Modeling Extreme Events from Computer Simulations
Part I

Emmanuel Vazquez

SUPELEC, Gif-sur-Yvette, France

Summer School CEA-EDF-INRIA, 2011

Synopsis

Part I
- Modeling Extreme Events from Computer Simulations: formalization of the issues at stake
- The Monte Carlo approach and its limitation
- Introduction to extreme value theory

Part II
- Introduction to structural reliability
- Elicitation of probability distributions

Part III
- Advanced Monte Carlo methods

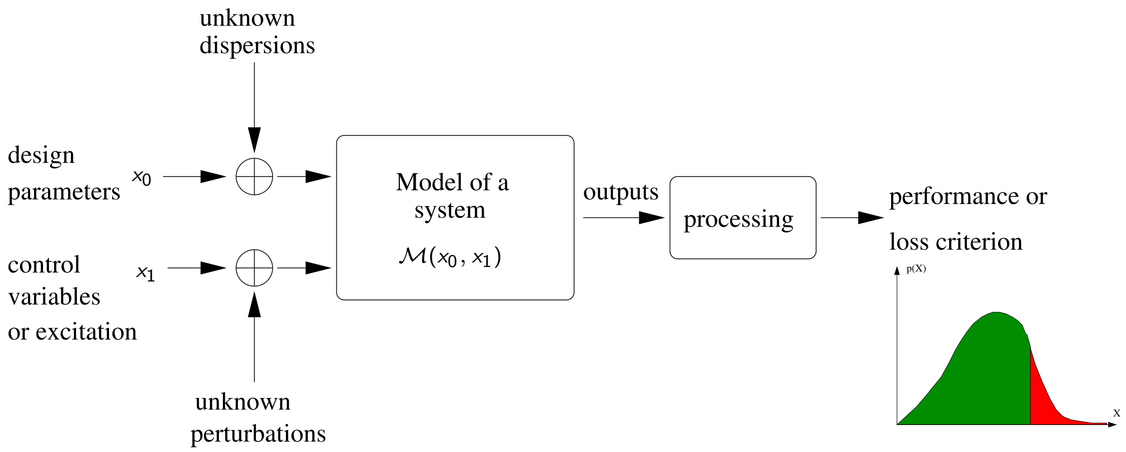
Part IV
- Sequential strategies

Outline of Lecture 1

1. Introduction

Modeling extreme events from computer simulations:
formalization of the issues at stake

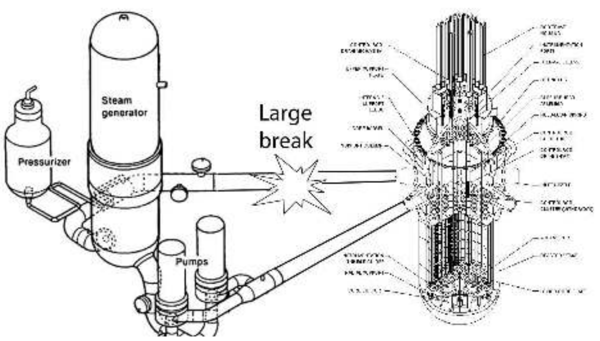
1.1 Context overview: computer experiments in engineering

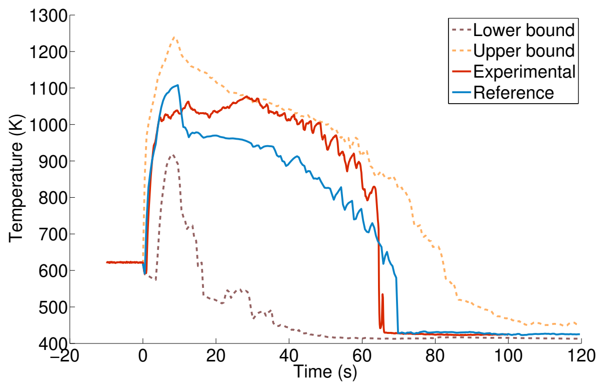


Model implemented under the form of a computer program (e.g., a finite element model).
A single run of the program may be time- and resource-consuming.

Example 1/2 – Risk analysis

- Computer simulations to assess the probability of undesirable events





(Courtesy of CEA)

- A serious accident: loss of coolant in a pressurized water nuclear reactor
- Under these conditions, temperature of fuel rods can be described by ~ 50 dimensioning factors, which are not known accurately
- Peak temperature can be estimated using complex and time-consuming simulations
- $f : \mathbb{X} \to \mathbb{R}$ peak temp. as a function of the factors
- Objective: estimate a probability of exceeding a critical value

$$\alpha = \mathsf{P}_{\mathbb{X}}\{f \geq u\}$$

or a quantile

$$q_\gamma = \inf\{u \in \mathbb{R};\ \mathsf{P}_{\mathbb{X}}\{f \leq u\} \geq \gamma\}$$

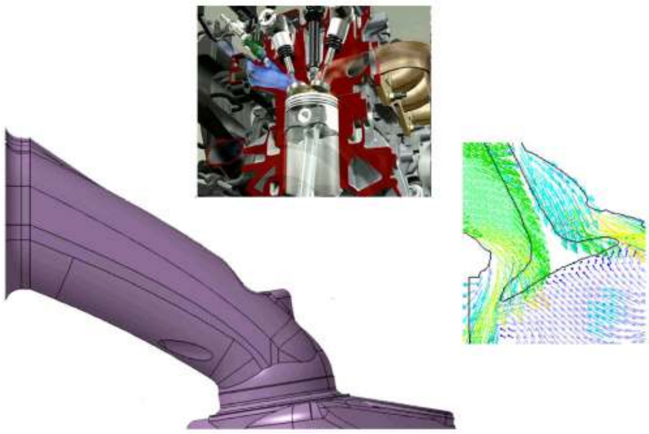
or a worst-case

$$M = \sup_{x \in \mathbb{X}} f(x)$$

Example 2/2 – Design optimization

- Computer simulations to design a product or a process, in particular
  - to find the best feasible values for design parameters (optimization problem)
  - to minimize the probability of failure of a product

- To comply with European emissions standards, the design parameters of combustion engines have to be carefully optimized
- The shape of intake ports controls airflow characteristics, which have direct impact on
  - the performances of the engine
  - emissions of NO_x and CO
- $f : \mathbb{X} \subset \mathbb{R}^d \to \mathbb{R}$ performance as a function of design parameters ($d = 20 \sim 100$)
- Computing $f(x)$ takes $5 \sim 20$ hours
- Objective: estimate $x^\star = \operatorname{argmax}_x f(x)$, or $x^\star = \operatorname{argmax}_x f(x)$ subject to $\mathrm{P}\{\text{pollutant emissions} \leq \text{threshold}\} > \gamma$



Simulation of an intake port (Navier-Stokes equations) (courtesy of Renault)

Uses of computer models in engineering

- $\mathbb{X} \subseteq \mathbb{R}^d$: input domain of the system, or *factor* (from Latin, “which acts”) space
- $f : \mathbb{X} \to \mathbb{R}$: a performance or cost function (function of the outputs of the system)
- Main classes of problems
  1. Optimization of the performances of a system, cost minimization...
  $$x^\star = \operatorname*{argmax}_{x \in \mathbb{X}} f(x)$$
  2. In presence of uncertain factors: minimize a probability of failure, i.e.,
  $$\begin{aligned} \mathbb{X} &= \mathbb{X}_0 \times \mathbb{X}_1 \\ x_0^\star &= \operatorname*{argmin}_{x_0 \in \mathbb{X}_0} \alpha(x_0) \\ \alpha(x_0) &:= \mathrm{P}_{\mathbb{X}_1}\{x_1 \in \mathbb{X}_1 : f(x_0, x_1) > u\} \end{aligned}$$
  where $\mathrm{P}_{\mathbb{X}_1}$ is some probability distribution on $(\mathbb{X}_1, \mathcal{B}(\mathbb{X}_1))$
  3. Performance assessment: estimation of a quantile
  $$q_\gamma(x_0) = \inf\{u \in \mathbb{R};\ \mathrm{P}_{\mathbb{X}_1}\{x_1 \in \mathbb{X}_1 : f(x_0, x_1) \leq u\} \geq \gamma\}$$

(This is a simplified view. Most real problems have several performance functions, and mix different objectives.)

Distinct properties of computer experiments

- The performance/cost function $f : \mathbb{X} \subseteq \mathbb{R}^d \to \mathbb{R}$ is only known through pointwise evaluations
- Most often, an evaluation of f is a deterministic experiment $\to$ one computer experiment consists in
  - choosing an $x \in \mathbb{X}$
  - running one or several deterministic computer programs to obtain the value $f(x)$
- Stochastic computer programs can also be used (noisy evaluations)
- In rare cases, ∇f may also be known
- The factor space $\mathbb{X}$ may be high-dimensional (typically $10 \sim 100$)
- Evaluation of f may be expensive (e.g., several hours) ➟ budget of experiments is limited (typically < 1000)
- f may have several local optima

- Some factors may have little influence
- f is smooth (a small number of localized discontinuities possible)
- Experts can provide a rough approximation of f (prior knowledge may be available)

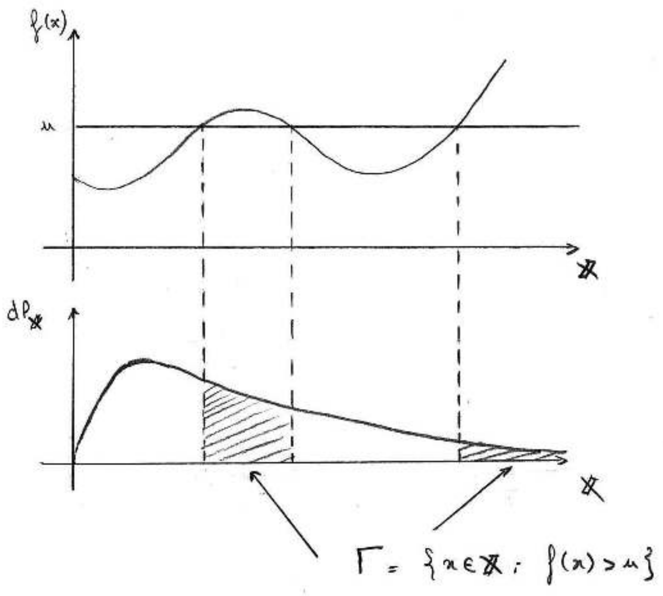
1.2 Probability of failure, quantiles: basic concepts

- Assume given
  - a domain $\mathbb{X} \subseteq \mathbb{R}^d$ (a factor space)
  - a function $f : \mathbb{X} \to \mathbb{R}$ (a loss function)
  - a distribution $\mathrm{P}_{\mathbb{X}}$ on $\mathbb{X}$ ($\mathrm{P}_{\mathbb{X}}$ models our uncertainty about the value of the factors)
  - a threshold $u \in \mathbb{R}$ (a critical value for the loss)
- The probability of failure of a system is the number

$$\begin{aligned} \alpha^u(f) &= \mathrm{P}_{\mathbb{X}}\{x \in \mathbb{X} : f(x) > u\} \\ &= \mathrm{P}_{\mathbb{X}}\{f > u\} \\ &= \int_{\mathbb{X}} \mathbb{1}_{f>u} \mathrm{d}\mathrm{P}_{\mathbb{X}} \\ &= \mathrm{E}(\mathbb{1}_{f>u}) \end{aligned}$$

- To simplify our notations: $\alpha = \alpha(f) = \alpha^u(f)$

- $\alpha(f)$ is the volume of the excursion set $\Gamma = \{x \in \mathbb{X}; f(x) > u\}$ of f above the threshold u
- 1D illustration



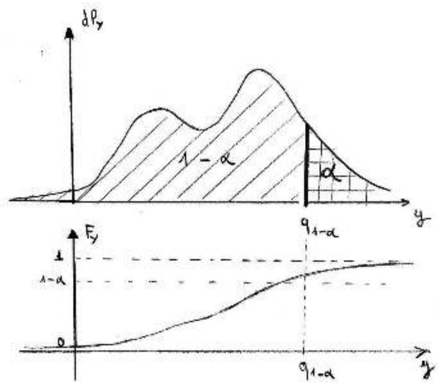
- If u is high, α is small

Quantile

- Let $(\Omega, \mathcal{B}, \mathrm{P})$ be a probability space, and consider the real-valued random variable $Y = f(X)$.

The quantile $q_{1-\alpha}$ of Y is the number

$$\begin{aligned} q_{1-\alpha} &= \inf\{u \in \mathbb{R}; \mathrm{P}\{Y \leq u\} \geq 1-\alpha\} \\ &= \inf\{u \in \mathbb{R}; F_Y(u) \geq 1-\alpha\} \end{aligned}$$



- $q_{1-\alpha}$ can be expressed directly in terms of $\mathrm{P}_{\mathbb{X}}$

$$\begin{aligned} q_{1-\alpha}(f) &= \inf\{u \in \mathbb{R}; \mathrm{P}_{\mathbb{X}}\{x \in \mathbb{X} : f(x) \leq u\} \geq 1-\alpha\} \\ &= \inf\{u \in \mathbb{R}; \alpha^u(f) \leq \alpha\} \end{aligned}$$

- How to compute a quantile? ➟ finding the largest threshold u such that the probability of failure $\alpha^u(f)$ is smaller than α, is an optimization problem
- Computing a quantile might be more difficult than computing a probability of failure

1.3 How to compute a probability of failure?

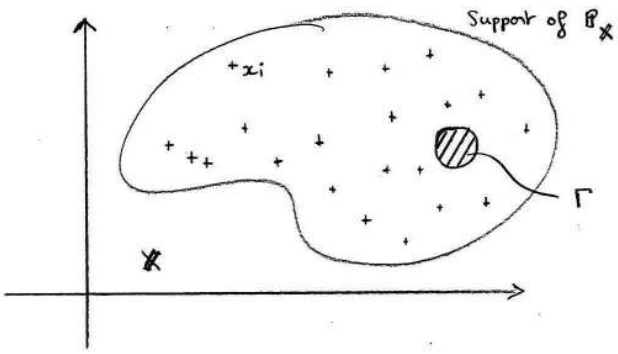
- The computation of $\alpha = \int_{\mathbb{X}} \mathbb{1}_{f>u} \mathrm{dP}_{\mathbb{X}}$ is a multidimensional integration problem
- f does not have a closed-form expression ➟ the objective is to obtain a numerical approximation of $\alpha(f)$
- Given $\phi : \mathbb{X} \to \mathbb{R} \in L^1(\mathbb{X}, \mathrm{P}_{\mathbb{X}})$, how to obtain a numerical approximation of an integral

$$\bar{\phi} = \int_{\mathbb{X}} \phi \mathrm{dP}_{\mathbb{X}}?$$

- Any numerical approximation of $\bar{\phi}$ will be constructed from a set of numerical evaluations of ϕ
  - Choose $x_1, x_2, \ldots, x_n \in \mathbb{X} \to$ get $\phi(x_1), \phi(x_2), \ldots \phi(x_n) \in \mathbb{R}$
  - Compute an approximation of $\bar{\phi}$ as a function of the evaluation results:

$$\bar{\phi}_n = h(x_1, \phi(x_1), \ldots, x_n, \phi(x_n))$$

- What makes the problem of computing α difficult?
  - The case of a probability of failure corresponds to $\phi = \mathbb{1}_{f>u} \in \{0, 1\}$
  - Whatever the integration technique, the difficulty is to choose evaluation points $x_1, \ldots, x_n$ in such a way that there is at least some points x_i for which $\phi(x_i) = 1$ (why?)
  - In practice, the volume of excursion $\alpha(f) = |\Gamma|$ is small, e.g. $\alpha(f) < 10^{-4}$



  - Γ small, unknown set ➟ a large number of function evaluations may be needed before finding at least one point in Γ
  - If f is expensive to evaluate ➟ getting an approximation of $\alpha(f) \approx 10^{-3}$ is already a challenging problem

2. Estimating a probability of failure by Monte Carlo

2.1 Monte Carlo integration with importance sampling

- Assume that $\mathsf{P}_{\mathbb{X}}$ has a density p with respect to the Lebesgue measure λ.
- Let $\phi \in L^1(\mathbb{X}, \mathsf{P}_{\mathbb{X}})$
- Let $\mathsf{Q}_{\mathbb{X}}$ be another probability distribution on $\mathbb{X}$, with density q with respect to λ, and assume that $\operatorname{supp} q \supset \operatorname{supp} p$

Proposition

Let $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} \mathsf{Q}_{\mathbb{X}}$ *be a random sample of size* $n \geq 1$.
Then

$$\bar{\phi}_n = \frac{1}{n} \sum_{i=1}^{n} \phi(X_i) \frac{p(X_i)}{q(X_i)} ,$$

is an unbiased estimator of $\bar{\phi}$. *Moreover,*

$$\operatorname{var} \bar{\phi}_n = O(n^{-1})$$

(does not depend on d*)*

- $\bar{\phi}_n$ is a weighted average (with weights $\frac{p(X_i)}{q(X_i)}$)
- NB: $\mathsf{Q}_{\mathbb{X}}$ is called an instrumental distribution

Proof:

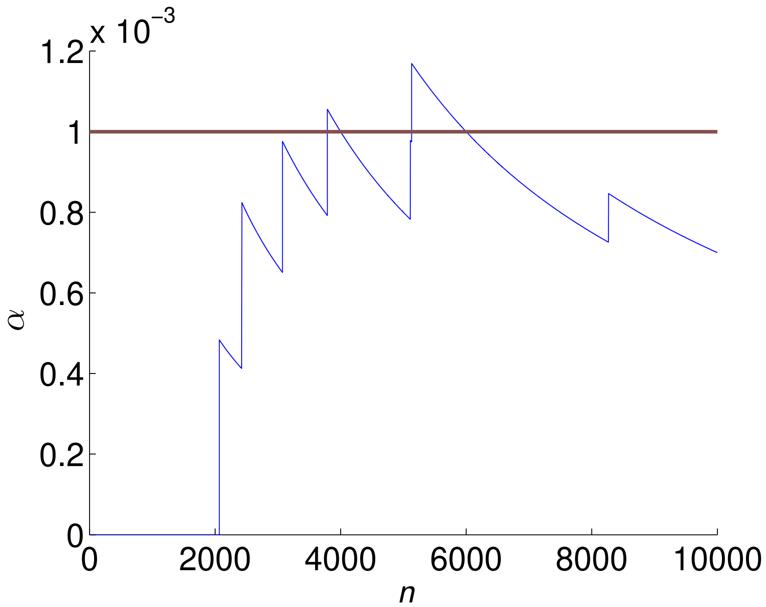
Application to the estimation of a probability of failure

- Set $\mathrm{Q}_{\mathbb{X}} = \mathrm{P}_{\mathbb{X}}$
- Let $X_1, X_2, \ldots, X_n \overset{\text{i.i.d}}{\sim} \mathrm{P}_{\mathbb{X}}$
- Then

$$\alpha_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{f(X_i)>u} = \frac{\#\{X_i; f(X_i) > u\}}{n}$$

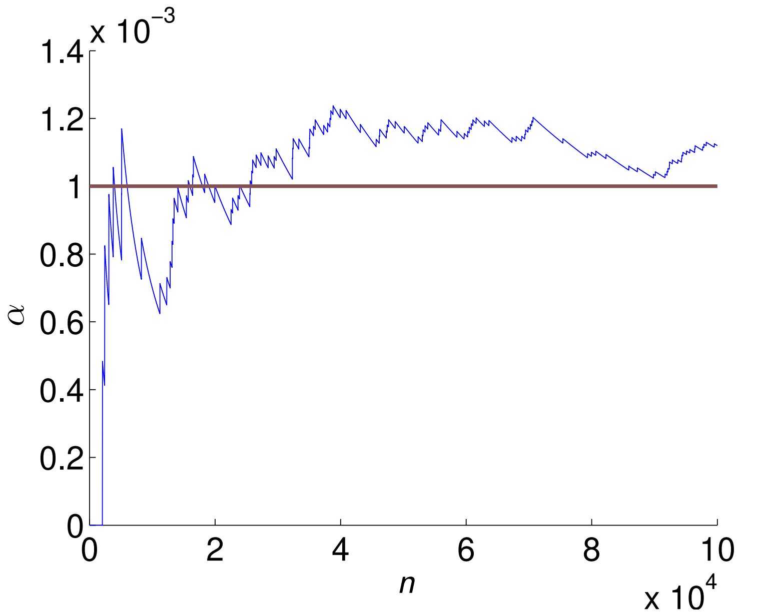
  is an unbiased estimator of α.
- In practice:
  1. Generate points $x_1, x_2, \ldots, x_n$ from $\mathrm{P}_{\mathbb{X}}$
  2. Evaluate f at $x_1, x_2, \ldots$ (may be resource- and time-consuming)
  3. Count the number of x_is such that $f(x_i) > u$ and divide by n to get an estimate of α.

Example: MC estimation of $\alpha = 10^{-3}$



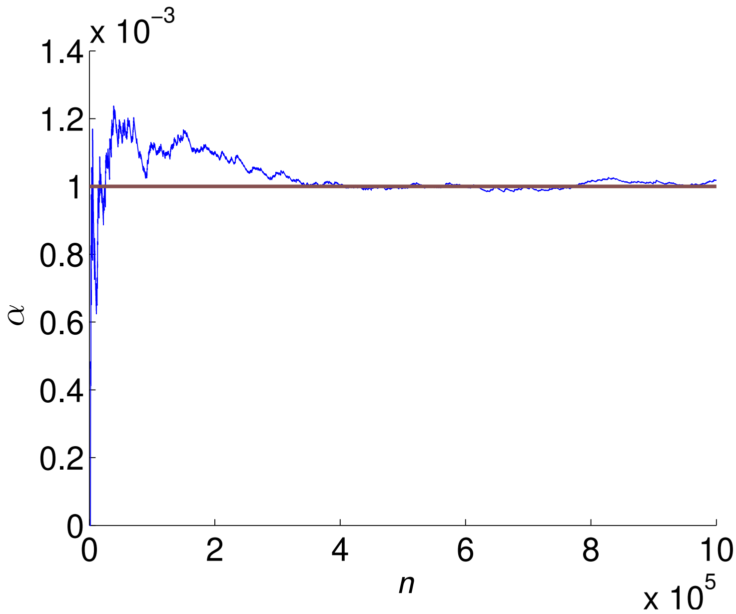
Monte Carlo estimation of a probability of failure Importance sampling

Example: MC estimation of $\alpha = 10^{-3}$



E. Vazquez Extreme events modeling I Summer School CEA-EDF-INRIA, 2011 19 / 54

Example: MC estimation of $\alpha = 10^{-3}$



How many function evaluations are needed in practice?

- Consider the MC estimator of α:

$$\alpha_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{f(X_i)>u)}$$

- The random variable $Z_i = \mathbb{1}_{f(X_i)>u}$ has distribution $\mathrm{Bernoulli}(\alpha)$, that is,

$$Z_i = \begin{cases} 0 & \text{with probability } 1-\alpha \\ 1 & \text{"} \qquad \text{"} \qquad \alpha \end{cases}$$

- Thus, $n\alpha_n \sim \mathrm{Binomial}(n, \alpha) \implies$ the variance of $n\alpha_n$ is $n\alpha(1-\alpha)$
- Thus

$$\operatorname{var} \alpha_n = \frac{\alpha(1-\alpha)}{n} \approx \frac{\alpha}{n} \text{ for } \alpha \text{ small enough}$$

- Define a notion of coefficient of variation δ as

$$\delta = \frac{\operatorname{std} \alpha_n}{\mathsf{E}(\alpha_n)}$$

- To achieve a given standard deviation $\delta\alpha$ thus requires approximately $1/(\delta^2\alpha)$ evaluations
- Examples:
  - Suppose $\alpha = 2 \times 10^{-3}$ and $\delta = 0.1$: we need $n = 50000$ evaluations.
    If one evaluation of f takes, say, one minute, then the entire estimation procedure will take about 35 days to complete!
  - Suppose $\alpha = 10^{-5}$ and $\delta = 0.1$: we need $n = 10^7$ evaluations.
    If one evaluation of f takes, say, one second, then the entire estimation procedure will take about 115 days to complete!
- When α is small, the computational cost of a MC estimation can be prohibitively high!

- Of course, we have chosen $\mathsf{Q}_{\mathbb{X}} = \mathsf{P}_{\mathbb{X}}$, and we can ask the question: what can be expected if one chooses the optimal instrumental distribution, that is, the distribution that will minimize $\operatorname{var} \alpha_n$
- What is the optimal instrumental distribution?

Proposition

The variance of the estimator

$$\bar{\phi}_n = \frac{1}{n} \phi(X_i) \frac{p(X_i)}{q(X_i)}$$

is minimum for $q = q^\star$, with $q^\star$ such that

$$q^\star(x) = \frac{|\phi(x)| p(x)}{\int_{\mathbb{X}} |\phi(y)| p(y) \mathrm{d}y}$$

Proof:

- In the case of the estimation of a probability of failure, this gives

$$q^\star(x) = \frac{p(x) \cdot \mathbb{1}_{f(x)>u}}{\alpha}$$

- NB: the normalizing constant of $q^\star$ is what we want to estimate ➟ this may not be a problem in practice (there exist sampling techniques that do not require to know the normalizing constant)
- The difficult part of the problem: $\operatorname{supp} q^\star \subset \Gamma$, which is unknown and maybe small $\Longrightarrow$ sampling from $q^\star$ can be a very difficult task

❑ Tentative conclusion:

- MC techniques to estimate a probability of failure are straightforward to implement but are generally very expensive whenever α is small
- (We shall see later that there do exist MC techniques that can still be considered, even if α is small)

3. Introduction to extreme value theory

The problem of estimation of a probability of failure revisited

- Let $X \sim \mathrm{P}_{\mathbb{X}}$ and $Z = f(X)$.
- We have

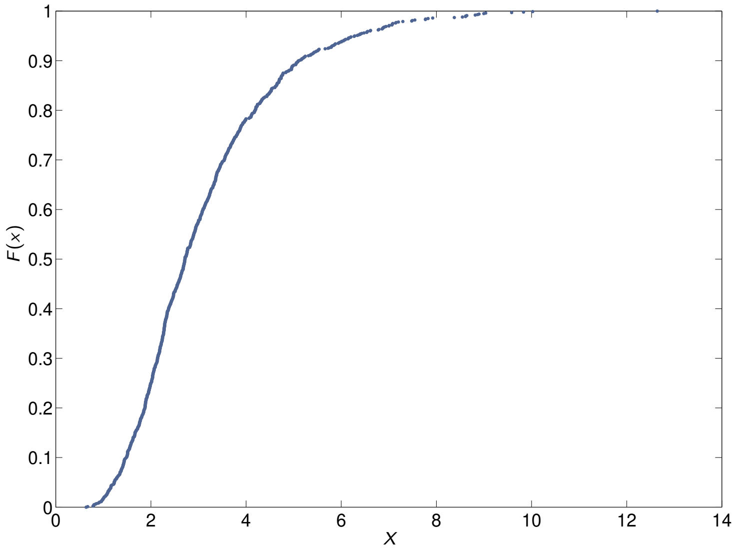
$$\alpha^u(f) = \mathrm{P}_{\mathbb{X}}\{x \in \mathbb{X}; f(x) > u\} = 1 - F_Z(u)$$

  with F_Z the cdf of Z.
- We might want to find an approximation $\widehat{F}_Z$ of F_Z, and use the plug-in estimator

$$\widehat{\alpha} = 1 - \widehat{F}_Z(u)$$

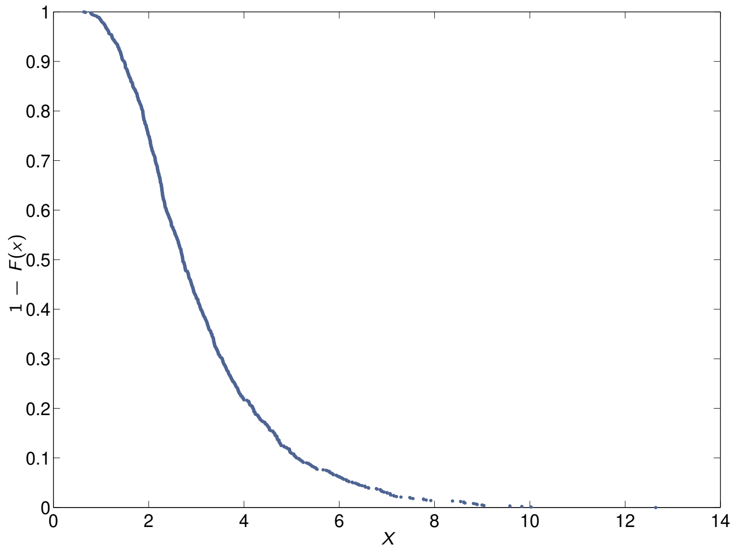
- However, since α is small, we are only interested in constructing an approximation of $u \mapsto F_Z(u)$ for high values of u, that is, when $F_Z(u) \approx 1$
- The idea is to construct an approximation of the tail of the distribution of Z

Example: $X \sim \mathrm{LN}(1, 1/2)$

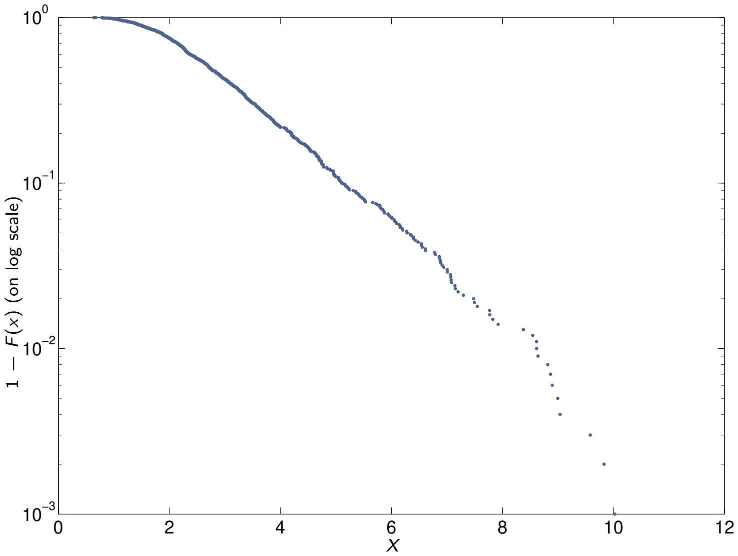


Introduction to extreme value theory Estimation of the tail of a distribution

Example: $X \sim \mathrm{LN}(1, 1/2)$

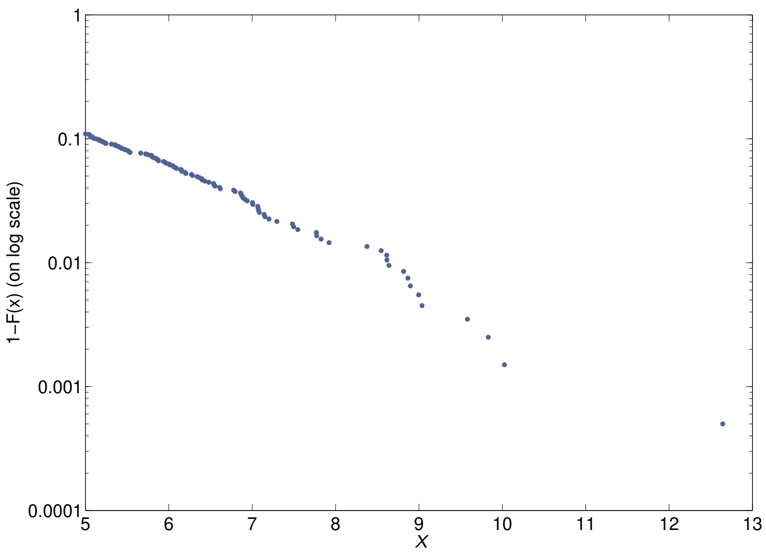


Example: $X \sim \mathrm{LN}(1, 1/2)$

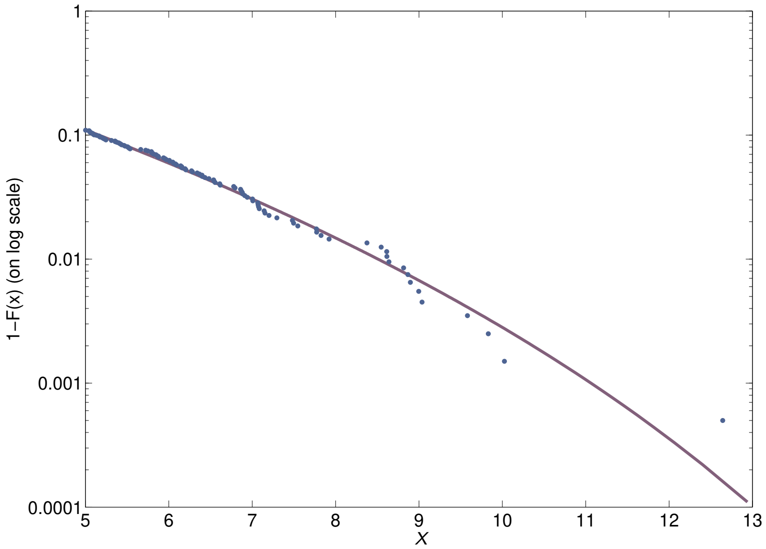


Introduction to extreme value theory Estimation of the tail of a distribution

Example: $X \sim \mathrm{LN}(1, 1/2)$



Example: $X \sim \mathrm{LN}(1, 1/2)$



Modeling the tail of a distribution from experimental data

- How to construct a model of the tail of the distribution of a real-valued random variable?
  ➟ Extreme Value Theory (EVT)
- EVT was formulated initially by Gumbel in the 50's
- EVT is a well-established branch in statistical modeling
- Applied to many fields
  - hydrology (Coles and Tawn 1996),
  - oceanography (Dawson 2000),
  - climatology (Carter and Chalenor 1981),
  - finance, insurance (Embrechts, Kluppelberg and Mikosch 1997)
  - ...
- General texts on the subject include (Resnick 1987), (Embrechts, Kluppelberg and Mikosch 1997), (Kotz and Nadarajah 2000), (Reiss and Thomas 2001), and (S. Coles 2001)

Fundamentals of EVT

- Let $X, X_1, X_2, \ldots$ be a sequence of i.i.d random variables from a distribution F
- Two main results:
  - Limit laws for maxima: for a large class of distributions F, the maximum

    $$M_n = \max X_1, \ldots, X_n$$

    converges, after proper renormalization, to a non-degenerate parametrized distribution called the Generalized Extreme Value distribution (Fisher–Tippett theorem)
  - Limit laws for excesses over a high threshold: for a large class of distributions F, the distribution of

    $$X - u \mid X > u$$

    can be approximated, for u large enough, by a parametrized distribution called the Generalized Pareto Distribution (Pickands–Balkema–de Hann theorem)
- The Fisher-Tippett theorem is the fundamental result of EVT, and can be considered as having the same status in EVT as the central limit theorem in the study of random sums
- Here, in the context of the estimation of a probability of failure, we are interested in the second result, which makes it possible to fitting a model to excesses over a threshold

Excess distribution function, mean excess function

- Excesses over thresholds play a fundamental role in EVT

Definition

Let X be a rv with cdf F and right endpoint x_F.
For $u < x_F$,

$$F_u : x \in \mathbb{R}_+ \mapsto \mathrm{P}(X - u \leq x \mid X > u)\,,$$

is called the excess cdf of X over the threshold u.

Definition

The function

$$e : u \mapsto \mathrm{E}[X - u \mid X > u]$$

is called the mean excess function of X

Example: calculation of a mean excess function

- Let $X \sim \text{Exp}(1)$
- Excess cdf, $u > 0$, $x \geq 0$,

$$\begin{aligned} F_u(x) &= \mathsf{P}(X - u \leq x \mid X > u) = \frac{\mathsf{P}(u < X \leq u + x)}{\mathsf{P}(X > u)} \\ &= \frac{\exp(-u) - \exp(-u-x))}{\exp(-u)} = 1 - \exp(-x) \end{aligned}$$

(does not depend on u ➟ memorylessness property)

- Mean excess function, $u > 0$,

$$\begin{aligned} e(u) &= \mathsf{E}[X - u \mid X > u] \\ &= \int (x-u) \mathrm{d}\mathsf{P}_{X|X>u}(x) = \int x \mathrm{d}\mathsf{P}_{X-u|X>u}(x) \\ &= \int x {F_u}'(x) \mathrm{d}x = 1 \end{aligned}$$

The Generalized Pareto Distribution

- The cdf of the GPD with shape parameter $\xi \in \mathbb{R}$, location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma > 0$, is defined as

$$G_{\xi,\mu,\sigma}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{x-\mu}{\sigma}\right), & \xi = 0 \end{cases}$$

- For $\xi \geq 0$, the support of $G_{\xi,\mu,\sigma}$ is $\mu \leq x < \infty$
- For $\xi < 0$, the support of $G_{\xi,\mu,\sigma}$ is $[0, x_F]$, with $x_F = \mu - \sigma/\xi$
- The GPD corresponds to the Pareto distribution for $\xi > 0$
- The probability density function is

$$g_{\xi,\mu,\sigma}(x) = \begin{cases} \frac{1}{\sigma}\left(1 + \xi \frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}-1}, & \xi \neq 0 \\ \frac{1}{\sigma} \exp\left(-\frac{x-\mu}{\sigma}\right) & \xi = 0 \end{cases}$$

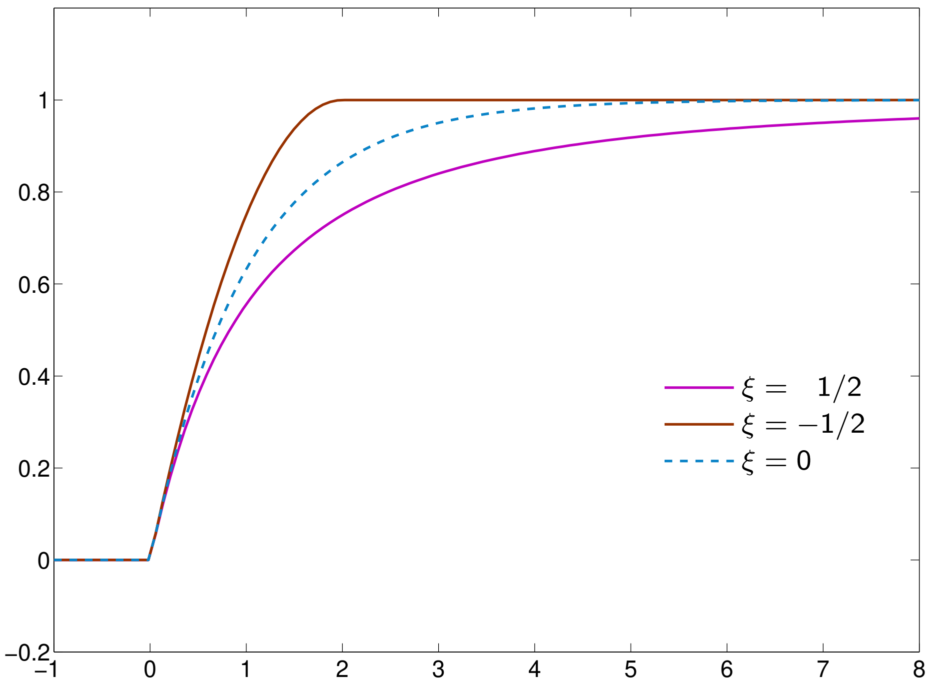
- The quantile function is

$$q_{1-\alpha} = \begin{cases} \mu - \frac{\sigma}{\xi}\left(1 - \alpha^{-\xi}\right) & \xi \neq 0 \\ \mu - \sigma \log \alpha & \xi = 0 \end{cases}$$

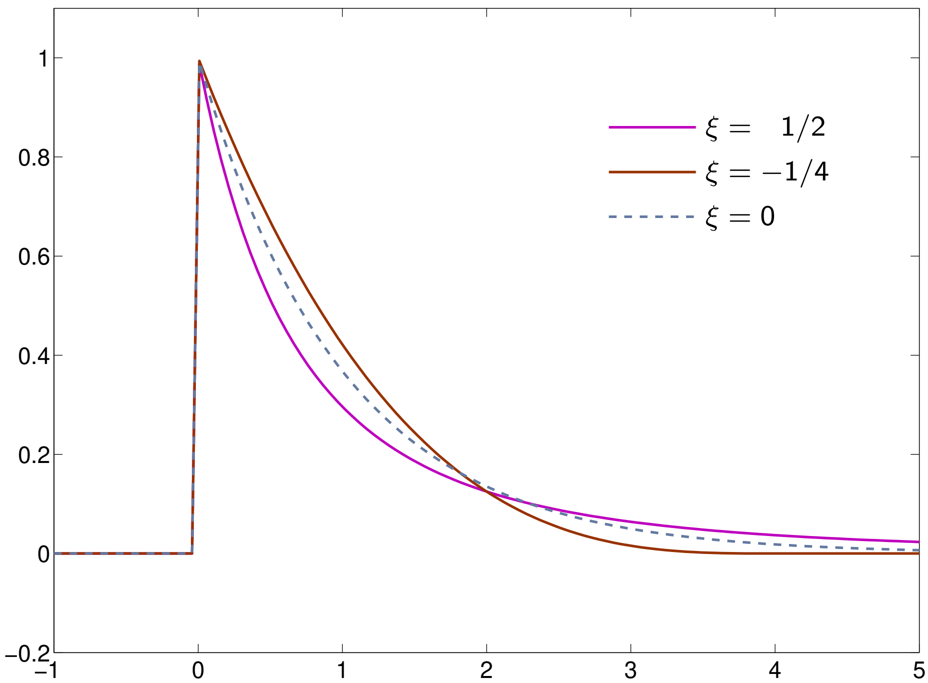
The two-parameter GPD

- In what follows, we will also consider the two-parameter GPD $G_{\xi,\sigma} := G_{\xi,0,\sigma}$
- Let $X \sim G_{\xi,\sigma}$
  For $\xi < 1$, $X \in L^1$
  $$E[X] = \frac{\sigma}{1-\xi}$$
  For $\xi < 1/2$, $X \in L^2$
  $$\mathrm{var}[X] = \frac{\sigma^2}{(1-\xi)^2(1-2\xi)}$$
- ξ high ➟ heavy tail
- ξ low/negative ➟ light tail

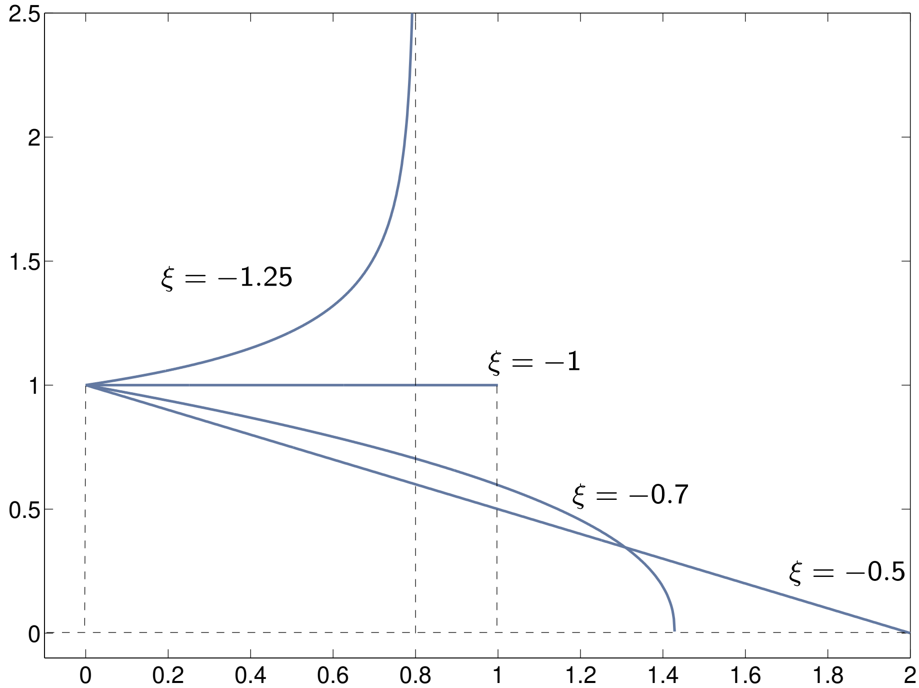
GPD for different shape parameters ξ and $\sigma = 1$



Densities of the GPD for different shape parameters ξ and $\sigma = 1$



Densities of the GPD for different shape parameters $\xi < 0$ and $\sigma = 1$



Properties of the GPD

- Log-transform:
  Let $X \sim \mathrm{GPD}(\xi, \sigma)$, then
  $$Y = \frac{1}{\xi} \log\left(1 + \frac{\xi}{\sigma} X\right)$$
  has a standard exponential distribution
- Stability with respect to excess-over-threshold operations
  $$X \sim \mathrm{GPD}(\xi, \sigma) \implies Y = X - u \mid X > u \sim \mathrm{GPD}(\xi, \sigma + \xi u)$$
  It is important to notice that the operation does not affect the shape parameter; it only alters the scale parameter of the distribution.
- the mean-excess function of the GPD
  If $X \sim \mathrm{GPD}(\xi, \sigma)$, with $\xi < 1$, the mean excess function $e(u) = \mathrm{E}(X - u \mid X > u)$ is given by
  $$e(u) = \frac{\sigma}{1-\xi} + u\frac{\xi}{1-\xi}, \quad u \leq x_F$$
  (The mean excess function is affine. Note that e is decreasing if $\xi < 0$, constant if $\xi = 0$, and increasing if $\xi > 0$)

Tail approximation

- GPD fitting is one of the most useful concept of EVT
- The Pickands–Balkema–de Haan theorem:
  For a large class of distributions, the conditional cdf F_u of the excesses over u can be approximated by the GPD, as $u \to x_F$
  $$\lim_{u \to x_F} \sup_{0<x<x_F-u} |F_u(x) - G_{\xi,\sigma(u)}(x)| = 0$$
  with
  $$F_u(y) = \mathrm{P}(X - u \leq y \mid X > u) = \frac{F(u+y) - F(u)}{1 - F(u)}, 0 \leq y \leq x_F - u$$
- Recall that raising the threshold of the GPD only changes the scale parameter of the GPD
- By setting $y = x - u$, we obtain, for $u \leq x \leq x_F$,
  $$F(x) = [1 - F(u)]F_u(x-u) + F(u)$$
- Since F_u converges to the GPD, for sufficiently large u, we obtain the approximation
  $$F(x) \approx [1 - F(u)]G_{\xi,\sigma(u)}(x-u) + F(u), \quad \text{for } u \leq x \leq x_F$$

❑ We want to take advantage of the approximation

$$F(x) \approx [1 - F(u)] G_{\xi,\sigma(u)}(x - u) + F(u)\,, \quad \text{for } u \leq x \leq x_F$$

❑ The EVT methodology:

- choose an appropriate value of the high threshold u
- replace $F(u)$ by an empirical estimator
- estimate the parameters ξ and $\sigma(u)$ of the GPD
- use the approximation of the distribution tail to estimate a small probability of failure, or a high quantile

Choice of high threshold

- How to select an adequate threshold above which it is appropriate to use the GPD?
- The choice of has been extensively addressed in the extremes literature ➟ see, e.g., de Haan (1990) and Beirlant, Teugels and Vynckier (1996)
- A difficult issue:
  - threshold too low generally increases the bias of the parameter estimators
  - threshold too high increases the variance of the parameter estimators, due to the reduced size of the corresponding sample of excesses.
- The MEF can assist a user in the search for an adequate threshold
  Recall that for a GPD

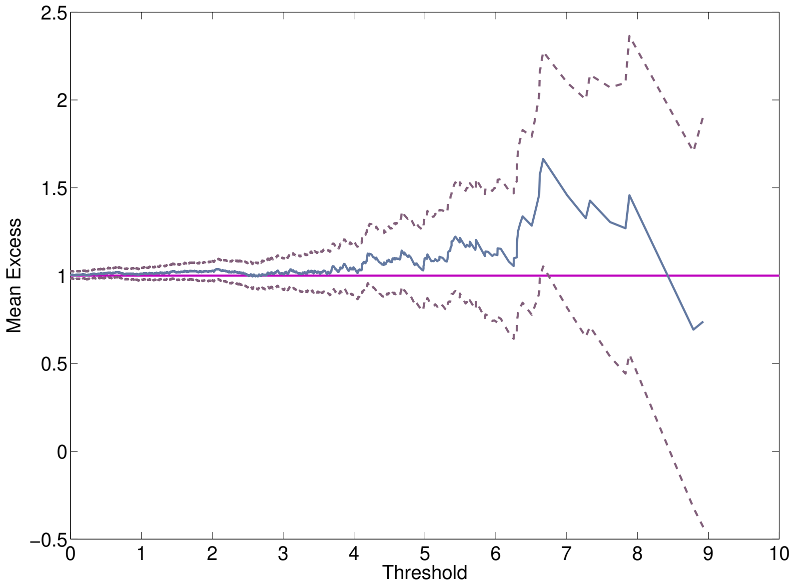
$$e(u) = \frac{\sigma}{1-\xi} + u\frac{\xi}{1-\xi}\,, \quad u \leq x_F$$

  Idea: compute an empirical estimate e_n, and check a region where the graph of e_n becomes roughly affine
- For a random sample of size n, the empirical MEF may be written as

$$e_n(u) = \frac{\sum_{i=1}^n (X_i - u)\mathbb{1}_{X_i > u}}{N_u}$$

  with $N_u = \sum_{i=1}^n \mathbb{1}_{X_i > u}$

Example: $X \sim \mathrm{Exp}(1)$

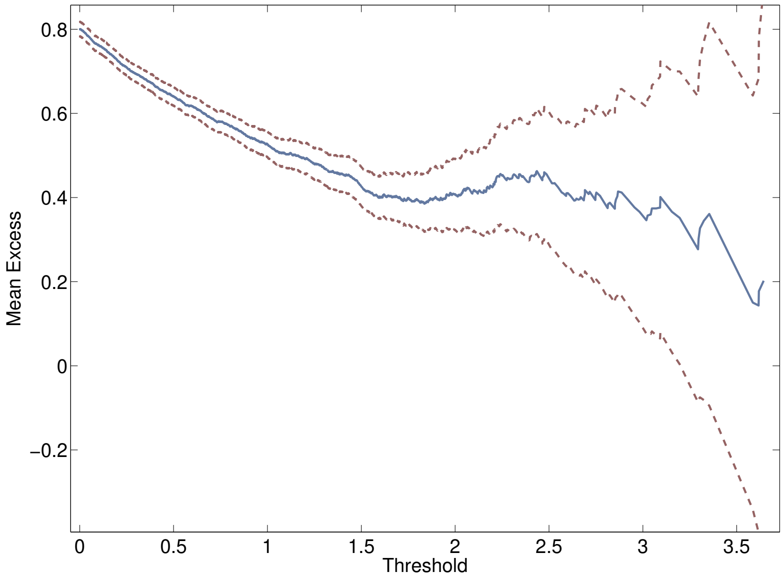


Choice of a threshold

- The statistical properties of e_n can be derived by using empirical process theory
- e_n is a useful tool to distinguish between light- and heavy-tailed models
- However, caution is called for when using such plots: due to the sparseness of the data available for large values of u, the plots are very sensitive to changes in the data towards the end of the range
- In choosing a threshold such that e_n is approximately affine, the key difficulty lies in the interpretation of "approximately"

  ➟ the user should never expect a unique choice of u to appear
- More robust versions like median excess plots have been suggested

Example: case of truncated normal $X \sim \sqrt{\frac{2}{\pi}} \exp(-x^2/2)$, $x \geq 0$,
The mean excess function is $e(u) \overset{u\to\infty}{\sim} u^{-1}(1+o(1))$



X belongs to the maximal domain of attraction of the Gumbel distribution ($\xi = 0$)

GPD fitting

- Having chosen a sufficiently large threshold u, we have the approximation

$$F(x) \approx [1 - F(u)] G_{\xi,\sigma(u)}(x-u) + F(u)\,, \quad \text{for } u \leq x \leq x_F$$

- $F(u)$ can be estimated by the empirical estimator

$$(n - N_u)/n$$

  where N_u is the number of observations above u, and n is the sample size
- An estimator of the tail of the cdf is therefore given by

$$\widehat{F}(x) = 1 - \frac{N_u}{n}\left(1 + \widehat{\xi}\frac{x-u}{\widehat{\sigma}}\right)^{-\frac{1}{\widehat{\xi}}}, \quad \text{for } u \leq x \leq x_F$$

  where $\widehat{\xi}$ and $\widehat{\sigma}$ are estimates of ξ and σ
- How to estimate ξ and σ?

Maximum likelihood estimation of ξ and σ

- Let $X_1, \ldots, X_n$ be a random sample of size n from a $GPD(\xi, \sigma)$.
- The logarithm of the likelihood can be expressed as

$$\ell(\xi, \sigma; X) = \begin{cases} -n \log \sigma - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^{n} \log\left(1 + \frac{\xi}{\sigma} X_i\right), & \xi \neq 0 \\ -n \log \sigma - \frac{1}{\sigma} \sum_{i=1}^{n} X_i, & \xi = 0 \end{cases}$$

- ML estimators exist only for $\xi \geq -1$ (for $\xi < -1$, the log-likelihood tends to ∞ as the maximum observed value $X_{(n)}$ approaches $-\sigma/\xi$)

Maximum likelihood estimation of ξ and σ

- Analytical maximization of the log-likelihood is not possible (numerical techniques are required, taking care to avoid numerical instabilities for $\xi \approx 0$)
- Asymptotic properties of the ML estimators of the GPD parameters, such as consistency, normality and efficiency, can be established for $\xi > -1/2$
- For $\xi > -1/2$,

$$n^{1/2}\left(\widehat{\xi}_n - \xi, \frac{\widehat{\sigma}_n}{\sigma} - 1\right) \xrightarrow{d} \mathrm{N}(0, \Sigma), \quad n \to \infty$$

with

$$\Sigma = (1 + \xi) \begin{pmatrix} 1 + \xi & -1 \\ -1 & 2 \end{pmatrix}$$

- The ML estimation of the GPD parameters can be a quite difficult task, even for $\xi \geq -1/2$
- Indeed, the algorithms used for computing the ML estimates can exhibit convergence problems, even for large sample sizes
- Note that other estimators of the parameters of the GPD are available in the literature

Threshold choice revisited

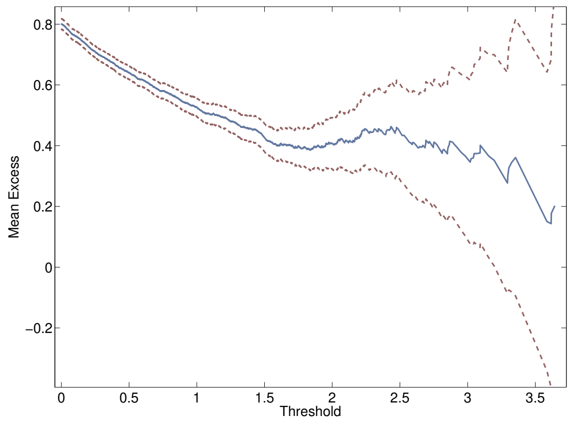
- We saw that choosing a high threshold based on the mean excess function can be difficult
- A complementary technique consists in fitting the GPD for different thresholds u, and to look for stability of parameter estimates
- The argument is based on the fact that if a GPD is a reasonable model for excesses of a threshold u_0, then excesses of a higher threshold u_1 should also follow a GPD, with the same shape parameter ξ
- However, denoting by $\sigma(u)$ the value of the scale parameter for a threshold $u > u_0$

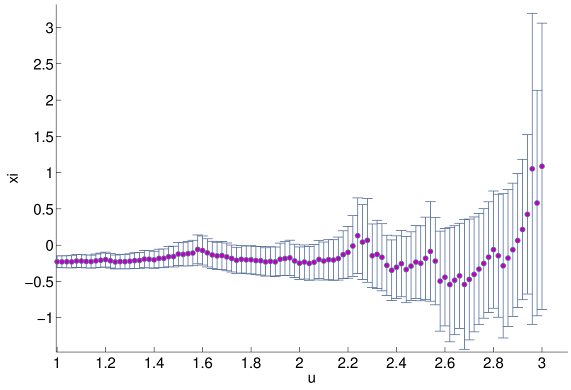
$$\sigma(u) = \sigma(u_0) + \xi(u - u_0)$$

so that the scale parameter changes with u unless $\xi = 0$
- The idea is then to plot σ and $\sigma^\star = \sigma(u) - \xi u$ with respect to u

Threshold choice revisited

Example: case of truncated normal $X \sim \sqrt{\frac{2}{\pi}}\exp(-x^2/2)$, $x \geq 0$





Estimation of a probability of failure

- Consider again the problem of the estimation of a probability of failure: let $X \sim \mathrm{P}_{\mathbb{X}}$ and $Z = f(X)$
- Given a threshold u, our objective is to estimate

$$\alpha^u(f) = 1 - F_Z(u)$$

by substituting $F_Z(u)$ with the approximation

$$\widehat{F}_Z(u) = 1 - \frac{N_{u_0}}{n}\left(1 + \widehat{\xi}\frac{u - u_0}{\widehat{\sigma}}\right)^{-\frac{1}{\widehat{\xi}}},$$

where u_0 is a high threshold chosen such that $u_0 < u$, N_{u_0} is the number of observations above u_0, and $\widehat{\sigma}$ and $\widehat{\xi}$ are estimates of the parameters of a GPD
- We obtain

$$\widehat{\alpha}^u(f) = \frac{N_{u_0}}{n}\left(1 + \widehat{\xi}\frac{u - u_0}{\widehat{\sigma}}\right)^{-\frac{1}{\widehat{\xi}}}$$

Confidence intervals

- Asymptotic standard errors, or asymptotic confidence intervals for $\widehat{\alpha}^u(f)$ can be derived from the delta method or bootstrap
- Delta method: Let T_n be an estimator of $\theta \in \mathbb{R}^d$, and let $\phi : \mathbb{R}^d \to \mathbb{R}$ be a differentiable function. If

$$\sqrt{n}(T_n - \theta) \to_d \mathrm{N}(0, \Sigma)$$

then

$$\sqrt{n}(\phi(T_n) - \phi(\theta)) \to_d \mathrm{N}(0, [\nabla\phi(\theta)]^{\mathsf{T}}\,\Sigma\,[\nabla\phi(\theta)])\,.$$

- The asymptotic distribution of $(\widehat{\xi}, \widehat{\sigma})$ has been given above
- The random variable N_u follows the binomial distribution $\mathrm{Binomial}(n, 1 - F(u))$, so that the estimator N_u/n of $[1 - F(u)]$ has variance $F(u)(1 - F(u))/n$
- Thus the complete asymptotic variance-covariance matrix for $(N_u/n, \widehat{\xi}, \widehat{\sigma})$ is $\frac{1}{n}\Sigma$, with

$$\Sigma = \begin{pmatrix} F(u)(1 - F(u)) & 0 & 0 \\ 0 & (1+\xi)^2 & -\sigma(1+\xi) \\ 0 & -\sigma(1+\xi) & 2\sigma^2(1+\xi) \end{pmatrix}$$

- The gradient of $\phi(\eta, \xi, \sigma) = \eta\left(1 + \xi\frac{u - u_0}{\sigma}\right)^{-\frac{1}{\xi}}$ can be easily expressed in closed form

Quantile estimation

- Similarly, an estimator of the quantile $q_{1-\alpha}$ is given by

$$\widehat{q}_{1-\alpha} = u + \frac{\widehat{\sigma}}{\widehat{\xi}}\left(\left(\frac{n}{N_u}\alpha\right)^{-\widehat{\xi}} - 1\right)$$

- Furthermore, for $\widehat{\xi} < 0$, an estimator of the right endpoint of F is given by

$$\widehat{x}_F = u - \frac{\widehat{\sigma}}{\widehat{\xi}}$$

Model checking

- We have seen that EVT allows us to extrapolate outside the range of available data
- In some sense, EVT is making the best use of available data for making inference about extreme values
- However, always be reluctant when it comes to estimate very small probabilities $\rightarrow$ the statistical reliability of these estimates are very difficult to judge in general
- Careful model checking, using cross-validation for instance, is always necessary to check the stability of the estimates
- Questions about the influence of single of few observations and model-robustness can be analyzed using simulations

Some references

- Beirlant, J., Teugels J.L., and Vynckier P. (1996). "Tail Index Estimation, Pareto Quantile Plots, and Regression Diagnostics" Journal of the American Statistical Association 91: 1659-1667
- Carter, D.J.T., and Chalenor, P.G. (1981) "Estimating return values of environmental parameters" Quarterly Journal of the Royal Meteorological Society 107: 259-266
- Coles, S.G, and Tawn, J.A (1996). "Modelling extremes of the areal rainfall process" Journal of the Royal Statistical Society B 58: 329-347
- Coles, S.G. (2001) An Introduction to Statistical Modeling of Extreme Values. Springer
- Dawson, T.H. (2000) "Maximum wave crests in heavy seas" Journal of Offshore Mechanics and Artic Engineering - Transactions of the AMSE 122: 222-224
- de Haan, L. (1990) "Fighting the arch-enemy with mathematics." Statistica Neerlandica 44: 45-68
- Embrechts, P., Kluppelberg C., and Mikosch T. (1997). Modelling Extremal Events for Insurance and Finance. Springer-Verlag
- Gumbel, E.J. (1958) Statistics of Extremes. Columbia University Press
- Kotz, S. and Nadarajah S. (2000) Extreme Value Distributions: Theory and Applications. World Scientific Publishing Company
- Pickands, J. (1975) "Statistical inference using extreme order statistics" Annals of Statistics 3: 119-131
- Reiss, R.D., and Thomas M. (2001) Statistical analysis of extreme values from insurance, finance, hydrology and other fields. Birkhäuser
- Resnick, S.I. (1987) Extreme Values, Regular Variation and Point Processes. Springer Verlag

Summing up

- The estimation of small probabilities of failure is a difficult task
- If the evaluation of the performance function f is time-consuming, estimating a probability of failure $\alpha \approx 10^{-3}$ is already challenging
- EVT is a useful tool: makes the best use of available data for analyzing extreme events