# Stat 446
# Advanced Short-Term Actuarial Modeling

## Chapter 4: Severity Models

# Outline

1. Overview of ASTAM Exam
2. Severity Models: Creating New Distribution
3. Tails of Distributions
4. Extreme Value Theory
   - Generalized Extreme Value Distribution
   - Generalized Pareto Distribution

# ASTAM Exam

SOA Advanced Short-Term Actuarial Mathematics (**ASTAM**) Exam:

- Severity models: tail behaviors, EVT
- Coverage modifications
- Aggregate models: recursive formula, discretization, sum of CP
- Parametric model: estimation and selection
- Credibility theory: Bayesian, Bühlmann and Bühlmann-Straub, semi/non-parametric estimations
- Ratemaking and reserving: stochastic reserve methods, classification rating

# Scaling: Multiplication by a Constant

- Multiplying by a constant is equivalent to applying inflation uniformly across all loss amounts.

**Change of Scale**

Let $X$ be a continuous R.V. with density $f_X(x)$ and df $F_X(x)$, and $Y = cX$ where $c > 0$. Then,

$$F_Y(y) = \Pr(Y \leq y) = \Pr(cX \leq y) = F_X\left(\frac{y}{c}\right).$$

**Scale Parameter**

In the distribution function, a parameter $\theta$ is called a **scale parameter** if $\theta$ only appears together with $x$ in the form $x/\theta$.

## Raising to a Power

Let $Y = X^{1/\tau}$. If $\tau > 0$, we have

$$F_Y(y) = \Pr(Y \leq y) = \Pr(X \leq y^\tau) = F_X(y^\tau),$$

and

$$f_Y(y) = \frac{d}{dy} F_X(y^\tau) = \tau y^{\tau-1} f_X(y^\tau), \quad y > 0$$

If $\tau < 0$, we have

$$F_Y(y) = 1 - F_X(y^\tau), \quad f_Y(y) = -\tau y^{\tau-1} f_X(y^\tau)$$

# Example: Weibull Distribution

## Example

If $X \sim \text{Exp}(\beta)$, find the distribution for $Y = X^{1/\tau}$ for $\tau > 0$.
[Weibull: $F_Y(y) = 1 - e^{-y^\tau/\beta}$, so $\theta = \beta^{1/\tau}$]

## Weibull Distribution

$Y \sim \text{WEI}(\theta, \tau)$, $\tau > 0$, is a Weibull distribution (transformed exponential) if the df

$$F_Y(y) = 1 - e^{-(y/\theta)^\tau}, \quad y \geq 0.$$

## Values of $\tau$

If $\tau > 0$, the resulting distribution is called **transformed**; if $\tau = -1$, called **inverse**; and if $\tau < 0$ ($\tau \neq -1$), called **inverse transformed**.

# Exponentiation

## Exponentiation

Let $Y = e^X > 0$, we have

$$F_Y(y) = \Pr(Y \le y) = \Pr(X \le \ln y) = F_X(\ln y),$$

and

$$f_Y(y) = \frac{d}{dy} F_X(\ln y) = \frac{1}{y} f_X(\ln y), \quad y > 0$$

## Lognormal Distribution

If $X$ is normally distributed with mean $\mu$ and std $\sigma$, $Y = e^X$ is lognormally distributed with

$$F_Y(y) = \Phi\left(\frac{\ln y - \mu}{\sigma}\right), \quad f_Y(y) = \frac{1}{y\sigma}\phi\left(\frac{\ln y - \mu}{\sigma}\right).$$

# Mixture of Distributions

- If $X$ is a distribution conditional on $\Theta$, which itself is a random variable, the resulting unconditional distribution is called a **mixed distribution**.

**Mixture**

Let $X$ have conditional pdf $f_{X|\Theta}(x|\theta)$ and $\Theta$ have pdf $f_{\Theta}(\theta)$.

$$F_X(x) = \Pr(X \leq x) = \int F_{X|\Theta}(x|\theta) f_{\Theta}(\theta)\, d\theta$$

and

$$f_X(x) = \int f_{X|\Theta}(x|\theta) f_{\Theta}(\theta)\, d\theta$$

# Mixture: Frailty Model

- **Frailty model** is an important mixture model, originated from survival analysis.

## Frailty Model

Let $\Lambda$ be a frailty R.V. quantifying hazard rate uncertainty with MGF $M_\Lambda(z) = \mathbb{E}\left[e^{z\Lambda}\right]$ and the conditional hazard rate be $h_{X|\Lambda}(x|\lambda) = \lambda a(x)$. The conditional survival function of $X$ is

$$S_{X|\Lambda}(x|\lambda) = e^{-\int_0^x h_{X|\Lambda}(t|\lambda)\,dt} = e^{-\lambda A(x)}, \quad A(x) = \int_0^x a(t)\,dt,$$

and, by the law of total expectation, the marginal survival function is

$$S_X(x) = \mathbb{E}\left[S_{X|\Lambda}(x|\Lambda)\right] = \mathbb{E}\left[e^{-\Lambda A(x)}\right] = M_\Lambda(-A(x)).$$

# Example: Pareto Distribution

## Example

Suppose that $X|\Lambda = \lambda$ is exponentially distributed with mean $1/\lambda$, and let $\Lambda$ be a gamma distributed R.V. with mean $\alpha/\theta$ and variance $\alpha/\theta^2$. Find the density function of $X$.
$[f_X(x) = \alpha\theta^\alpha/(x+\theta)^{\alpha+1}]$

## Two-Parameter Pareto Distribution

$X \sim \text{Pareto}(\alpha, \theta)$ is a two-parameter Pareto distribution if

$$F_X(x) = 1 - \left(\frac{\theta}{\theta + x}\right)^\alpha \quad \text{and} \quad f_X(x) = \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}, \quad x \geq 0.$$

# Splicing

- Mixing: each simple distribution operates on subsets of the population.
- **Splicing**: break the domain interval into pieces, and each model governs the behavior of the losses in a particular piece.

**$k$-Component Spliced Distribution**

A $k$-component spliced distribution has pdf

$$f_X(x) = \begin{cases} a_1 f_1(x), & c_0 < x < c_1, \\ a_2 f_2(x), & c_1 < x < c_2, \\ \vdots & \vdots \\ a_k f_k(x), & c_{k-1} < x < c_k, \end{cases}$$

where $a_j$'s are the probability weights with $\sum_{j=1}^{k} a_j = 1$ and $f_j(x)$ must be a proper pdf with all probability on $(c_{j-1}, c_j)$.

# Example: STAM Sample Q211

**Example**

An actuary for a medical device manufacturer initially models the failure time for a particular device with an exponential distribution with mean 4 years.
This distribution is replaced with a spliced model whose density function:

(i) is uniform over $[0, 3]$

(ii) is proportional to the initial modeled density function after 3 years

(iii) is continuous

Calculate the probability of failure in the first 3 years under the revised distribution. $[3/7 = 0.4286]$

(Exercise: ASTAM S24 Q6(a)(b))

# Tail of the Distribution

- We emphasize the impact of large losses on insurers and focus on the **tail of the loss distribution**.

**Tail of the Distribution**

The tail for a loss random variable $X$ is defined as

$$\Pr(X > x) = \int_x^{\infty} f_X(t)\,dt = \overline{F}_X(x) = 1 - F_X(x)$$

Some tools to compare and assess the tail risk:

- **Reliability criteria** (relative concepts): moments, limiting ratios, hazard rate, and residual lifetime.
- **Tail risk measures**: VaR and TVaR.

# Reliability Criteria

- Existence of moments $\mathbb{E}\left[X^k\right]$
  - Moments do not always exist, especially when the density admits large values for large $x$.
  - Existence of moments only up to a certain threshold indicates a heavy-tailed distribution.
- Limiting ratios of two tails, i.e. $c = \lim_{x\to\infty} \overline{F}(x)/\overline{G}(x)$.
  - $c = 0$ indicates $F$ has a lighter tail than $G$.
  - $c = \infty$ indicates $F$ has a heavier tail than $G$.
  - If $0 < c < \infty$ is finite, $F$ and $G$ have similar tails.

# Reliability Criteria (cont.)

- Hazard rate $h(x) = f_X(x)/\overline{F}_X(x)$
  - The probability of failure at $x + dx$, given $X > x$ is

$$h(x)\,dx \approx \Pr\left(X \in (x, x + dx) \,|\, X > x\right)$$

  - Decreasing function $h(x)$ indicates heavy tail.
- Residual lifetime $T_x = X - x \,|\, X > x$.
  - $\Pr(T_x > t) = \overline{F}_X(x + t)/\overline{F}_X(x)$ with the **mean excess loss**

$$e(x) = \mathbb{E}\left[T_x\right] = \int_0^\infty \frac{\overline{F}_X(x + t)}{\overline{F}_X(x)}\,dt$$

  - An increasing mean residual lifetime (IMRL) $e(x)$ indicates heavy tail.

# Example: STAM Sample Q162

## Example

A loss, $X$, follows a 2-parameter Pareto distribution with $\alpha = 2$ and unspecified parameter $\theta$. You are given:

$$\mathbb{E}[X - 100 \,|\, X > 100] = \frac{5}{3}\,\mathbb{E}[X - 50 \,|\, X > 50]$$

Calculate $\mathbb{E}[X - 150 \,|\, X > 150]$. [175]

## Important Property

If $X \sim \text{Pareto}(\alpha, \theta)$, the residual lifetime

$$X - d \,|\, X > d \sim \text{Pareto}(\alpha, \theta + d),$$

with the mean excess loss $e(d) = \frac{\theta + d}{\alpha - 1}$.

# Example: ASTAM Fall 2023 Q2(a)(b)

**Example**

An insurer is analyzing its auto insurance claims with a focus on the right tails of the distributions.

(a) Losses are categorized into two categories: bodily injury and vehicle repairs.
State with reasons which of these categories is more likely to have a heavy-tailed loss distribution. [bodily injury]

(b) (i) State whether an increasing mean excess loss function indicates a heavy-tailed or light-tailed distribution. [heavy-tailed]
(ii) State whether an increasing hazard rate function indicates a heavy-tailed or light-tailed distribution. [light-tailed]

# Example: ASTAM Fall 2023 Q2(c)-(e)

## Example

(c) You are given $X \sim \text{Pareto}(\alpha, \theta_X)$.

(i) Prove that $e_X(d) = \frac{\theta_X + d}{\alpha - 1}$, for $\alpha > 1$.

(ii) Determine an expression for the hazard rate function for $X$, $h_X(x)$, in terms of $\alpha$ and $\theta_X$. $[h_X(x) = \alpha/(\theta_X + x)]$

(d) You are given $Y \sim \text{Weibull}(\tau, \theta_Y)$.

(i) Determine an expression for the hazard rate function for $Y$, $h_Y(x)$, in terms of $\tau$ and $\theta_Y$. $[h_Y(x) = \tau x^{\tau-1}/\theta_Y^\tau]$

(ii) Determine the range of values of $\tau$ for which $Y$ is heavy-tailed, based on the hazard rate function. $[\tau \leq 1]$

(e) You are given that $\alpha = 2$, $\tau = 1/2$, $\theta_X = 20$, and $\theta_Y = 10$. Use the limiting ratio of the survival function of $X$ to the survival function of $Y$ to compare their relative tail weights. [the limit is $\infty$, so $X$ has the heavier tail]

(Exercise: ASTAM Sample Qs 8(a)-(c) & 14; F24 Q4(a)-(c))

# Extreme Value Theory (EVT)

- Insurers might be concerned about the impact of very rare but large losses: the extreme tails of the distributions.
- Conventional statistical methods typically give a good *overall* fit but are less satisfactory in the *extreme* right tail of the loss distribution.
- We therefore introduce **Extreme Value Theory** (EVT) to infer the extreme tail behavior:
  - **Block maxima** models: maximum value in a random sample of losses, e.g., generalized extreme value distribution
  - **Points over threshold** models: losses exceeding some high threshold, e.g., generalized Pareto distribution

# Distributions of Block Maxima

- Consider an i.i.d. sample of $n$ values, $X_1, X_2, \ldots, X_n$ with common CDF $F(x)$.
- The maximum of the sample $M_n = \max\{X_1, X_2, \ldots, X_n\}$ has distribution function:

$$F_n(m) = \Pr(M_n \leq m) = \prod_{i=1}^{n} \Pr(X_i \leq m) = (F(m))^n$$

- Limiting property with two possibilities

$$\lim_{n\to\infty} F_n(m) = 0, \quad \text{if and only if } F(m) < 1$$
$$\lim_{n\to\infty} F_n(m) = 1, \quad \text{if and only if } F(m) = 1$$

# Maximum Domain of Attraction (MDA)

- We may further *normalize* the block maximum with deterministic functions of $n$ (involving parameters of $F(x)$), $c_n > 0$, and $d_n$:

$$M_n^* = \frac{M_n - d_n}{c_n}$$

**Maximum Domain of Attraction (MDA)**

$F(x)$ is in the **maximum domain of attraction** (MDA) of $H$ if the limit of $M_n^*$ goes to a non-degenerate $H(x)$, i.e.,

$$\lim_{n\to\infty} \Pr(M_n^* \leq x) = \lim_{n\to\infty} F_n(c_n x + d_n) = H(x)$$

# Choosing the Normalizing Constants

- $d_n$ locates the typical size of the maximum $M_n$ and $c_n$ sets its scale. They are not unique: any choice that gives a non-degenerate limit works.
- A standard choice is the $(1 - 1/n)$ quantile and the reciprocal of the hazard rate there:

$$d_n = F^{-1}\left(1 - \tfrac{1}{n}\right), \qquad c_n = \frac{1}{h(d_n)} = \frac{\overline{F}(d_n)}{f(d_n)}.$$

- Exponential with mean $\beta$: $d_n = \beta \ln n$ and $c_n = \beta$.
- Pareto with parameters $\alpha$ and $\theta$: $d_n = \theta\left(n^{1/\alpha} - 1\right)$ and $c_n = \dfrac{\theta + d_n}{\alpha} = \dfrac{\theta}{\alpha} n^{1/\alpha}$.

# Limits of Normalized Block Maxima

There are three important distributions for limits of normalized block maxima:

- The **Gumbel distribution**

$$F(x) = \exp\left\{-\exp\left(-\frac{x-\mu}{\theta}\right)\right\}, \quad \theta > 0.$$

- The **Fréchet distribution**

$$F(x) = \exp\left\{-\left(\frac{x-\mu}{\theta}\right)^{-\alpha}\right\}, \quad x > \mu,\ \alpha > 0,\ \theta > 0.$$

- The **Weibull EV distribution**

$$F(x) = \exp\left\{-\left(\frac{\mu-x}{\theta}\right)^{\tau}\right\}, \quad x < \mu,\ \tau > 0,\ \theta > 0.$$

# Generalized Extreme Value Distribution

**Fisher-Tippett-Gnedenko Theorem**

If $F$ lies in the MDA of a non-degenerate $H$, then $H$ must be the **generalized extreme value (GEV) distribution** $H_\xi(x)$.

- The **generalized extreme value (GEV) distribution** $H_\xi(x)$ with shape parameter $\xi$:

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

  - $\xi = 0$: Gumbel distribution with $\mu = 0$ and $\theta = 1$.
  - $\xi > 0$: Fréchet distribution with $\mu = -1/\xi$, $\theta = 1/\xi$, $\alpha = 1/\xi$.
  - $\xi < 0$: Weibull EV distribution with $\mu = -1/\xi$, $\theta = -1/\xi$, $\tau = -1/\xi$.

# Generalized Extreme Value Distribution (cont.)

- The GEV can be adjusted for scale $\theta$ and location $\mu$:

$$H_{\xi,\mu,\theta}(x) = \begin{cases} \exp\left(-\left(1+\frac{\xi(x-\mu)}{\theta}\right)^{-1/\xi}\right), & \xi \neq 0, 1+\frac{\xi(x-\mu)}{\theta} > 0, \\ \exp\left(-e^{-(x-\mu)/\theta}\right), & \xi = 0. \end{cases}$$

# GEV Illustration

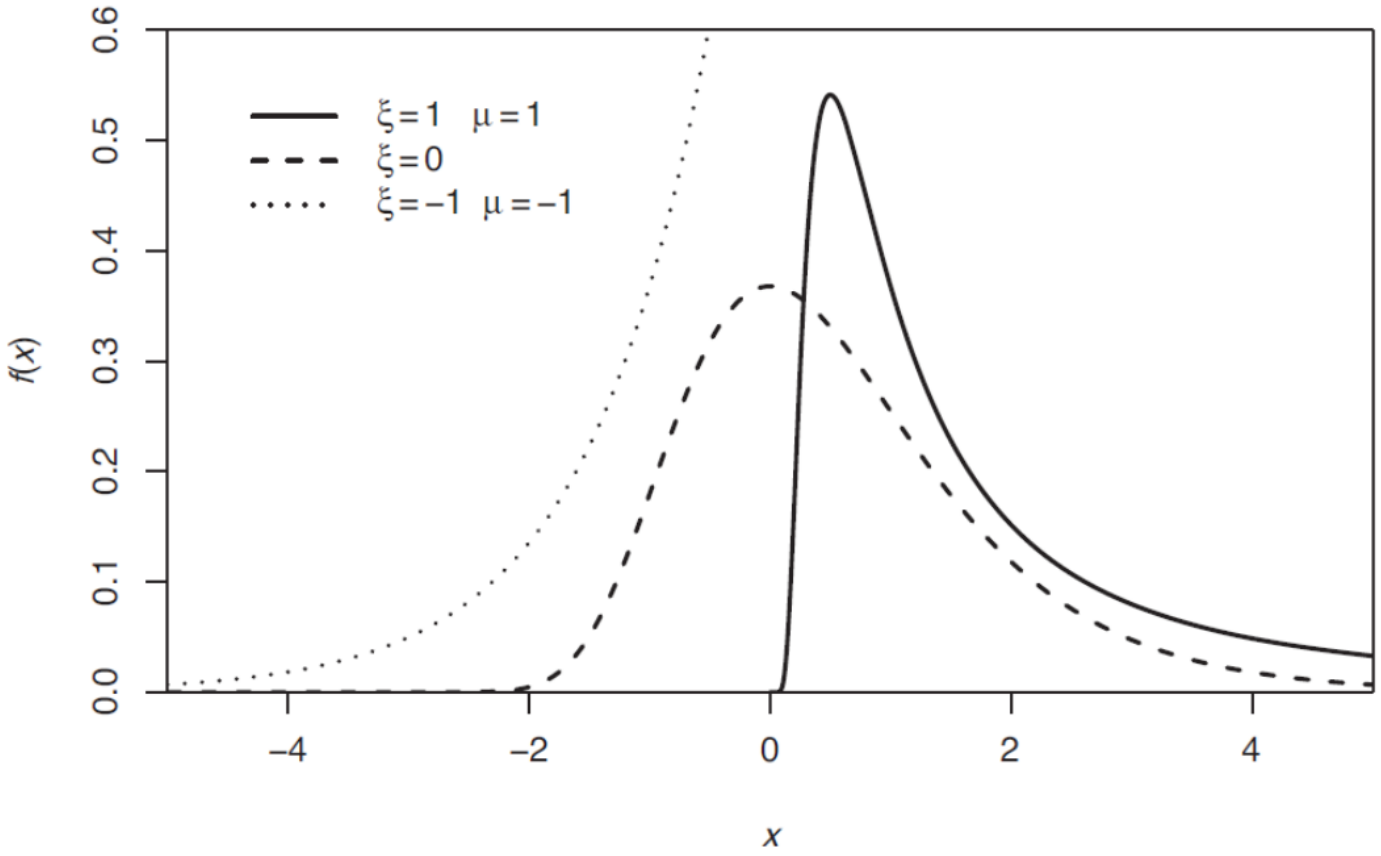



Figure 5.1 Generalized extreme value probability density functions; $\theta = 1$.

# Example: ASTAM-22-23, QERM 5.1 & 5.2

**Example**

For the exponential distribution with mean 1, you are given that $c_n = 1$ and $d_n = \log n$.

- Show that this distribution lies in the Gumbel MDA. $[(1 - e^{-x}/n)^n \to e^{-e^{-x}}]$
- Calculate the probability that $M_n \leq 5$ for $n = 10$ and for $n = 100$, using the exponential distribution function directly, and using the limiting extreme value distribution. [$n = 10$: 0.9346 vs. 0.9348; $n = 100$: 0.5086 vs. 0.5098]

(Exercises: ASTAM Sample Qs 5, 8(d); F23 Q2(f) & S24 Q6(c))

# Further Notes on GEV

- Fréchet distribution ($\xi > 0$, bounded below):
  - Fat-tailed: larger $\xi$ (smaller $\alpha = 1/\xi$) indicates a fatter tail.
  - Any distribution in the Fréchet MDA has a **finite number of moments**, up to (not including) $\alpha = 1/\xi$ (e.g., Pareto, Student's $t$, and Burr distributions).
- Gumbel distribution (unbounded, same shape $\xi = 0$):
  - Distributions in the Gumbel MDA have an **infinite number of moments**, i.e., $\mathbb{E}\left[X^k\right] < \infty$ for any $k \in \mathbb{Z}^+$.
  - Distributions in the Gumbel MDA range from quite thin-tailed (e.g., normal and exponential) to fatter-tailed (e.g., gamma, lognormal, and inverse Gaussian).
- Weibull EV distribution (bounded above):
  - Less useful for large loss modeling.
  - Weibull EV, beta, and uniform distributions are in the Weibull EV MDA.

# Excess Losses Over a Threshold

- For distributions lying in the GEV MDA, the non-normalized $M_n$ approximately follows GEV distributions.
- The key parameter $\xi$ is estimated by fitting the GEV to *block maxima* of our data, using maximum likelihood estimation.
- However, there are some limitations on block maxima (and hence an inefficient way to model tail behavior).
  - Block maximum may not be very extreme in some cases.
  - A block may have several values that would be considered extreme, and only the maximum is actually valued.
- We alternatively select extreme values by considering *all the values falling beyond some threshold*.

# Excess Losses: Distributional Properties

- For the excess loss $Y_d = X - d \,|\, X > d$, we have

$$\overline{F}_{Y_d}(y) = \Pr(X - d > y \,|\, X > d) = \frac{\overline{F}_X(y+d)}{\overline{F}_X(d)}$$

- The mean excess loss (MEL) of $X$ is

$$e_X(d) = \mathbb{E}\left[X - d \,|\, X > d\right] = \frac{\mathbb{E}[X] - \mathbb{E}[X \wedge d]}{\overline{F}_X(d)} = \int_d^\infty \frac{\overline{F}_X(x)}{\overline{F}_X(d)}\, dx$$

- The general asymptotic distribution for excess losses is the **generalized Pareto distribution** (GPD).

## Generalized Pareto Distribution

The CDF of the **generalized Pareto distribution** (GPD) is, for a scale parameter $\beta > 0$,

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

where $x \geq 0$ for $\xi \geq 0$, and $0 \leq x \leq -\beta/\xi$ for $\xi < 0$.

- $\xi > 0$: regular Pareto with $\alpha = 1/\xi$ and $\theta = \beta/\xi$. The $k$th moment only exists for $k < 1/\xi$. In particular, the mean only exists when $0 < \xi < 1$ and is $\beta/(1-\xi)$.
- $\xi = 0$: exponential with mean $\beta$.
- $\xi < 0$: generalized beta distribution (left and right bounded).

# Relationship between GPD and GEV

**Pickands-Balkema-De Haan Theorem**

Let $F$ denote the CDF of a R.V. $X$ which is bounded above at $x^{\text{sup}} \leq \infty$. Then

$$F \in \text{MDA}(H_\xi) \iff \lim_{d \to x^{\text{sup}}} \sup_{0 \leq x \leq x^{\text{sup}} - d} \left| F_{Y_d}(x) - G_{\xi,\beta(d)}(x) \right| = 0$$

for some positive function $\beta(d)$.

- For every distribution in the MDA of $H_\xi$, the excess losses will converge to the $G_{\xi,\beta}$ distribution as the threshold goes to the maximum loss.
- For a sufficiently high threshold $d$, the excess loss $Y_d$ will (approximately) follow the GPD with the same shape parameter $\xi$ as the GEV.

# Determination of Threshold $d^*$

- We examine the empirical MEL for the ordered sample $\left\{x_{(j)}\right\}_{j=1}^{n}$:

$$\hat{e}\left(x_{(j)}\right) = \sum_{k=j+1}^{n} \frac{x_{(k)}}{n-j} - x_{(j)}, \quad j = 1, \ldots, n-1.$$

- The MEL of the GPD over a threshold $k$ is

$$e(k) = \begin{cases} \frac{\beta+\xi k}{1-\xi}, & 0 < \xi < 1, \\ \beta, & \xi = 0. \end{cases}$$

- The first term of $\hat{e}(x_{(j)})$ is the average of the observations above $x_{(j)}$. When $x_{(j)}$ is the empirical VaR, that average is the empirical Expected Shortfall, so $\hat{e} = \widehat{\mathrm{ES}} - \widehat{\mathrm{VaR}}$.

# Determination of Threshold $d^*$ (cont.)

- The GPD MEL is a **straight line** (positive gradient or flat), even if the MEL of the underlying distribution is not.
- The region where the empirical MEL becomes approximately linear indicates where the GPD becomes a good approximation to the distribution of excess losses.
- We then fit the GPD to the reduced sample $X_{(j)} - d^*$ for all $X_{(j)} > d^*$.

# Example: ASTAM Spring 2023 Q4(a)

## Example

A company is modeling the severity of flood losses. The following table shows the largest 20 values from a sample of 750 observations in decreasing order, with their ranks.

| Rank | 750 | 749 | 748 | 747 | 746 | 745 | 744 | 743 | 742 | 741 |
|---|---|---|---|---|---|---|---|---|---|---|
| Loss | 28,250 | 14,782 | 12,427 | 9,676 | 7,005 | 4,322 | 4,057 | 3,773 | 3,504 | 3,123 |

| Rank | 740 | 739 | 738 | 737 | 736 | 735 | 734 | 733 | 732 | 731 |
|---|---|---|---|---|---|---|---|---|---|---|
| Loss | 2,924 | 2,780 | 2,656 | 2,483 | 2,441 | 2,357 | 2,067 | 2,045 | 2,008 | 1,977 |

For convenience, you are also given that the total of the largest 10 losses is 90,919.

(a) (i) Estimate the 98.4% Value at Risk (VaR) of the losses, using the smoothed empirical method. [2778]

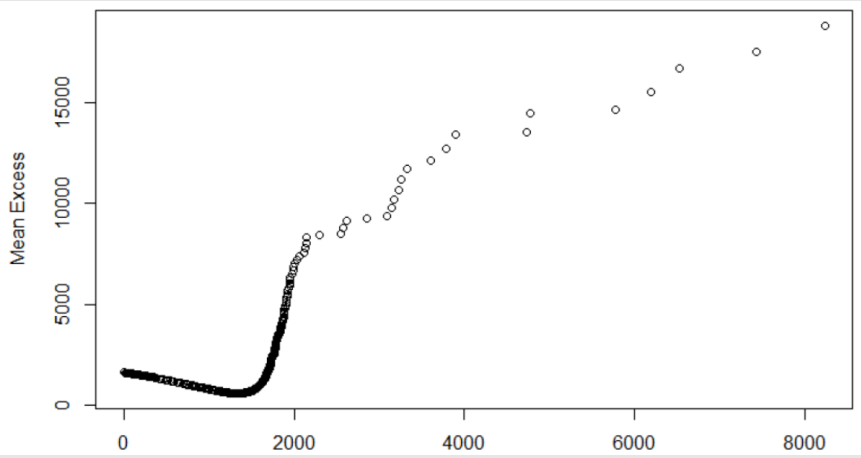
(ii) Estimate the 98.4% Expected Shortfall of the losses. [8052]

(iii) Calculate the empirical Mean Excess Loss, $\hat{e}(x)$, at $x = 3000$. [6091.9]

# Example: ASTAM Spring 2023 Q4(b)

## Example

The actuary decides to estimate the risk measures using the Generalized Pareto Distribution (GPD) for the excess loss distribution. She has created the following plot of the empirical mean excess loss function. (The last five points are excluded.)



# Example: ASTAM Spring 2023 Q4(b) (cont.)

**Example**

(b) The actuary proposes a threshold of $d = 2{,}100$ for fitting the GPD. Her colleague points out that there is a gradient change at 1,500. He states that this would be a better threshold because it provides more data to fit the GPD.
Critique the colleague's suggestion. [the GPD fits only where the MEL plot is linear in the tail; the 1500 kink comes before that]

# Risk Measures for GPD

- We may estimate the VaR and TVaR of $X$ (assuming VaR $> d$) if the excess loss of $X$ over $d$ follows a GPD, i.e.,

$$\begin{aligned}\Pr(X > x) &= \overline{F}_X(d)\,\overline{F}_{Y_d}(x-d) \\ &= \overline{F}_X(d)\left(1+\frac{\xi}{\beta}(x-d)\right)^{-1/\xi}, \quad \xi > 0\end{aligned}$$

- The $\mathrm{VaR}_\alpha(X)$ or $Q_\alpha(X)$ (assuming $Q_\alpha > d$) is

$$\begin{aligned}1-\alpha &= \overline{F}_X(d)\left(1+\frac{\xi}{\beta}(Q_\alpha-d)\right)^{-1/\xi} \\ \Longrightarrow \quad Q_\alpha &= d+\frac{\beta}{\xi}\left(\left(\frac{\overline{F}_X(d)}{1-\alpha}\right)^{\xi}-1\right)\end{aligned}$$

# Risk Measures for GPD (cont.)

- **Remark**: if $X$ is unknown, $\overline{F}_X(d)$ may be empirically estimated as the number of $x_i > d$ over $n$.
- The $\mathrm{TVaR}_\alpha(X)$ or $\mathrm{ES}_\alpha$ (when $\xi < 1$) is

$$\mathrm{ES}_\alpha = Q_\alpha + e_X(Q_\alpha) = Q_\alpha + e_{Y_d}(Q_\alpha - d) = \frac{1}{1-\xi}(Q_\alpha + \beta - \xi d)$$

# Example: ASTAM Spring 2023 Q4(c)(d)

## Example

(c) The MLE estimators with $d = 2{,}100$ are $\hat{\xi} = 0.68$ and $\hat{\beta} = 2{,}000$.

(i) Show that the 98.4% VaR of the loss using the fitted GPD is 2,700 to the nearest 100. You should calculate the value to the nearest 1. [2735]

(ii) Calculate the 98.4% Expected Shortfall (ES) using the fitted GPD. [10,334]

(d) (i) Explain why the VaR estimates in part (a) and in part (c) are similar, while the Expected Shortfall estimates are quite different. [VaR is near the threshold; ES depends on the far tail, which only the GPD extrapolates]

(ii) Recommend which Expected Shortfall estimate the company should use for assessing its potential risk. Justify your answer. [the GPD estimate, since it allows for losses beyond the data]

(Exercises: ASTAM S24 Q6(d)(e) & F25 Q3)