Nonparametric Tail Extrapolation

F.W. Scholz
Research & Technology
Boeing Information & Support Services

Abstract

Given a pure random sample, $X_1, \ldots, X_n$, from a population with a continuous distribution function F we are interested in the extrapolation problem. Namely, we wish to estimate or provide confidence bounds for extreme quantiles x_p of F, in particular for quantiles which fall beyond the range of the sample. Building on the fact that the i^{th} sample extreme Y_i can act as exact $100\gamma\%$ upper confidence bound for x_{p_i} for some appropriate $p_i = p_i(\gamma, n)$, we extend the scope of these nonparametric confidence bounds by introducing an adaptive type of QQ-plot, which plots the sample extremes $Y_1, \ldots, Y_k$ against appropriate transforms of $p_1, \ldots, p_k$. In such a plot the sample quantiles are expected to form an approximately straight line which then becomes the vehicle for extrapolation. The adaptability of the QQ-plot arises from the assumption that F is in the domain of attraction of one of the three classical extreme value distribution types. Estimating the extreme-value index c by the generalization of Hill's estimator (Dekkers et al., 1989) we then use a probability transform $f_c(p_i) = ([-n\log(p_i)]^{-c}-1)/c$ to act as the abscissa values of the QQ-plot. A straight line or quadratic is fitted by weighted least squares, using the estimated approximate covariance structure suggested by extreme value theory. The extrapolation depth k is pushed as far in as is consistent with the domain of attraction assumption. Under the latter the slope of the fitted line and the scatter of the points around that line are intimately linked leading to a ratio criterion that can then be used to motivate the proper depth k. By duality the confidence bounds for quantiles can be inverted to confidence bounds for tail probabilities for given thresholds.

1 Introduction and Summary

One often is interested and tempted to make inferences about the extreme tail behavior of distributions without having sufficient data, as is the case when we consider extrapolation from the data to regions beyond the data. To make inferences in the face of such difficulties is a risky undertaking. Traditionally such extrapolations often take the form of fitting an "appropriate" probability distribution to the observed data and using that distribution in the extrapolation step. Of course, the fit of the employed probability distribution can only be judged over the range of the observed data and not beyond. There are many distributions, in fact infinitely many, that agree with the fitted distribution over the observed range but show vastly different behavior in the tail regions beyond that range. If however a particular distribution fits the data very well over that range and if that distribution is from a list of well known distributions, then one might be more inclined to take that extrapolatory step with some faith. The reason for this is that these distributions are well known for certain intrinsic modeling properties which, with some reflection, would make them a natural choice in particular applications. Of course, such reflection sometimes comes after the fact of examining the data. These modeling properties often describe the mechanics of how the random data may be generated. Given that the model fits the data reasonably well so far, there is then reason to believe that the underlying mechanics will continue to work in similar fashion, if we were to collect such data ad infinitum, i.e., get the whole population. Ultimately, the act of fitting a distribution to the data already constitutes an extrapolation step, although some practitioners may not be fully aware of it.

When judging the fit of a particular probability distribution to a set of data it is worthwhile contrasting the following two graphical techniques: In the first we superimpose the fitted density function over the histogram of the data. In the second we compare the quantiles of the observed data to the corresponding quantiles of the fitted distribution. The p–quantile, x_p, is the division point between the lower $100p\%$ and the upper $100(1-p)\%$ of all population values of this distribution. Similarly, the sample p–quantile, $\hat{x}_p$, is the division point between the lower $100p\%$ and the upper $100(1-p)\%$ of all sample values. This is not very precise but should suffice at this point. For a good fit such a quantile-quantile plot (QQ-plot) should approximately follow a straight line, namely the main diagonal. What may look innocuous or only somewhat suspect in a histogram/density comparison may become quite glaring in the QQ-plot.

On the other hand, if the fit looks good in either case, then the extrapolation along the fitted straight line looks much more inviting than the continuation along a density curve which has followed the histogram closely over the data range. We are doing an extrapolation step in either case. Much of the difference between the two steps is only a matter of perception. However, one advantage of the QQ-plot is that it does not require a choice for width and location of bins as they are needed for histograms. Further, since the sample quantiles are related in direct fashion to the empirical distribution function we also inherit the latter's greater stability as an estimator, since the empirical distribution function involves averaging and thus some smoothing.

Of course, a straight line extrapolation only seems to be less biased than committing oneself to some kind of curvature beyond the data range, because the curvature issue has in actuality only been transformed into an equivalent distortion of the plotting axes.

Ultimately, the extrapolation step is one of good faith
and not statistical in nature.

If one were perverse and pessimistic, but not necessarily pragmatic, one could easily generate the same kind of data from distributions with vastly different extreme tail behavior beyond the range of the data, and thus without being able to detect it. Having understood this and being pragmatic and somewhat optimistic we will continue by offering an alternate scheme which is a mostly nonparametric in nature, i.e., it is not based on a particular population model for the data.

Throughout we will focus on the extreme right tail of the distribution. The left tail of the distribution can be treated by reflection around zero. First we will develop upper confidence bounds for the upper quantiles. Such upper confidence bounds are also called upper tolerance bounds. By duality one can invert these methods to also obtain upper bounds for right tail probabilities corresponding to given thresholds. Lower confidence bounds are obtained by complementing the confidence level, i.e., a $100(1-\gamma)\%$ upper bound for some target θ serves also as a $100\gamma\%$ lower bound for the same target.

One feature of the proposed method is that it is based only on the observed tail behavior of the sample, i.e., the extrapolation results are not influenced by the middle nor by the opposite tail of the data. One advantage is that we no longer have to fit a distribution over the full range of the data. With larger samples of real data, e.g., with $n = 1000$ or more data points, it becomes very difficult to pass any of the formal tests of fit over the full data range for any of the well known distributions. Of course, one could devise tests of fit over the tail of a given distribution using only the k relevant sample extremes. Some such tests are available and they are discussed by Michael and Schucany (1986).

The nonparametric extrapolation scheme makes two basic assumptions about the sampled distribution, namely that it be continuous and that it should be in the domain of attraction of one of the three possible extreme value distribution types. Although we require a continuous parent distribution, which excludes ties in the sample data with probability one, we feel that such ties due to rounding or recording accuracy limitations should not invalidate the method. What is excluded here are data that are intinsically limited to discrete points. The assumption of being in the domain of attraction of one of the extreme value distributions is mainly used in motivating the linearization of the extrapolation plot.

The scheme is based on exact upper confidence bounds for selected population quantiles. Here "exact" refers to the fact that no approximations are involved prior to the extrapolatory step. The latter is taken in such a way that its quality can to some extent be judged visually where it counts and its effect is not hidden or propagated by further procedural steps. First an appeal is made to linear extrapolation against a somewhat ad hoc choice for a transformed percentage scale, i.e., using the log-odds scale. Through extreme value theory this ad hoc choice can be motivated when the sampled population is in the domain of attraction of the extreme value or Gumbel distribution. When the sampled population is in the domain of attraction of either of the other extreme value distribution types we can use an adaptively transformed scale for linearization. The choice of this scale is based on the extreme-value index estimate proposed by Dekkers et. al. (1989).

Using this index estimate we also estimate the approximate covariance structure of the extreme order statistics and fit a line to the point pattern using generalized least squares. This avoids being unduly influenced by the sometimes wild swings in the most extreme order statistics. We can use the residual standard deviation, $\widehat{\sigma}_k$, to guide us in the choice of the data tail proportion to be used for extrapolation. Because of the domain of attraction assumption $\widehat{\sigma}_k$ estimates not only the residual variability but also the slope δ of the fitted line. Since the weighted least squares fit gives us another estimate $\widehat{\delta}$ of δ, we can use the ratio $T = \widehat{\delta}/\widehat{\sigma}_k$ as a natural criterion for deciding on the sample tail depth to be used for extrapolation.

The domain of attraction assumption essentially says that the tail behavior of the sampled distribution F be of some broad type and it seems only reasonable to use as many of the sample extremes as are consistent with that stipulated tail behavior type. As long as we are in the range of that validity it seems reasonable to treat T as a known multiple of a noncentral Student-t random variable for which we can give expected ranges of variation around one. As long as T is within that range it would appear safe to use the sample tail data to that depth. Iterative search over several tail depths will eventually lead to a reasonable choice for the appropriate depth. Of course, there are still certain arbitrary choices and the noncentral t distribution assumption is somewhat tenuous. Thus it is mandatory to test the method extensively by Monte Carlo methods against samples drawn from many diverse but known populations. The goal of this validation is to establish to what extent the intended confidence level is maintained, how far out from the data can we expect its deterioration, how much variability can we expect from such confidence bounds, and to what extent do these results depend on the sample size and the sampled distribution.

Linear extrapolation is reasonably motivated only for estimates (median unbiased, confidence level $\gamma = .5$). For general confidence bounds it was thus felt worthwhile to try not only linear, but also quadratic extrapolation, since quadratic polynomials represent the simplest form of model deviation from linearity. However, only curvature away from the linear extrapolation fit was allowed.

A major feature of the method is that the judgment of extrapolation quality has a strong graphical component. Aside from our more objective criterion T, it also allows us to judge the degree of linearity visually and provides a visual awareness of the extent to which extrapolation goes beyond the reach of the data.

2 Nonparametic Tolerance Bounds

Let $X_1, \ldots, X_n$ be a random sample from a population with continuous cumulative distribution function $F(x) = P(X_i \le x)$, and denote by $Y_1 \ge Y_2 \ge \ldots \ge Y_n$ the ordered sample, in order from largest to smallest. The p–quantile x_p of F is defined as the smallest value for which $F(x_p) = p$, i.e. $x_p = \inf\{x : F(x) \ge p\}$. Hence $P(X_i \le x_p) = p$.

It is well known that each Y_i can, for given p, serve as $100\gamma\%$ upper confidence bound for the quantile x_p, when the confidence level γ is determined from the following identity

$$\gamma = P(Y_i \geq x_p) = \sum_{j=i}^{n} \binom{n}{j} (1-p)^j p^{n-j} = I_{1-p}(i, n-i+1) \ , \tag{1}$$

where

$$I_x(a,b) = \frac{\int_0^x \ t^{a-1}(1-t)^{b-1} \ dt}{\int_0^1 \ t^{a-1}(1-t)^{b-1} \ dt} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x \ t^{a-1}(1-t)^{b-1} \ dt$$

is the incomplete beta function ratio.

Usually the confidence level γ and the value p of the target quantile x_p are specified and one tries to choose i such that Equation (1) is satisfied as closely as possible. Unfortunately the choice of i is limited to few discrete values, $i = 1, \ 2, \ 3, \ \ldots$, and for high γ or $p \approx 1$ not even $i = 1$ will give a satisfactory choice. Thus we will approach this problem from a slightly different angle.

In Equation (1) treat i and γ as fixed and determine $p = p_i = p_{i,\gamma,n}$ such that this equation is satisfied. Foe each i and γ there is a unique and exact solution to this problem. For $\gamma = .5$ an excellent approximation of $p_{i,.5,n}$ is given by

$$p_{i,.5,n} \approx 1 - \frac{i - \frac{1}{3}}{n + \frac{1}{3}} \ ,$$

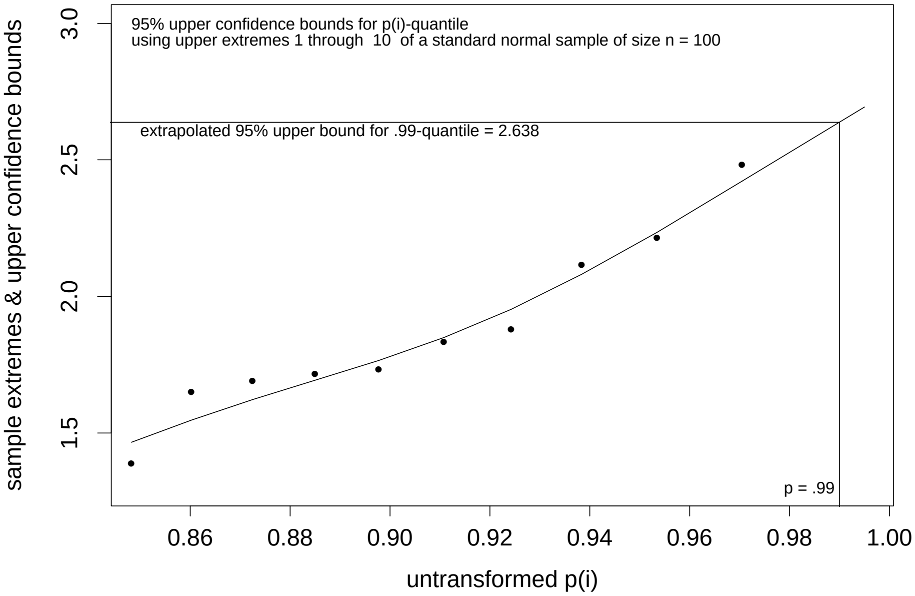
see Hoaglin (1983), and also Filliben (1975) for a similar approximation.

Thus Y_i can serve as an exact $100\gamma\%$ upper confidence bound for x_{p_i}. This can be done for $i = 1, \ 2, \ 3, \ \ldots$. Plotting Y_i against p_i as in Figure 1a for $\gamma = .95$ might suggest a smoothing curve (as shown there by a smoothing spline) for interpolation, or even extrapolation, on the p–axis to a value p_0 of interest. Using such a smoothing curve one reads off (again as shown in Figure 1a for $p_0 = .99$) the corresponding $Y_0 = Y_{p_0} = 2.638$ on the ordinate scale and then treats Y_0 as an extrapolated/interpolated $100\gamma\%$ upper confidence bound for x_{p_0}.

Unfortunately the p_i on the abscissa are naturally bounded between 0 and 1, whereas the Y_i are usually unbounded. Thus the relationship between Y_i and p_i will usually be quite curved and will not be very amenable to taking the extrapolatory step, especially not by linear extrapolation as was done in Figure 1a.

Although the 95% upper confidence bound turned out to be 2.638 and is on the correct side of the true standard normal .99-quantile = 2.326, it is quite obvious

Figure 1a: Exact Nonparametric Tolerance Bounds with $p_{i,\gamma,n}$ not transformed

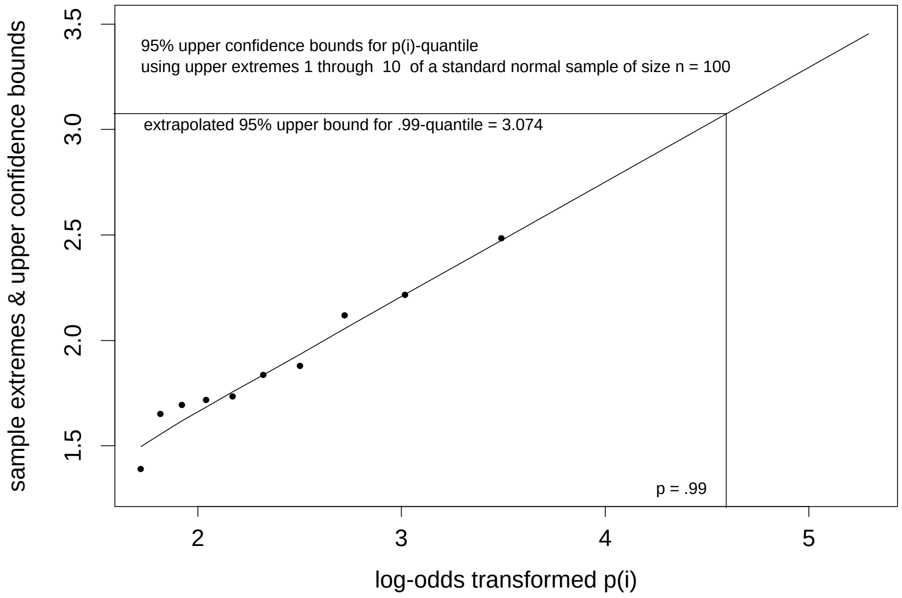


that for p much closer to 1, the linearly extrapolated upper bound would never exceed the value 3, whereas the true standard normal quantiles are unbounded for p close to 1.

To avoid this difficulty one could employ the often used device of viewing the p_i on the unbounded log–odds scale, where linear extrapolation is less problematic. Here we extrapolate the pattern of Y_i versus $\log(p_i/(1-p_i))$, that pattern often having more of a straight line character.

Using the same data as in Figure 1a, this is illustrated in Figure 1b, where again a smoothing spline was fitted and linearly extrapolated. Note that the spline itself is almost linear within the data range, although that is somewhat accidental. Instead of fitting a smoothing spline one could also simply fit a straight line. The circumstances under which this is reasonable will be explored in the next section.

Figure 1b: Exact Nonparametric Tolerance Bounds with log-odds transformed $p_{i,\gamma,n}$



3 Extreme Value Theory

As pointed out we need to understand the situations for which the above ad hoc choice of the log-odds transform can be justified as a linearizing transform. We also need to find out what to do in many other, if not all other, situations. Extreme value theory provides answers to both both questions.

To set the stage, we review the notion of *domain of attraction*. A distribution function F, giving rise to a random sample $X_1, \ldots, X_n$, is in the domain of attraction of a distribution H, if there are sequences of constants $\{a_n\}$ and $\{b_n > 0\}$ such that, as $n \to \infty$,

$$\begin{aligned} P\left(\frac{\max(X_1, \ldots, X_n) - a_n}{b_n} \le y\right) &= P\left(\max(X_1, \ldots, X_n) \le a_n + b_n y\right) \\ &= \left[P\left(X_i \le a_n + b_n y\right)\right]^n \\ &= F^n\left(a_n + b_n y\right) \quad \longrightarrow \quad H(y) \qquad (2) \end{aligned}$$

for all $-\infty < y < \infty$. This translates to

$$F(a_n + b_n y) \approx H^{1/n}(y)$$

for large n. For those y which are of interest in the limit, i.e., for which $0 < H(y) < 1$, and for large enough n we have that $H^{1/n}(y) \approx 1$ and thus

$$F(z) \approx H^{1/n}\left(\frac{z - a_n}{b_n}\right) \qquad (3)$$

for $F(z) \approx 1$, writing $z = a_n + b_n y$. This characterizes the right tail behavior of F in terms of the distribution H. The depth, i.e., how close to one $F(z)$ must be, for the approximation (3) to hold, depends on the distribution F. For the domain of attraction condition to be meaningful, it would seem that (3) ought to hold reasonably well for $F(z) \ge 1 - k/n$ for some k. It would be of little practical use if approximation (3) becomes valid only for $F(z) > 1 - 1/n$. The asymptotic arguments would still hold, but the contemplated sample sizes in the limiting argument are way beyond the given sample size. Thus the limiting results can then not be appealed to for finite sample approximation. Our attitude in appealing to extreme value theory is that the limiting arguments are useful as approximations for the sample size at hand.

According to extreme value theory, see Castillo (1988), the above limiting distribution H can only be one of three different types, namely the *Frèchet* type

$$H(x) = \Phi_\nu(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\nu}) & x > 0 \end{cases} \quad \text{for some } \nu > 0 \ ,$$

the *Weibull* type (its traditional form reflected around zero)

$$H(x) = \Psi_\nu(x) = \begin{cases} \exp(-(-x)^{\nu}) & x < 0 \\ 1 & x \geq 0 \end{cases} \quad \text{for some } \nu > 0 \ ,$$

and the *Gumbel* type

$$H(x) = \Lambda(x) = \exp(-\exp(-x)) \quad \text{for } -\infty < x < \infty \ .$$

These three forms can be represented simultaneously in one analytic form, namely in the von Mises form:

$$H_{c,\lambda,\delta}(x) = \exp\left\{ -\left[1 + c\left(\frac{x-\lambda}{\delta}\right)\right]^{-1/c} \right\} \quad \text{for} \quad 1 + c\left(\frac{x-\lambda}{\delta}\right) \geq 0 \ .$$

The above three types are distinguished by the extreme-value index c. For $c > 0$, $\nu = 1/c$, $\delta = c$, $\lambda = 1$ we obtain the Frechét form $\Phi_\nu(x)$, for $c < 0$, $\nu = -1/c$, $\delta = -c$, $\lambda = -1$ we have the Weibull form $\Psi_\nu(x)$, and $c = 0$ (interpreted as $c \to 0$), $\delta = 1$, $\lambda = 0$ yields the Gumbel form or extreme value distribution $\Lambda(x)$. If F satisfies the above limiting property (2), we can write for large enough x (i.e., $F(x) \approx 1$)

$$F(x) \approx H^{1/n}_{c,\lambda,\delta}(x)$$

for appropriate location and scale parameters λ and δ. For $p \approx 1$ we can thus find the p-quantile x_p of F approximately by solving

$$H_{c,\lambda,\delta}\left(x_p\right) = p^n \ ,$$

i.e.,

$$x_p \approx \delta \ \frac{(-n\ln(p))^{-c} - 1}{c} + \lambda$$

for $c \neq 0$ and for $c = 0$ ($c \to 0$) we get

$$x_p \approx \lambda + \delta(-\ln(-n\ln(p))) = \lambda + \delta(-\ln(n)) - \delta\ln(-\ln(p)) \ .$$

Note that for $p \approx 1$ we have $-\ln(p) \approx 1-p \approx (1-p)/p$ and thus (for $c = 0$)

$$x_p \approx \lambda - \delta \ln(n) + \delta \ln\left(\frac{p}{1-p}\right),$$

which is a linear function of the log-odds of p. By choosing $\gamma = .5$ and finding the respective $p_i = p_{i,\gamma,n}$, the right tail order statistics $Y_1 \geq Y_2 \geq \ldots \geq Y_k$ will be median unbiased estimates of $x_{p_1}, x_{p_2}, \ldots, x_{p_k}$. Thus in the case $c = 0$, i.e., when F is in the domain of attraction of the Gumbel distribution, we can expect the $Y_1, \ldots, Y_k$ to show an approximately linear pattern when plotted against the log-odds of $p_1, \ldots, p_k$.

The above argument suggests the appropriate transform for any other value of c, namely transform p_i to

$$f_c(p_i) = \frac{(-n\ln(p_i))^{-c} - 1}{c}$$

and plot $Y_1, \ldots, Y_k$ against $f_c(p_1), \ldots, f_c(p_k)$. Again one would expect a near linear plot pattern, since the p_i-quantiles are in approximately linear relation to the transforms $f_c(p_i)$ of p_i. The validity of extending this latter relation far beyond the range of the data and thus being able to extrapolate it to the extreme quantiles of interest depends on the quality of the extreme value approximation that is being appealed to. This is similar to using the central limit theorem in justifying a normality assumption and then using the normal curve to perform extreme tail extrapolations. If the normal approximation is poor in the tails, then the extrapolation will be poor as well.

In using the transformation $f_c(p_i)$ it is assumed that it is known which value of c to choose. One could choose the value c so that the plotted points become most linear, using a generalized least squares approach in each linear fit. This is still an option worth pursuing. Here we estimate c from the data directly using the moment estimate proposed by Dekkers et al. (1989), namely:

$$\widehat{c}_k = M_{1,k} + 1 - .5\left(1 - \frac{M_{1,k}^2}{M_{2,k}}\right)^{-1}$$

with

$$M_{1,k} = \frac{1}{k-1}\sum_{i=1}^{k-1}\log(\tilde{Y}_i/\tilde{Y}_k) \qquad \text{and} \qquad M_{2,k} = \frac{1}{k-1}\sum_{i=1}^{k-1}\left[\log(\tilde{Y}_i/\tilde{Y}_k)\right]^2 ,$$

where $\tilde{Y}_i = Y_i - \text{median}(X_1, \ldots, X_n)$. Here it is assumed that $k < n/2$. The shifting of the data by $\text{median}(X_1, \ldots, X_n)$ is motivated by two considerations. The first is that the arguments in the logarithms should be positive and the second is that the estimate $\widehat{c}_k$ then becomes properly location and scale invariant. The median shift was suggested to me by Laurent De Haan (via e-mail) after earlier attempts of shifting the whole sample to be positive failed miserably for Cauchy samples. De Haan argues that the asymptotic results in their paper (Dekkers et al., 1989) should still hold, since the rate of convergence for the median is $1/\sqrt{n}$ whereas the extreme value limiting result rates are tied to k which is assumed to be small compared to n, i.e., $k/n \to 0$ as $n \to \infty$. Thus for all practical purposes the estimated median acts like a constant, i.e., as though it were the known population median.

Since the Y_i are highly correlated and typically show inhomogeneous variability it is appropriate that the generalized least squares approach should be used to fit the straight line to $(f_{\widehat{c}_k}(p_{i,.5,n}), Y_i)$, $i = 1, \ldots, k$. In order to do this we need to motivate and employ the proper weights for the extreme order statistics. These are given in the next section.

In order to avoid numerical stability problems during the least squares fitting it was decided to limit the estimated $\widehat{c}_k$ from below by -1.5. An extremal index of $c = -1.5$ indicates a very sharp upper bound on the sample distribution. Actually the density of $H_{c,\lambda,\delta}(y)$ becomes infinite at its upper support end point. The density behavior is like $(-y)^{-1/3}$ as $y \nearrow 0$. We feel that such cases are of lesser interest in practice. In such cases one gets a fairly quick idea from the sample that there is a hard upper bound to the sampled distribution. The main issue then is to locate that upper bound with high precision. Maybe more relevant would be to find the reason for the hard upper bound, and thus most likely also its exact value.

The above argument for a linearity transform was based on using $\gamma = .5$ and may not be valid when $\gamma \neq .5$. In practice it appears that the value $\widehat{c}_k$, that works for $\gamma = .5$, also seems to linearize the plotted points reasonably well for other values of γ. However, one also may try to fit a quadratic in $f_{\widehat{c}_k}(p_{i,\gamma,n})$ to capture some of the mild Y_i curvature that may be present.

4 Extreme Order Statistics Covariance Structure

Using the above approximate representation of the extreme p-quantiles of F we can use a one term Taylor expansion and get the approximate covariance structure of the extremes $Y_1, \ldots, Y_k$ from that of the uniform order statistics as follows. If $U_{(1)} \leq \ldots \leq U_{(n)}$ are the order statistics of a sample of size n from $U(0,1)$, then

$$E\left(U_{(i)}\right) = \pi_i = \frac{i}{n+1} \quad \text{and} \quad \operatorname{cov}\left(U_{(i)}, U_{(j)}\right) = \frac{\pi_i(1-\pi_j)}{n+2} \text{ for } i \leq j \; .$$

Assuming that the high sample order statistics ($i/n \approx 1$) can be represented approximately as

$$X_{(i)} = F^{-1}\left(U_{(i)}\right) \approx g\left(U_{(i)}\right) \approx g(\pi_i) + \left(U_{(i)} - \pi_i\right) g'(\pi_i)$$

with (for $p \approx 1$)

$$g(p) = \lambda + \delta f_c(p) \quad \text{and} \quad g'(p) = \frac{n\delta}{p} \left(\frac{1}{-n \log p}\right)^{c+1} ,$$

we get for $Y_i = X_{(n-i+1)}$ and $i \geq j$ (i/n and $j/n \approx 0$)

$$\begin{aligned} \sigma_{ij} = \operatorname{cov}(Y_i, Y_j) \;\; &\approx \;\; \frac{(1-\pi_i)\pi_j}{n+2} \cdot \left[\frac{n\delta}{1-\pi_i} \; \frac{1}{[-n\log(1-\pi_i)]^{c+1}}\right] \\ &\qquad\qquad \cdot \left[\frac{n\delta}{1-\pi_j} \; \frac{1}{[-n\log(1-\pi_j)]^{c+1}}\right] \\ &\approx \;\; \frac{jn^2\delta^2}{(n+1)(n+2)} \, \frac{1}{i^{c+1}j^{c+1}} \approx \delta^2 \, i^{-c-1} \, j^{-c} \; . \end{aligned}$$

The matrix $\mathbf{\Sigma} = (\sigma_{ij}/\delta^2)$, using the above approximations for σ_{ij} and with c replaced by its estimate $\widehat{c}_k$, can now be used in the generalized least squares fit of a straight line to $\left(f_{\widehat{c}_k}(p_{i,.5,n}), Y_i\right)$, $i = 1, \ldots, k$. For reasons to become clear when discussing the proper choice of k in the next section, it is worthwhile to spell out the notational details of this generalized least squares fit. Let

$$\mathbf{Y}' = (Y_1, \; \ldots, \; Y_k), \qquad \mathbf{f}' = (f_{\widehat{c}_k}(p_{1,.5,n}), \; \ldots, \; f_{\widehat{c}_k}(p_{k,.5,n})), \qquad \mathbf{1}' = \overbrace{(1, \; \ldots, \; 1)}^{k-\text{vector}} ,$$

$$\mathbf{e}' = (e_1, \ \ldots, \ e_k), \qquad \mathbf{b}' = (b_1, b_2), \qquad \text{and} \qquad \mathbf{X} = (\mathbf{1}, \mathbf{f}),$$

then the vector representation of the least squares model

$$Y_i = b_1 + b_2 f_{\widehat{c}_k}(p_{i,.5,n}) + e_i \ , \quad i = 1, \ldots, k$$

is

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{e} \ , \qquad \text{with} \qquad \mathbf{E}(\mathbf{e}) = \mathbf{0} \quad \text{and} \qquad \mathbf{var}(\mathbf{e}) = \delta^2 \mathbf{\Sigma} \ .$$

Note that $b_2 = \delta$, so that there is a structural link between the slope of the regression line and the covariance matrix.

To reduce this problem to the setting of ordinary least squares we find a symmetric matrix $\mathbf{C}$ such that $\mathbf{C}^t\mathbf{C} = \mathbf{C}\mathbf{C}^t = \mathbf{\Sigma}^{-1}$ and thus $\mathbf{C}^{-1}\mathbf{C}^{-t} = \mathbf{\Sigma}$ or $\mathbf{C}\mathbf{\Sigma}\mathbf{C}^t = \mathbf{I}_k$. For natural reasons one calls $\mathbf{C}$ also the square root of $\mathbf{\Sigma}^{-1}$. With

$$\mathbf{Z} = \mathbf{C}\mathbf{Y}, \qquad \mathbf{W} = \mathbf{C}\mathbf{X}, \qquad \text{and} \qquad \mathbf{d} = \mathbf{C}\mathbf{e}$$

we have

$$\mathbf{Z} = \mathbf{W}\mathbf{b} + \mathbf{d} \qquad \text{with} \qquad \mathbf{E}(\mathbf{d}) = \mathbf{0} \qquad \text{and} \qquad \mathbf{var}(\mathbf{d}) = \delta^2 \mathbf{I}_k \ ,$$

i.e., a regression model with uncorrelated errors, but same parameters $\mathbf{b}$ and δ. The ordinary least squares theory, applied to this transformed model, yields the following estimates for $\mathbf{b}$ and $\sigma^2 = \delta^2$:

$$\widehat{\mathbf{b}} = \left(\mathbf{W}^t\mathbf{W}\right)^{-1} \mathbf{W}^t\mathbf{Z} = \left(\mathbf{X}^t\mathbf{\Sigma}^{-1}\mathbf{X}\right)^{-1} \left(\mathbf{X}^t\mathbf{\Sigma}^{-1}\mathbf{Y}\right)$$

and

$$\widehat{\sigma}_k^2 = \frac{1}{k-2} \sum_{i=1}^{k} \left(Z_i - \widehat{b}_1 w_{i1} - \widehat{b}_2 w_{i2}\right)^2$$

where (w_{i1}, w_{i2}) is the i^{th} row of $\mathbf{W}$ and Z_i is the i^{th} component of $\mathbf{Z}$. The variance covariance matrix of $\widehat{\mathbf{b}}$ is

$$\mathbf{var}\left(\widehat{\mathbf{b}}\right) = \delta^2 \left(\mathbf{W}^t\mathbf{W}\right)^{-1} = \delta^2 \left(\mathbf{X}^t\mathbf{\Sigma}^{-1}\mathbf{X}\right)^{-1} \ .$$

In particular, $\text{var}(\widehat{b}_2) = \delta^2\kappa_k^2$, where κ_k^2 is the $(2,2)$-element in the matrix $(\mathbf{X}^t\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}$.

5 The Choice of k

An issue, that has not yet been addressed, is the number k of extreme data values to use in the estimation of c and in the extrapolation step. There is not much in the extreme value literature that addresses this issue. In a recent paper De Haan (1994) discusses the optimal choice for k. It is an issue of balancing the bias (for k too large) against the variability (for k too small) of the estimate $\widehat{c}_k$ of c. Unfortunately the optimal choice of k depends on the distribution F. The relation of the "optimal" $k = k(n)$ to n can take dramatically different forms, depending on F. This is of little help, because if one knew the form of F one could use that form to do the extrapolation. In choosing k one should keep in mind that in order to have consistency for $\widehat{c}_k$ one needs $k \to \infty$ and $k/n \to 0$. Of course, it is not clear how close to ∞ and 0 is close enough in practical cases and that may again depend on the underlying distribution F. We will take a very simpleminded approach for a good choice of k in that we probe as deep into the sample as we can, provided those sample extremes appear to agree with the domain of attraction assumption.

The procedure is based on the following observation. The estimate $\widehat{\sigma}_k^2$ from the generalized least squares fit will estimate δ^2. Since $\delta = b_2$ also represents the slope of the ideal line that is being estimated, we can use $T_k = \widehat{b}_{2,k}/\widehat{\sigma}_k$ as a criterion for judging the appropriateness of the choice of k. Here we added the subscript k on $\widehat{b}_2$ to emphasize its dependence on k.

As long as T_k remains near 1 we can increase k. Once T_k deviates too far from 1, we presumably have gone too deep into the sample. To provide a rational/heuristic yardstick for judging such deviations we apeal to the noncentral Student-t distribution. This is motivated by

$$\widehat{b}_{2,k} \approx N(\delta, \delta^2 \kappa_k^2) \ ,$$

appealing losely to the central limit theorem, and (with weaker backing, namely, simply by normal theory analogy)

$$\widehat{\sigma}_k^2 \approx \frac{\delta^2 \chi_{k-2}^2}{k-2} \ .$$

Thus

$$T_k = \frac{\widehat{b}_{2,k}}{\widehat{\sigma}_k} = \frac{\widehat{b}_{2,k}/\delta}{\widehat{\sigma}_k/\delta} \approx \frac{N(1,\kappa_k^2)}{\sqrt{\chi_{k-2}^2/(k-2)}} = \kappa_k \frac{N(1/\kappa_k, 1)}{\sqrt{\chi_{k-2}^2/(k-2)}} = \kappa_k \ t_{k-2,1/\kappa_k} \ ,$$

where $t_{f,\alpha}$ denotes a noncentral Student-t random variable with f degrees of freedom and noncentrality parameter α. Since the distribution of $\kappa_k t_{k-2,1/\kappa_k}$ is completely known, we can set probability limits for the ordinary variation of T_k assuming our domain of attraction assumption is valid. For lack of a better understanding of the sensitivity of T_k to deviations from our domain of attraction assumption we propose to set equal tailed probability limits.

In searching for the proper extrapolation depth k we will restrict ourselves to $k \in [K_1, K_2]$, where $K_1 = \max(6, \lfloor 1.3\sqrt{n} \rfloor)$ and $K_2 = 2\lfloor \log_{10}(n)\sqrt{n} \rfloor$. These choices, although somewhat arbitrary, satisfy the requirement that $K_2/n \longrightarrow 0$ as $n \longrightarrow \infty$. Through a search algorithm, described in detail below, we replace the original interval $[K_1, K_2]$ by a strictly decreasing sequence of subintervals. Each subsequent subinterval is again denoted by $[K_1, K_2]$.

We also fix a grid resolution k_{res} below which it is felt not worthwhile to refine the search for k, i.e., when the search for k is narrowed down to an interval $[K_1, K_2]$ with width $K_2 - K_1 \leq k_{\text{res}}$, then the search is stopped. See Table 1 for some numerical values of K_1, K_1/n, K_2, K_2/n, and k_{res} for a selection of sample sizes.

Table 1

n	K_1	K_1/n	K_2	K_2/n	k_{res}	k_Δ	k_{span}
100	13	.130	40	.400	1	1	5
500	29	.058	120	.240	1	1	11
1000	41	.041	188	.188	1	2	15
5000	91	.018	522	.104	3	4	35
10000	130	.013	800	.080	5	7	50

As k gets larger, it becomes computationally more and more tedious to compute the square root, $\mathbf{C}$, of $\mathbf{\Sigma}^{-1}$. One may therefore set an upper limit, M, on the dimensionality of $\mathbf{\Sigma}^{-1}$. In individual applications we may let M be fairly large. However, when trying to simulate many scenarios using a large number of replications to get accurate assessments of coverage rates, one may want to be content with lower values of M, say $M = 50$. This issue typically arises only when the sample sizes are large, say in the thousands.

Maintaining the bound M on the dimensionality of $\mathbf{\Sigma}^{-1}$ can be accomplished fitting the regression line to a thinned out subset of the points $(f_{\widehat{c}_k}(p_{i,.5,n}), Y_i)$, $i = 1, \ldots, k$. This subset is of size at most M, i.e., if $k \le M$ we fit all these k observations, and when $k > M$ we will thin that set of points, so that it is of size M. Out of many possible thinning strategies we have employed the following. Given a value $k > M$ we choose from $\{1, 2, \ldots, k\}$ the following subsequence of approximate length M:

$$i_1 = 1, \;\; i_2, \;\; \ldots, \;\; i_M$$

with spacings

$$i_2 - i_1 = 1 + \Delta \cdot 1, \quad i_3 - i_2 = 1 + \Delta \cdot 2, \quad \ldots, \quad i_M - i_{M-1} = 1 + \Delta \cdot (M-1) \; .$$

By concatenating these spacings we get

$$i_j - i_1 = j - 1 + \Delta \cdot \frac{j(j-1)}{2} \; , \qquad \text{i.e.,} \qquad i_j = j + \Delta \cdot \frac{j(j-1)}{2} \; .$$

The requirement $i_M \le k$ yields

$$\Delta \le \frac{2(k-M)}{M(M-1)} \; .$$

For i_M to come as close as possible to k we take

$$\Delta = \frac{2(k-M)}{M(M-1)}$$

and, to maintain the integer character of the subsequence, we take

$$i_j = \left\lfloor j + \Delta \frac{j(j-1)}{2} \right\rfloor \; .$$

Note that the thinned sequence is strictly increasing, starts at $i_1 = 1$ and ends at $i_M = k$. The latter may not happen due to rounding problems, in which case we add k to the thinned sequence, i.e., increase M by one. Although the sequence $\{1, 2, \ldots, k\}$ is thinned for purposes of fitting a regression line (or a quadratic) no such thinning takes place for estimating c, i.e., all k extremes $Y_1, \ldots, Y_k$ are used in computing $\widehat{c}_k$.

The actual search for $k \in [K_1, K_2]$ proceeds as follows. Starting at $k = K_1$ we proceed in steps of size $k_\Delta = \max(1, \lfloor .07\sqrt{n} \rfloor)$, i.e., $k_i = K_1 + (i-1)k_\Delta$,

$i = 1, \ldots 11$, and $k_{12} = K_2$ (see Table 1 for some sample values of k_Δ). For each of these 12 k-values we obtain the weighted least squares fit of a straight line to the points $(f_{\widehat{c}_k}(p_{i,.5,n}), Y_i)$, $i = 1, \ldots, k$, or, if k is larger than 50, of a suitably thinned subset of these points, and calculate the noncentral Student-t criterion T_k. This criterion is compared against two intervals, namely

$$I_{0k} = \left[\kappa_k t_{k-2,1/\kappa}\,(.25)\,, \quad \kappa_k t_{k-2,1/\kappa}\,(.75)\right]$$

and

$$I_{1k} = \left[\kappa_k t_{k-2,1/\kappa}\,(.025)\,, \quad \kappa_k t_{k-2,1/\kappa}\,(.975)\right] \;.$$

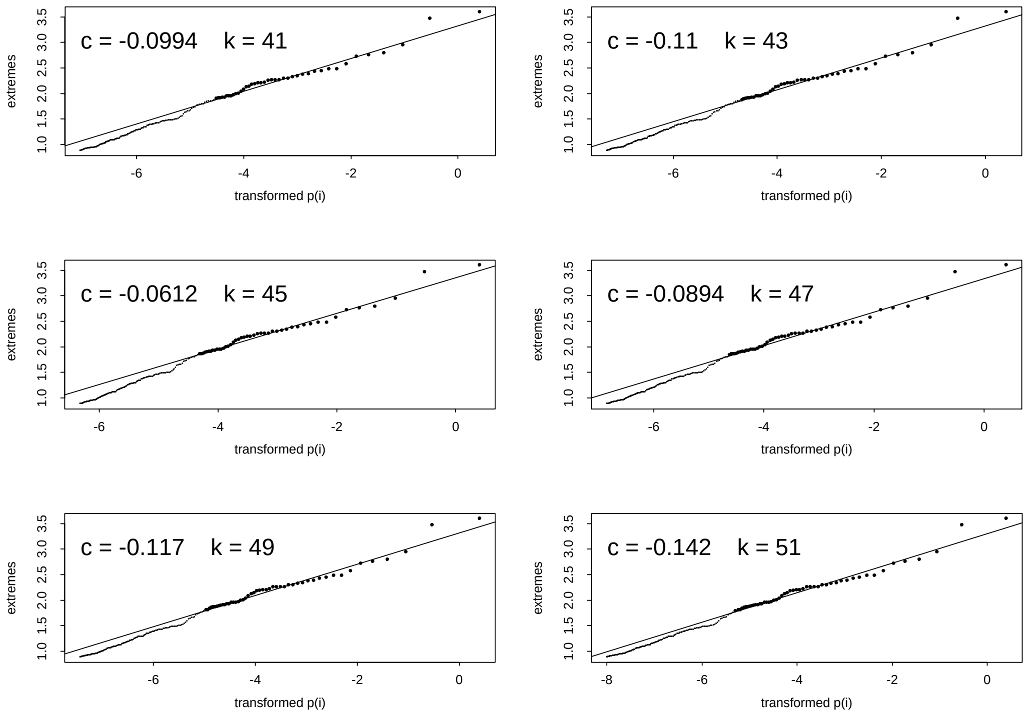
These two intervals contain respectively 50% and 95% of the involved noncentral Student-t distribution. The idea is to stop searching as soon as T_k falls outside of I_{1k}. Unfortunately the last k prior to that is not necessarily a good choice for k either, because it often is nearly outside of I_{1k}. This motivates the use of the other interval I_{0k}. We will consider those k values as "good" choices for which T_k falls inside of I_{0k}. As we go to larger and larger choices of k and while we still have $T_k \in I_{1k}$, we will typically find stretches of k values for which $T_k \in I_{0k}$ interspersed with stretches for which $T_k \notin I_{0k}$. We will keep track of the longest contiguous stretch of "good" choices of k.

For the above 12 initial values of k there usually is a large gap between $k_{11} = K_1 + 10k_\Delta$ and $k_{12} = K_2$. If among the first 11 values of k we always have $T_k \in I_{1k}$, then we continue the search by treating $K'_1 = K_1 + 10k_\Delta$ as our new left search interval endpoint and $K'_2 = K_2$. Another contingency for arriving at a narrower search interval $[K'_1, K'_2]$ occurs when for some first k_i we find $T_k \notin I_{1k}$. Then we take $K'_1 = k_{i-1}$ and $K'_2 = k_i$.

Again we try out 12 values for k, namely $k_i = K'_1 + 2(i-1)k_\Delta$, $i = 1, \ldots, 11$ and $k_{12} = K'_2$. Note that the incrementing value is now $2k_\Delta$. We continue in this fashion to a grid of 12 values over $[K''_1, K''_2]$, with the incrementing value increased to $3k_\Delta$, until the span of the decreased search interval is $\leq k_{\text{res}}$ or the span of good values is long enough, namely at least $k_{\text{span}} = \max(2, \lfloor .5\sqrt{n} \rfloor)$ (see Table 1 for some sample values of k_{span}).

Once we stop the search we take the best k out of the longest stretch of good k values. Here the best k is that which yields a T_k closest to the median (≈ 1) of the involved noncentral Student-t distribution. The metric of closeness is p_Δ and is described below. Although the above essentially describes the search strategy, there are few more minor details, which will not be discussed here.

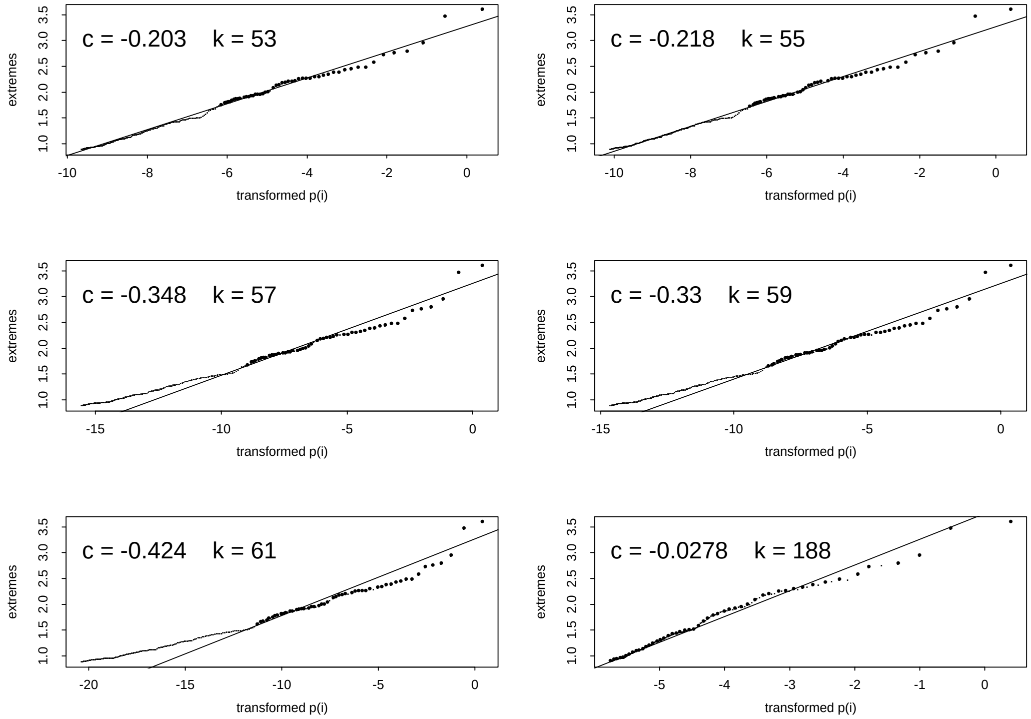
Figure 2a: Normal Tail Extrapolation
sample of size $\mathbf{n = 1000}$ from $\mathbf{N(0,1)}$
extrapolation for first 12 choices of $\mathbf{k \in [41, 188]}$
using at most 50 points for weighted least squares fit



Figures 2a-d illustrate this search for the proper extrapolation depth k using the upper tail of a sample of size $n = 1000$ drawn from a standard normal population. Initially the chosen range for k is $[K_1, K_2] = [41, 188]$. For $n = 1000$ we have $k_\Delta = 2$. Thus the first 12 trial values of k are $41, 43, 45, \ldots, 61, 188$. The corresponding weighted least squares fits and the respective estimated value of c are shown in Figures 2a.

In each of these plots all 188 high extremes $Y_1, \ldots, Y_{188}$ are plotted as tiny dots against the corresponding transformed $p_{i,.5,n}$-values $f_{\widehat{c}_k}(p_{i,.5,n})$. The fat dots represent those points that were used for fitting the weighted least squares line. Note that the abscissa scale changes from plot to plot, since the depth k for

Figure 2a: Normal Tail Extrapolation (continued)
extrapolation for first 12 choices of k $\in [41, 188]$



computing $\widehat{c}_k$ and thus the transformation $f_{\widehat{c}_k}(\cdot)$ changes.

The number of fat dots, used for weighted least squares fitting in each plot, is always 50 or less. In the 12th plot for $k = 188$ one can clearly see the effect of thinning out the points that are fitted. Note the (unfitted) small dots interspersed with the fat dots. Observe that the fitted line is above the bulk of the data in the first six plots and for $k = 57, 59, 61$ the situation is reversed.

Figure 2b: Noncentral t Criteria for Plots in Figure 2a

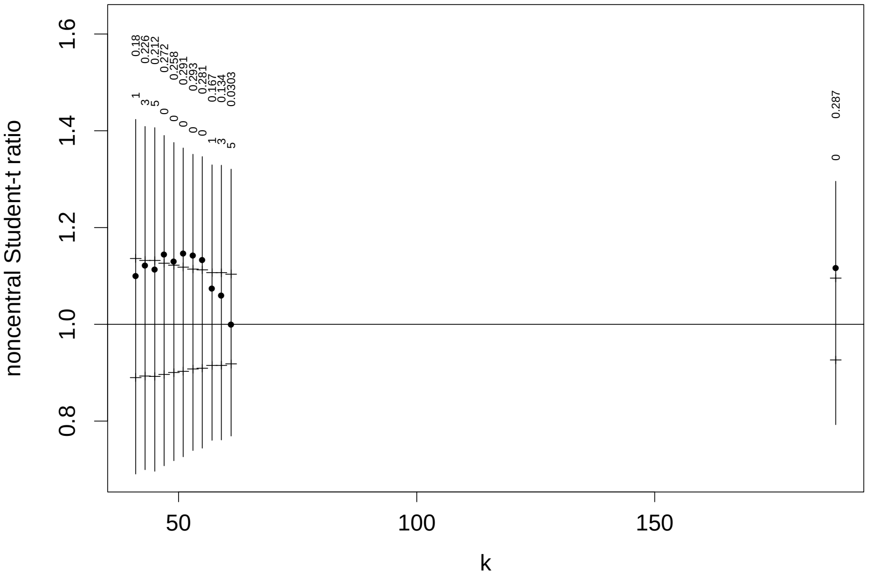


Figure 2c: Normal Tail Extrapolation
sample of size $\mathbf{n = 1000}$ from $\mathbf{N(0,1)}$
extrapolation for next 6 choices of $\mathbf{k \in [61, 188]}$
using at most 50 points for weighted least squares fit

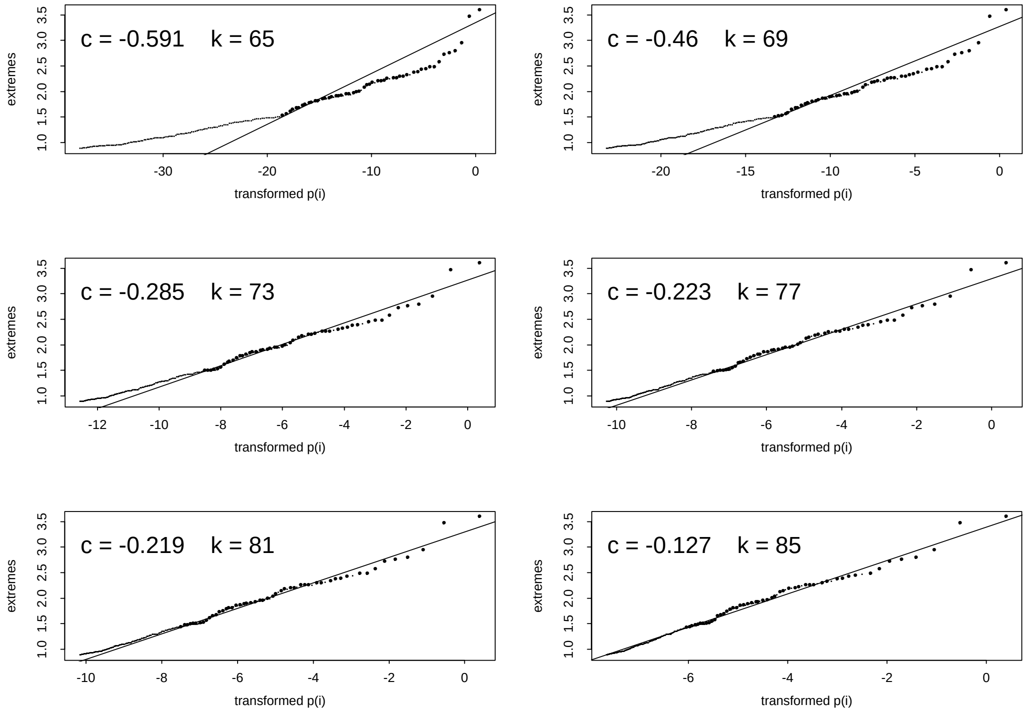
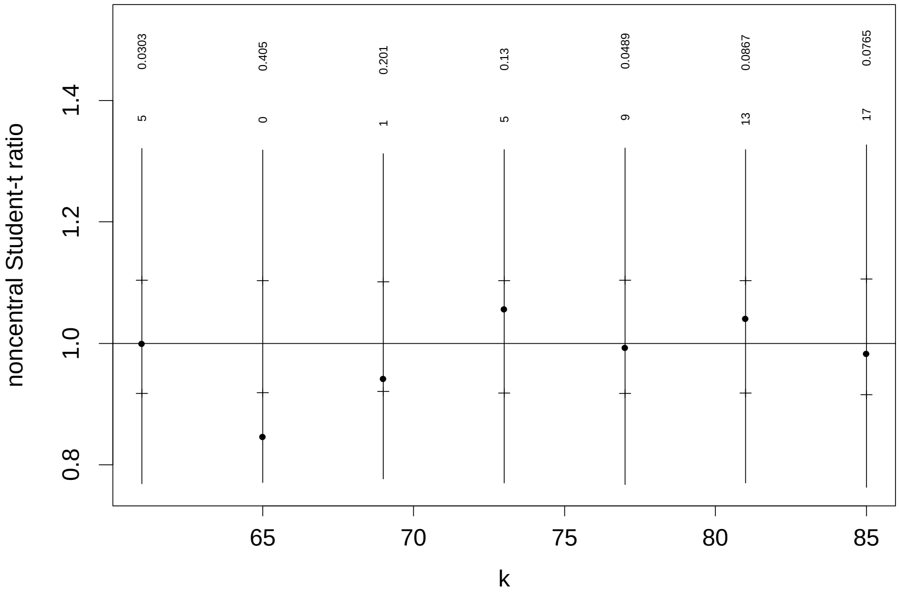


Figure 2b gives the plot of the 12 corresponding ratios T_k. The vertical lines represent the 95% intervals I_{1k} and the narrower marks on them indicate the 50% intervals I_{0k}. Above each vertical line are two numbers, a count and a probability. The count is the number of consecutive "good" T_k values up to that value k, i.e., for which we have $T_k \in I_{0k}$. The probability indicates the closeness of T_k to its median. That closeness probability is computed as follows:

$$p_\Delta = \left| P\left(T_k \le t_k^\star\right) - \frac{1}{2} \right| ,$$

where $t_k^\star$ is the observed value of T_k. Note that $p_\Delta = 0$ just then when $t_k^\star$ falls on the median of the distribution for T_k.

Figure 2d: Noncentral t Criterion for each plot in Figure 2c

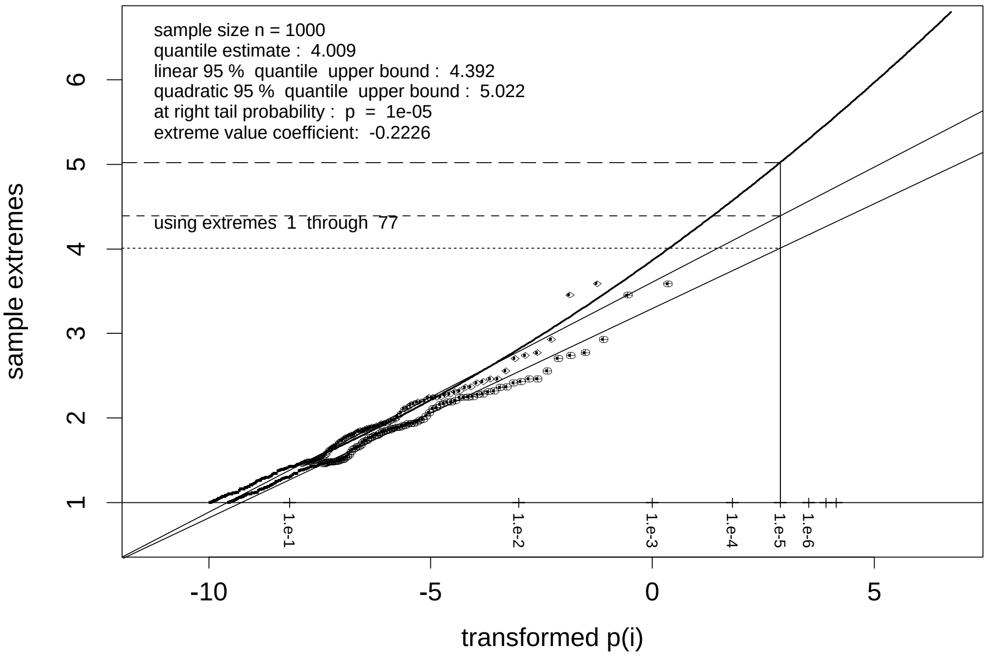


Note that in Figure 2b we start out with a stretch of 5 "good" values, although technically only 3 of these 5 were checked. This is followed by 5 checked values outside their respective I_{0k} intervals and after that we run into another stretch of 5 "good" k values. A stretch of 5 is not sufficient to end the search for k. For that to happen we would need a stretch of length at least $k_{\text{span}} = \lfloor .5\sqrt{1000} \rfloor = 15$.

The next search interval is thus $[K_1, K_2] = [61, 188]$. The least squares fits for the first 6 values of k after 61 are shown in Figure 2c together with the estimates for c. This is followed by Figure 2d which shows again the observed T_k values together with the respective intervals. Over the stretch from $k = 69$ to $k = 85$ we accumulate a count of 17 "good" k values. Thus we stop in our search. Over that span of 17 "good" k values we find that p_Δ is smallest for $k = 77$, which thus becomes our final choice.

Finally, Figure 3 shows the actual quantile extrapolation for the chosen sample

Figure 3: Normal Right Tail Quantile Extrapolation from sample of size $\mathbf{n = 1000}$ from $\mathbf{N(0,1)}$, estimates and 95% upper confidence bounds for $\mathbf{x_p = x_{1-q}}$ with $\mathbf{q = 10^{-5}}$



tail depth of $k = 77$. The target in this figure is the quantile $x_p = x_{1-q}$ for tail probability $q = 10^{-5}$. Two linear extrapolations are shown. The lower line corresponds to a 50% confidence level and thus results in a median unbiased estimate of x_p. The upper line extrapolates 95% upper confidence bounds for x_p.

This is accomplished by plotting the same data $Y_1, \ldots, Y_k$ twice, namely against the two transformed scales $f_{\widehat{c}_k}(p_{i,.5,n})$ and $f_{\widehat{c}_k}(p_{i,.95,n})$, respectively. In both cases we fit straight lines by the method of weighted least squares, as outlined previously. The extrapolation then proceeds by marking $f_{\widehat{c}_k}(p)$ (for the desired $p = 1 - q$) on the abscissa, moving up vertically to intercept the respective fitted lines, and moving left from those intercepts to find the estimate/confidence bound on the ordinate axis, as indicated by the dashed and dotted lines.

Also shown is a quadratic extrapolation curve for extrapolating 95% upper confidence bounds for x_p. It too is fit by weighted least squares. If the curvature coefficient of the quadratic is such that the extrapolation becomes more conservative than the linear extrapolation, we force the quadratic to reduce to the linear extrapolation.

There is considerable graphical appeal in this extrapolation procedure. Not only can one see upfront how well the fitted line fits the data, but one also gets an appreciation of how far out from the data one tries to extrapolate. To get a better sense of the extrapolation a probability scale with a few order of magnitude tick marks is superimposed.

6 Confidence Bounds for Tail Probabilities

Upper confidence bounds for quantiles x_p are intrinsically related to lower confidence bounds for $F(t)$ for a given t. By subtraction from 1 such lower bounds become upper bounds to the tail probability $1 - F(t)$. Similarly, lower bounds for x_p relate to upper bounds for $F(t)$, and, by subtraction from 1, to lower bounds to $1 - F(t)$. This makes it possible to get upper and lower bounds for left or right tail probabilities. We will describe this relationship in generic terms only for upper tolerance bounds and lower confidence bounds for $F(t)$. It is assumed that F is continuous.

Let $\widehat{U}(p, \gamma)$ be an approximately $100\gamma\%$ upper confidence bounds for x_p, i.e. $\widehat{U}(p, \gamma) = \widehat{U}(X_1, \ldots, X_n;\ p, \gamma)$ is a function of the sample with the following property

$$P_F(\widehat{U}(p, \gamma) \geq x_p) \approx \gamma \ .$$

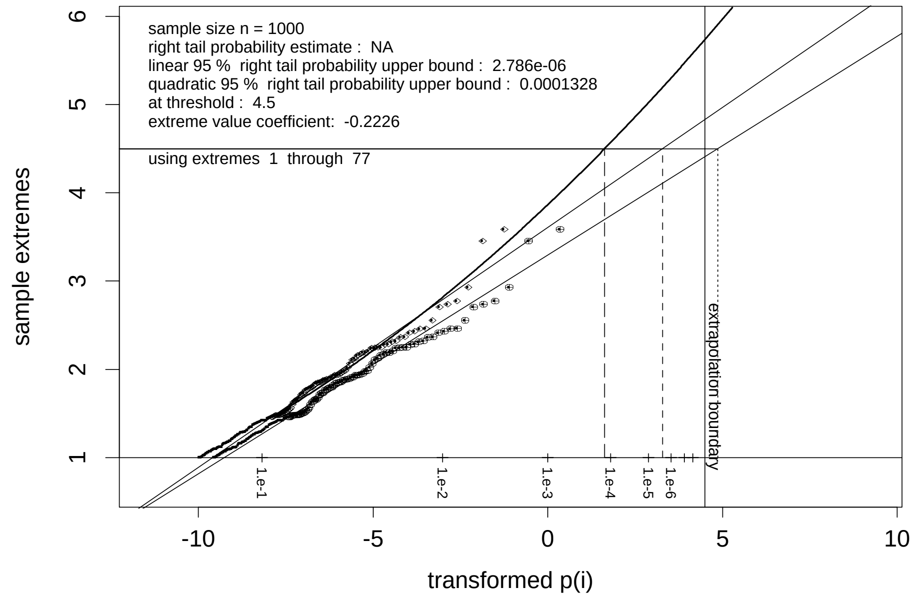
Here F denotes the distribution from which the sample is drawn and x_p is the p–quantile of F. Assume that $\widehat{U}(p, \gamma)$ is strictly increasing in p and suppose further that $\widehat{U}(p, \gamma) = t$ has a unique solution $p = \widehat{p} = \widehat{p}(t, \gamma) = \widehat{p}(X_1, \ldots, X_n; t, \gamma)$, then

$$\widehat{p}(t, \gamma) \leq p \quad \Longleftrightarrow \quad t \leq \widehat{U}(p, \gamma) \ .$$

For t and p such that $t = F^{-1}(p) = x_p$ ($\Longrightarrow\ F(t) = p$ if F is continuous) we then have

$$P_F(\widehat{p}(t, \gamma) \leq F(t)) = P_F(x_p \leq \widehat{U}(p, \gamma)) \approx \gamma \ .$$

Figure 4: Normal Right Tail Probability Extrapolation from sample of size $\mathbf{n = 1000}$ from $\mathbf{N(0,1)}$, estimates and 95% upper confidence bounds for $\mathbf{q_0 = P(X \geq t_0)}$ with $\mathbf{t_0 = 4.5}$



For the nonparametric tolerance bounds this inversion to confidence bounds for $F(t)$ is easy enough by simply extrapolating from t on the Y–scale to the transformed p_i scale to obtain the lower confidence bound $\widehat{p}(t, \gamma)$ for $F(t)$ and the upper confidence bound
$1 - \widehat{p}(t, \gamma) = \widehat{q}(t, \gamma)$ for $1 - F(t)$.

Using the same normal data as before this idea is illustrated in Figure 4 for the median unbiased estimate and confidence upper bound for the right tail probability $P(X \geq t_0) = 3.4\ 10^{-6}$ corresponding to the threshold $t_0 = 4.5$. The estimate is given as NA, which results from the fact that the estimate for c is negative, namely $\widehat{c}_{77} = -.2226$. This is discussed in more detail below.

Note that the 95% upper bound of $2.786\ 10^{-6}$ understates the target of $3.4\ 10^{-6}$

by 18%. This is not too bad considering that we are trying to extrapolate to a tail probability of magnitude 3.4 10^{-6} based on only 1000 observations.

In the context of estimates and confidence bounds for tail probabilities we should point out the following possibility. For $c < 0$ the implied (reflected) Weibull distribution has a hard upper bound t_u. If we choose a threshold $t \geq t_u$, then it is not possible to perform the backtransformation to the extrapolated value of p corresponding to t. If the threshold ordinate t leads to the abscissa value f_t, then the formal backtransform, from f_t to p_t, would yield

$$f_t = \frac{(-\log p_t)^{-c} - 1}{c} \qquad \Longrightarrow \qquad p_t = \exp\left(-\frac{1}{n}\left(cf_t + 1\right)^{-1/c}\right) \ ,$$

which runs into problems when $cf_t + 1 \leq 0$, i.e., when $f_t \geq -1/c$. We thus can view $-1/c$ as the extrapolation boundary on the abscissa. In that case we could presumably infer that the right tail probability corresponding to t is zero. However, we should keep in mind that the extrapolation boundary itself is estimated and a zero chance is a rather strong statement. We thus prefer to designate such cases by the symbol NA as in the estimation case of Figure 4.

7 Simulations

Although the proposed method may have many attractive features, there is a great need to build an experience base for using this method. This can be done by performing extensive simulations, namely by sampling a variety of different distributions, using a range of sample sizes, and trying different extrapolation depths.

There should be no illusions concerning the results of such simulations. They will be of mixed quality and are presented in Scholz and Tjoelker (1995).

References

Castillo, E. (1988). *Extreme Value Theory in Egineering.* Academic Press, San Diego.

Michael, J.R. and Schucany, W.R. (1986). Chapter 11 in *Goodness-Of-Fit Techniques* edited by R.B. D'Agostino M.A. Stephens, Dekker, New York.

De Haan, L. (1994). "Extreme value statistics." *Extreme Value Theory and Applications* eds. J. Galambos, J. Lechner, and E. Simiu, Kluwer Academic Publishers, Dordrecht.

Dekkers, A.L.M., Einmahl, J.H.J., and De Haan, L. (1989). "A moment estimator for the index of an extreme-value distribution." *The Annals of Statist.* **17**, 1833-1855.

De Haan, L. (1971), *On Regular Variation and its Application to Weak Convergence of Sample Extremes,* Mathematical Centre Tract, **32**, Amsterdam: Mathematisch Centrum.

Filliben, J.J. (1975). "The probability plot correlation coefficient test for normality." *Technometrics* **17**, 111-117.

Hoaglin, D.C. (1983). "Letter values: A set of selected order statistics." in *Understanding Robust and Exploratory Data Analysis*, eds. D.C. Hoaglin, F. Mosteller, and J.W. Tukey, John Wiley & Sons, New York, 33-57.

Scholz, F.W. and Tjoelker, R.A. (1995). "Nonparametric Tail Extrapolation, Simulation Results." *ISSTECH-95-015*, Boeing Information & Support Services, P.O. Box 3707, MS 7L-22, Seattle WA 98124-2207.