
The construction of confidence intervals for frequency analysis using resampling techniques: a supplementary note

H.F.P. van den Boogaard¹ and M.J. Hall²

¹WL|Delft Hydraulics, PO Box 177, 2600 MH Delft, The Netherlands

²UNESCO-IHE Institute for Water Education, PO Box 3015, 2601 DA Delft, The Netherlands

Email for corresponding author: m.hall@unesco-ihe.org

Abstract

In a recent contribution, Hall *et al.* (2004) examined the use of the Bootstrap resampling technique as a means of constructing confidence limits for the quantiles of the (two-parameter) Gumbel and the (three-parameter) Weibull distributions. Particular emphasis was placed on the behaviour of sample sizes of the order of 30, which are typical of those encountered in hydrological frequency analysis. The resampled confidence limits obtained for the Gumbel distribution were found to be comparable with those based upon a well-known theoretical approximation. However, those for samples of size 30 from the Weibull distribution were shown to be more problematical, with the results dependent upon the skewnesses of the resampled distributional parameters. For a further and more quantitative assessment of the suitability of Bootstrap resampling for constructing confidence intervals, so-called coverage rates were evaluated for the Weibull distribution in a supplementary study. The results show a satisfactory performance when using the percentile method but do not really mitigate the conclusion of the original study that resampled confidence limits should be employed with caution when sample sizes are of the order of 30.

Keywords: Bootstrap, Jack-knife, frequency analysis, maximum likelihood method, maximum product of spacings method, confidence intervals, coverage rates.

Introduction

Resampling techniques such as the Bootstrap and the Jack-knife (Efron and Tibshirani, 1993; Davison and Hinkley, 1997) are data-based simulation methods for statistical inference. In effect, resampling creates an *ensemble* of data sets, each of which is a replication from the original sample. The *Jack-knife algorithm* generates the new samples by deleting one (or more) specific data points, while the *Bootstrap algorithm* creates the new data sets by sampling with replacement. In the latter case, one or more data points may be absent, and one or more may be repeated once or more than once in any resampled data set. For each resampled data set the actual statistic is evaluated and, from this ensemble, an estimate of the statistic and its uncertainty (spread, quantiles, or confidence interval) can be derived.

Despite their considerable attraction in providing estimates of standard errors avoiding the need for theoretical calculation and manipulation of complicated forms of estimator, resampling techniques appear to have attracted

only comparatively limited attention in hydrology and water resources engineering (see Hall *et al.*, 2004, and the references therein). In a recent contribution (Hall *et al.*, 2004), the construction of confidence limits to quantile estimates obtained from two- and three-parameter frequency distributions was investigated using Bootstrap resampling techniques. The *Gumbel* (Extreme Value type I) distribution was adopted as an example of a typical two-parameter distribution for which the standard errors of quantile estimates can be derived using a well-known theoretical approximation. Monte Carlo methods were applied to generate different sample sizes from a Gumbel population and their parameters were estimated using the *Method of Maximum Likelihood* (MLH). The Bootstrap was then employed to compute standard errors of quantile estimates which could be compared with those generated from the aforementioned theoretical approximation, particular attention being paid to the number of resamples required for different sample sizes. Confidence levels were computed

using both the *Gaussian method*, assuming that the resampled quantiles were normally distributed, and the *percentile method*, for which the appropriate limiting values were read from a ranked list of values provided by all resamples. Similar results were achieved with both methods, largely because the location and scale parameters of the resamples were approximately normally distributed, and were found to produce confidence limits that were comparable with the theoretical approximation.

These results were sufficiently encouraging for the same approach to be applied to the three-parameter *Weibull* distribution, for which estimates for standard errors of quantile estimates are difficult to determine by analytical methods. In this case, three initial sample sizes were generated by Monte Carlo methods for two distributions with location parameter of zero and a scale parameter of unity, but with shape parameters of 0.75 and 2 respectively. The parameters of the resamples were estimated by the *Maximum product of spacings* (MPS) method (Cheng and Amin, 1983), which has been found to be superior to MLH with distributions such as the Weibull with a shifted origin as a location parameter. In such cases, and especially when the distribution is J-shaped, MPS may provide consistent estimates where MLH is bound to fail (Cheng and Amin, 1983). A comparison was made of the sample distribution function (SDF), based on MPS estimates of the parameters of the original sample, and the Bootstrap ensemble distribution function (EnDF), obtained by averaging 1000 Bootstrap resamples, along with their 95% confidence limits. For discussion purposes, the latter were computed with both the Gaussian approximation and the percentile method.

For both values of the shape parameter and an initial sample size of 30, the confidence limits displayed a marked divergence as the probability of non-exceedance increased from 0.99 to 0.9999. There was little agreement between the limits given by the Gaussian method and the percentile method, and the lower bound of the (for example 95%) confidence interval limit of the Gaussian method may even be *decreasing* for an *increasing* probability of non-exceedance.

This discrepancy was explained partly by the distributional properties of both the quantiles for a given plotting position variate and the resampled parameters of the parent distributions. For an initial sample size of 100 and both values of the shape parameter, much better agreement was discernable between the SDF and the EnDF, and the confidence levels obtained with the Gaussian and percentile methods. This agreement improved even further when the initial sample size was increased to 400. Unfortunately, the confidence limits obtained by Bootstrapping with the Weibull distribution with sample sizes of 30, which are more

representative of the data sets generally available for hydrological frequency analysis, must be subject to critical evaluation. In any case, the percentile method is to be preferred to the Gaussian approach.

During the review of the original paper, an anonymous referee commented that the conclusions concerning the behaviour of the confidence levels constructed by the different procedures would be reinforced if consideration were given to their coverage properties. Unfortunately, this extra information could not be made available in time to meet the deadline for publication of the paper. This supplementary note is therefore offered by way of an extension to the original discussion. The following sections therefore describe the methodology adopted for the generation of coverage rates, along with a summary of the results and some concluding remarks.

Coverage properties

Coverage properties can be computed using the following procedure (see, e.g. Stark, 2003). Firstly, a random sample of size N , $X^{(1)} = (x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, \dots, x_N^{(1)})$ is generated from a parent distribution, $F(\cdot | \Theta)$. The chosen estimation technique (e.g. MLH, MPS) is then applied to find estimates $\hat{\Theta}^{(1)}$ and $\hat{x}_Q^{(1)} = F^{-1}(Q | \hat{\Theta}^{(1)})$ for the population parameters Θ and quantiles $x_Q = F^{-1}(Q | \Theta)$ for one or more probabilities of non-exceedance $Q \in (0,1)$. In addition, the γ -confidence intervals are computed for one or more confidence levels γ . In the present case, these confidence intervals are based on the standard Bootstrap resampling scheme using either the percentile method or the Gaussian method. Confirmation is then sought as to whether the true quantile x_Q (from the parent distribution) is contained within the γ -confidence interval, $C_{Q,\gamma}^{(1)}$, for the Q th quantile from sample $X^{(1)}$. This procedure is repeated many times, i.e. for an ensemble of independently-created samples, $X^{(1)}, X^{(2)}, X^{(3)}, \dots, X^{(K)}$, each of the same (fixed) size N drawn from the same parent distribution, $F(\cdot | \Theta)$. In this way, an ensemble of confidence intervals, $\{C_{Q,\gamma}^{(k)}\}_{k=1}^K$ is obtained. From this ensemble, a *coverage rate*, $\hat{\rho}_{Q,\gamma}$, is determined for every combination of a quantile, Q , and a confidence level γ . This coverage rate is defined as the fraction of the K experiments for which the true quantile x_Q is (*covered*) in the computed confidence interval according to:

$$\hat{\rho}_{Q,\gamma} = \frac{1}{K} \sum_{k=1}^K I_k \quad \text{where } I_k = 1 \text{ if } x_Q \in C_{Q,\gamma}^{(k)},$$

$$\text{and } I_k = 0 \text{ if } x_Q \notin C_{Q,\gamma}^{(k)} \quad (1)$$

For large K , the rate $\hat{\rho}_{Q,\gamma}$ should converge (or should at least be close) to the true confidence level γ . The agreement

between $\hat{\rho}_{Q,\gamma}$ and γ therefore provides a quantitative validation of the accuracy of the procedures for the computation of the confidence intervals.

In the present assessment of the coverage rates of the confidence intervals of quantiles, a three parameter Weibull distribution was chosen for the parent distribution $F(\cdot|\Theta)$, i.e.

$$F(x|\Theta) = 1 - \exp\left(-\left(\frac{x-x_0}{\sigma}\right)^\alpha\right) ;$$

$$\Theta := (\sigma, \alpha, x_0), \sigma > 0, \alpha > 0 ; x \geq x_0 \quad (2)$$

An ensemble of $K = 10,000$ independent samples $\{X^{(k)}\}_{k=1}^K$ of size $N = 30$ was generated, and MPS was used as estimation technique. For each sample $X^{(k)}$, $B = 1000$ Bootstrap resamples were created to derive the confidence intervals $C_{Q,\gamma}^{(k)}$ for five variations of the quantile Q (0.5, 0.9, 0.99, 0.999, 0.9999) and five variations of the confidence level γ (50%, 80%, 90%, 95%, 99%), giving a set of 25 coverage rates $\hat{\rho}_{Q,\gamma}$. These experiments were done three times, each with a different value of the Weibull shape parameter ($\alpha = 0.75, 1, 2$), but with the location and scale parameters held at zero and one respectively ($x_0 = 0, \sigma = 1$).

With regard to the accuracy that is achieved in these experiments for the coverage rates $\hat{\rho}_{Q,\gamma}$ of Eqn. (1), for a given confidence level γ , the estimator $\hat{\rho}_{Q,\gamma}$ is based on K independent Bernoulli trials. Under the hypothesis that the procedure for the assessment of confidence intervals is correct, the probabilities for ‘success’ and ‘failure’ for each trial are γ and $1 - \gamma$ respectively. In this way, the total number of successful coverages (i.e. $K \cdot \hat{\rho}_{Q,\gamma}$) follows a binomial distribution. As a result, the expectation value for

the coverage rate $\hat{\rho}_{Q,\gamma}$ will be γ and its spread is given by $\sqrt{\frac{\gamma(1-\gamma)}{K}}$. For the present $K=10000$, these spreads are then 0.5% for $\gamma=50\%$, 0.4% for $\gamma=80\%$, 0.3% for $\gamma=90\%$, 0.2% for $\gamma=95\%$, and 0.1% for $\gamma=99\%$. When the 2.5% and 97.5% quantiles of the binomial distribution are taken as limits for the range of likely values of $\hat{\rho}_{Q,\gamma}$ this would lead to the ranges $\left[\gamma - 1.96 \cdot \sqrt{\frac{\gamma(1-\gamma)}{K}}, \gamma + 1.96 \cdot \sqrt{\frac{\gamma(1-\gamma)}{K}} \right]$ when using a Gaussian approximation, which for $K=10000$ is sufficiently accurate, even for $\gamma=99\%$ where the binomial distribution may be highly non-symmetrical when K is small (and much less than 10000). Indeed, for $\gamma=99\%$ and $K=10000$, the exact 2.5%-97.5% quantile range is [98.79%, 99.19%] while the Gaussian approximation yields [98.80%, 99.20%].

The results of the three experiments with different shape parameter for the Weibull distribution are presented in Tables 1–3. The coverage rates (given in the form of a percentage) using the percentile method are seen to provide a satisfactory performance, particularly with the shape parameter of 0.75 which applies to a J-shaped distribution. However, for all three distributions, the coverage rates for the $Q = 0.9$ quantile tend to underestimate the confidence level. The reason for this particular discrepancy is unclear, but may well be related to the small sample size of the experiments. Indeed, a further set of tests for a sample size of 100 (not reported here) did not display the same discrepancy, providing better coverage rates for the 0.9 quantile as well as the other four.

Tables 1–3 also indicate that the coverage properties using the percentile method are notably better than those of the Gaussian approximation, particularly for the 80 and 90% confidence levels and for the larger quantiles. Once again,

Table 1. Coverage rates for the confidence intervals of five quantiles of the Weibull distribution and five confidence levels γ , using both the percentile (P) and the Gaussian (G) methods. Results are presented for values of the Weibull shape parameter α of 0.75.

| Quantile | Conf. level method | Confidence level γ (per cent) | | | | |
|----------|--------------------|--------------------------------------|------|------|------|------|
| | | 50 | 80 | 90 | 95 | 99 |
| 0.5 | P | 50.4 | 80.2 | 89.9 | 94.8 | 98.8 |
| | G | 51.1 | 81.5 | 90.5 | 94.6 | 97.9 |
| 0.9 | P | 45.5 | 74.7 | 85.5 | 91.1 | 96.7 |
| | G | 44.4 | 75.4 | 86.4 | 91.6 | 96.2 |
| 0.99 | P | 46.2 | 77.3 | 88.0 | 93.7 | 98.5 |
| | G | 46.0 | 81.0 | 91.0 | 94.6 | 97.1 |
| 0.999 | P | 47.4 | 79.2 | 89.2 | 94.6 | 99.1 |
| | G | 48.2 | 85.0 | 93.2 | 95.7 | 97.5 |
| 0.9999 | P | 48.4 | 79.9 | 89.8 | 95.1 | 99.3 |
| | G | 49.9 | 87.6 | 94.5 | 96.2 | 97.8 |

Table 2. Coverage rates for the confidence intervals of five quantiles of the Weibull distribution and five confidence levels γ , using both the percentile (P) and the Gaussian (G) methods. Results are presented for values of the Weibull shape parameter α of 1.0.

| Quantile | Conf. level method | Confidence level γ (per cent) | | | | |
|----------|--------------------|--------------------------------------|------|------|------|------|
| | | 50 | 80 | 90 | 95 | 99 |
| 0.5 | P | 51.1 | 80.9 | 90.5 | 95.2 | 98.9 |
| | G | 51.7 | 80.9 | 90.0 | 94.4 | 98.1 |
| 0.9 | P | 45.4 | 74.5 | 85.4 | 91.1 | 96.8 |
| | G | 44.2 | 74.9 | 86.1 | 91.5 | 96.4 |
| 0.99 | P | 45.6 | 77.1 | 88.0 | 94.2 | 98.9 |
| | G | 47.1 | 81.9 | 91.4 | 95.4 | 97.7 |
| 0.999 | P | 46.4 | 78.6 | 89.3 | 95.0 | 99.4 |
| | G | 49.7 | 85.8 | 94.1 | 96.6 | 98.3 |
| 0.9999 | P | 47.1 | 79.4 | 89.9 | 95.3 | 99.5 |
| | G | 51.6 | 88.6 | 95.4 | 97.1 | 98.6 |

Table 3. Coverage rates for the confidence intervals of five quantiles of the Weibull distribution and five confidence levels γ , using both the percentile (P) and the Gaussian (G) methods. Results are presented for values of the Weibull shape parameter α of 2.0.

| Quantile | Conf. level method | Confidence level γ (per cent) | | | | |
|----------|--------------------|--------------------------------------|------|------|------|------|
| | | 50 | 80 | 90 | 95 | 99 |
| 0.5 | P | 51.4 | 81.3 | 90.4 | 95.0 | 98.9 |
| | G | 51.7 | 81.1 | 90.0 | 94.7 | 98.5 |
| 0.9 | P | 43.8 | 73.2 | 84.1 | 90.1 | 96.2 |
| | G | 44.2 | 73.8 | 84.6 | 90.3 | 96.1 |
| 0.99 | P | 43.5 | 76.1 | 87.7 | 94.1 | 99.3 |
| | G | 48.9 | 83.2 | 92.4 | 96.0 | 98.3 |
| 0.999 | P | 45.4 | 78.0 | 89.9 | 95.5 | 99.6 |
| | G | 54.3 | 88.6 | 95.4 | 97.7 | 99.0 |
| 0.9999 | P | 46.8 | 79.3 | 90.8 | 96.2 | 99.7 |
| | G | 57.9 | 91.3 | 96.8 | 98.3 | 99.3 |

increasing the sample size to 100 gave similar or better coverage rates than those shown in Tables 1–3. However, the differences between the percentile and Gaussian methods were smaller, especially for the shape parameter $\alpha = 2$ providing further agreement with the results presented in Hall *et al.* (2004).

Concluding remarks

In general, the coverage rates summarised in Tables 1–3 provide corroboration for the confidence intervals for the Weibull distribution constructed by Bootstrap resampling,

as shown in Hall *et al.* (2004). As before, the confidence limits developed by the percentile method were to be preferred to those constructed by the Gaussian method. Nevertheless, the problems encountered with the resampled confidence limits for samples of size 30, particularly the anomalous behaviour of the lower confidence limits at low probabilities of exceedance, would appear to mitigate against the use of Bootstrapping in these circumstances. Similar problems were not evident, however, when the samples size was increased to 100. Unfortunately, sample sizes of size 30 are typical of those to which hydrological frequency analysis is generally applied. A more comprehensive

investigation of Bootstrapping small samples should therefore be undertaken before the technique is more widely applied.

Acknowledgements

The suggestion by an anonymous referee to extend the original paper with an analysis of coverage rates is gratefully acknowledged.

References

- Cheng, R.C.H. and Amin, N.A.K., 1983. Estimating parameters in continuous univariate distributions with a shifted origin. *J. Roy. Stat. Soc., Ser. B- Stat. Met.*, **45**, 394–403.
- Davison, A.C. and Hinkley, D.V., 1997. *Bootstrap methods and their application*. Cambridge University Press, Cambridge, UK. 582pp.
- Efron, B. and Tibshirani, R.J., 1993. *An introduction to the Bootstrap*. Chapman and Hall, London, UK. 436pp.
- Hall, M.J., van den Boogaard, H.F.P., Fernando, R.C. and Mynett, A.E., 2004. The construction of confidence intervals for frequency analysis using resampling techniques. *Hydrol. Earth Syst. Sci.*, **8**, 235–246.
- Stark, P.B., 2003. *SticiGui: Statistics tools for internet and classroom instruction with a graphical user interface. Chapter 18 Confidence intervals*. <http://stat-www.berkeley.edu/users/stark/SticiGui/text/ch18.htm>