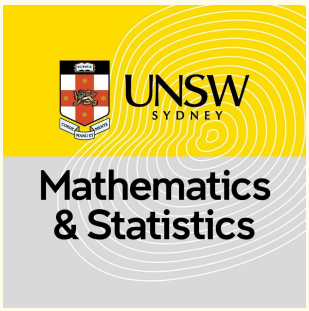


MATH5805
A Modern Introduction to Extreme Value Theory
Term 1, 2025

Dr. Boris Béranger

Week 2 - Lab Solutions

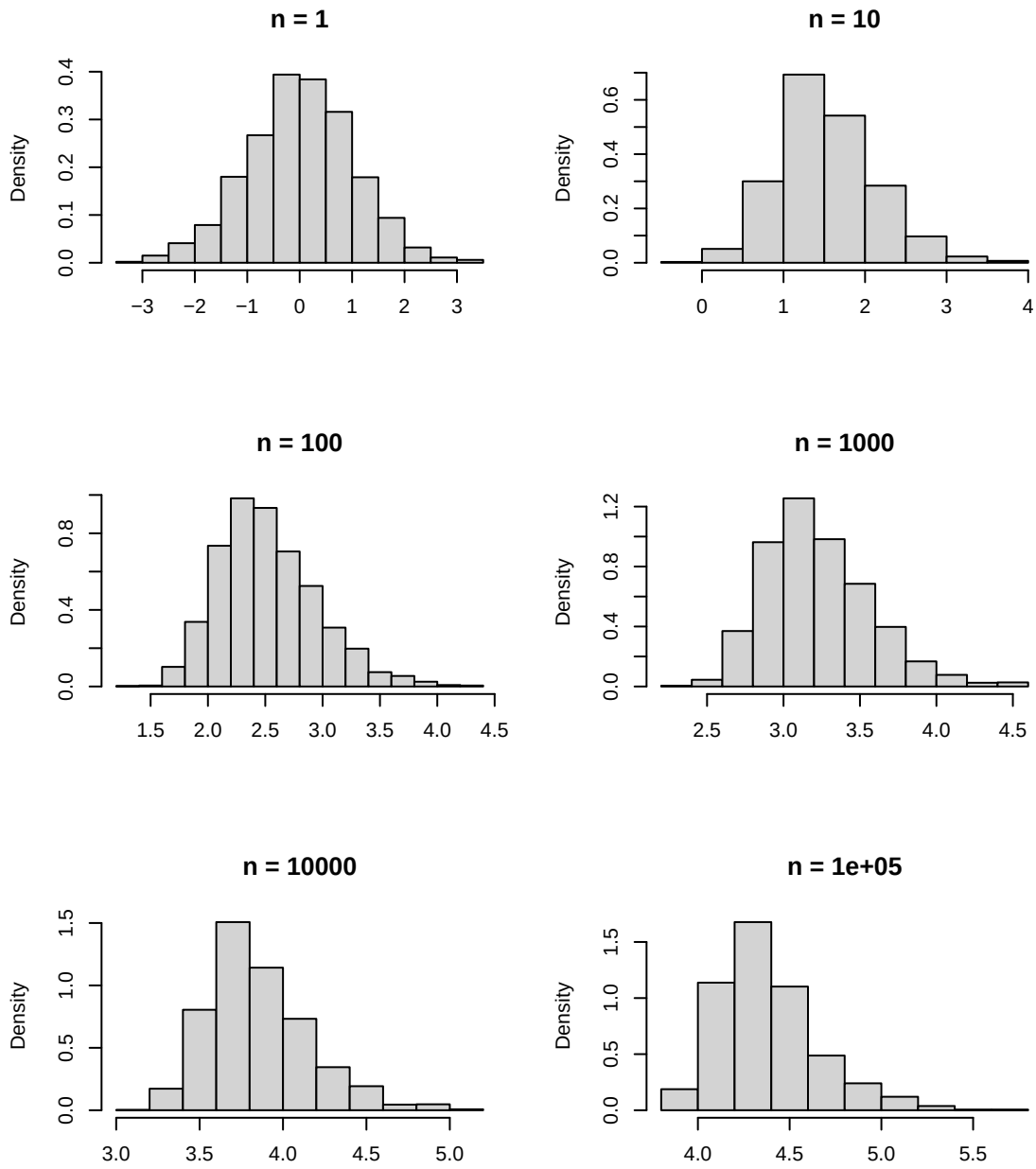
Exercise 1

Suppose that $X_1, \ldots, X_n \sim N(0,1)$.

a) Generate 2,000 samples of $M_n = \max\{X_1, \ldots, X_n\}$ for $n = 1$, 10, 100, 1,000, 10,000, 100,000, and produce histograms of these block maxima. Note the change in location, scale and shape of the distribution as n increases.

Solution:
As n increases, the location is clearly (and intuitively) increasing, the scale appears to be decreasing, and the shape appears to be becoming more right-skew.

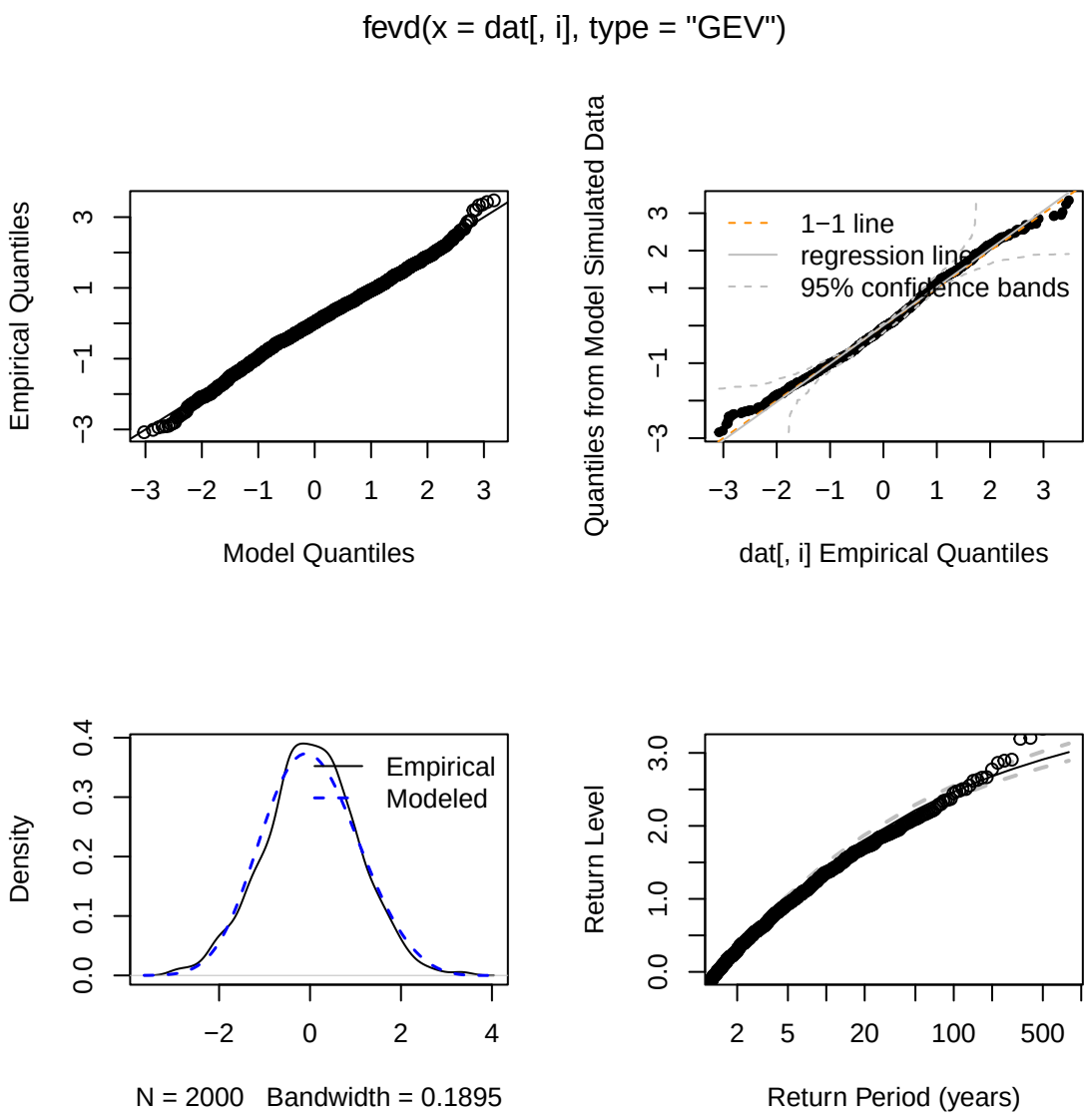
```
nn <- c(1,10,100,1000,10000,100000)
L <- length(nn)
nrep <- 2000
dat <- matrix(NA,ncol=L,nrow=nrep)
par(mfrow = c(3,2))
for (i in 1:L) {
  temp <- matrix(rnorm(nn[i]*nrep),nrow=nrep)
  dat[,i] <- apply(temp,1,max)
  hist(dat[,i], probability=T, xlab="", main=paste("n =",nn[i]))
}
```



b) For each n fit the GEV distribution to the 2,000 block maxima. Explore how the distribution of the block maxima approaches the GEV limit as $n \to \infty$ by a) superimposing the fitted density on the histograms b) examining goodness of fit diagnostics (such as qq-plots) and c) producing profile likelihood confidence intervals for ξ (what value should ξ approach?). How fast does the GEV approximation become a good one?

Solution:

```
library(extRemes)
i <- 1 # Change this manually i=1, 2, 3, 4, 5, 6
fit <- fevd(dat[,i], type="GEV")
plot(fit)
```




```
ci(fit, alpha=0.05, type="parameter", which.par=3)
```

```
## fevd(x = dat[, i], type = "GEV")
##
## [1] "Normal Approx."
##
## [1] "shape: -0.242"
##
## [1] "95% Confidence Interval: (-0.2601, -0.2247)"
```

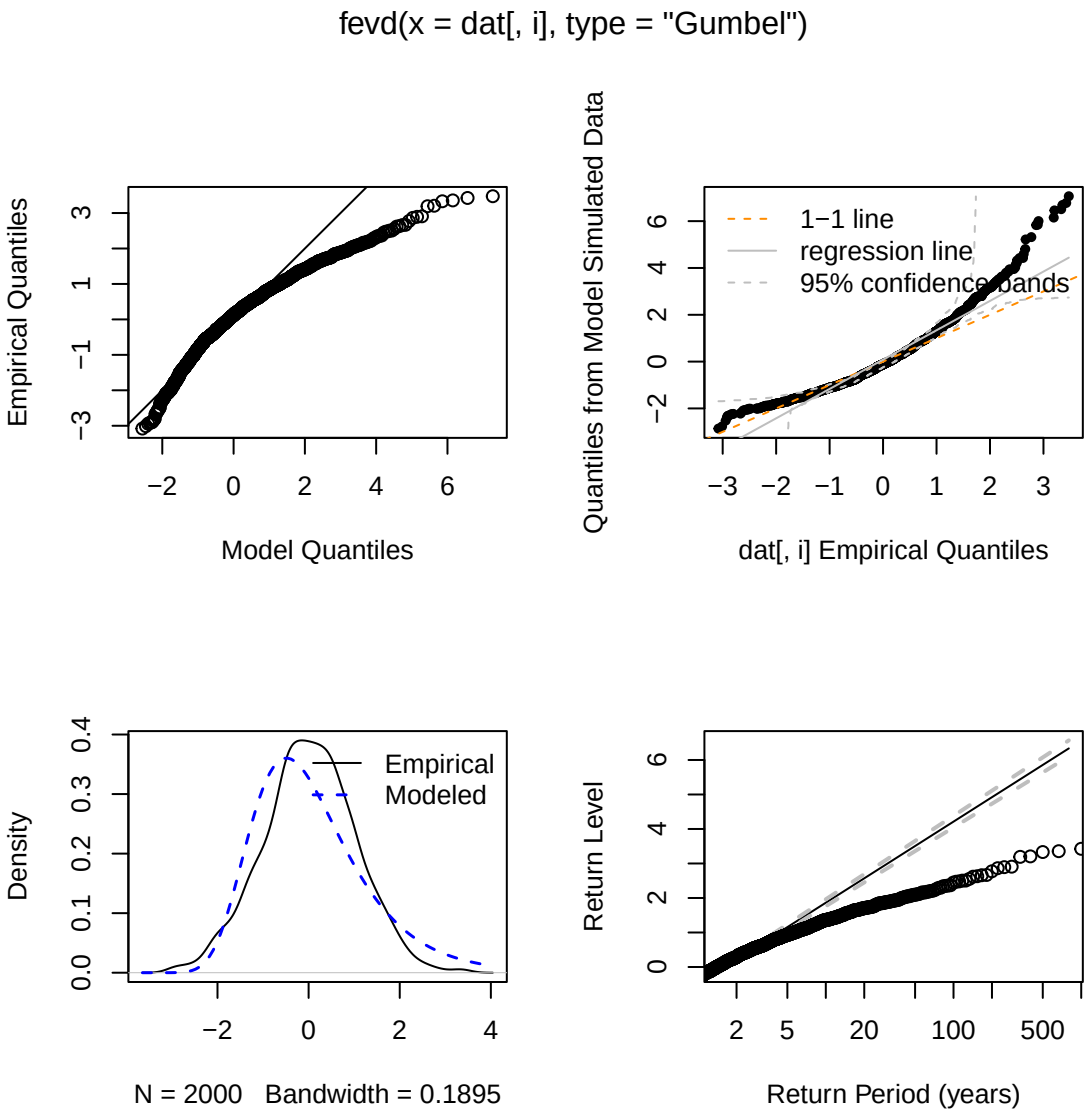
Visually each of these fits (with $n = 1$, and $n = 100,000$) don't appear to be unreasonable. But it is different when looking at the confidence intervals.In this case, as we know that the normal distribution is in the domain of attraction of the Gumbel distribution, we know that $\xi \to 0$ as $n \to \infty$. From the above 95% confidence intervals it seems clear that convergence is in fact slow, and we do not appear to be there (statistically) even for $n = 100,000$.

c) The normal distribution is in the domain of attraction of the Gumbel distribution. Repeat question 2 but fitting the Gumbel distribution instead of the GEV. Contrast your results

with the GEV fit. Is there an advantage in knowing that the domain of attraction is Gumbel, compared to not knowing this?

Solution:

```
i <- 1 # Change this manually i=1, 2, 3, 4, 5, 6
fit <- fevd(dat[,i], type="Gumbel")
plot(fit)
```



At first glance it seems that the fit is far worse with the Gumbel than the GEV, especially for smaller n. Of course this is perhaps not surprising as the GEV is a more flexible model. So what we are seeing is the GEV using the extra parameter ξ (which we *know* to be 0) to account for the actual lack of convergence of the block maxima to the Gumbel. Fixing $\xi = 0$, and comparing to the correct limit, makes this lack of convergence actually more apparent. So which model (GEV or Gumbel) should we use here for finite n? In principle, if we know that the Gumbel is the limiting distribution, then by fixing $\xi = 0$ and estimating one less parameter, we should be doing better statistically. (i.e. better parameter estimates.)

However, in the present case, because convergence to the Gumbel is very (very!) slow for the normal, we are barely in a position that the Gumbel is a good approximation for any of these sample sizes. When $n = 100,000$ the approximation is acceptable, however for this sample size, the gains obtained by estimating one less parameter are vanishingly small. For smaller n we actually do better by fitting the GEV and model the data *statistically* rather than *mathematically*!

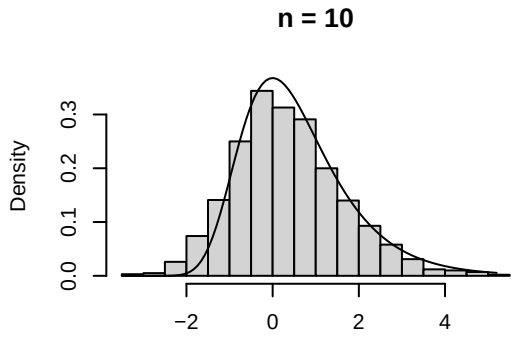
d) Calculate the rescaled maxima $M_n^* = (M_n - b_n)/a_n$ where

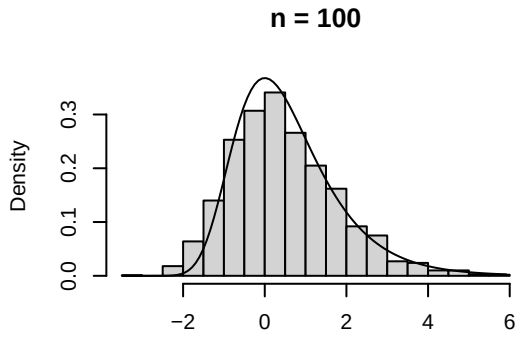
$$\begin{aligned} a_n &= (2\log n)^{-1/2} \\ b_n &= (2\log n)^{1/2} - 1/2(\log\log n + \log 4\pi)/(2\log n)^{1/2}. \end{aligned}$$

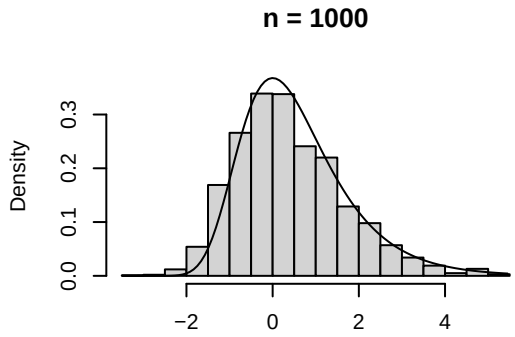
Show that the location, scale and shape of M_n^* stabilises as $n \to \infty$ and that the distribution approaches a standardised GEV(0,1,0).

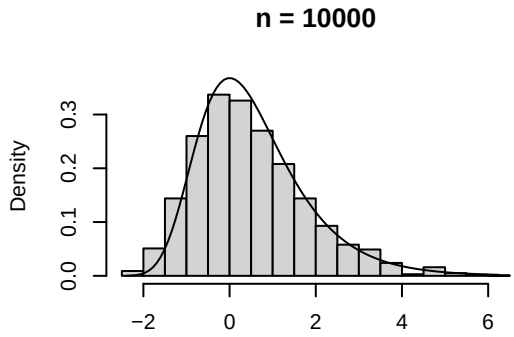
Solution:

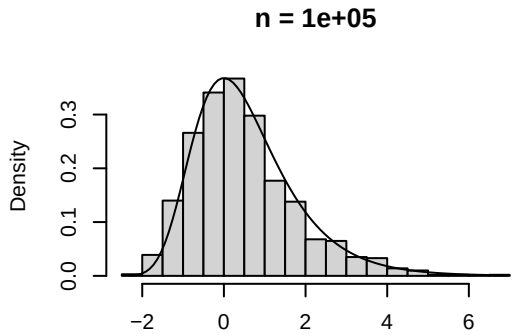
```
a <- (2*log(nn))^(-0.5)
b <- (2*log(nn))^0.5-0.5*(log(log(nn))+log(4*pi))/(2*log(nn))^0.5
par(mfrow=c(3,2))
for (i in 2:L) {
  temp <- (dat[,i]-b[i])/a[i]
  hist(temp, probability=T, xlab="", main=paste("n =",nn[i]), nclass=21,
       ylim=c(0,0.37))
  xx <- seq(min(temp), max(temp), length=100)
  lines(xx,devd(xx,type="GEV"))
}
```











From the results, the norming constants appear to be correct in that the location and scale of the maxima seem to be stable. The convergence of the normalised maxima to the standard GEV appears to be improving with increasing n.

Exercise 2

Repeat question a), b) & d) of Exercise 1 but with $X_1, \ldots, X_n \sim U(0,1)$, noting that:

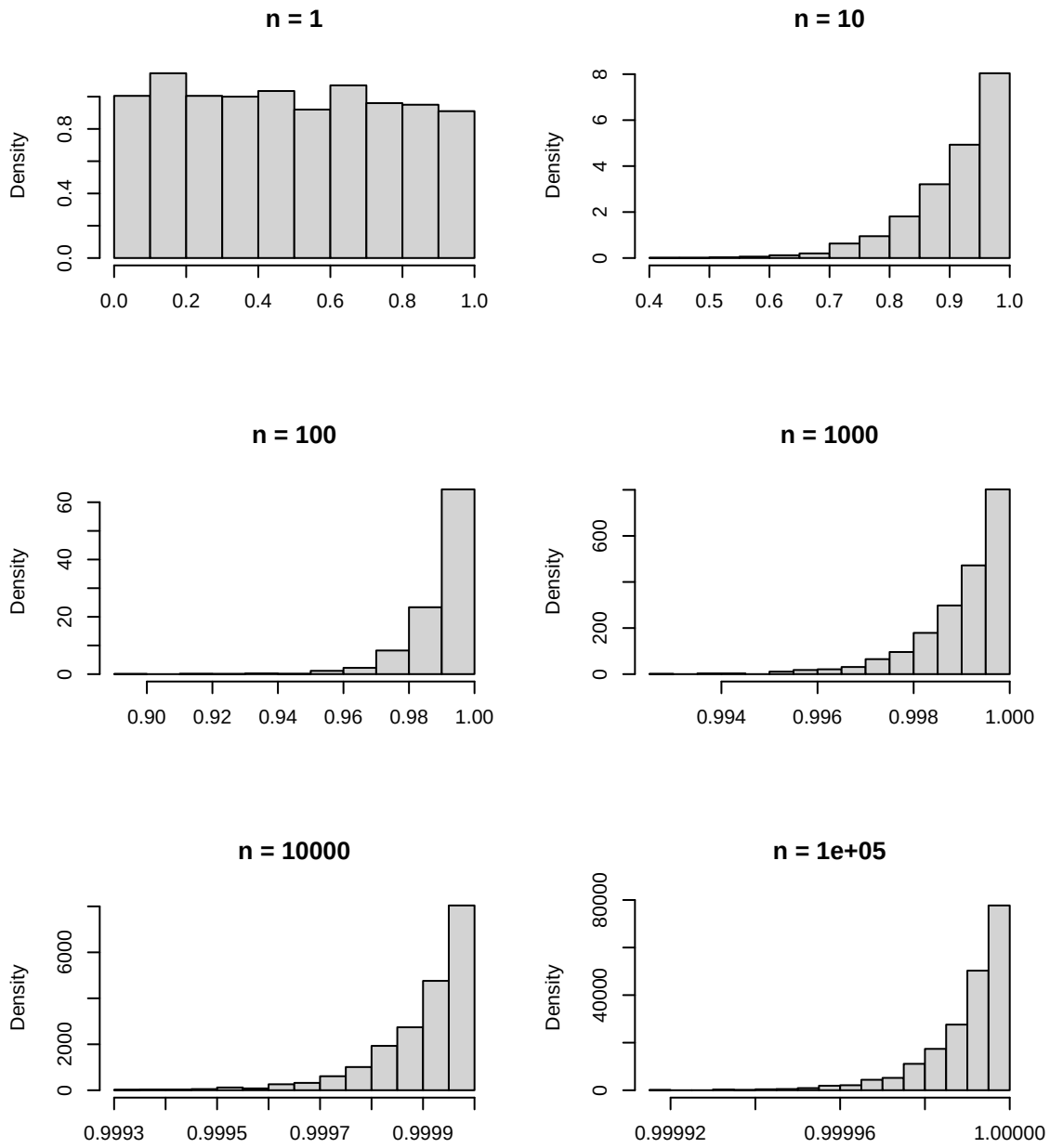
- For question b), the uniform distribution is in the domain of attraction of the Weibull distribution.
- For question d) use the norming constants

$$a_n = 1/n \qquad \text{and} \qquad b_n = 1$$

to show that the distribution of M_n^* approaches a standardised Weibull distribution (= GEV$(-1, 1, -1)$).

a) **Solution:**
Using similar code to previously:

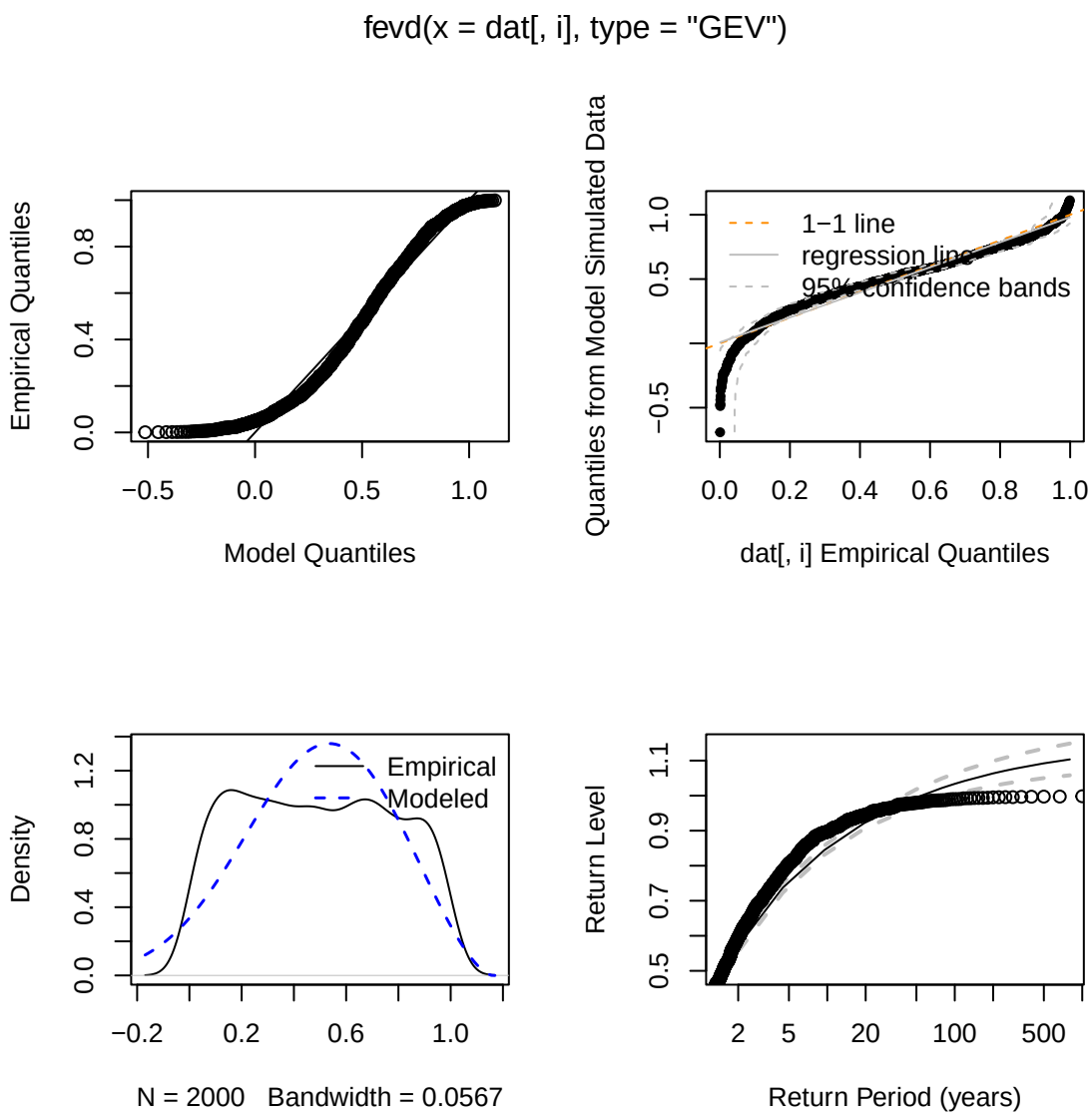
```
nn <- c(1,10,100,1000,10000,100000)
L <- length(nn)
nrep <- 2000
dat <- matrix(NA, ncol=L, nrow=nrep)
par(mfrow=c(3,2))
for (i in 1:L) {
  temp <- matrix(runif(nn[i]*nrep), nrow=nrep)
  dat[,i] <- apply(temp, 1, max)
  hist(dat[,i], probability=T, xlab="", main=paste("n =",nn[i]))
}
```



As n increases, the distribution of sample maxima is becoming increasingly pushed against the upper endpoint of the domain of the uniform distribution.

b) **Solution:**

```
library(extRemes)
i <- 1 # Change this manually i=1, 2, 3, 4, 5, 6
fit <- fevd(dat[,i], type="GEV")
plot(fit)
```



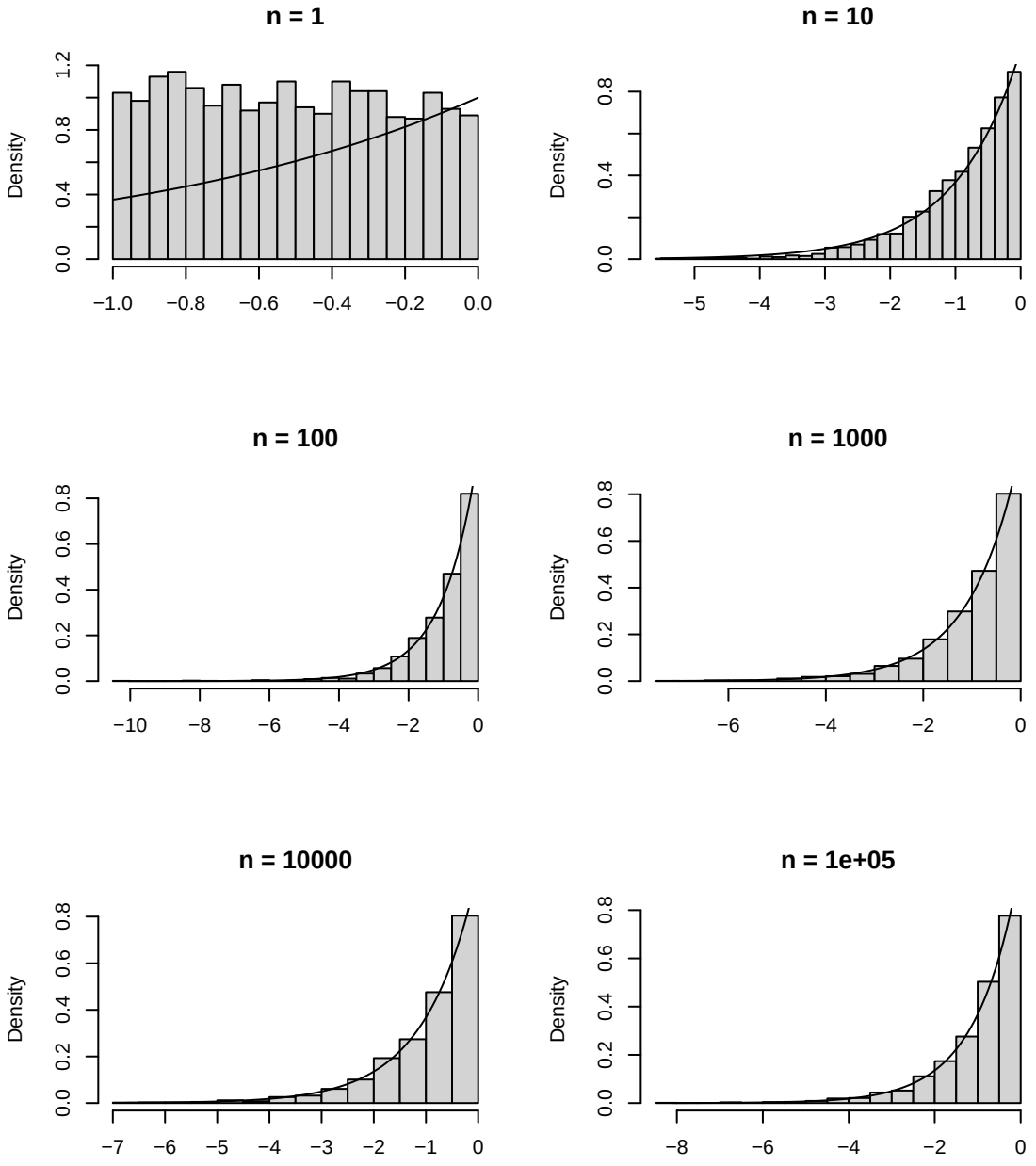

```
ci(fit, alpha=0.05, type="parameter", which.par=3)
```

```
## fevd(x = dat[, i], type = "GEV")
##
## [1] "Normal Approx."
##
## [1] "shape: -0.391"
##
## [1] "95% Confidence Interval: (-0.4413, -0.3399)"
```

For moderate n, the block maxima appears to be rapidly well approximated by the GEV. For the estimates of ξ, there seem to be some issue with the CI: the point estimates are clearly negative and $\to -1$ as $n \to \infty$ indicating a clear Weibull limit.

c) **Solution:**

```r
a <- 1/nn
b <- rep(1,L)
par(mfrow=c(3,2))
for(i in 1:L) {
  temp <- (dat[,i]-b[i])/a[i]
  hist(temp, probability=T, xlab="", main=paste("n =",nn[i]), nclass=21)
  xx <- seq(min(temp), max(temp), length=100)
  lines(xx, evd::dgev(xx, loc=-1, scale=1, shape=-1))
}
```



Clearly the normalised maxima are rapidly converging to a GEV$(-1, 1, -1)$ (standard Weibull) distribution.

Exercise 3

Repeat question a), b) & d) of Exercise 1 but with $X_1, \ldots, X_n \sim$ log-Gamma$(1, 1)$, noting that:

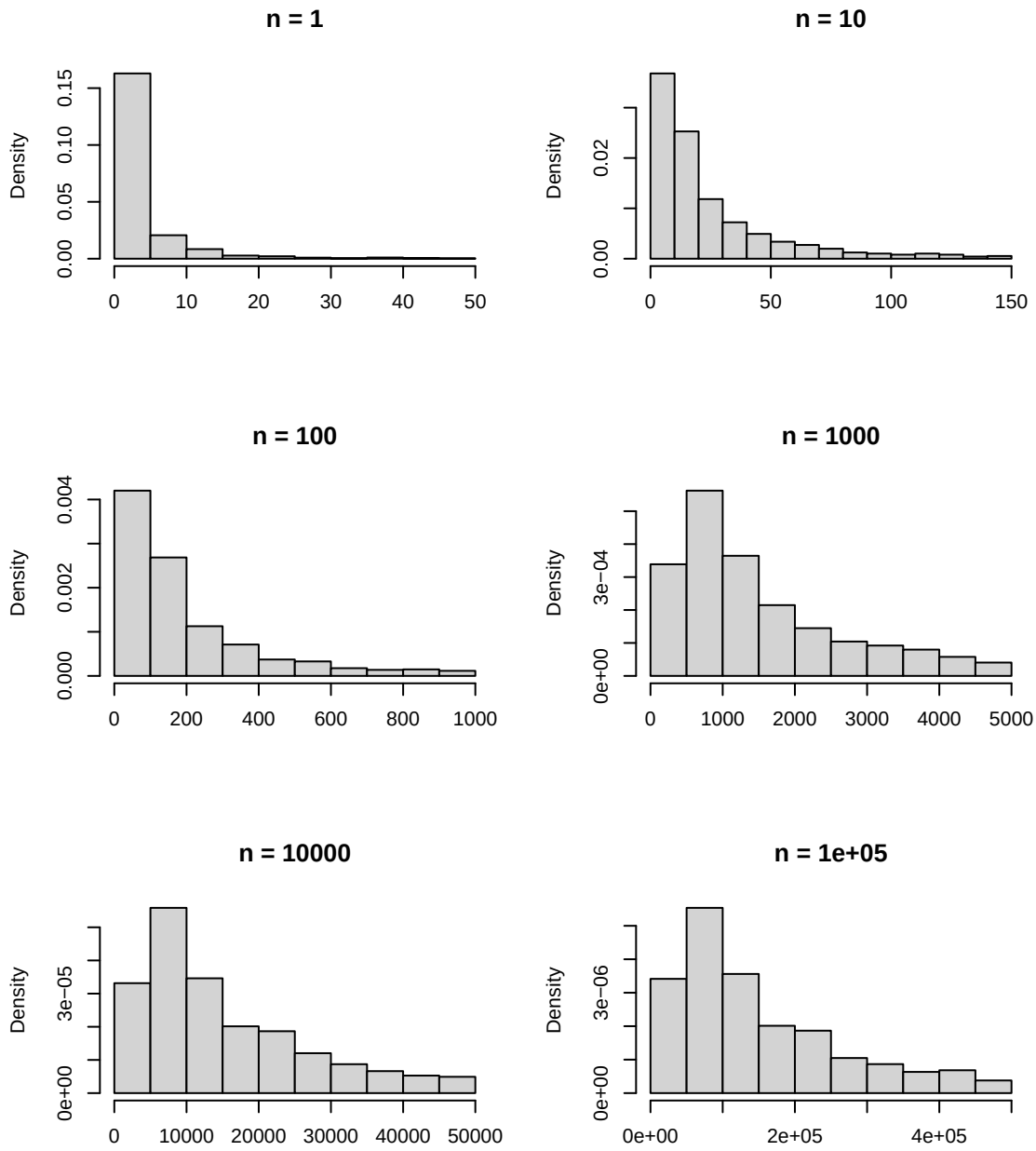
- For question a), to generate X_i note that if $Y \sim$ Gamma(a, b) then $Z = \exp Y \sim$ log-Gamma(a, b). Also note that as the log-Gamma distribution is heavy tailed, you may wish to restrict the x-axis before producing the histogram. You might use the upper end points 50, 150, 1000, 5000, 50000, 500000 for the n given above.
- For question b), the log-Gamma distribution is in the domain of attraction of the Fréchet distribution.
- For question d) use the norming constants

$$a_n = n \qquad \text{and} \qquad b_n = 0$$

to show that the distribution of M_n^* approaches a standardised Fréchet distribution (= GEV$(1, 1, 1)$). Note: You may wish to restrict the upper domain of the histograms to 50 for your plots.

a) **Solution:**
Using similar code to previously:

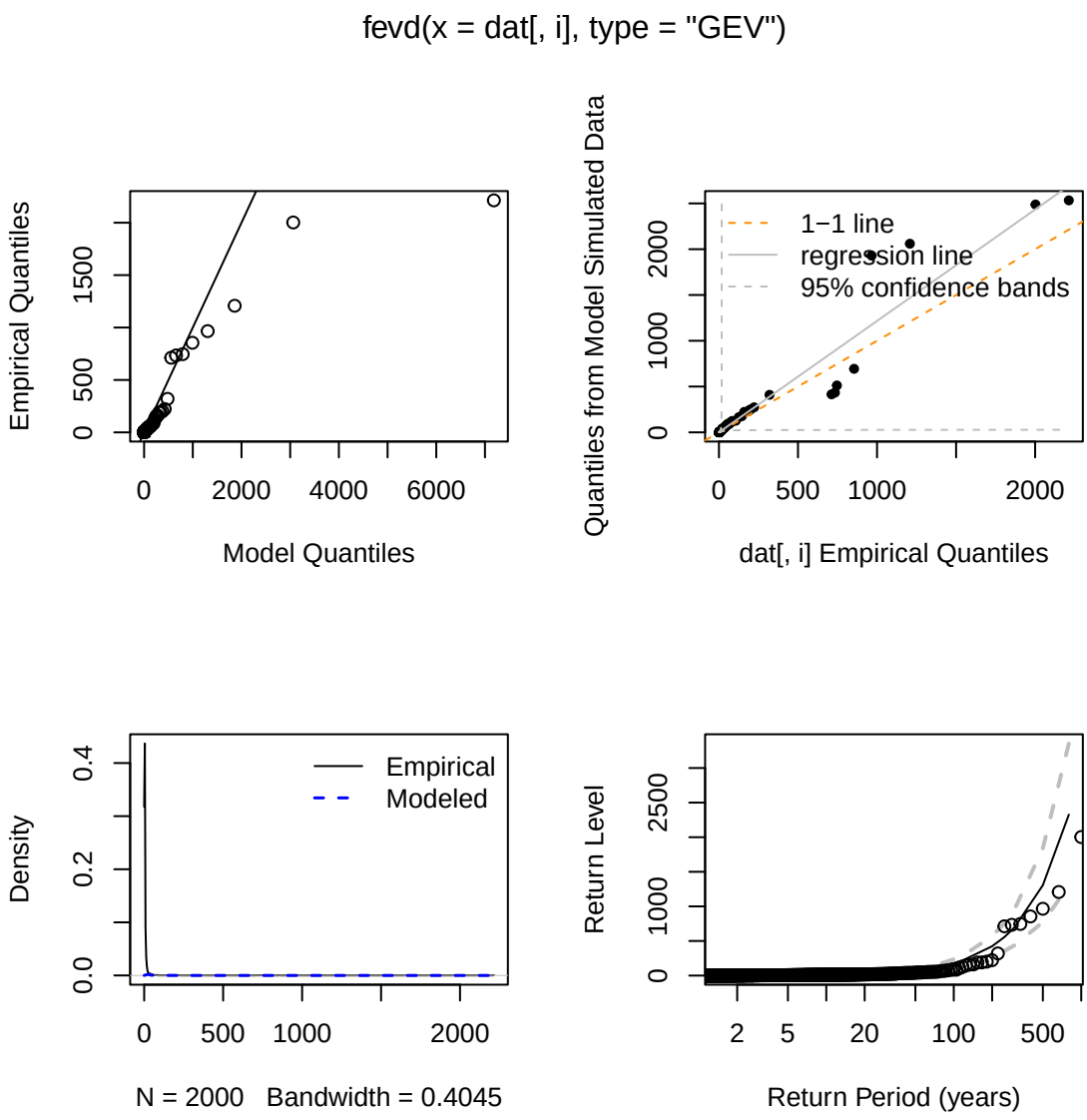
```
nn <- c(1,10,100,1000,10000,100000)
L <- length(nn)
nrep <- 2000
dat <- matrix(NA,ncol=L,nrow=nrep)
aa <- 1; bb <- 1
llim <- c(50, 150, 1000, 5000, 50000, 500000)
par(mfrow=c(3,2))
for (i in 1:L) {
  temp <- matrix(exp(rgamma(nn[i]*nrep, aa, bb)), nrow=nrep)
  dat[,i] <- apply(temp, 1, max)
  hist(dat[,i][dat[,i] < llim[i]], probability=T, xlab="", main=paste("n =",nn[i]))
}
```



These are heavy tailed distributions, so there should be some very large values generated in the upper distributional tails. While the scale of the distribution is drastically increasing with n, the shape appears to be stabilising.

b) **Solution:**

```
library(extRemes)
i <- 1 # Change this manually i=1, 2, 3, 4, 5, 6
fit <- fevd(dat[,i], type="GEV")
plot(fit)
```

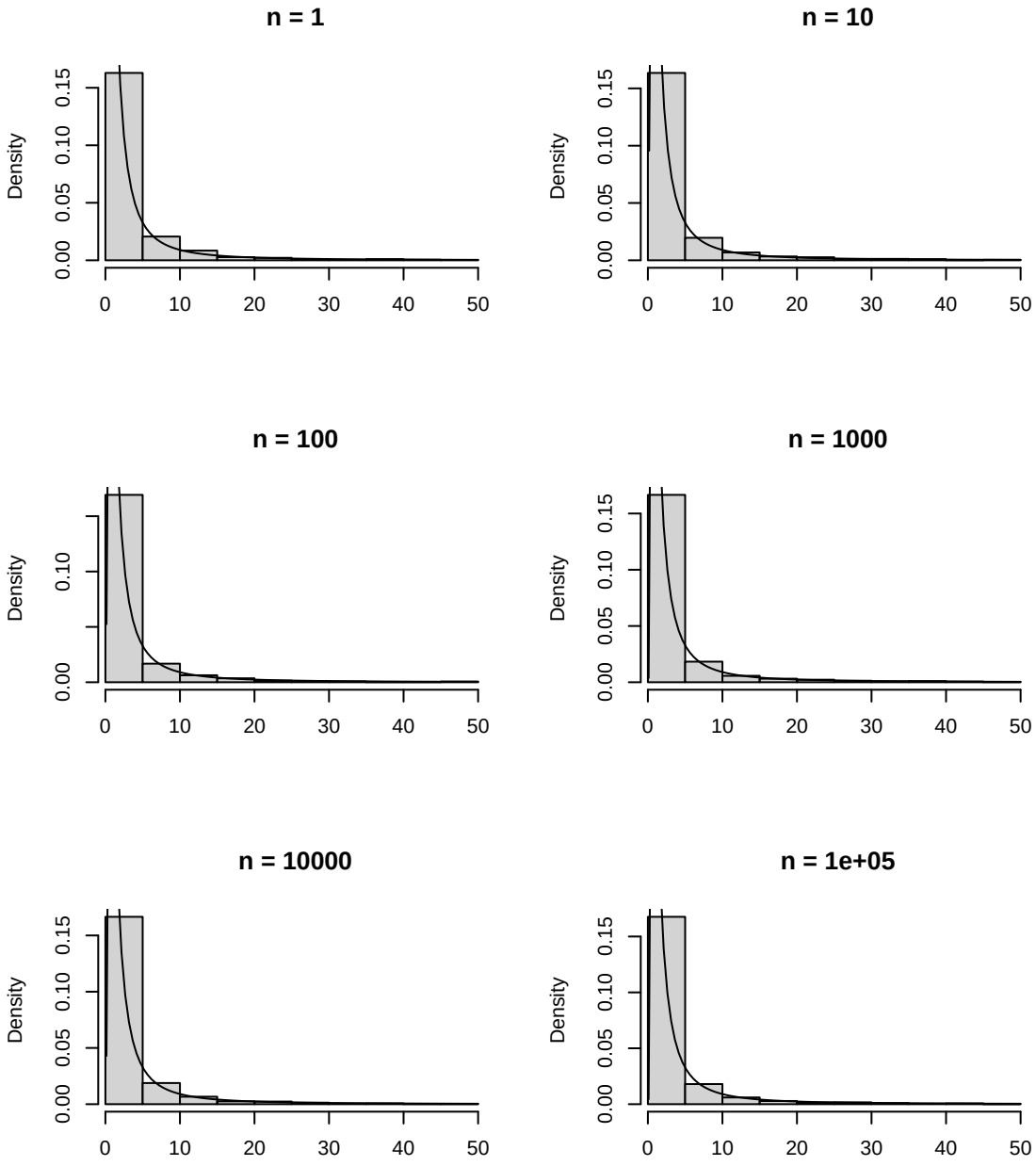


It is difficult to see what is happening here as there is a large amount of variability in the data generated from a distribution with a heavy tail. The estimates of ξ for $i = 1, \ldots, 6$ indicate that we are clearly in the Fréchet domain of attraction. Let's see what happens when we normalise and restrict the domain of the plots.

c) **Solution:**

```
a <- (1/gamma(bb)*nn*(log(nn))^(bb-1))^{1/aa}
b <- rep(0,L)
par(mfrow=c(3,2))
for(i in 1:L) {
  temp <- (dat[,i]-b[i])/a[i]
  hist(temp[temp<50], probability=T, xlab="", main=paste("n =",nn[i]),
       xlim=c(1,50))
  xx <- seq(min(temp), 50, length=100)
  lines(xx, evd::dgev(xx, loc=1, scale=1, shape=1))
```

```
}
```



The normalisation has clearly stabilised the distribution of block maxima to the standardised Fréchet distribution (though there are still some heavy tailed realisations).