Computation of Confidence Limits for the Two Populations Extreme Value Type I Distribution

Jose A. Raynal-Villasenor
Civil and Environmental Engineering Department, Universidad de las Americas Puebla
72820 Cholula, Puebla, Mexico

M. Elena Raynal-Gutierrez and
Department of Civil Engineering, Monterrey Institute of Technology Campus Puebla
Via Atlixcayotl # 2301
72453 Puebla, Puebla, Mexico

Abstract

A methodology for obtaining the confidence limits for the two populations extreme value type I distribution is presented. The methodology is based on the application of the maximum likelihood method for estimating the parameters of the distribution and the confidence limits of the design events. The confidence limits are obtained by using of the variance-covariance matrix of the parameters and assuming a normality of the design events to compute them. Given the complexity of the likelihood function, its logarithmic form is used and a non-linear multivariable constrained optimization method is applied to maximize such function to produce the maximum likelihood estimators of the parameters and its confidence limits of the distribution. An example of application of the proposed methodology is contained in the paper. The results showed an improvement in the standard error of fit and confidence limits narrower than those produced by the one population procedure.

1. Introduction

The method of maximum likelihood has been recognized like one of the best methods of estimation of parameters of functions of probability distribution, the properties of its estimators like asymptotic unbiasness and sufficiency, as well as the consistency and efficiency, have been briefed frequently in technical literature, [1], [2], etc. This method also has the virtue of being able to handle very complex likelihood functions with an amazing flexibility. The use of mixed functions of probability to fit samples coming from two or more populations has been proposed from time back, [1]. In the particular case of the functions of distribution of extreme values, TCEV, [3], [4], [5] and [6], the mixed Gumbel, [7], [8], [9], and the mixed general of extreme value, [10] and [11]. [12] used different mixtures of normal, gamma and Gumbel distributions to test the relevance of using mixture models, by computing the marginal likelihoods of single distribution models, and to verify the presence of a persistence in the time series by comparing independent and identically distributed and Markovian mixture models.

2. The Extreme Value Type I Distribution Function

The extreme value type I distribution function, for the maxima, is[13]:

$$F(x) = \exp\left[-\exp\left(-\frac{x - x_{01}}{\alpha_1}\right)\right] \tag{1}$$

where α and x_{01} are the scale and location parameters, respectively.

The probability density function is given by, [13]:

$$f(x) = \frac{1}{\alpha_1}\exp\left(-\exp\left[-\frac{x - x_{01}}{\alpha_1}\right]\right) \bullet \exp\left[-\frac{x - x_{01}}{\alpha_1}\right] \tag{2}$$

where: $-\infty < x < \infty$ and $\alpha > 0$

3. Two Populations Extreme Value Type I Distribution Function

Based in the general form for distribution function for two populations, [1] proposed:

ISSN: 2278-0181
Vol. 2 Issue 9, September - 2013

$$F(x)_{mix} = (1-p)F(x;\theta_1) + pF(x;\theta_2) \qquad (3)$$

where p is the proportion of the second population in the sample.

The two populations extreme value type I distribution function can be expressed as:

$$F(x)_{mix} = (1-p)\exp\left(-\exp\left[-\frac{(x-x_{01})}{\alpha_1}\right]\right) + p\exp\left(-\exp\left[-\frac{(x-x_{01}}{\alpha_2}\right]\right) \qquad (4)$$

and the corresponding probability density function is:

$$f(x)_{mix} = \frac{(1-p)}{\alpha_1}\exp\left[-\frac{(x-x_{01})}{\alpha_1}\right] \bullet \exp\left(-\exp\left[-\frac{(x-x_{01})}{\alpha_1}\right]\right) \bullet + \frac{p}{\alpha_2}\exp\left[-\frac{x-x_{02}}{\alpha_2}\right]\exp\left[-\exp\left(-\frac{x-x_{02}}{\alpha_2}\right)\right] \qquad (5)$$

4. The Method of Maximum Likelihood

The likelihood function for N independent and identically distributed X1, X2,..., Xn can be obtained as the joint probability density function, that is, [1]:

$$L(x,\theta) = \prod_{i=1}^{N} f(x_i) \qquad (6)$$

where θ is the parameter vector and f (.) is the probability density function.

The logarithmic version of the former equation is:

$$LnL(x,\theta) = \sum_{i=1}^{N} Lnf(x) \qquad (7)$$

Based in the statements of the previous section, the logarithmic likelihood function of the two populations extreme value type I distribution is:

$$LnL(x;x_{01},\alpha_1,x_{02},\alpha_2,p) = \sum_{i=1}^{N} \mathrm{Ln}\left\{\frac{1-p}{\alpha_1}\exp\left(-\frac{x-x_{01}}{\alpha_1}\right) \bullet \exp\left[-\exp\left(-\frac{x-x_{01}}{\alpha_1}\right)\right] + \frac{p}{\alpha_2}\exp\left(-\frac{x-x_{02}}{\alpha_2}\right)\exp\left[-\exp\left(-\frac{x-x_{02}}{\alpha_2}\right)\right]\right\} \qquad (8)$$

In the proposed procedure, eq. (8) has been maximized directly by using the well-known non-linear mutlivariable constrained Rosenbrock optimization method, [14].

5. Estimation of the Confidence Limits for the Two Populations Extreme Value Type I Distribution

The variance-covariance matrix for the two populations extreme value type I distributions, may be expressed as:

$$V = \begin{bmatrix} Var(x_{01}) & Cov(x_{01},\alpha_1) & Cov(x_{01},x_{02}) & Cov(x_{01},\alpha_2) & Cov(x_{01},p) \\ Cov(x_{01},\alpha_1) & Var(\alpha_1) & Cov(\alpha_1,x_{02}) & Cov(\alpha_1,\alpha_2) & Cov(\alpha_1,p) \\ Cov(x_{01},x_{02}) & Cov(\alpha_1,x_{02}) & Var(x_{02}) & Cov(x_{02},\alpha_2) & Cov(x_{02},p) \\ Cov(x_{01},\alpha_2) & Cov(\alpha_1,\alpha_2) & Cov(x_{02},\alpha_2) & Var(\alpha_2) & Cov(\alpha_2,p) \\ Cov(x_{01},p) & Cov(\alpha_1,p) & Cov(x_{02},p) & Cov(\alpha_2,p) & Var(p) \end{bmatrix} \qquad (9)$$

and its elements are:

$$V = \begin{bmatrix} E\left(-\frac{\partial^2 LL}{\partial x_{01}^2}\right) & E\left(-\frac{\partial^2 LL}{\partial x_{01}\partial\alpha_1}\right) & E\left(-\frac{\partial^2 LL}{\partial x_{01}\partial x_{02}}\right) & E\left(-\frac{\partial^2 LL}{\partial x_{01}\partial\alpha_2}\right) & E\left(-\frac{\partial^2 LL}{\partial x_{01}\partial p}\right) \\ E\left(-\frac{\partial^2 LL}{\partial x_{01}\partial\alpha_1}\right) & E\left(-\frac{\partial^2 LL}{\partial\alpha_1^2}\right) & E\left(-\frac{\partial^2 LL}{\partial\alpha_1\partial x_{02}}\right) & E\left(-\frac{\partial^2 LL}{\partial\alpha_1\partial\alpha_2}\right) & E\left(-\frac{\partial^2 LL}{\partial\alpha_1\partial p}\right) \\ E\left(-\frac{\partial^2 LL}{\partial x_{01}\partial x_{02}}\right) & E\left(-\frac{\partial^2 LL}{\partial\alpha_1\partial x_{02}}\right) & E\left(-\frac{\partial^2 LL}{\partial x_{02}^2}\right) & E\left(-\frac{\partial^2 LL}{\partial x_{02}\partial\alpha_2}\right) & E\left(-\frac{\partial^2 LL}{\partial x_{02}\partial p}\right) \\ E\left(-\frac{\partial^2 LL}{\partial x_{01}\partial\alpha_2}\right) & E\left(-\frac{\partial^2 LL}{\partial\alpha_1\partial\alpha_2}\right) & E\left(-\frac{\partial^2 LL}{\partial x_{02}\partial\alpha_2}\right) & E\left(-\frac{\partial^2 LL}{\partial\alpha_2^2}\right) & E\left(-\frac{\partial^2 LL}{\partial\alpha_2\partial p}\right) \\ E\left(-\frac{\partial^2 LL}{\partial x_{01}\partial p}\right) & E\left(-\frac{\partial^2 LL}{\partial\alpha_1\partial p}\right) & E\left(-\frac{\partial^2 LL}{\partial x_{02}\partial p}\right) & E\left(-\frac{\partial^2 LL}{\partial\alpha_2\partial p}\right) & E\left(-\frac{\partial^2 LL}{\partial p^2}\right) \end{bmatrix}^{-1} \qquad (10)$$

The second partial derivatives, which expected values must be obtained to evaluate the variance-covariance matrix for the two populations extreme value type I distribution, are:

$$\frac{\partial LnL}{\partial x_{01}} = \frac{1-p}{\alpha_1^2} \sum_{i=1}^{N} \left\{ \frac{F(x;\theta_1)\exp\left(-\frac{x-x_{01}}{\alpha_1}\right) \bullet \left[1-\exp\left(-\frac{x-x_{01}}{\alpha_1}\right)\right]}{DEN} \right\} \tag{11}$$

$$\frac{\partial LnL}{\partial \alpha_1} = \frac{1-p}{\alpha_1^2} \sum_{i=1}^{N} \left\{ \frac{F(x;\theta_1)\exp\left(-\frac{x-x_{01}}{\alpha_1}\right) \bullet (x-x_{01}) \bullet \left[\exp\left(-\frac{x-x_{01}}{\alpha_1}\right)-1\right]}{DEN} \right\} \tag{12}$$

$$\frac{\partial LnL}{\partial x_{02}} = \frac{1-p}{\alpha_2^2} \sum_{i=1}^{N} \left\{ \frac{F(x;\theta_2)\exp\left(-\frac{x-x_{02}}{\alpha_2}\right) \left[1-\exp\left(-\frac{x-x_{02}}{\alpha_2}\right)\right]}{DEN} \right\} \tag{13}$$

$$\frac{\partial LnL}{\partial \alpha_2} = \frac{1-p}{\alpha_2^2} \sum_{i=1}^{N} \left\{ \frac{F(x;\theta_2)\exp\left(-\frac{x-x_{02}}{\alpha_2}\right) \bullet (x-x_{02}) \bullet \left[\exp\left(-\frac{x-x_{02}}{\alpha_2}\right)-1\right]}{DEN} \right\} \tag{14}$$

$$\frac{\partial LnL}{\partial p} = \sum_{i=1}^{N} \left\{ \frac{-f(x;\theta_1)+f(x;\theta_2)}{DEN} \right\} \tag{15}$$

where:

DEN = f(x)mix (16)

$$F(x;\theta_1) = \exp\left[-\exp\left(-\frac{x-x_{01}}{\alpha_1}\right)\right] \tag{17}$$

$$F(x;\theta_2) = \exp\left[-\exp\left(-\frac{x-x_{02}}{\alpha_2}\right)\right] \tag{18}$$

$$f(x,\theta_1) = \frac{1}{\alpha_1}\exp\left[-\exp\left(-\frac{x-x_{01}}{\alpha_1}\right)\right] \bullet \exp\left(-\frac{x-x_{01}}{\alpha_1}\right) \tag{19}$$

$$f(x,\theta_2) = \frac{1}{\alpha_2}\exp\left[-\exp\left(-\frac{x-x_{02}}{\alpha_2}\right)\right] \bullet \exp\left(-\frac{x-x_{02}}{\alpha_2}\right) \tag{20}$$

6. Results and Discussion

The gauging station Jaina in the state of Sinaloa, located in Northwestern Mexico, with period of record (1941-1991), has been chosen to show the procedure for obtaining the values of the confidence limits based in the two populations extreme value type I distribution. The initial values required by the procedure were estimated by using the computer code FLODRO 4.0 in [15]. The design values and their confidence limits for the one and two populations' approaches are shown in Tables 1 and 2, respectively, and were obtained by using the computer code contained in [16].

Table 1. One population design values and their confidence limits for gauging station Jaina, Sin (SE = 371.42)

(1)	(2)	(3)	(4)	(5)
5	821	1002	1184	363
10	1044	1277	1509	465
20	1256	1541	1825	569
50	1528	1881	2235	707
100	1731	2137	2543	812

(1) Return Period (years)
(2) Lower Limit (m^3/s)
(3) Design Value (m^3/s)
(4) Upper Limit (m3/s)
(5) Interval Width between Confidence Limits (m3/s)

Table 2. Two populations design values and their confidence limits for gauging station Jaina, Sin (SE = 276.15)

(1)	(2)	(3)	(4)	(5)
5	886	1007	1129	243
10	1345	1498	1651	306
20	1756	1938	2120	364
50	2261	2479	2698	437
100	2631	2876	3121	490

(1) Return Period (years)
(2) Lower Limit (m^3/s)
(3) Design Value (m^3/s)
(4) Upper Limit (m3/s)
(5) Interval Width between Confidence Limits (m3/s)

A graphical representation of these results can be observed in figures 1 and 2, for the one population approach and for the two populations model, respectively.

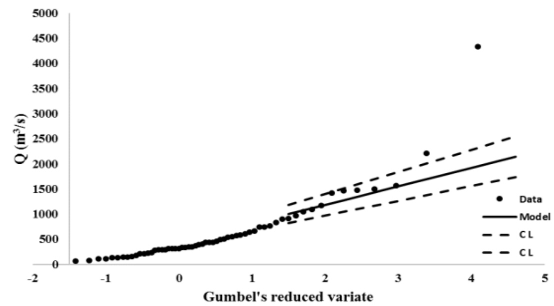

Figure 1. Empirical and One Population Theoretical Probability Distribution Function and Confidence Limits for Gauging Station Jaina, Mexico

The standard error of fitting (SE) has been computed as, [17]:

$$SE = \left[\frac{\sum_{i=1}^{N} (x_i - y_i)^2}{(N - m_j)} \right]^{1/2} \quad (21)$$

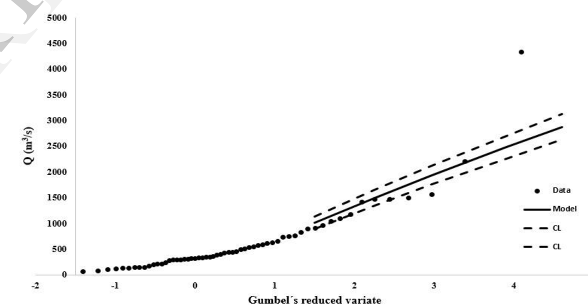

Figure 2. Empirical and Two Populations Theoretical Probability Distribution Function and Confidence Limits for Gauging Station Jaina, Mexico

The Gumbel's reduced variate, required to produce the abscissa axis in graphical displays of flood data, models applied and its confidence limits, is obtained as follows:

$$y = - Ln(-Ln(1-1/Tr)) \quad (22)$$

where Tr is the return period in years.

It is observed that the two populations model fits the flood sample much better (SE = 276.15) compared with the one population model (SE= 371.42). The two populations model produced a narrower confidence limits, too.

The application of the proposed approach is restricted to the fact that the computer code for the Rosenbrock's constrained multivariable method must be available, given that performing the required computations for such method without a computer code is just out of the question.

7. Conclusions

A procedure for the obtaining of the confidence limits for the two populations extreme value type I distribution has been described here, based on the method of maximum likelihood. The procedure has given good results so far with the samples of data analyzed until now, one of which was used as an example of application of the proposed methodology. It can be observed that in this example of application, the standard error of fitting has been reduced significantly and the width of the confidence limits was reduced, too. Based on this arguments, the authors recommend the procedure here shown as an effective tool for annual flood frequency analysis when two populations are detected within a sample of flood data.

8. Acknowledgements

The authors wishes thank to the Universidad de las Americas Puebla for the support granted to make this publication possible.

9. References

[1] A. M. Mood, F. Graybill, F and D. C. Boes, *Introduction to the theory of statistics*, McGraw-Hill Book Co., Third Ed., New York, N Y. 1974

[2] C. T. Haan, *Statistical methods in hydrology*, The Iowa State University Press, Ames, Iowa. 1977

[3] E. J. Gumbel, *Statistics of extremes*, Columbia University Press, New York, N Y. 1958

[4] P. Todorovic and J. Rousselle "Some problems of flood analysis", *Wat. Resour. Res.*, 7(5):1144-1150, 1971

[5] R. V. Canfield, The Distribution of the extremes of a mixture of random variable with applications to hydrology, in *Input for Risk Analysis in Water Systems*, E A McBean, K W Hipel T E Unny, eds. , 77-84, Water Resources Publications, Littleton, Colorado. 1979

[6] F. Rossi, M. Florentino and P. Versace, "Two component extreme value distribution for flood frequency analysis", *Wat. Resour. Res*., 20(7):847-856. 1984

[7] F. J. Gonzalez-Villareal, *Contribution to the frequency analysis of the extreme values of the floods in a river, Report # 277*, Instituto de Ingenieria, Universidad Nacional Autonoma de Mexico, Mexico, D.F., Mexico (In Spanish). 1970

[8] J. A. Raynal-Villasenor, Maximum likelihood estimators of the parameters of the mixed Gumbel distribution, *Proc. XII Congress of the National Academy of Engineering*, Saltillo, Coah., Mexico, 468-471. (In Spanish). 1986

[9] J. A. Raynal-Villasenor and J. L. Guevara-Miranda, Maximum likelihood estimators for the two populations Gumbel distribution, *Hydrological Science and Technology Journal*, 13(1-4):47-56. 1997

[10] J. A. Raynal-Villasenor and O. D. Santillan-Hernandez, Maximum likelihood estimators of the parameters of the mixed general extreme value distribution, *Proc. IX National Congress on Hydraulics*, Queretaro, Qro., Mexico, 79-90. (In Spanish). 1986

[11] C. Gutierrez-Ojeda and J. A. Raynal-Villasenor, Mixed distributions in flood frequency analysis, *Proc. X National Congress on Hydraulics*, Morelia, Mich., Mex., 220-228. (In Spanish). 1988

[12] G. Evin, J. Merleau, and L. Perreault, "Two components mixtures of normal, gamma and Gumbel distributions for hydrological applications", *Wat. Resour. Res*., 47, 2011, Art. No: W08525, DOI: 10.1029/2010WR010266.

[13] Natural Environment Research Council (NERC) *Flood studies report, vol. 1 Hydrological studies*, Whitefriars LDT. London. UK. 1975

[14] J. L.Kuester and J. H. Mize, *Optimization with FORTRAN*, 386-398, McGraw-Hill Book Co., New York. 1973

[15] J. A. Raynal-Villasenor, *Frequency Analysis of Hydrologic Extremes*, Lulu.com. 2010

[16] M. E. Raynal-Gutierrez, *An interactive computer package for the estimation of the confidence limits for the two populations Gumbel distribution*, B. Sc. Thesis, Department of Civil Engineering, Universidad de las Americas Puebla, Puebla, Mexico. (In Spanish). 2001

[17] G. W. Kite, *Frequency and risk analyses in hydrology*, Water Resources Publications, Littleton, Colorado. 1988