Optimization-based Approaches to Extreme Event Analysis

Henry Lam
Industrial Engineering & Operations Research
Columbia University

Joint work with Clementine Mottet (TripAdvisor), Xinyu Zhang (Columbia University) and Soumyadip Ghosh (IBM Research AI)

Overview

This talk is about: Estimating tail quantities..

Examples of extreme event/risk analysis:

- Environmental sciences, e.g. probability of flood etc. (Davison & Smith '90...)
- Engineering reliability (Nicola et al. '93, Heidelberger '95...)
- Large loss in insurance and finance (Kulik & Palmowski '11, McNeil '97, Beirlant & Teugels '92, Embrechts et al. '97, Glasserman & Li '05, Glasserman et al. '07…)

Overview

This talk is about: Estimating tail quantities..

Examples of extreme event/risk analysis:

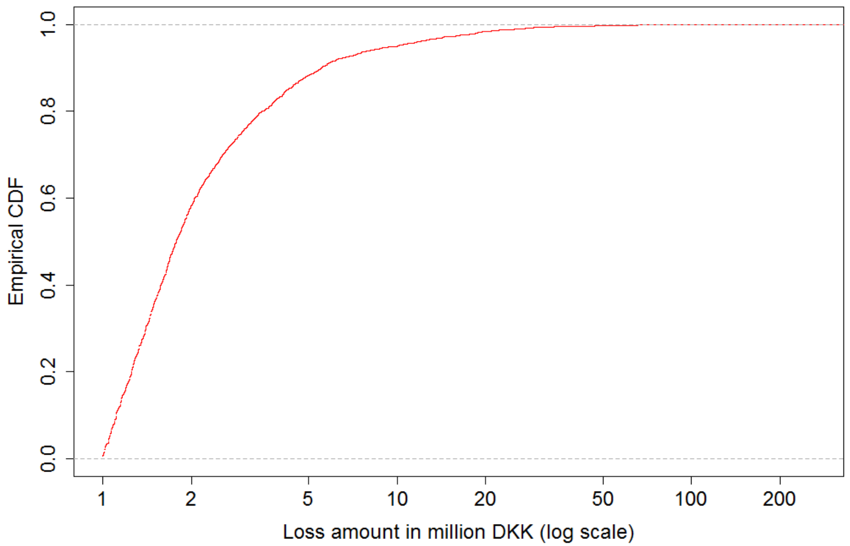
- Environmental sciences, e.g. probability of flood etc. (Davison & Smith ’90…)
- Engineering reliability (Nicola et al. ‘93, Heidelberger ’95…)
- Large loss in insurance and finance (Kulik & Palmowski ‘11, McNeil ‘97, Beirlant & Teugels ‘92, Embrechts et al. ’97, Glasserman & Li ‘05, Glasserman et al. ’07…)

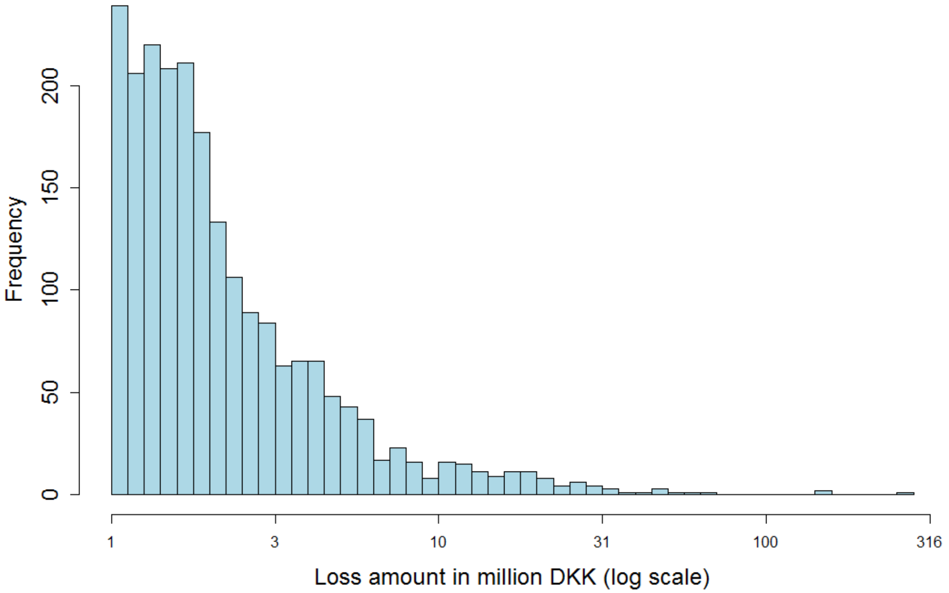
Recurring issue: few data in the tail

Example

Example: Fire losses over one million Danish Krone (DKK) during 1980-1990 (McNeil '97)

data = 2156



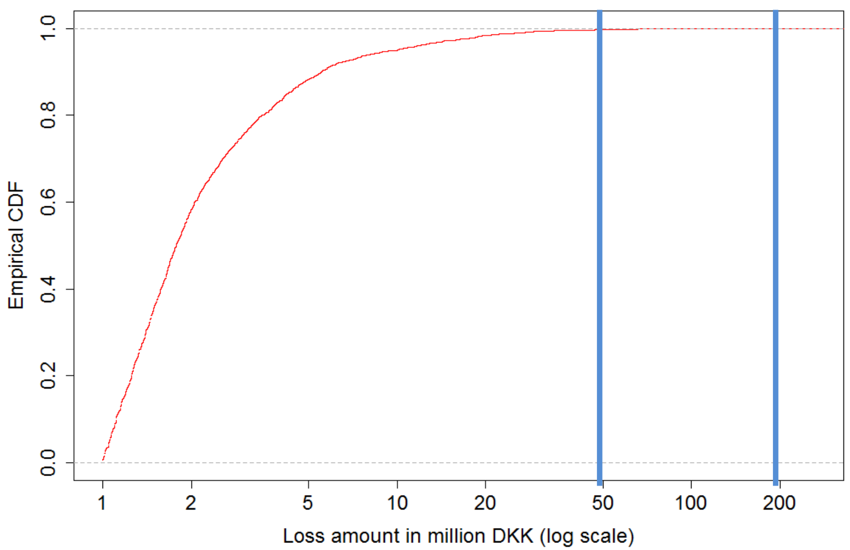


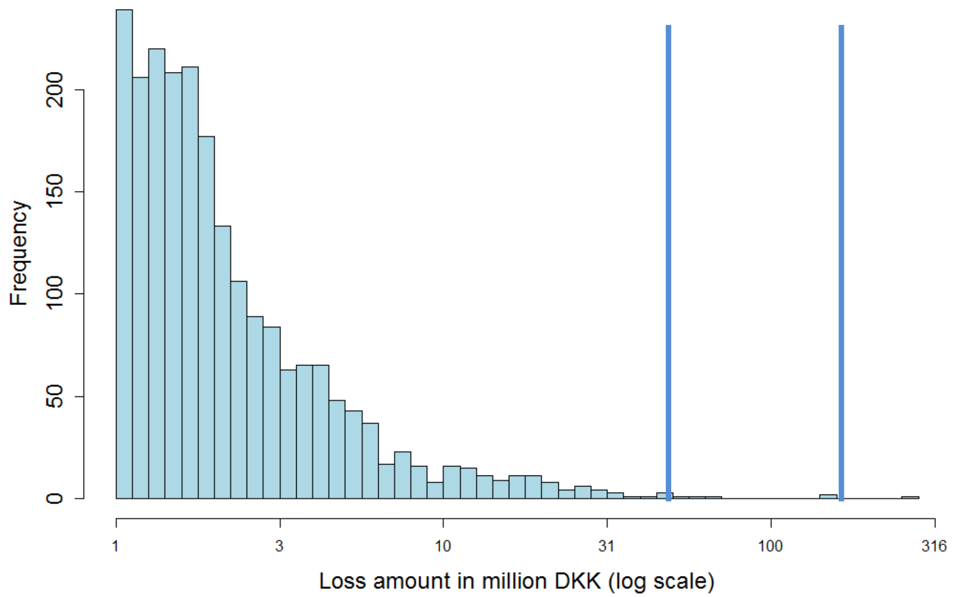
Example

High-excess insurance policy (McNeil '97):

$$\text{Payout} = \begin{cases} 0 & \text{if } \ 0 < X < 50 \\ X - 50 & \text{if } \ 50 \le X < 200 \\ 150 & \text{if } \ 200 \le X < \infty \end{cases}$$

Insurance Price $= E[\text{Payout}]$





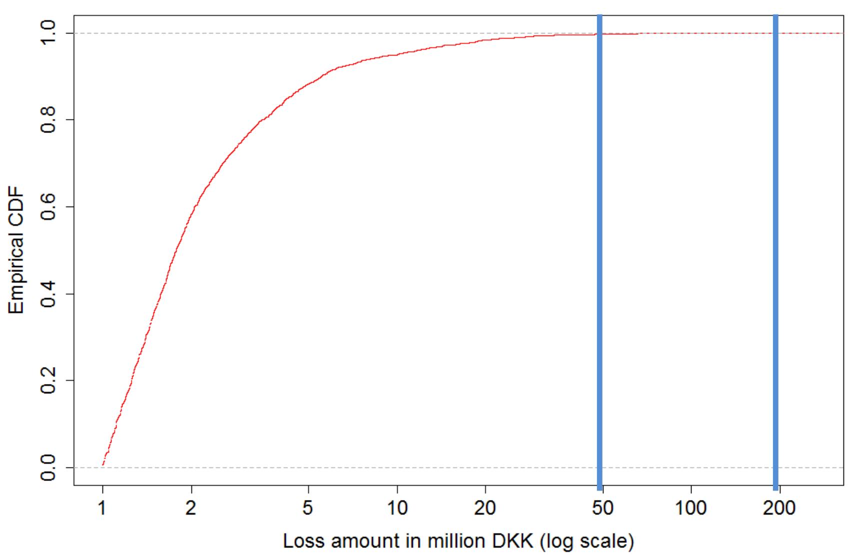
Example

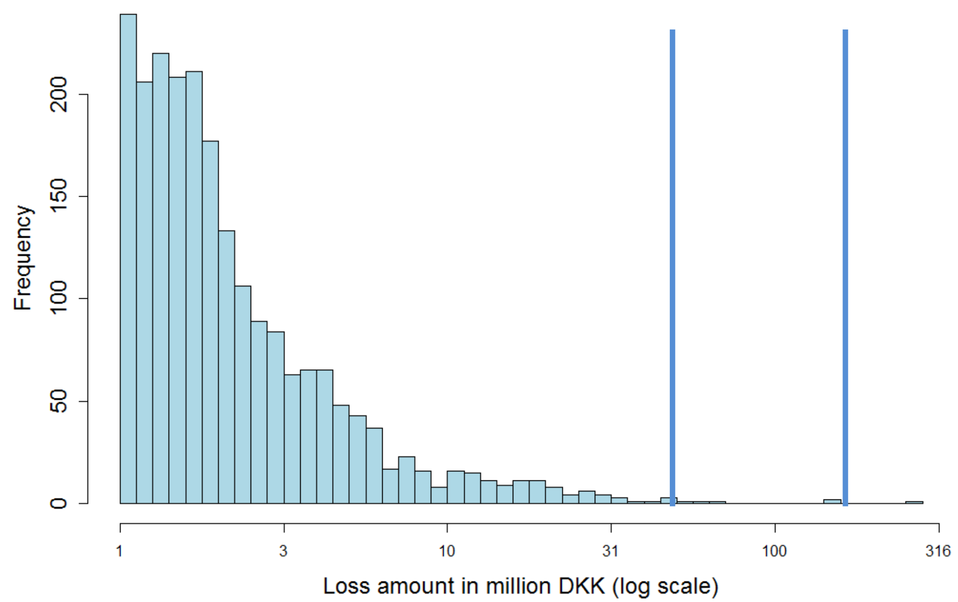
High-excess insurance policy (McNeil '97):

$$\text{Payout} = \begin{cases} 0 & \text{if } 0 < X < 50 \\ X - 50 & \text{if } 50 \le X < 200 \\ 150 & \text{if } 200 \le X < \infty \end{cases}$$

Insurance Price $= E[\text{Payout}]$

Only seven data points above 50





Example

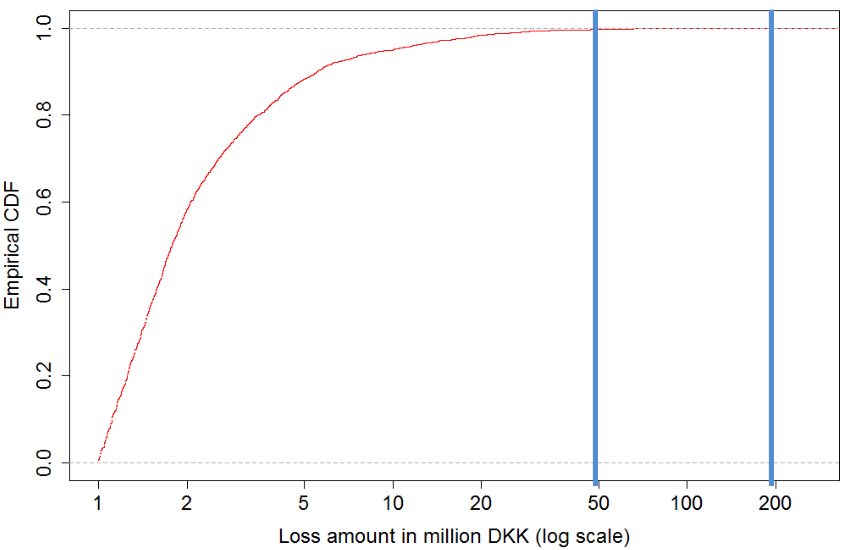
High-excess insurance policy (McNeil '97):

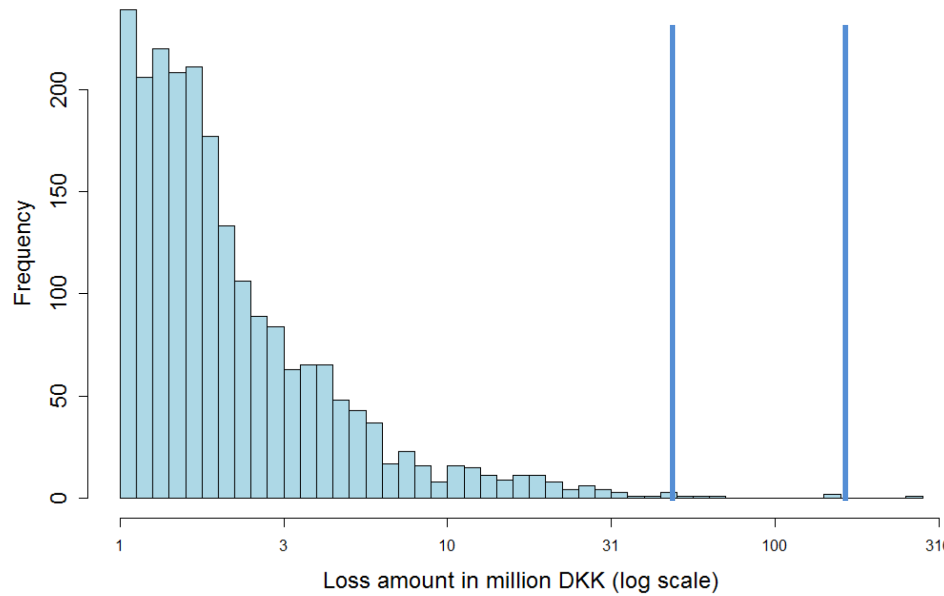
$$\text{Payout} = \begin{cases} 0 & \text{if } \; 0 < X < 50 \\ X - 50 & \text{if } \; 50 \le X < 200 \\ 150 & \text{if } \; 200 \le X < \infty \end{cases}$$

Insurance Price $= E[\text{Payout}]$

Only seven data points above 50

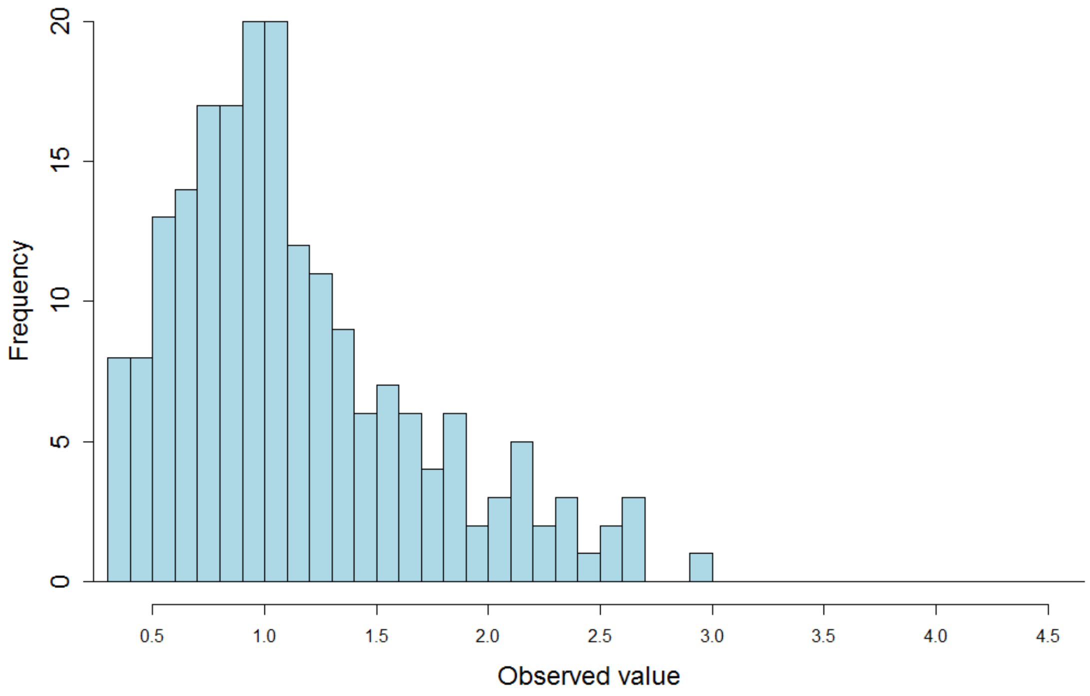
What is a good estimate of the insurance price?





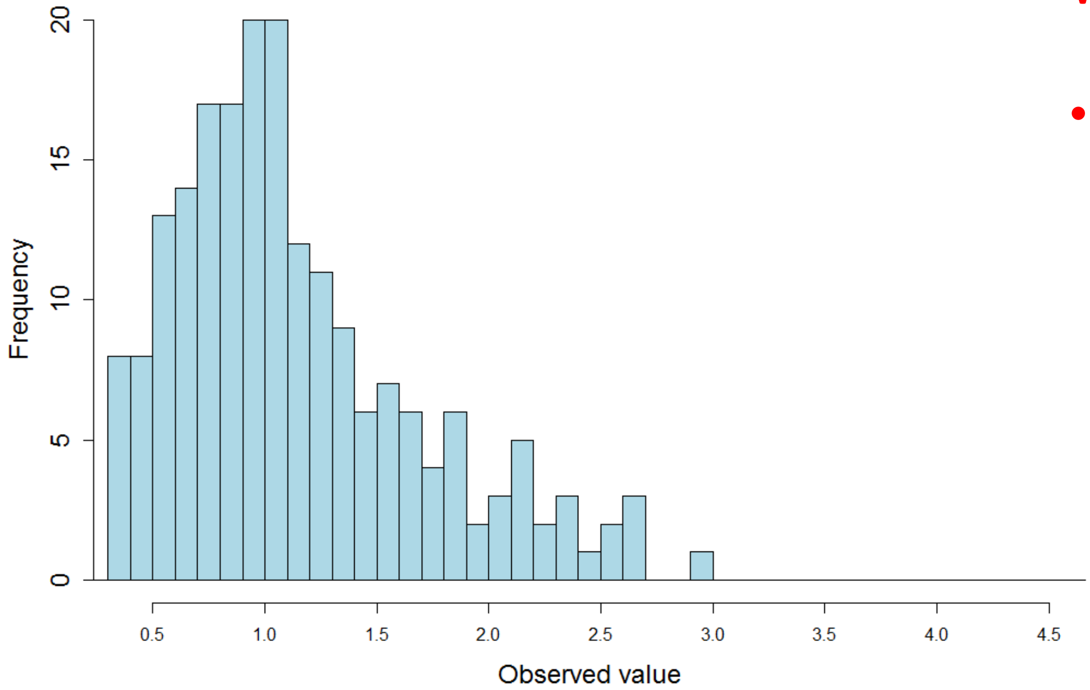
Example 2

200 data points from unknown distribution



Example 2

200 data points from unknown distribution

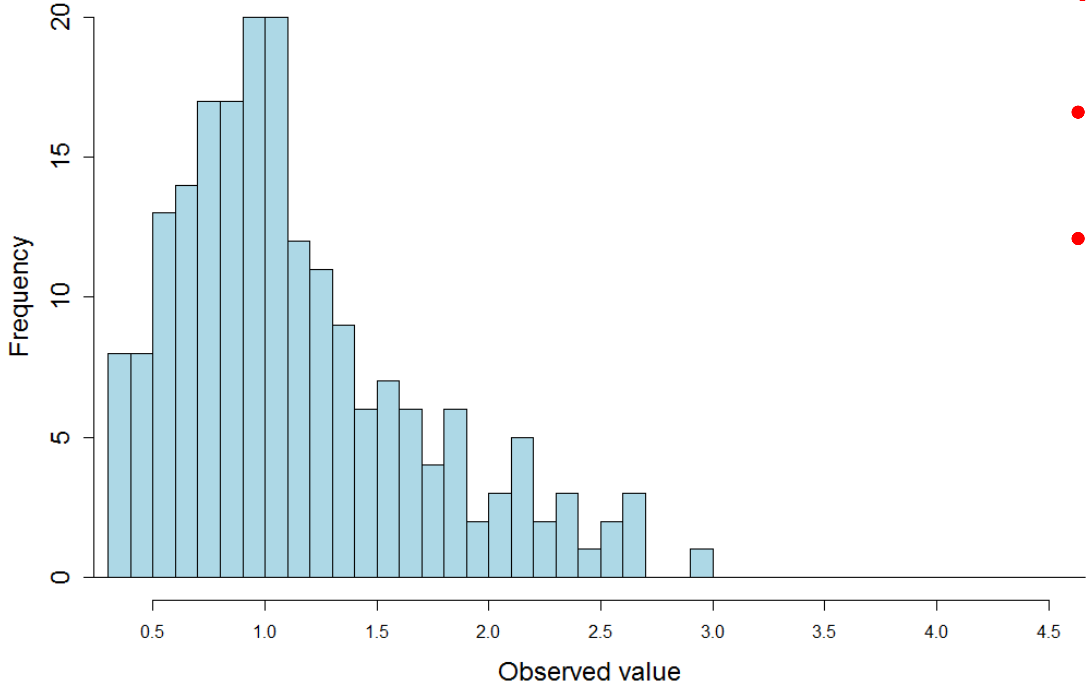


What is an estimate of

- $P(3.5 < X < 4.5)$

Example 2

200 data points from unknown distribution

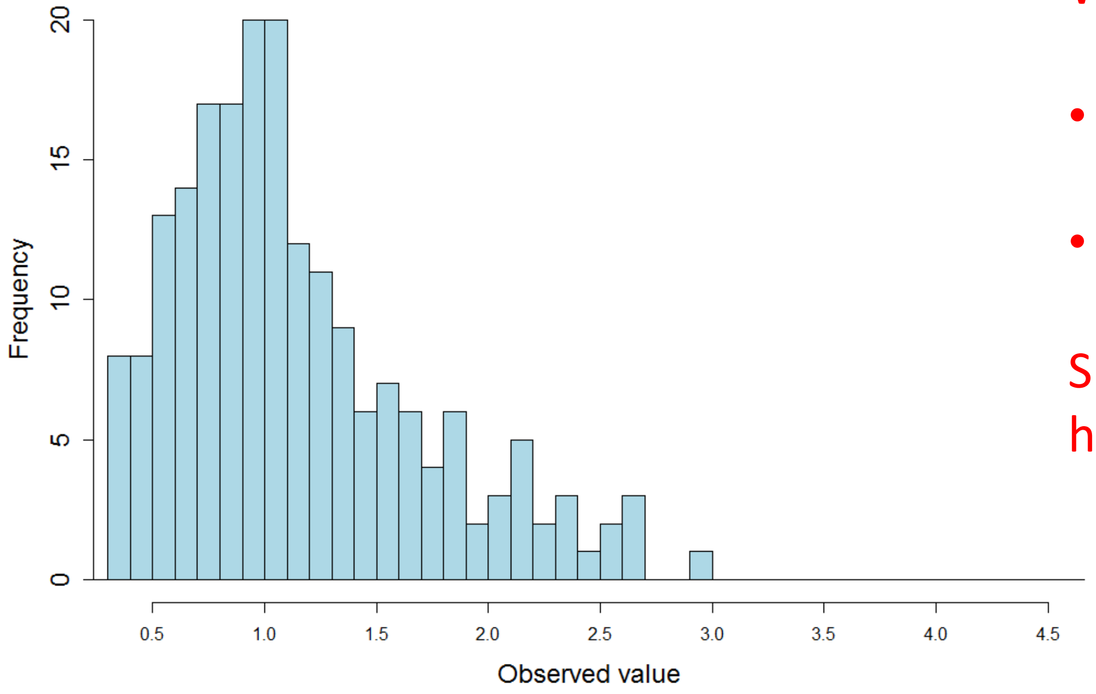


What is an estimate of

- $P(3.5 < X < 4.5)$
- $P(X_1 + \cdots + X_{50} > 100)$

Example 2

200 data points from unknown distribution



What is an estimate of

- $P(3.5 < X < 4.5)$
- $P(X_1 + \cdots + X_{50} > 100)$

Should you fit a light or heavy tail?

Conventional Methods

"Goodness-of-fit":

- Quantile plot, mean excess plot, etc. (Embrechts et al. '97…)
- Shape/properties of parametric families (Hogg & Klugman '84…)

Extreme value theory:

- Pickands-Balkema-de Haan Theorem (Pickands '75, Davison & Smith '90…):

The distribution function of an excess loss over high threshold u, i.e.

$$F_u(x) = P(X - u \le x | X > u)$$

converges to a generalized Pareto distribution (GPD) as $u \to \infty$

Conventional Methods

“Goodness-of-fit”:

- Quantile plot, mean excess plot, etc. (Embrechts et al. ’97…)
- Shape/properties of parametric families (Hogg & Klugman ’84…)

Extreme value theory:

- Pickands-Balkema-de Haan Theorem (Pickands ‘75, Davison & Smith ’90…):

The distribution function of an excess loss over high threshold u, i.e.

$$F_u(x) = P(X - u \leq x | X > u)$$

converges to a generalized Pareto distribution (GPD) as $u \to \infty$

Implication:

- GPD is justified to fit the tail portion of data

Challenges:

- Need u to be large enough for convergence to GPD (control bias), and
- Sizable data above u to estimate the parameters (control variance)

Main Goal of the Talk

Introduce an alternate method that is robust and nonparametric

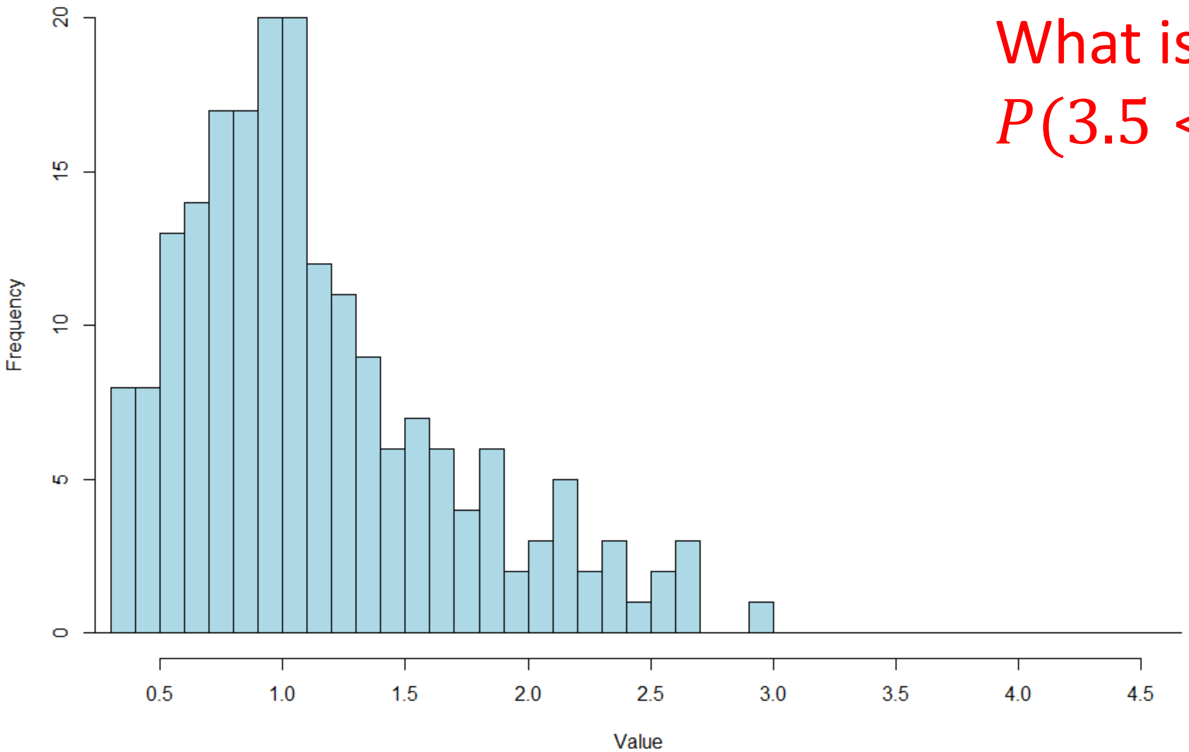
Main Goal of the Talk

Introduce an alternate method that is robust and nonparametric

No free lunch. Some price to pay…

A Robust Optimization Approach

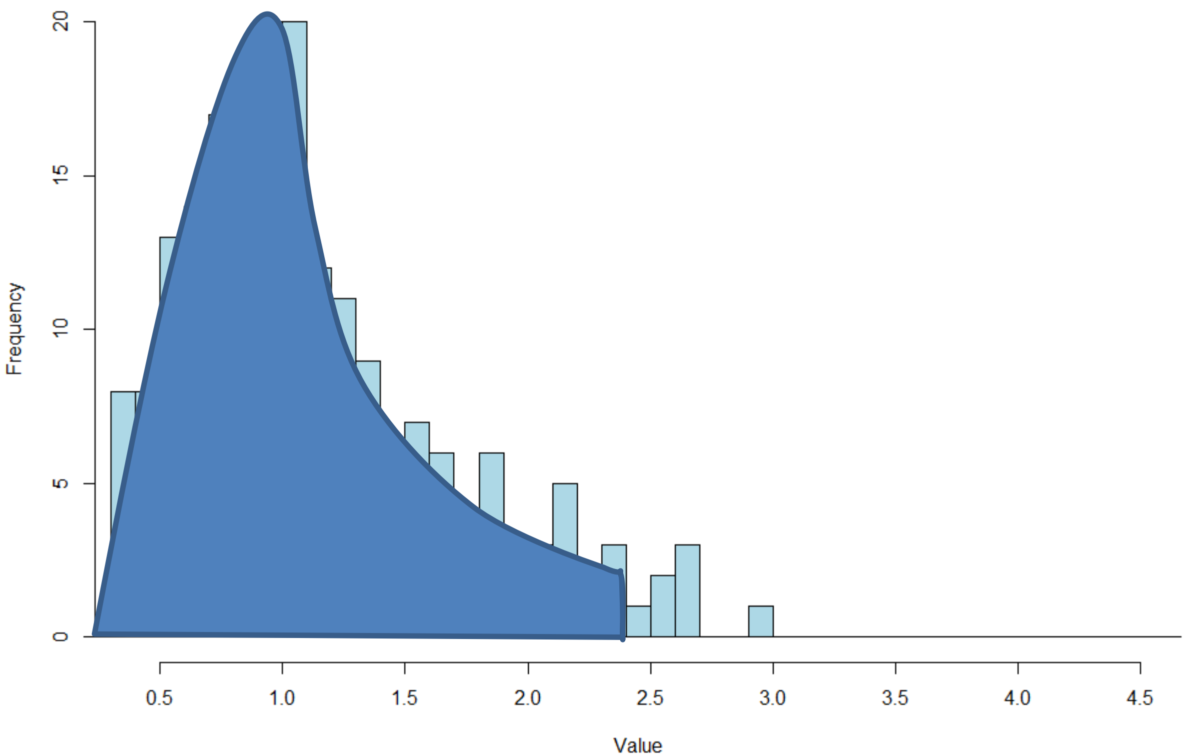
Can we use information from the region that has data to help predict for the region that has no data?



What is an estimate of $P(3.5 < X < 4.5)$?

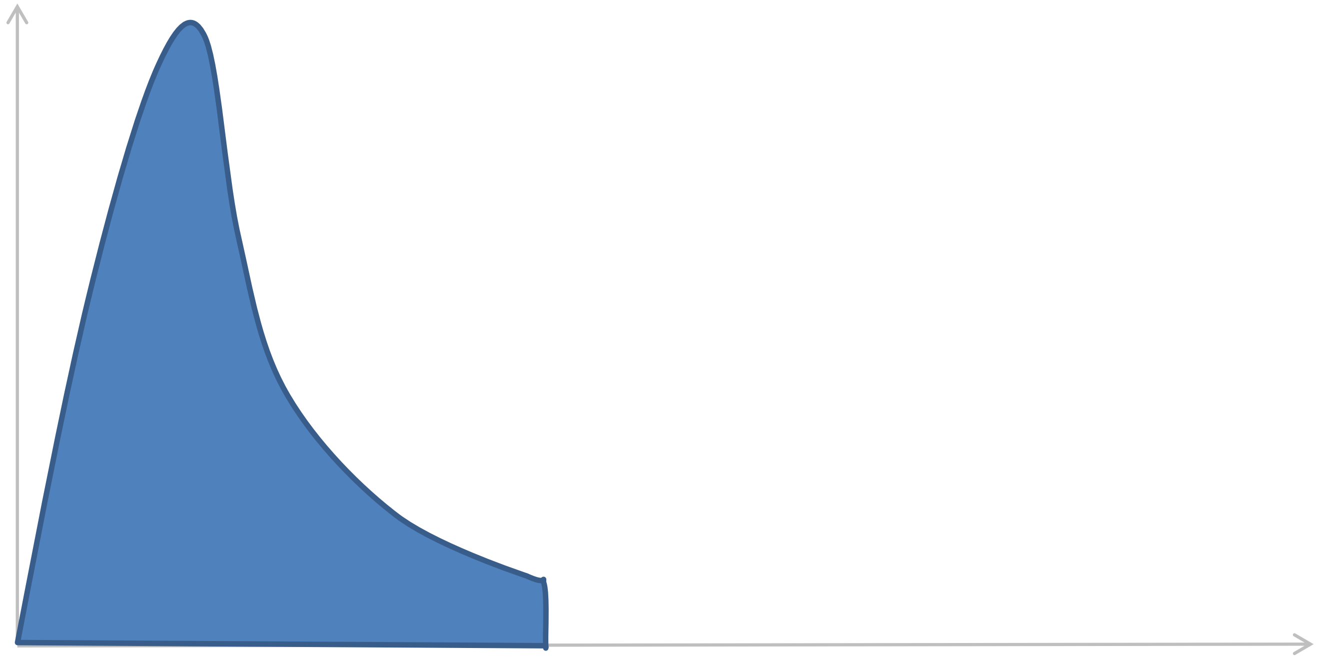
A Robust Optimization Approach

Can we use information from the region that has data to help predict for the region that has no data?



A Robust Optimization Approach

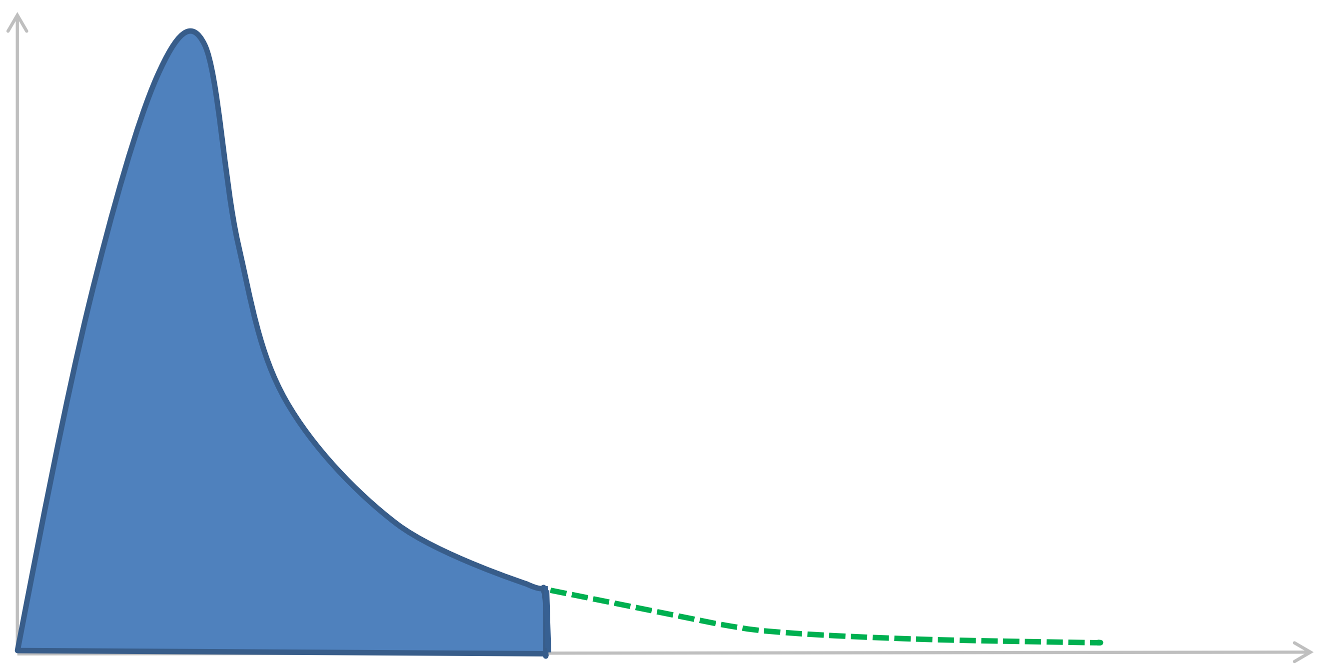
Can we use information from the region that has data to help predict for the region that has no data?



A Robust Optimization Approach

Assume the tail density is convex

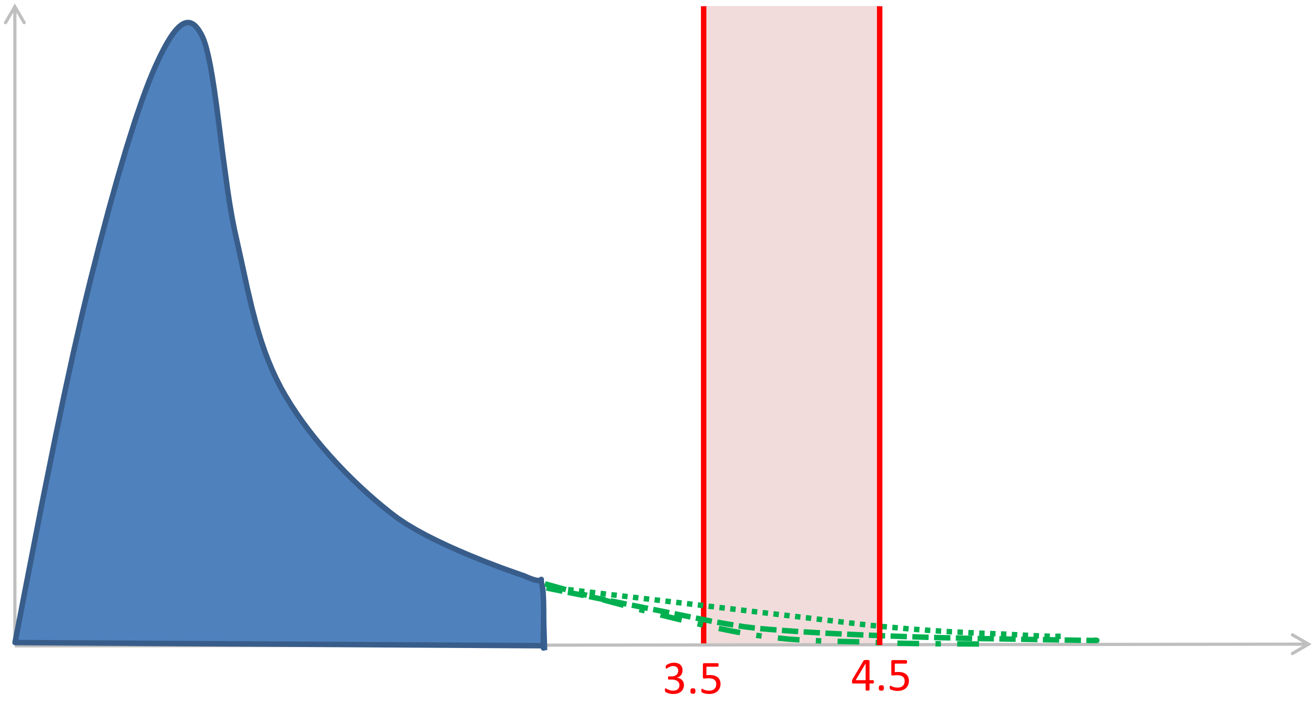
This assumption is satisfied by all known parametric density functions that has support on real line, e.g. normal, exponential, Weibull, Cauchy, GPD etc.



A Robust Optimization Approach

Worst-case estimation:

Maximize a given target performance measure, e.g. $P(3.5 < X < 4.5)$ among all convex tail extrapolations



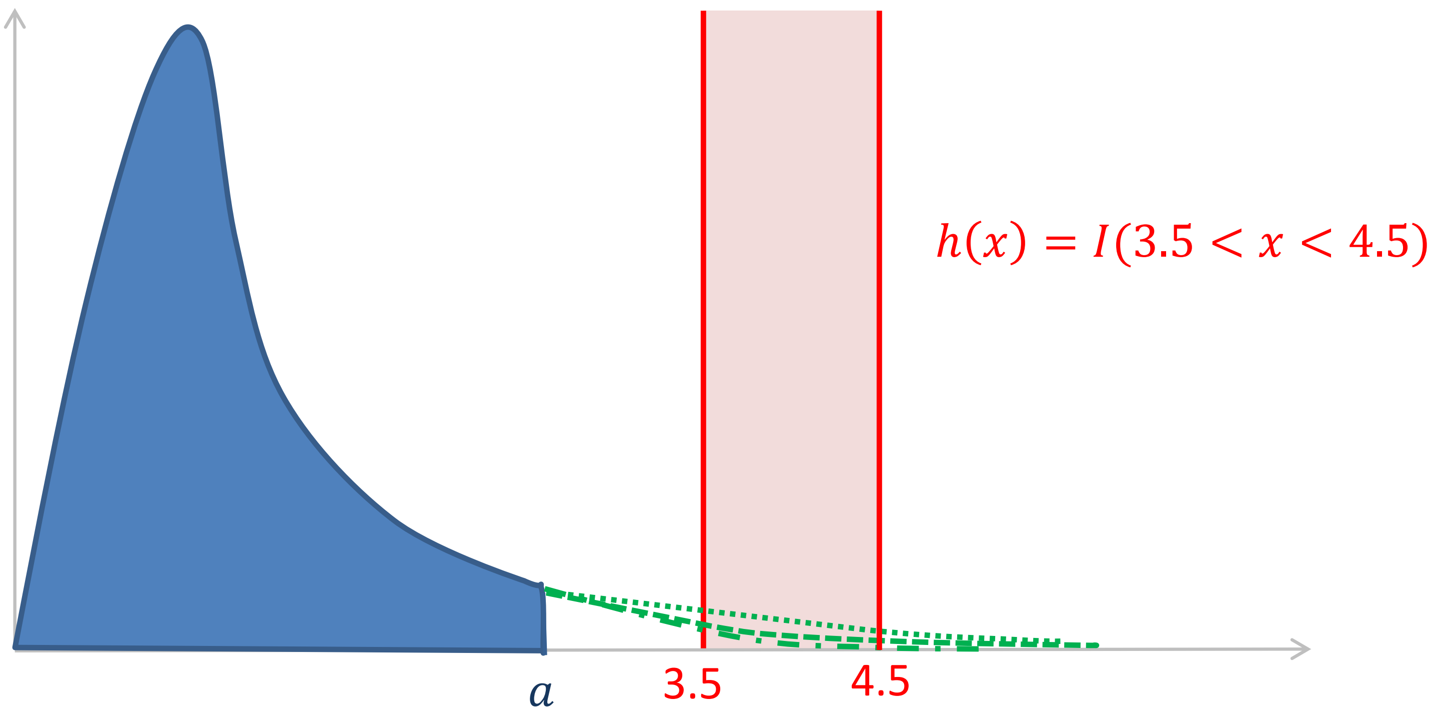
A Robust Optimization Approach

In general, given a bounded function $h(x)$, compute:

$$\max_f E_f[h(X)]$$

among all densities $f(x)$ that are

$$\begin{cases} \text{known for } x \le a \\ \text{convex for } x > a \end{cases}$$



Main Qualitative Results

Roughly speaking, worst-case convex tail is

- either extremely light or extremely heavy

Main Qualitative Results

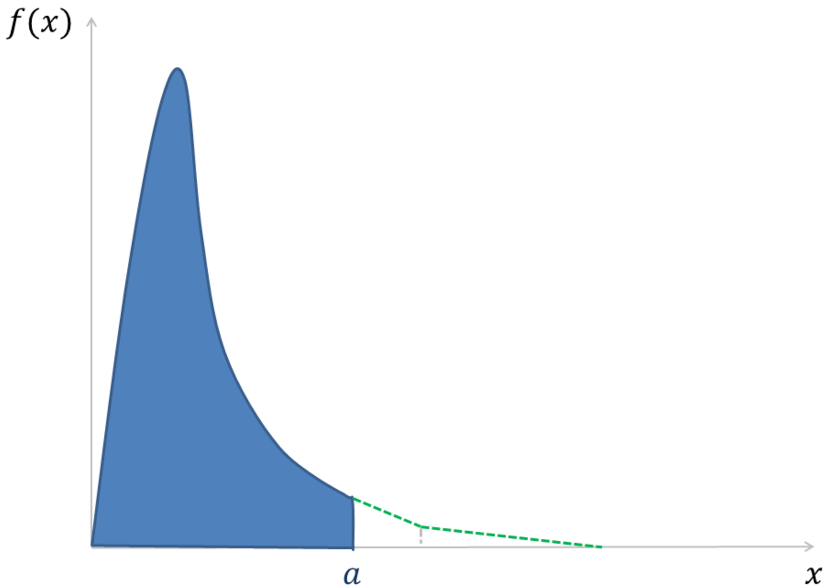
Roughly speaking, worst-case convex tail is

- either extremely light or extremely heavy
- in the form of piecewise linear density

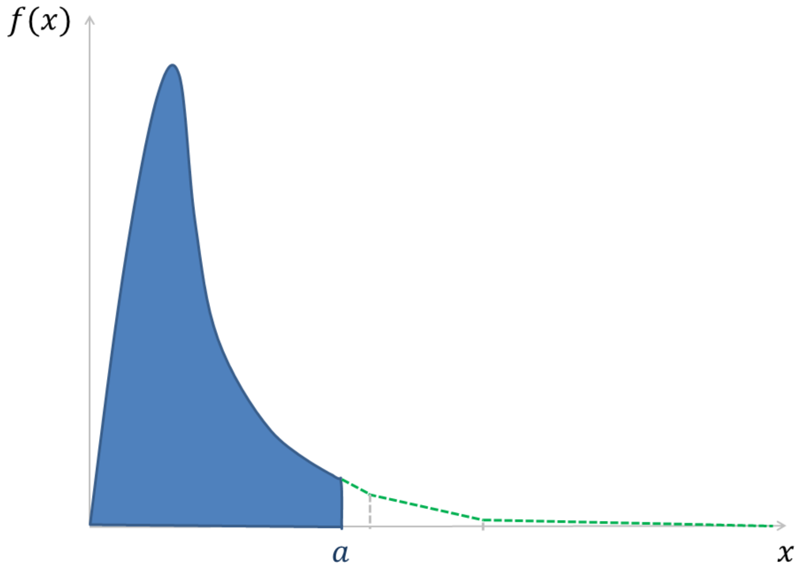
Main Qualitative Results

Roughly speaking, worst-case convex tail is
- either extremely light or extremely heavy
- in the form of piecewise linear density

Either of the following two closely related cases happen:
1. There exists an optimal tail that has three (or less) line segments
2. There exists a sequence of tails that has three (or less) line segments, the last segment getting "flatter and flatter", whose objective values converge to the optimum



Case 1: Extreme light tail



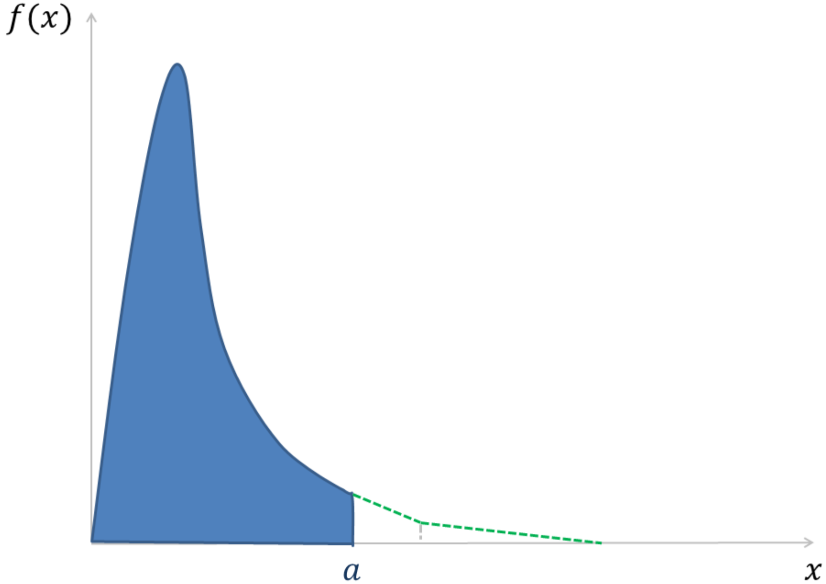
Case 2: "Extreme heavy" tail

Main Computational Results

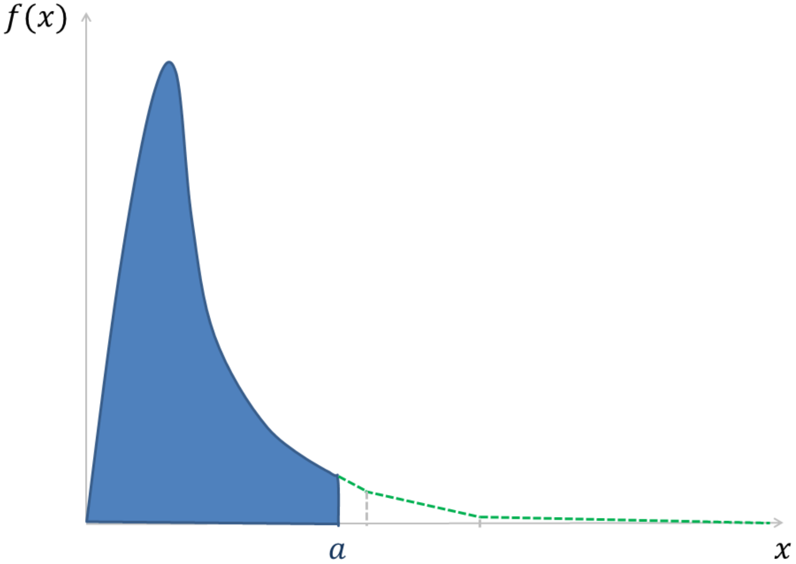
If $h(x)$ is bounded and satisfies an increasing-decreasing property, the worst-case tail can be found by two low-dimensional nonlinear programs:

Step 1: 1-variable optimization to distinguish between light and heavy tails

Step 2: 2-variable optimization to find the sequence of tails in the heavy-tail case



Case 1: Extreme light tail



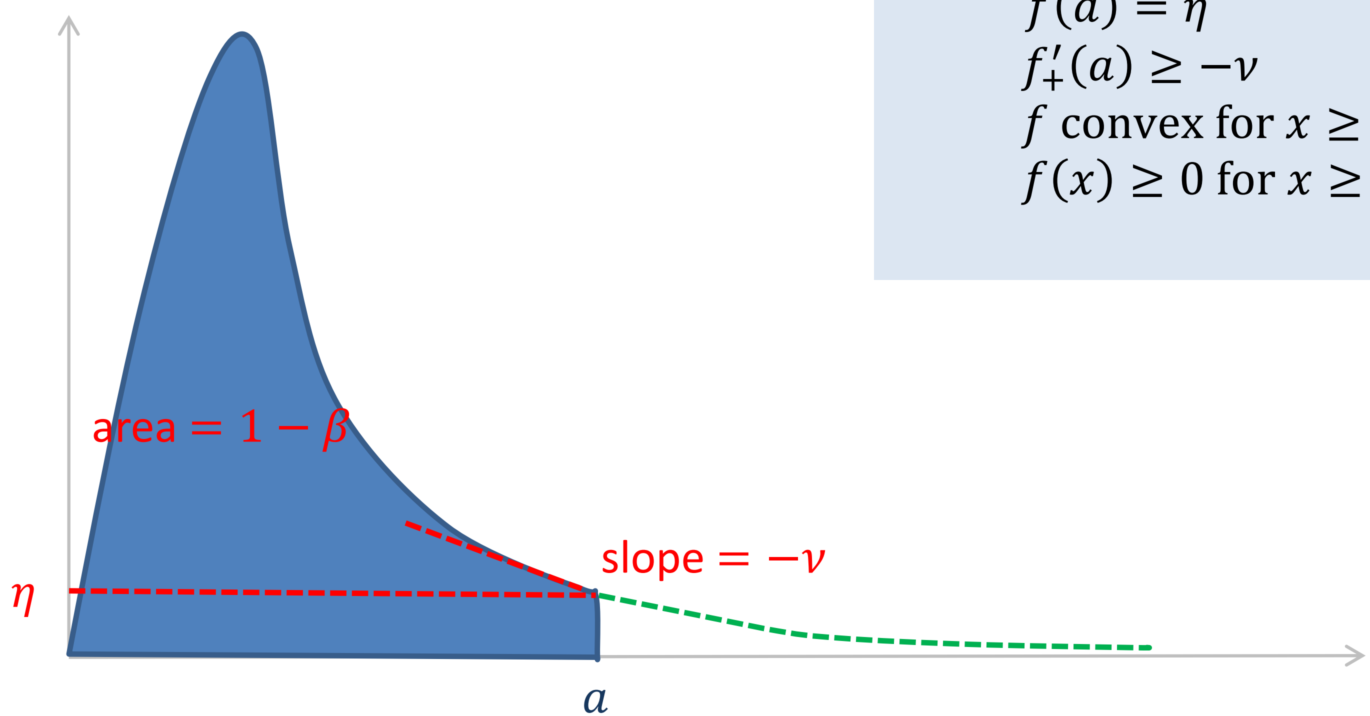
Case 2: "Extreme heavy" tail

Information for Convex Extrapolation

Given a threshold $\boldsymbol{a}$, a bounded performance function $\boldsymbol{h(x)}$

Three information needed for extrapolation:

- $\boldsymbol{\eta}$: value of density at $x = a$
- $\boldsymbol{-\nu}$: left derivative of f at $x = a$
- $\boldsymbol{\beta}$: tail probability for $X > a$



Optimization formulation:

max $\int_a^\infty h(x)f(x)dx$

subject to

$$\int_a^\infty f(x)\,dx = \beta$$
$$f(a) = \eta$$
$$f'_+(a) \geq -\nu$$
$$f \text{ convex for } x \geq a$$
$$f(x) \geq 0 \text{ for } x \geq a$$

Explanation of the Main Qualitative Results

Optimization formulation:

max $\int_a^\infty h(x)f(x)dx$
subject to

$$\int_a^\infty f(x)\,dx = \beta$$
$$f(a) = \eta$$
$$f'_+(a) \geq -\nu$$
$$f \text{ convex for } x \geq a$$
$$f(x) \geq 0 \text{ for } x \geq a$$

$\longleftrightarrow$

1-to-1 map from $f'(x)$ to a probability distribution on $[0, \infty)$

Moment problem:

max $\nu E[H(X)]$
subject to

$$E[X] = \frac{\eta}{\nu}$$
$$E[X^2] = \frac{2\beta}{\nu}$$

Explanation of the Main Qualitative Results

Optimization formulation:

max $\int_a^\infty h(x)f(x)dx$
subject to

$$\int_a^\infty f(x)\,dx = \beta$$
$$f(a) = \eta$$
$$f'_+(a) \geq -\nu$$
$$f \text{ convex for } x \geq a$$
$$f(x) \geq 0 \text{ for } x \geq a$$

$\longleftrightarrow$

1-to-1 map from $f'(x)$ to a probability distribution on $[0, \infty)$

Moment problem:

max $\nu E[H(X)]$
subject to

$$E[X] = \frac{\eta}{\nu}$$
$$E[X^2] = \frac{2\beta}{\nu}$$

- For the moment problem, it suffices to consider probability measures that have discrete support on at most three points

Explanation of the Main Qualitative Results

Optimization formulation:

max $\int_a^\infty h(x)f(x)dx$
subject to

$$\int_a^\infty f(x)\,dx = \beta$$
$$f(a) = \eta$$
$$f'_+(a) \geq -\nu$$
$$f \text{ convex for } x \geq a$$
$$f(x) \geq 0 \text{ for } x \geq a$$

⟷

1-to-1 map from $f'(x)$ to a probability distribution on $[0,\infty)$

Moment problem:

max $\nu E[H(X)]$
subject to

$$E[X] = \frac{\eta}{\nu}$$
$$E[X^2] = \frac{2\beta}{\nu}$$

- For the moment problem, it suffices to consider probability measures that have discrete support on at most three points
- Kinks in the piecewise linear density correspond to the support points of the optimal distribution in the moment problem

Step 1: Distinguishing between Light and Heavy Tails

Given $\boldsymbol{h(x)}$, and $\boldsymbol{\eta, \nu, \beta}$, distinguishing light vs heavy tail requires a one-dimensional line search (L. & Mottet '17)

Define

$$\mu = \frac{\eta}{\nu}, \qquad \sigma = \frac{2\beta}{\nu}, \qquad H(x) = \int_0^x \int_0^u h(v+a)\, dv\, du,$$

$$\lambda = \text{limsup}_{x\to\infty} \frac{H(x)}{x^2} < \infty$$

Solve

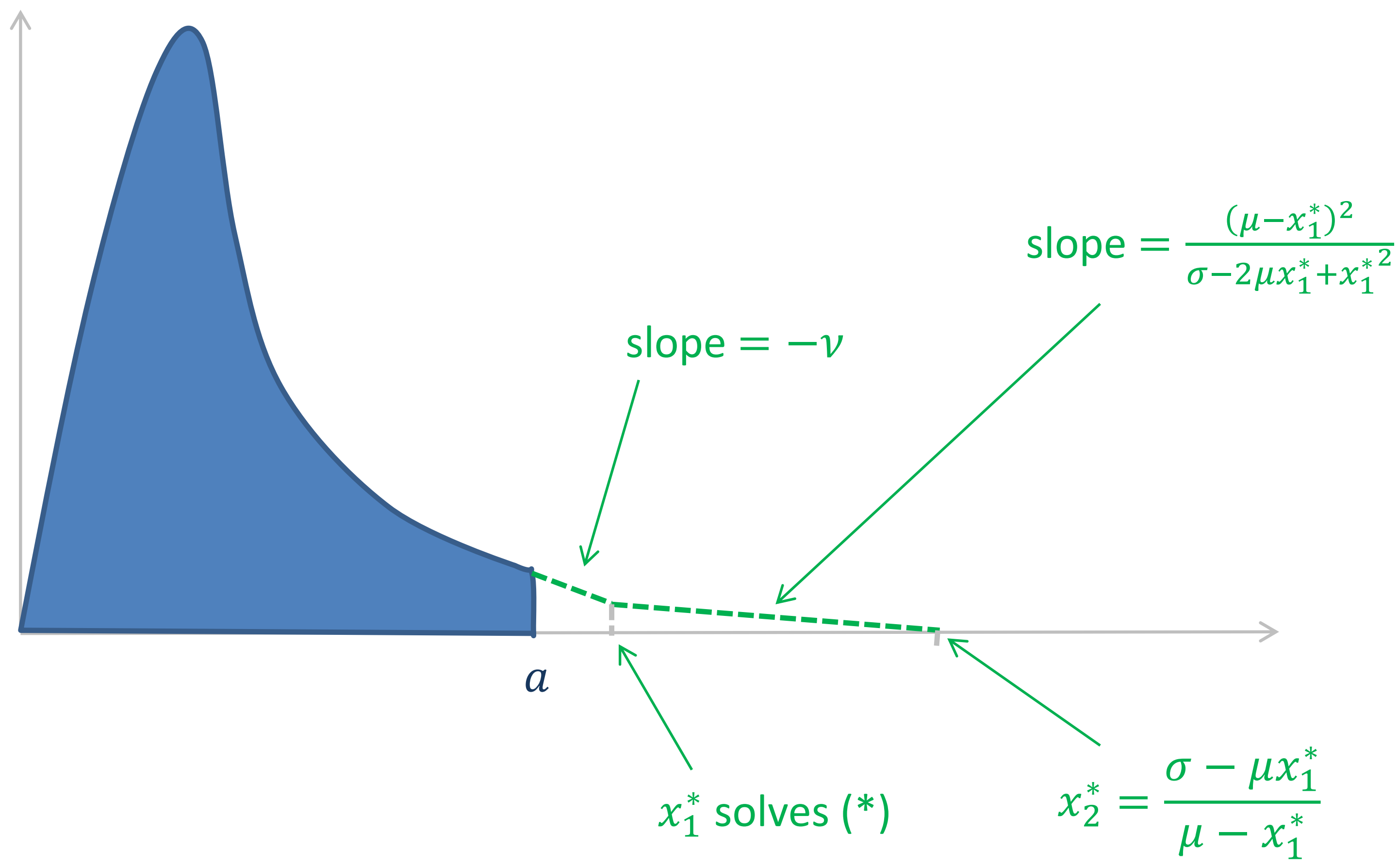
$$\max_{x_1 \in [0,\mu)} \frac{\sigma - \mu^2}{\sigma - 2\mu x_1 + x_1^2} H(x_1) + \frac{(\mu - x_1)^2}{\sigma - 2\mu x_1 + x_1^2} H\left(\frac{\sigma - \mu x_1}{\mu - x_1}\right) \qquad (*)$$

If there is an optimal solution $x_1^* \Rightarrow$ light tail

If there is no optimal solution, i.e. optimality occurs at $\mu \Rightarrow$ heavy tail

Step 1 (cont'd): Finding the Worst-Case Light Tail

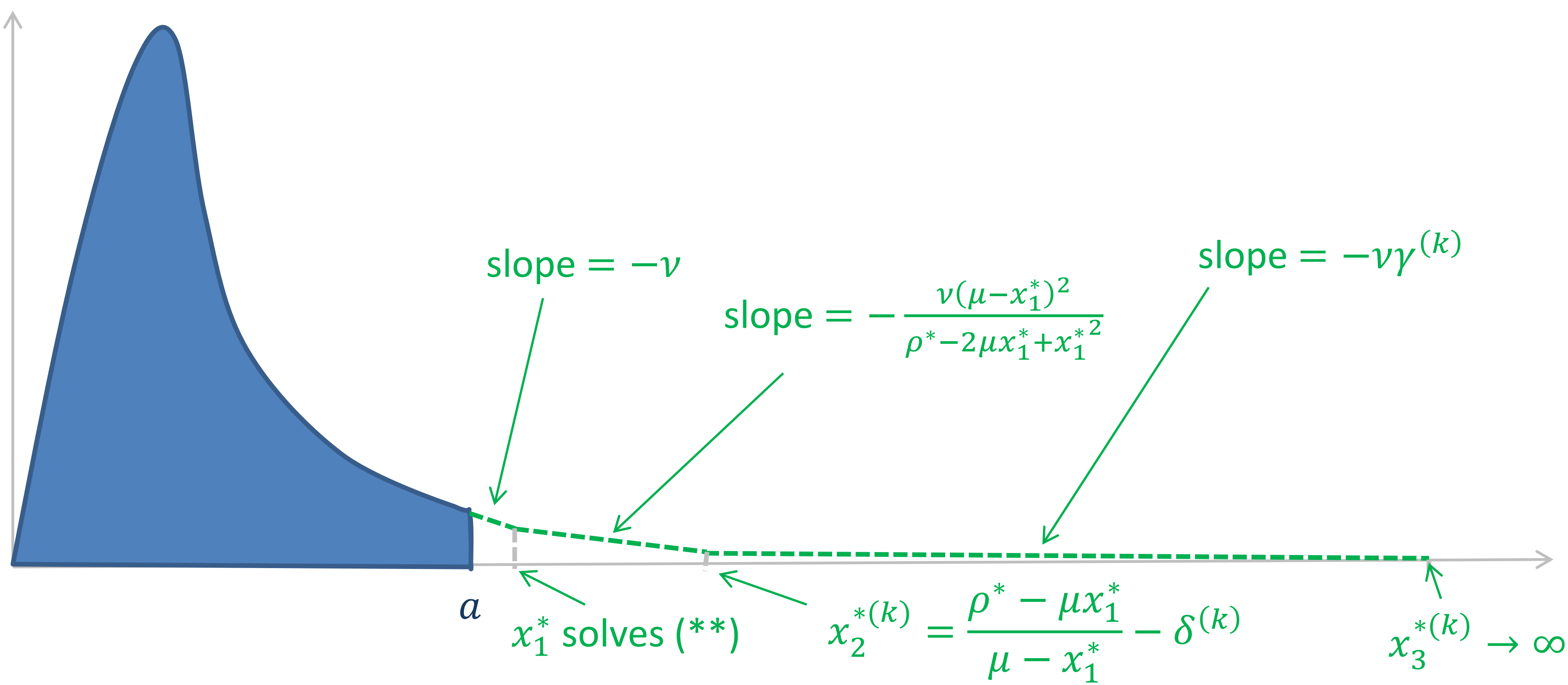
Light-tail Case: The optimization (*) characterizes the optimality



Step 2: Finding the Worst-case Heavy-tailed Sequence

Heavy-tail case: Solve an additional two-dimensional optimization

$$\max_{x_1\in[0,\mu),\rho\in[\mu^2,\sigma]} \frac{\rho-\mu^2}{\rho-2\mu x_1+x_1^2}\left(H(x_1)-\lambda x_1^2\right)+\frac{(\mu-x_1)^2}{\rho-2\mu x_1+x_1^2}\left(H\left(\frac{\sigma-\mu x_1}{\mu-x_1}\right)\right) \quad (**)$$



Related Literature

Robust extremal analysis:

– Worst-case copula: Balkema & Embrechts ’07, Wang & Wang ’11, Puccetti ’13, Puccetti & Ruschendorf ‘13…
– Distance-based constraints for extremal bound: Atar et al. ‘13, Engelke & Ivanovs ‘17, Blanchet & Murthy ’16
– Worst-case extremal coefficient: Stoev et al. ?

Optimization formulations and techniques:

– Shape-constrained problems: Li et al. ‘16, Popescu ’05, van Parys et al. ‘15
– Moment problems: Birge & Wets ‘87, Birge ‘90, Bertsimas & Popescu ‘05, Popescu ‘05, Karr ‘83, Winkler ‘88, Smith ’95, Bertsimas & Natajaran ’07…
– Distributionally robust optimization: Delage & Ye ‘10, Goh & Sim ‘09, Ben-Tal et al. ‘13, Goldfarb & Iyengar ’03, Wiesemann et al. ’14, Li et al. ‘18…

Back to Data..

In practice, η, ν, β need to be estimated

Optimization formulation:

max $\int_a^\infty h(x) f(x) dx$

subject to

$$\int_a^\infty f(x)\, dx = \beta$$

$$f(a) = \eta$$

$$f'_+(a) \geq -\nu$$

$$f \text{ convex for } x \geq a$$

$$f(x) \geq 0 \text{ for } x \geq a$$

Back to Data..

In practice, η, ν, β need to be estimated

Optimization formulation:

max $\int_a^\infty h(x)f(x)dx$
subject to

$$\underline{\beta} \leq \int_a^\infty f(x)\,dx \leq \bar{\beta}$$

$$\underline{\eta} \leq f(a) \leq \bar{\eta}$$

$$f'_+(a) \geq -\bar{\nu}$$

$$f \text{ convex for } x \geq a$$

$$f(x) \geq 0 \text{ for } x \geq a$$

Relax the worst-case optimization where $\left[\underline{\beta}, \overline{\beta}\right], \left[\underline{\eta}, \bar{\eta}\right], \bar{\nu}$ are the 95% joint confidence intervals for the respective quantities

Back to Data..

In practice, η, ν, β need to be estimated

Optimization formulation:

max $\int_a^\infty h(x) f(x) dx$
subject to

$$\underline{\beta} \leq \int_a^\infty f(x)\, dx \leq \bar{\beta}$$
$$\underline{\eta} \leq f(a) \leq \bar{\eta}$$
$$f'_+(a) \geq -\bar{\nu}$$
$$f \text{ convex for } x \geq a$$
$$f(x) \geq 0 \text{ for } x \geq a$$

Relax the worst-case optimization where $\left[\underline{\beta}, \overline{\beta}\right]$, $\left[\underline{\eta}, \bar{\eta}\right]$, $\bar{\nu}$ are the 95% joint confidence intervals for the respective quantities

The optimal value provides a valid upper bound with 95% confidence

Back to Data..

In practice, η, ν, β need to be estimated

Optimization formulation:

max $\int_a^\infty h(x)f(x)dx$
subject to

$$\underline{\beta} \leq \int_a^\infty f(x)\,dx \leq \bar{\beta}$$
$$\underline{\eta} \leq f(a) \leq \bar{\eta}$$
$$f'_+(a) \geq -\bar{\nu}$$
$$f \text{ convex for } x \geq a$$
$$f(x) \geq 0 \text{ for } x \geq a$$

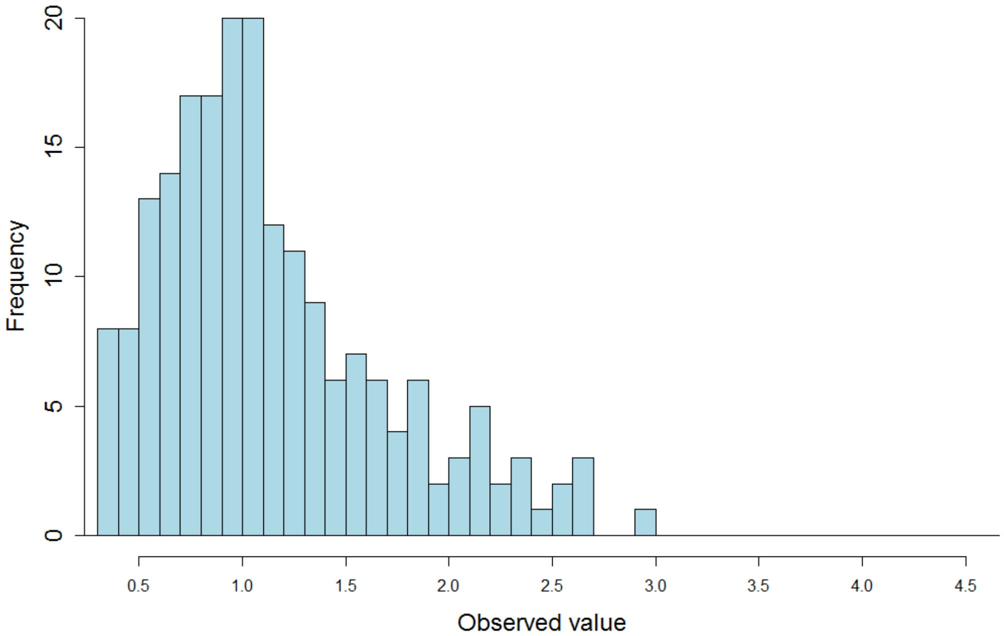
Relax the worst-case optimization where $\left[\underline{\beta}, \overline{\beta}\right]$, $\left[\underline{\eta}, \bar{\eta}\right]$, $\bar{\nu}$ are the 95% joint confidence intervals for the respective quantities

The optimal value provides a valid upper bound with 95% confidence

The nonlinear programs can be modified with one more variable

Example

200 data points from unknown distribution



Estimate of $P(4 < X < 5)$

Method	95% Confidence Upper Bound
Truth	2.1×10^{-3}
GPD fitting	1.5×10^{-3}
Robust approach	6.6×10^{-3}
Empirical	N/A

Example

GPD fitting

c	d	Truth	Mean upper bound	Coverage probability
4	5	2.13×10^{-3}	4.06×10^{-3}	0.67
5	6	4.73×10^{-4}	9.93×10^{-4}	0.54
6	7	1.19×10^{-4}	3.01×10^{-4}	0.44
7	8	3.37×10^{-5}	1.06×10^{-4}	0.37
8	9	1.04×10^{-5}	4.38×10^{-5}	0.30

The robust approach

c	d	Truth	Mean upper bound	Coverage probability
4	5	2.13×10^{-3}	1.09×10^{-2}	1
5	6	4.73×10^{-4}	6.87×10^{-3}	1
6	7	1.19×10^{-4}	4.98×10^{-3}	1
7	8	3.37×10^{-5}	3.90×10^{-3}	1
8	9	1.04×10^{-5}	3.20×10^{-3}	1

Choosing Threshold for the "Tail Region"

Pickands-Balkema-de Haan Theorem: A threshold where the excess loss data are believed to follow GPD

- Bias-variance tradeoff

Robust Optimization: A threshold where the constraints are believed to hold (e.g., convexity)

- This threshold can be chosen lower than POT
- Conservativeness-variance tradeoff

Coping with Conservativeness

Add additional constraints, e.g., power moments, quantiles, or constraints implied by the Kolmogorov-Smirnov statistic, etc.

max $E_f[h(X)]$

subject to

$$\underline{\beta} \leq P_f(X \geq a) \leq \bar{\beta}$$

$$\underline{\eta} \leq f(a) \leq \bar{\eta}$$

$$f_+'(a) \geq -\bar{\nu}$$

$$-z + \frac{i}{n} \leq P\left(X \leq X_{(i)} \middle| X \geq a\right) \leq z + \frac{i-1}{n}, i = 1, \dots, n$$

$$f \text{ convex for } x \geq a$$

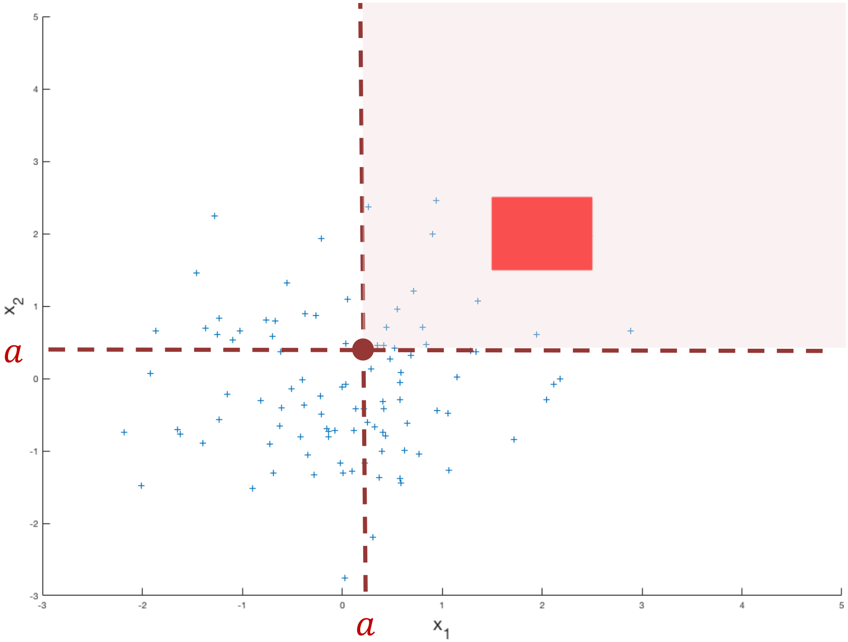
$$f(x) \geq 0 \text{ for } x \geq a$$

Solution via:

- Duality and semidefinite programming (e.g., Bertsimas et al. ‘17, Bertsimas & Popescu ’05, Van Parys et al. ‘16)
- Generalized linear programming / column generation with a “heavy-tail” slack variable (Mottet & L. ’18)

Multivariate Generalizations

Estimate a probability involving a multivariate extremes, e.g.,
$P(1.5 \leq X_1 \leq 2.5, 1.5 \leq X_2 \leq 2.5)$



$\max_f P_f(1.5 \leq X_1 \leq 2.5, 1.5 \leq X_2 \leq 2.5)$
subject to

$\underline{\beta} \leq P_f(X_1 \geq a, X_2 \geq a) \leq \overline{\beta}$

$\underline{\eta}_1 \leq f(X_1 = a; X_2 \geq a) \leq \overline{\eta}_1$

$\underline{\eta}_2 \leq f(X_2 = a; X_1 \geq a) \leq \overline{\eta}_2$

$f(x_1, x_2)$ coordinate-wise non-increasing for $x_i \geq a$

Robust Simulation

Estimate $P(X_1 + \cdots + X_n > b)$

No closed form but simulable

max $\quad P_f(X_1 + \cdots + X_n > b)$

subject to

$$\underline{\beta} \leq P_f(X \geq a) \leq \bar{\beta}$$

$$\underline{\eta} \leq f(a) \leq \bar{\eta}$$

$$f'_+(a) \geq -\bar{\nu}$$

$$f \text{ convex for } x \geq a$$

$$f(x) \geq 0 \text{ for } x \geq a$$

Robust Simulation

Estimate a simulation-based performance measure $\psi(f)$ that depends on the input distribution f

No closed form but simulable

max $\psi(f)$
subject to
$f \in feasible\ region$

Solution via stochastic gradient descent / sequential linearization
(L. & Mottet '15, Ghosh & L. '18)

Conclusion

A robust and nonparametric approach to extremal estimation:

- Alternative to conventional extreme value theory
- Worst-case optimization subject to tail constraints, e.g., convex tail extrapolation
- Qualitative characterization (extreme light and heavy tail) and quantitative solution methods (e.g., line search, semidefinite programming, column generation)

References

Lam, H. & Mottet, C., "Tail analysis without parametric models: A worst-case perspective", *Operations Research*, 65(6), 1696-1711, 2017

Lam, H. & Mottet, C., "Simulating tail events with unspecified tail models", *Proceedings of the Winter Simulation Conference,* 2015

Ghosh, S. & Lam, H., "Robust analysis in stochastic simulation: Computation and performance guarantees", *forthcoming in Operations Research*, 2018.

Mottet, C. & Lam, H., "On optimization over tail distributions", *under review,* 2018, available at: https://arxiv.org/abs/1711.00573 .