Stochastic approaches for extreme phenomena: Theoretical tools, numerical challenges, applications

Denis Talay

INRIA Sophia Antipolis, France
TOSCA Project-team

Nice – Novembre 2013

Outline

Outline

Standard extreme values distributions

See, e.g., Embrechts, Klüppelberg, Mikosch, Springer-Verlag,1997.
Let X_i be a sequence of independent and identically distributed random variables. Let F be the d.f. of X_i, $M_n := \max(X_i, 1 \leq i \leq n)$.
One has

$$\mathbb{P}(M_n \leq x) = F^n(x).$$

Theorem (Fisher–Tippett).

Suppose there exist real numbers c_n and d_n and a random variable H such that

$$\frac{M_n - d_n}{c_n} \longrightarrow H \text{ in distribution.}$$

Then the probability distribution of H is of one of the three following types:

$$\textit{(Fréchet:)} : \Phi_\alpha(x) := \begin{cases} 0, & x < 0, \\ \exp(-x^{-\alpha}), & x > 0, \end{cases}$$

$$\textit{(Weibull:)} : \Psi_\alpha(x) := \begin{cases} \exp(-|x|^{-\alpha})0, & x < 0, \\ 1, & x > 0, \end{cases}$$

$$\textit{(Gumbel:)} : \Lambda(x) := \exp(-e^{-x}),\ x \in \mathbb{R}.$$

Von Mises sufficient conditions

If F has density f and

$$\lim_{x \to \infty} \frac{xf(x)}{1 - F(x)} = \alpha > 0,$$

then H is Fréchet (α)-distributed.
Examples: Pareto-like distributions ($1 - F(x) \equiv Kx^{-\alpha}$), log-gamma distributions ($1 - F(x) \equiv \frac{\alpha^{\beta-1}}{\Gamma(\beta)}(\log(x))^{\beta-1}x^{-\alpha}$).

If F has density f which is positive on a finite interval (a, b) and

$$\lim_{x \to b} \frac{b - x)f(x)}{1 - F(x)} = \alpha > 0,$$

then H is Weibull (α)-distributed.
Examples: uniform distributions, beta distributions.

Suppose that, for some $a \geq 0$,

$$1 - F(x) = C \exp\left(-\int_a^x \frac{1}{q(t)} dt\right), \; x > a,$$

where q' tends to 0 at infinity (or, more generally, at the end point of F). Then H is Gumbel distributed.
Examples: Exponential, Weibull, Normal and log-normal distributions.

Outline

Extremal events modelling

Statistical issues

Monte Carlo methods

Approximation of quantiles

Risk measures

The three standard extreme values distributions may be represented by

$$F_H(x) = \exp\left(-\left(1+\xi\frac{x-\mu}{\psi}\right)^{-1/\xi}\right),\ 1+\xi\frac{x-\mu}{\psi} > 0.$$

The case $\xi = 0$ is the Gumbel case.
Given independent samples of H, the parameters of F_H can be estimated by the maximum likelihood procedure.
Specific estimators are efficient, all of them based on based on order statistics: Pickands estimator and Hill's estimators for ξ, quantile estimators, etc.
Goodness-of-fit tests for heavy tailed distributions, etc.

Tauberian theory

Let μ be a measure on $[0, \infty)$ and finite on bounded sets. The Laplace transform $\hat{\mu}(\lambda)$ of μ is the real-valued function defined for $\lambda \geq c$ by

$$\hat{\mu}(\lambda) := \int_0^\infty e^{-\lambda x} \mu(dx),$$

where $c := \inf\{\lambda \in \mathbb{R} : \int_0^\infty e^{-\lambda x} \mu(dx) < \infty\}$.
Let $R_\alpha(0+)$ denote the class of regularly varying functions at origin with index α and $\overleftarrow{f}$ denote the generalized inverse of f defined as $\overleftarrow{f}(x) := \sup\{t : f(t) > x\}$.

Bruijn's Tauberian theorem.

Let μ be a measure on $(0,\infty)$ which is finite on bounded sets and whose Laplace transform is well defined for all $\lambda > 0$. For $\alpha < 0$ and $\phi \in R_\alpha(0+)$, set $\psi(\lambda) := \frac{\phi(\lambda)}{\lambda}$. For all $B > 0$,

$$-log\mu(0,x] \underset{x\to 0+}{\sim} \frac{B}{\overleftarrow{\phi}(1/x)}$$

if and only if

$$-\log\nu(\lambda) \underset{\lambda\to\infty}{\sim} \frac{(1-\alpha)\left(\frac{B}{-\alpha}\right)^{\frac{\alpha}{\alpha-1}}}{\overleftarrow{\psi}(\lambda)}.$$

Let X be a non-negative random variable with probability distribution function F and complex valued Laplace-Stieltjes transform Φ defined as

$$\forall s \in \mathbb{C},\ \Phi(s) := \int_0^\infty e^{-sx} dF(x).$$

The abscissa of convergence of $\Phi(s)$ is defined as the real number a_0 such that the preceding integral converges for $\mathrm{Re}(s) > a_0$ and diverges for $\mathrm{Re}(s) < a_0$. The line $\mathrm{Re}(s) = a_0$ is called the axis of convergence of $\Phi(s)$.

Nakagawa's Analytic Tauberian Theorem.

If $-\infty < a_0 < 0$ and the singularities of $\Phi(s)$ on the axis of convergence $Re(s) = a_0$ have a finite number of poles, then

$$\lim_{x\to\infty} \frac{1}{x} \log \mathbb{P}[X > x] = a_0.$$

Outline

Rare event simulation: compute $\mathbb{P}(A)$ with $\mathbb{P}(A)$ (very) small given a prescribed relative error .
Variance reduction: importance sampling .
Consider the stochastic differential equation model with smooth coefficients

$$X_t(x) = x + \int_0^t A_0(s, X_s(x))ds + \sum_{i=1}^{r} \int_0^t A_i(s, X_s(x))dW_s^i,$$

and the Euler scheme

$$\bar{X}^h_{(p+1)h}(x) = \bar{X}^h_{ph}(x) + A_0(ph, \bar{X}^h_{ph}(x))h + \sum_{i=1}^{r} A_i(ph, \bar{X}^h_{ph}(x))(W^i_{(p+1)h} - W^i_{ph}).$$

Aim: compute $M := \mathbb{E}f(X_T)$ or $M := \mathbb{E}f(\sup_{t\leq T} X_t)$ using a Monte Carlo method. Example: $f(x) = \mathbb{P}[|X_T| > x]$.
Under probability $\mathbb{P}$, the variance of the simulation is

$$\mathbb{E}[(f(X_T))^2] - M^2.$$

In theory one can construct a process (θ_t^*) and the corresponding Girsanov transform which lead one to a **perfect** importance sampling variance reduction method.
Actually, suppose that $u(t,x) := \mathbb{E}[f(X_T)|X_0 = x]$ is strictly positive. Then, for all (θ_t),

$$\mathbb{E}f(X_T) = \mathbb{E}^\theta\left[f(X_T)\frac{\mathbb{E}f(X_T)}{f(X_T)}\right].$$

We have

$$\frac{\mathbb{E}f(X_T)}{f(X_T)} = \exp(\log u(0,x_0) - \log u(T,X_T)),$$

from which, applying Itô's formula to $\log u(t,X_t)$, using a PDE characterization of $u(t,x)$ or, equivalently, the Markov property of X_t), and choosing

$$\theta_t := \frac{\sigma(X_t)\frac{\partial}{\partial x}u(t,X_t)}{u(t,X_t)},$$

we get

$$\frac{\mathbb{E}f(X_T)}{f(X_T)} = \exp\left(-\int_0^T \theta_s\,dW_s + \frac{1}{2}\int_0^T |\theta_s|^2\,ds\right).$$

The variance of the simulation under $\mathbb{P}^\theta$ then is

$$\mathbb{E}^\theta\left(\left(f(X_T)\frac{d\mathbb{P}}{d\mathbb{P}^\theta}\right)^2\right) - M^2 = \mathbb{E}\left((f(X_T))^2\frac{d\mathbb{P}}{d\mathbb{P}^\theta}\right) - M^2$$
$$= \mathbb{E}\left((f(X_T))^2\frac{\mathbb{E}f(X_T)}{f(X_T)}\right) - M^2 = 0.$$

We thus have constructed a **perfect** variance reduction.Unfortunately, it requires to know the function $u(t,x)$ which we aim to approximate. Open issue for complex coefficients: a priori estimates on $u(t,x)$ and control variate methods.

We thus now limit ourselves to a class of processes (θ_t) which can be simulated.
Under any probability measure $\mathbb{P}^\theta$, the variance of the simulation is

$$\mathbb{E}^\theta \left((f(X_T))^2 \exp(-2 \int_0^T \theta_s \cdot dW_s^\theta - \int_0^T |\theta_s|^2 \, ds) \right) - M^2,$$

where now

$$X_t = X_0 + \int_0^t (b(X_s) + \sigma(X_s)\theta_s) \, ds + \int_0^t \sigma(X_s) \, dW_s^\theta.$$

This variance is also equal to

$$\mathbb{E} \left((f(X_T))^2 \exp(- \int_0^T \theta_s \cdot dW_s + \frac{1}{2} \int_0^T |\theta_s|^2 \, ds) \right) - M^2.$$

Therefore our objective is to seek a process (θ_t) within a prescribed class such that

$$\mathbb{E} \left((f(X_T))^2 \exp(- \int_0^T \theta_s \cdot dW_s + \frac{1}{2} \int_0^T |\theta_s|^2 \, ds) \right) \ll \mathbb{E}[(f(X_T))^2].$$

Optimization problem:

$$\min_{(\theta_t)\in\mathcal{A}} \mathbb{E}\left((f(X_T))^2 \exp(-\int_0^T \theta_s \cdot dW_s + \frac{1}{2}\int_0^T |\theta_s|^2\, ds)\right),$$

where $\mathcal{A}$ is a class of adapted processes satisfying suitable integrability conditions.
Consider the class of non constant deterministic functions (θ_t). The Euler scheme is a **deterministic** (and complex!) function F_n of the vector of the Brownian increments

$$\Delta W := (W_{(p+1)T/n} - W_{pT/n},\ p = 0, \cdots, n-1).$$

Therefore, we are led to determine the piecewise constant function

$$\Theta := (\theta_{(p-1)T/n},\ p = 1, \cdots, n)$$

which minimizes

$$H(\Theta) := \mathbb{E}\left[(f \circ F_n(\Delta W))^2 \exp\left(-\Theta \cdot \Delta W + \frac{T}{2n}\|\Theta\|^2\right)\right].$$

Stochastic algorithms provide approximations of the optimal value of Θ. One may modify the Monte Carlo method to **simultaneously** optimize the choice of Θ and approximate M.

Outline

Suppose that (X_t) is d-dimesnional.

Objective: Estimates on the approximation by the Euler scheme of the quantile $\rho(x,\delta)$ of the law of $X_T^d(x)$: for $0<\delta<1$ set

$$\boxed{\rho(x,\delta) := \inf\{\rho \in \mathbb{R};\ \mathbb{P}[X_T^d(x) \leq \rho] = \delta\}}$$

and

$$\boxed{\bar{\rho}^h(x,\delta) := \inf\{\rho \in \mathbb{R};\ \mathbb{P}[\bar{X}_T^{h,d}(x) \leq \rho] = \delta\}.}$$

Constraint: The Malliavin covariance matrix of $(X_t(x))$ is NOT NECESSARILY invertible. We need to consider cases where the variance of $X^d_T(x)$ only can be controlled:
We thus consider partially hypoelliptic diffusions . Under suitable assumptions, the d-th marginal distribution of $X_T(x)$ has a smooth density $p^d_T(x, y)$ which is strictly positive at all point y in the interior of its support, and

$$|\rho(x,\delta) - \bar{\rho}^h(x,\delta)| \leq \frac{K(T)}{T^q} \, \frac{1 + \|x\|^Q}{\overline{p}^d_T(\rho(x,\delta))} \, \Psi_\lambda(x) \, h,$$

where

$$\overline{p}^d_T(\rho(x,\delta)) = \inf_{y \in (\rho(x,\delta)-1, \rho(x,\delta)+1)} p^d_T(x,y).$$

(see Zheng and T.)

Major issue: a priori estimates on the tail of the density $p_T^d(x,y)$.

Large deviation techniques available in certain situations (small noises, e.g.)

Theoretical results by Kusuocka-Stroock, Bally, Delarue, Menozzi, T. and Zheng, etc.

New challenges: a priori estimates, or convergence rate estimates, for the distribution of passage times of diffusion processes at critical thresholds (cf. Neurosciences).

Outline

Notions coming from Financial Mathematics (see Föllmer, Schied, El Karoui, etc.)
Elementary examples:

- Quantiles,
- $VaR_\lambda(X) := -\inf\{x;\ \mathbb{P}(X \leq x) \geq \lambda\}$.

VaR does not capture extreme risks and does not satisfy the inverse monotonicity condition below.

Axiomatic Risk measures

The real-valued function M is a Risk measure if

- $X_1 \leq X_2 \Longrightarrow M(X_1) \geq M(X_2)$ (inverse monotonicity)
- $M(X + a) = M(X) - a$ for all deterministic a
- $M(aX) = aM(X)$ for all $a \geq 0$
- If X_1 and X_2 have same probability distribution, then $M(X_1) = M(X_2)$
- M is convex
- Invariance under randomization (by heads and toss) of the choice between positions X_1 and X_2 such that $M(X_i) \leq 0$.

Quasi-example: Average Value at Risk

$$\begin{aligned} AVaR_\lambda(X) &:= \frac{1}{\lambda} \int_0^\lambda VaR_u(X) du \\ &= -\mathbb{E}(X \mid X \leq \rho_X(\lambda)) \end{aligned}$$

(not invariant under randomization).

Conclusion

New challenges:

- New risk measures adapted to physical and biological events (earthquakes, neurons, turbulent winds, climate, diseases, etc.)
- A priori estimates on tails of stochastic processes distributions,
- Statistical procedures for these tails
- New variance reduction methods for rare event simulation
- Convergence rates for numerical approximations of passage times
- Fascinating issues for McKean-Vlasov stochastic interacting particle systems (see Bossy et al. e.g.)