A}\}$ is invariant w.r.t. the random transformation $\mathcal{M}^{\mathcal{K}}$: for any bounded $\varphi : \mathbb{S} \rightarrow \mathbb{R}$

$$\mathbb{E}(\varphi(\mathcal{M}^{\mathcal{K}}(\mathbf{Y})) \mid \mathbf{Y} \in \mathcal{A}) = \mathbb{E}(\varphi(\mathbf{Y}) \mid \mathbf{Y} \in \mathcal{A}).$$

✓ **Birkhoff theorem**: given an initial position $\mathbf{X}_0 \in \mathcal{A}$ define

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

Then, if $(X^{(m)})_m$ is an ergodic Markov chain,

$$\frac{1}{M} \sum_{m=1}^M \mathbf{f}(\mathbf{X}^{(m)}) \xrightarrow{M \rightarrow +\infty} \mathbb{E}[\mathbf{f}(\mathbf{Y}) \mid \mathbf{Y} \in \mathcal{A}].$$

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

```

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

```

Full algorithm with M data, $M \geq L$.

2.4 Convergence results about regression

Notations:

- ✓ Let $\mu \, d\lambda$ be the distribution of $X \in \mathbb{R}^d$, 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 $\psi_\star = \arg \inf_{\varphi \in \text{Span}(\phi_1, \dots, \phi_L)} |\phi_\star - \varphi|_{L_2(\mu)}$ be the projection of ϕ_\star on the basis functions.

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

Assume that

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

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

(ii) the conditional distribution Q satisfies

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

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

$$\begin{aligned} \Delta_M &:= \mathbb{E} \left[\frac{1}{M} \sum_{m=1}^M \left(\hat{\phi}_M(\mathbf{X}^{(m)}) - \phi_\star(\mathbf{X}^{(m)}) \right)^2 \right] \\ &\leq \frac{\sigma^2 \mathbf{L}}{M} + |\psi_\star - \phi_\star|_{\mathbf{L}_2(\mu)}^2 + \frac{C_P}{M} \sum_{m=1}^M \rho(m). \end{aligned}$$

Remarks.

- ✓ Finite dimensional Gaussian shaker $\implies \rho(m) = \text{Cst } e^{-cm}$ with $c > 0$.
- ✓ Usually $\sum_{m=1} \rho(m) < +\infty$
- ✓ The use of MCMC design does not impact significantly the statistical error.

Theorem (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

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

Then

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

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

```

1 /* Simulation of the design and the observations */
2  $X^{(0)} \sim \xi$ , where  $\xi$  is a distribution on  $\mathcal{A}$  ;
3 for  $m = 1$  to  $M$  do
4    $X^{(m)} \sim \mathbf{P}(X^{(m-1)}, dx)$  ;
5   for  $k = 1$  to  $N$  do
6      $R^{(m,k)} \sim \mathbf{Q}(X^{(m)}, dr)$ ;
7 /* Conditional expectation by crude Monte Carlo */
8 Compute  $\bar{R}_N^{(m)} = \frac{1}{N} \sum_{k=1}^N R^{(m,k)}$ ;
9 /* Final estimator using ergodic average */
10 Return  $\tilde{\mathcal{J}}_M := \frac{1}{M} \sum_{m=1}^M f(X^{(m)}, \bar{R}_N^{(m)})$ .

```

Full algorithm with M outer samples, and N inner samples for each outer one.

Theorem (Convergence analysis). Assume that

(i) the second/fourth conditional moments of Q are bounded: for $p = 2, 4$,

$$\sigma_p := \left(\sup_{x \in \mathcal{A}} \int \left| r - \int r Q(x, dr) \right|^p Q(x, dr) \right)^{1/p} < \infty.$$

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

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

Then,

$$\left(\mathbb{E} \left[\left| \tilde{J}_M - J \right|^2 \right] \right)^{1/2} \leq \begin{cases} \text{Cst} \left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{M}} \right) & \text{if } f : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R} \text{ is globally Lipschitz} \\ & \text{in the second variable} \\ \text{Cst} \left(\frac{1}{N} + \frac{1}{\sqrt{M}} \right) & \text{if } f \text{ is continuously differentiable in the} \\ & \text{second variable, with Lipschitz derivative} \end{cases}$$

2.6 Asymptotic analysis MCMC/Regression versus MCMC/Crude MC

Assume that the unknown regression function $\phi_\star : \mathbb{R}^d \rightarrow \mathbb{R}$ is C^p . Then

$$\begin{aligned} \text{Error}_{\text{MCMC/Regression}} &= O\left(\text{Cost}^{-\frac{p}{2p+d}}\right), \\ \text{Error}_{\text{MCMC/Crude MC}} &= \begin{cases} O\left(\text{Cost}^{-\frac{1}{4}}\right) & \text{if } f \text{ Lipschitz} \\ O\left(\text{Cost}^{-\frac{1}{3}}\right) & \text{if } f \text{ smoother.} \end{cases} \end{aligned}$$

In the standard case of Lipschitz f :

- ✓ Low dimension ($p \geq d/2$): use MCMC/Regression
- ✓ Large dimension: use MCMC/Crude MC

3 Numerical examples

Goal: to approximate

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

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.

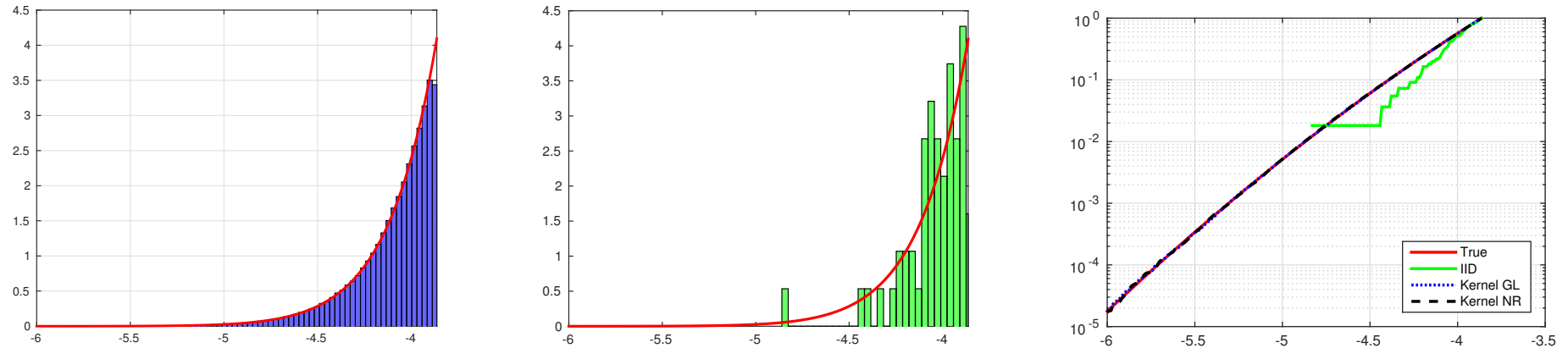
3.1 A toy example in dimension 1

Here $h(y) = y$ and $\mathcal{S} = \{s \in \mathbb{R}_+ : s \leq s_\star\}$ so that

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

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

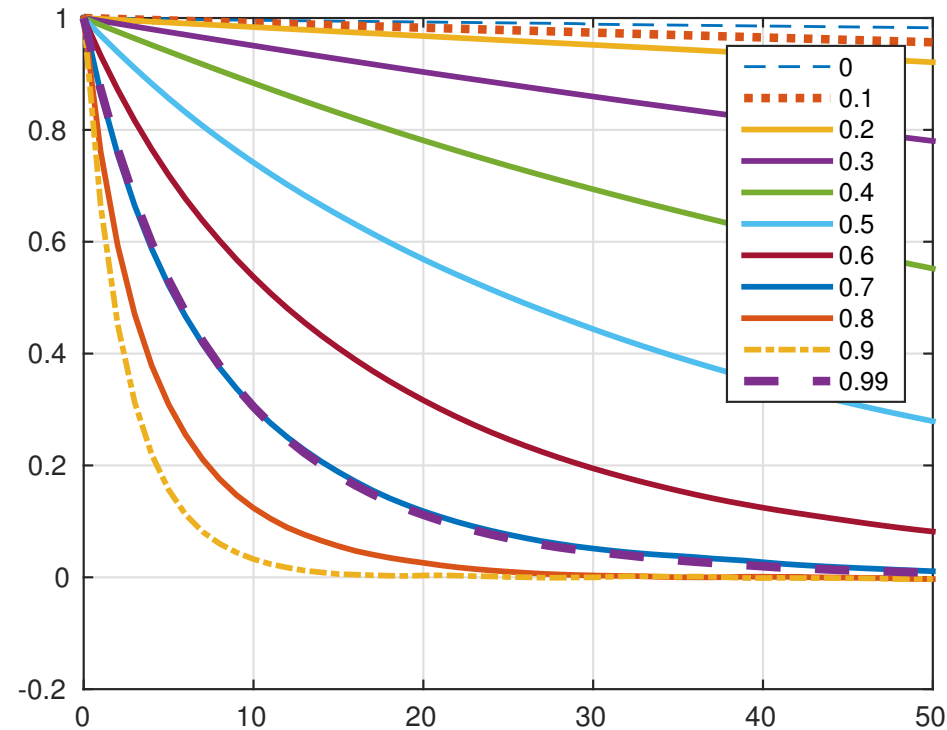
Here $\mathbb{P}(Y \in \mathcal{A}) \approx 5.6e-5$



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

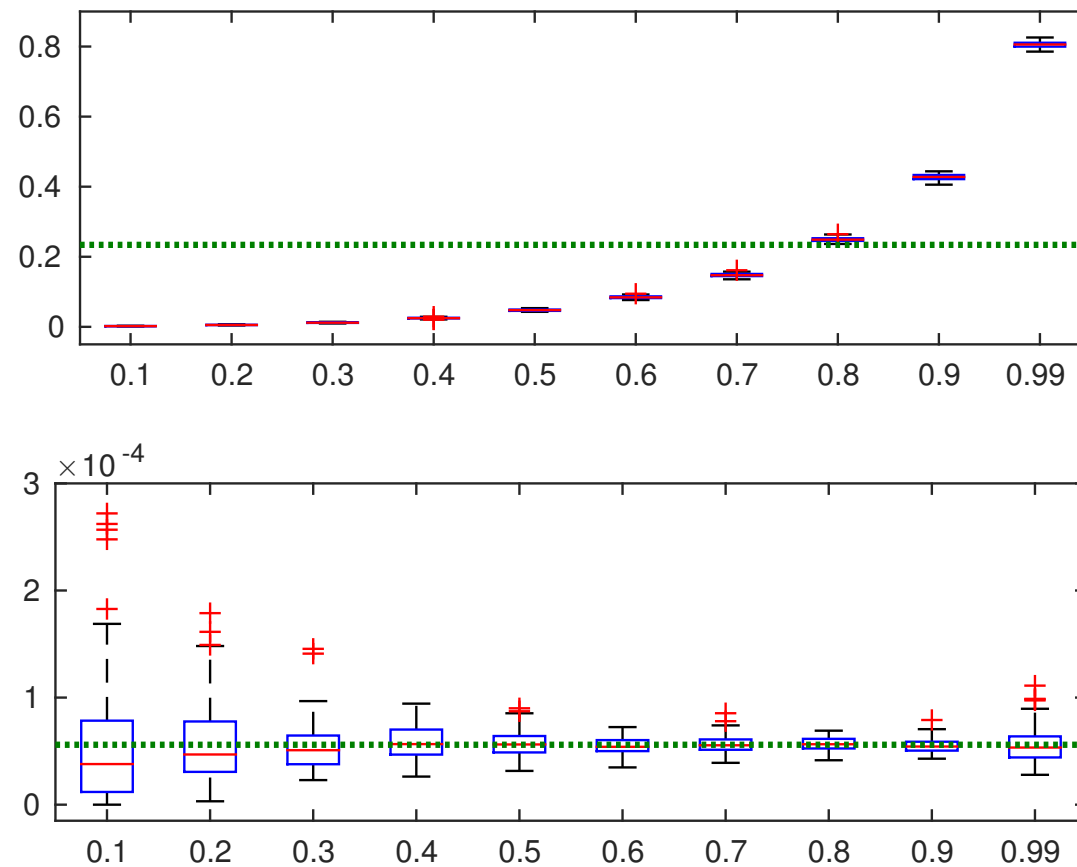
Choosing the shaking parameter is important!



For different values of ρ , estimation of the autocorrelation function (over 100 independent runs) of the chain P_{GL} .

Each curve is computed using $1e6$ sampled points.

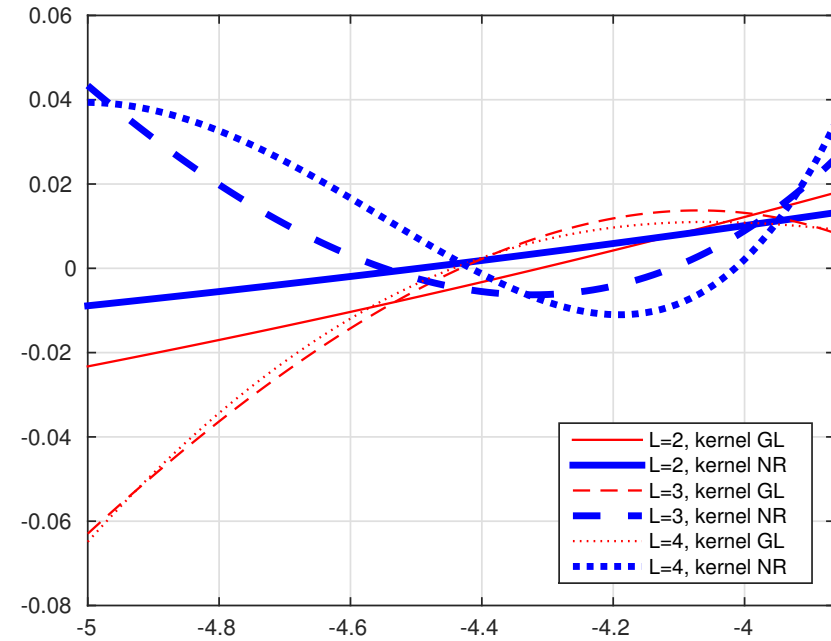
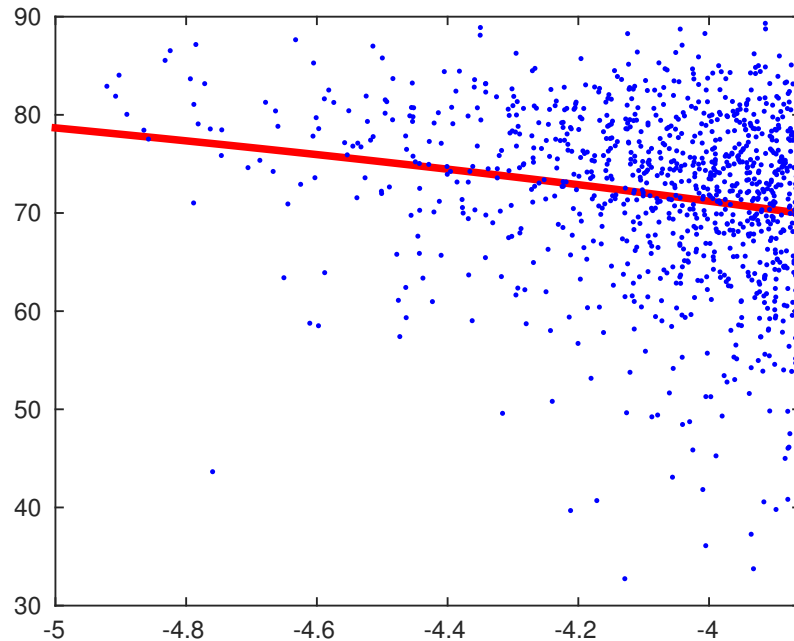
Tune the shaking parameter according to the acceptance rate



(top) Mean acceptance rate after M chain iterations of the chain

(green = 23.4% [Rosenthal 2008])

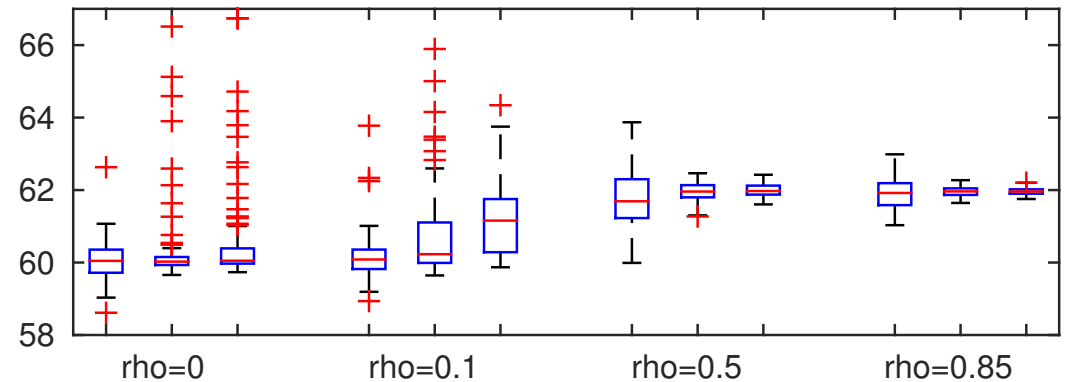
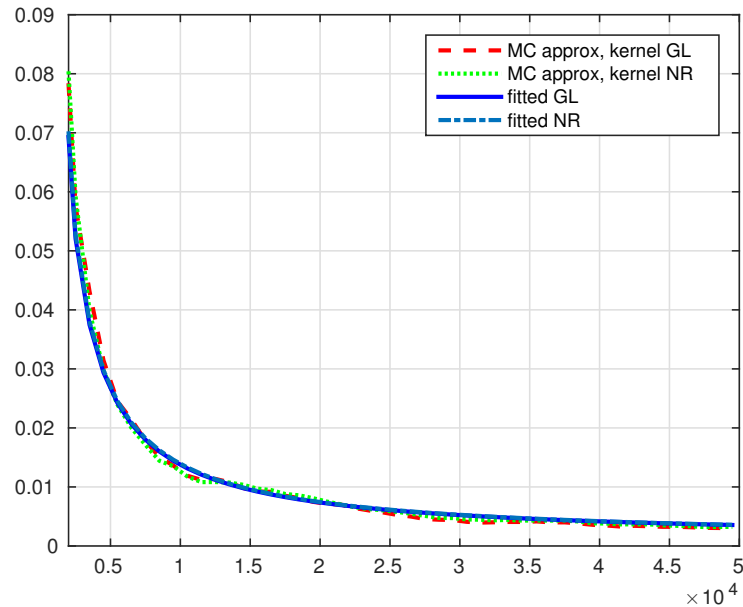
(bottom) Estimation of $\mathbb{P}(Y \in \mathcal{A})$ by combining splitting and MCMC



(left) 1000 sampled points $(X^{(m)}, R^{(m)})$ (using the MCMC sampler), together with ϕ_* ;

(right) A realization of the error function $x \mapsto \hat{\phi}_M(x) - \phi_*(x)$ on $[-5, y_*]$, for different values of $L \in \{2, 3, 4\}$ and two different kernels when sampling $X^{(1:M)}$.

Empirical evidence of the theoretical error bounds



(left) Monte Carlo approximations of $M \mapsto \Delta_M$, and fitted curves of the form $M \mapsto \alpha + \beta/M$.

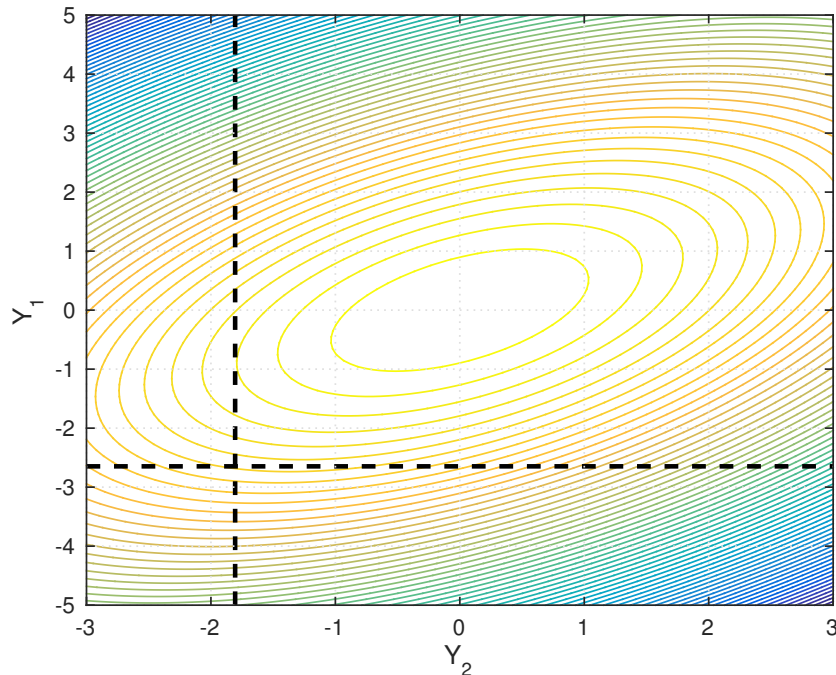
(right) For different values of ρ , and for three different values of M , boxplot of 100 independent estimates $\hat{\mathcal{J}}_M$ when $X^{(1:M)}$ is sampled from Markov chain.

3.2 Correlated geometric Brownian motions in dim. 2

We consider

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

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



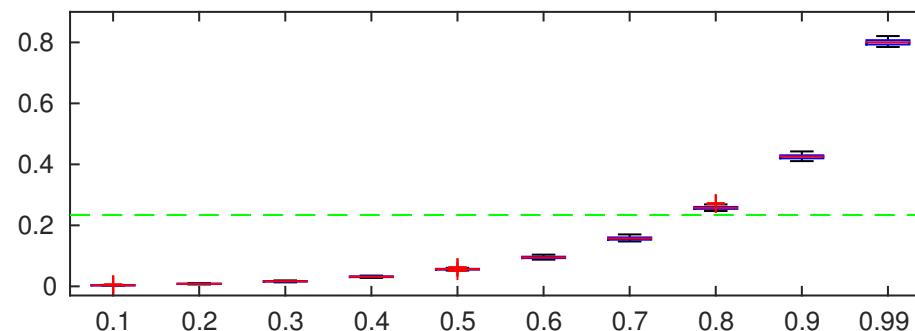
Level curves of the density function of $(S_{1,T}, S_{2,T})$ and the rare set in the lower left area delimited by the two hyperplanes.

For the basis functions, we take

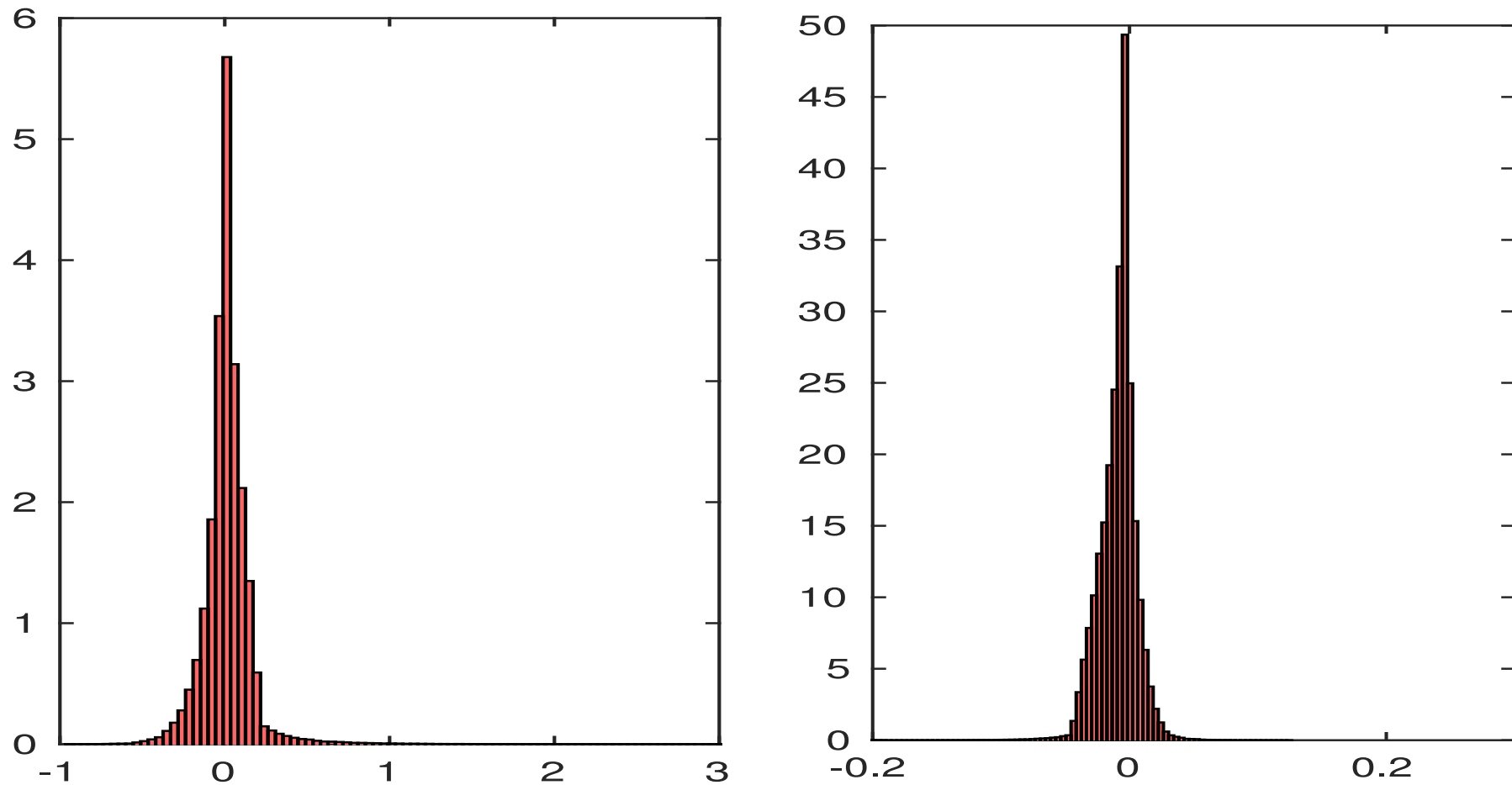
$$\begin{aligned}\varphi_1(x) &= 1, & \varphi_2(x) &= \sqrt{x_1}, & \varphi_3(x) &= \sqrt{x_2}, \\ \varphi_4(x) &= x_1, & \varphi_5(x) &= x_2, & \varphi_6(x) &= \sqrt{x_1 x_2}.\end{aligned}$$

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

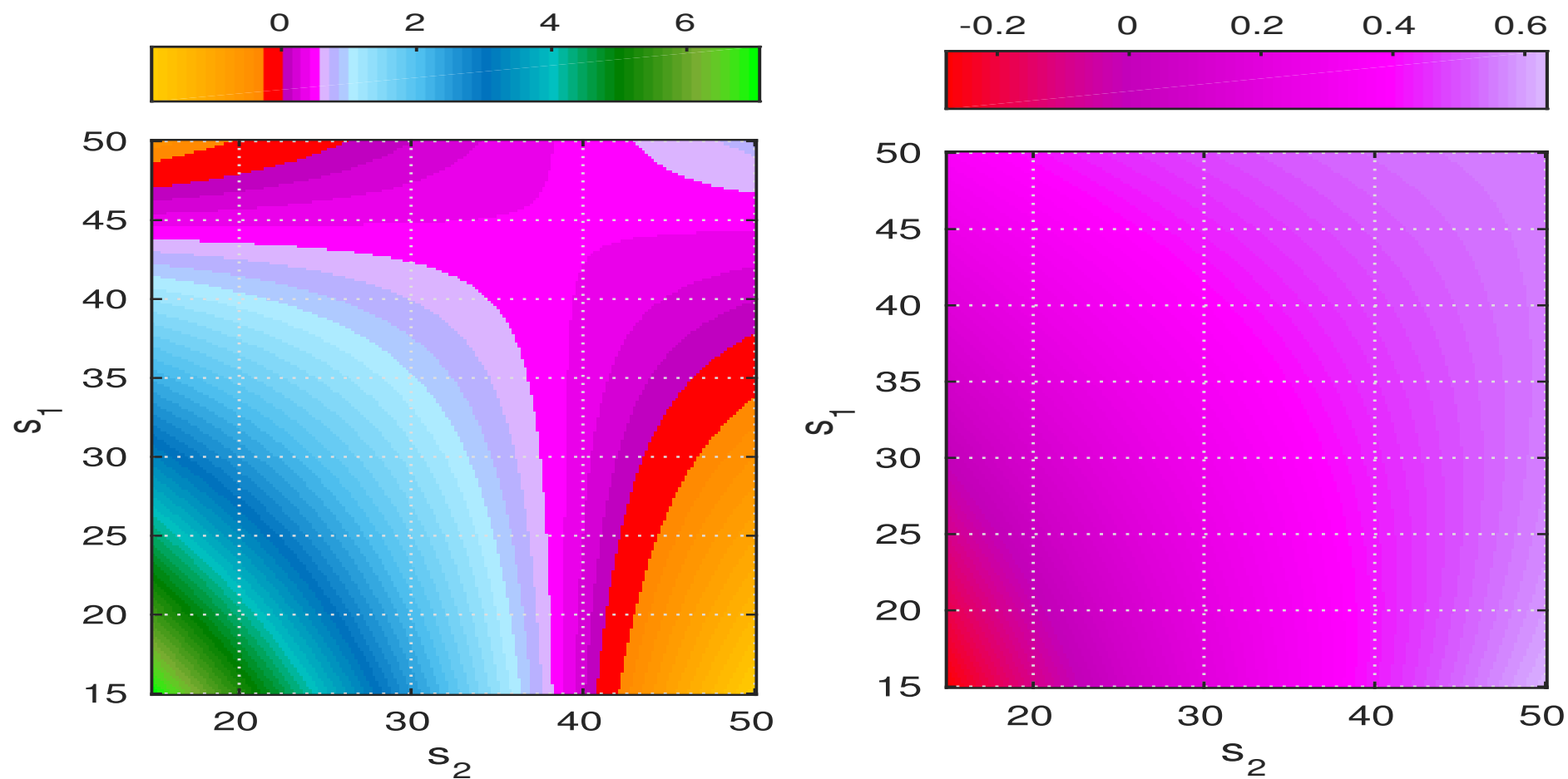
Tuning the shaking parameter according acceptance rate



Boxplot over 100 independent runs, of the mean acceptance rate after $M = 1e4$ iterations for the Markov chain kernel. Different values of ρ are considered.



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



(left) Error function with $L = 3$.

(right) The same case with $L = 6$.

4 Conclusion

- ✓ Design of a regression method suitable for accurate computations in the tails
- ✓ Tail distribution is sampled using MCMC
- ✓ MCMC/Regression \gg MCMC/Crude Monte Carlo for small ratio $\frac{\text{dimension}}{\text{smoothness}}$
- ✓ What next?
 - ▶ Ongoing works (with David Barrera, Postdoc at Ecole Polytechnique, and Gersende Fort)
 - ▶ Uniform concentration of measure estimates for non stationary ergodic chains
 - ▶ Rare event set \mathcal{A} depending itself on conditional expectations (adaptive MCMC-regression scheme)

References

- [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.
- [BDM15] M. Broadie, Y. Du, and C.C. Moallemi. Risk estimation via regression. *Operations Research*, 63(5):1077–1097, 2015.
- [BKS10] D. Belomestny, A. Kolodko, and J. Schoenmakers. Regression methods for stochastic control problems and their convergence analysis. *SIAM Journal on Control and Optimization*, 48(5):3562–3588, 2010.
- [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.
- [Egl05] D. Egloff. Monte-Carlo algorithms for optimal stopping and statistical learning. *Ann. Appl. Probab.*, 15:1396–1432, 2005.
- [FGM16] G. Fort, E. Gobet, and E. Moulines. MCMC design-based non-parametric regression for rare-event. Application to nested risk computations. *hal-01394833 preprint*, November 2016.
- [GJ10] M.B. Gordy and S. Juneja. Nested simulation in portfolio risk measurement. *Management Science*, 56(10):1833–1848, 2010.
- [GL15] E. Gobet and G. Liu. Rare event simulation using reversible shaking transformations. *SIAM Scientific Computing*, 37(5):A2295–A2316, 2015.
- [HJ09] L.J. Hong and S. Juneja. Estimating the mean of a non-linear function of conditional

expectation. In *Simulation Conference (WSC), Proceedings of the 2009 Winter*, pages 1223–1236. IEEE, 2009.

- [LGW06] J-P. Lemor, E. Gobet, and X. Warin. Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations. *Bernoulli*, 12(5):889–916, 2006.
- [LS01] F. Longstaff and E.S. Schwartz. Valuing American options by simulation: A simple least squares approach. *The Review of Financial Studies*, 14:113–147, 2001.
- [LS10] M. Liu and J. Staum. Stochastic kriging for efficient nested simulation of expected shortfall. *The Journal of Risk*, 12(3):3, 2010.
- [MFE05] A.J. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management*. Princeton series in finance. Princeton University Press, 2005.
- [TR01] J.N. Tsitsiklis and B. Van Roy. Regression methods for pricing complex american-style options. *IEEE Transactions on Neural Networks*, 12(4):694–703, 2001.