The Mathematics of Risk

Paul Embrechts

Department of Mathematics

Director of RiskLab, ETH Zurich

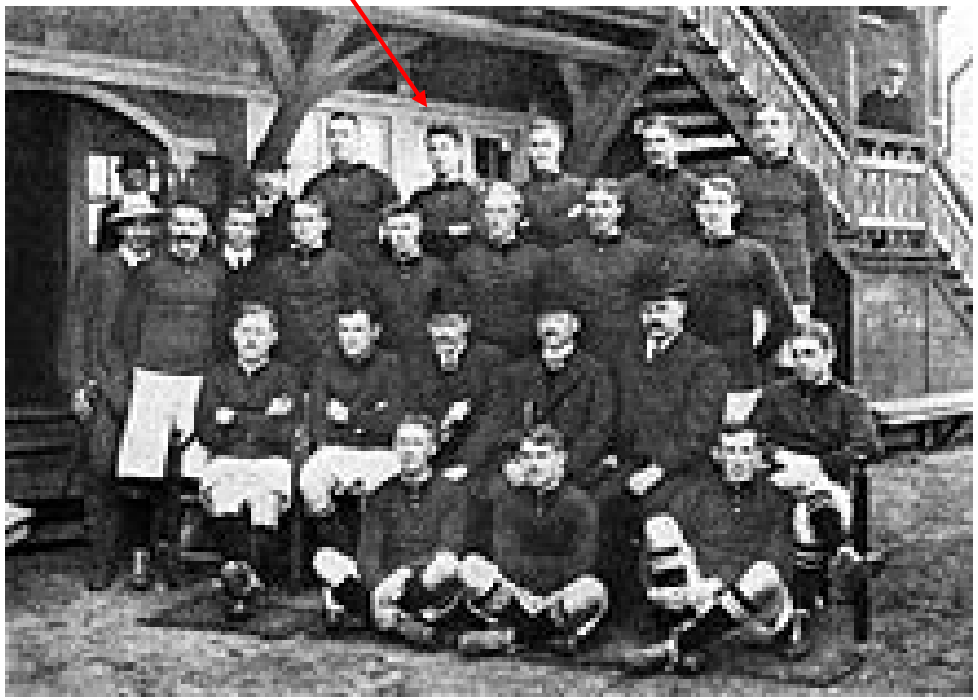
Senior SFI Chair

www.math.ethz.ch/~embrechts

The Harald Bohr Lecture

Copenhagen, October 10, 2017

Harald August Bohr: Danish National Football Team
1908, Olympic Games in London (*)



Opening game: Denmark - France B : 9 – 0, HB scored twice!
Semi final: Denmark - France A : 17 – 1, till today, **Olympic record**!
Final: Great Britain - Denmark: 2 – 0, Denmark Olympic Silver Medal!

Informal Analytic Number Theory Seminar – D-MATH ETH Zurich Kowalski Emmanuel (e-mail: Sunday 8 October, 2017, 18:11)

Dear all,

To avoid a time conflict, the Informal Analytic Number Theory seminar has been moved to Mondays, 16:15, in HG G 19.2.

The first talk in this new slot will end a bit before to give time to attend the Einführungsvorlesung of V. Tassion. **I will finish the discussion of the Bohr-Pál Theorem.**

Best wishes,

Emmanuel

A more realistic title would be:

On some Mathematics of Risk

I like to start on a personal note:

31. Jan. 1953 – 1. Feb. 1953 flooding at Dutch coast

- 1836 people killed
- 72’000 people evacuated
- 49’000 houses and farms floaded
- 201’000 cattle drowned
- 500 km coastal defenses destroyed; more than 400 breaches of dikes
- 200’000 ha land floaded

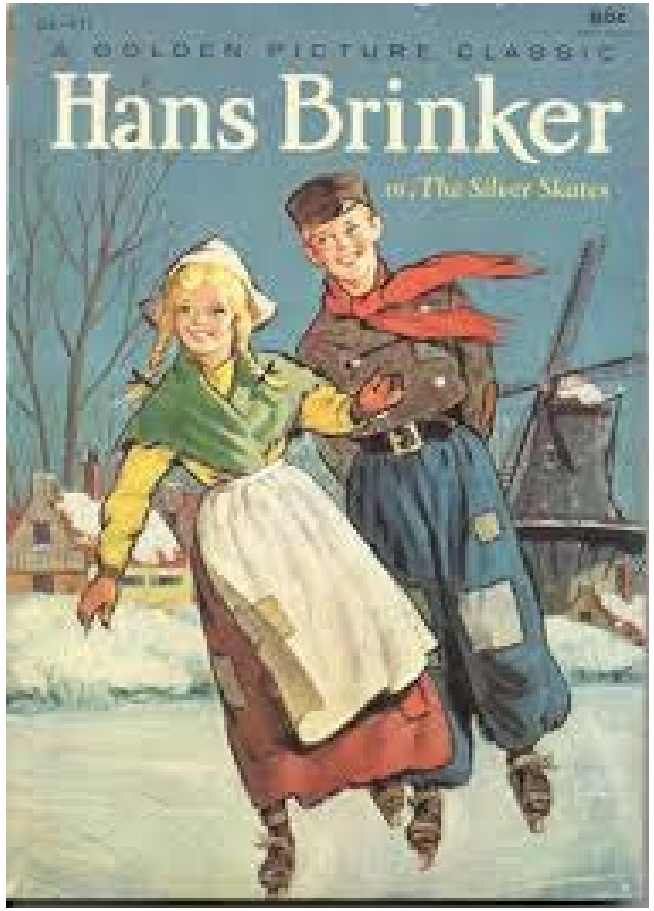
Two days later, on **February 3, 1953** I was born in Schoten, near **Antwerp** (which was also partly floaded):

Schoten

How can we protect the population from such events?

For instance by **building dikes**!

Like in **Mary Mapes Dodge: «Hans Brinker or The Silver Skates», 1865?**



If only!

But more by building dikes like:

(Maeslantkering)

The Delta-Project

- Coastal flood-protection
- Requested dike height at l: $h_d(l)$
- Safety margin at l: **MYSS**(l) = **M**aximal **Y**early **S**ea **S**urge at l
- Probability(**MYSS**(l) > $h_d(l)$)

  should be „small“, whereby „small“ is defined as (Risk):

  - **1 / 10’000** in the Randstad
  - 1 / 250 in the Deltaregion to the North
  - Similar requirements for rivers, but with 1/10 – 1/100

- For the Randstad (Amsterdam-Roterdam):

Dike height = Normal-level (= NAP) + 5.14 m

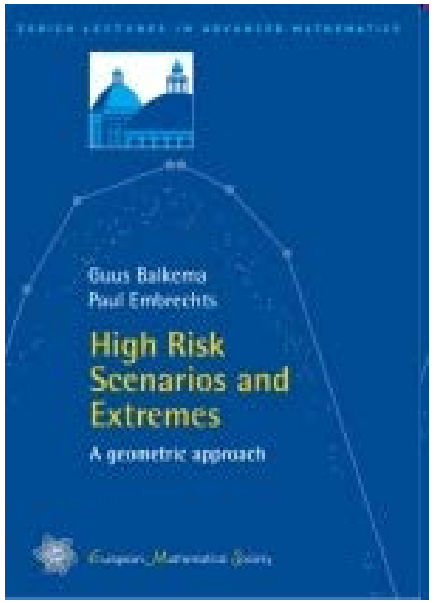
Dike-height = NAP + 5.14 m

1953: NAP + 4.5 m

PE

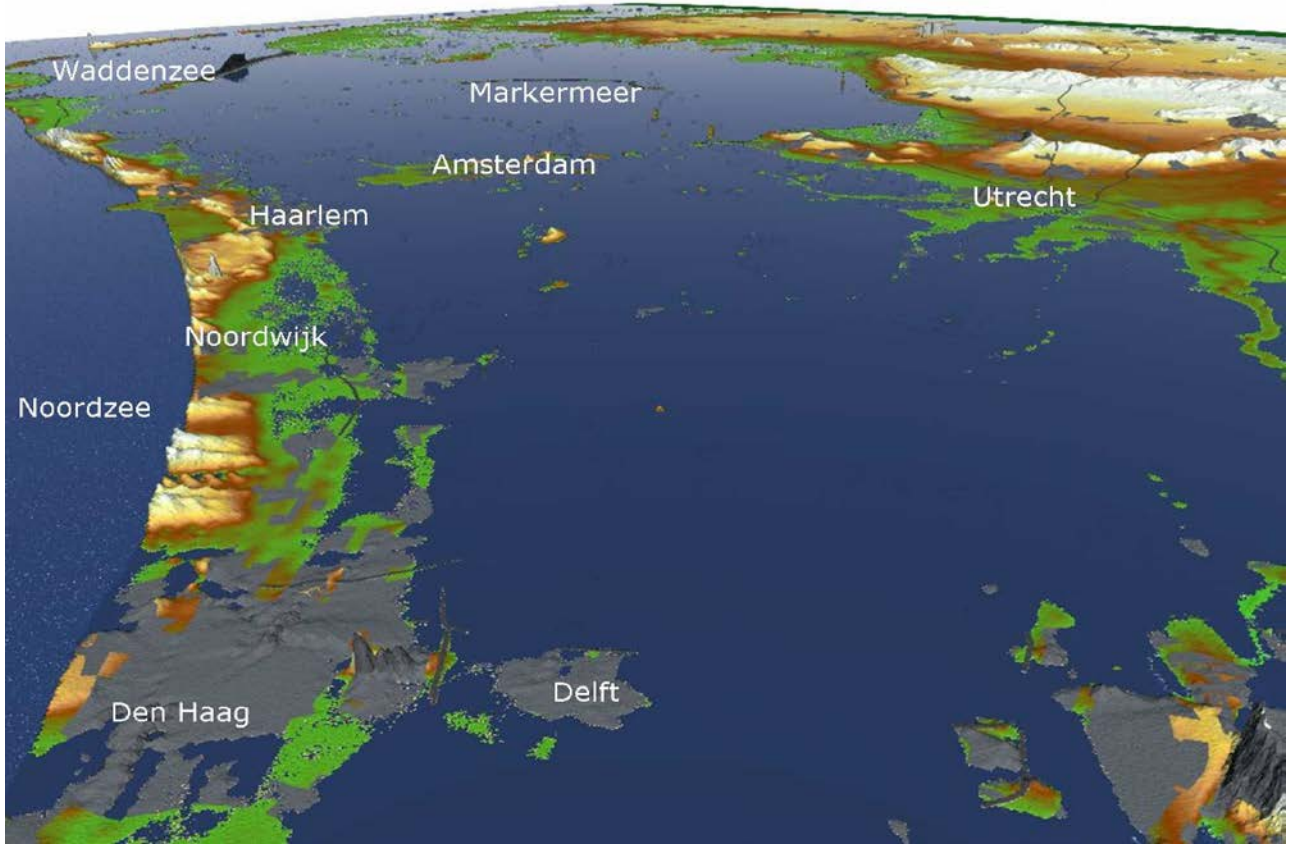
NAP

Guus Balkema



Oosterschelde dike, “Neeltje Jans”

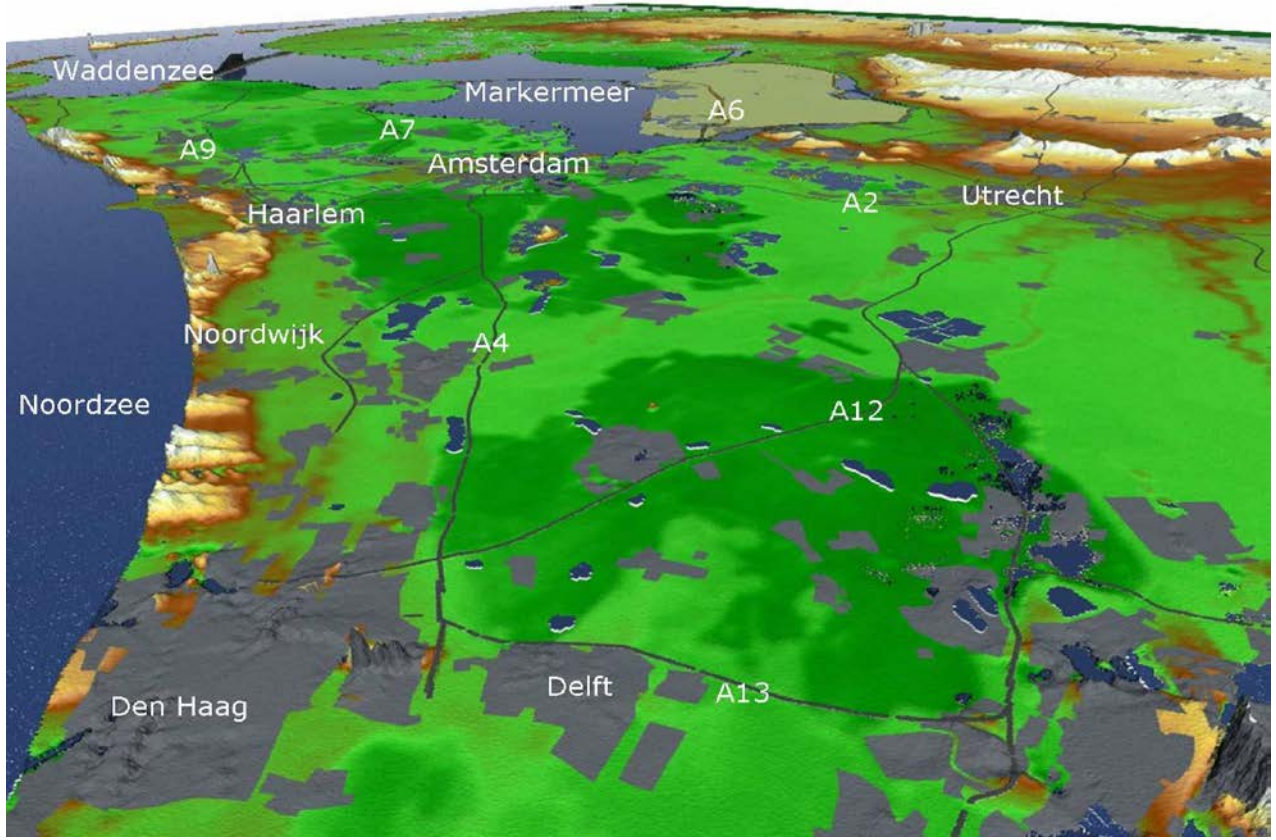
The Netherlands **without** dikes!



26% area below sea level
with about a population of 60%

Henk van den Brink, KNMI, Fighting the arch-enemy with **mathematics** and **climate models**

The Netherlands **with** dikes:



(Henk van den Brink, KNMI)

Remark: **1/10’000** → 1/100’000

And the situation in Denmark?

- https://www.climatechangepost.com/denmark/coastal-floods/
- **Copenhagen**, though overall **well protected**: "The level of flood **protection** in Copenhagen is based on a water level of **150 cm above the mean sea level**, which is associated to **a return-period of 120 years**. The level of protection will have to be upgraded in case of sea level rise: even **with only a 25 cm sea level rise, the city protection level would decline from 1-in-120 years to 1-in-10 years**."
- "Past experience demonstrates that the retrofit of coastal defense structures is **a lengthy process** requiring forward thinking and planning. For example, there was **a 30 year lag** between the decision to build the Thames barrier and its actual implementation**. It is necessary to start thinking about long-term adaptation in coastal cities today, even if the risks of climate change are not imminent**."
- Damage protection against **excessive rain** … 14/8/10, 2/7/11 … !

We have unfortunately seen many more examples since!

- Lothar, Katrina (*), Matthew
- Ike, Sandy (*)
- Harvey, Irma, Maria, …
- Sendai, Tohoku Earthquake (*)
- 2011 Thai flood
- …

… as well as financial “storms” on Wall Street:

FORECAST: MORE FLOODING IN NEW YORK

Some methodological ingredients **from a mathematical point of view:**

- Data: historical, expert, simulation, hence **(mathematical) statistics**
- Numerically simulated climate models (so-called ensembles): **numerical analysis** (Navier-Stokes equations)
- One is interested in yearly (say) maximal storm surges/heights, not the yearly **average** one, hence **Extreme Value Theory** (EVT) and the stochastic modelling of rare events enter
- Risk is interpreted as a 1 in 10’000 year event, hence **the theory of risk measures** plays a crucial role
- **Interdisciplinarity** is a fact and excellent **communication** skills are absolutely indispensable … e.g. L’Aquila!
- …

An important research distinction (very broadly)

- All fields of application need a proper **definition** of risk
- All fields of application need statistical **estimation** of risk measures
- **(EE)** For **Environmental/Engineering** applications one typically faces one random variable (r.v.), of course based on measurements of typically multivariate stochastic processes, or measurements of several such rvs for different sites leading to statistical estimation**:** **STATISTICS**
- **(EF)** For applications in **Economics/Finance** one typically combines measures of risk across different rvs (e.g. portfolio factors) in a functional way leading to risk aggregation/diversification/allocation: **FUNCTIONAL ANALYSIS**

As a consequence, classical mathematical theorems (e.g.!) in use in **(EE)** and **(EF)** differ:

(EE) The **Fisher-Tippett-Gnedenko** Theorem

The **Pickands-Balkema-de Haan** Theorem, **etc …**

from the realm of Extreme Value Theory (EVT)

(EF) The **Hahn-Banach** Theorem **(*)**

The **Fenchel-Moreau** Theorem, **etc …**

from the realm of Functional Analysis and Convex Analysis

(*) during the financial crisis some banks actually "achieved"

the financial equivalent of the Banach –Tarski paradox!

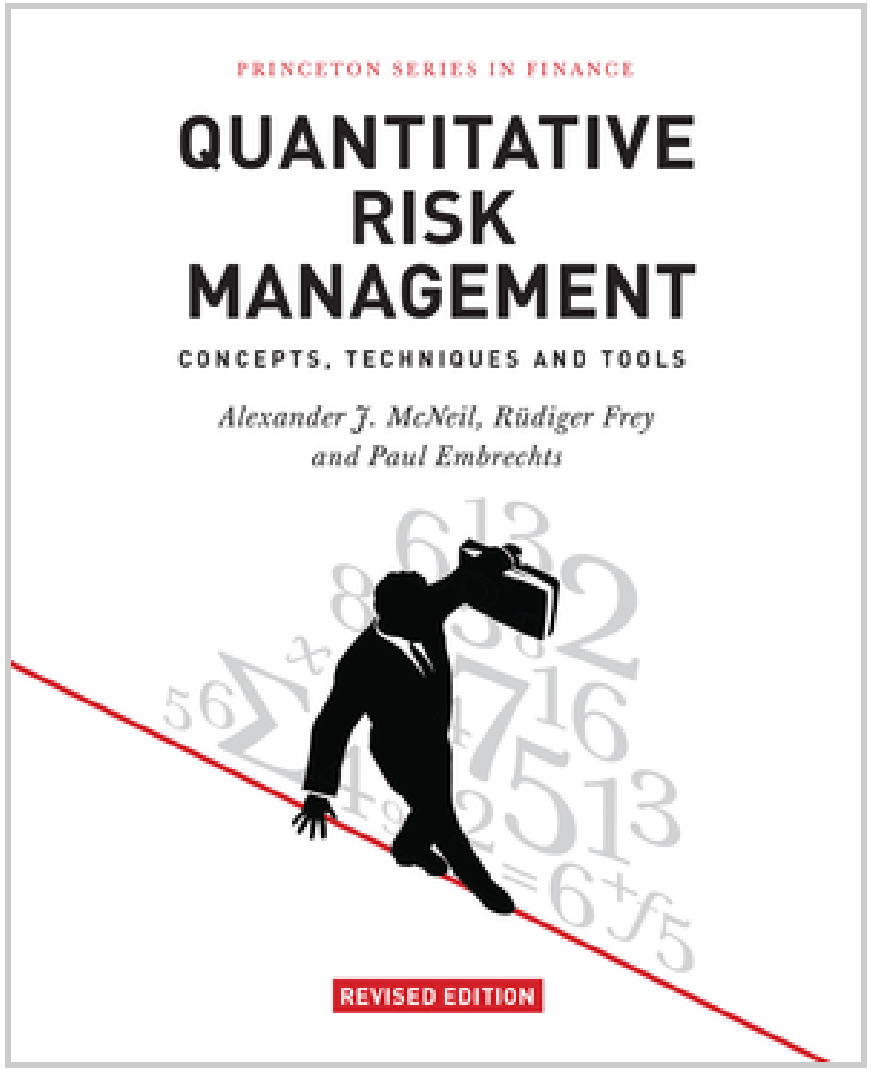
Werner Fenchel
(1905-1988)

Fenchel's Theorem (1929): The total curvature of a space curve is greater than or equal to 2π with equality only for convex curves.
Proof of Konrad Voss (*) (1955): The total curvature of a space curve is one halve the total absolute Gaussian curvature of a circular tube around the curve.
(*) Konrad Voss (ETH): 1928 – 2017

In QRM the relevant result is Fenchel's Duality Theorem from convex analysis.

A sample of mathematical results from:

(2005, 2015)



(700 pages)

Taken from the support-website

www.qrmtutorial.org

Containing:

- Book(s) info
- **Full course pdf slides (~ 800)**
- All R-programs
- Extra course information
- Videos
- News, etc.

2.3.2 Value-at-risk

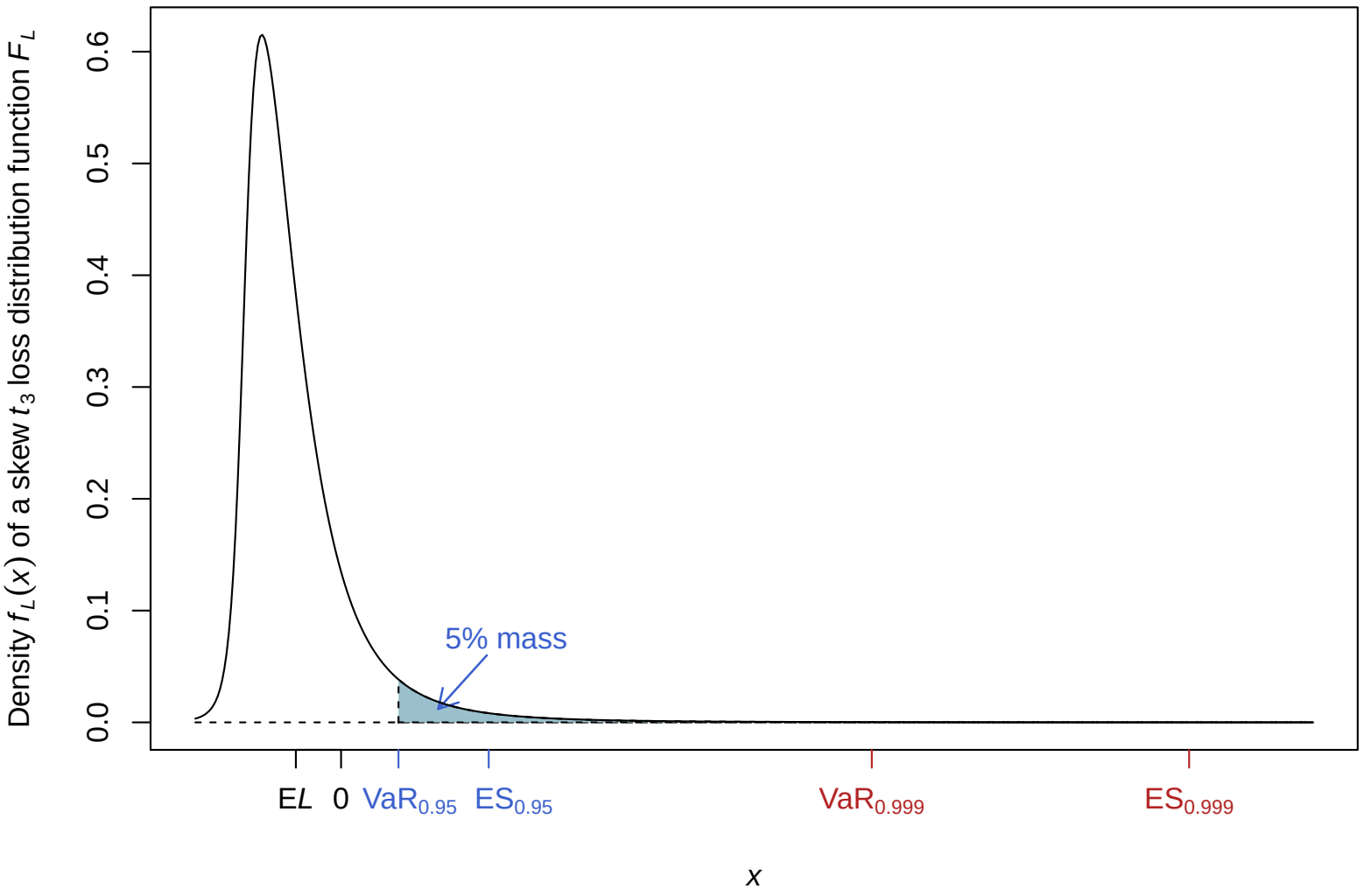
Definition 2.5 (Value-at-risk)
For a loss $L \sim F_L$, *value-at-risk* (VaR) at confidence level $\alpha \in (0,1)$ is defined by $\mathrm{VaR}_\alpha = \mathrm{VaR}_\alpha(L) = F_L^{\leftarrow}(\alpha) = \inf\{x \in \mathbb{R} : F_L(x) \geq \alpha\}$.

- VaR_α is simply the α-quantile of F_L. As such, $F_L(x) < \alpha$ for all $x < \mathrm{VaR}_\alpha(L)$ and $F_L(\mathrm{VaR}_\alpha(L)) = F_L(F_L^{\leftarrow}(\alpha)) \geq \alpha$.
- Known since 1994: Weatherstone 4^{15} report (J.P. Morgan; RiskMetrics)
- VaR is the most widely used risk measure (by Basel II or Solvency II)
- $\mathrm{VaR}_\alpha(L)$ is not a what if risk measure: It does not provide information about the severity of losses which occur with probability $\leq 1 - \alpha$.

Definition 2.8 (Expected shortfall)
For a loss $L \sim F_L$ with $\mathbb{E}(L_+) < \infty$, *expected shortfall* (ES) at confidence level $\alpha \in (0,1)$ is defined by

$$\mathrm{ES}_\alpha = \mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L)\,\mathrm{d}u. \tag{7}$$

- ES_α is the average over VaR_u for all $u \geq \alpha \Rightarrow \mathrm{ES}_\alpha \geq \mathrm{VaR}_\alpha$.
- Besides VaR, ES is the most important risk measure in practice.
- ES_α looks further into the tail of F_L, it is a "what if" risk measure (VaR_α is frequency-based; ES_α is severity-based).
- ES_α is more difficult to estimate and backtest than VaR_α (the variance of estimators is typically larger; larger sample size required).
- $\mathrm{ES}_\alpha(L) < \infty$ requires $\mathbb{E}(L_+) < \infty$.
- Subadditivity and elicitability (see the appendix). One can show:
  - In contrast to VaR_α, ES_α is subadditive (more later).


Density $f_L(x)$ of a skew t_3 loss distribution function F_L
0.0
0.1
0.2
0.3
0.4
0.5
0.6
5% mass
EL
0
VaR$_{0.95}$
ES$_{0.95}$
VaR$_{0.999}$
ES$_{0.999}$
x

8 Aggregate risk

8.1 Coherent and convex risk measures

- Consider a linear space $\mathcal{M} \subseteq \mathcal{L}^0(\Omega, \mathcal{F}, \mathbb{P})$ (a.s. finite rvs).
- Each $L \in \mathcal{M}$ (incl. constants) represents a loss over a fixed time horizon.
- A *risk measure* is a mapping $\varrho : \mathcal{M} \to \mathbb{R}$; $\varrho(L)$ gives the total amount of capital needed to back a position with loss L.
- $C \subseteq \mathcal{M}$ is *convex* if $(1-\gamma)x + \gamma y \in C$ for all $x, y \in C$, $0 < \gamma < 1$. C is a *convex cone* if, additionally, $\lambda x \in C$ when $x \in C$, $\lambda > 0$.
- Axioms for ϱ we consider are:

  Monotonicity: $L_1 \leq L_2 \Rightarrow \varrho(L_1) \leq \varrho(L_2)$.

  Translation invariance: $\varrho(L + m) = \varrho(L) + m$ for all $m \in \mathbb{R}$.

  Subadditivity: $\varrho(L_1 + L_2) \leq \varrho(L_1) + \varrho(L_2)$ for all $L_1, L_2 \in \mathcal{M}$.

  Positive homogeneity: $\varrho(\lambda L) = \lambda \varrho(L)$ for all $\lambda \geq 0$.

  Convexity: $\varrho(\gamma L_1 + (1-\gamma)L_2) \leq \gamma \varrho(L_1) + (1-\gamma)\varrho(L_2)$ for all $0 \leq \gamma \leq 1$, $L_1, L_2 \in \mathcal{M}$.

Definition 8.1 (Convex, coherent risk measures)

- A risk measure which satisfies monotonicity, translation invariance and convexity is called *convex*.
- A risk measure which satisfies monotonicity, translation invariance, subadditivity and positive homogeneity is called *coherent*.

A coherent risk measure is convex; the converse is not true, see below. On the other hand, for a positive-homogeneous risk measure, convexity and coherence are equivalent.

8.1.1 Risk measures and acceptance sets

Definition 8.2 (Acceptance set)
For a monotone and translation-invariant risk measure ϱ the *acceptance set of ϱ* is $A_\varrho = \{L \in \mathcal{M} : \varrho(L) \leq 0\}$ (so it contains the positions that are acceptable without any backing capital).

8.1.2 Dual representation of convex measures of risk

Theorem 8.10 (Dual representation for risk measures)
Suppose $|\Omega| = n < \infty$. Let $\mathcal{F} = \mathcal{P}(\Omega)$ (power set) and $\mathcal{M} := \{L : \Omega \to \mathbb{R}\}$. Then:

1) Every convex risk measure ϱ on $\mathcal{M}$ can be written in the form

$$\varrho(L) = \max\{\mathbb{E}_{\mathbb{Q}}(L) - \alpha_{\min}(\mathbb{Q}) : \mathbb{Q} \in \mathcal{S}^1(\Omega, \mathcal{F})\}, \tag{39}$$

   where $\mathcal{S}^1(\Omega, \mathcal{F})$ denotes the set of all probability measures on Ω, and where the penalty function $\alpha_{\min}$ is given by $\alpha_{\min}(\mathbb{Q}) = \sup\{\mathbb{E}_{\mathbb{Q}}(L) : L \in A_\varrho\}$.

2) If ϱ is coherent, it has the representation

$$\varrho(L) = \max\{\mathbb{E}_{\mathbb{Q}}(L) \colon \mathbb{Q} \in \mathcal{Q}\}$$

   for some set $\mathcal{Q} = \mathcal{Q}(\varrho) \subseteq \mathcal{S}^1(\Omega, \mathcal{F})$.

One can show that $\alpha_{\min}(\mathbb{Q}) = \sup_{L \in \mathcal{M}}\{\mathbb{E}_{\mathbb{Q}}(L) - \varrho(L)\}$.

8.1.3 Examples of dual representations

Proposition 8.11 (ES formulas)
For $\alpha \in (0,1)$,

1) $\mathrm{ES}_\alpha(L) = \dfrac{\mathbb{E}((L - F_L^{\leftarrow}(\alpha))_+)}{1-\alpha} + F_L^{\leftarrow}(\alpha)$;

2) $\mathrm{ES}_\alpha(L) = \dfrac{\mathbb{E}(L I_{\{L > F_L^{\leftarrow}(\alpha)\}}) + F_L^{\leftarrow}(\alpha)(1 - \alpha - \bar{F}_L(F_L^{\leftarrow}(\alpha)))}{1-\alpha}$.

Corollary 8.12 (ES formulas under continuous F_L)
Let F_L be continuous at $F_L^{\leftarrow}(\alpha)$. Then

1) $\mathrm{ES}_\alpha(L) = \frac{\mathbb{E}(L I_{\{L > F_L^{\leftarrow}(\alpha)\}})}{1-\alpha}$

2) $\mathrm{ES}_\alpha(L) = \mathbb{E}(L \,|\, L > F_L^{\leftarrow}(\alpha))$ (i.e. *conditional VaR (CVaR)*)

With dual representations one can give a proof for ES_α being subadditive; see the following result.

Theorem 8.13

For $\alpha \in [0,1)$, ES_α is coherent on $\mathcal{M} = \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. The dual representation is given by

$$\mathrm{ES}_\alpha(L) = \max\{\mathbb{E}^{\mathbb{Q}}(L) : \mathbb{Q} \in \mathcal{Q}_\alpha\}, \tag{40}$$

where $\mathcal{Q}_\alpha$ is the set of all probability measures on $(\Omega, \mathcal{F})$ that are absolutely continuous with respect to $\mathbb{P}$ and for which the measure-theoretic density $d\mathbb{Q}/d\mathbb{P}$ is bounded by $1/(1-\alpha)$.

8.3.2 Elliptically distributed risk factors

Theorem 8.24 (Risk measurement for elliptical risk factors)
Let $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ and ϱ be any positive-homogeneous, translation-invariant and law-invariant risk measure on $\mathcal{M}$. Then:

1) For any $L = m + \boldsymbol{\lambda}'\boldsymbol{X} \in \mathcal{M}$, $\varrho(L) = m + \boldsymbol{\lambda}'\boldsymbol{\mu} + \sqrt{\boldsymbol{\lambda}'\Sigma\boldsymbol{\lambda}}\varrho(Y_1)$ for $Y_1 \sim S_1(\psi)$.
2) If $\varrho(Y_1) \geq 0$, then ϱ is subadditive on $\mathcal{M}$ (e.g., VaR_α for $\alpha \geq 0.5$).
3) If $\mathbb{E}\boldsymbol{X}$ exists then, $\forall\, L = m + \boldsymbol{\lambda}'\boldsymbol{X} \in \mathcal{M}$ and $\rho_{ij} = \wp(\Sigma)_{ij} = P_{ij}$,

$$\varrho(L - \mathbb{E}L) = \sqrt{\sum_{i=1}^{d}\sum_{j=1}^{d} \rho_{ij}\lambda_i\lambda_j\varrho(X_i - \mathbb{E}X_i)\varrho(X_j - \mathbb{E}X_j)}.$$

4) If $\mathrm{cov}(\boldsymbol{X})$ exists and $\varrho(Y_1) > 0$ then, for every $L \in \mathcal{M}$, $\varrho(L) = \mathbb{E}(L) + k_\varrho\sqrt{\mathrm{var}(L)}$ for some $k_\varrho > 0$ depending on ϱ.
5) If Σ^{-1} ex., $\varrho(Y_1) > 0$ then $S_\varrho = \{\boldsymbol{x} : (\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \leq \varrho(Y_1)^2\}$.

8.4.4 Risk aggregation and Fréchet problems

- Consider the margins-plus-copula approach where $L_j \sim F_j$, $j \in \{1, \dots, d\}$, are treated as known (estimated or postulated) and C is unknown.
- Consider $L = L_1 + \cdots + L_d$. Due to the unknown C (*dependence uncertainty*), risk measures can no longer be computed explicitly.
- Our goal is to find bounds on VaR_α and ES_α under all possible C. Let

$$\mathcal{S}_d := \mathcal{S}_d(F_1, \dots, F_d) := \left\{ L = \sum_{j=1}^d L_j : L_j \sim F_j,\ j = 1, \dots, d \right\}$$

and consider

$$\overline{\varrho}(L) := \overline{\varrho}(\mathcal{S}_d) := \sup\{\varrho(L) : L \in \mathcal{S}_d(F_1, \dots, F_d)\} \quad \text{(worst } \varrho\text{)}$$
$$\underline{\varrho}(L) := \underline{\varrho}(\mathcal{S}_d) := \inf\{\varrho(L) : L \in \mathcal{S}_d(F_1, \dots, F_d)\} \quad \text{(best } \varrho\text{)}$$

- If $\varrho = \mathrm{ES}_\alpha$, $\overline{\mathrm{ES}}_\alpha(L) = \sum_{j=1}^d \mathrm{ES}_\alpha(L_j)$ (subadditivity, com. additivity). $\underline{\mathrm{ES}}_\alpha$, $\underline{\mathrm{VaR}}_\alpha$, $\overline{\mathrm{VaR}}_\alpha$ depend on whether the portfolio is *homogeneous* (that is, $F_1 = \cdots = F_d$); we focus on $\overline{\mathrm{VaR}}_\alpha$.

Summary of existing results

$d = 2$: Fully solved analytically

$d \geq 3$: Here we distinguish:

- Homogeneous case ($F_1 = \cdots = F_d$):
  - $\underline{\mathrm{ES}}_\alpha(L)$ solved analytically for decreasing densities (e.g. Pareto, Exponential)
  - $\underline{\mathrm{VaR}}_\alpha(L)$, $\overline{\mathrm{VaR}}_\alpha(L)$ solved analytically for tail-decreasing densities (e.g. Pareto, Log-normal, Gamma)
- Inhomogeneous case:
  - Few analytical results: current research
  - Numerical methods: (Adaptive/Block) Rearrangement Algorithm

8.5 Capital allocation

How can the overall capital requirement may be disaggregated into additive contributions/units/investments? Motivation: How can we measure the risk-adjusted performance of different investments?

8.5.1 The allocation problem

- The performance of investments is usually measured using a *RORAC (return on risk-adjusted capital) approach* by considering

$$\frac{\text{expected profit of investment } j}{\text{risk capital for investment } j}.$$

- The risk capital of investment j with loss L_j can be computed as follows: Compute $\varrho(L) = \varrho(L_1 + \cdots + L_d)$. Then allocate $\varrho(L)$ to the investments according to a *capital allocation principle* such that

$$\varrho(L) = \sum_{j=1}^{d} \mathrm{AC}_j,$$

where the *risk contribution* AC_j is the capital allocated to investment j.

The formal set-up

- Consider an open set $\mathbf{1} \in \Lambda \subseteq \mathbb{R}^d \setminus \{\mathbf{0}\}$ of portfolio weights and define

$$L(\boldsymbol{\lambda}) = \boldsymbol{\lambda}'\boldsymbol{L} = \sum_{j=1}^{d} \lambda_j L_j, \quad \boldsymbol{\lambda} \in \Lambda.$$

- For a risk measure ϱ, define the *associated risk-measure function*

$$r_\varrho(\boldsymbol{\lambda}) = \varrho(L(\boldsymbol{\lambda})),$$

so that $r_\varrho(\mathbf{1}) = \varrho(L)$.

8.5.2 The Euler principle and examples

- If r_ϱ is positive homogeneous and differentiable at $\boldsymbol{\lambda} \in \Lambda$, Euler's rule (see the appendix)implies that

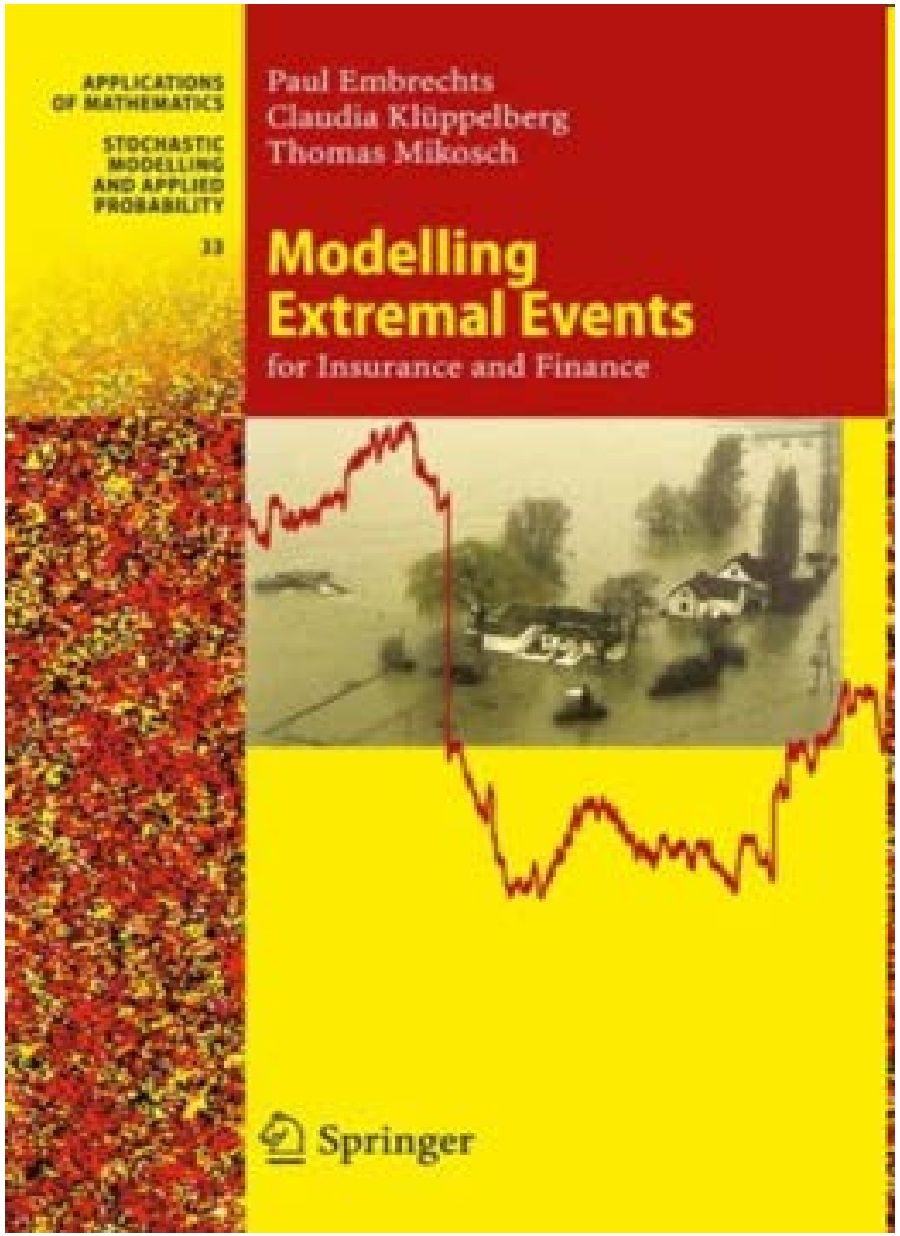
$$r_\varrho(\boldsymbol{\lambda}) = \sum_{i=1}^{d} \lambda_i \frac{\partial r_\varrho}{\partial \lambda_i}(\boldsymbol{\lambda}) \quad \text{so} \quad \varrho(L) = r_\varrho(\mathbf{1}) = \sum_{j=1}^{d} \frac{\partial r_\varrho}{\partial \lambda_j}(\mathbf{1}).$$

Note that r_ϱ is positive homogeneous if ϱ is.

5 Extreme value theory

5.1 Maxima

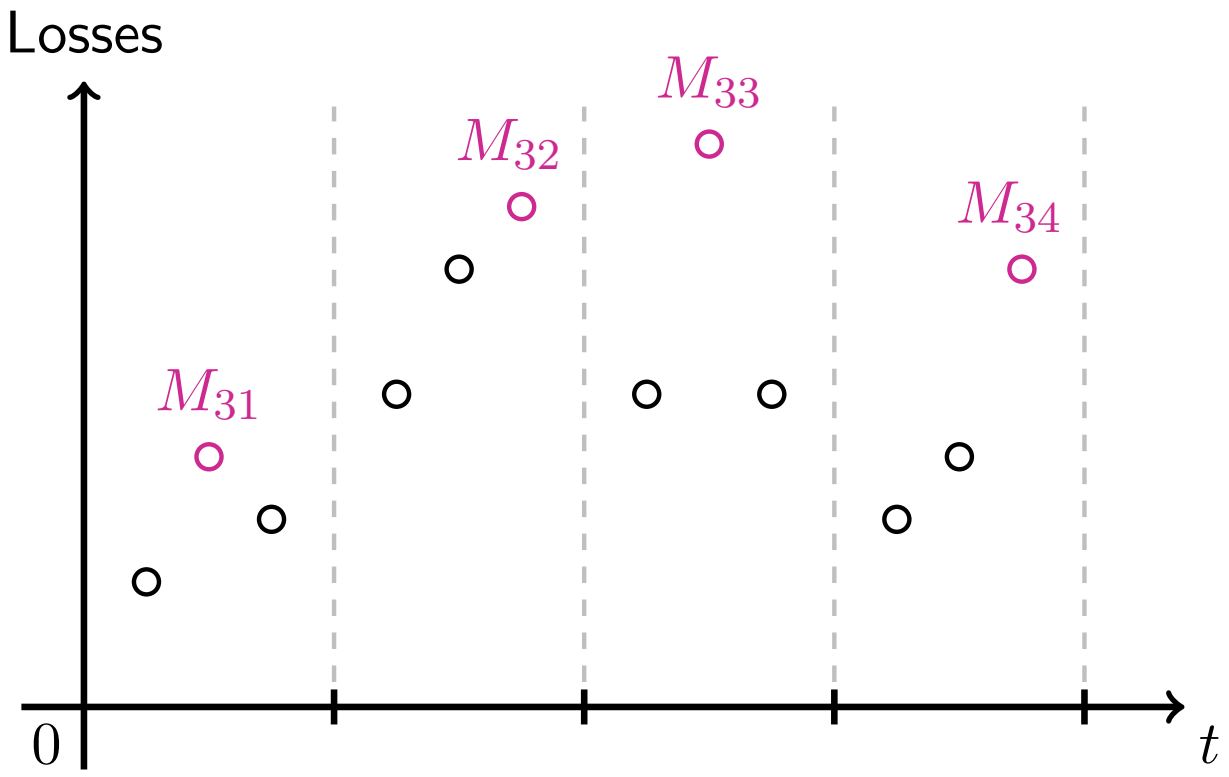
5.2 Threshold exceedances



An introduction to 1-d EVT

5.1.4 The block maxima method (BMM)

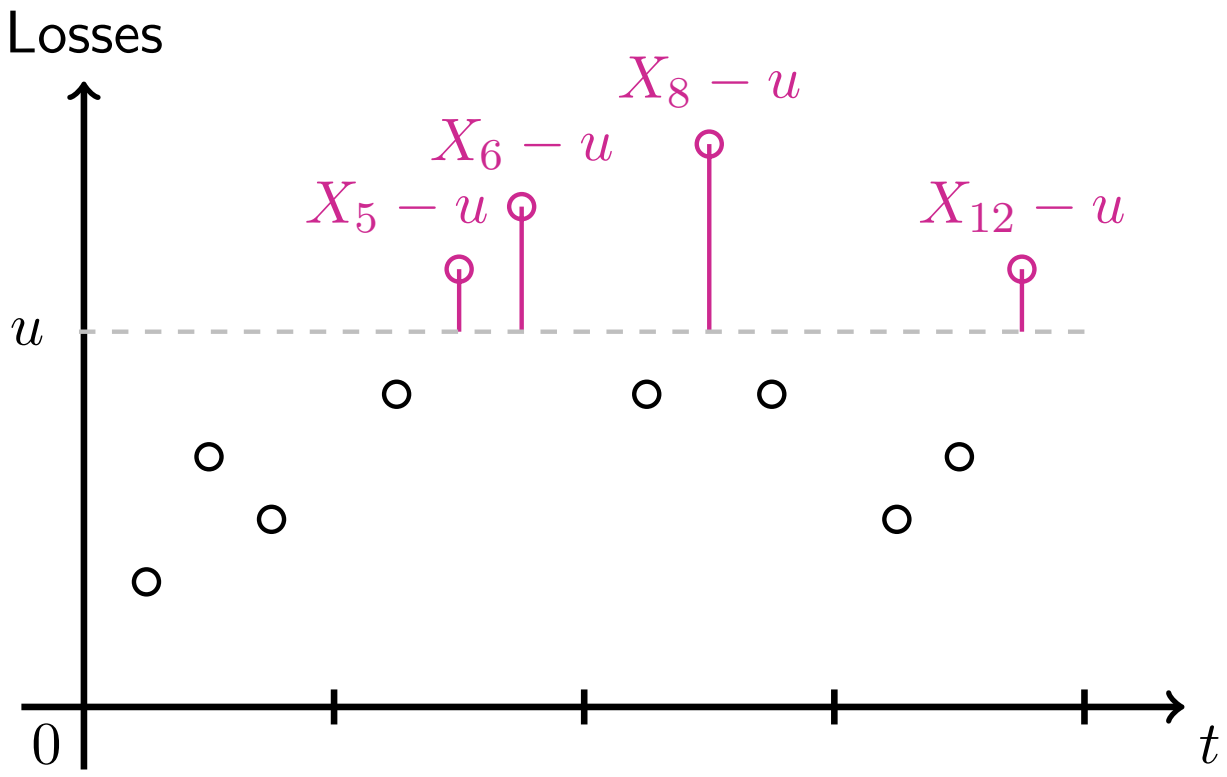
The basic idea in a picture based on losses $X_1, \ldots, X_{12}$:



Consider the maximal loss from each block and fit $H_{\xi,\mu,\sigma}$ to them.

5.2.2 Modelling excess losses

The basic idea in a picture based on losses $X_1, \ldots, X_{12}$.



Consider all excesses over u and fit $G_{\xi,\beta}$ to them.

Convergence of maxima

QRM is concerned with maximal losses (worst-case losses). Let $(X_i)_{i\in\mathbb{N}} \overset{\text{ind.}}{\sim} F$ (can be relaxed to a strictly stationary time series) and F continuous. Then the *block maximum* is given by

$$M_n = \max\{X_1, \dots, X_n\}.$$

One can show that $M_n \underset{n\to\infty}{\overset{\text{a.s.}}{\to}} x_F$ (similar as in the SLLN; due to monotone convergence to a constant) where

$$x_F := \sup\{x \in \mathbb{R} : F(x) < 1\} = F^{\leftarrow}(1) \leq \infty$$

denotes the *right endpoint of F*.

Question: Is there a "CLT" for block maxima?

Idea CLT: What about linear transformations (the simplest possible)?

Definition 5.1 (Maximum domain of attraction)
Suppose we find *normalizing sequences* of real numbers $(c_n) > 0$ and (d_n) such that $(M_n - d_n)/c_n$ converges in distribution, i.e.

$$\mathbb{P}((M_n - d_n)/c_n \leq x) = \mathbb{P}(M_n \leq c_n x + d_n) = F^n(c_n x + d_n) \underset{n\uparrow\infty}{\rightarrow} H(x),$$

for some *non-degenerate* df H (not a unit jump). Then F is in the *maximum domain of attraction* of H ($F \in \mathrm{MDA}(H)$).

The convergence to types theorem (see the appendix)guarantees that H is determined up to location/scale, i.e. H specifies a unique *type* of distribution.

Question: What does H look like?

Theorem 5.3 (Fisher–Tippett–Gnedenko)
If $F \in \mathrm{MDA}(H)$ for some non-degenerate H, then H must be of GEV type, i.e. $H = H_\xi$ for some $\xi \in \mathbb{R}$.

Proof. Non-trivial. For a sketch, see Embrechts et al. (1997, p. 122). □

- **Interpretation:** If location-scale transformed maxima of iid random variables converge in distribution to a non-degenerate limit, the limiting distribution must be a location-scale transformed GEV distribution (that is, of GEV type).
- One can always choose normalizing sequences $(c_n) > 0$, (d_n) such that H_ξ appears in standard form (although from a statistical point of view, $(c_n) > 0$, (d_n) are simply estimated).
- All commonly encountered continuous distributions are in the MDA of some GEV distribution.

Definition 5.2 (Generalized extreme value (GEV) distribution)
The (standard) *generalized extreme value (GEV) distribution* is given by

$$H_\xi(x) = \begin{cases} \exp(-(1+\xi x)^{-1/\xi}), & \text{if } \xi \neq 0, \\ \exp(-e^{-x}), & \text{if } \xi = 0, \end{cases}$$

where $1 + \xi x > 0$ (MLE!). A three-parameter family is obtained by a location-scale transform $H_{\xi,\mu,\sigma}(x) = H_\xi((x-\mu)/\sigma)$, $\mu \in \mathbb{R}$, $\sigma > 0$.

- The parameterization is continuous in ξ (simplifies statistical modelling).
- The larger ξ, the heavier tailed H_ξ (if $\xi > 0$, $\mathbb{E}(X^k) = \infty$ iff $k \geq \frac{1}{\xi}$).
- ξ is the *shape* (determines moments, tail). Special cases:
  1) $\xi < 0$: the Weibull df, short-tailed, $x_{H_\xi} < \infty$;
  2) $\xi = 0$: the Gumbel df, $x_{H_0} = \infty$, decays exponentially;
  3) $\xi > 0$: the Fréchet df, $x_{H_\xi} = \infty$, heavy-tailed ($\bar{H}_\xi(x) \approx (\xi x)^{-1/\xi}$), most important case for practice

Definition 5.13 (Excess distribution over u, mean excess function)
Let $X \sim F$. The *excess distribution over the threshold u* is defined by

$$F_u(x) = \mathbb{P}(X - u \leq x \,|\, X > u) = \frac{F(x+u) - F(u)}{1 - F(u)}, \quad x \in [0, x_F - u).$$

If $\mathbb{E}|X| < \infty$, the *mean excess function* is defined by

$$e(u) = \mathbb{E}(X - u \,|\, X > u) \quad \text{(i.e. the mean w.r.t. } F_u\text{)}$$

- **Interpretation:** F_u is the distribution of the excess loss $X - u$ over u, given that $X > u$. $e(u)$ is the mean of F_u as a function of u.
- One can show the useful formula $e(u) = \frac{1}{\bar{F}(u)} \int_u^{x_F} \bar{F}(x)\,\mathrm{d}x$.
- For continuous $X \sim F$ with $\mathbb{E}|X| < \infty$, the following formula holds:

$$\mathrm{ES}_\alpha(X) = e(\mathrm{VaR}_\alpha(X)) + \mathrm{VaR}_\alpha(X), \quad \alpha \in (0,1) \tag{12}$$

Theorem 5.15 (Pickands–Balkema–de Haan (1974/75))
There exists a positive, measurable function $\beta(u)$, such that

$$\lim_{u \uparrow x_F} \sup_{0 \le x < x_F - u} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0.$$

if and only if $F \in \mathrm{MDA}(H_\xi)$, $\xi \in \mathbb{R}$.

Proof. Non-trivial; see, e.g. Pickands (1975) and Balkema and de Haan (1974). □

Interpretation

- The GPD is the canonical df for excess losses over high u.
- The result is also a characterization of $\mathrm{MDA}(H_\xi)$, $\xi \in \mathbb{R}$. All $F \in \mathrm{MDA}(H_\xi)$ form a set of df for which the excess distribution converges to the GPD $G_{\xi,\beta}$ with the same ξ as in H_ξ when u is raised.

5.2 Threshold exceedances

The BMM is wasteful of data (only the maxima of large blocks are used). It has been largely superseded in practice by methods based on threshold exceedances (*peaks-over-threshold (POT)* approach), where all data above a designated high threshold u are used.

5.2.1 Generalized Pareto distribution

Definition 5.12 (Generalized Pareto distribution (GPD))
The *generalized Pareto distribution (GPD)* is given by

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - \exp(-x/\beta), & \text{if } \xi = 0, \end{cases}$$

where $\beta > 0$, and the support is $x \geq 0$ when $\xi \geq 0$ and $x \in [0, -\beta/\xi]$ when $\xi < 0$.

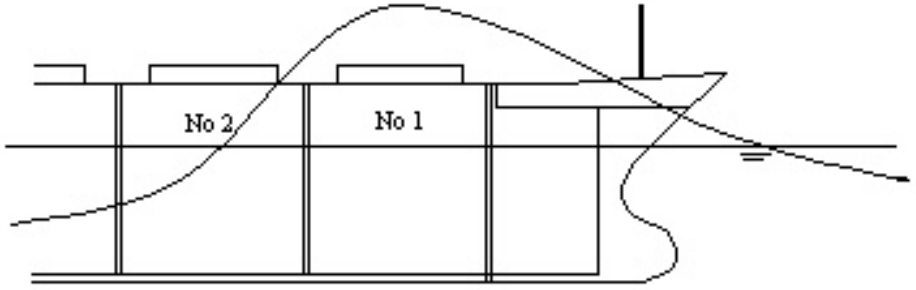
The sinking of the MV Derbyshire

- Massive ore-bulk-oil combination carrier, built in 1976
- On 9 Sep 1980, she sinks in the Pacific Ocean, close to Japan, during Typhoon Orchid
- All 44 people on board died
- Sunk suddenly: no mayday calls, no lifeboats
- Largest UK ship to have been lost at sea

J.E. Heffernan & J.A. Tawn (2003) *JRSS(C)* 52(3), 337–354

Why did the Derbyshire sink?

- 1990 Formal investigation. Hypothesis: structural failure (evidence from sister ships)? Inconclusive evidence.
- 1994 Ship found in 4200m deep ocean
- 1997–1998 underwater expedition: sinking due to crew negligence?
- 2000 two-month formal high court investigation: hatch cover on hold 1 failed as bow deck encountered ‘green seas’



No. 1 Cargo hold hatch covers on the Derbyshire (The vessel was built in 1976 as the Liverpool Bridge)

Crown copyright

The question: Probability of large wave impact?

- Safety standard for hatch cover on hold 1: 42 kPa
- Uncertainty about:
  - ship: speed, preliminary damage leading to loss of free-board
  - wave conditions (height, period): hindcast via satellite data

Statistical problem

$$\mathrm{P}(\text{wave impact} > 42 \text{ kPa on hatch cover} \mid \text{ship \& see conditions}) = ?$$

How to model the distribution of large wave impact?

- Wave impacts arrive according to a Poisson process
- $\mu(z)$ is expected number of impacts per hour larger than u
- Size of such an impact:

$$F_u(x; z) = \mathrm{P}(\text{impact} \leq x \mid \text{impact} > u; z)$$

- Previous analyses: model uncertainty about F_u is much larger than uncertainty about the state of the ship or the sea!

Conclusions and final outcome of the investigation

- Sinking due to initial damage, leading to increased probability of wave impact causing front hatch cover to collapse
- Judge's report: Statistical evidence of

  'absolutely fundamental importance to the outcome of this investigation'

- Judge recommends further extensive model testing to strengthen regulations on hatch cover strengths
- Eventually led to new worldwide mandatory design standards for carriers: 35% stronger than previously

J.E. Heffernan & J.A. Tawn (2001) *Extremes* 4, 359–378
J.E. Heffernan & J.A. Tawn (2003) *JRSS(C)* 52(3), 337–354

A final comment:

L‘Aquila April 6, 2009

David Spiegelhalter

Question: Why this combination?

Thank you!

ETH Zürich

UNIVERSITY OF COPENHAGEN