PARAMETER ESTIMATION OF THE GENERALZIED EXTREME VALUE DISTRIBUTION USING QUADRATIC PROGRAMMING

Hyun Woo PARK [1] and Hoon SOHN [2]

1) *Korea Bridge Design & Engineering Research Center, Seoul National University, Seoul 151-742, KOREA*
2) *The Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

Corresponding Author: Hyun Woo PARK, hwpark91@snu.ac.kr

ABSTRACT

Structural health monitoring can be defined as a statistical pattern recognition problem which necessitates establishing a decision boundary based on measured data. Choosing the decision boundary is often based on the assumption that the distribution of measured data is Gaussian in nature. This unwarranted assumption impairs the performance of structural health monitoring significantly by increasing false positive and negative indications of damage. This paper attempts to address the issue of the decision boundary establishment using extreme value statistics (EVS) so that the tails of a distribution associated with damage can be properly modeled. Combining three extreme value distributions (Gumbel, Weibull, and Frechet distributions) into a single one, the generalized extreme value distribution (GEV) is adopted for establishing the decision boundary. A parameter estimation technique based on a nonlinear optimization scheme is developed to estimate the model parameters of the GEV and choose the most appropriate domain of attraction.

GENERALIZED EXTREME VALUE DISTRIBUTION

Suppose that one is given a vector of samples $\{X_1, X_2, \ldots, X_n\}$ from an arbitrary *parent distribution*. The most relevant statistic for studying the tails of the parent distribution is the maximum operator, max($\{X_1, X_2, \ldots, X_n\}$), which selects the point of maximum value from the sample vector. Note that this statistic is relevant for the right tail of a univariate distribution only. For the left tail, the minimum should be used. The pivotal theorem of EVS states that in the limit as the number of vector samples tends to infinity, the induced distribution on the maxima of the samples can only take one of three forms: Gumbel, Weibull, or Frechet [1,2]. Combining these three forms of extreme value distributions into a unified one, generalized extreme value distribution (GEV) is introduced in this paper [3]:

$$\text{MAXIMA}: \ \bar{\phi}(x;\mu,\sigma,\gamma) = \exp\left\{-\left[1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\gamma}\right\},\ \text{-}1-\gamma\left(\frac{x-\mu}{\sigma}\right) \le 0 \qquad (1)$$

$$\text{MAXIMA}: \ \bar{\phi}(x;\mu,\sigma,\gamma) = \exp\left\{-\left[1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\gamma}\right\},\ \text{-}1-\gamma\left(\frac{x-\mu}{\sigma}\right) \le 0 \qquad (2)$$

where $\overline{\phi}$ and $\underline{\phi}$ are the GEVs for maxima and minima while μ, σ, and γ are the location, scale, and shape parameters of the GEV, respectively. Note that the GEVs for maxima and minima can be converted to the associated Frechet and Weibull distributions by using simple transformations while the Gumbel distribution corresponds to the limiting case of the GEV when the shape parameter of the GEV approaches 0, i.e. $\gamma \rightarrow 0$ [4].

When samples of maximum or minimum data from a number of *n*-point populations are given, it is possible to fit a parametric model of the GEV to the data. It is also possible to fit the model to portions of the parent distribution's tails, as the distribution of the tails is equivalent to the appropriate extreme value distribution. Once the parametric model is obtained, it can be used to compute an effective threshold for novelty based on the true statistics of the data as opposed to statistics based on an unwarranted assumption of a Gaussian distribution.

PARAMETER ESTIMATION

A least squares method using the generalized weighted least squares is employed to estimate the best-fit GEV and the associated model parameters [4].

$$\underset{\theta}{\text{Min}}\ \pi = \frac{1}{2}[\phi(\mathbf{x};\theta) - \mathbf{p}]^T \mathbf{W}[\phi(\mathbf{x};\theta) - \mathbf{p}] \quad \text{subject to} \quad \mathbf{R}(\mathbf{x};\theta) \leq \mathbf{0} \qquad (3)$$

where ϕ, $\mathbf{p}$, $\mathbf{x}$, θ, $\mathbf{W}$, and $\mathbf{R}$ represent an assumed cumulative density function vector from one of Eqs. (1)-(2), an empirical cumulative density function vector, a probability position vector, a model parameter vector of ϕ, a weighting matrix, and a constraint vector with respect to θ, respectively. The model parameter vector consists of μ, σ, and γ in Eqs.(1)-(2). Regarding the constraint vector **R**, nonlinear constraints are imposed on the model parameters to ensure that the model parameters stay in a feasible domain defined in Eqs. (1)-(2).

Eq.(3) is a nonlinear optimization problem being subject to multiple constraints. In this study, sequential quadratic programming (SQP) [5] is used to solve the optimization problem iteratively until a converged solution is obtained.

The validity of the proposed method will be demonstrated in numerical studies using real sample data sets and in the context of delamination detection in a composite plate.

REFERENCES

1. Fisher, R.A. and Tippett, L.H.C., "Limiting Forms of the Frequency Distributions of the Largest or Smallest Members of a Sample," *Proceedings of the Cambridge Philosophical Society*, **24**, 180-190, 1928.
2. Castillo, E., *Extreme Value Theory in Engineering*, Academic Press Series in Statistical Modeling and Decision Science, San Diego, CA, 1998.
3. Jenkinson, A. F., "The frequency distribution of the annual maximum (or minimum) of meteorological elements," *Quarterly Journal of the Royal Meteorological Society*, 81, 158–171, 1955.
4. Sohn, H., Park, H.W., Law, K.H., Farrar, C.R., "Minimizing Misclassification of Damage using Extreme Values Statistics," *the US-Korea Workshop on Smart Structure Technologies*, Seoul, Korea, September 2-4, 2004.
5. Luenberger, D.G., "Linear and Nonlinear Programming, Second Edition", Kluwer Academic Publishers, 1989.