# ACM Winter School 2025

# Lecture 3: Extreme Value Theory: Combating Model Risk

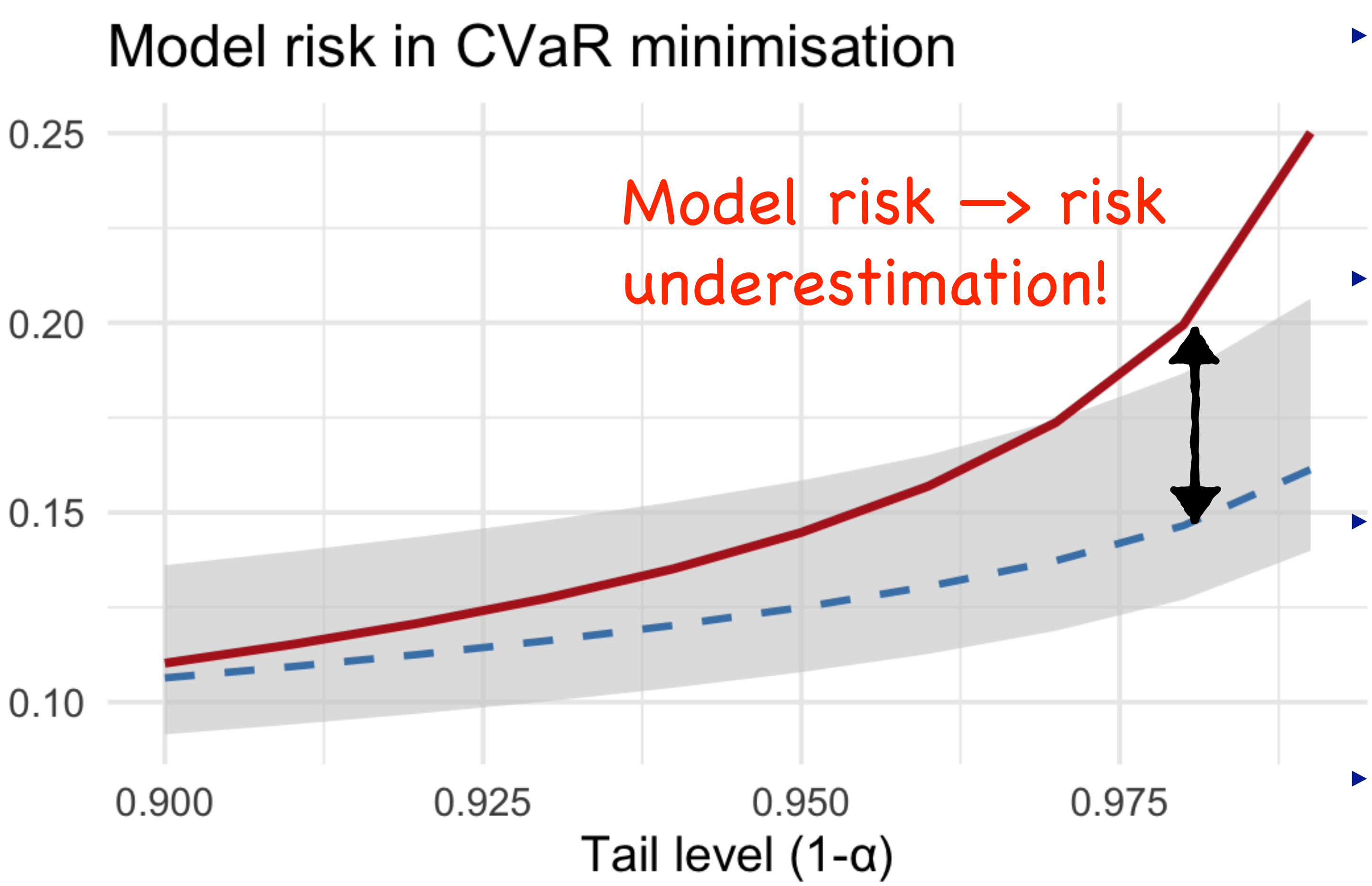
# Recap: how errors in input modelling propagate

(Intuition via portfolio optimisation example)

Problem: Find the investment strategy that minimises

$$\rho(x, P) = \text{CVaR}_{1-\alpha}(x, P)$$

where $\text{CVaR}_{1-\alpha}(x, P) = E_P[\varphi(x, \xi) \mid \varphi(x, \xi) \geq \text{VaR}_{1-\alpha}(x, P)]$




- Gaussian distribution underestimates outliers in data (Model error)
- Leads to underestimation of risk (percolation of error)
- Alternative: sample distribution + bootstrap
- Catch: Paucity of tail samples $\Longrightarrow$ fragility to data realisation

Samples relevant to risk evaluation are mis-represented in model!

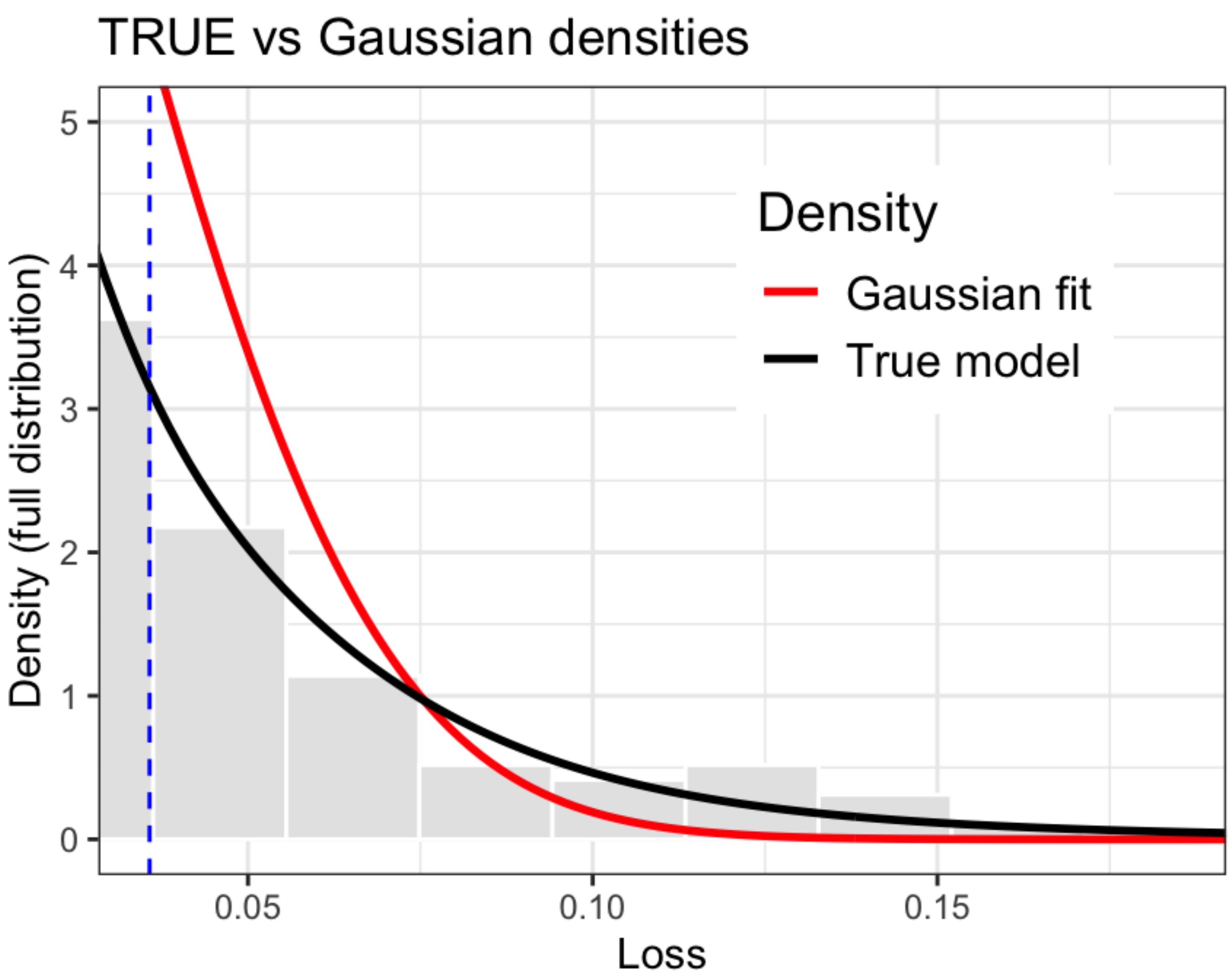
# Recap: how errors in input modelling propagate

(Intuition via portfolio optimisation example)

Problem: Find the investment strategy that minimises

$$\rho(x, P) = \text{CVaR}_{1-\alpha}(x, P)$$

where $\text{CVaR}_{1-\alpha}(x, P) = E_P[\varphi(x, \xi) \mid \varphi(x, \xi) \geq \text{VaR}_{1-\alpha}(x, P)]$




- Gaussian distribution underestimates outliers in data (Model error)
- Leads to underestimation of risk (percolation of error)
- Alternative: sample distribution + bootstrap
- Catch: Paucity of tail samples $\Longrightarrow$ fragility to data realisation

Samples relevant to risk evaluation are mis-represented in model!

# Model Risk: how errors in input modelling propagate

(Comparing Rates)

Problem: Find the investment strategy that minimises

$$\rho(x, P) = \text{CVaR}_{1-\alpha}(x, P)$$

where $\text{CVaR}_{1-\alpha}(x, P) = E_P[\varphi(x, \xi) \mid \varphi(x, \xi) \geq \text{VaR}_{1-\alpha}(x, P)]$

What went wrong: The risk measure used was CVaR, which for a Gaussian portfolio has a closed form!

- For Gaussian $\xi$: $C_{1-\alpha}(x, \hat{P}) = \mu^\intercal x + \sqrt{x^\intercal \Sigma x}\ \dfrac{\phi(\Phi^{-1}(1-\alpha))}{\alpha}$
- For true data (heavy tailed): $C_{1-\alpha}(x, P) \asymp \alpha^{-1/\nu} c(x)$ , $\nu > 0$ and $c(x) > 0$

Observation: For every $x$, $C_{1-\alpha}(x, P) \gg C_{1-\alpha}(x, \hat{P})$

Implication: Minimum possible risk underestimated!!

# Model Risk: how errors in input modelling propagate

(A mathematical Theorem)

**Model Risk due to tail mis-specification:** Suppose the data distribution satisfies

$$\liminf_{t \to \infty} t^{-\rho} P(\|\xi\| > t) > 0. \quad \text{—— (i)}$$

for some $\rho > 0$. Let $\hat{P}$ be any distribution that satisfies

$$\limsup_{t \to \infty} t^{-\rho} P(\|\xi\| > t) = 0 \quad \text{—— (ii)}$$

Then $C_{1-\alpha}(x, \hat{P}) \ll C_{1-\alpha}(x, P)$, and hence $v^{\star}_{1-\alpha}(\hat{P}) \ll v^{\star}_{1-\alpha}(P)$ as $\alpha \to 0$ where for any distribution $Q$, $v^{\star}_{1-\alpha}(Q) = \min C_{1-\alpha}(x, Q)$

How to interpret the above theorem:

- Equation (i) $\rightarrow$ Extent of outliers in data
- Equation (ii) $\rightarrow$ Outliers captured by the nominal model

Mismatch in outliers $\Longrightarrow$ mismatch in minimum risk

# Model Risk: how errors in input modelling propagate

(Proof Sketch)

Model Risk due to tail mis-specification: Suppose the data distribution satisfies

$$\liminf_{t\to\infty} t^{-\rho} P(\|\xi\| > t) > 0. \quad \text{—— (i)}$$

for some $\rho > 0$. Let $\hat{P}$ be any distribution that satisfies

$$\limsup_{t\to\infty} t^{-\rho} P(\|\xi\| > t) = 0 \quad \text{—— (ii)}$$

Then $C_{1-\alpha}(x, \hat{P}) \ll C_{1-\alpha}(x, P)$, and hence $v^{\star}_{1-\alpha}(\hat{P}) \ll v^{\star}_{1-\alpha}(P)$ as $\alpha \to 0$ where for any distribution $Q$, $v^{\star}_{1-\alpha}(Q) = \min C_{1-\alpha}(x, Q)$

Proof sketch:

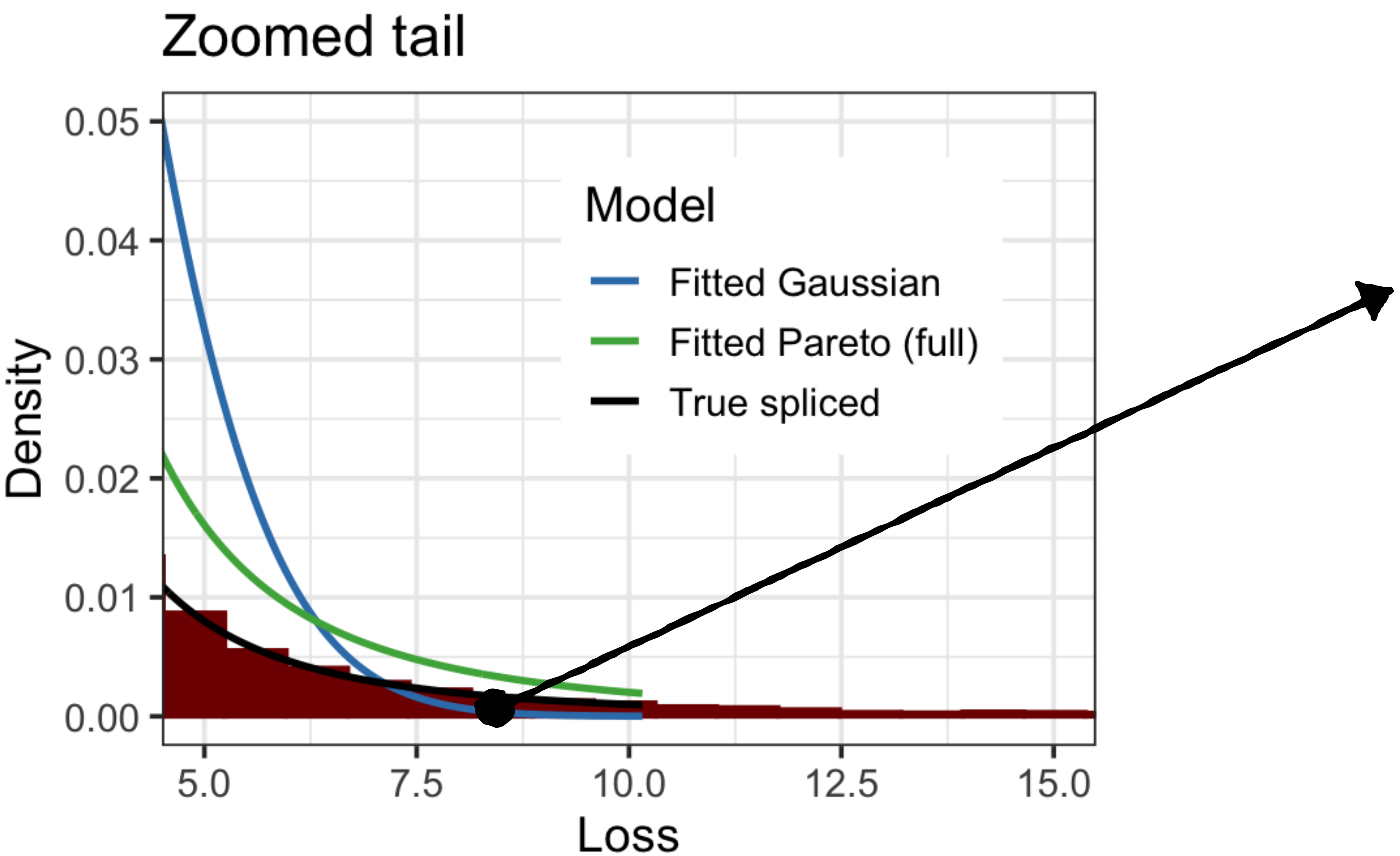
- Equation (i) $\Longrightarrow$ that for all $x$, $C_{1-\alpha}(x, P)$ is of order $t^{\rho}$
- Equation (ii) $\Longrightarrow$ that for all $x$, $C_{1-\alpha}(x, \hat{P})$ is $o(t^{\rho})$
- The above two $\Longrightarrow$ $v^{\star}_{1-\alpha}(P)$ is of order $t^{\rho}$, while $v^{\star}_{1-\alpha}(\hat{P}) = o(t^{\rho})$ ∎

# Model Risk: mitigation strategies

(How to reduce model risk?)

Risk minimisation: $\min_x \rho(x, \hat{P})$ where $\hat{P} \to$ nominal distribution

Issue: Nominal does not capture features of data relevant to risk evaluation



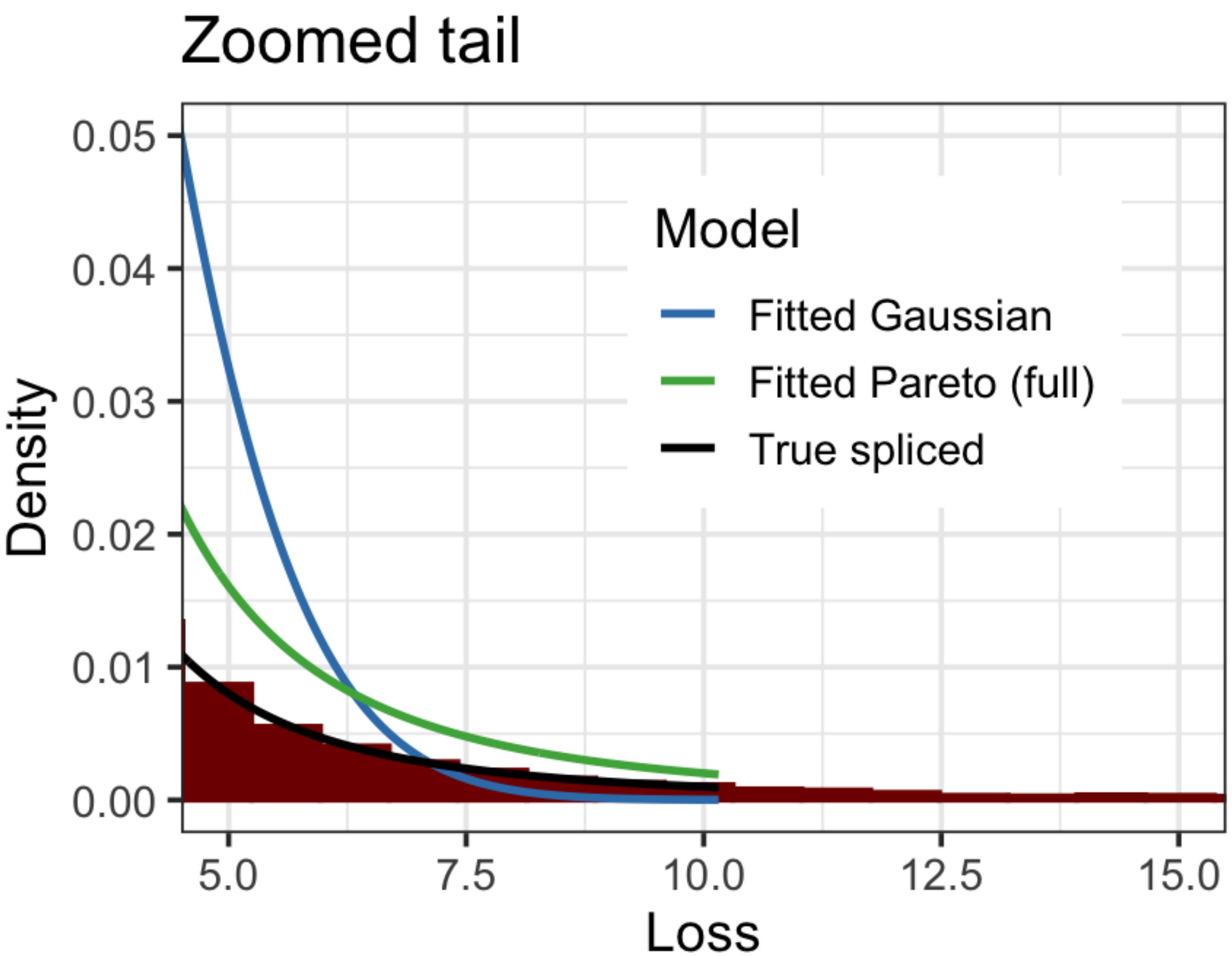

Gaussian model underestimates risk

# Model Risk: mitigation strategies

(How to reduce model risk?)

Risk minimisation: $\min_x \rho(x, \hat{P})$ where $\hat{P} \rightarrow$ nominal distribution

Issue: Nominal does not capture features of data relevant to risk evaluation




- Learning I: Let objective inform model choice.

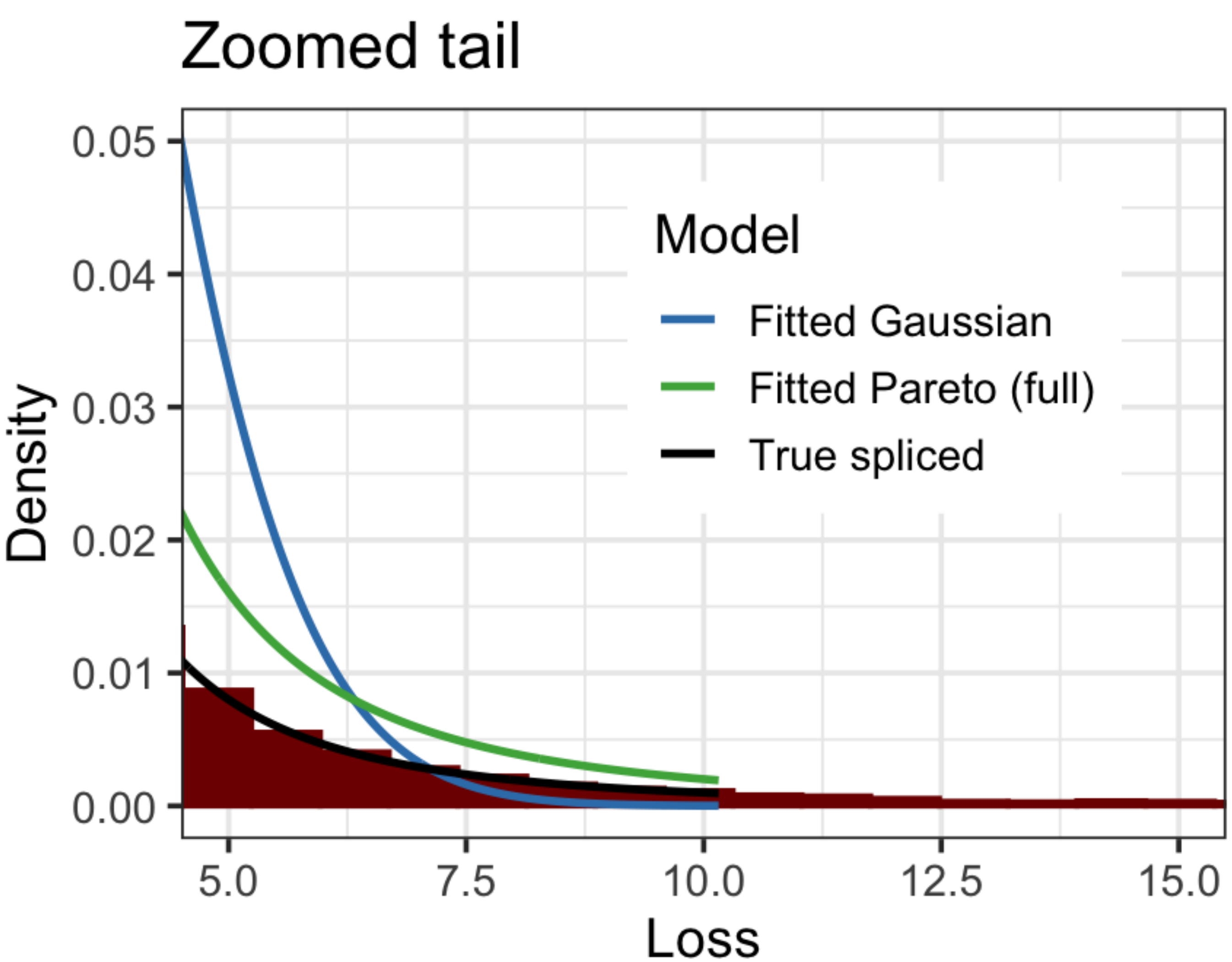
  (Gaussian fit to the bulk of the data. Objective → evaluating tail risk)

# Model Risk: mitigation strategies

(How to reduce model risk?)

Risk minimisation: $\min_x \rho(x, \hat{P})$ where $\hat{P} \to$ nominal distribution

Issue: Nominal does not capture features of data relevant to risk evaluation




- Learning I: Let objective inform model choice.
- Learning II: In evaluation of tail-risks, the extent of outlier captured in data and by the model should match

Catch: Outliers are rarely observed in sample!!

Up next: Extreme Value Theory - A structured way of modelling tails of data!

# Agenda for the course

| Session 1 | Session 2 | Session 3 |
|---|---|---|
| Input Modelling 🚂 | Copulas and Model Risk | Extreme Value Theory 🚂 |
| The bootstrap | The copula | Model Risk due to tail uncertainty |
| Inverse Transforms | Model Risk | |
| Sampling from multivariate distributions | | |

# Modelling extremes: A motivating example

(Conditional Value at Risk is Fragile)

Problem: In a risk management setting, you are required to assess the conditional value at risk (CVaR) of their portfolios. You have access to 500 samples of data, and regulatory requirements necessitate them to compute the CVaR at a risk level of 99%. They try the following two strategies:

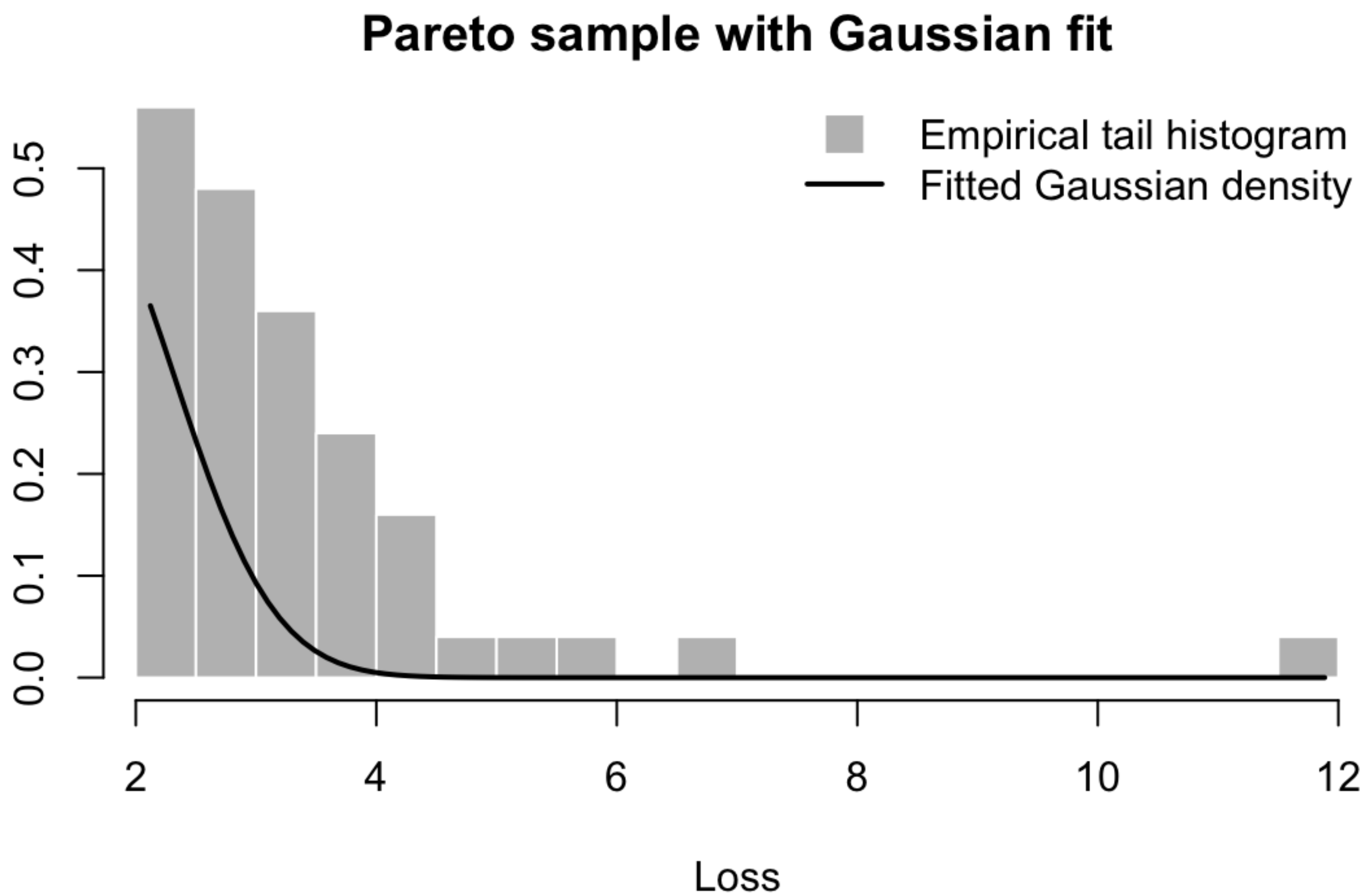
i) Use the data directly
ii) Use a Gaussian model to capture portfolio returns

You having seen some amount of QRM feel that something is fishy, and decide to intervene to improve their model

i) What could go wrong with either approaches?
ii) Suggest some ways in which this could be fixed

# Modelling extremes: A motivating example

(Pitfall - 1: Underestimation of tail)



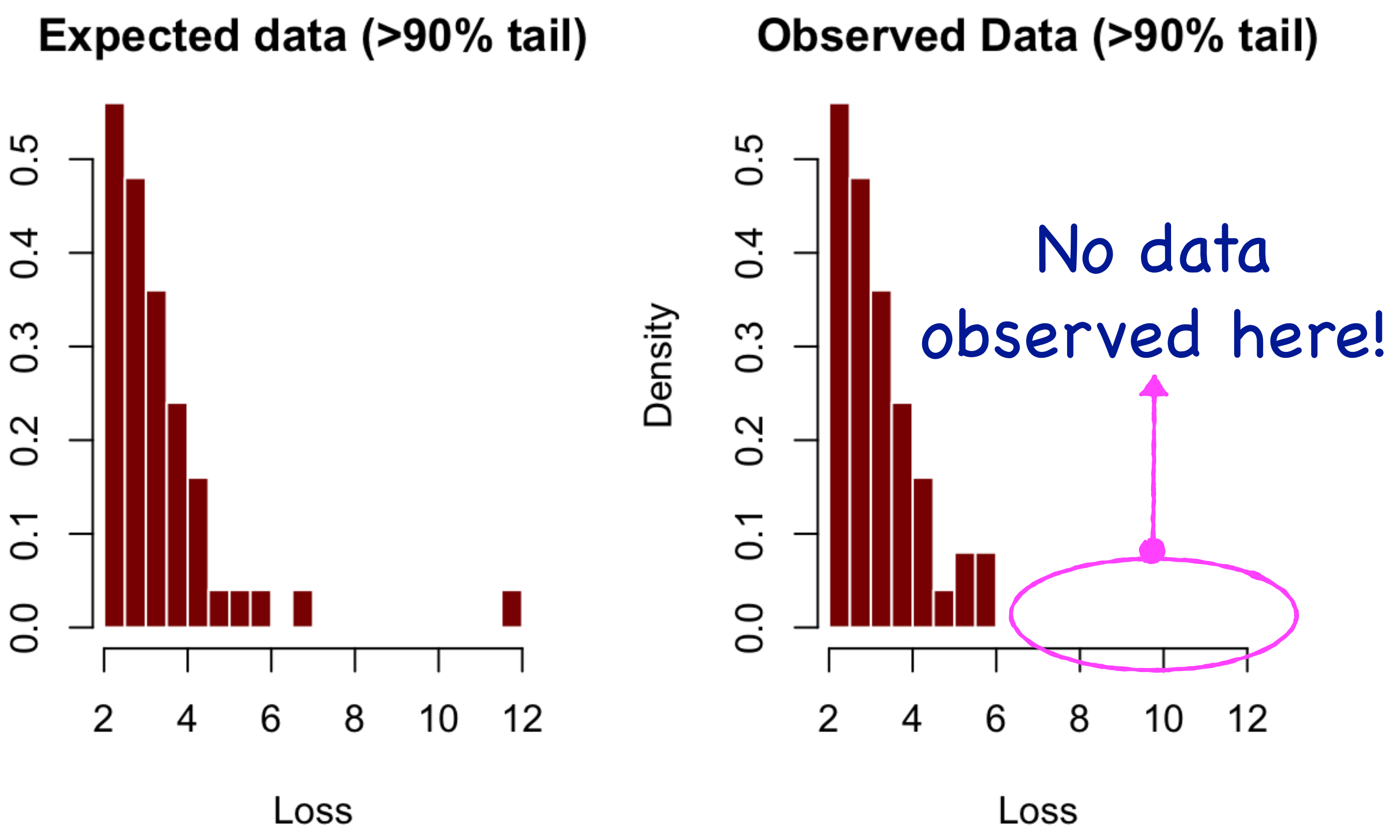

Attempt 1: Gaussian fit

- Tails severely under-represented
- Leads to a downward bias in CVaR estimation.

| Method | CVaR Estimate |
|---|---|
| Gaussian Fit | 3.78 |
| | |
| True CVaR | 6.96 |

# Modelling extremes: A motivating example

(Pitfall - 2: Fragility to tail samples)



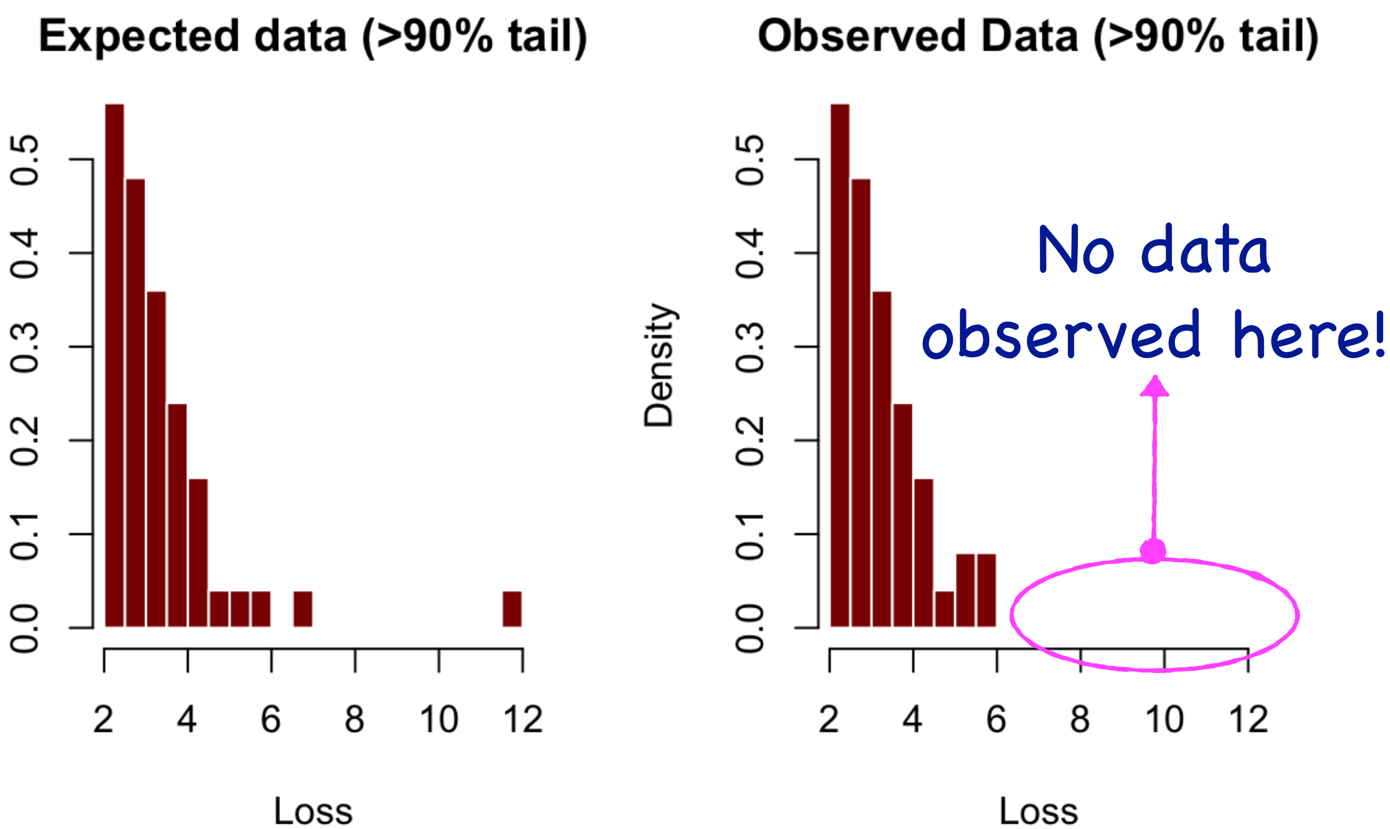

Attempt 2: Bootstrap the data

- Bootstrap cannot generate "new" samples
- Leads to a fragility of estimates to actual tail realisations

| Method | CVaR Estimate |
|---|---|
| Gaussian Fit | 3.78 |
| Bootstrap from data (95% CI) | (4.32 , 5.59) |
| True CVaR | 6.96 |

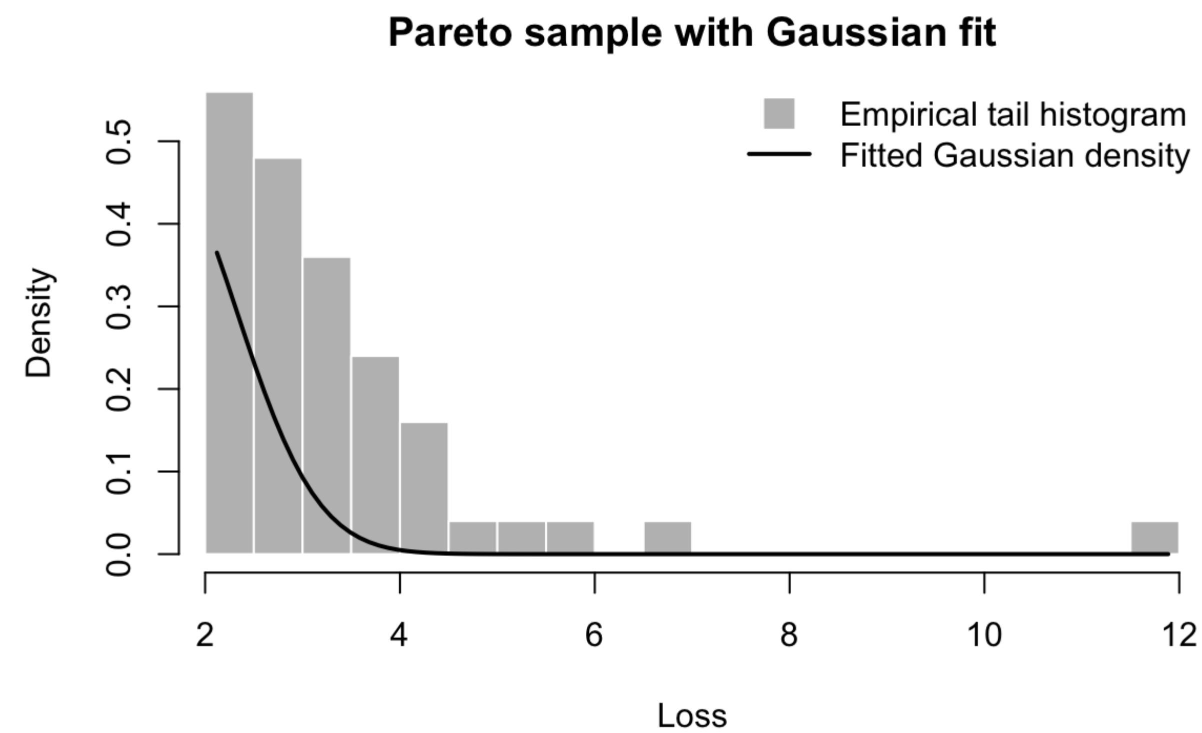
# Modelling extremes: A motivating example

(Summary of pitfalls)




Attempt: Bootstrap (fully model-free)

- Severe fragility to actual tail samples
- Statistically noisy risk evaluations (high variance estimates)




Attempt: distributional assumption

- Tails severely under-represented
- Systematic underestimation (biased estimates)

Question: Is there a systematic way of achieving a bias variance trade-off in problems involving tail risk estimation?

# Agenda for the course

## Session 1

### Input Modelling 🚂

- The bootstrap
- Inverse Transforms
- Sampling from multivariate distributions

## Session 2

### Copulas and Model Risk

- The copula
- Model Risk

## Session 3

### Extreme Value Theory 🚂

- Model Risk due to tail uncertainty
- The Pareto distribution

# Extreme Value Theory: A first pass

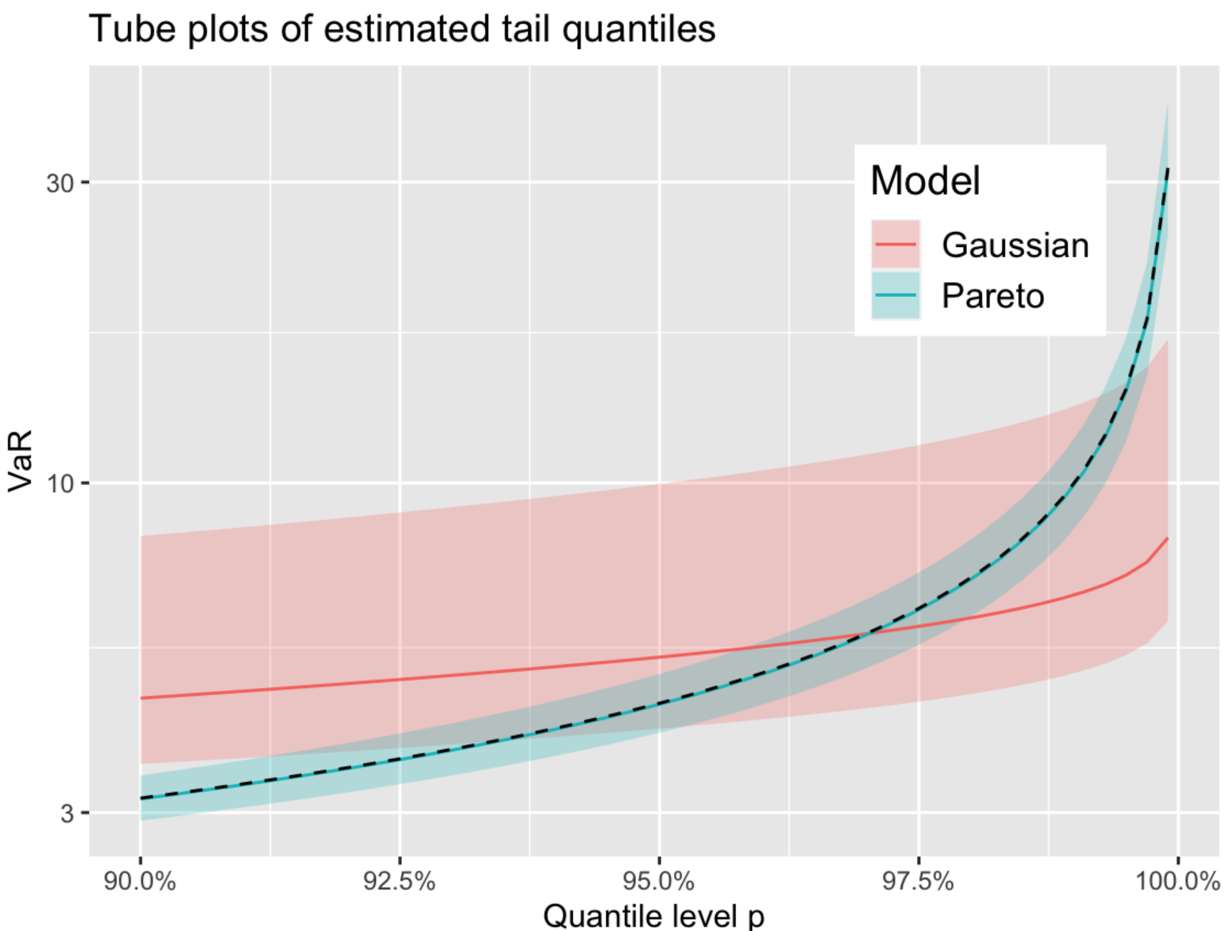
(Intuition)

Toy Assumption : Losses are Pareto distributed, for some $(\xi, \sigma, \mu) > 0$,

$$P(Z \geq t) = \left(1 + \frac{t - \mu}{\sigma}\right)^{-1/\xi}$$

- Tail behaviour is parametric. Therefore, standard estimation techniques suffice to capture extreme behaviour!

- $\xi \rightarrow$ captures heaviness of tail




- Gaussian distribution underestimates quantile —> model risk

- Parametric estimation can resolve the issue if the Pareto model is correct.

- Estimation of $\xi$ allows for quantification of tail strength

# Extreme Value Theory: A first pass

(Failure of naive approach)

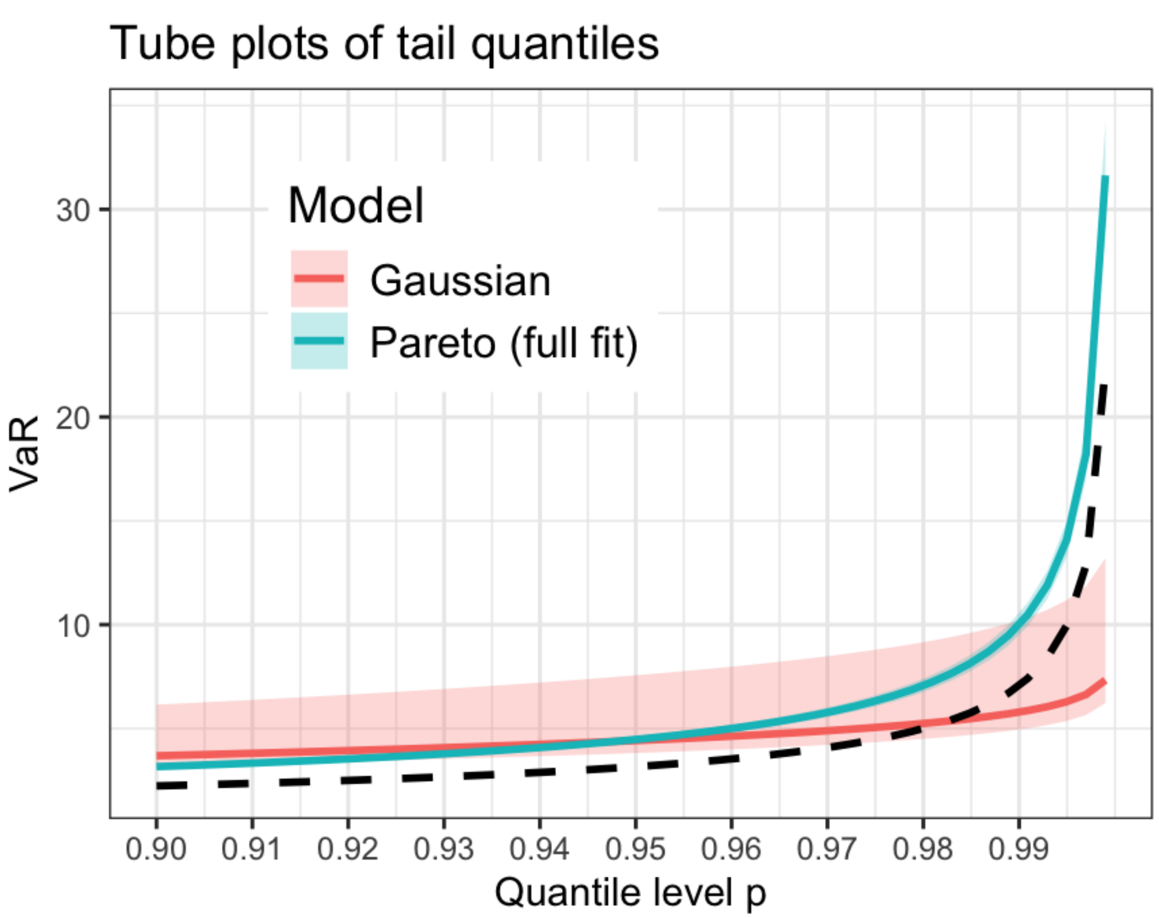
Toy Assumption : ~~Losses are Pareto distributed, for some $(\xi, \sigma, \mu) > 0$,~~

$$\cancel{P(Z \geq t) = \left(1 + \frac{t - \mu}{\sigma}\right)^{-1/\xi}}$$

Tail behaviour is parametric. Therefore, standard estimation techniques suffice to capture extre

More realistic: There are small deviations from Pareto!

$\xi \to$ captures heaviness of tail




- Suppose the density of $Z$ equals

$$f(z) = \phi(z)\mathbf{1}\,(z \leq 2) + c_2 z^{-(1/\xi+1)}\,\mathbf{1}(z > 2)$$

- You fit a Pareto distribution to this data distribution and computed VaR...

# Extreme Value Theory: A first pass

(Failure of naive approach)

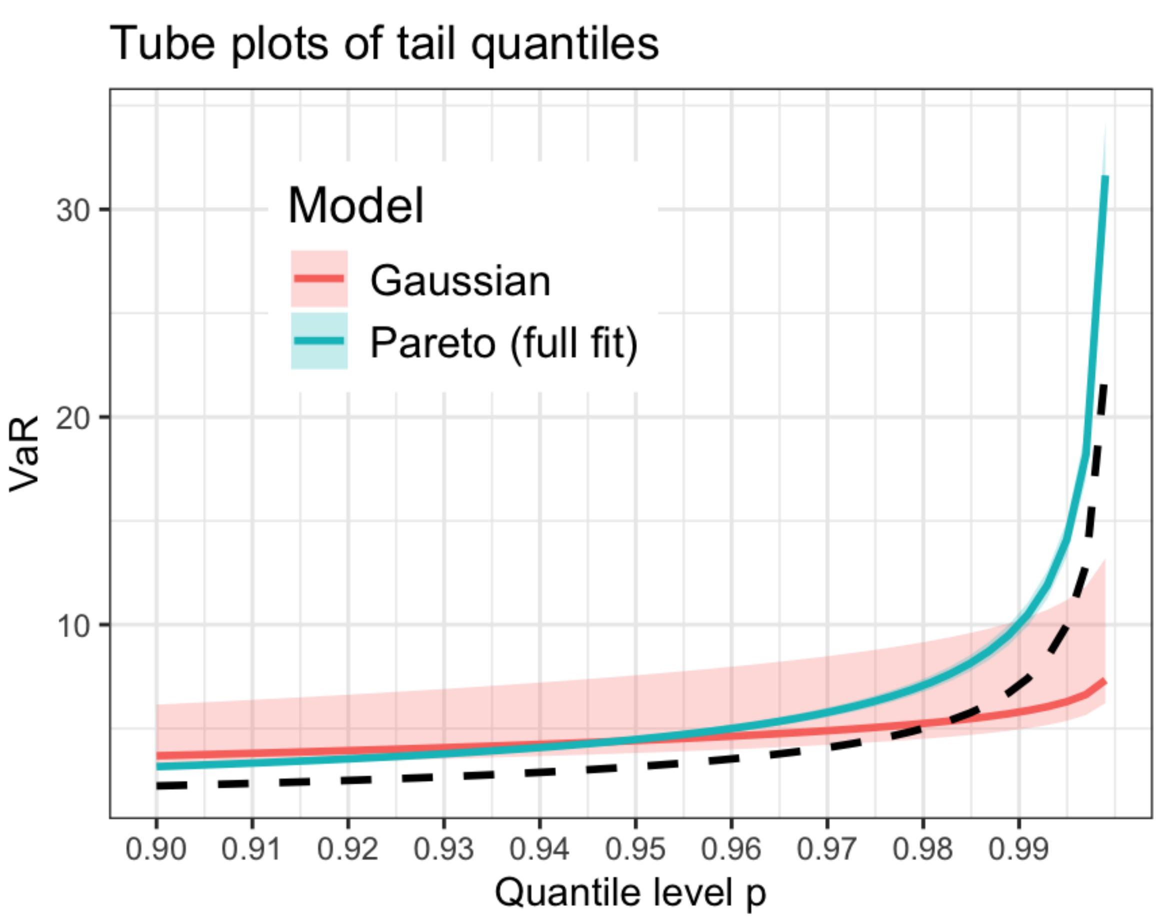
Toy Assumption : ~~Losses are Pareto distributed, for some $(\xi, \sigma, \mu) > 0$,~~

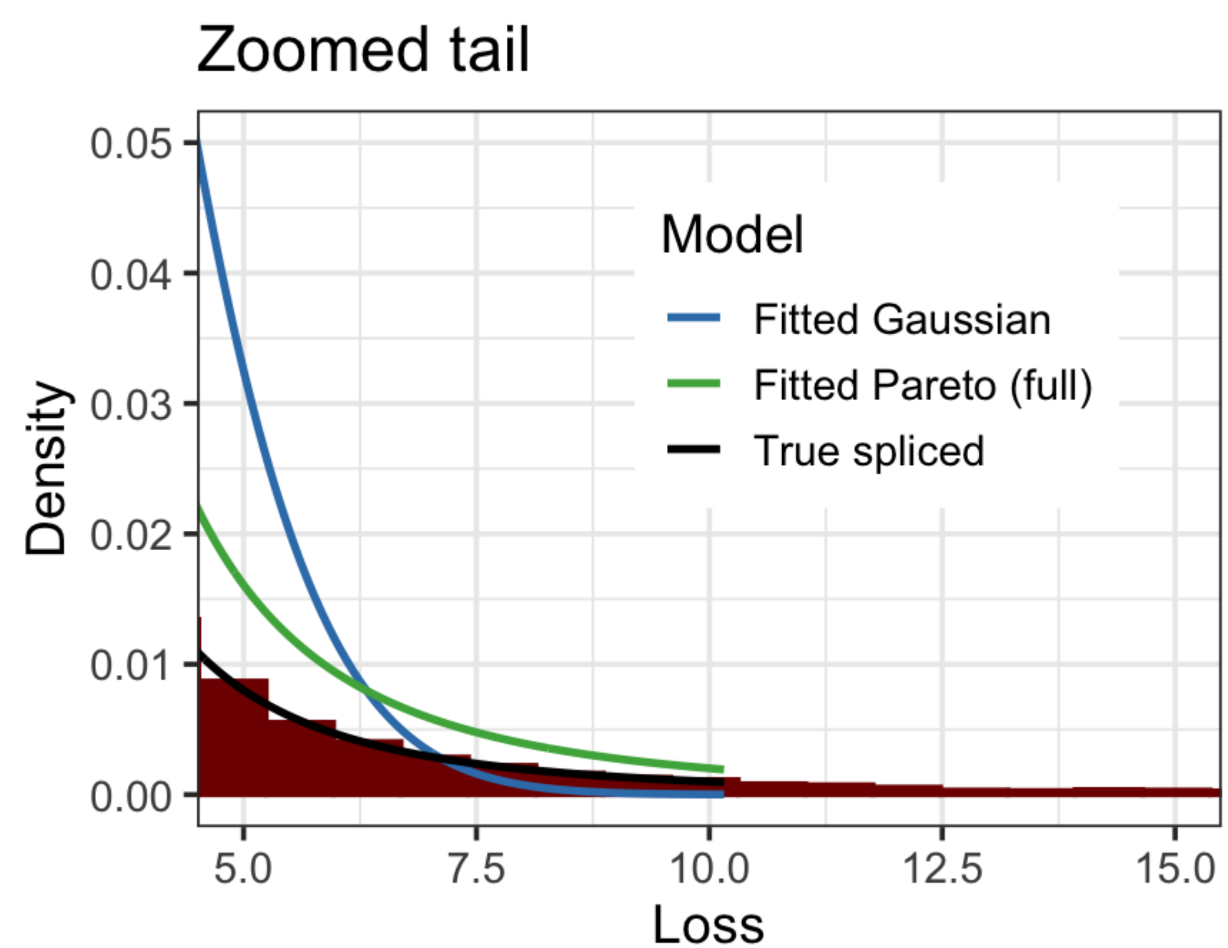
$$\cancel{P(Z \geq t) = \left(1 + \frac{t - \mu}{\sigma}\right)^{-1/\xi}}$$

Tail behaviour is parametric. Therefore, standard estimation techniques suffice to capture extre

**More realistic: There are small deviations from Pareto!**

$\xi \rightarrow$ captures heaviness of tail

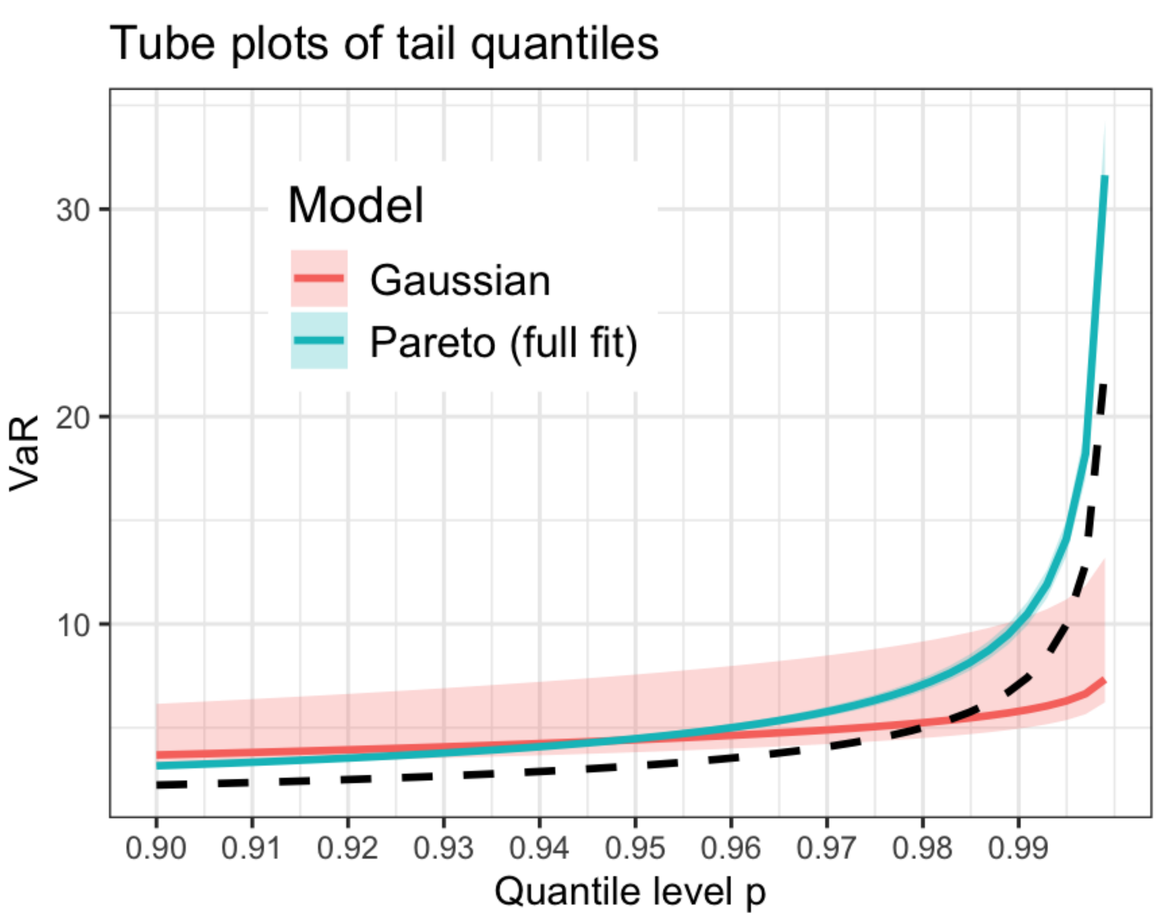


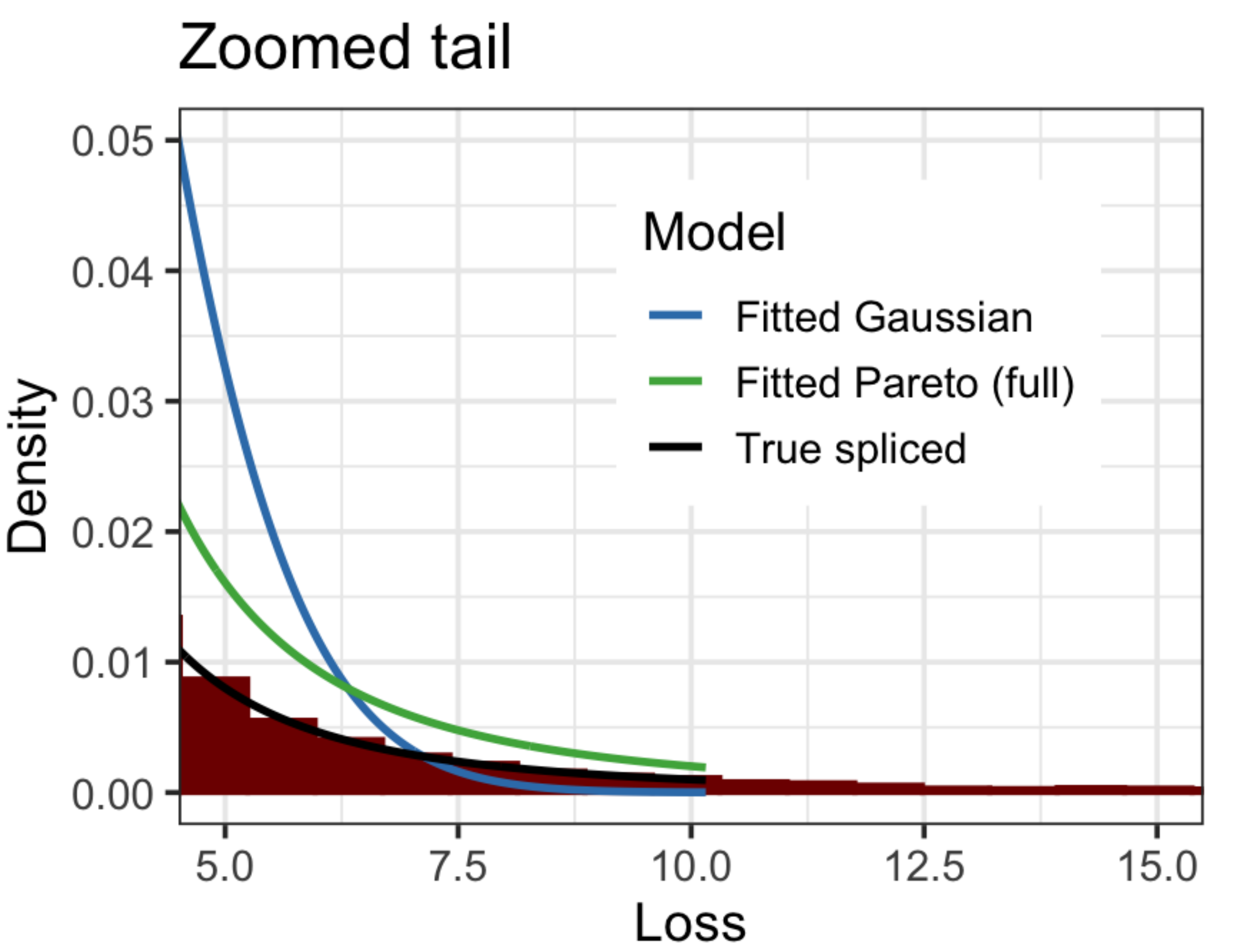



# Extreme Value Theory: A first pass

(Failure of naive approach)

- Data has a Pareto-like distribution, but only in the tail region!
- Therefore, fitting a model to the bulk of the data may lead to model error.
- Moral: The Pareto story is not all that crazy! It can serve as a basis for more nuanced models!



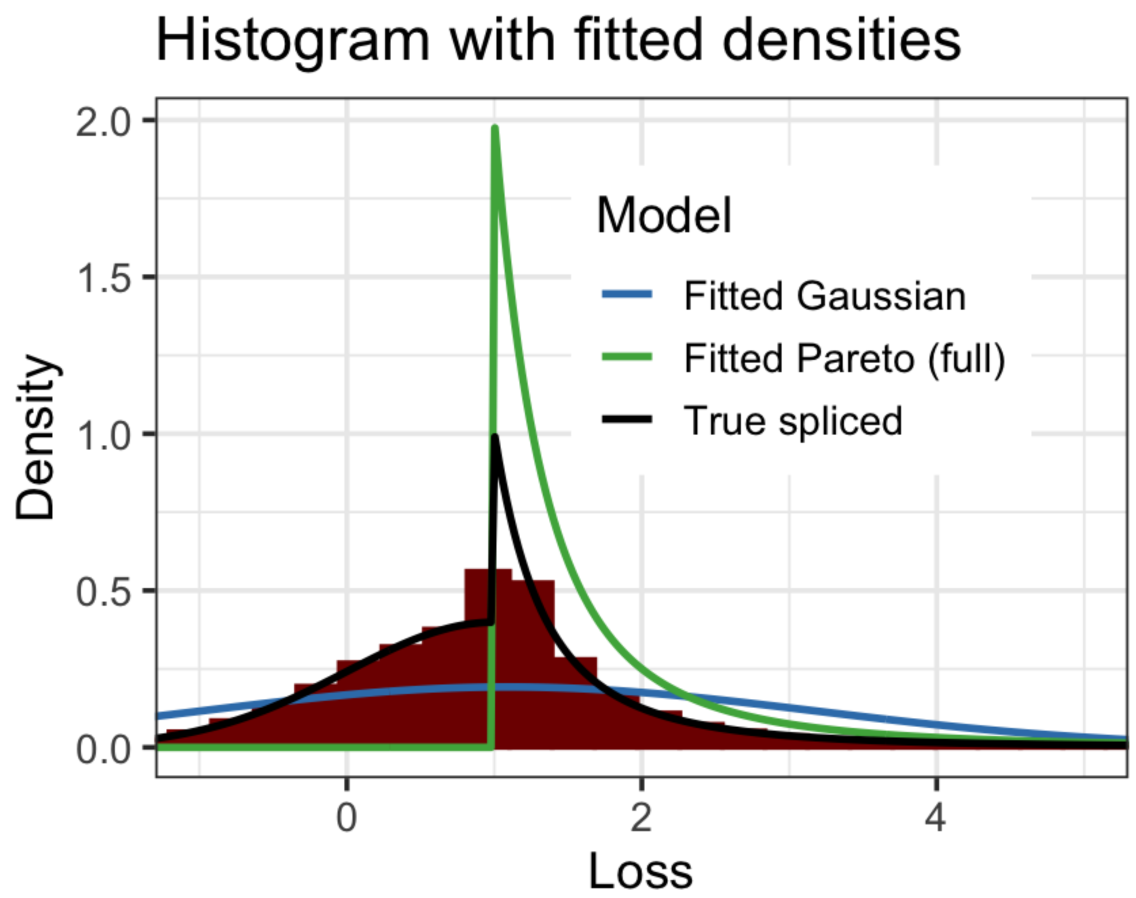
# Agenda for the course

| Session 1 | Session 2 | Session 3 |
| --- | --- | --- |
| Input Modelling 🚂 | Copulas and Model Risk | Extreme Value Theory 🚂 |
| The bootstrap | The copula | Model Risk due to tail uncertainty |
| Inverse Transforms | Model Risk | The Pareto distribution |
| Sampling from multivariate distributions | | The class RV |

# Extreme Value Theory: what all needs to be captured?

(Extending beyond the naive Pareto approach)

- Example: the density of $Z$ equals
$$f(z) = \phi(z)\mathbf{1}\ (z \le 2) + c_2 z^{-(1/\xi+1)}\ \mathbf{1}(z > 2)$$
- In the region $\{t \le 2\}$ → loss distribution is Gaussian
- In the region $\{t \ge 2\}$ → loss distribution is Pareto!




- Two "design principles" from the above example:
  - Fit a Pareto-like distribution to the tail of the data (mitigate bias)
  - Choose a tail level for which sufficient data for estimation (mitigate variance)

Extreme Value Theory formalises above principles!

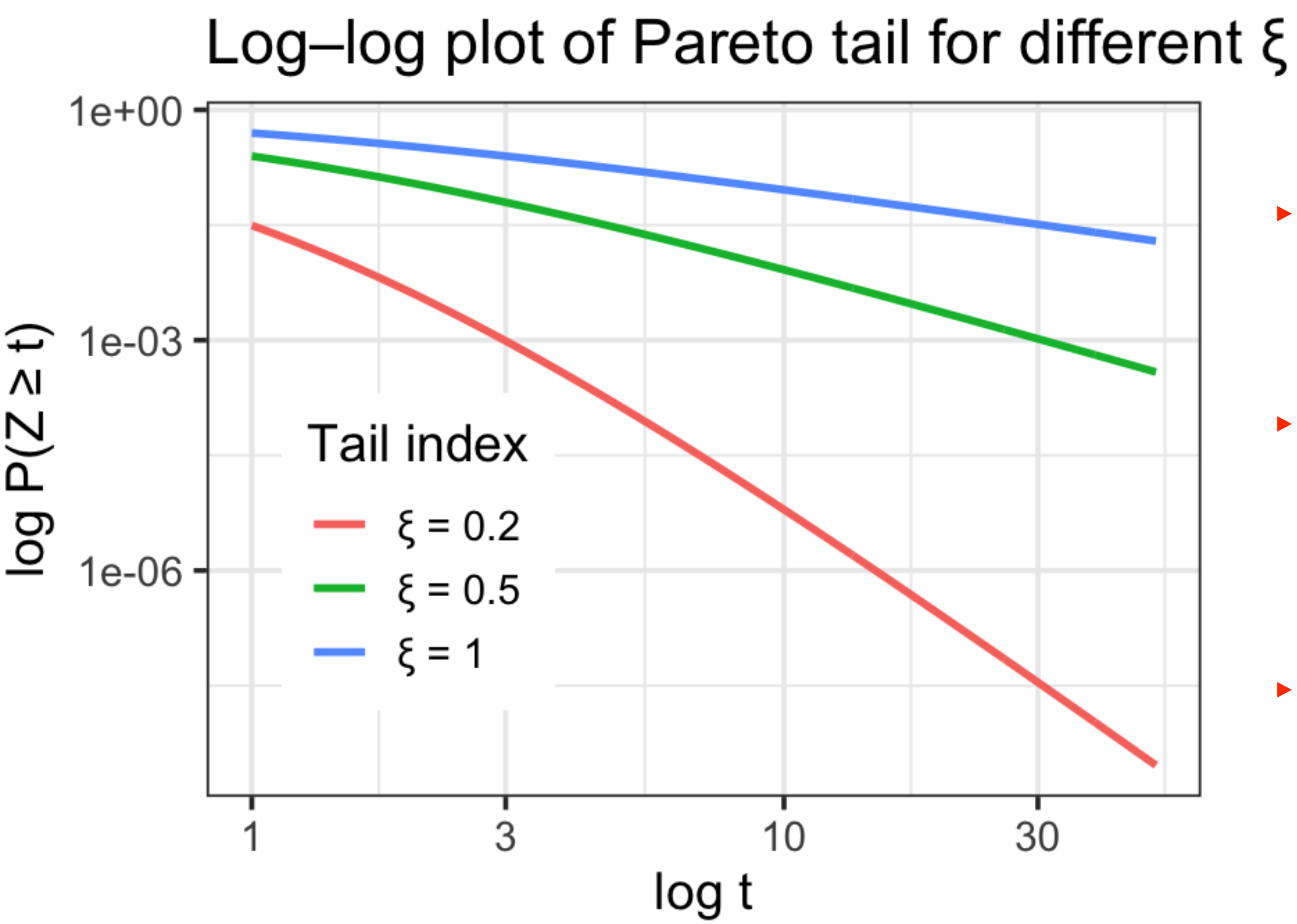
# Extreme Value Theory: the basic assumption

(Modelling "Pareto-like" distributions)

- Recall: Pareto distribution has $P(Z \geq t) = \left(1 + \frac{t-\mu}{\sigma}\right)^{-1/\xi}$.

- For a large $t$, this implies that $P(Z \geq t) \sim \sigma^{1/\xi} t^{-1/\xi}$

- Pareto distributions are "pure" polynomials: $P(Z \geq ct) \sim c^{-1/\xi} P(Z \geq t)$




- This implies $\frac{\log P(Z \geq t)}{\log t} \sim -\frac{1}{\xi}$

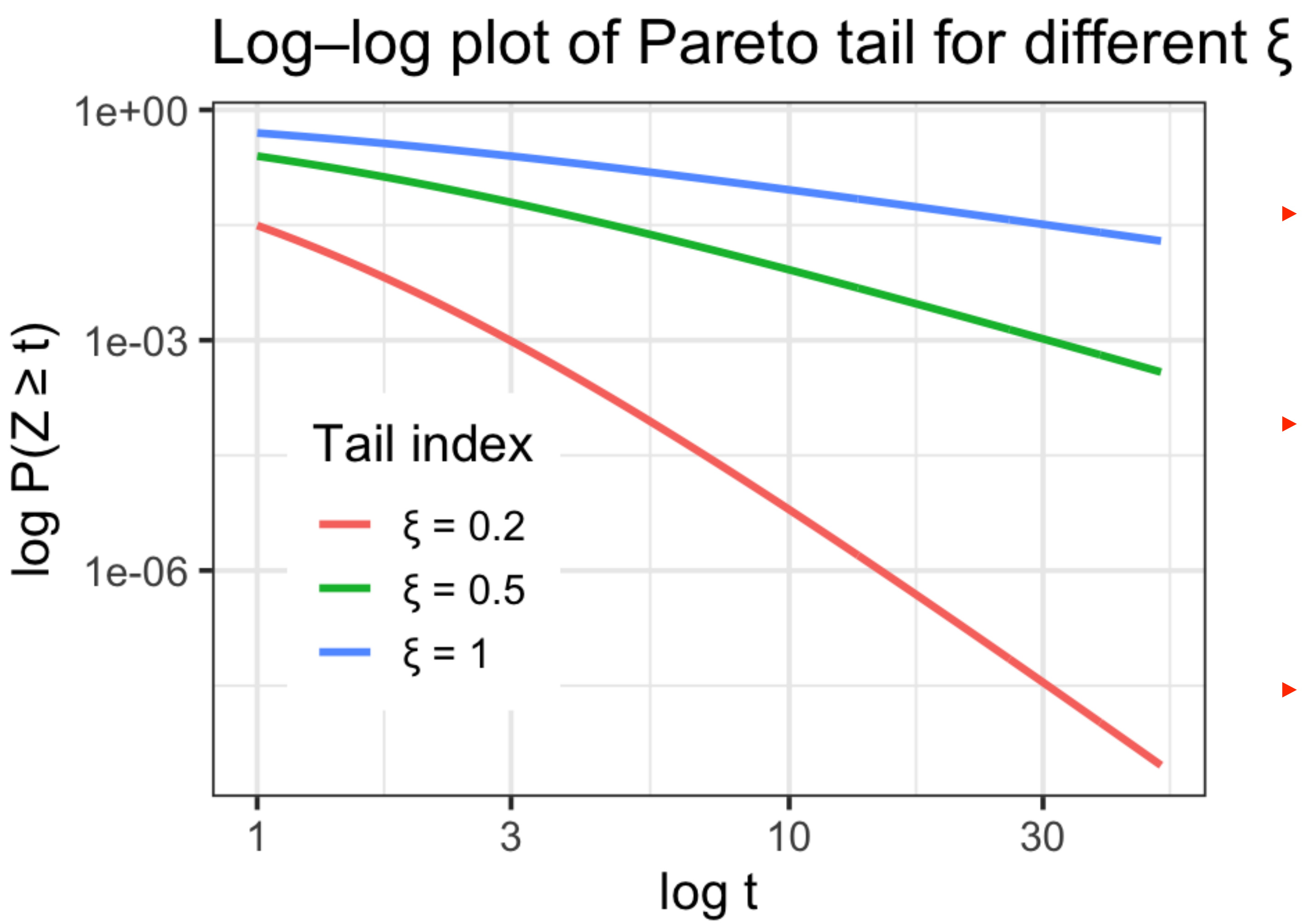
- Larger the above ratio, thicker the tail of $Z$.

- A single parameter $\xi$ controls tail strength!

# Extreme Value Theory: the basic assumption

(Modelling "Pareto-like" distributions)

More realistic: Class of models whose slope of log-probability tends to a constant in the tails!




- This implies $\dfrac{\log P(Z \geq t)}{\log t} \sim -\dfrac{1}{\xi}$
- Larger the above ratio, thicker the tail of $Z$.
- A single parameter $\xi$ controls tail strength!

# Technical preliminaries: Regular variation

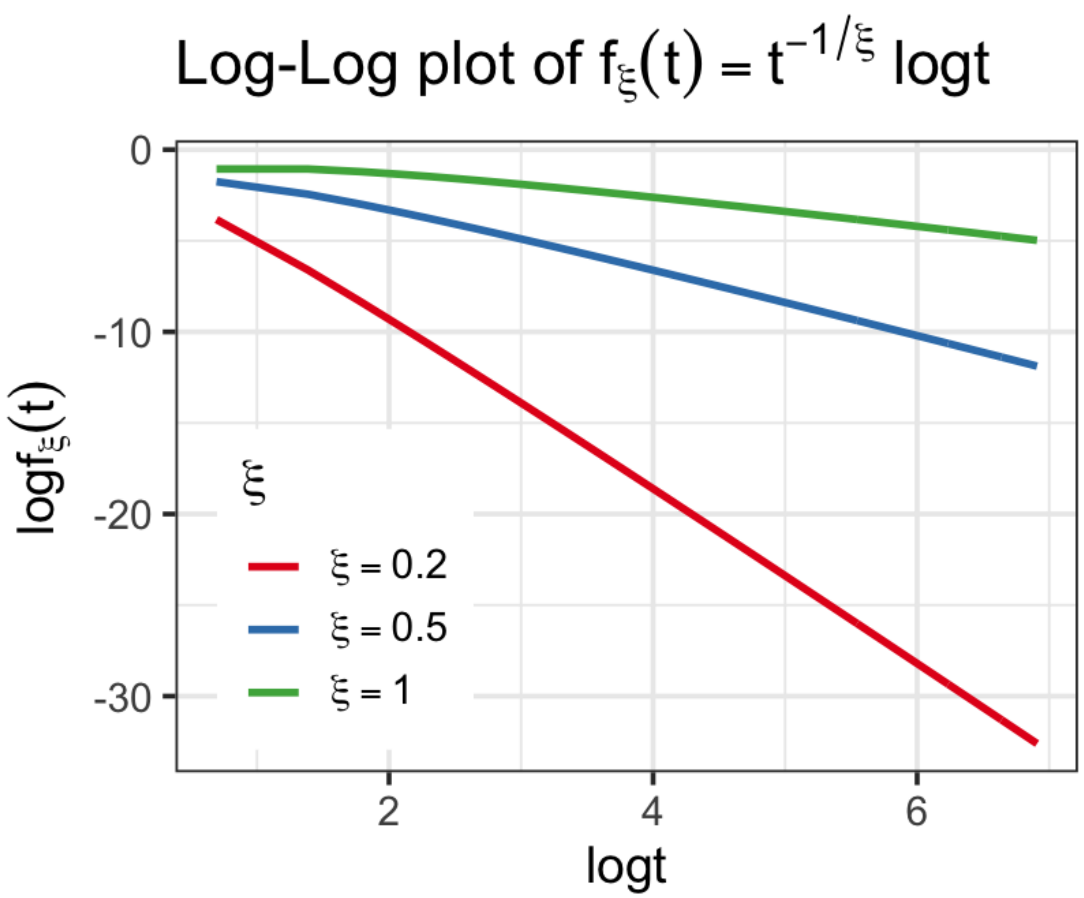
(Mathematics behind heavy tails)

Definition (Regular Variation): A function $f$ is said to be regularly varying with index $\rho$ if for any $x > 0$,

$$\lim_{t\to\infty} \frac{f(tx)}{f(t)} = x^{\rho}.$$

We denote this by $f \in \mathrm{RV}(\rho)$.

Note: we can write $f(tx) \sim x^p f(t)$ when $f \in \mathrm{RV}(p)$




- Functions that are RV are exactly the class which have linear log-log plots!
- It is such tails that we will use to generalise Pareto distributions

# Technical preliminaries: Regular variation

(Mathematics behind heavy tails)

Definition (Regular Variation): A function $f$ is said to be regularly varying with index $\rho$ if for any $x > 0$,

$$\lim_{t\to\infty} \frac{f(tx)}{f(t)} = x^{\rho}.$$

We denote this by $f \in \mathrm{RV}(\rho)$.

Note: we can write $f(tx) \sim x^{p} f(t)$ when $f \in \mathrm{RV}(p)$

Example 1a: $f(x) = x^{-2}$

- $f(x) \in \mathrm{RV}(-2)$ since $f(tx) = x^{-2} f(t)$
- Side note: All homogeneous functions are regularly varying

Example 1b: $f(x) = \log(x)$

- $f(x) \in \mathrm{RV}(0)$ since $f(tx) \sim f(t)$ for all $x$

# Technical preliminaries: Regular variation

(Mathematics behind heavy tails)

Definition (Regular Variation): A function $f$ is said to be regularly varying with index $\rho$ if for any $x > 0$,

$$\lim_{t\to\infty} \frac{f(tx)}{f(t)} = x^{\rho}.$$

We denote this by $f \in \mathrm{RV}(\rho)$.

Note: we can write $f(tx) \sim x^{p} f(t)$ when $f \in \mathrm{RV}(p)$

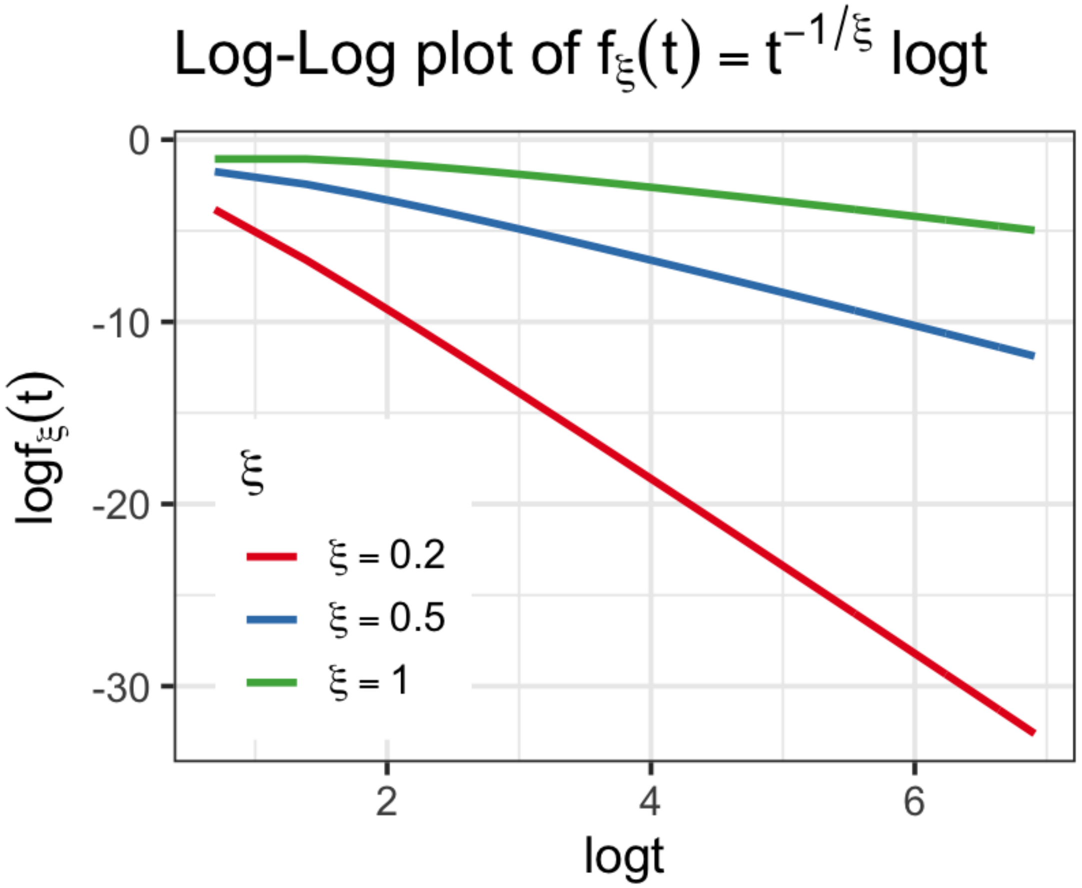
Example 2: $f(x) = x^{-2}\log(x)$

- Write $\dfrac{f(tx)}{f(t)} = \dfrac{(tx)^{-2}\log(tx)}{t^{-2}\log t} \to x^{-2}$ as $t \to \infty$
- Therefore $f \in \mathrm{RV}(-2)$
- General Property: if $f \in \mathrm{RV}(p)$ and $g \in \mathrm{RV}(q)$, then $f \times g \in \mathrm{RV}(p+q)$

# The RV class of distributions

(A precise quantification of "approximately" Pareto distributions)

Assumption (Regular Variation): There exists a $\xi > 0$ such that the tail cdf $\bar{F}(t) = P(Z > t)$ is regularly varying with index $-1/\xi$




- When $\bar{F} \in$ RV, its log-log plot is approximately a straight line (Pareto-like)
- Mathematically: $\bar{F}(t) = t^{-1/\xi}L(t)$ where $L \in \mathrm{RV}(0)$
- Results in same tail-index as Pareto

Question: Is the RV class a sufficiently general mathematical characterisation "Pareto like" tail behaviour?

# Agenda for the course

## Session 1

**Input Modelling 🚂**

- The bootstrap
- Inverse Transforms
- Sampling from multivariate distributions

## Session 2

**Copulas and Model Risk**

- The copula
- Model Risk

## Session 3

**Extreme Value Theory 🚂**

- Model Risk due to tail uncertainty
- The Pareto distribution
- The class RV
- FTG Theorem and Pareto exceedence over threshold

# Characterising Tails: Max Domain of Attraction

(Asymptotic theory for tails of distributions)

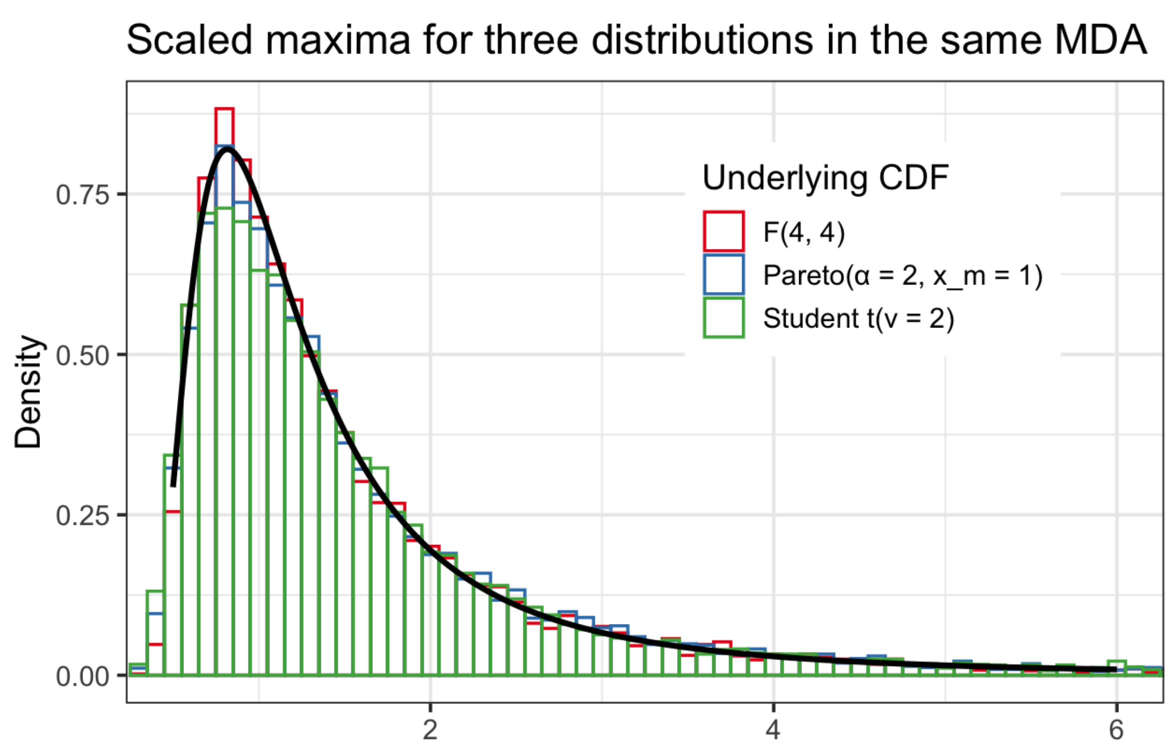
Definition: We say that a probability distribution $P$ is in the max-domain of attraction of a distribution $G$ if there exist $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$\max_{i=1}^{n} \left( \frac{X_i - b_n}{a_n} \right) \to Y \quad \text{where } X_i \sim P \text{ and } Y \sim G$$

Notation: $P \in \text{MDA}(G)$.

Equivalently: $F^n(a_n x + b_n) \to G(x)$ for all $x$ for which $G$ is continuous at $x$

- CLT $\rightarrow$ statistical behaviour of deviations of mean.
- MDA $\rightarrow$ statistical behaviour of extremes




- (Q1): Is there a classification of all possible $G$? (For CLT, this is Normal)
- (Q2): Do the parameters of $G$ depend on the tail of $P$?

# Characterising Tails: Max Domain of Attraction

(Asymptotic theory for tails of distributions)

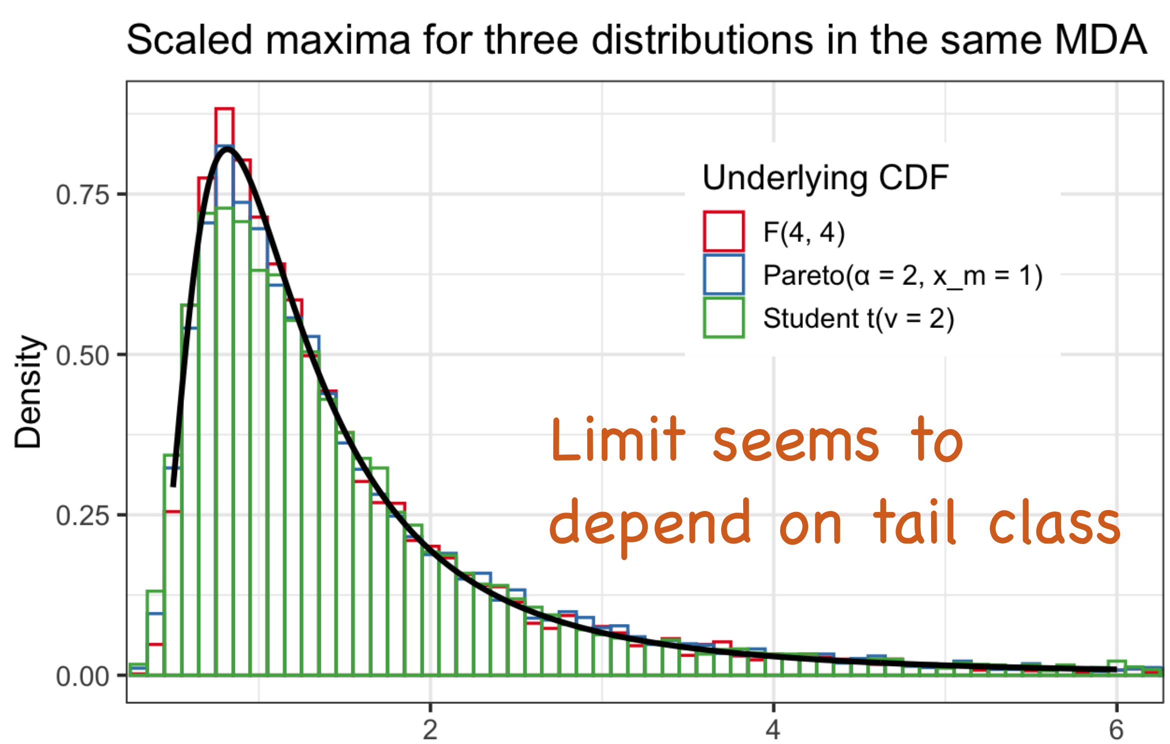
**Definition:** We say that a probability distribution $P$ is in the max-domain of attraction of a distribution $G$ if there exist $a_n > 0$ and $b_n \in \mathbb{R}$ such that

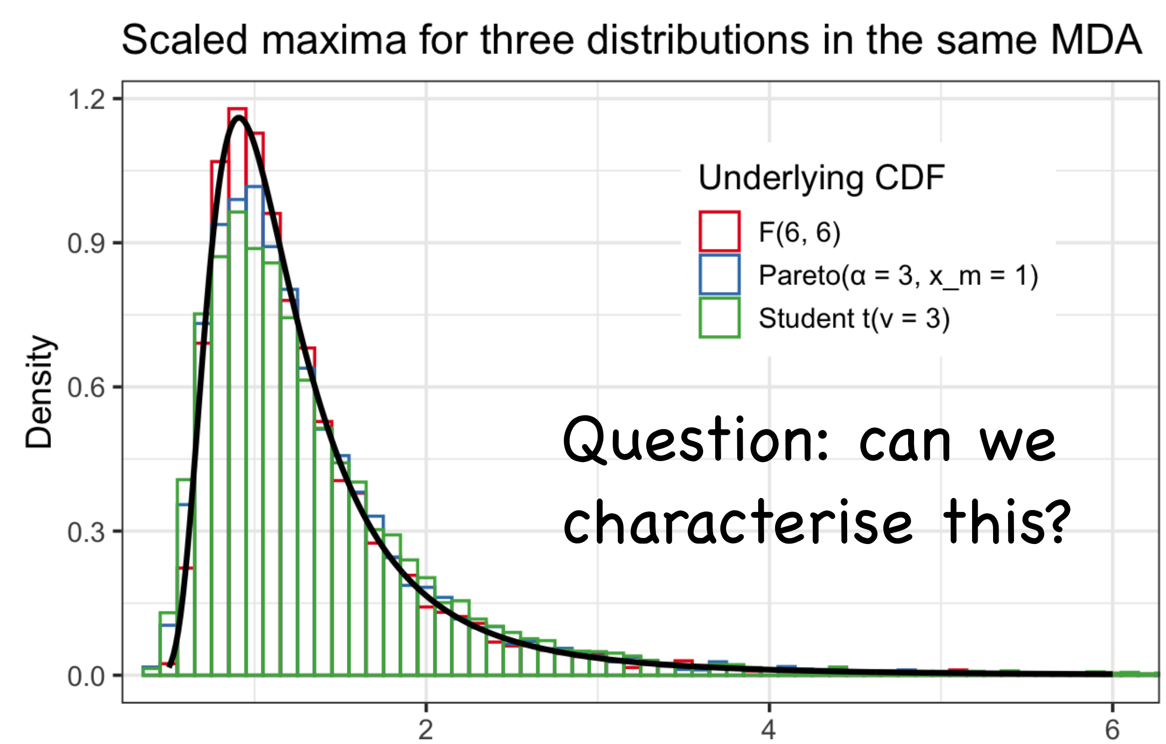
$$\max_{i=1}^{n} \left( \frac{X_i - b_n}{a_n} \right) \to Y \quad \text{where } X_i \sim P \text{ and } Y \sim G$$

**Notation:** $P \in \text{MDA}(G)$.

**Equivalently:** $F^n(a_n x + b_n) \to G(x)$ for all $x$ for which $G$ is continuous at $x$

- CLT $\to$ statistical behaviour of deviations of mean.
- MDA $\to$ statistical behaviour of extremes



# Fisher-Tippet-Gnedenko Therorem: GEV Limit class

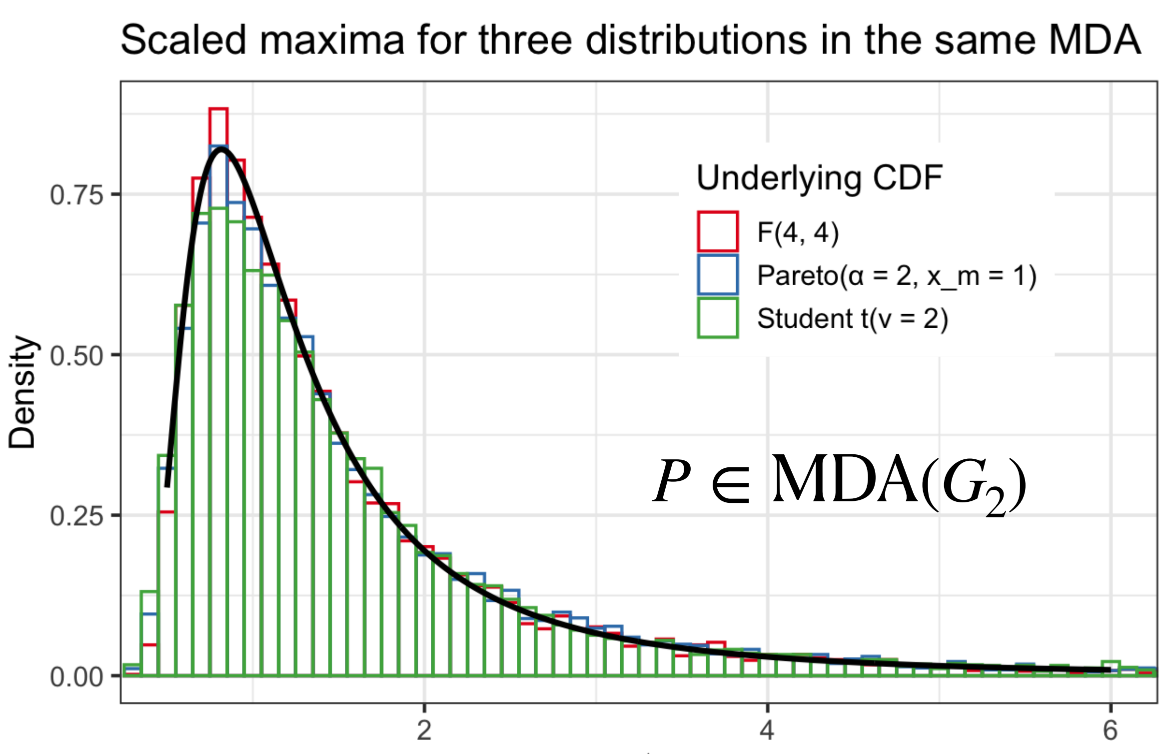
(A full characterisation of the MDA condition)

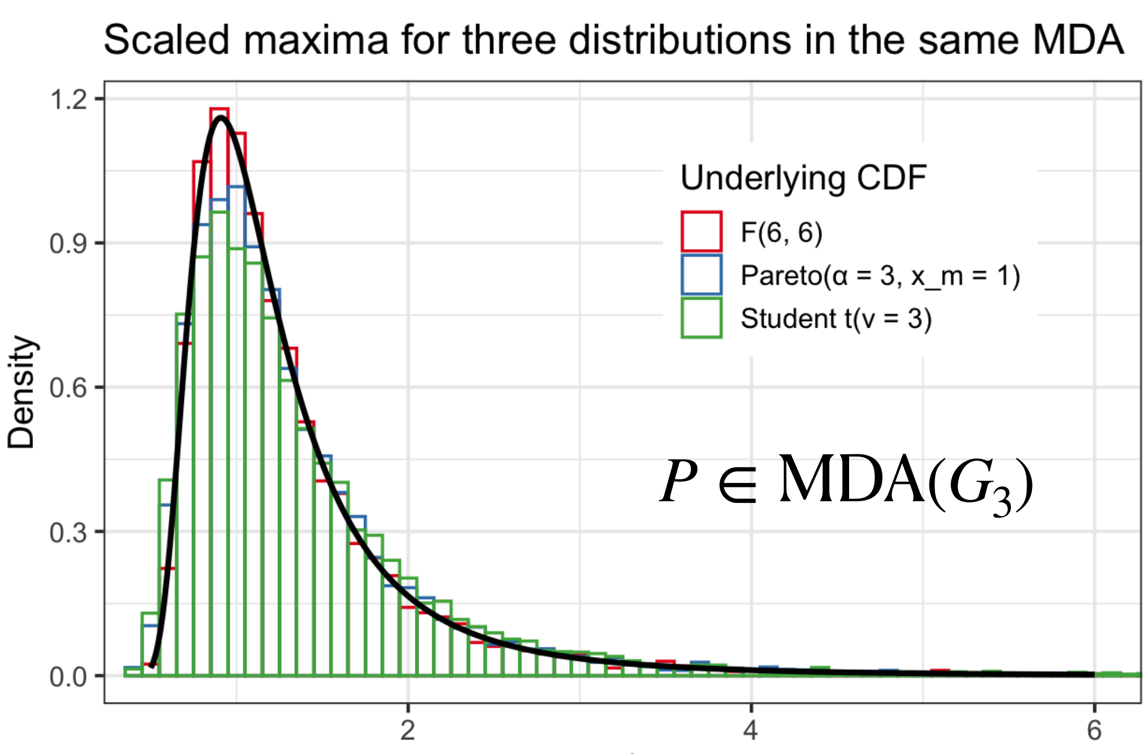
Theorem: Suppose that $P \in \text{MDA}(G)$. Then, necessarily $G$ is a generalised Extreme Value (GEV) distribution with parameter $\xi \in \mathbb{R}$:

1. $G_\xi(t) = \exp(-t^{-1/\xi})$ if $\xi > 0$ for $t \geq 0$ (Frechet DOA)

2. $G_\xi(t) = \exp(-e^{-t})$ if $\xi = 0$, for $t \in \mathbb{R}$ (Gumbel DOA) and

3. $G_\xi(t) = \begin{cases} \exp(-(-t)^{1/\xi}), & t \leq 0 \\ 1, & t > 0 \end{cases}, \quad \xi < 0$



# Fisher-Tippet-Gnedenko Therorem: GEV Limit class

(A full characterisation of the MDA condition)

**Theorem**: Suppose that $P \in \mathrm{MDA}(G)$. Then, necessarily $G$ is a generalised Extreme Value (GEV) distribution with parameter $\xi \in \mathbb{R}$:

1. $G_\xi(t) = \exp(-t^{-1/\xi})$ if $\xi > 0$ for $t \geq 0$ (Frechet DOA)

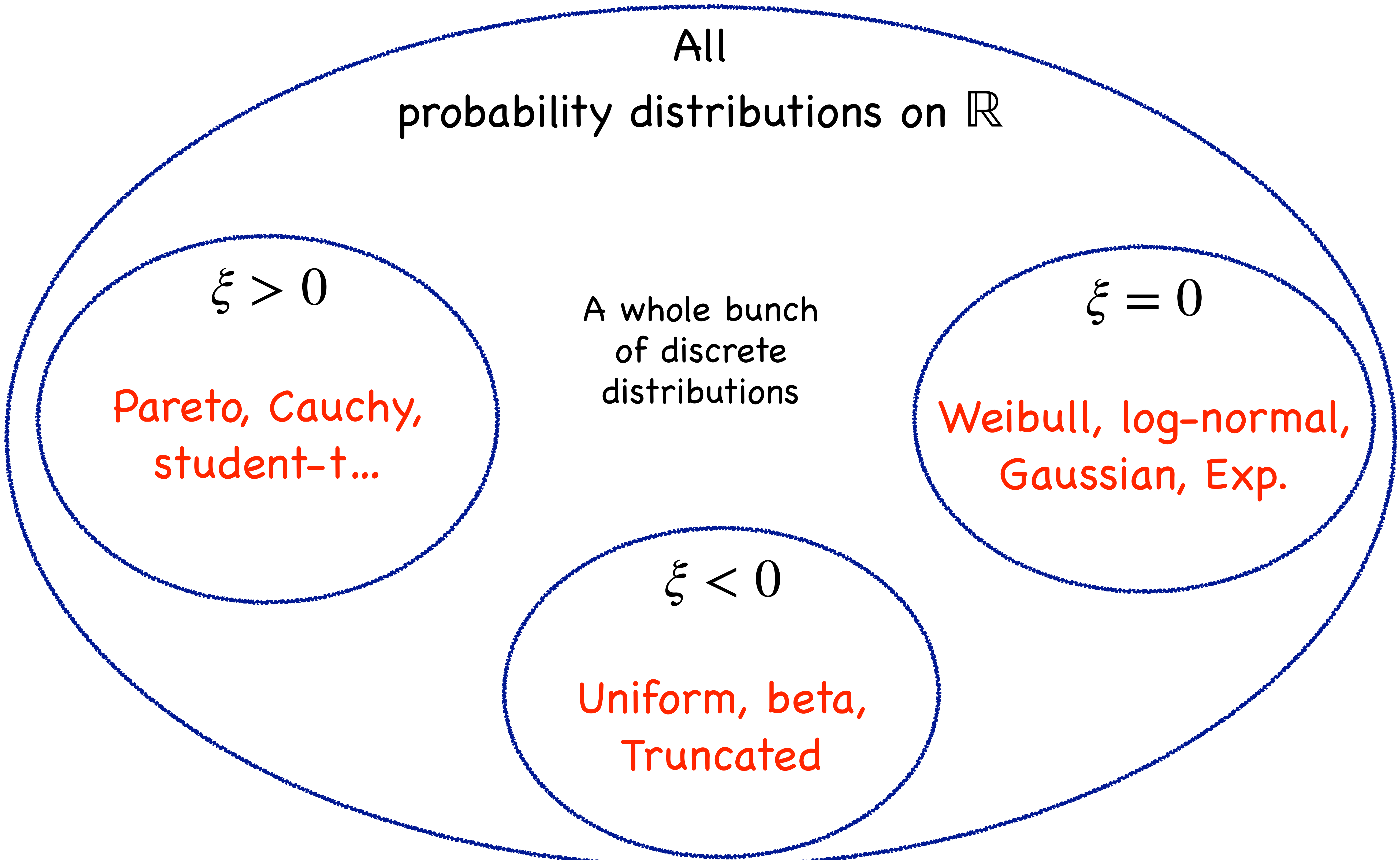
2. $G_\xi(t) = \exp(-e^{-t})$ if $\xi = 0$, for $t \in \mathbb{R}$ (Gumbel DOA) and

3. $G_\xi(t) = \begin{cases} \exp(-(-t)^{1/\xi}), & t \leq 0 \\ 1, & t > 0 \end{cases}, \quad \xi < 0$

- Implication: the limiting statistical behaviour of maxima is parametric!
- Characterises "approximately" Pareto behaviour with $\xi \to$ Pareto index
- $\xi > 0 \to$ heavy tails (Pareto, Cauchy, F distributions)
- $\xi = 0 \to$ light tails (Gaussian, Lognormal, Gamma distributions)
- $\xi < 0 \to$ bounded distributions (Beta, Uniform, truncated distribution)

# Fisher-Tippet-Gnedenko Therorem: GEV Limit class

(A full characterisation of the MDA condition)



# End of Lecture 3

# Appendix: Technical Material on Regular variation

(Mathematics behind heavy tails)

Definition (Regular Variation): A function $f$ is said to be regularly varying with index $\rho$ if for any $x > 0$,

$$\lim_{t \to \infty} \frac{f(tx)}{f(t)} = x^{\rho}.$$

We denote this by $f \in \mathrm{RV}(\rho)$.

Note: we can write $f(tx) \sim x^p f(t)$ when $f \in \mathrm{RV}(p)$

Important Property: If a monotone function $f \in \mathrm{RV}(p)$, then $f^{\leftarrow} \in \mathrm{RV}(1/p)$

Proof Idea:

- Let $g = f^{\leftarrow}$. Then $g(tx) = f^{\leftarrow}\left(f(y)x\right)$ where $f(y) = t$.
- Then, $f(y)x \sim f(yx^{1/p})$ since $f \in \mathrm{RV}(p)$
- Now, $g(tx) \sim x^{1/p}y$. But $f(y) = t \implies y = g(t)$.
- Hence $g(tx) \sim x^{1/p}g(t) \implies g \in \mathrm{RV}(1/p)$.