Generalized Maximum Likelihood GEV quantile estimators from Partial Duration and Annual Maximum Series employing complete and censored information

Jery Stedinger and Eduardo Martins
School of Civil and Environmental Engineering
Cornell University
Hollister Hall
Ithaca, NY 14853-3501
tel: 607-255-2351; fax: 607-255-9004
e-mail: jrs5@cornell.edu

The 3-parameter GEV distribution has found wide application for describing annual flood, rainfall, wind speed, wave height and other maxima. Hosking, et al. (1985) show that small-sample MLE parameter estimators are very unstable, and recommended probability weighted moment (PWM) estimators, which are equivalent to L-moment estimators (Hosking, 1990). Both are linear combinations of order statistics. Reexamination of the behavior of MLEs in small samples demonstrates that indeed absurd values of the GEV shape parameter κ are generated. Use of Bayesian prior distributions to restrict estimated κ-values to a statistically/physically reasonable range in a Generalized Maximum Likelihood (GMLE) analysis eliminated this problem. For example, the GMLE obtained with a relatively uniform prior over [-0.3, 0.5] and L-moment estimators had nearly equivalent precision in samples of 15 to 25 for $-0.2 \leq \kappa \leq +0.3$. As sample sizes increase, the relative performance of GMLEs increases. When regional information from a number of examples is be pooled to develop a more informative prior distribution for κ (as in Madsen and Rosbjerg, 1997), then substantial improvements in extreme quantile estimators (exceedance probability $\leq$ 1%) results.

A recent series of papers by Madsen and others (1997ab) address the relative precision of GEV flood-quantile estimators obtained with either partial duration or annual maximum series using MLE and L-moment estimators. Their results are reexamined with the new GMLE procedure. Generally the GMLE-PDS procedures were competitive with corresponding L-moment PDS estimators. For

sufficiently large arrival rates for peaks over the threshold, quantile estimators based on partial duration series can be more accurate and robust, though the accuracy gains are often modest.

An advantage of maximum likelihood estimators is that they can efficiently employ censored information. Often records of regularly collected extreme values can be augmented with censored data. For example,often the magnitude of the largest historical floods can be determined because their extraordinary values results in physical or social evident which is still available. Because only large values are recorded, historical flood records represent a censored sample. Stedinger and Cohn (1986), Jin and Stedinger (1989), Frances, Salas and Boes (1994) and others have explored the value of historical flood information in the framework of annual maximum series. The greatest values were observed when estimating distributions with a third shape parameter (Cohn and Stedinger, 1987), as is done here. The precise magnitude of historical floods can be difficult to determine so that Stedinger and Cohn considered cases wherein the magnitudes of the flood peaks were know with relatively high precision, and when all that was know is that an event exceeded the perception threshold. The later is called binomially censored data.

The value of censored information in a partial duration series framework employing GMLEs with the GEV distribution is explored. Again, historical flood information can be of tremendous value depending on the length of the historical period and the magnitude of the perception threshold. These performance gains were described in terms of the effective PDS record length for the estimators, which also yields an average gain (AG) in effective PDS record length per year of historical record. For annual flood series the AG is essentially a simple function of the perception threshold for historical events. In a partial duration series analysis the AG was found to also depend upon the threshold level for the systematic PDS data. Again, as the magnitude of the perception threshold increased, the differences between the case when the magnitude of the historical peaks are well recorded and binomial censored data decreases.

References

Cohn, T. A., and J. R. Stedinger, Use of historical information in a maximum-likelihood framework, *J. of Hydrol., 96* (1-4), 215-23, 1987.

Hosking, J. R. M., J. R. Wallis, and E. F. Wood, Estimation of the generalized extreme-value distribution by the method of probability weighted moments, *Technometrics, 27(*3), 251-261, 1985.

Hosking, J. R. M., L-Moments: Analysis and estimation of distributions using linear combinations of order statistics, *J. of Royal Statistical Society, B, 52* (2), 105-124, 1990.

Frances, F., J.D. Salas, and D.C. Boes, Flood frequency analysis with systematic and historical or paleoflod data based on the two-parameter general extreme vale models, *Water Resour. Res., 29*(6) , 1653-64, 1994.

Jin, M., and J. R. Stedinger, Flood frequency analysis with regional and historical information, *Water Resour. Res., 25*(5), 925-36, 1989.

Madsen, H., C.P. Pearson, and P.F. Rasmussen, and D. Rosbjerg, Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 1. At-site modeling, *Water Resour Res., 33*(4), 747-758, 1997a.

Madsen, H., C.P. Pearson, and D.. Rosbjerg, Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 2. Regional modelling, *Water Resour Res., 33*(4), 759-770, 1997b.

Madsen, H., and D. Rosbjerg, Generalized least squares and empirical Bayes estimation in regional partial duration series index-flood modeling,*Water Resour Res., 33*(4), 771-782, 1997.