ESTIMATES OF EXTREMES IN THE BEST OF ALL POSSIBLE WORLDS

By

R.R.P. van Nooijen[1] and A.G. Kolechkina[1,2]

[1] *Department of Water Management, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, 2628 CN, Delft, The Netherlands (r.r.p.vannooyen@tudelft.nl; a.g.kolechkina@tudelft.nl)*
[2] *Aronwis, Den Hoorn, The Netherlands*

ABSTRACT

In applied hydrology the question of the probability of exceeding a certain value occurs regularly. Often it is in a context where extrapolation from a relatively short time series is needed. It is well known that in its simplest form extreme value theory applies to independent identically distributed random variables. It is also well known that more advanced theory allows for some degrees of correlation and that techniques for coping with trends are available. However, the problem of extrapolation remains. To isolate the effect of extrapolation we generate synthetic time series of length 20, 50 and 100 from known distributions to derive empirical distributions for the 1:100 and 1:1000 exceedance.

Keywords: Extremes, Estimators, Optimization, Statistical distributions.

1 INTRODUCTION

There is pressure on hydrologists to provide predictions of extreme system behaviour based on extrapolation. For an extreme view of the results of this pressure we refer to Klemeš (2000a,b). In this paper we try to establish empirically how bad things actually are. The experiments are done in the R language R Core Team (2012). We hope that the resulting images will provide a stimulus for a more skeptical approach to the prediction of extreme values.

2 FORMAL PROBLEM STATEMENT

There is always a certain tension between theory and practice. As the title of this paper indicates, here we resolve the tension by positing an ideal world where there theory and practice live in harmony. In that case our times series of maximum discharges corresponds directly to a random vector X of *n independent identically distributed* (i.i.d.) random variables. Moreover in this ideal world our curiosity about future maxima can be satisfied through the study of a random variable Y that has the same distribution as one of the components of X. The components X_k are random variables on a probability space $\langle \Omega, \mathscr{A}, P \rangle$ (where Ω is the set of all possible outcomes, $\mathscr{A}$ is the σ-algebra of events of interest and P is a probability measure P defined on $A \in \mathscr{A}$). We assume that, although P is unknown, there is a collection $\mathscr{P}$ of probability measures such that $P \in \mathscr{P}$. If we use X to represent a generic component of X then X induces a measure μ_X on $\langle \mathbb{R}, \mathscr{B}(\mathbb{R}) \rangle$ (where $\mathscr{B}(\mathbb{R})$ is the σ-algebra).

We will assume that we can label the elements of $\mathscr{P}$ with a parameter vector $\theta \in L$ for a certain set $L \subset \mathbb{R}^m$ in such a way that these vectors can also be used to label the induced measures. We assume that for the induced measures μ_P we have well defined *probability density functions* (pdf) $f_X(. \mid \theta)$ and *cumulative distribution functions* $F_X(. \mid \theta)$ on the positive real number axis. For the proper definition of X we construct the product measure in the usual way. This results in an induced measure P_X on $\mathscr{B}(\mathbb{R}^n)$ that satisfies

$$P_X([a_1,b_1] \times [a_2,b_2] \times \ldots \times [a_n,b_n]) = P_X([a_1,b_1]) \cdot P_X([a_2,b_2]) \cdot \ldots \cdot P_X([a_n,b_n])$$

For the Bayesian case we assume that we also have a prior probability measure π on $\langle L, \mathscr{B}(L) \rangle$.

Given this model of reality we can use it to model the effects of methods of estimation of extreme system behaviour. In the next section we perform an experiment to get a feel for the problem. We pick a set of parameters that resemble parameters found for real systems. We then generate K samples of size n, use the maximum likelihood estimator to estimate the parameters based on the sample and determine an empirical distribution function for the parameters. Given these parameters we can also determine an empirical distribution function for the 1 : 20, 1 : 100 and 1 : 1000 event probability or equivalently distributions for the the 95%, 99% and the 99.9% quantiles. We will then look for theory that can provide information on the spread of predictions we may expect.

3 DISTRIBUTIONS OF INTEREST

The standard extreme value distributions for maxima are the the Gumbel (type I),

$$F_{\mathrm{G}}\left(x \mid \langle \xi, \zeta \rangle\right) = \exp\left(-\exp\left(-\frac{x-\xi}{\zeta}\right)\right)$$

the Fréchet (type II)

$$F_{\mathrm{F}}\left(x \mid \langle \xi, \zeta, \alpha \rangle\right) = \begin{cases} 0 & : x \leq \xi \\ \exp\left(-\left(\frac{x-\xi}{\zeta}\right)^{-\alpha}\right) & : x > \xi \end{cases}$$

and the reverse Weibull (type III)

$$F_{\mathrm{W}}\left(x \mid \langle \xi, \zeta, \alpha \rangle\right) = \begin{cases} \exp\left(-\left(-\frac{x-\xi}{\zeta}\right)^{-\alpha}\right) & : x \leq \xi \\ 1 & : x > \xi \end{cases}$$

distributions. In all cases $\zeta > 0$ and $\alpha > 0$. Note that the reverse Weibull (type III) has an upper limit on the maxima. If Z is Fréchet distributed with parameters $\langle 0, \zeta, \alpha \rangle$ then $Y = \log Z$ has a Gumbel distribution with parameters $\langle \log \zeta, 1/\alpha \rangle$,

$$\begin{aligned} F_Y\left(y \mid \langle \xi, \zeta, \alpha \rangle\right) &= \exp\left(-\left(\frac{\exp y}{\zeta}\right)^{-\alpha}\right) \\ &= \exp\left(-\exp\left(-\alpha\left\{y - \log \zeta\right\}\right)\right) \end{aligned}$$

Therefore in this paper when providing formulas we limit ourselves to the Gumbel distribution.

$$f_{\mathrm{G}}\left(x \mid \langle \xi, \zeta \rangle\right) = \frac{1}{\zeta}\exp\left(-\frac{x-\xi}{\zeta} - \exp\left(-\frac{x-\xi}{\zeta}\right)\right)$$

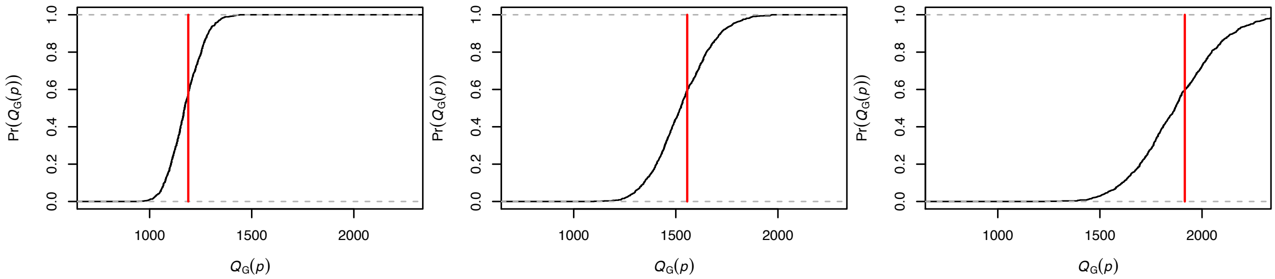
$$F_{\mathrm{G}}\left(x \mid \langle \xi, \zeta \rangle\right) = \exp\left(-\exp\left(-\frac{x-\xi}{\zeta}\right)\right)$$

$$Q_{\mathrm{G}}\left(p \mid \langle \xi, \zeta \rangle\right) = -\zeta \log\left(-\log p\right) + \xi$$

4 AN EXPERIMENT

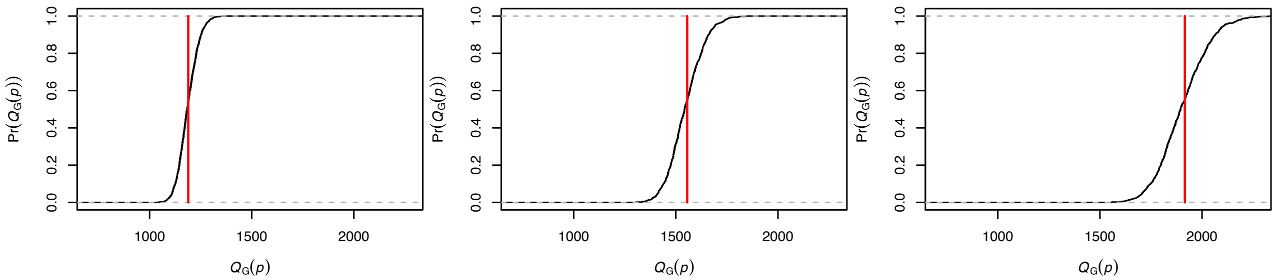
We take the `Nile` data set supplied with R 2.15.0 (R Development Core Team, 2012) and fit a Gumbel distribution to it using from package `fgev` from package `evd` with the parameter `shape` (our α) set to zero. We find $\xi = 838.2, \zeta = 156.0$. For completeness we note that `fgev` without restrictions gives $\xi = 853.6, \zeta = 157.8, \alpha = -0.1977$ which corresponds to a reverse Weibull distribution. Fitting a reverse Weibull gives similar parameters with $\xi' = 1649.6, \zeta' = 795.51, \alpha' = 5.037$, where as expected $\xi' = 1649.6 \approx \xi - \zeta/\alpha = 1651.9$, $\zeta' = 795.51 \approx -\zeta/\alpha = 798.3$ and $\alpha' = 5.037 \approx -1/\alpha = 5.058$.

Next we take a Gumbel distribution with parameters $\xi = 838.2, \zeta = 156.0$ and derive 1000 vectors of length m. We fit the Gumbel distribution to the data using the maximum likelihood method and determine $Q_{\mathrm{G}}(p)$. This results in an empirical cdf (ecdf). We do this for $m = 20, 50, 100$ and $p = 0.9, 0.99, 0.999$. The results are given in Figure 4.1, Figure 4.2 and Figure 4.3 . The vertical lines show the true quantiles.



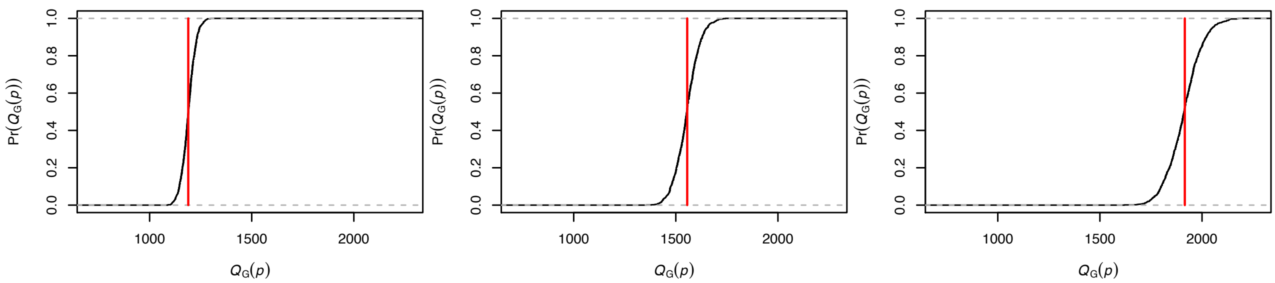
(a) $p = 0.9$ **(b)** $p = 0.99$ **(c)** $p = 0.999$

Figure 4.1 – ecdf of quantiles for $m = 20$



(a) $p = 0.9$ **(b)** $p = 0.99$ **(c)** $p = 0.999$

Figure 4.2 – ecdf of quantiles for $m = 50$



(a) $p = 0.9$ **(b)** $p = 0.99$ **(c)** $p = 0.999$

Figure 4.3 – ecdf of quantiles for $m = 100$

5 A STEP TOWARDS PREDICTIONS

If we have just one sample, the maximum likelihood method will give us just one parameter vector, θ_0. Bootstrapping might give us an impression of the spread in possible parameters, but for extremes the samples are often relatively small and leave little room for bootstrapping. Another way to examine that spread would be to repeatedly sample from the distribution with label θ_0 and examine the spread of the parameter vectors fit to these samples. However, this method would also suffer from the smallness of the sample and additionally introduce a dependence on the method used to derive θ_0.

There are several theoretical approaches we could take. We could take a confidence interval approach and look for a function that provides an upper bound $b_{\text{ub}}(\alpha, p, x)$ on the theoretical quantile location $F_X^{-1}(p)$ with probability α

$$\Pr(Q_Y(p) \leq b_{\text{ub}}(\alpha, p, X)) = \alpha$$

where $1-\alpha$ is the risk we are willing to take that our upper bound is incorrect.

We could also take a prediction interval approach and look for an upper bound $b_{\text{ub}}(p,x)$ such that

$$\Pr(Y \leq b_{\text{ub}}(p,X)) = p$$

More advanced techniques look for prediction intervals for the maximum over n observations based on $m < n$ observations. An overview of many different techniques can be found in Patel (1989) and Barndorff-Nielsen and Cox (1996).

A basic problem with estimates of prediction intervals is discussed in Barndorff-Nielsen and Cox (1996) and Vidoni (2009), where, for the general case, an error term of $O(1/n)$ crops up. Although this term might not be sharp, see for example Chen and Hall (1993), it suggests that extrapolation will be problematical.

A useful tool used in the remainder of the paper are *ancillary statistics*, according to Ghosh et al. (2010): “the usual definition is that a statistic is ancillary if its distribution does not depend on the parameters in the assumed model”.

6 FREQUENTIST THEORY

From a frequentist point of view, the parameter vector is unknown but fixed. We can only examine the probability that, for a given sample, the corresponding interval will contain the quantile of interest. For a quantile this would be a prediction interval, a special case of a confidence interval. Such an interval would be a random set $D_n(\alpha,X)$ such that $\Pr(Y \in D_n(\alpha,X)) = \alpha$. There is extensive literature on this subject, see for example Lawless and Fredette (September 2005), Barndorff-Nielsen and Cox (1996), Beran (1990) and the review paper on prediction intervals by Patel (1989). For the upper quantiles this can be reduced to finding a function $b_{\text{ub}}(\alpha,x)$ such that

$$\Pr(Y \leq b_{\text{ub}}(\alpha,X)) = \alpha$$

Now for a series of experiments sampling *both* Y and X and with null hypothesis that $y \leq Q_Y(p)$ whenever $y \leq b_{\text{ub}}(p,x)$ we have

$$\Pr(Y \leq b_{\text{ub}}(\alpha,X) \text{ and } Y \leq Q_Y(p)) = \alpha p$$

$$\Pr(Y \leq b_{\text{ub}}(\alpha,X) \text{ and } Y > Q_Y(p)) = \alpha(1-p) = \Pr(\text{type II error})$$

$$\Pr(Y > b_{\text{ub}}(\alpha,X) \text{ and } Y \leq Q_Y(p)) = (1-\alpha)p = \Pr(\text{type I error})$$

$$\Pr(Y > b_{\text{ub}}(\alpha,X) \text{ and } Y > Q_Y(p)) = (1-\alpha)(1-p)$$

Formulated in terms of exceedance probability $p_{\text{ex}} = 1-p$ we find

$$\Pr(Y \leq b_{\text{ub}}(\alpha,X) \text{ and } Y > Q_Y(p)) = \alpha p_{\text{ex}} = \Pr(\text{type II error})$$

$$\Pr(Y > b_{\text{ub}}(\alpha,X) \text{ and } Y \leq Q_Y(p)) = (1-\alpha)(1-p_{\text{ex}}) = \Pr(\text{type I error})$$

6.1 A simple example for the confidence interval approach

Purely for the purpose of illustration we provide a simple example where we can construct a prediction interval for a quantile. Consider the one parameter exponential distribution, defined for $x \geq 0$,

$$f_{\exp}(x \mid \lambda) = \lambda \exp(-\lambda x)$$

$$F_{\exp}(x \mid \lambda) = 1 - \exp(-\lambda x)$$

$$Q_{\exp}(p \mid \lambda) = -\frac{1}{\lambda}\log(1-p)$$

Suppose we are interested in determining limits on the quantiles. We use some standard properties of the exponential distribution:

- if X is exponentially distributed with parameter $\lambda = \lambda_0$ then X/λ_0 is exponentially distributed with parameter $\lambda = 1$.
- the sum of n independent exponentially distributed variables Z_k, each with parameter $\lambda = \lambda_0$ is a Γ distribution with parameters $\langle 1/\lambda_0, n\rangle$, so

$$\Pr\left(\sum_{k=1}^{n} Z_k \leq t\right) = F_{\Gamma}\left(t \mid \left\langle \frac{1}{\lambda_0}, n\right\rangle\right)$$

From these properties follows immediately that, for $0 < a < b$,

$$\begin{aligned}
\Pr\left(\left(\frac{-\log(1-p)}{\frac{1}{n}\sum_{k=1}^{n} X_k}\right) a \leq Q_Y(p,\lambda) \leq \left(\frac{-\log(1-p)}{\frac{1}{n}\sum_{k=1}^{n} X_k}\right) b\right) &= \\
\Pr\left(\left(\frac{-\log(1-p)}{\frac{1}{n}\sum_{k=1}^{n} X_k}\right) a \leq -\frac{1}{\lambda}\log(1-p) \leq \left(\frac{-\log(1-p)}{\frac{1}{n}\sum_{k=1}^{n} X_k}\right) b\right) &= \\
\Pr\left(na \leq \sum_{k=1}^{n} \frac{X_k}{\lambda} \leq nb\right) &= \\
F_{\Gamma}(nb \mid \langle 1,n\rangle) - F_{\Gamma}(na \mid \langle 1,n\rangle)
\end{aligned}$$

where we used that for $c > 0$ and $X > 0$

$$\Pr\left(\frac{1}{X} \leq c\right) = \Pr(Xc \geq 1) = \Pr\left(X \geq \frac{1}{c}\right)$$

This provides us with the probability that, for a long series of identical experiments, each involving taking a sample of size n, the quantile lies in an interval derived from the sample. We used the following notation:

$$F_{\Gamma}(x \mid \langle \zeta, \alpha\rangle) = \int_{t=-\infty}^{x} f_{\Gamma}(x, \langle \zeta, \alpha\rangle)\, dx$$

where

$$f_{\Gamma}(t \mid \langle \zeta, \alpha\rangle) = \frac{t^{\alpha-1}\exp\left(-\frac{t}{\zeta}\right)}{\zeta^{\alpha}\Gamma(\alpha)}$$

6.2 A simple example for the prediction interval approach

In Example 1 of Lawless and Fredette (September 2005) a predictive distribution is provided, which can be used to derive a prediction interval. We repeat here the main ingredients and the result. If X_k and Y are independent random variables with an exponential distribution then

$$\frac{Y}{\sum_{i=1}^{n} X_i} = \frac{\frac{Y}{\lambda}}{\sum_{i=1}^{n} \frac{X_i}{\lambda}}$$

so this neither quantity nor its distribution depends on λ. In this fraction the numerator has pdf $f_{\text{exp}}\left(. \mid 1\right)$ and the denominator has distribution $f_{\Gamma}\left(. \mid \langle 1, n\rangle\right)$ so

$$\Pr\left(\frac{Y}{\frac{1}{n}\sum_{i=1}^{n} X_i} \leq c\right) = \Pr\left(\frac{Y}{\sum_{i=1}^{n} X_i} \leq \frac{c}{n}\right)$$

and

$$\begin{aligned}
\Pr\left(\frac{Y}{\sum_{i=1}^{n} X_i} \leq \frac{c}{n}\right) &= \int_{s=0}^{\infty}\int_{y=0}^{cs/n} f_{\text{exp}}\left(y \mid 1\right) f_{\Gamma}\left(s \mid \langle 1, n\rangle\right) dy ds \\
&= \int_{s=0}^{\infty} \left(1 - \exp\left(-\frac{c}{n}s\right)\right) f_{\Gamma}\left(s \mid \langle 1, n\rangle\right) ds \\
&= 1 - \int_{s=0}^{\infty} \frac{s^{n-1}\exp\left(-\left(1+\frac{c}{n}\right)s\right)}{\Gamma(n)} ds \\
&= 1 - \frac{1}{\left(1+\frac{c}{n}\right)^n} \int_{s=0}^{\infty} \frac{\left(\left(1+\frac{c}{n}\right)s\right)^{n-1}\exp\left(-\left(1+\frac{c}{n}\right)s\right)}{\Gamma(\alpha)} \left(1+\frac{c}{n}\right) ds \\
&= 1 - \frac{1}{\left(1+\frac{c}{n}\right)^n}
\end{aligned}$$

so for $c_1 < c_2$

$$\Pr\left(c_1 \frac{1}{n}\sum_{i=1}^{n} X_i \leq Y \leq c_2 \frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{\left(1+\frac{c_1}{n}\right)^n} - \frac{1}{\left(1+\frac{c_2}{n}\right)^n}$$

6.3 Example with Gumbel with confidence interval approach

We wish to study

$$\Pr\left(Q_Y\left(p, \langle \xi, \zeta\rangle\right) \leq b_{\text{ub}}\left(p, \alpha, X\right)\right)$$

We define

$$\hat{\sigma}(x) = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(x_j - \frac{1}{n}\sum_{k=1}^{n} x_k\right)^2}$$

and

$$\hat{\mu}(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Note that

$$\hat{\mu}(x) = \zeta\hat{\mu}\left(\frac{x-\xi}{\zeta}\right) + \xi$$

and

$$\hat{\sigma}(x) = \zeta\hat{\sigma}\left(\frac{x-\xi}{\zeta}\right)$$

Define

$$W_{p,a} = (a - \log(-\log p))\,\hat{\sigma}\left(\frac{X-\xi}{\zeta}\right) + \hat{\mu}\left(\frac{X-\xi}{\zeta}\right)$$

this random variable induces a distribution $F_W(w \mid \langle p,a \rangle)$ that does not depend on ξ or ζ. If, for given p, we can determine $a(p,\alpha)$ such that

$$1 - F_W(-\log(-\log p) \mid \langle p, a(p,n,\alpha)\rangle) = \alpha$$

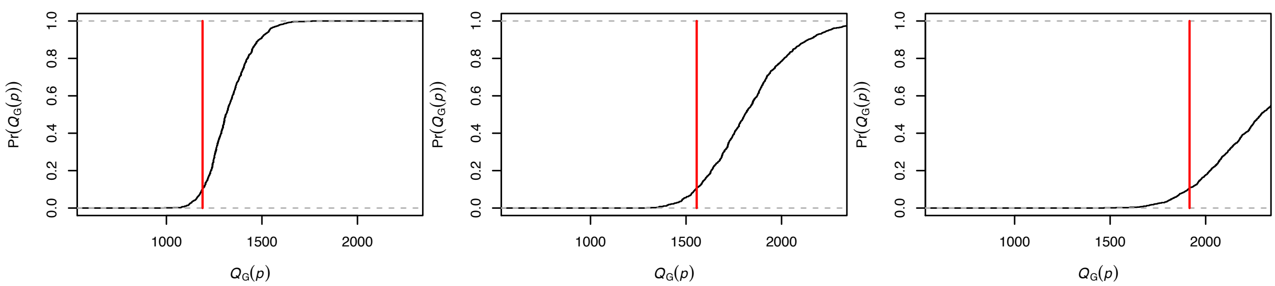
so the probability that the true quantile is greater than our estimate is α, and we take

$$b_{\text{ub}}(p,\alpha,x) = (a(p,\alpha) - \log(-\log p))\,\hat{\sigma}(x) + \hat{\mu}(x)$$

then

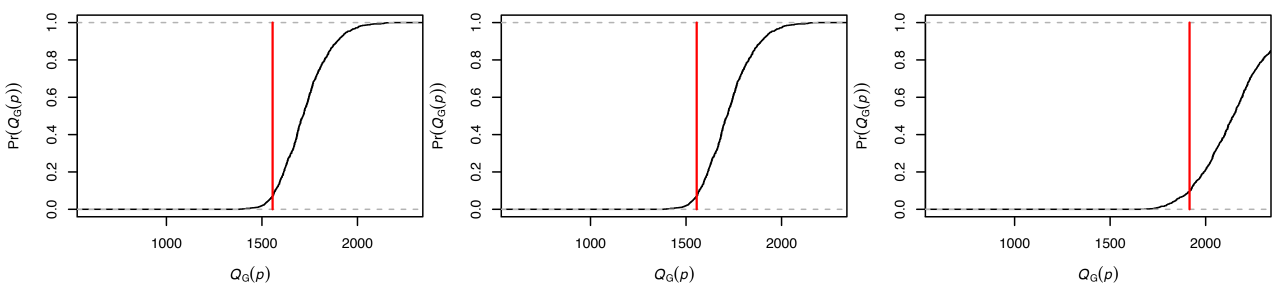
$$\begin{aligned}
\Pr(Q_Y(p,\langle\xi,\zeta\rangle) \le b_{\text{ub}}(p,\alpha,X)) &= \\
\Pr(-\zeta\log(-\log p) + \xi \le (a(p,\alpha) - \log(-\log p))\,\hat{\sigma}(X) + \hat{\mu}(X)) &= \\
\Pr\left(-\log(-\log p) \le (a(p,\alpha) - \log(-\log p))\,\hat{\sigma}\left(\frac{X-\xi}{\zeta}\right) + \hat{\mu}\left(\frac{X-\xi}{\zeta}\right)\right) &= \\
1 - F_W(-\log(-\log p) \mid \langle p,a\rangle) &= \alpha
\end{aligned}$$

In Figures (6.1,6.1,6.3) $a(p,n,\alpha)$ was taken to match $\alpha = 0.9$. They show the correct behaviour of our quantile estimator.



(a) $p = 0.9$ **(b)** $p = 0.99$ **(c)** $p = 0.999$

Figure 6.1 – ecdf of quantile estimates for $m = 20$

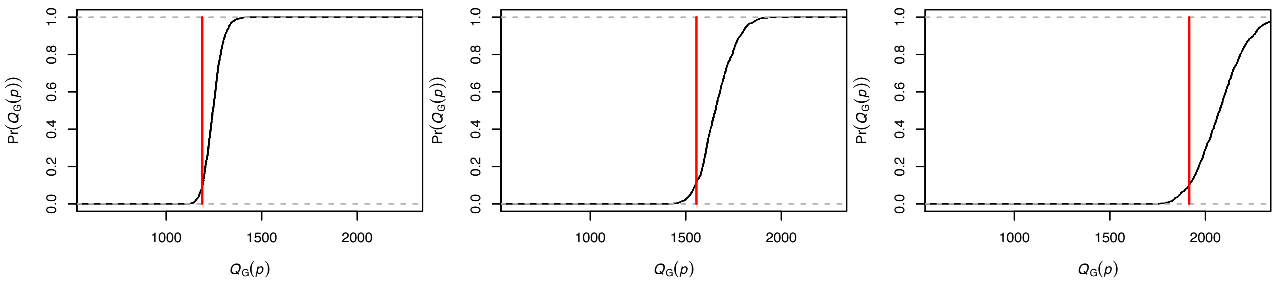


(a) $p = 0.9$ **(b)** $p = 0.99$ **(c)** $p = 0.999$

Figure 6.2 – ecdf of quantile estimates for $m = 50$

7 BAYESIAN THEORY

For the Bayesian approach we need a prior on our parameter space. Once we have that the Bayesian analogue of a frequentist confidence interval is a credible set Carlin and Luis (2000). It is defined as follows. A $100\times$



(a) $p = 0.9$ **(b)** $p = 0.99$ **(c)** $p = 0.999$

Figure 6.3 – ecdf of quantile estimates for $m = 100$

$(1-\alpha)$ percent credible set C for a parameter (vector) θ given a sample y is a set C such that

$$1-\alpha \leq \Pr(\theta \in C \mid x) = \int_C \pi(\theta \mid x)\, d\theta$$

where $\pi(\theta \mid y)$ is the posterior density. Next let us consider how to get information on the distribution of a quantile. Given our earlier definitions we have

$$\Pr(Y \leq y \mid \theta) = F_X(y \mid \theta)$$

We introduce the notation $Q_X(p \mid \theta)$ for the quantile function and we assume that for $p \in]0,1[$

$$F_X(Q_X(p \mid \theta) \mid \theta) = p$$

Now suppose Θ is the random vector corresponding to θ. For a given quantile p_0 we are looking for

$$\Pr(Q_X(p_0 \mid \Theta) \leq t)$$

which is the distribution of that particular quantile given a (posterior) distribution for the parameter.

7.1 A toy example

Again we use the one parameter exponential distribution. For known λ, in other words $\lambda \in L = \{\lambda_0\}$ we have

$$\Pr(Q_X(p_0 \mid \Lambda) \leq t) = \begin{cases} 0 & : t < Q_X(p_0 \mid \lambda_0) \\ 1 & : t \geq Q_X(p_0 \mid \lambda_0) \end{cases}$$

For unknown λ we have $\lambda \in L =]0,\infty[$ we take as prior distribution for the parameter the traditional improper prior distribution for a scale parameter

$$\pi(\lambda) = \frac{1}{\lambda}$$

A priori we have

$$\Pr(Q_X(p_0 \mid \Lambda) \leq t) = \int_{\{\lambda>0: Q_X(p_0|\lambda) \leq t\}} \pi(\lambda)\, d\lambda$$

and a posteriori we have

$$\Pr(Q_X(p_0 \mid \Lambda) \leq t) = \int_{\{\lambda>0: Q_X(p_0|\lambda) \leq t\}} \pi(\lambda \mid x)\, d\lambda$$

In both cases

$$\begin{aligned} \{\lambda > 0 : Q_X(p_0 \mid \lambda) \leq t\} &= \left\{\lambda > 0 : -\frac{1}{\lambda}\log(1-p) \leq t\right\} \\ &= \left\{\lambda > 0 : -\frac{1}{t}\log(1-p) \leq \lambda\right\} \end{aligned}$$

so the a priori case is not integrable. For the a posteriori case we see that

$$\begin{aligned}\pi(\theta \mid x) &\propto \pi(\lambda) \prod_{j=1}^{n} \lambda \exp(-\lambda x_j) \\ &= \lambda^{n-1} \exp\left(-\lambda \sum_{j=1}^{n} x_j\right)\end{aligned}$$

so

$$\pi(\lambda \mid x) = \frac{\left(\sum_{j=1}^{n} x_j\right)^n}{\Gamma(n)} \lambda^{n-1} \exp\left(-\lambda \sum_{j=1}^{n} x_j\right)$$

with the cdf of the Γ distribution

$$F_\Gamma\left(\lambda \sum_{k=1}^{n} x_k \mid \langle 1, n\rangle\right)$$

so

$$\begin{aligned}\Pr(Q_X(p_0 \mid \Lambda) \leq t) &= \int\limits_{-\frac{1}{t}\log(1-p)\leq\lambda} \pi(\lambda \mid x)\, d\lambda \\ &= 1 - F_\Gamma\left(-n\frac{\frac{1}{n}\sum_{k=1}^{n} x_k}{t} \log(1-p), \mid \langle 1, n\rangle\right)\end{aligned}$$

As a posteriori distribution for Y we find

$$\begin{aligned}f_Y(y \mid x) &\propto \int_{\lambda=0}^{\infty} \lambda \exp(-\lambda y) \frac{\left(\sum_{k=1}^{n} x_k\right)^n}{\Gamma(n)} \lambda^{n-1} \exp\left(-\lambda \sum_{k=1}^{n} x_k\right) d\lambda \\ &= \frac{\left(\sum_{k=1}^{n} x_k\right)^n}{\left(y+\sum_{k=1}^{n} x_k\right)^n} \int_{\lambda=0}^{\infty} \frac{\left(y+\sum_{k=1}^{n} x_k\right)^n}{\Gamma(n)} \lambda^n \exp\left(-\lambda\left(y+\sum_{k=1}^{n} x_k\right)\right) d\lambda \\ &= \frac{\left(\sum_{k=1}^{n} x_k\right)^n}{\left(y+\sum_{k=1}^{n} x_k\right)^{n+1}} \int_{t=0}^{\infty} \frac{1}{\Gamma(n)} t^n \exp(-t)\, dt \\ &= \frac{\left(\sum_{k=1}^{n} x_k\right)^n}{\left(y+\sum_{k=1}^{n} x_k\right)^{n+1}} \frac{\Gamma(n+1)}{\Gamma(n)} = \frac{n}{\left(\sum_{k=1}^{n} x_k\right)} \left(\frac{y}{\sum_{k=1}^{n} x_k}+1\right)^{-(n+1)}\end{aligned}$$

so

$$F_X(y \mid x) = 1 - \left(1 + \frac{1}{n} \frac{y}{\frac{1}{n}\sum_{k=1}^{n} x_k}\right)^{-n}$$

which remarkably enough is equal to the frequentist bound.

8 CONCLUSIONS

We saw that for small samples the spread in estimates of the upper quantiles is considerable. Moreover, both literature references and common sense (error inversely proportional to number of measurements) suggest that it will be very hard to obtain reliable estimates of the upper quantiles. It seems unlikely that we can estimate the $1 - 1/k$ quantile with sufficient accuracy without at least k observations. However, many Bayesian and frequentist tools to obtain either point or interval estimates exist and research into better, more accurate methods for quantile estimation is alive and well.

9 REFERENCES

Barndorff-Nielsen O.E. and Cox D.R. (1996). Prediction and asymptotics. *Bernoulli*, 2: 319–340.

Beran R. (1990). Calibrating prediction regions. *Journal of the American Statistical Association*, 85: pp. 715–723.

Carlin B.P. and Luis T.A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall/CRC, second edition.

Chen S.X. and Hall P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *The Annals of Statistics*, 21: pp. 1166–1181.

Ghosh M., Reid N. and Fraser D.A.S. (2010). Ancillary statistics: a review. *Statist. Sinica*, 20: 1309–1332.

Klemeš V. (2000a). Tall tales about tails of hydrological distributions. I. *Journal of Hydrologic Engineering*, 5: 227–231.

Klemeš V. (2000b). Tall tales about tails of hydrological distributions. II. *Journal of Hydrologic Engineering*, 5: 232–239.

Lawless J.F. and Fredette M. (September 2005). Frequentist prediction intervals and predictive distributions. *Biometrika*, 92: 529–542.

Patel J.K. (1989). Prediction intervals - a review. *Communications in Statistics - Theory and Methods*, 18: 2393–2465.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Vidoni P. (2009). Improved prediction intervals and distribution functions. *Scandinavian Journal of Statistics*, 36: 735–748.