ACM Winter School 2025

Lecture 4: Resolving Model Risk - an EVT Approach

© Anand Deo (Indian Institute of Management, Bangalore)

Characterising Tails: Max Domain of Attraction

(Asymptotic theory for tails of distributions)

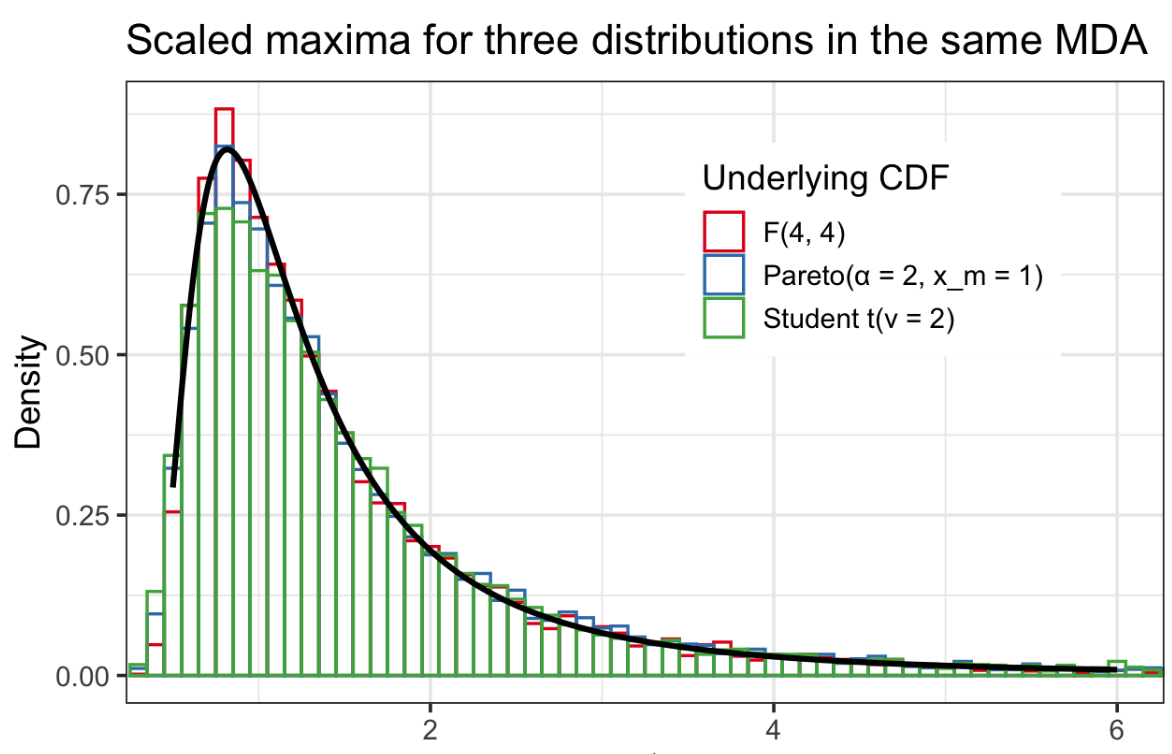
Definition: We say that a probability distribution P is in the max-domain of attraction of a distribution G if there exist $a_n > 0$ and $b_n \in \mathbb{R}$ such that

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

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 $\longrightarrow$ statistical behaviour of deviations of mean.
- MDA $\longrightarrow$ 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?

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

- 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)

Connecting the Dots: FTG Theorem, RV class, Pareto tails

(Why Pareto was a good model to begin with)

- FTG Theorem → Characterises behaviour of maxima
- RV survival function → characterises tail behaviour

Question: Is there an equivalence between these two?

Toy Example: Suppose $\bar{F}(z) = z^{-1/\xi}$ for $z \geq 1$

- Write $F^n(a_n x + b_n) = \left(1 - \frac{1}{(a_n x + b_n)^{1/\xi}}\right)^n$
- Set $a_n = n^{1/\xi}$ and $b_n = 0$ above and take limits as $n \to \infty$:

$$\lim_{n\to\infty} F^n(n^{1/\xi} x) = \left(1 - n^{-1} x^{-1/\xi}\right)^n = \exp(-x^{-1/\xi})$$

- This implies that $P \in \text{MDA}(G_\xi)$

Connecting the Dots: FTG Theorem, RV class, Pareto tails

(Why Pareto was a good model to begin with)

It appears that if $\bar{F} \in \text{RV}(-1/\xi)$, then $P \in \text{MDA}(G_\xi)$

Theorem: If $\xi > 0$, then $\bar{F} \in \text{RV}(-1/\xi) \iff P \in \text{MDA}(G_\xi)$

(Implication: All distributions with regular tails are "Pareto-like")

Toy Example: Suppose $\bar{F}(z) = z^{-1/\xi}$ for $z \geq 1$

- Write $F^n(a_n x + b_n) = \left(1 - \dfrac{1}{(a_n x + b_n)^{1/\xi}}\right)^n$
- Set $a_n = n^{1/\xi}$ and $b_n = 0$ above and take limits as $n \to \infty$:

$$\lim_{n\to\infty} F^n(n^{1/\xi}x) = \left(1 - n^{-1}x^{-1/\xi}\right)^n = \exp(-x^{-1/\xi})$$

- This implies that $P \in \text{MDA}(G_\xi)$

Connecting the Dots: FTG Theorem, RV class, Pareto tails

(Why Pareto was a good model to begin with)

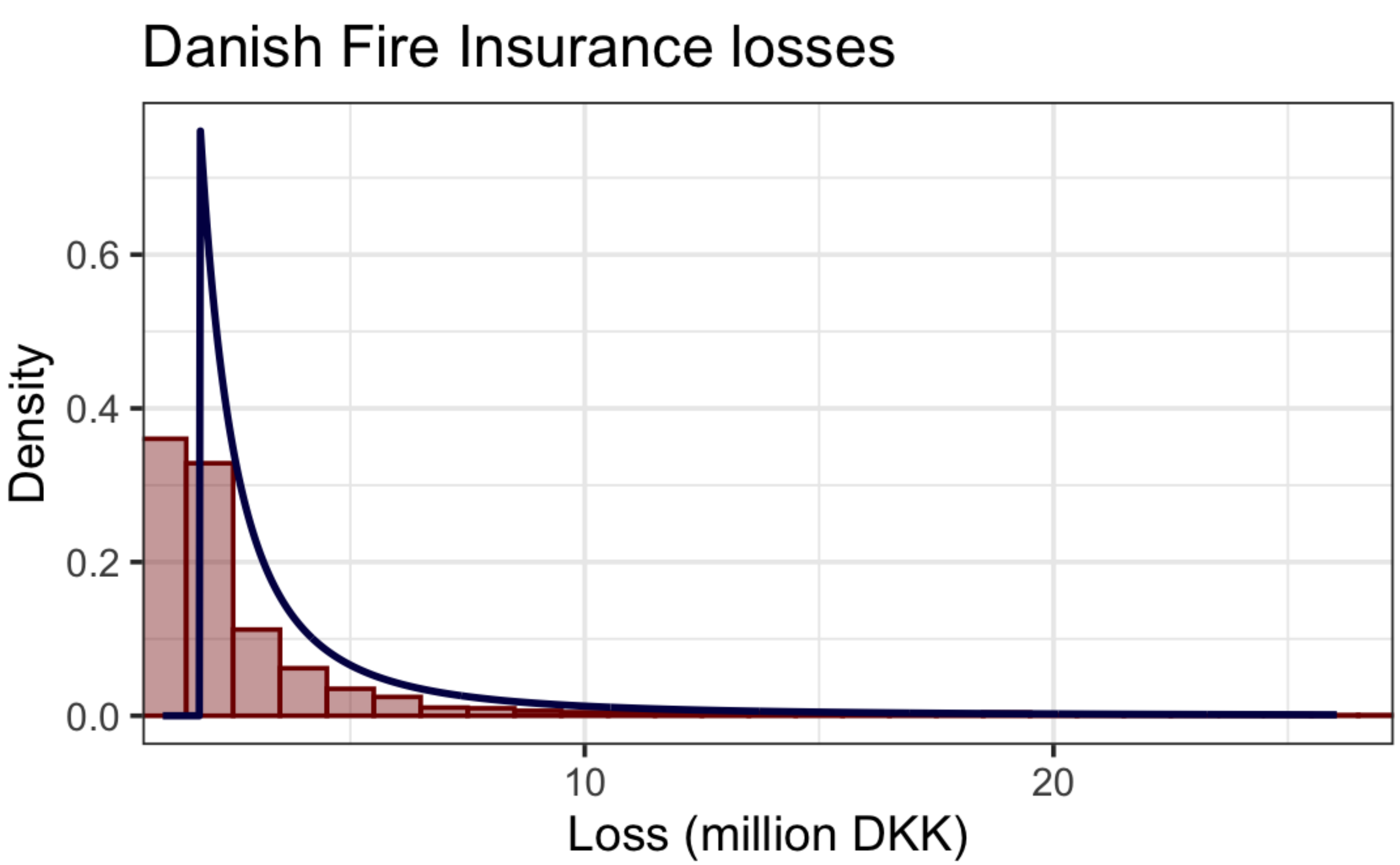
Observations so far:

i) FTG Theorem: Characterises distributions with "regular tails"

ii) The RV class: Characterises "Pareto like" tails

iii) Pareto exceedances: Tail behaviour of RV class

Taken together: Any distribution with regular tails has Pareto exceedances!



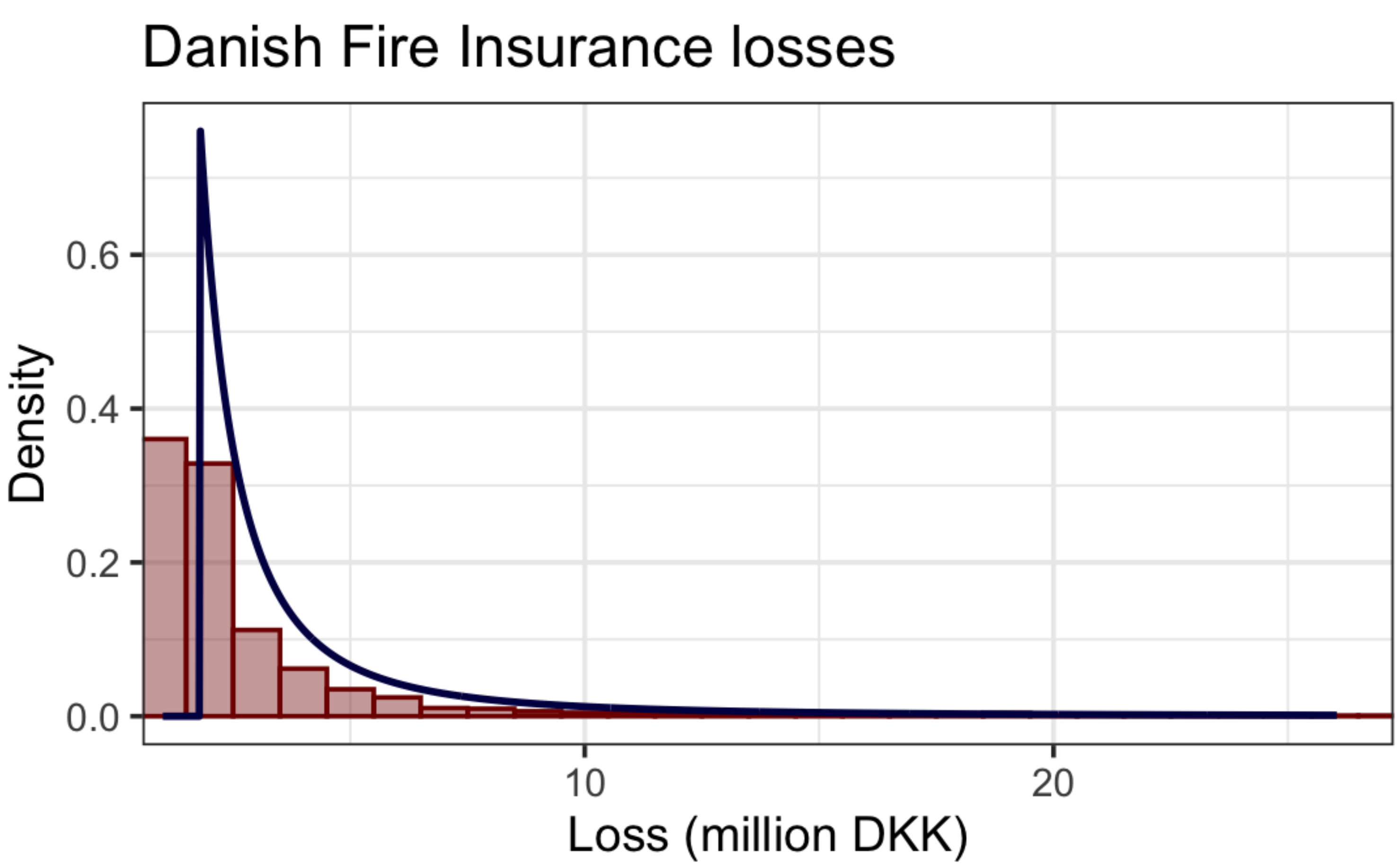
- Dataset: Fire insurance claims in Denmark
- Data seems to have regular tail behaviour
- Pareto exceedances fit well to data

Connecting the Dots: FTG Theorem, RV class, Pareto tails

(Why Pareto was a good model to begin with)

Combating Model Risk

- Pareto Exceedences are a good working model for extreme behaviour
- Therefore to mitigate model risk for tail risk
  - Fix a threshold beyond which there is sufficient data (lowers stat. Error)
  - Calibrate a Pareto model to data beyond this threshold (lowers bias)



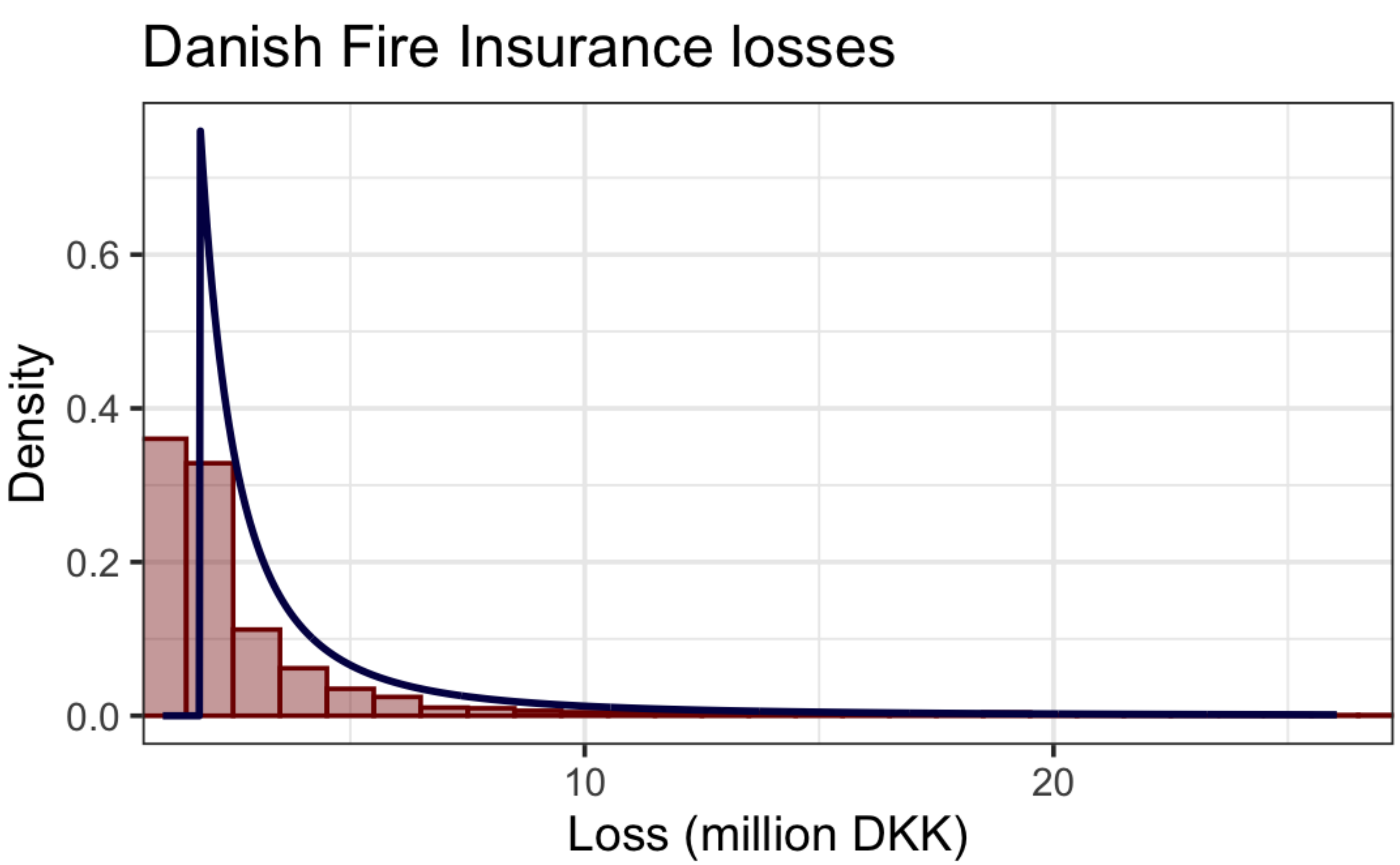
- Dataset: Fire insurance claims in Denmark
- Data seems to have regular tail behaviour
- Pareto exceedances fit well to data

Connecting the Dots: FTG Theorem, RV class, Pareto tails

(Why Pareto was a good model to begin with)

Combating Model Risk

- Pareto Exceedences are a good working model for extreme behaviour
- Therefore to mitigate model risk for tail risk
  - Fix a threshold beyond which there is sufficient data (lowers stat. Error)
  - Calibrate a Pareto model to data beyond this threshold (lowers bias)



Next up: Combating tail-fragility: Using EVT to compute tail risk measures

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
- Tail Risk Measures

Session 3

Extreme Value Theory

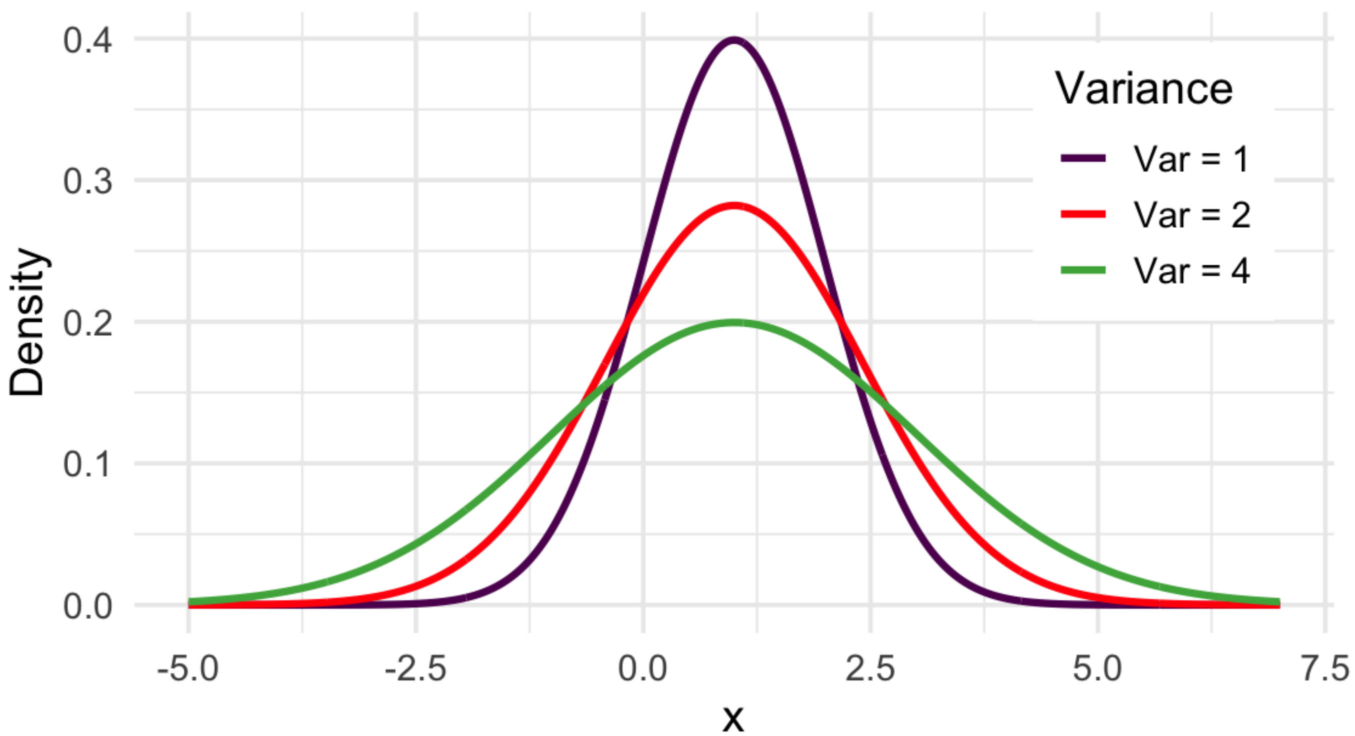
- Model Risk due to tail uncertainty
- The Pareto distribution
- The class RV
- FTG Theorem and Pareto exceedence over threshold

Session 4

Risk Measures 🚂

Variance as a measure of risk

(A first attempt to capture risk)



- Observation: Higher variance implies higher potential for large losses
- Variance →often used as a proxy for downside risk!
- Regulators favours designs with low variance

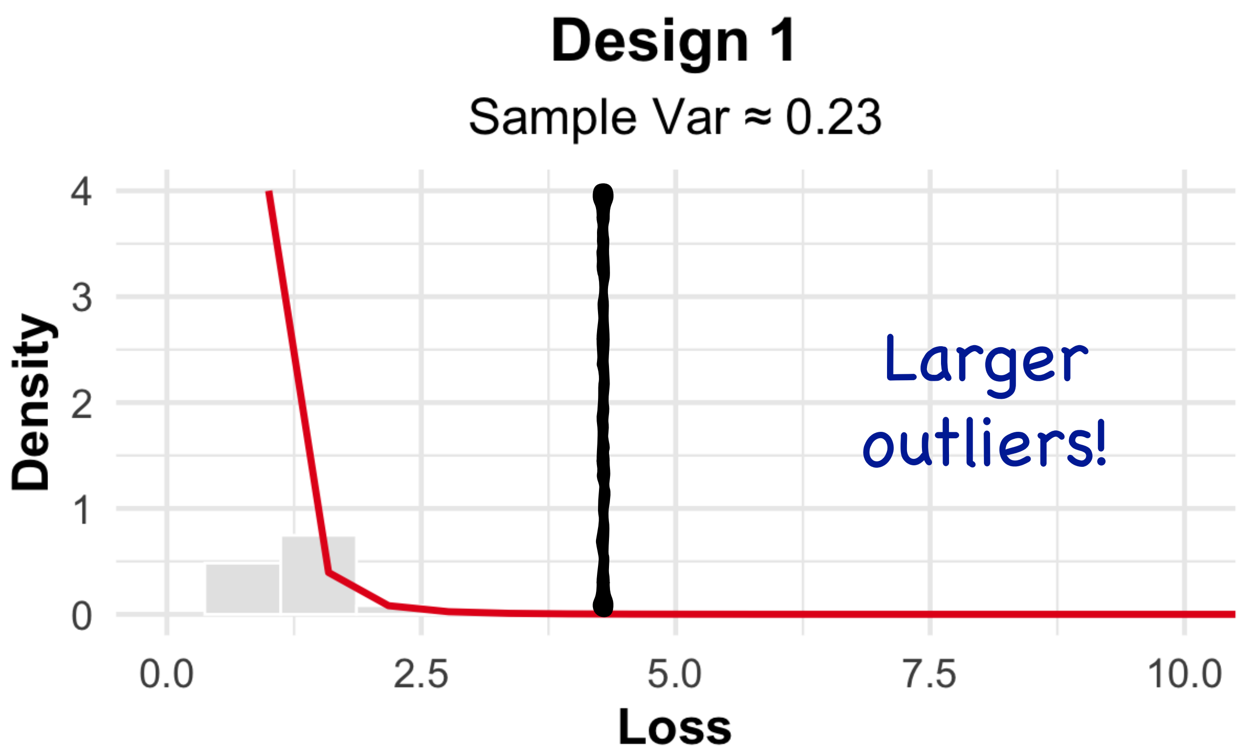
In class example: A regulator chooses to model losses due to power outages. Their in-house engineers propose 2 designs, one with variance 0.23cr and the other with variance 1cr (mean loss for both designs is the same).

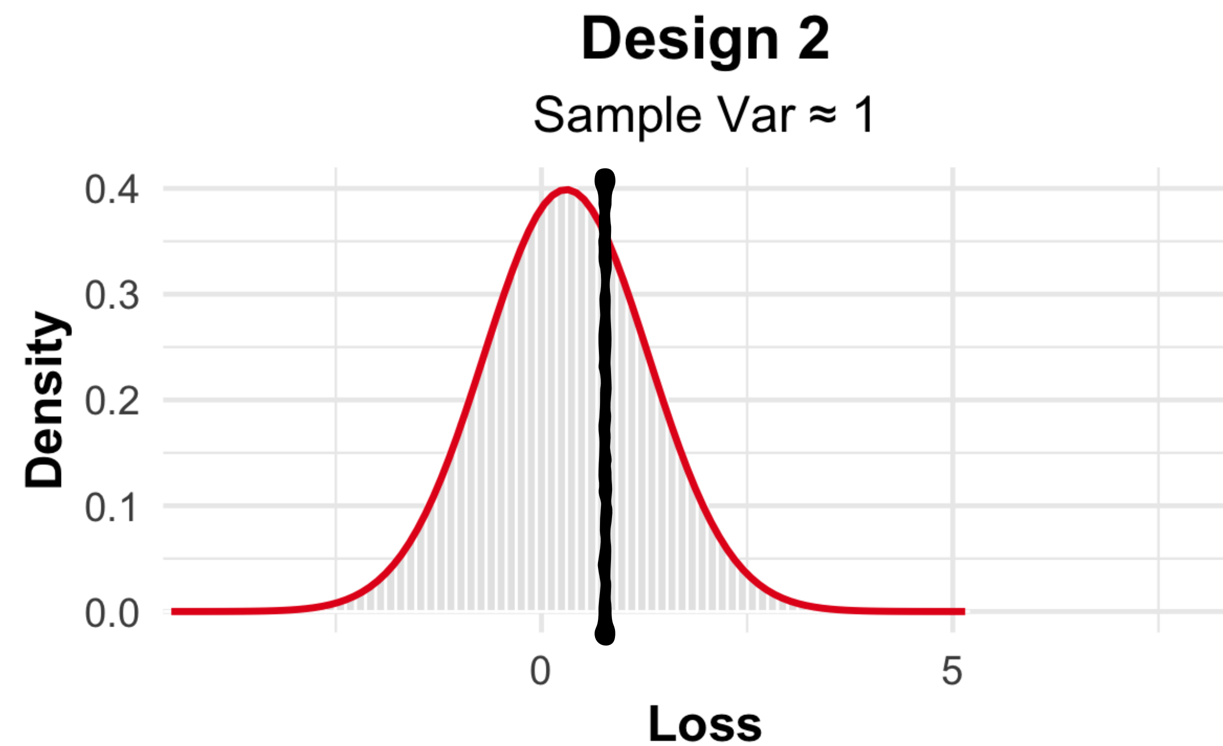
As per their current modelling practice which design will the regulator accept, and why?

Answer: They will accept design 1 (smaller variance)

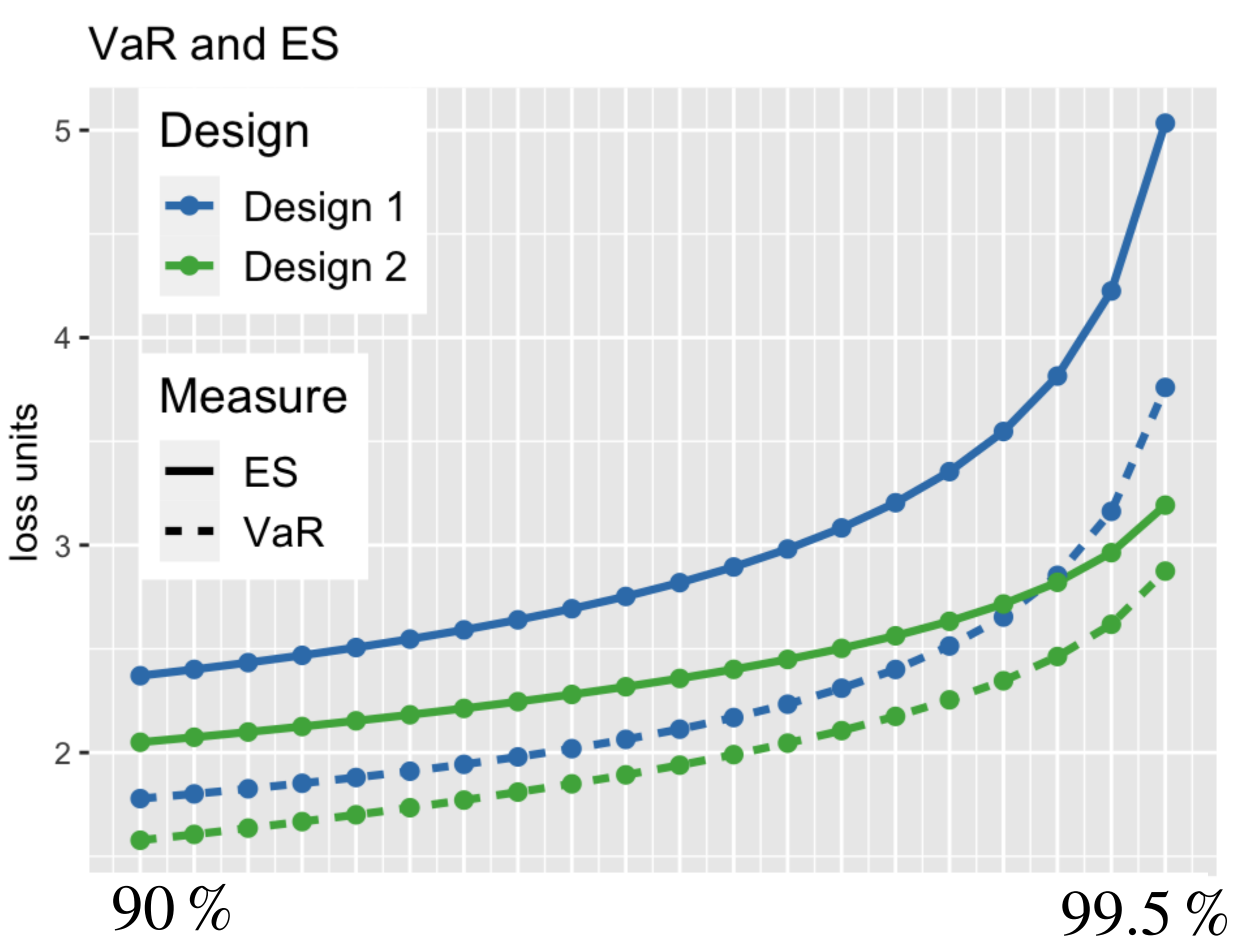
Variance as a measure of risk: Pitfalls

(A first attempt to capture risk)





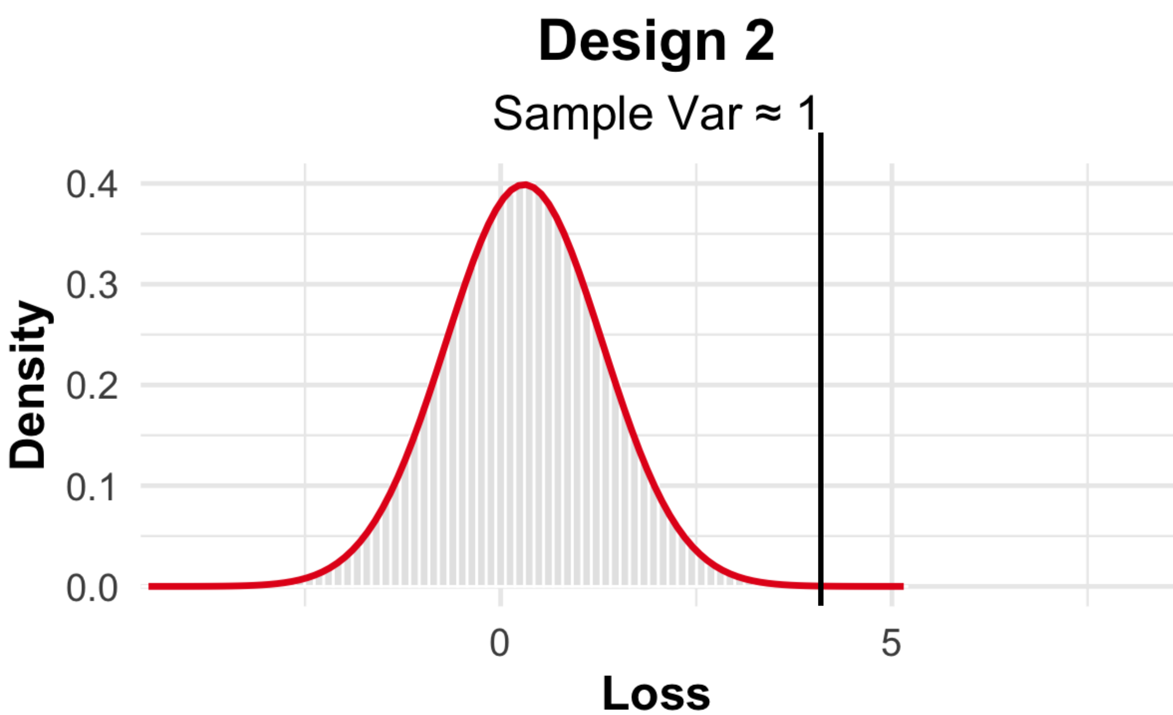
- Smaller variance design faces a greater risk of outliers!



- Regulators care about extreme losses
- $\text{Var}(X) = E[X - \mu]^2$ —> deviation about the mean
- Quantiles/CVaR computed over tail —> more representative of the risk the regulators wants to asses!

Measures of Tail Risk: VaR and CVaR

(Going beyond variance)



- Value at risk of a distribution P at a risk level $(1-\alpha)$ equals its $(1-\alpha)$th quantile.
- Mathematically: $\text{VaR}_{1-\alpha}(X) = Q_X(1-\alpha)$

Question 1: Does VaR capture outliers?

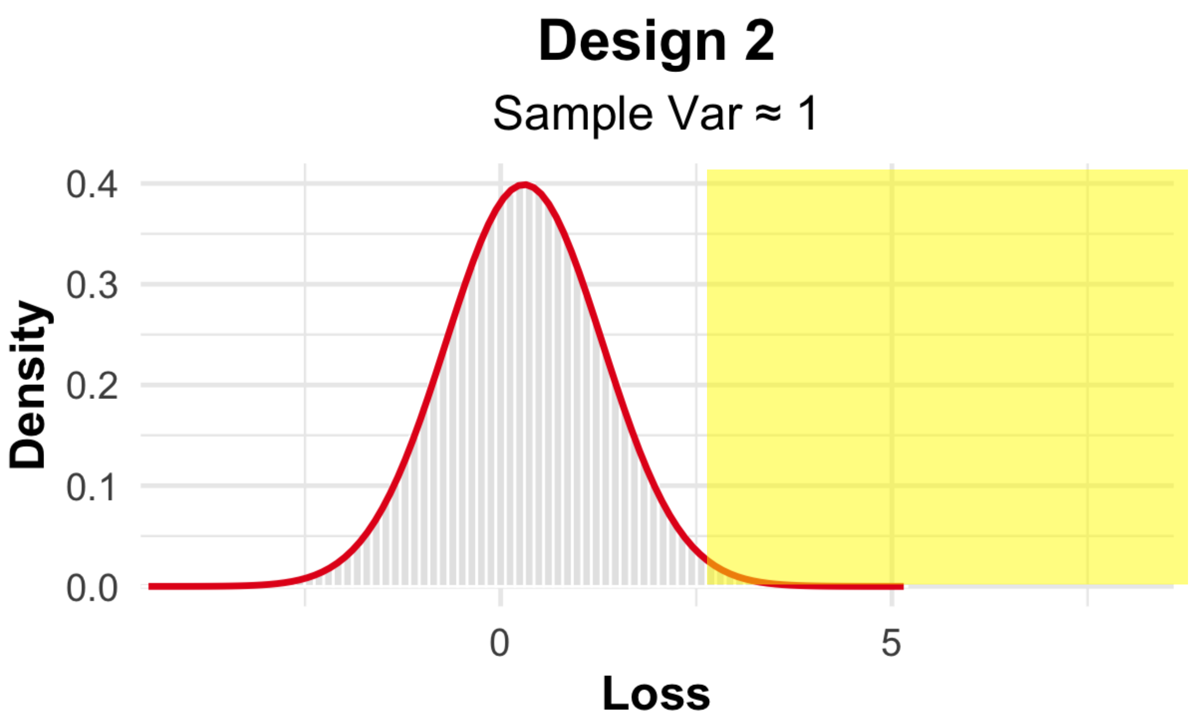
- Yes. VaR is the level that is exceeded in about α fraction of days
- In terms of regulatory capital: it is that capital required to survive all but the largest α fraction of losses

Question 2: Is VaR always adequate

- No. VaR tells you where the tail starts, not how bad it gets beyond that

Measures of Tail Risk: VaR and CVaR

(Going beyond variance)



- Conditional value at risk of a distribution P at a risk level $(1-\alpha)$ equals the average of the largest α fraction of loss realisations from P
- Mathematically: $\text{CVaR}_{1-\alpha}(X) = E[X \mid X > v_{1-\alpha}(X)]$

Question 1: Does CVaR capture outliers?

- Yes, it is the conditional expectation over large losses

Question 2: Is CVaR adequate?

- Yes, since it also incorporates the extent of large losses beyond the quantile
- Has better properties from a risk measure standpoint (see appendix)

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
- Tail Risk Measures
- Curse of Rarity

Session 3

Extreme Value Theory

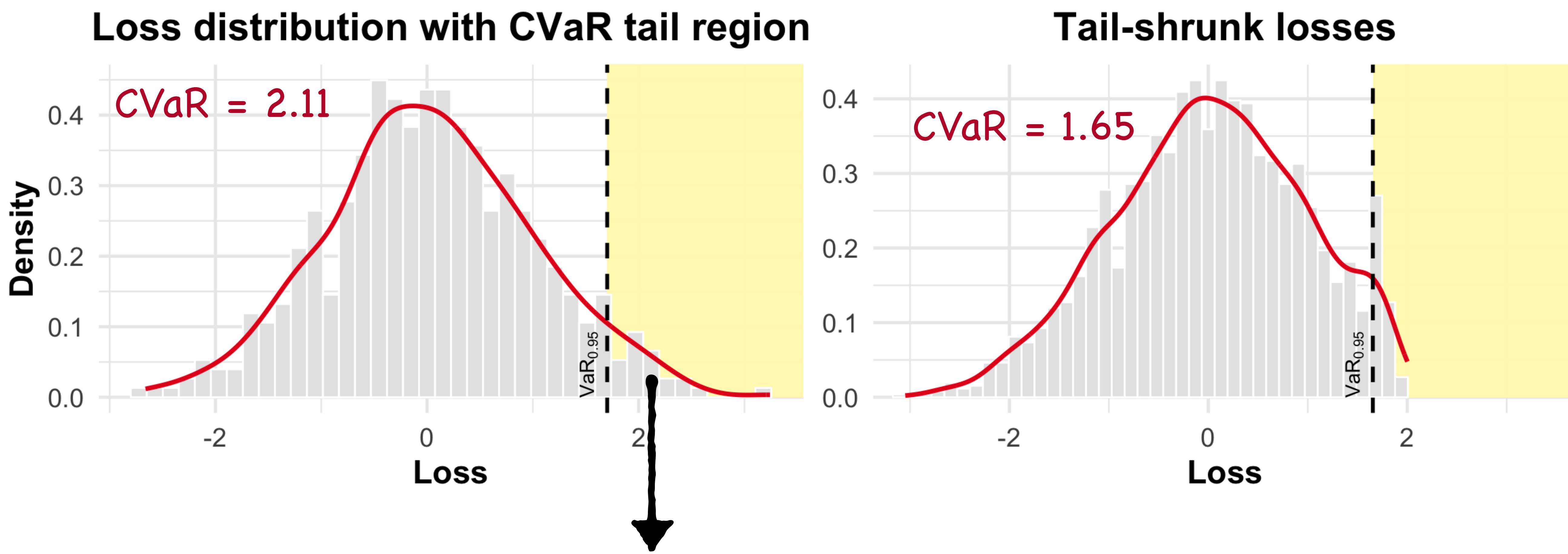
- Model Risk due to tail uncertainty
- The Pareto distribution
- The class RV
- FTG Theorem and Pareto exceedence over threshold

Session 4

Risk Measures 🚂

Fragility of Tail Risk Measures to Tails

(How tail samples affect risk evaluations)

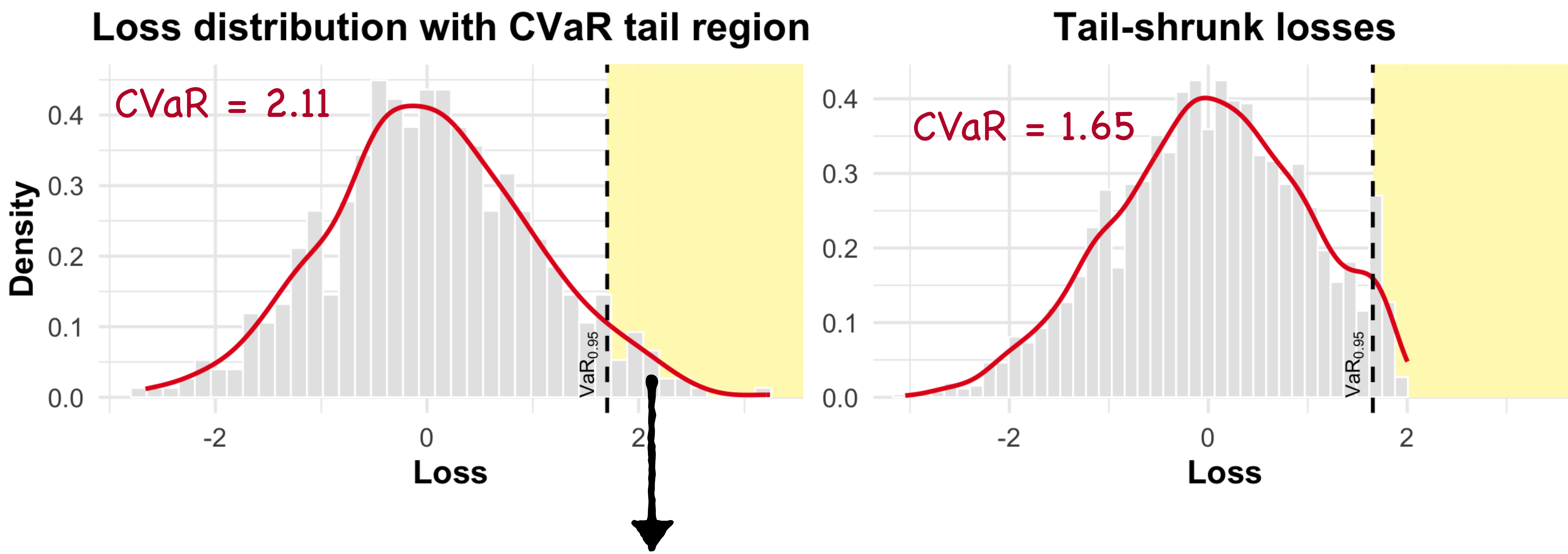


- Lack of samples in the tail —> sensitivity to actual data realisation
- Leads to model risk: if the data realisations are not representative of tails —> underestimation of CVaR

Small tweaks to tail observations —> huge change in CVaR —> CVaR is highly sensitive to model errors!

Fragility of Tail Risk Measures to Tails

(Recourse: Use EVT)



- Lack of samples in the tail —> sensitivity to actual data realisation
- Leads to model risk: if the data realisations are not representative of tails —> underestimation of CVaR
- Recourse: we know that if the tails are "regular", they are eventually Pareto-like. Why not utilise this observation?

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
- Tail Risk Measures
- Curse of Rarity
- EVT to the rescue

Session 3

Extreme Value Theory

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

Session 4

Risk Measures

Using EVT for accurate assessment of tail risk

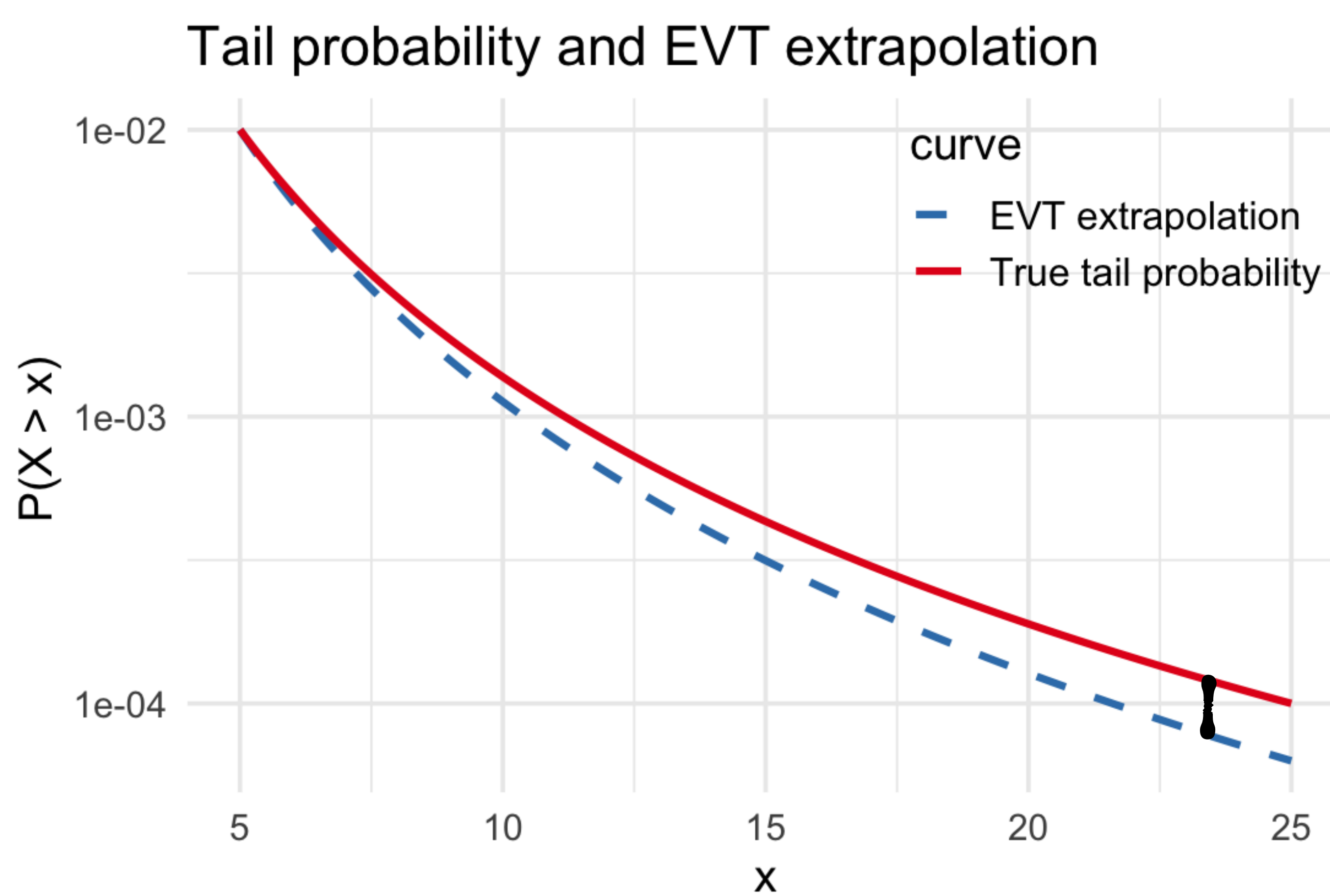
(Why does it work)

- Basic Example: Let us try to estimate the probability of observing a loss in excess of u using EVT.

- Standing Assumption: $\bar{F} \in \mathrm{RV}(-1/\xi)$ for some $\xi > 0$

- Observation: $\bar{F}(u) \approx (u/u_0)^{-1/\xi}\bar{F}(u_0)$, whenever BOTH u and u_0 are large and $u \gg u_0$.

- Approximation strategy: Find an appropriate intermediate level u_0 and set $\bar{F}_{\text{app}}(u) = (u/u_0)^{-1/\xi}\bar{F}(u_0)$



If u_0 chosen properly, $\bar{F}_{\text{app}}(u)$ is "close" to $\bar{F}(u)$

Using EVT for accurate assessment of tail risk

(What all is required?)

- Basic Example: Let us try to estimate the probability of observing a loss in excess of u using EVT.

- Standing Assumption: $\bar{F} \in \mathrm{RV}(-1/\xi)$ for some $\xi > 0$

Proposed Approximation: $\bar{F}_{\text{app}}(u) = (u/u_0)^{-1/\xi}\bar{F}(u_0)$

- Q1: How to choose intermediate level u_0?
  - Choice of u_0 –> bias-variance in EVT:
  - Small $u_0 \Longrightarrow$ RV does not kick in $\Longrightarrow$ large bias
  - Large $u_0 \Longrightarrow$ noisy estimates of $\bar{F}(u_0) \Longrightarrow$ large variance

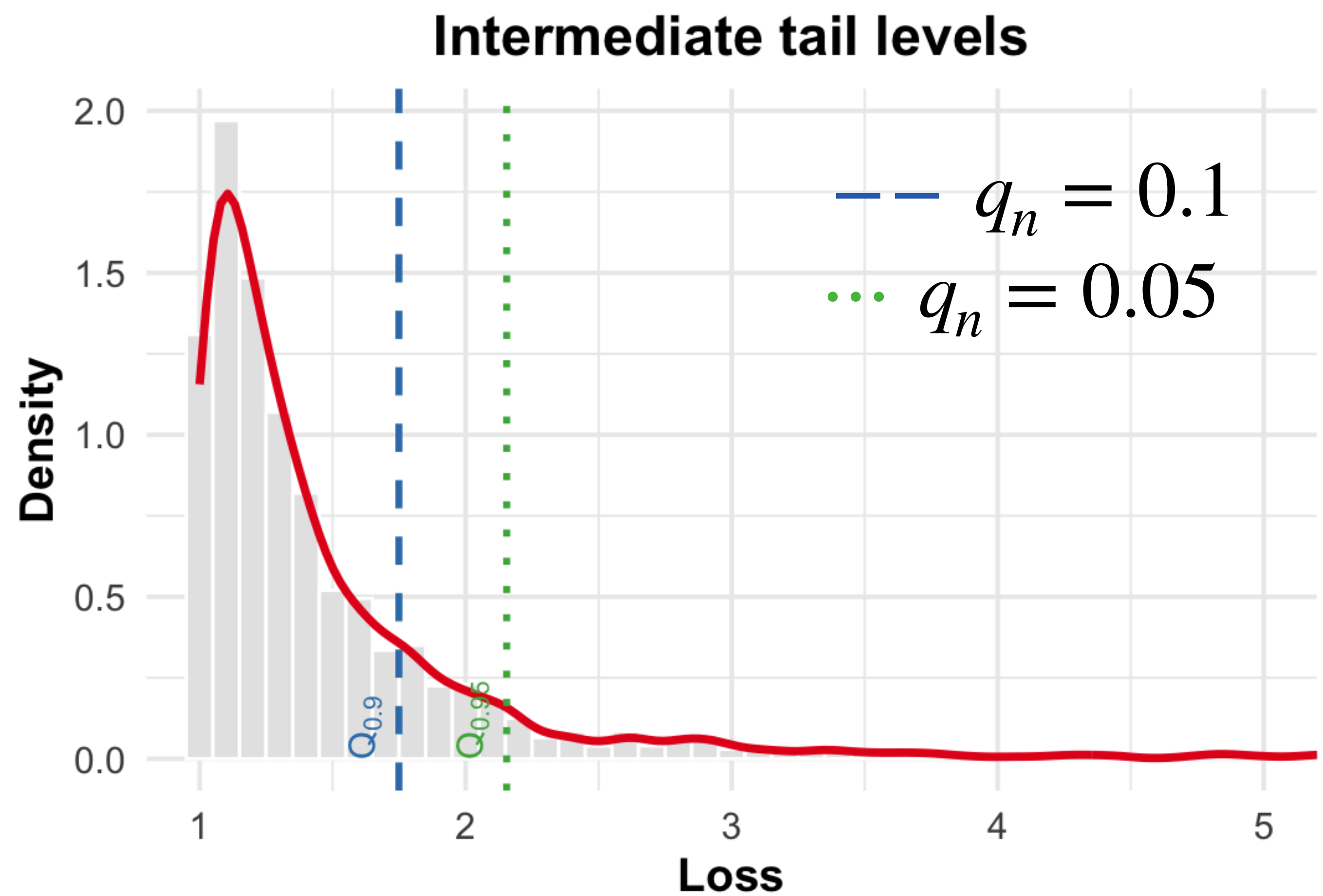
- Q2: How to estimate the tail parameter ξ?
  - ξ associated with the Pareto distribution approximating the tail of data
  - Need to choose threshold from which the tail "starts".

Using EVT for accurate assessment of tail risk

(The baseline level: intermediate quantiles)

Proposed Approximation: $\bar{F}_{\text{app}}(u) = (u/u_0)^{-1/\xi}\bar{F}(u_0)$

- For proposed approximation to work:
  - u_0 has to be large enough that it is in the tail
  - It should be chosen such that there is sufficient data to estimate $\bar{F}(u_0)$



Choice of u_0: Given $\{X_1, \ldots, X_n\}$ choose $u_0 = \lfloor nq_n \rfloor$th largest element.

Typical choice: $q_n = n^{-\kappa}$ for $\kappa \in (0,1)$

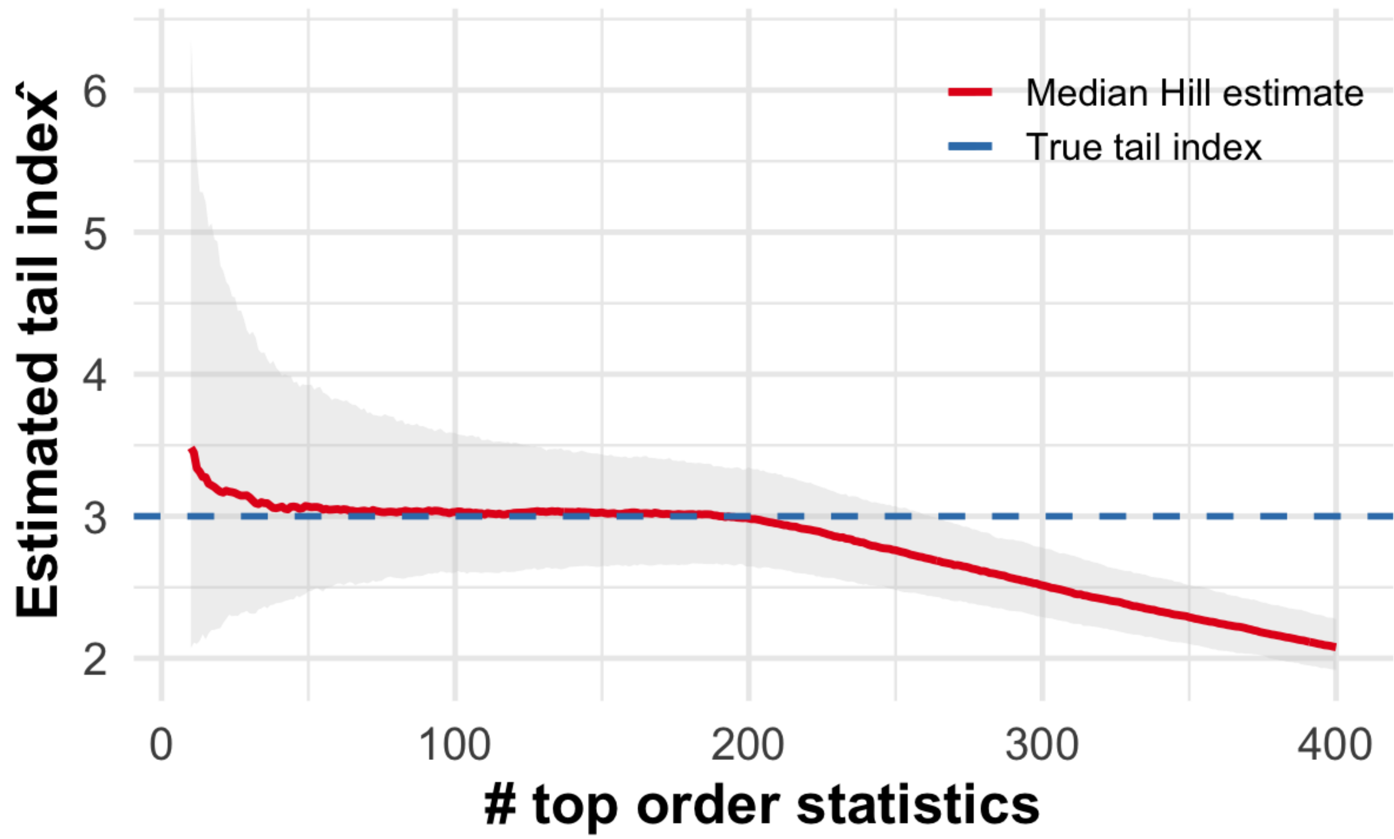
($nq_n \to \infty$ and $q_n \to 0$)

Using EVT for accurate assessment of tail risk

(Estimation of the tail index - The Hill Estimator)

Proposed Approximation: $\bar{F}_{\text{app}}(u) = (u/u_0)^{-1/\xi}\bar{F}(u_0)$

- $\xi \rightarrow$ property of the tail of the data
- Therefore estimated from "large" samples



Hill estimator

Step 1: $\{X_{(1)}, \ldots, X_{(n)}\} \rightarrow$ data $\{X_1, \ldots, X_n\}$ arranged in descending order

Step 2: Return

$$\xi = \frac{1}{k_n}\sum_{i=1}^{k_n} \log\left(\frac{X_{(i)}}{X_{(k_n)}}\right)$$

Using EVT for accurate assessment of tail risk

(Estimation of Tail Probabilities)

Proposed Approximation: $\bar{F}_{\text{app}}(u) = (u/u_0)^{-1/\xi}\bar{F}(u_0)$

Data driven estimation of tail probability: Given n data samples $\{X_1, \ldots, X_n\}$, $u := u_n$

- Step 1 (Filter Tail): Choose $u_0 = X_{(k_n)}$ where $k_n = \lfloor nq_n \rfloor$. Let $\mathcal{T}_n = \{X_{(1)}, \ldots, X_{(k_n)}\}$.
- Step 2 (Tail Index): Output estimate $\hat{\xi}_n$ using Hill estimator on $\mathcal{T}_n$
- Step 3 (Estimate Probability): Output $\hat{p}_n = (u_n/X_{(k_n)})^{-1/\hat{\xi}_n} q_n$.

Big Theorem: Suppose $q_n = n^{-\kappa}$ for $\kappa \in (0,1)$ and let u_n be such that $\bar{F}(u_n) = n^{-\tau}$ where $\tau \geq 1$. Then,

$$\frac{\hat{p}_n}{\bar{F}(u_n)} - 1 \to 0 \text{ almost surely}$$

- Takeaway: EVT provides a structured way of estimating tail probabilities in presence of limited data —> model risk averted!

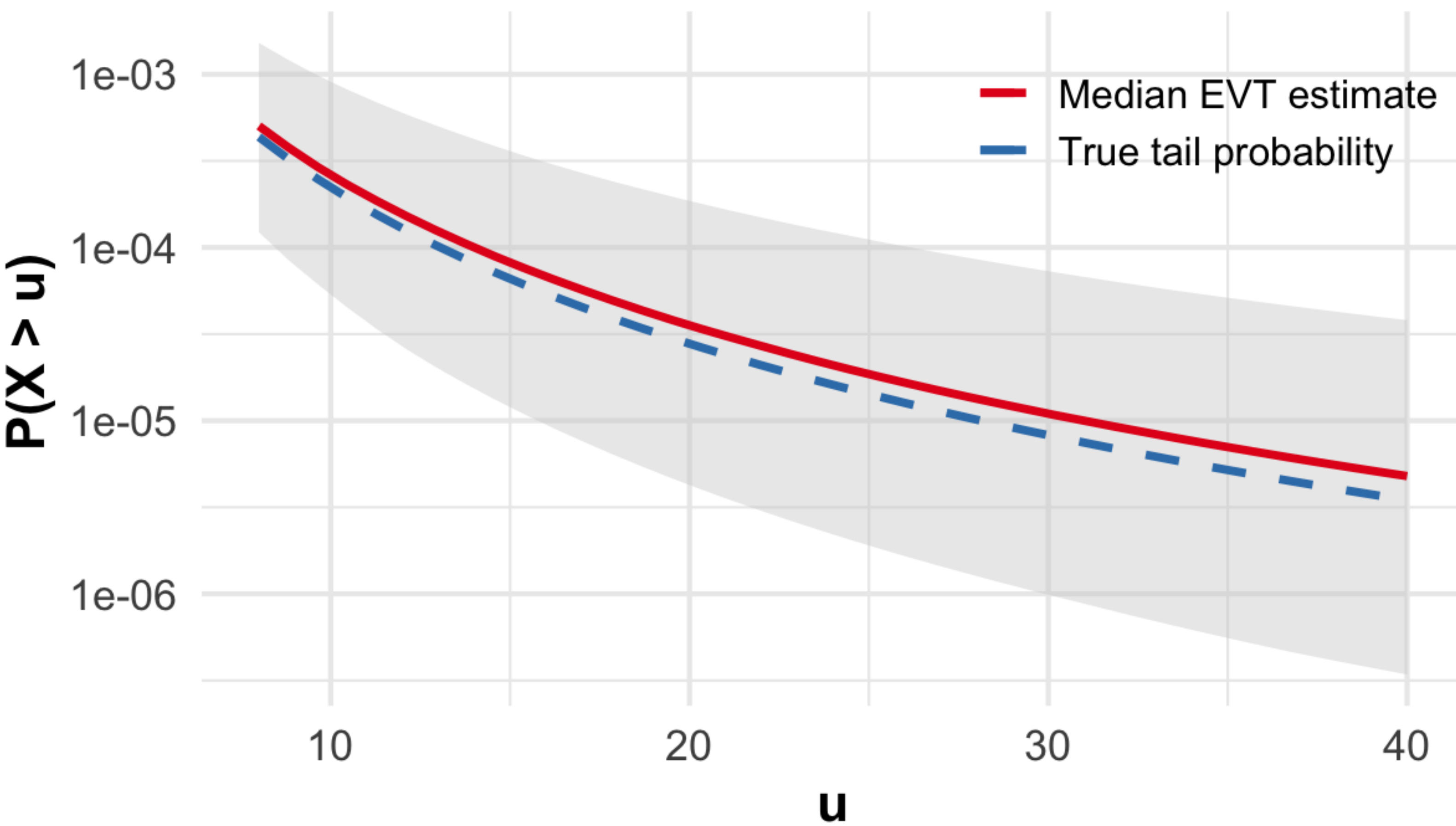
Using EVT for accurate assessment of tail risk

(Estimation of Tail Probabilities)

Proposed Approximation: $\bar{F}_{\text{app}}(u) = (u/u_0)^{-1/\xi} \bar{F}(u_0)$

Data driven estimation of tail probability: Given n data samples $\{X_1, \ldots, X_n\}$, $u := u_n$

- **Step 1 (Filter Tail):** Choose $u_0 = X_{(k_n)}$ where $k_n = \lfloor n q_n \rfloor$. Let $\mathcal{T}_n = \{X_{(1)}, \ldots, X_{(k_n)}\}$.
- **Step 2 (Tail Index):** Output estimate $\hat{\xi}_n$ using Hill estimator on $\mathcal{T}_n$
- **Step 3 (Estimate Probability):** Output $\hat{p}_n = (u_n / X_{(k_n)})^{-1/\hat{\xi}_n} q_n$.



- Number of input samples: 500
- Target probability: 1e-3 - 1e-5
- Observation: Model risk taken care of even though there are 0 samples observed in tail region!

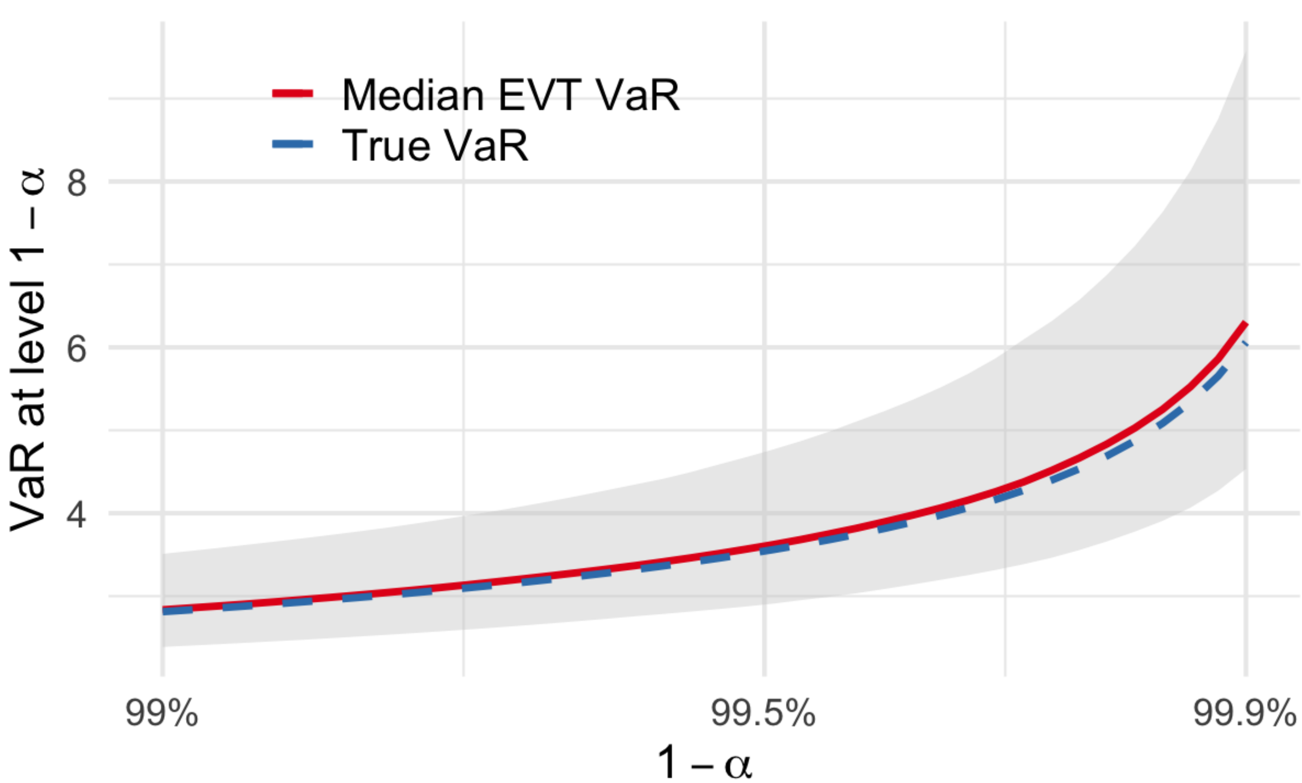
Using EVT for accurate assessment of tail risk

(Estimation of VaR/CVaR)

Theorem: Let $\bar{F} \in \mathrm{RV}(-1/\xi)$. Then as $\alpha \to 0$,

$$\mathrm{VaR}_{1-\alpha}(P) \in \mathrm{RV}(\xi) \qquad \text{and} \qquad \mathrm{CVaR}_{1-\alpha}(P) \sim \frac{1}{1-\xi}\mathrm{VaR}_{1-\alpha}(P)$$

- **Implication:** VaR/CVaR at level α can be approximated using VaR/CVaR at level α_0 where $\alpha_0 \gg \alpha$



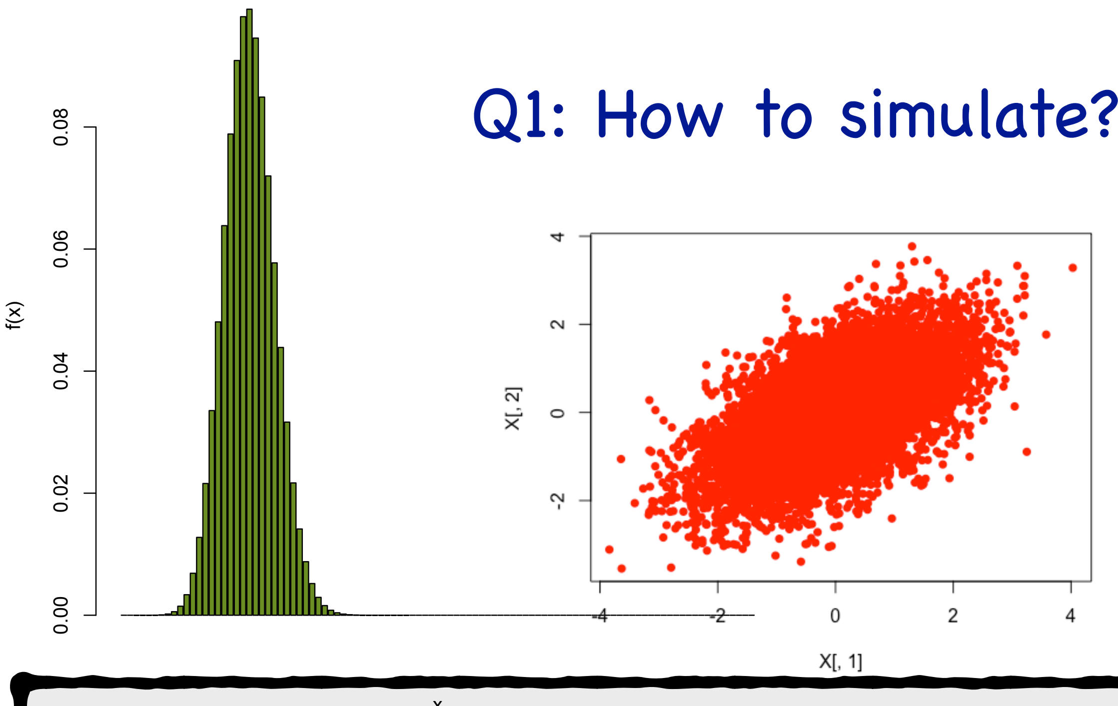
Data driven estimation of VaR/CVaR: Given n data samples $\{X_1, \ldots, X_n\}$, target α.

- **Step 1 (Filter Tail):** Set $k_n = \lfloor n\alpha_0 \rfloor$ and estimate VaR/CVaR at level $1 - q_n$
- **Step 2 (Tail Index):** Estimate ξ using Hill estimator
- **Step 3:** Output $\widehat{\mathrm{VaR}}_{1-\alpha} = (\alpha_0/\alpha)^{\hat{\xi}_n} X_{(k_n)}$ (and same for CVaR)

A recap of the course

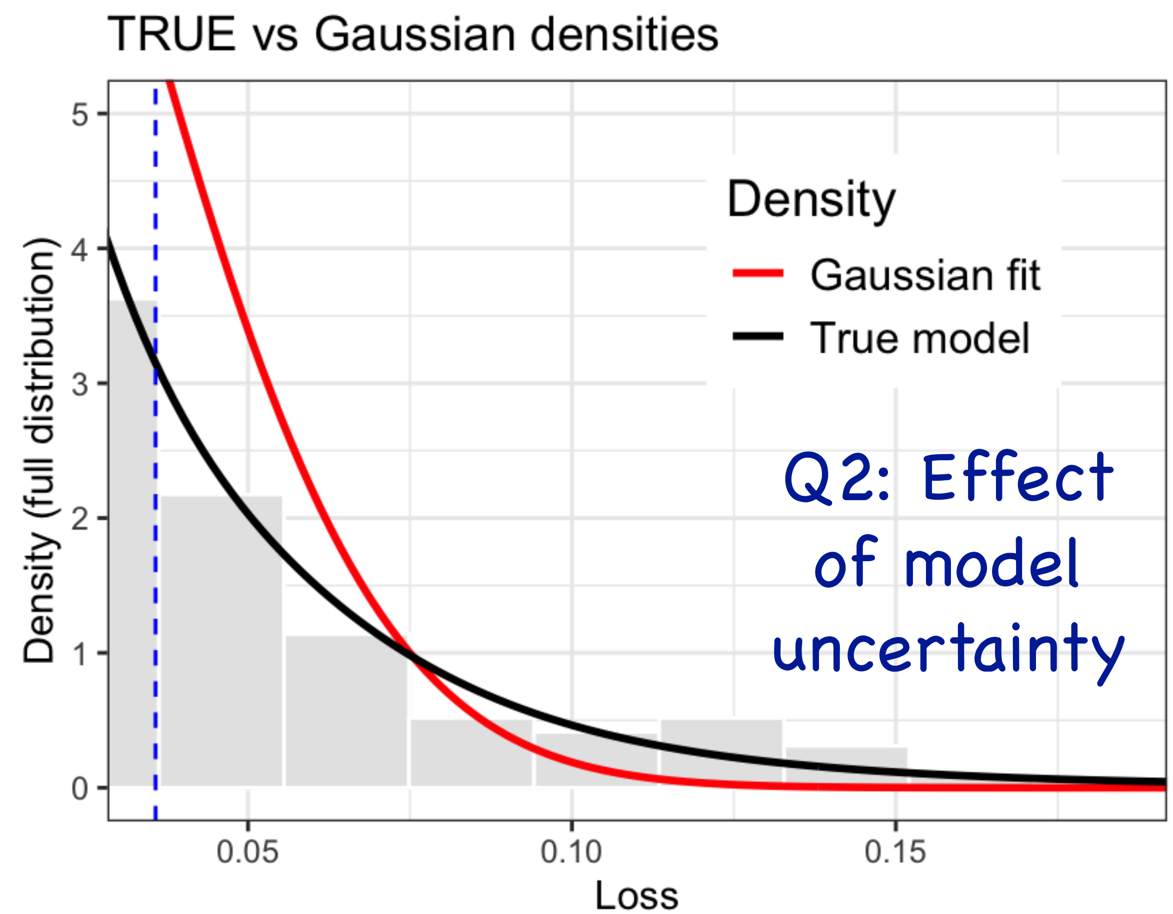
(What did we learn in these sessions?)

Q1: How to simulate?



Solution: Input Modelling

- Bootstrap —> non-parametric
- For multivariate data —> use copula
- However, beware model risk!



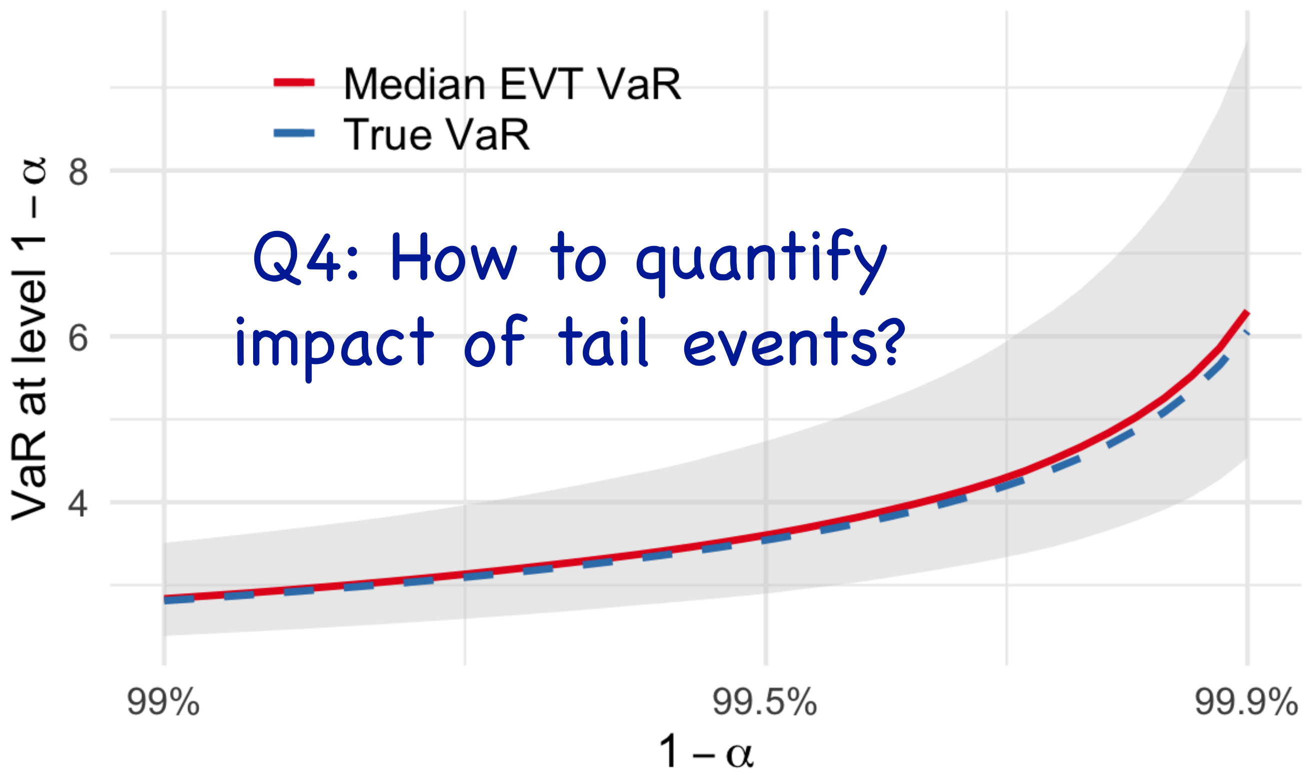
Q2: Effect of model uncertainty

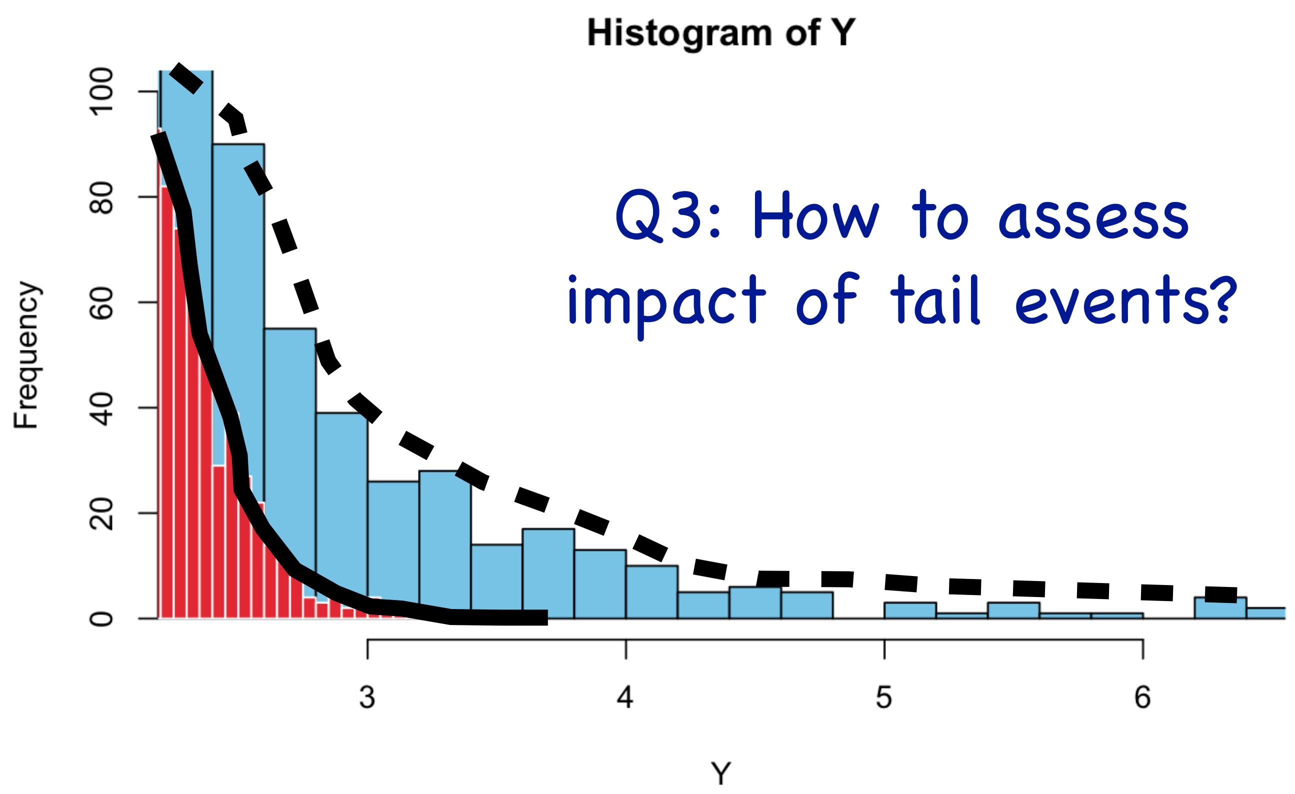
Solution: Analyse Model Risk

- Model Risk: samples —> decisions
- Pick a model that captures the features of data that you care about!

A recap of the course

(What did we learn in these sessions?)





Solution: Use EVT to estimate risk

- From data —> decisions via EVT + statistical guarantees

Solution: Extreme Value Theory

- Overcoming the model risk in estimation of tails
- Basic principles behind EV design

The End!

Thank you for attending!

Coherent Risk Measures

(What properties do we want risk measures to have)

Assumption: $\rho(X)$ —> risk measure depending on the distribution of X

Property 1 (Monotone): If $X_1 \geq X_2$ almost surely, then $\rho(X_1) \geq \rho(X_2)$

- Interpretation: If the loss of an investment always exceeds the another, then the first investment is more risky than the second

Property 2 (Linear): $\rho(\lambda X + c) = \lambda \rho(X) + c$

- Interpretation: Risk is linear in X

Property 3 (Sub-additive): $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$

- Interpretation: Diversification reduces risk

If properties 1,2 and 3 hold then ρ is called a coherent measure of risk

Coherent Risk Measures

(What properties do we want risk measures to have)

Assumption: $\rho(X)$ —> risk measure depending on the distribution of X

Property 1 (Monotone): If $X_1 \geq X_2$ almost surely, then $\rho(X_1) \geq \rho(X_2)$

- Interpretation: If the loss of an investment always exceeds the another, then the first investment is more risky than the second

Property 2 (Linear): $\rho(\lambda X + c) = \lambda \rho(X) + c$

- Interpretation: Risk is linear in X

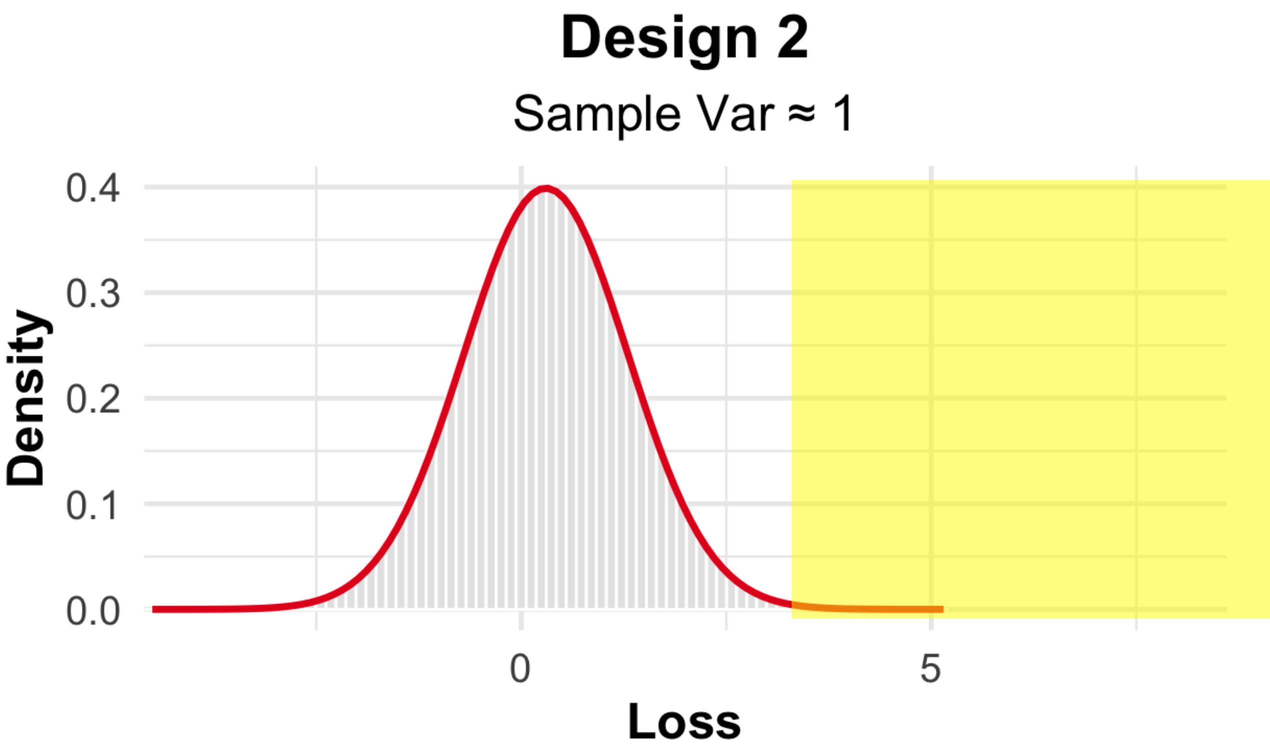
Property 3 (Sub-additive): $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$

- Interpretation: Diversification reduces risk

Theorem: CVaR is coherent, but VaR is not (Property 3 fails)

Measures of Tail Risk

(A general theory of extreme monetary risk)



- VaR/CVaR are tail focussed: they ignore the center and are determined by what happens in the worst α fraction of losses

Definition: Let $X \sim F$. Then, the α-restriction of X, is defined as

$$F_\alpha(x) = F(x \mid X \geq \text{VaR}_{1-\alpha}(X))$$

- Interpretation: α-restriction captures tail behaviour.

Definition: ρ is said to be a $\alpha-$tail risk measure if for any random variables $X \sim F_X$ and $Y \sim F_Y$ with $F_{\alpha,X} = F_{\alpha,Y}$, $\rho(X) = \rho(Y)$

- Interpretation: ρ depends only on the largest α fraction losses!

Measures of Tail Risk

(A general theory of extreme monetary risk)

Definition: ρ is said to be a $\alpha-$tail risk measure if for any random variables $X \sim F_X$ and $Y \sim F_Y$ with $F_{\alpha,X} = F_{\alpha,Y}$, $\rho(X) = \rho(Y)$

- Interpretation: ρ depends only on the largest α fraction losses!

Theorem: VaR$_{1-\alpha}$ and CVaR$_{1-\alpha}$ are both α-tail risk measures

Proof sketch: Assume densities exist!

i) For VaR: Let X and Y have the same $\alpha-$restrictions. Then, at the very least, they must have the same VaR at $(1-\alpha)$ level!

ii) For CVaR: CVaR has the representation CVaR$_{1-\alpha}(X) = \displaystyle\int_0^1$ VaR$_{1-\alpha t}(X)dt$. Since α restrictions of X and Y coincide, so must their VaRs at level $(1-s)$ for all $s \in (0,\alpha]$. This implies that CVaR$_{1-\alpha}(X) =$ CVaR$_{1-\alpha}(Y)$ ∎