Non-Life Insurance: Mathematics and Statistics

Solution sheet 5

Solution 5.1 Large Claims

(a) The density of a Pareto distribution with threshold $\theta = 50$ and tail index $\alpha > 0$ is given by

$$f(x) = f_\alpha(x) = \frac{\alpha}{\theta}\left(\frac{x}{\theta}\right)^{-(\alpha+1)},$$

for all $x \geq \theta$. Using the independence of $Y_1, \ldots, Y_n$, the joint likelihood function $\mathcal{L}_{\mathbf{Y}}(\alpha)$ for the observation $\mathbf{Y} = (Y_1, \ldots, Y_n)$ can be written as

$$\mathcal{L}_{\mathbf{Y}}(\alpha) = \prod_{i=1}^n f_\alpha(Y_i) = \prod_{i=1}^n \frac{\alpha}{\theta}\left(\frac{Y_i}{\theta}\right)^{-(\alpha+1)} = \prod_{i=1}^n \alpha\theta^\alpha Y_i^{-(\alpha+1)},$$

whereas the joint log-likelihood function $\ell_{\mathbf{Y}}(\alpha)$ is given by

$$\ell_{\mathbf{Y}}(\alpha) = \log \mathcal{L}_{\mathbf{Y}}(\alpha) = \sum_{i=1}^n \log\alpha + \alpha\log\theta - (\alpha+1)\log Y_i = n\log\alpha + n\alpha\log\theta - (\alpha+1)\sum_{i=1}^n \log Y_i.$$

The MLE $\widehat{\alpha}_n^{\mathrm{MLE}}$ is defined as

$$\widehat{\alpha}_n^{\mathrm{MLE}} = \arg\max_{\alpha>0} \mathcal{L}_{\mathbf{Y}}(\alpha) = \arg\max_{\alpha>0} \ell_{\mathbf{Y}}(\alpha).$$

Calculating the first and the second derivative of $\ell_{\mathbf{Y}}(\alpha)$ with respect to α, we get

$$\frac{\partial}{\partial\alpha}\ell_{\mathbf{Y}}(\alpha) = \frac{n}{\alpha} + n\log\theta - \sum_{i=1}^n \log Y_i \quad \text{and}$$
$$\frac{\partial^2}{\partial\alpha^2}\ell_{\mathbf{Y}}(\alpha) = \frac{\partial}{\partial\alpha}\left(\frac{n}{\alpha} + n\log\theta - \sum_{i=1}^n \log Y_i\right) = -\frac{n}{\alpha^2} < 0,$$

for all $\alpha > 0$, from which we can conclude that $\ell_{\mathbf{Y}}(\alpha)$ is strictly concave in α. Thus $\widehat{\alpha}_n^{\mathrm{MLE}}$ can be found by setting the first derivative of $\ell_{\mathbf{Y}}(\alpha)$ equal to 0. We get

$$\frac{n}{\widehat{\alpha}_n^{\mathrm{MLE}}} + n\log\theta - \sum_{i=1}^n \log Y_i = 0 \qquad \Longleftrightarrow \qquad \widehat{\alpha}_n^{\mathrm{MLE}} = \left(\frac{1}{n}\sum_{i=1}^n \log Y_i - \log\theta\right)^{-1}.$$

(b) Let $\widehat{\alpha}$ denote the unbiased version of the MLE for the storm and flood data given on the exercise sheet. Since we observed 15 storm and flood events, we have $n = 15$. Thus $\widehat{\alpha}$ can be calculated as

$$\widehat{\alpha} = \frac{n-1}{n}\left(\frac{1}{n}\sum_{i=1}^n \log Y_i - \log\theta\right)^{-1} = \frac{14}{15}\left(\frac{1}{15}\sum_{i=1}^{15} \log Y_i - \log 50\right)^{-1} \approx 0.98,$$

where for $Y_1, \ldots, Y_{15}$ we plugged in the observed claim sizes given on the exercise sheet. Note that with $\widehat{\alpha} = 0.98 < 1$, the expectation of the claim sizes does not exist.

(c) We define $N_1, \ldots, N_{20}$ to be the number of yearly storm and flood events during the twenty years 1986 – 2005. By assumption, we have

$$N_1, \ldots, N_{20} \overset{\text{i.i.d.}}{\sim} \text{Poi}(\lambda).$$

Using Estimator 2.32 of the lecture notes with $v_1 = \cdots = v_{20} = 1$, the MLE $\widehat{\lambda}$ of λ is given by

$$\widehat{\lambda} = \frac{1}{\sum_{i=1}^{20} 1} \sum_{i=1}^{20} N_i = \frac{1}{20} \sum_{i=1}^{20} N_i.$$

Since we observed 15 storm and flood events in total, we get

$$\widehat{\lambda} = \frac{15}{20} = 0.75.$$

(d) Using Proposition 2.11 of the lecture notes, the expected yearly claim amount $\mathbb{E}[S]$ of storm and flood events is given by

$$\mathbb{E}[S] = \lambda \mathbb{E}[\min\{Y_1, M\}].$$

The expected value of $\min\{Y_1, M\}$ can be calculated as

$$\begin{aligned}
\mathbb{E}[\min\{Y_1, M\}] &= \mathbb{E}[\min\{Y_1, M\} 1_{\{Y_1 \leq M\}}] + \mathbb{E}[\min\{Y_1, M\} 1_{\{Y_1 > M\}}] \\
&= \mathbb{E}[Y_1 1_{\{Y_1 \leq M\}}] + \mathbb{E}[M 1_{\{Y_1 > M\}}] \\
&= \mathbb{E}[Y_1 1_{\{Y_1 \leq M\}}] + M\mathbb{P}[Y_1 > M],
\end{aligned}$$

where for $\mathbb{E}[Y_1 1_{\{Y_1 \leq M\}}]$ and $M\mathbb{P}[Y_1 > M]$ we have

$$\begin{aligned}
\mathbb{E}[Y_1 1_{\{Y_1 \leq M\}}] &= \int_\theta^\infty x 1_{\{x \leq M\}} f(x)\, dx = \int_\theta^M x \frac{\alpha}{\theta} \left(\frac{x}{\theta}\right)^{-(\alpha+1)} dx = \alpha\theta^\alpha \left[\frac{1}{1-\alpha} x^{1-\alpha}\right]_\theta^M \\
&= \frac{\alpha}{1-\alpha} \theta^\alpha M^{1-\alpha} - \frac{\alpha}{1-\alpha}\theta = \frac{\alpha}{1-\alpha}\theta \left(\frac{M}{\theta}\right)^{1-\alpha} - \frac{\alpha}{1-\alpha}\theta \\
&= \theta \frac{\alpha}{1-\alpha} \left[\left(\frac{M}{\theta}\right)^{1-\alpha} - 1\right] = \theta \frac{\alpha}{\alpha - 1} \left[1 - \left(\frac{M}{\theta}\right)^{1-\alpha}\right]
\end{aligned}$$

and

$$M\mathbb{P}[Y_1 > M] = M\left(1 - \mathbb{P}[Y_1 \leq M]\right) = M\left(1 - \left[1 - \left(\frac{M}{\theta}\right)^{-\alpha}\right]\right) = \theta \left(\frac{M}{\theta}\right)^{1-\alpha}.$$

Hence, we get

$$\mathbb{E}[\min\{Y_1, M\}] = \theta \frac{\alpha}{\alpha - 1}\left[1 - \left(\frac{M}{\theta}\right)^{1-\alpha}\right] + \theta\left(\frac{M}{\theta}\right)^{1-\alpha} = \theta\frac{\alpha}{\alpha-1} - \frac{\theta}{\alpha-1}\left(\frac{M}{\theta}\right)^{1-\alpha}.$$

Replacing the unknown parameters by their estimates, we get for the estimated expected total yearly claim amount $\widehat{\mathbb{E}}[S]$:

$$\widehat{\mathbb{E}}[S] = \widehat{\lambda}\left[\frac{\theta}{1-\widehat{\alpha}}\left(\frac{M}{\theta}\right)^{1-\hat{\alpha}} - \frac{\widehat{\alpha}}{1-\widehat{\alpha}}\theta\right] \approx 0.75\left[\frac{50}{1-0.98}\left(\frac{2'000}{50}\right)^{1-0.98} - \frac{0.98 \cdot 50}{1-0.98}\right] \approx 180.4.$$

(e) Since $S \sim \text{CompPoi}(\lambda, G)$, we can write S as

$$S = \sum_{i=1}^N Y_i,$$

where $N \sim \text{Poi}(\lambda)$, $Y_1, Y_2, \ldots$ are i.i.d. with distribution function G and N and $Y_1, Y_2, \ldots$ are independent. Since we are only interested in events that exceed the level of $M = 2$ billion CHF, we define S_M as

$$S_M = \sum_{i=1}^{N} Y_i 1_{\{Y_i > M\}}.$$

Due to Theorem 2.14 of the lecture notes, we have $S_M \sim \text{CompPoi}(\lambda_M, G_M)$ for some distribution function G_M and

$$\lambda_M = \lambda \mathbb{P}[Y_1 > M] = \lambda\left(1 - \mathbb{P}[Y_1 \leq M]\right) = \lambda\left(1 - \left[1 - \left(\frac{M}{\theta}\right)^{-\alpha}\right]\right) = \lambda\left(\frac{M}{\theta}\right)^{-\alpha}.$$

Defining a random variable $N_M \sim \text{Poi}(\lambda_M)$, the probability that we observe at least one storm and flood event in a particular year is given by

$$\mathbb{P}[N_M \geq 1] = 1 - \mathbb{P}[N_M = 0] = 1 - \exp\{-\lambda_M\} = 1 - \exp\left\{-\lambda\left(\frac{M}{\theta}\right)^{-\alpha}\right\}.$$

If we replace the unknown parameters by their estimates, this probability can be estimated by

$$\widehat{\mathbb{P}}[N_M \geq 1] = 1 - \exp\left\{-\widehat{\lambda}\left(\frac{M}{\theta}\right)^{-\widehat{\alpha}}\right\} \approx 1 - \exp\left\{-0.75\left(\frac{2'000}{50}\right)^{-0.98}\right\} \approx 0.02.$$

Note that in particular such a storm and flood event that exceeds the level of 2 billion CHF is expected roughly every $1/0.02 = 50$ years.

Solution 5.2 Claim Size Distributions (R Exercise)
All of the four considered distributions (gamma, Weibull, log-normal and Pareto) depend on two parameters. Thus, by fixing the mean and the standard deviation, these parameters are uniquely determined. The R code used to create the plots in the Figures 1, 2 and 3 can be found below.

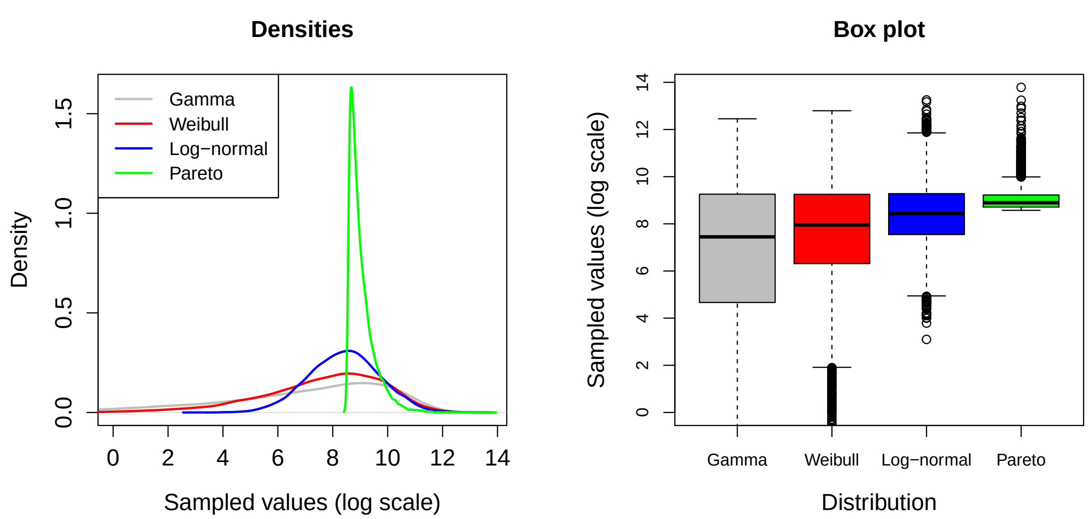

Figure 1: Plot of the densities of the four i.i.d. samples (left). Box plots of the four i.i.d. samples (right).

In Figure 1 we show the densities (left) of the generated i.i.d. samples as well as the corresponding box plots (right), both on a log scale. We only consider logarithmic values starting from 0. We see for example that we have a lot of very small values in case of the gamma distribution (and also in case of the Weibull distribution). The smallest values observed are considerably bigger for the log-normal and especially the Pareto distribution. Moreover, the value of the biggest value observed increases in going from the gamma over the Weibull and the log-normal to the Pareto distribution. We can not say much about the tails from looking at these two plots.

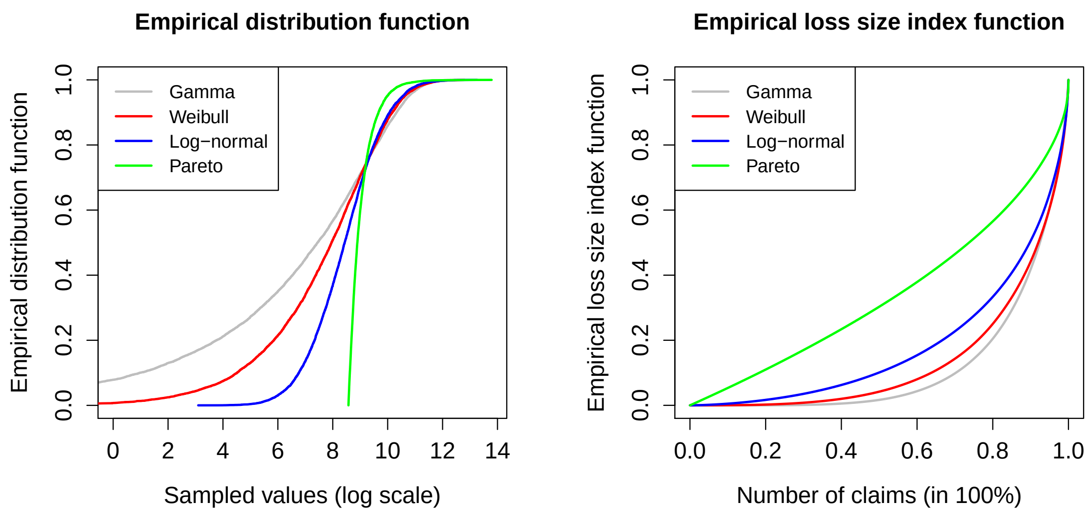

Figure 2: Plot of the empirical distribution functions of the four i.i.d. samples (left). Plot of the empirical loss size index functions of the four i.i.d. samples (right).

In Figure 2 we show the plots of the empirical distribution functions (left, on a log scale) and of the empirical loss size index functions (right) of the generated i.i.d. samples. For the plot of the empirical distribution functions we only consider logarithmic values starting from 0. We observe that the empirical distribution functions almost perfectly intersect at the point with x-coordinate equal to log(10'000)$\approx$ 9.21. This means that for all of the four considered distributions approximately the same percentage of observations is smaller than the expected value. This percentage is roughly equal to 75%, indicating that three quarters of the observations are smaller than the expected value and one quarter of the observations are above the expected value. Thus, not surprisingly, the large claims are the main driver of the expected value. We get confirmed the observations from Figure 1, namely that the smallest values observed are considerably bigger for the log-normal and especially the Pareto distribution, compared to the gamma and the Weibull distribution. This carries over to the plot of the empirical loss size index function. Also these two plots do not tell us much about the tails of the distributions.
In Figure 3 we show the log-log plots (left) and the plot of the empirical mean excess functions (right) of the generated i.i.d. samples. These two plots can be used for studying the tails of the distributions. We see in both plots that the gamma distribution is the most light-tailed distribution. The Weibull distribution and the log-normal distribution have a similar tail behaviour, with slightly heavier tails of the log-normal distribution. Note that this similar tail behaviour is due to the value of the parameter τ of the Weibull distribution being smaller than 1. With a value $\tau \geq 1$ the distribution gets (even) more light-tailed. The most heavy-tailed distribution among the four distributions we analyzed here is the Pareto distribution.
Summarizing, we can say that although we fixed the mean and the standard deviation to be the same, all of the four considered distributions behave differently, implying that one has to carefully

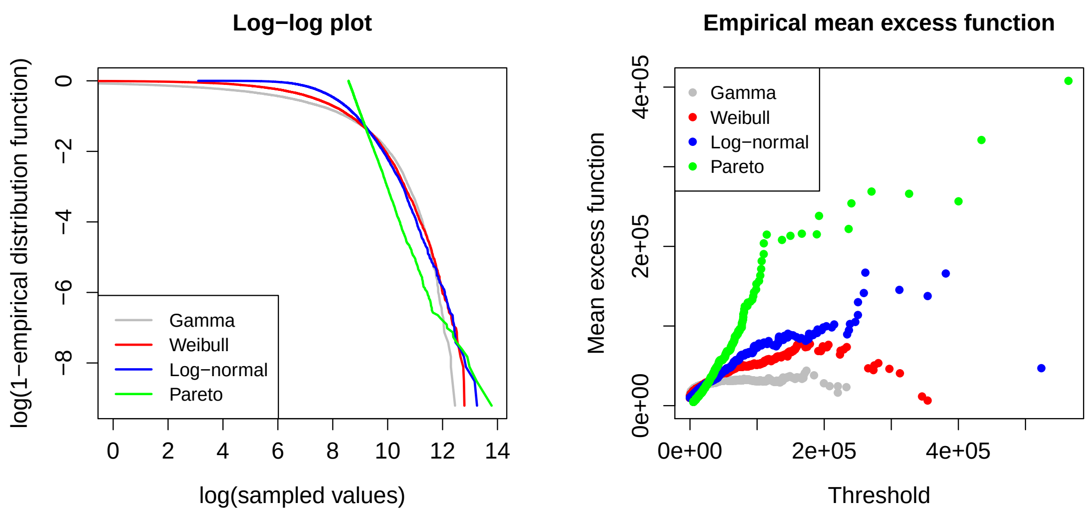

Figure 3: Log-log plots of the four i.i.d. samples (left). Plot of the empirical mean excess functions of the four i.i.d. samples (right).

choose the claim size distribution in order to find the one which suits best in a particular problem at hand.

```
### Size of the i.i.d. sample
n <- 10000

### Generate the gamma i.i.d. sample
gamma <- 1/4
c <- 1/40000
set.seed(100)
gamma.sample <- rgamma(n=n, shape=gamma, rate=c)

### Generate the Weibull i.i.d. sample
tau <- 0.54
c <- 0.000175
set.seed(200)
weibull.sample <- rgamma(n=n, shape=1, rate=1)^(1/tau)/c

### Generate the log-normal i.i.d. sample
mu <- log(2000*sqrt(5))
sigma.squared <- log(5)
set.seed(300)
lognormal.sample <- exp(rnorm(n=n, mean=mu, sd=sqrt(sigma.squared))
    )

### Generate the Pareto i.i.d. sample
theta <- 10000*(sqrt(5)/(2+sqrt(5)))
alpha <- 1+sqrt(5)/2
set.seed(400)
pareto.sample <- theta*exp(rgamma(n=n, shape=1, rate=alpha))

```

```r
### Density plot
ymax <- max(density(log(gamma.sample))$y, density(log(weibull.sample))$y, density(log(lognormal.sample))$y, density(log(pareto.sample))$y)
ymax2 <- max(log(gamma.sample),log(weibull.sample),log(lognormal.sample),log(pareto.sample))
plot(density(log(gamma.sample)), xlim=c(0,ymax2), col="grey", ylim=c(0,ymax), main="Densities", xlab="Sampled values (log scale)", cex.lab=1.25, cex.main=1.25, cex.axis=1.25, lwd=2)
lines(density(log(weibull.sample)), col="red", xlim=c(0,ymax2), lwd=2)
lines(density(log(lognormal.sample)), col="blue", xlim=c(0,ymax2), lwd=2)
lines(density(log(pareto.sample)), col="green", xlim=c(0,ymax2), lwd=2)
legend("topleft", lty=1, lwd=2, col=c("grey","red","blue","green"), legend=c("Gamma","Weibull","Log-normal","Pareto"))


### Boxplot
boxplot(log(gamma.sample), log(weibull.sample), log(lognormal.sample), log(pareto.sample), ylim=c(0,ymax2), col=c("grey","red","blue","green"), main="Box plot", names=c("Gamma","Weibull","Log-normal","Pareto"), xlab="Distribution", ylab="Sampled values (log scale)", cex.lab=1.25, cex.main=1.25, cex.axis=0.95)


### Plot of the empirical distribution function
plot(log(gamma.sample[order(gamma.sample)]), 1:10000/10001, xlim=c(0,ymax2), type="l", col="grey", main="Empirical distribution function", xlab="Sampled values (log scale)", ylab="Empirical distribution function", cex.lab=1.25, cex.main=1.25, cex.axis=1.25, lwd=2)
lines(log(weibull.sample[order(weibull.sample)]), 1:10000/10001, xlim=c(0,ymax2), col="red", lwd=2)
lines(log(lognormal.sample[order(lognormal.sample)]), 1:10000/10001, xlim=c(0,ymax2), col="blue", lwd=2)
lines(log(pareto.sample[order(pareto.sample)]), 1:10000/10001, xlim=c(0,ymax2), col="green", lwd=2)
legend("topleft", lty=1, lwd=2, col=c("grey","red","blue","green"), legend=c("Gamma","Weibull","Log-normal","Pareto"))


### Plot of the empirical loss size index function
plot(1:n/n, cumsum(gamma.sample[order(gamma.sample)])/sum(gamma.sample), type="l", col="grey", main="Empirical loss size index function", xlab="Number of claims (in 100%)", ylab="Empirical loss size index function", cex.lab=1.25, cex.main=1.25, cex.axis=1.25, lwd=2)
lines(1:n/n, cumsum(weibull.sample[order(weibull.sample)])/sum(weibull.sample), type="l", col="red", lwd=2)
```

```r
lines(1:n/n, cumsum(lognormal.sample[order(lognormal.sample)])/sum(lognormal.sample), type="l", col="blue", lwd=2)
lines(1:n/n, cumsum(pareto.sample[order(pareto.sample)])/sum(pareto.sample), type="l", col="green", lwd=2)
legend("topleft", lty=1, lwd=2, col=c("grey","red","blue","green"), legend=c("Gamma","Weibull","Log-normal","Pareto"))


### Log-log plot
plot(log(gamma.sample[order(gamma.sample)]), log(1-1:n/(n+1)), xlim=c(0,ymax2), type="l", col="grey", main="Log-log plot", xlab="log(sampled values)", ylab="log(1-empirical distribution function)", cex.lab=1.25, cex.main=1.25, cex.axis=1.25, lwd=2)
lines(log(weibull.sample[order(weibull.sample)]), log(1-1:n/(n+1)), xlim=c(0,ymax2), type="l", col="red", lwd=2)
lines(log(lognormal.sample[order(lognormal.sample)]), log(1-1:n/(n+1)), xlim=c(0,ymax2), type="l", col="blue", lwd=2)
lines(log(pareto.sample[order(pareto.sample)]), log(1-1:n/(n+1)), xlim=c(0,ymax2), type="l", col="green", lwd=2)
legend("bottomleft", lty=1, lwd=2, col=c("grey","red","blue","green"), legend=c("Gamma","Weibull","Log-normal","Pareto"))


### Plot of the empirical mean excess function
mean.excess.function <- Vectorize(function(threshold,input.sample){
mean(input.sample[input.sample>threshold])-threshold
},"threshold")
xmax <- pareto.sample[order(pareto.sample)][n-1]
ymax3 <- max(pareto.sample)-xmax
plot(gamma.sample[order(gamma.sample)][-n],mean.excess.function(gamma.sample[order(gamma.sample)][-n],gamma.sample), pch=16, col="grey", xlim=c(0,xmax), ylim=c(0,ymax3), main="Empirical mean excess function", xlab="Threshold", ylab="Mean excess function", cex.lab=1.25, cex.main=1.25, cex.axis=1.25)
points(weibull.sample[order(weibull.sample)][-n], mean.excess.function(weibull.sample[order(weibull.sample)][-n],weibull.sample), pch=16, col="red", ylim=c(0,ymax3))
points(lognormal.sample[order(lognormal.sample)][-n], mean.excess.function(lognormal.sample[order(lognormal.sample)][-n],lognormal.sample), pch=16, col="blue", ylim=c(0,ymax3))
points(pareto.sample[order(pareto.sample)][-n], mean.excess.function(pareto.sample[order(pareto.sample)][-n],pareto.sample), pch=16, col="green", ylim=c(0,ymax3))
legend("topleft", pch=16, col=c("grey","red","blue","green"), legend=c("Gamma","Weibull","Log-normal","Pareto"))
```

Solution 5.3 Pareto Distribution

The density g and the distribution function G of Y are given by

$$g(x) = \frac{\alpha}{\theta}\left(\frac{x}{\theta}\right)^{-(\alpha+1)} \qquad \text{and} \qquad G(x) = 1 - \left(\frac{x}{\theta}\right)^{-\alpha},$$

for all $x \geq \theta$.

(a) The survival function $\bar{G} = 1 - G$ of Y is

$$\bar{G}(x) = 1 - G(x) = \left(\frac{x}{\theta}\right)^{-\alpha},$$

for all $x \geq \theta$. Hence, for all $t > 0$ we have

$$\lim_{x\to\infty} \frac{\bar{G}(xt)}{\bar{G}(x)} = \lim_{x\to\infty} \frac{(xt/\theta)^{-\alpha}}{(x/\theta)^{-\alpha}} = t^{-\alpha}.$$

Thus, by definition, the survival function of Y is regularly varying at infinity with tail index α.

(b) Let $\theta \leq u_1 < u_2$. Then, the expected value of Y within the layer $(u_1, u_2]$ can be calculated as

$$\mathbb{E}[Y 1_{\{u_1<Y\leq u_2\}}] = \int_\theta^\infty x 1_{\{u_1<x\leq u_2\}} g(x)\, dx = \int_{u_1}^{u_2} x \frac{\alpha}{\theta} \left(\frac{x}{\theta}\right)^{-(\alpha+1)} dx = \alpha\theta \int_{u_1}^{u_2} \frac{1}{\theta} \left(\frac{x}{\theta}\right)^{-\alpha} dx.$$

In the case $\alpha \neq 1$, we get

$$\mathbb{E}[Y 1_{\{u_1<Y\leq u_2\}}] = \alpha\theta \left[-\frac{1}{\alpha-1}\left(\frac{x}{\theta}\right)^{-\alpha+1}\right]_{u_1}^{u_2} = \theta\frac{\alpha}{\alpha-1}\left[\left(\frac{u_1}{\theta}\right)^{-\alpha+1} - \left(\frac{u_2}{\theta}\right)^{-\alpha+1}\right],$$

and if $\alpha = 1$, we get

$$\mathbb{E}[Y 1_{\{u_1<Y\leq u_2\}}] = \theta \int_{u_1}^{u_2} \frac{1}{x}\, dx = \theta \log\left(\frac{u_2}{u_1}\right).$$

(c) Let $\alpha > 1$ and $y > \theta$. Then, the expected value μ_Y of Y is given by

$$\mu_Y = \theta\frac{\alpha}{\alpha-1}$$

and, similarly as in part (b), we get

$$\mathbb{E}[Y 1_{\{Y\leq y\}}] = \mathbb{E}[Y 1_{\{\theta<Y\leq y\}}] = \theta\frac{\alpha}{\alpha-1}\left[\left(\frac{\theta}{\theta}\right)^{-\alpha+1} - \left(\frac{y}{\theta}\right)^{-\alpha+1}\right] = \mu_Y\left[1 - \left(\frac{y}{\theta}\right)^{-\alpha+1}\right].$$

Hence, for the loss size index function for level $y > \theta$ we have

$$\mathcal{I}[G(y)] = \frac{1}{\mu_Y}\mathbb{E}[Y 1_{\{Y\leq y\}}] = 1 - \left(\frac{y}{\theta}\right)^{-\alpha+1} \in [0,1].$$

(d) Let $\alpha > 1$ and $u > \theta$. The mean excess function of Y above u can be calculated as

$$e(u) = \mathbb{E}[Y - u|Y > u] = \mathbb{E}[Y|Y > u] - u = \frac{\mathbb{E}[Y 1_{\{Y>u\}}]}{\mathbb{P}[Y > u]} - u = \frac{\mathbb{E}[Y 1_{\{Y>u\}}]}{\bar{G}(u)} - u,$$

where for $\mathbb{E}[Y 1_{\{Y>u\}}]$ we have, similarly as in part (b),

$$\mathbb{E}[Y 1_{\{Y>u\}}] = \alpha\theta\left[-\frac{1}{\alpha-1}\left(\frac{x}{\theta}\right)^{-\alpha+1}\right]_u^\infty = \frac{\alpha}{\alpha-1}\theta\left(\frac{u}{\theta}\right)^{-\alpha+1} = \frac{\alpha}{\alpha-1}u\bar{G}(u).$$

Thus, we get

$$e(u) = \frac{\alpha}{\alpha-1}u - u = \frac{1}{\alpha-1}u.$$

Note that the mean excess function $u \mapsto e(u)$ has slope $\frac{1}{\alpha-1} > 0$.

Solution 5.4 Kolmogorov-Smirnov Test

The distribution function G_0 of a Weibull distribution with shape parameter $\tau = \frac{1}{2}$ and scale parameter $c = 1$ is given by

$$G_0(y) = 1 - \exp\{-y^{1/2}\},$$

for all $y \geq 0$. Note that since G_0 is continuous, we are allowed to apply a Kolmogorov-Smirnov test. If $x = (-\log u)^2$ for some $u \in (0, 1)$, we have

$$G_0(x) = 1 - \exp\left\{-\left[(-\log u)^2\right]^{1/2}\right\} = 1 - \exp\{\log u\} = 1 - u.$$

Hence, if we apply G_0 to $x_1, \ldots, x_5$, we get

$$G_0(x_1) = \frac{2}{40}, \quad G_0(x_2) = \frac{3}{40}, \quad G_0(x_3) = \frac{5}{40}, \quad G_0(x_4) = \frac{6}{40}, \quad G_0(x_5) = \frac{30}{40}.$$

Moreover, the empirical distribution function $\widehat{G}_5$ of the sample $x_1, \ldots, x_5$ is given by

$$\widehat{G}_5(y) = \begin{cases} 0 & \text{if } y < x_1, \\ 1/5 & \text{if } x_1 \leq y < x_2, \\ 2/5 & \text{if } x_2 \leq y < x_3, \\ 3/5 & \text{if } x_3 \leq y < x_4, \\ 4/5 & \text{if } x_4 \leq y < x_5, \\ 1 & \text{if } y \geq x_5. \end{cases}$$

The Kolmogorov-Smirnov test statistic D_5 is defined as

$$D_5 = \sup_{y \in \mathbb{R}} \left| \widehat{G}_5(y) - G_0(y) \right|.$$

Since G_0 is continuous and strictly increasing with range $[0, 1)$ and $\widehat{G}_5$ is piecewise constant and attains both the values 0 and 1, it is sufficient to consider the discontinuities of $\widehat{G}_5$ to find D_5. We define

$$f(s-) = \lim_{r \nearrow s} f(r),$$

for all $s \in \mathbb{R}$, where the function f stands for G_0 and $\widehat{G}_5$. Since G_0 is continuous, we have $G_0(s-) = G_0(s)$ for all $s \in \mathbb{R}$. The values of G_0 and $\widehat{G}_5$ and their differences (in absolute value) can be summarized in the following table:

x_i, x_i-	x_1-	x_1	x_2-	x_2	x_3-	x_3	x_4-	x_4	x_5-	x_5
$\widehat{G}_5(\cdot)$	0	8/40	8/40	16/40	16/40	24/40	24/40	32/40	32/40	1
$G_0(\cdot)$	2/40	2/40	3/40	3/40	5/40	5/40	6/40	6/40	30/40	30/40
$\lvert\widehat{G}_5(\cdot) - G_0(\cdot)\rvert$	2/40	6/40	5/40	13/40	11/40	19/40	18/40	26/40	2/40	10/40

Table 1: Values of G_0 and $\widehat{G}_5$, and their differences (in absolute value).

From this table we see that $D_5 = 26/40 = 0.65$. Let $q = 5\%$. By writing $K^{\leftarrow}(1-q)$ for the $(1-q)$-quantile of the Kolmogorov distribution, we have $K^{\leftarrow}(1-q) = 1.36$. Since

$$\frac{K^{\leftarrow}(1-q)}{\sqrt{5}} \approx 0.61 < 0.65 = D_5,$$

we can reject the null hypothesis (at significance level of 5%) of having a Weibull distribution with shape parameter $\tau = \frac{1}{2}$ and scale parameter $c = 1$ as claim size distribution.