Solution 1

(a) The following code produces the plots for normal samples of sizes 10, 20 and 50, shown in Figure 1.

```
par(mfrow=c(3,3),pty="s")
tp() # without line
tp(line=TRUE) # with line
```

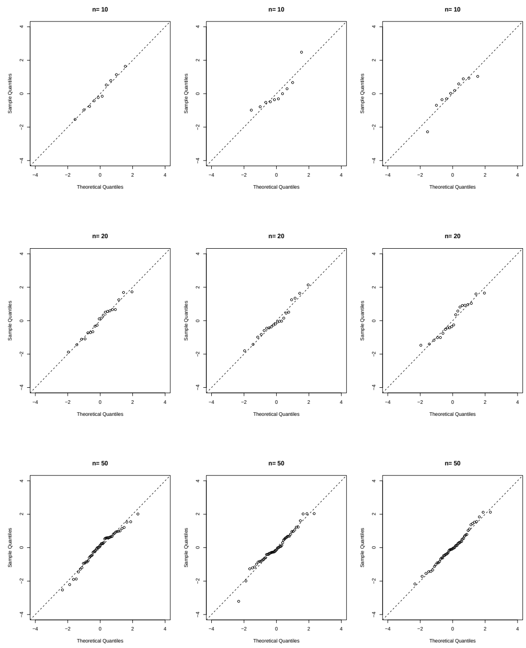

Figure 1: Normal QQ-plots for (a)

As the same size grows, we see a better alignment between the sample and theoretical quantiles along the diagonal.

(b) For the gamma distribution with shape parameter 4 and from the t distribution with 5 degrees of freedom, the code below gives Figure 2.

```
tp(ran.gen=rgamma,shape=4,line=TRUE)
tp(ran.gen=rt,df=5,line=TRUE)
```

In Figure 2 we notice a difference between the sample and theoretical quantiles. The Gamma distribution is right-skewed, and the $t-$distribution is characterised by heavy tails. The differences become more apparent for large sample sizes.

(c) Figure 3 contains the quantile plots for the different distributions. The correct choices are rounded, light-tailed, with outliers, skewed, light-tailed, rounded, with outliers, skewed, and with outliers. In most cases, we notice that, if the distribution is skewed, has outliers or heavy tails, or is rounded, the difference with the normal quantiles is more noticeable.

(d) Using `ggplot` we obtain Figure 4. We observe that precipitation quantities are mostly below 10mm. Precipitations between 10mm and 20mm are less frequent, and those above 30mm are rather few. For precipitations exceeding 20mm we also notice an increasing trend, which may be a reflection of the change in climate over time.

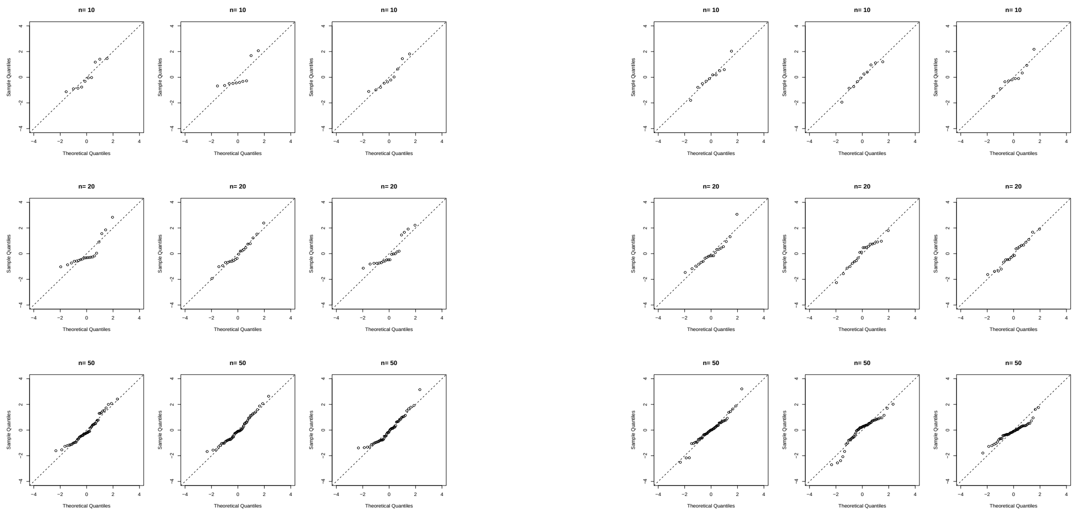

Figure 2: Normal QQ-plots for the Gamma distribution (left) and $t-$ distribution (right)

(e) The following function, modified to include also the yearly maxima, gives the quantile plots in Figure 5. Here we observe differences between the monthly and yearly maxima. In particular, the Gumbel quantiles provide a better description of the yearly maxima, whereas the monthly maxima is affected by the presence of small precipitation quantities .

```
plot_abisko2 <- function (){
abisko.max <- matrix(NA, 102, 12)
year <- c(1913:2014)
for (i in 1:102) for (j in 1:12)
{ k <- (year(abisko$date)-1912==i & month(abisko$date)==j)
abisko.max[i,j] <- max(abisko$precip[k]) }
mon.max <- c(abisko.max)
mon.n <- length(mon.max)
 year.max <- apply(abisko.max,1,max)
year.n <- length(year.max)
plot1 <- ggplot()+
geom_point(aes(x=qgumbel(c(1:mon.n)/(mon.n+1)),
 y=sort(mon.max)), pch=16,cex=0.7)+
geom_point(aes(x=qgumbel(c(1:year.n)/(year.n+1)),
 y=sort(year.max)), col="blue", pch=16, cex=0.7)+
labs(y="Ordered maxima (mm)", x="Gumbel plotting positions")+
theme_classic(base_size=11)+
theme(axis.text = element_text(size = 10),
panel.background = element_rect(fill = "white",
colour = "white",
size = 0.5, linetype = "blank"))

pl <- cowplot::plot_grid(plotlist = list(plot1),
labels = c(""),
ncol = 1)
ggsave(filename = "figures/abisko2.png", plot = pl,
bg = "white", width = 900, height = 900, unit = 'px', dpi=250)
}
```

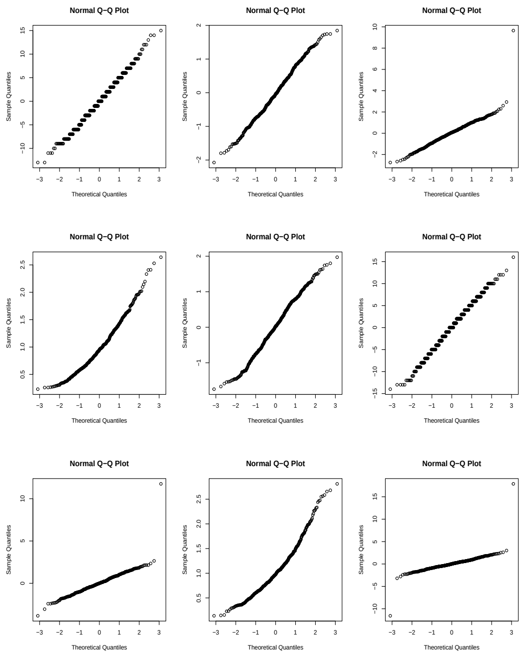

Figure 3: Normal QQ-plots for (c)

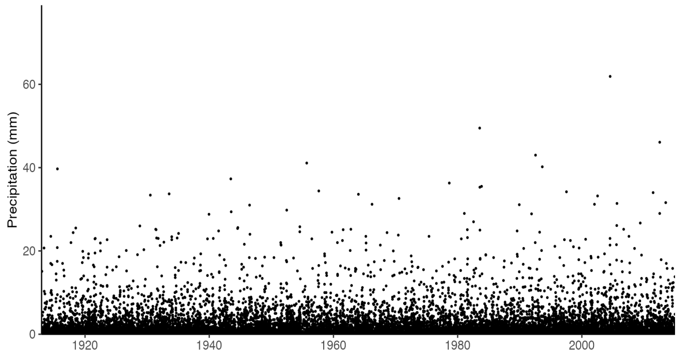

Figure 4: Precipitation in Abisko for the years 1913–2014.

Solution 2

(a) The following code gives the quantile plots in Figure 6, where we again notice a better alignment between the sample and theoretical quantiles for a larger sample size.

```
qqexp<-function(n){
qqplot(qexp(c(1:n)/(n+1)), rexp(n),
 ylab="Sample quantiles",
 xlab="Theoretical quantiles")
}

par(mfrow=c(1,2))
qqexp(50)
qqexp(100)
```

(b) Note that, for $x > 0$, $X_1 > x$ if and only if $N(x) = 0$, so we may write

$$\mathrm{P}(X_1 > x) = \mathrm{P}(N(x) = 0) = \exp(-\lambda x).$$

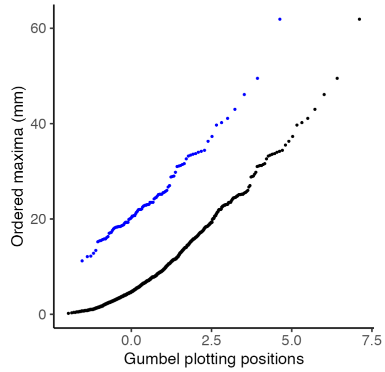

Figure 5: Gumbel quantile plots of the monthly (black) and yearly (blue) maxima.

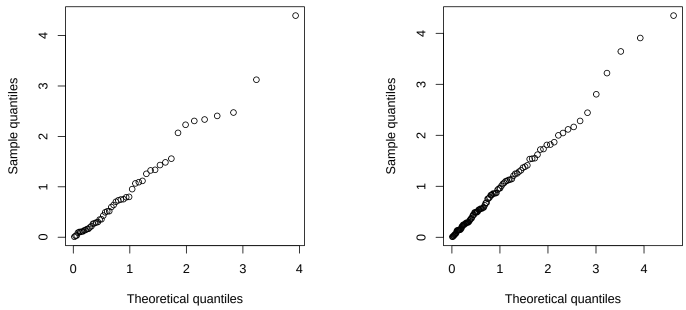

Figure 6: Quantile plots for the exponential distribution, with $n = 50$ (left) and $n = 100$ (right).

More generally, for the k-th interval, we have

$$\begin{aligned} P(X_k > x) &= P(X_k > x | S_{k-1} = s_{k-1}) \\ &= P(N(s_{k-1} + x) - N(s_{k-1}) = 0) \\ &= P(\text{no new events in } (s_{k-1}, s_{k-1} + x]) \\ &= P(N(s_{k-1}, s_{k-1} + x] = 0) = \\ &= \exp(-\mu((s_{k-1}, s_{k-1} + x])) = \\ &= \exp(-(\lambda(s_{k-1} + x) - \lambda(s_{k-1}))) \\ &= \exp(-\lambda x), \end{aligned}$$

where we have used independence of the intervals in the first line. The second line uses the relation between the Poisson process N and S_k, and follows upon noting that, given $S_{k-1} = s_{k-1}$, the event $X_k > x$ occurs if and only if $N(s_{k-1} + x) - N(s_{k-1}) = 0$. Finally, the fifth line uses the intensity function of the Poisson process (see slide 32), i.e., that $\dot{\mu} = \lambda$.

(c) Use of the probability integral transform shows that, for $y > 0$

$$P(Y \le y) = P(1/X \le y) = P(X \ge 1/y) = \exp(-\lambda y^{-1}),$$

which is known as the *Fréchet* distribution with shape parameter $\alpha = 1$, scale λ, and location $m = 0$.

The following code produces Figure 7. As these are heavy-tailed random variables, we see that the visual assessment of the quantiles is more challenging, the reason being that the maxima (or extremes) can take values that are much larger relative to the bulk of the distribution.

```
qqinvexp<-function(n){
qqplot(1/qexp(c(1:n)/(n+1)), 1/rexp(n),
```

```
 ylab="Sample quantiles",
 xlab="Theoretical quantiles")
}

par(mfrow=c(1,2))
qqinvexp(50)
qqinvexp(100)
```

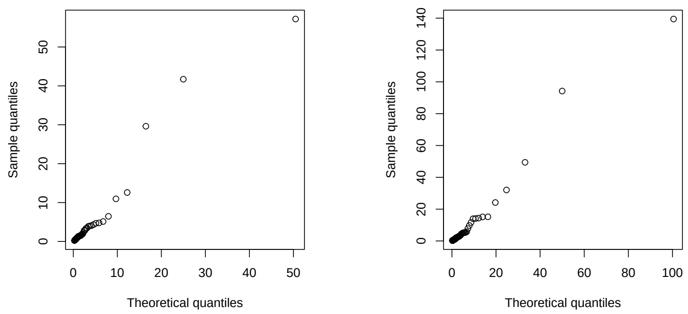

Figure 7: Quantile plots for the exponential distribution, with $n = 50$ (left) and $n = 100$ (right).

Solution 3

(a) The log-likelihood equals

$$\ell(\alpha, \beta) := \ell(x_1, \dots, x_n; \alpha, \beta) = n \log(\alpha/\beta) - (\alpha + 1) \sum_{i=1}^{n} \log(1 + x_i/\beta).$$

```
#take only positive values from Abisko (note support of the Lomax r.v.)
abplus<-abisko[abisko[,2]>0,2]

#define theta as the vector containing the parameters
theta<-vector()

nlogl_lomax<-function(theta, obs=abplus){
#negative log-likelihood
n<-length(obs);
nl<- -(n*log(theta[1]/theta[2])-(theta[1]+1)*sum(log(1+obs/theta[2])))
}
```

(b) We use the following code, which gives the rounded values $(\widehat{\alpha}_{\mathrm{MLE}}, \widehat{\beta}_{\mathrm{MLE}}) = (1.86, 1.97)$.

```
#starting values set to (1, 0.5)
(fit<-optim(c(1,.5), nlogl_lomax, hessian = T))

$par
[1] 1.864351 1.972466

$value
[1] 24102.7
```

```
...

$hessian
          [,1]      [,2]
[1,]  4353.527 -2678.408
[2,] -2678.408  1853.264
```

For the standard errors we find:

```
(se <- sqrt(diag(solve(fit$hessian))))

[1] 0.04552088 0.06976902
```

(c) The identity in (a) equals the given expression with $S(\beta) = \sum_{i=1}^{n} \log(1 + y_j/\beta)$. Differentiation with respect to α results in

$$\frac{\partial \ell(\alpha,\beta)}{\partial \alpha} = \frac{n}{\alpha} - S(\beta), \quad \frac{\partial^2 \ell(\alpha,\beta)}{\partial \alpha^2} = -\frac{n}{\alpha}^2 < 0,$$

so that $\widehat{\alpha}_\beta$ is the MLE of α for a fixed β. Using $\widehat{\alpha}_\beta$ in $\ell(\alpha, \beta)$ gives

$$\ell_{\mathrm{p}}(\beta) = \ell(\widehat{\alpha}_\beta, \beta) = n \log(n/\beta S(\beta)) - (n/S(\beta) + 1)S(\beta) \equiv -n \log S(\beta) - n \log \beta - S(\beta),$$

where the last equivalence drops the constants that do not depend on β.

The following code computes the profile log-likelihood and creates a plot of $\ell_{\mathrm{p}}(\beta)$ over a range of β, shown in Figure 8.

```
# function for profile likelihood wrt beta
plogl_lomax<-function(beta, obs=abplus){
n<-length(obs)
pl<- -n*log(sum(log(1+obs/beta)))-n*log(beta)-sum(log(1+obs/beta))
}

#evaluate at x which is a vector of dimension k
plogl_lomax_ev<-function(x){
k<-length(x)
p<-vector()
for(i in 1:k){p[i]<-plogl_lomax(beta=x[i])}
return(p)
}

#plot
plot(seq(.01,10,length=100), plogl_lomax_ev(seq(.01,10,length=100)),
 type='l', ylim=c(-156000,-154000), xlab=TeX(paste0("$\\beta$")),
 ylab="Profile \u2113")
```

The figure shows that the maximum of $\ell_{\mathrm{p}}(\beta)$ is attained at the value $\widehat{\beta}_{\mathrm{MLE}}$ obtained in (b).

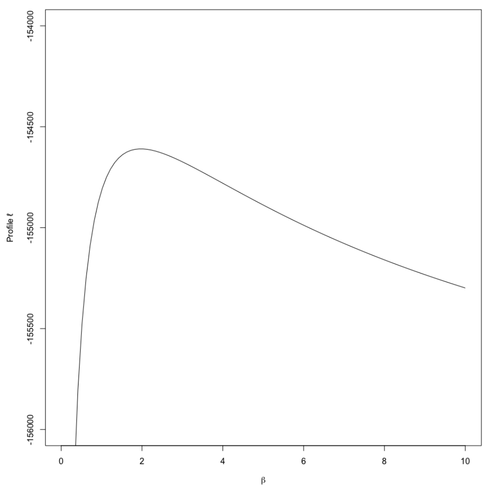

Figure 8: Plot of $\ell_{\mathrm{p}}(\beta)$.