Queen Mary University of London

MTH5126 Statistics for Insurance

Dr Lei Fang

lei.fang@qmul.ac.uk

Week 5

Great credit and thanks to D. Boland, G. Ng and F. Parsa, previous lecturers, for their excellent work in producing the original version of these notes.

Extreme Value Theory

Extreme Value Theory

- Extreme events
- Key idea

Generalised Extreme Value (GEV) distribution

- Fréchet-type, Weibull-type and Gumbel-type GEV distribution

Generalised Pareto Distribution (GPD)

- Threshold exceedances
- Calculating threshold exceedances using R
- GPD

Measures of tail weight:

Existence of moments

Limiting density ratios

- Plotting the limiting density ratios using R

Hazard rate

- Example: Pareto distribution
- Plotting hazard rate using R

Mean residual life

- Example: Pareto distribution
- Plotting mean residual life using R

Extreme Value Theory

Extreme events

Q. Define an 'extreme' event in terms of its frequency and severity.

Answer: An extreme event is one that occurs with very low frequency and very high severity.

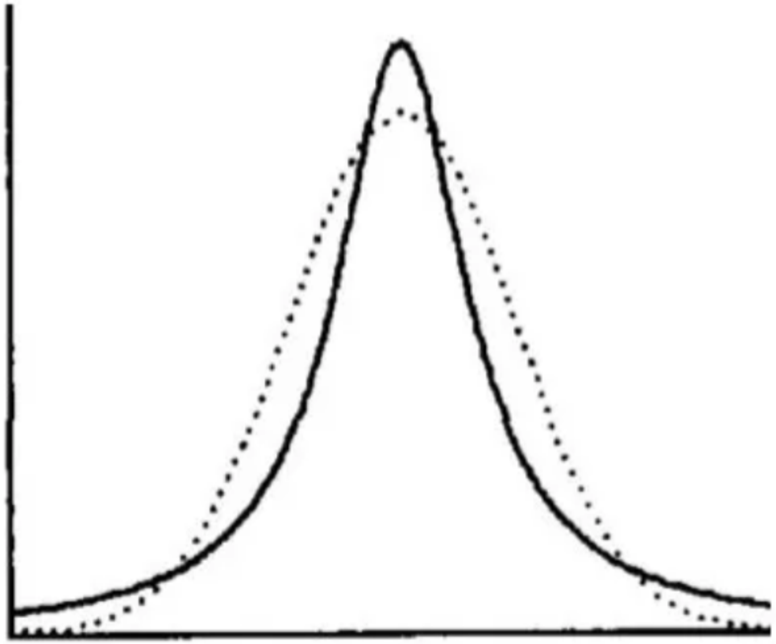
Q. Why is it difficult to model extreme events?

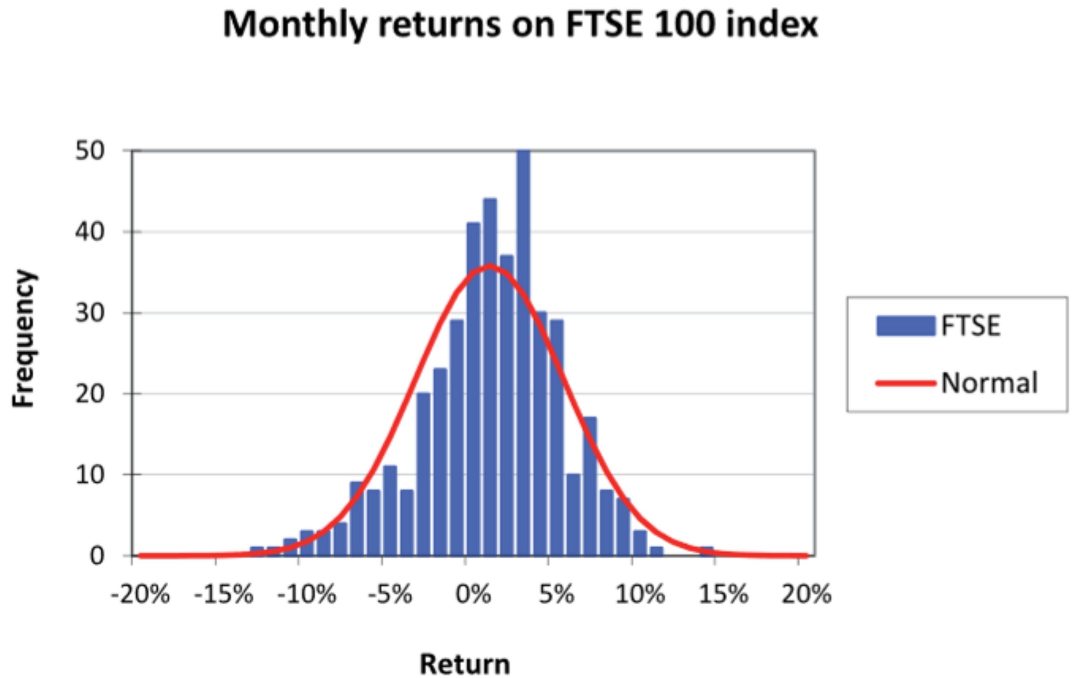
Answer:

- The low frequency of these events means there is relatively **little data** to model their effects accurately.
- The 'true' **distribution** of many types of financial data are more leptokurtic (more narrowly peaked, with fatter tails. Having greater kurtosis than the normal distribution; more concentrated about the mean) than a normal distribution.
- The **volatility** of financial variables does not remain constant, but varies stochastically over **t**. This property is called heteroscedasticity.
- Even if we select an appropriate form of fat-tailed distribution, if we attempt to fit the distribution using the whole of our dataset, this is unlikely to result in a good model for the tails because **parameter** estimates are inappropriately **affected by** the main bulk of the data in the **middle** of the distribution.

Extreme Value Theory

Extreme events





Extreme Value Theory

Key idea

- The key idea of extreme value theory is that the asymptotic behavior of the tails of most distributions can be accurately described by certain families of distributions.

  Q. What does ‘asymptotic behaviour of the tails of a distribution’ mean?

  Answer: The phrase is referring to how the distribution behaves in the limit, as a certain parameter (such as the number of observations in a sample) tends to infinity.

- Two approaches:
  - Modelling the maximum values of a distribution – with the generalised extreme value distributions,
  - Modelling the values exceeding a threshold – with the generalised Pareto distributions.

Generalised Extreme Value (GEV) distribution

- Let losses X_i be IID with cumulative distribution $F(x_i)$.
- Divide your observations into blocks of size n, i.e., in each block you have n observations, X_1 to X_n. Find the maximum in each block. Let

  $X_M = \max\{ X_1 , X_2 , X_3 , \ldots, X_n \}$ be the block maxima

- Then

  $P(X_M \leq x) = P(X_1 \leq x, X_2 \leq x, X_3 \leq x, \ldots, X_n \leq x)$
  $= P(X_1 \leq x)\ P(X_2 \leq x)\ P(X_3 \leq x) \ldots P(X_n \leq x)$, because X_i's are independent
  $= [P(X \leq x)]^n$, because X_i's are identical
  $= [F(x)]^n$

Generalised Extreme Value (GEV) distribution

- Let $\alpha_1, \alpha_2, \ldots, \alpha_n, \ \beta_1, \beta_2, \ldots, \beta_n > 0$ be suitable sequences of real constants.
- We standardise the values of X_M and call the following the standardised block maxima:

$$\frac{X_M - \alpha_n}{\beta_n}$$

- We can attempt to standardise the values of X_M by finding a sequence of constants α_1, $\alpha_2, \ldots, \beta_1, \beta_2$ >0 so that the limiting distribution depends only on x:

$$\lim_{n\to\infty} P\left(\frac{X_M - \alpha_n}{\beta_n} \le x\right) = \lim_{n\to\infty}\left[F(\beta_n x + \alpha_n)\right]^n$$

- The Extreme Value Theorem tells us that it is possible to find such values of α_n, β_n for most common distributions of X.

Generalised Extreme Value (GEV) distribution

- More generally, whatever the distribution of the underlying individual claims, as n increases, the distribution of the standardised maximum values $\frac{X_M - \alpha_n}{\beta_n}$ will converge to a distribution called the GEV distribution with CDF:

$$\lim_{n\to\infty}\left[F\left(\beta_n x + \alpha_n\right)\right]^n = H(x)$$

- The CDF of the GEV distribution is:

$$H(x) = \begin{cases} \exp\left(-\left(1+\dfrac{\gamma(x-\alpha)}{\beta}\right)^{-\frac{1}{\gamma}}\right) & \gamma \neq 0 \\ \exp\left(-\exp\left(-\dfrac{(x-\alpha)}{\beta}\right)\right) & \gamma = 0 \end{cases}$$

- The GEV distribution has 3 parameters:
  1. a location parameter α
  2. a scale parameter $\beta > 0$
  3. a shape parameter γ

Generalised Extreme Value (GEV) distribution

- The parameters α and β rescale (shift and stretch) the distribution. They are analogous to, but do not usually correspond, to the mean and standard deviation.
- The parameter γ determines the overall shape of the distribution (analogous to the skewness). The sign (positive, negative or zero) results in 3 different shaped distribution: Fréchet-type GEV distribution, Weibull-type GEV distribution and Gumbel-type GEV distribution.
- Note that the Weibull-type GEV distribution is not the same thing as the Weibull distribution.

Generalised Extreme Value (GEV) distribution

Fréchet-type, Weibull-type and Gumbel-type GEV distribution

➢ Fréchet-type GEV distributions:

suitable for modelling extreme financial (loss) events because

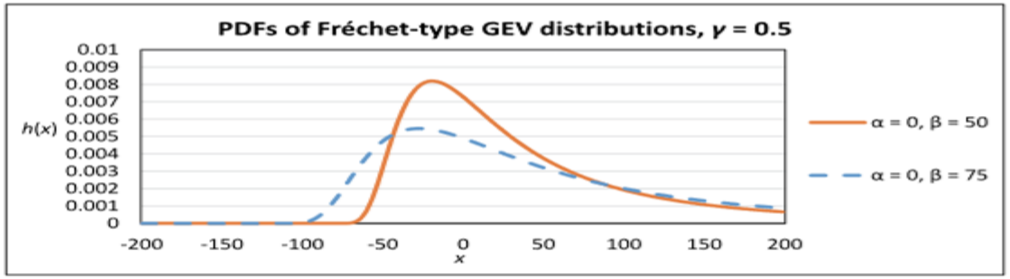
- there is no upper bound to the loss events
- Fréchet-type GEV distributions have a heavier tail (i.e. a tail that decays more slowly to 0) than other types of GEV distribution.

➢ Weibull-type GEV distributions:

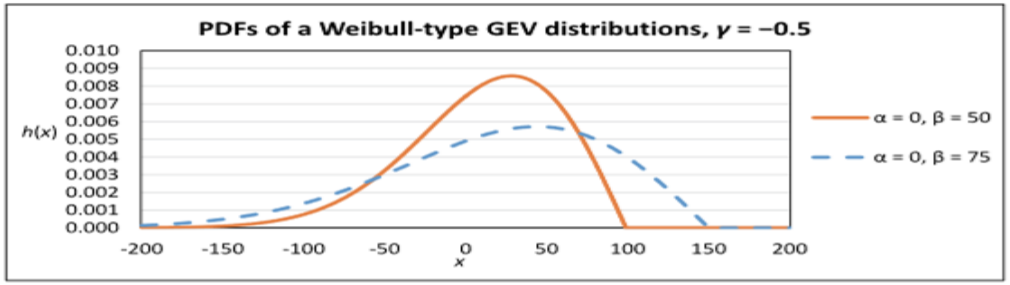
could fit such a distribution to, for example:

- the ages of a human population (indicating an upper bound to possible age) or
- where a loss is certain not to exceed a certain value (for example, if such losses are reinsured).

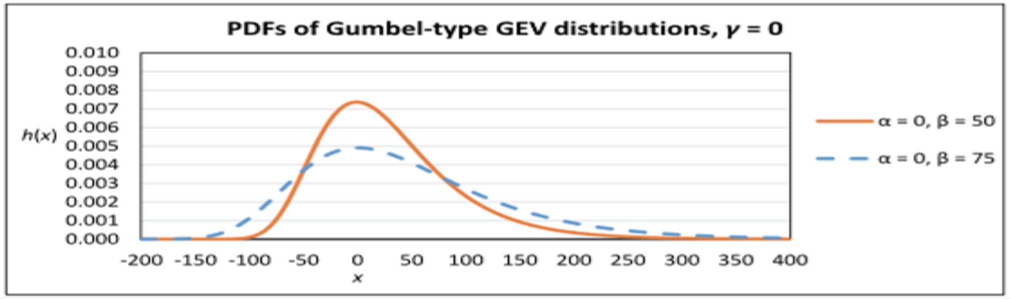
Fréchet-type GEV distributions



Weibull-type GEV distribution



Gumbel-type GEV distributions



Generalised Extreme Value (GEV) distribution

Fréchet-type, Weibull-type and Gumbel-type GEV distribution

- If we know the form of the underlying distribution, it is possible to work out the limiting distribution of the standardised maximum value. We can then use the appropriate member of the GEV family to model the tail of the distribution.
- Note the distinction between the underlying **distribution** (the distribution that applies to the full dataset) and the distribution that we are using to model the extreme values (the **GEV distribution**).
- The underlying distribution will determine which of the three different types of GEV distribution will arise, as shown in the table below. The three types are distinguished by the sign of the shape parameter γ.

Generalised Extreme Value (GEV) distribution

Fréchet-type, Weibull-type and Gumbel-type GEV distribution

	GEV distributions (for the maximum value) corresponding to common loss distributions		
Type **Shape parameter**	**WEIBULL** $\gamma < 0$	**GUMBEL** $\gamma = 0$	**FRÉCHET** $\gamma > 0$
Underlying distribution	**Beta** **Uniform** **Triangular**	**Chi-square** **Exponential** **Gamma** **Lognormal** **Normal** **Weibull**	**Burr** ***F*** **Log-gamma*** **Pareto** ***t***
Range of values permitted	$x < \alpha - \dfrac{\beta}{\gamma}$	$-\infty < x < \infty$	$x > \alpha - \dfrac{\beta}{\gamma}$

Generalised Extreme Value (GEV) distribution

Fréchet-type, Weibull-type and Gumbel-type GEV distribution

As a rough guide:

- For underlying distributions that have **finite upper limit** (eg: uniform dist), the limiting distribution of the standardised max value (i.e., the tail) is of the **Weibull**-type GEV (which also has an upper limit).
- For underlying distributions that are '**light tail**' distributions with finite moments of all orders (e.g.: exponential, normal, lognormal), the limiting distribution of the standardised max value (i.e., the tail) is of the **Gumbel**-type GEV.
- For underlying distributions that are '**heavy tail**' distributions whose higher moments can be **infinite**, the limiting distribution of the standardised max value (i.e., the tail) is of the **Fréchet**-type GEV.

Generalised Pareto Distribution (GPD)

Threshold exceedances

- Let X be a r.v. which follows some underlying distribution. Instead of focusing on the maximum values, we consider the distribution of the values of X that **exceed a specified threshold**, for example, all claims exceeding £100,000.
- For large samples, the distribution of these extreme values converges to the **generalised Pareto distribution**.
- So, we can model the tail of a distribution by choosing a suitably high threshold and then fitting a GPD to the observed values in excess of that threshold.
- Example: Under excess of loss reinsurance, GPD can be used to model the **claim amounts above retention limit** (i.e., the amounts that will be passed to the reinsurer).
- Let X be a r.v. with cumulative distribution function F, then the **excess** over the threshold u is:

$$X - u \mid X > u$$

Generalised Pareto Distribution (GPD)

Threshold exceedances

Choice of *u*:

- Sometimes, the value of the threshold, *u*, may be specified, e.g., if it is a reinsurance retention limit.
- Other times, we may need to make a judgement as to where the threshold should be. Usually, we pick the threshold to be say the 90th or 95th percentile of the underlying distribution.
- The choice of *u* also depends on there being a sufficient volume of data available above the selected threshold.

Generalised Pareto Distribution (GPD)

Threshold exceedances

- If the maximum possible value of X is $x_F \leq \infty$, the cumulative distribution function of the excess is (for $0 \leq x < x_F - u$):

$$F_u(x) = P(X - u \leq x \mid X > u) = \frac{P(X - u \leq x, X > u)}{P(X > u)}$$

$$= \frac{P(X \leq x + u, X > u)}{P(X > u)}$$

$$= \frac{P(X \leq x + u) - P(X \leq u)}{P(X > u)}$$

$$= \frac{F(x + u) - F(u)}{1 - F(u)}$$

Generalised Pareto Distribution (GPD)

Calculating threshold exceedances using R

- Example: Suppose we have monthly claim data stored in a data frame `data` with the first column `month` and the second column `claim`.

  To calculate the threshold exceedances, `xe`, for these claims, at the threshold `u`, we can use the following R code:

  - `x <- data$claim`
  - `xe <- x[x>u] - u`

Generalised Pareto Distribution (GPD)

Generalised Pareto distribution

- More generally, we find that whatever the underlying distribution of *X*, the distribution of the threshold exceedances will converge to a GPD as the threshold *u* increases, i.e.

$$\lim_{u\to\infty} F_u(x) = G(x)$$

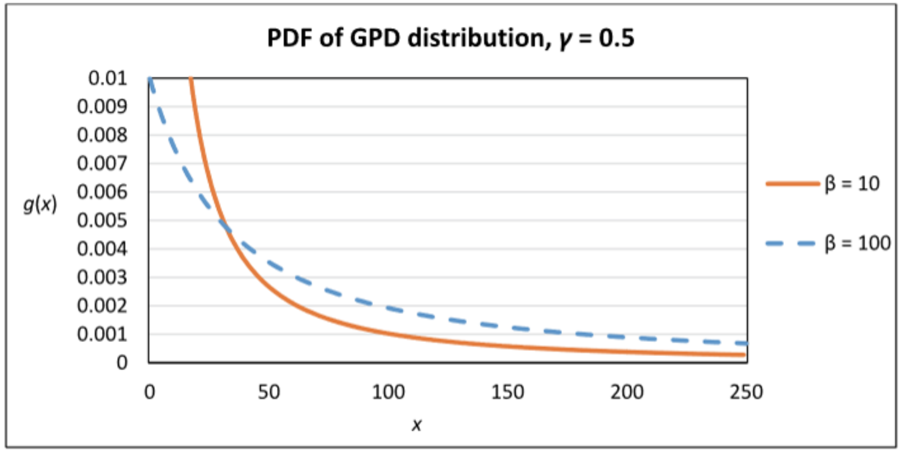
- The generalised Pareto distribution is a two-parameter distribution with CDF:

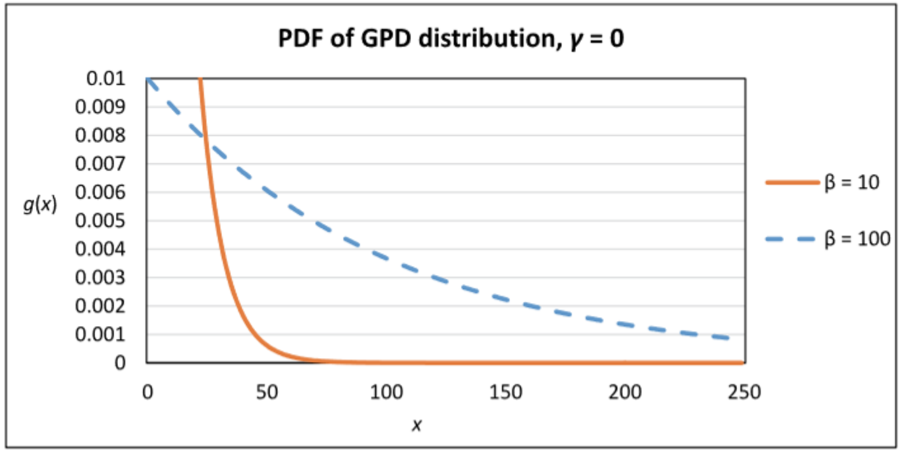
$$G(x) = \begin{cases} 1-\left(1+\dfrac{x}{\gamma\beta}\right)^{-\gamma} & \gamma \neq 0 \\ 1-\exp\left(-\dfrac{x}{\beta}\right) & \gamma = 0 \end{cases}$$

- The GPD distribution has 2 parameters:
    - a scale parameter $\beta > 0$
    - a shape parameter γ

Generalised Pareto Distribution (GPD)

Generalised Pareto distribution





Queen Mary University of London

Measures of tail weight

- Tail weight is a measure of how quickly the (upper) tail of a PDF tends to 0.
- If the PDF of random variable, X_1 , **tends to 0 as x** $\rightarrow \infty$ **more slowly** than the PDF of random variable, X_2 , then X_1 is said to have a **heavier tail** than X_2 .
- Depending on the context, an exponential, normal or lognormal distribution may be considered to be a suitable baseline to use for comparison.
- Let us look at 4 ways of measuring tail weight:
  1) The existence of moments
  2) Limiting density ratios
  3) Hazard rate
  4) Mean residual life

The existence of moments

Reminder:

➢ $E(X^k)$, the kth (non-central) moment of a continuous positive-valued distribution with density function *f*(*x*) is given below and the integral below must converge (i.e. take a finite value) in order for $E(X^k)$ to exist.

$$\boldsymbol{E(X^k) = \int_0^\infty x^k\, f(x) dx}$$

Example: For the Gamma(α, λ) distribution with the following pdf

$$\boldsymbol{f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}}$$

the kth moment exists for all values of k, indicating that it has a relatively light tail.

Queen Mary University of London

The existence of moments

- However, for some distributions, the value of the kth moment doesn't exist beyond a certain value of k (i.e. its value becomes infinite).

Example: For a Pareto(α, λ) distribution with the following pdf

$$f(x) = \frac{\alpha\lambda^{\alpha}}{(\lambda + x)^{\alpha+1}}$$

the kth moment only exists for $\alpha < \lambda$.

So a Pareto distribution (with a low value of the parameter α) will have a much thicker tail.

Limiting density ratios

- We can compare the thickness of the tail of two distributions by calculating the relative values of their density functions at the far end of the upper tail:
- Limiting density ratio = $\lim_{x\to\infty} \frac{f_{X_1}(x)}{f_{X_2}(x)}$
- Comparing the gamma distribution with the Pareto distribution, the presence of the exponential factor in the gamma density results in a limiting density ratio of zero which confirms that the gamma distribution has a lighter tail.

Limiting density ratios

Plotting the limiting density ratios using R

Using the R software:

- We can obtain the values of the PDF of two distributions `d1` and `d2`, for say, x values 1 to 1000 and then calculate the ratio `d1/d2`.
- We can then plot the graph of `d1/d2` against x to determine which of `d1` and `d2` has the thicker tail.

Hazard rate

For a distribution with pdf $f(x)$, cdf $F(x)$,

$$\mathbf{hazard\ rate, h(x)} = \frac{\mathbf{f(x)}}{\mathbf{1 - F(x)}}$$

- The hazard rate is the rate of failure given survival up until that point.
- To interpret the hazard rate, you can think of it as the “force of mortality” at age x.
- If the force of mortality increases as a person’s age increases, relatively few people will live to old age (corresponding to a light tail). If, on the other hand, the force of mortality decreases as the person’s age increases, there is the potential to live to a very old age (corresponding to a heavier tail).

Hazard rate

Example: Pareto distribution

- For the Pareto distribution, we find that the hazard rate is always a decreasing function of x, confirming that it has a heavy tail.
- For the Pareto distribution, the hazard rate:

$$h(x)=\frac{f(x)}{1-F(x)}=\frac{\alpha\lambda^{\alpha}}{(\lambda+x)^{\alpha+1}}\bigg/\left(\frac{\lambda}{\lambda+x}\right)^{\alpha}=\frac{\alpha}{\lambda+x}$$

Queen Mary University of London

Hazard rate

Plotting hazard rate using R

Using the R software:

We can plot the hazard rate for a distribution against x to determine the thickness of its tail.

Example: The R code to calculate the hazard rate, *H,* for a Weibull distribution with shape parameter g and scale parameter $b = c^{-1/g}$ is given by:

```
H <- dweibull(x,g,b)/(1-pweibull(x,g,b))
```

We can then plot the graph of `H` against `x` to determine the thickness of its tail.

Queen Mary
University of London

Mean residual life

For a distribution with pdf $f(x)$, cdf $F(x)$, mean residual life, $e(x)$ is defined as:

$$e(x) = \frac{\int_x^\infty (y-x)f(y)\,dy}{\int_x^\infty f(y)\,dy} = \frac{\int_x^\infty \{1-F(y)\}\,dy}{1-F(x)}$$

- As with hazard rates, we can interpret this in terms of mortality as the expected future lifetime, in other words, the mean residual life gives the expected remaining survival time given survival up until this point.
- If the expected future lifetime decreases with age, relatively few people will live to old age (corresponding to a light tail), but if it increases, there is the potential to live to a very old age (corresponding to a heavier tail).

Mean residual life

Example: Pareto distribution

For the Pareto distribution, we find that the mean residual life is always an increasing function of *x*, confirming that it has a heavy tail.

The mean residual life for the Pareto distribution with $\alpha > 1$ is:

$$e(x) = \frac{\int_x^\infty \{1 - F(y)\}\,dy}{1 - F(x)} = \frac{\int_x^\infty \left(\frac{\lambda}{\lambda + y}\right)^\alpha dy}{\left(\frac{\lambda}{\lambda + x}\right)^\alpha} = (\lambda + x)^\alpha \int_x^\infty (\lambda + y)^{-\alpha}\,dy$$

$$= (\lambda + x)^\alpha \left[\frac{1}{-\alpha + 1}(\lambda + y)^{-\alpha + 1}\right]_x^\infty = (\lambda + x)^\alpha \left[0 - \left(\frac{1}{-\alpha + 1}\right)(\lambda + x)^{-\alpha + 1}\right]$$

$$= \frac{\lambda + x}{\alpha - 1}$$

Queen Mary
University of London

Mean residual life

Plotting mean residual life using R

Using the R software:

We can plot the mean residual life against *x*.

Example: The R code for the survival function of a Weibull distribution with shape parameter g and scale parameter $b = c^{-1/g}$ is given by:

```
Sy <- function(y) {(1-pweibull(y,g,b))}
```

Hence the mean residual life for `x` is given by `ex` as follows:

```
int <- integrate(Sy,x,Inf)
ex <- int$value/(1-pweibull(x,g,b))
```

We can then plot the graph of `ex` against `x`.