ETH Zürich, D-MATH
HS 2017
Prof. Dr. Mario V. Wüthrich

Coordinator
A. Gabrielli

Non-Life Insurance: Mathematics and Statistics

Solution sheet 7

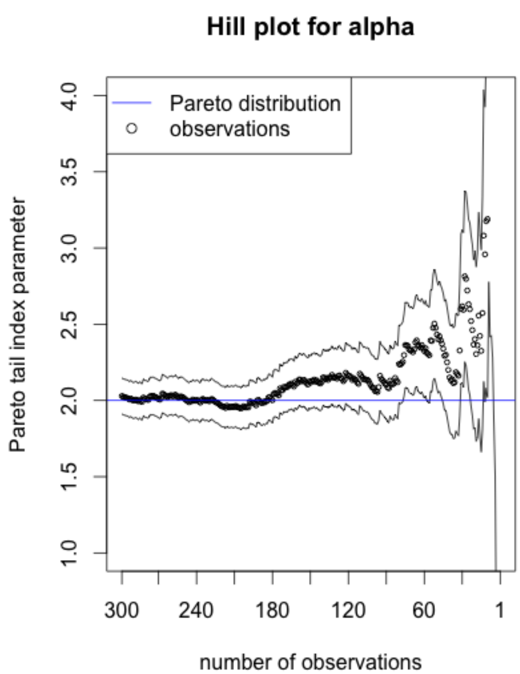
Solution 7.1 Hill Estimator

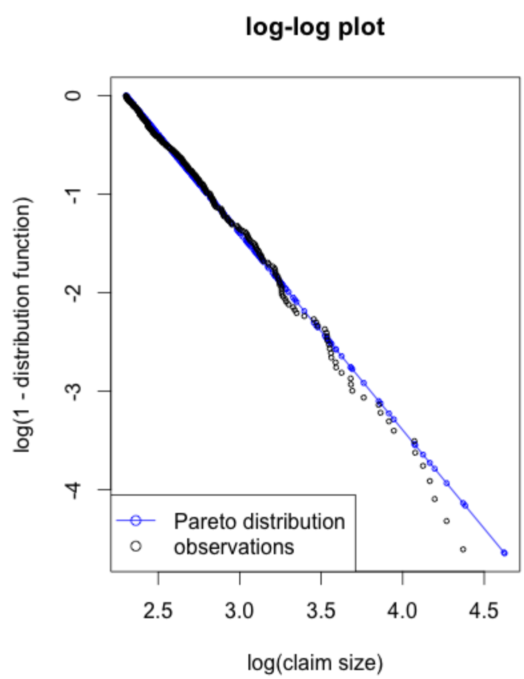
An example of a possible R-code is given below.

```
### Generate 300 independent observations coming from a
### Pareto distribution with threshold theta = 10 millions
### and tail index alpha = 2.
### We use that if Z ~ Gamma(1,alpha),
### then theta*exp{Z} ~ Pareto(theta, alpha).
### Note that for the Gamma distribution we have:
### scale parameter in R = 1/(scale parameter in lecture notes)
n <- 300
theta <- 10   ### in millions
alpha <- 2
set.seed(100)   ### for reproducibility
data.1 <- rgamma(n, shape = 1, scale = 1 / alpha)
data <- theta * exp(data.1)

### Order the data
data.ordered <- data[order(data, decreasing = FALSE)]

### Take the logarithm
log.data.ordered <- log(data.ordered)

### Number of observations
n.obs <- n:1

### Hill estimator
hill.estimator <- ((sum(log.data.ordered)
    - cumsum(log.data.ordered) + log.data.ordered) / n.obs
    - log.data.ordered)^(-1)

### Confidence bounds (see Lemma 3.7 of the lecture notes)
upper.bound <- hill.estimator + sqrt(n.obs^2 / ((n.obs - 1)^2
    * (n.obs - 2)) * hill.estimator^2)
lower.bound <- hill.estimator - sqrt(n.obs^2 / ((n.obs - 1)^2
    * (n.obs - 2)) * hill.estimator^2)

### Hill plot and log-log plot next to each other
par(mfrow=c(1,2))

### Hill plot
plot(hill.estimator, ylim = c(alpha-1,alpha+2), xaxt="n",
    xlab = "number of observations",
    ylab = "Pareto tail index parameter", cex = 0.5)
```

```
title(main = "Hill plot for alpha")
axis(1,at=c(1,seq(from = n / 10, to = n, by = n / 10)),
     c(seq(from = n, to = n / 10, by = -n / 10),1))
lines(upper.bound)
lines(lower.bound)
abline(h = alpha, col = "blue")
legend("topleft", col = c("blue","black"), lty = c(1,NA),
     pch = c(NA,1),
     legend = c("Pareto distribution","observations"))

### True survival function (= 1 - true distribution function)
true.sf <- (data.ordered / theta)^(-alpha)

### Empirical survival function (= 1 - empirical survival function)
empirical.sf <- 1 - (1:n) / n

### Log-log plot
plot(log.data.ordered,log(true.sf), xlab = "log(claim size)",
     ylab = "log(1 - distribution function)",
     cex= 0.5, col = "blue")
title(main = "log-log plot")
lines(log.data.ordered, log(true.sf), col = "blue")
points(log.data.ordered, log(empirical.sf), col = "black",
     cex= 0.5)
legend("bottomleft", col = c("blue","black"), lty = c(1,NA),
     pch = c(1,1), legend = c("Pareto distribution","observations"))
```

The Hill plot (on the left) and the log-log plot (on the right) look as follows:





Note that even though we sampled from a Pareto distribution with tail index $\alpha = 2$, it is not at all clear to see that the data comes from a Pareto distribution. In the Hill plot we see that, first, the estimates of α seem more or less correct, but starting from the 180 largest observations, the plot suggests a higher α or even another distribution. In the log-log plot we see that for small-sized and medium-sized claims the fit seems to be fine. But looking at the largest claims, we would conclude that our data is not as heavy-tailed as a true Pareto distribution with threshold $\theta = 10$ millions and tail index $\alpha = 2$ would suggest. We are confronted with these problems even though we sampled directly from a Pareto distribution. This might indicate the difficulties one faces when trying to fit such a distribution to a real data set, which, to make matters even worse, often contains far less than 300 observations as in this example and moreover the observations may be contaminated by other distributions.

Solution 7.2 Approximations

Note that if $Y \sim \Gamma(\gamma = 100, c = \frac{1}{10})$, then

$$\begin{aligned}
\mathbb{E}[Y] &= \frac{\gamma}{c} = \frac{100}{1/10} = 1'000, \\
\mathbb{E}[Y^2] &= \frac{\gamma(\gamma+1)}{c^2} = \frac{100 \cdot 101}{1/100} = 1'010'000 \quad \text{and} \\
\mathbb{E}[Y^3] &= \frac{\gamma(\gamma+1)(\gamma+2)}{c^3} = \frac{100 \cdot 101 \cdot 102}{1/1000} = 1'030'200'000.
\end{aligned}$$

Let M_Y denote the moment generating function of Y. According to formula (1.3) of the lecture notes, we have

$$M_Y'''(0) = \frac{d^3}{dr^3} M_Y(r) \bigg|_{r=0} = \mathbb{E}[Y^3].$$

For the total claim amount S, we can use Proposition 2.11 of the lecture notes to get

$$\begin{aligned}
\mathbb{E}[S] &= \lambda v \mathbb{E}[Y] = 1'000 \cdot 1'000 = 1'000'000, \\
\mathrm{Var}(S) &= \lambda v \mathbb{E}[Y^2] = 1'000 \cdot 1'010'000 = 1'010'000'000 \quad \text{and} \\
M_S(r) &= \exp\{\lambda v [M_Y(r) - 1]\}.
\end{aligned}$$

In order to get the skewness ς_S of S, which we will need for the translated gamma and the log-normal approximations, we can use the third equation given in the formulas (1.5) of the lecture notes:

$$\varsigma_S \cdot \mathrm{Var}(S)^{3/2} = \frac{d^3}{dr^3} \log M_S(r) \bigg|_{r=0} = \lambda v \frac{d^3}{dr^3} M_Y(r) \bigg|_{r=0} = \lambda v M_Y'''(0) = \lambda v \mathbb{E}[Y^3],$$

from which we can conclude that

$$\varsigma_S = \frac{\lambda v \mathbb{E}[Y^3]}{(\lambda v \mathbb{E}[Y^2])^{3/2}} = \frac{\mathbb{E}[Y^3]}{\sqrt{\lambda v} \mathbb{E}[Y^2]^{3/2}} = \frac{1'030'200'000}{\sqrt{1'000}(1'010'000)^{3/2}} \approx 0.0321.$$

Let F_S denote the distribution function of S. Then, since F_S is continuous and strictly increasing, the quantiles $q_{0.95}$ and $q_{0.99}$ can be calculated as

$$q_{0.95} = F_S^{-1}(0.95) \qquad \text{and} \qquad q_{0.99} = F_S^{-1}(0.99).$$

(a) According to Section 4.1.1 of the lecture notes, the normal approximation is given by

$$F_S(x) \approx \Phi\left(\frac{x - \lambda v \mathbb{E}[Y]}{\sqrt{\lambda v \mathbb{E}[Y^2]}}\right),$$

for all $x \in \mathbb{R}$, where Φ is the standard Gaussian distribution function. For all $\alpha \in (0,1)$, we have

$$\begin{aligned} F_S^{-1}(\alpha) &= \lambda v \mathbb{E}[Y] + \sqrt{\lambda v \mathbb{E}[Y^2]} \cdot \Phi^{-1}(\alpha) \\ &= 1'000 \cdot 1'000 + \sqrt{1'000 \cdot 1'010'000} \cdot \Phi^{-1}(\alpha) \\ &\approx 1'000'000 + 31'780.5 \cdot \Phi^{-1}(\alpha). \end{aligned}$$

In particular, we get

$$q_{0.95} = F_S^{-1}(0.95) \approx 1'000'000 + 31'780.5 \cdot \Phi^{-1}(0.95) \approx 1'000'000 + 31'780.5 \cdot 1.645 = 1'052'279$$

and

$$q_{0.99} = F_S^{-1}(0.99) \approx 1'000'000 + 31'780.5 \cdot \Phi^{-1}(0.99) \approx 1'000'000 + 31'780.5 \cdot 2.325 = 1'073'890.$$

Note that the normal approximation also allows for negative claims S, which under our model assumption is excluded. The probability for negative claims S in the normal approximation can be calculated as

$$F_S(0) \approx \Phi\left(\frac{0 - \lambda v \mathbb{E}[Y]}{\sqrt{\lambda v \mathbb{E}[Y^2]}}\right) \approx \Phi\left(-\frac{1'000'000}{31'780.5}\right) \approx \Phi(-31.5) \approx 4.34 \cdot 10^{-218},$$

which of course is positive, but very close to 0.

(b) According to Section 4.1.2 of the lecture notes, in the translated gamma approximation we model S by the random variable

$$X = k + Z,$$

where $k \in \mathbb{R}$ and $Z \sim \Gamma(\tilde{\gamma}, \tilde{c})$. The three parameters $k, \tilde{\gamma}$ and $\tilde{c}$ can be determined by solving the equations

$$\mathbb{E}[X] = \mathbb{E}[S], \qquad \operatorname{Var}(X) = \operatorname{Var}(S) \qquad \text{and} \qquad \varsigma_X = \varsigma_S, \tag{1}$$

where ς_X is the skewness parameter of X. Since $Z \sim \Gamma(\tilde{\gamma}, \tilde{c})$, we can use the results given in Section 3.2.1 of the lecture notes to calculate

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[k+Z] = k + \mathbb{E}[Z] = k + \frac{\tilde{\gamma}}{\tilde{c}}, \\ \operatorname{Var}(X) &= \operatorname{Var}(k+Z) = \operatorname{Var}(Z) = \frac{\tilde{\gamma}}{\tilde{c}^2} \quad \text{and} \\ \varsigma_X &= \frac{\mathbb{E}\left[(X - \mathbb{E}[X])^3\right]}{\operatorname{Var}(X)^{3/2}} = \frac{\mathbb{E}\left[(k + Z - \mathbb{E}[k+Z])^3\right]}{\operatorname{Var}(k+Z)^{3/2}} = \frac{\mathbb{E}\left[(Z - \mathbb{E}[Z])^3\right]}{\operatorname{Var}(Z)^{3/2}} = \varsigma_Z = \frac{2}{\sqrt{\tilde{\gamma}}}. \end{aligned}$$

Using equations (1), we get

$$\begin{aligned} \frac{2}{\sqrt{\tilde{\gamma}}} = \varsigma_S \quad &\iff \quad \tilde{\gamma} = \frac{4}{\varsigma_S^2} \approx 3'883, \\ \frac{\tilde{\gamma}}{\tilde{c}^2} = \operatorname{Var}(S) \quad &\iff \quad \tilde{c} = \sqrt{\frac{\tilde{\gamma}}{\operatorname{Var}(S)}} \approx 0.002 \quad \text{and} \\ k + \frac{\tilde{\gamma}}{\tilde{c}} = \mathbb{E}[S] \quad &\iff \quad k = \mathbb{E}[S] - \frac{\tilde{\gamma}}{\tilde{c}} = \mathbb{E}[S] - \sqrt{\tilde{\gamma} \operatorname{Var}(S)} \approx -980'392. \end{aligned}$$

If we write F_Z for the distribution function of $Z \sim \Gamma(\tilde{\gamma} \approx 3'883, \tilde{c} \approx 0.002)$, using the translated gamma approximation, we get

$$F_S(x) = \mathbb{P}[S \le x] \approx \mathbb{P}[X \le x] = \mathbb{P}[k + Z \le x] = \mathbb{P}[Z \le x - k] = F_Z(x-k),$$

for all $x \in \mathbb{R}$. Now, for all $\alpha \in (0,1)$, we have

$$F_S^{-1}(\alpha) \approx k + F_Z^{-1}(\alpha)$$

In particular, we get

$$q_{0.95} = F_S^{-1}(0.95) \approx k + F_Z^{-1}(0.95) \approx -980'392 + 2'032'955 = 1'052'563$$

and

$$q_{0.99} = F_S^{-1}(0.99) \approx k + F_Z^{-1}(0.99) \approx -980'392 + 2'055'074 = 1'074'682.$$

Note that since $k < 0$, the translated gamma approximation in this example also allows for negative claims S, which under our model assumption is excluded. The probability for negative claims S can be calculated as

$$F_S(0) \approx F_Z(0-k) \approx F_Z(980'392) \approx 4.87 \cdot 10^{-320},$$

which is basically 0.

(c) According to Section 4.1.2 of the lecture notes, in the translated log-normal approximation we model S by the random variable

$$X = k + Z,$$

where $k \in \mathbb{R}$ and $Z \sim \mathrm{LN}(\mu, \sigma^2)$. Similarly as in part (b), the three parameters k, μ and σ^2 can be determined by solving the equations

$$\mathbb{E}[X] = \mathbb{E}[S], \qquad \mathrm{Var}(X) = \mathrm{Var}(S) \qquad \text{and} \qquad \varsigma_X = \varsigma_S. \tag{2}$$

Since $Z \sim \mathrm{LN}(\mu, \sigma^2)$, we can use the results given in Section 3.2.3 of the lecture notes to calculate

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[k+Z] = k + \mathbb{E}[Z] = k + \exp\left\{\mu + \sigma^2/2\right\}, \\ \mathrm{Var}(X) &= \mathrm{Var}(k+Z) = \mathrm{Var}(Z) = \exp\left\{2\mu + \sigma^2\right\}\left(\exp\left\{\sigma^2\right\} - 1\right) \quad \text{and} \\ \varsigma_X &= \varsigma_Z = \left(\exp\left\{\sigma^2\right\} + 2\right)\left(\exp\left\{\sigma^2\right\} - 1\right)^{1/2}. \end{aligned}$$

Using the third equation in (2), we get

$$\left(\exp\left\{\sigma^2\right\} + 2\right)\left(\exp\left\{\sigma^2\right\} - 1\right)^{1/2} = \varsigma_S \approx 0.0321 \qquad \Longleftrightarrow \qquad \sigma^2 \approx 0.00012,$$

which was found using a computer software. Using the second equation in (2), we get

$$\exp\left\{2\mu + \sigma^2\right\}\left(\exp\left\{\sigma^2\right\} - 1\right) = \mathrm{Var}(S) \iff \mu = \frac{1}{2}\left(\log\left[\left(\exp\left\{\sigma^2\right\} - 1\right)^{-1}\mathrm{Var}(S)\right] - \sigma^2\right),$$

which implies

$$\mu \approx 14.875.$$

Finally, using the first equation in (2), we get

$$k + \exp\left\{\mu + \sigma^2/2\right\} = \mathbb{E}[S] \qquad \Longleftrightarrow \qquad k = \mathbb{E}[S] - \exp\left\{\mu + \sigma^2/2\right\} \approx -2'391'769.$$

If we write F_W for the distribution function of $W = \log Z \sim \mathcal{N}(\mu \approx 14.875, \sigma^2 \approx 0.00012)$, using the translated log-normal approximation, we get

$$F_S(x) = \mathbb{P}[S \le x] \approx \mathbb{P}[X \le x] = \mathbb{P}[k + Z \le x] = \mathbb{P}[\log Z \le \log(x-k)] = F_W(\log[x-k]),$$

for all $x \in \mathbb{R}$. Now, for all $\alpha \in (0,1)$, we have

$$F_S^{-1}(\alpha) \approx k + \exp\{F_W^{-1}(\alpha)\}.$$

In particular, we get

$$q_{0.95} = F_S^{-1}(0.95) \approx k + \exp\{F_W^{-1}(0.95)\} \approx -2'391'769 + 3'444'295 = 1'052'527$$

and

$$q_{0.99} = F_S^{-1}(0.99) \approx k + \exp\{F_W^{-1}(0.99)\} \approx -2'391'769 + 3'466'359 = 1'074'590.$$

Note that since $k < 0$, the translated log-normal approximation in this example also allows for negative claims S, which under our model assumption is excluded. The probability for negative claims S can be calculated as

$$F_S(0) \approx F_Z(0-k) = F_W(\log[-k]) \approx F_W(\log 2'391'769) \approx 1.92 \cdot 10^{-304},$$

which is basically 0.

(d) We observe that with all the three approximations applied in parts (a) - (c) we get almost the same results. In particular, the normal approximation does not provide estimates that deviate significantly from the ones we get using the translated gamma and the translated log-normal approximations. This is due to the fact, that $\lambda v = 1'000$ is large enough and the gamma distribution assumed for the claim sizes is not a heavy tailed distribution. Moreover, the skewness $\varsigma_S = 0.0321$ of S is rather small, hence the normal approximation is a valid model in this example. Note that in all the three approximations we allow for negative claims S, which actually should not be possible under our model assumption. However, the probability of observe a negative claim S is vanishingly small.

Solution 7.3 Akaike Information Criterion and Bayesian Information Criterion

(a) By definition, the MLEs $\left(\hat{\gamma}^{\mathrm{MLE}}, \hat{c}^{\mathrm{MLE}}\right)$ maximize the log-likelihood function $\ell_{\mathbf{Y}}$. In particular, we have

$$\ell_{\mathbf{Y}}\left(\hat{\gamma}^{\mathrm{MLE}}, \hat{c}^{\mathrm{MLE}}\right) \geq \ell_{\mathbf{Y}}(\gamma, c),$$

for all $(\gamma, c) \in \mathbb{R}_+ \times \mathbb{R}_+$.

If we write $d^{(\mathrm{MLE})}$ and $d^{(\mathrm{MM})}$ for the number of estimated parameters in the MLE model and in the method of moments model, respectively, we have $d^{(\mathrm{MLE})} = d^{(\mathrm{MM})} = 2$. The AIC value $\mathrm{AIC}^{(\mathrm{MLE})}$ of the MLE model and the AIC value $\mathrm{AIC}^{(\mathrm{MM})}$ of the method of moments model are then given by

$$\mathrm{AIC}^{(\mathrm{MLE})} = -2\ell_{\mathbf{Y}}\left(\hat{\gamma}^{\mathrm{MLE}}, \hat{c}^{\mathrm{MLE}}\right) + 2d^{(\mathrm{MLE})} = -2 \cdot 1264.013 + 2 \cdot 2 = -2524.026 \quad \text{and}$$
$$\mathrm{AIC}^{(\mathrm{MM})} = -2\ell_{\mathbf{Y}}\left(\hat{\gamma}^{\mathrm{MM}}, \hat{c}^{\mathrm{MM}}\right) + 2d^{(\mathrm{MM})} = -2 \cdot 1264.171 + 2 \cdot 2 = -2524.342.$$

According to the AIC, the model with the smallest AIC value should be preferred. Since $\mathrm{AIC}^{(\mathrm{MLE})} < \mathrm{AIC}^{(\mathrm{MM})}$, we choose the MLE fit.

(b) If we write $d^{(\mathrm{gam})}$ and $d^{(\exp)}$ for the number of estimated parameters in the gamma model and in the exponential model, respectively, we have $d^{(\mathrm{gam})} = 2$ and $d^{(\exp)} = 1$. The AIC value $\mathrm{AIC}^{(\mathrm{gam})}$ of the gamma model and the AIC value $\mathrm{AIC}^{(\exp)}$ of the exponential model are then given by

$$\mathrm{AIC}^{(\mathrm{gam})} = -2\ell_{\mathbf{Y}}^{(\mathrm{gam})}\left(\hat{\gamma}^{\mathrm{MLE}}, \hat{c}^{\mathrm{MLE}}\right) + 2d^{(\mathrm{gam})} = -2 \cdot 1264.013 + 2 \cdot 2 = -2524.026 \quad \text{and}$$
$$\mathrm{AIC}^{(\exp)} = -2\ell_{\mathbf{Y}}^{(\exp)}\left(\hat{c}^{\mathrm{MLE}}\right) + 2d^{(\exp)} = -2 \cdot 1264.169 + 2 \cdot 1 = -2526.338.$$

Since $\mathrm{AIC}^{(\mathrm{gam})} > \mathrm{AIC}^{(\mathrm{exp})}$, we choose the exponential model.

The BIC value $\mathrm{BIC}^{(\mathrm{gam})}$ of the gamma model and the BIC value $\mathrm{BIC}^{(\mathrm{exp})}$ of the exponential model are given by

$$\begin{aligned}\mathrm{BIC}^{(\mathrm{gam})} &= -2\ell_{\mathbf{Y}}^{(\mathrm{gam})}\left(\hat{\gamma}^{\mathrm{MLE}}, \hat{c}^{\mathrm{MLE}}\right) + d^{(\mathrm{gam})} \cdot \log 1000 \\ &= -2 \cdot 1264.013 + 2 \cdot \log 1000 \\ &\approx -2514.21\end{aligned}$$

and

$$\begin{aligned}\mathrm{BIC}^{(\mathrm{exp})} &= -2\ell_{\mathbf{Y}}^{(\mathrm{exp})}\left(\hat{c}^{\mathrm{MLE}}\right) + d^{(\mathrm{exp})} \cdot \log 1000 \\ &= -2 \cdot 1264.169 + \log 1000 \\ &\approx -2521.43.\end{aligned}$$

According to the BIC, the model with the smallest BIC value should be preferred. Since $\mathrm{BIC}^{(\mathrm{gam})} > \mathrm{BIC}^{(\mathrm{exp})}$, we choose the exponential model. Note that the gamma model gives the better in-sample fit than the exponential model. But if we adjust this in-sample fit by the number of parameters used, we conclude that the exponential model probably has the better out-of-sample performance (better predictive power).