Parameters Estimation of Generalized Extreme Value Distribution under Progressive Type II Censored

El Sherpieny, E. A.1, Assar, S. M. 2 and Amer, N. M. 3

[1] Department of Mathematical Statistics, Institute of Statistical Studies & Research,
Cairo University, (Giza, Egypt).

[2] Department of Mathematical Statistics, Institute of Statistical Studies & Research,
Cairo University, (Giza, Egypt).

[3] M. Sc. Student, Institute of Statistical Studies & Research,
Cairo University, (Giza, Egypt).

ABSTRACT— *In this paper, the estimation for the three unknown parameters of the generalized extreme value distribution under progressive type-II censored will be considered. The corresponding asymptotic variance covariance matrix for the parameters and also asymptotic confidence intervals for the parameters are obtained. Finally, A numerical illustration will be carried out. Some literatures may be considered as special cases from our results.*

Keywords— Generalized extreme value distribution, Progressive censoring, Variance- covariance matrix.

---

1. INTRODUCTION

Life testing often consists of conducting a test on an item (under specified conditions) to determine the time it takes for a failure to occur. An analysis of the data from replicate tests includes both the times to failure for the items that failed and the time of test termination for those that did not fail. An efficient way of collecting lifetime data that results in a saving in experimental time and cost is progressive type-II censoring [Balakrishnan & Aggarwala (2000)] [1], so it is useful in both industrial and clinical settings. It allows for the removal of surviving experimental units at points other than the termination of the experiment while retaining mathematical tractability in the data analysis in many cases.

The generalized extreme value (GEV) distribution is a flexible three parameters distribution that combines three extreme-value distributions within a single framework: the Gumbel, Frechet and Weibull. The distribution of extremes is one of the common interests to many disciplines such as the natural scientist and financial scientist. Many authors have considered the applications of extremes distributions. Coles (2001) [3] introduced a general introduction to the analysis of extreme values; Katz et al. (2002) [5] showed a review of the analysis of hydrological extremes. Embrechts et al. (1997) [4] and McNeil et al. (2005) [6] considered the modeling of extremes in insurance, finance and quantitative risk management.

The progressive type II censoring scheme is conducted as follows, suppose that *n* independent and identically distributed units taken from a continuous distribution are placed on a life test experiment. Let a censoring scheme $r_1, r_2, \ldots, r_m$ be prefixed in such a manner that immediately after the first failure, r_1 surviving units are removed from the experiment randomly. Similarly when the second failure occurs, r_2 surviving units are removed from the experiment, again randomly. The test is continued until the $mth(1 \leq m \leq n)$ failure takes place. By this time remaining units r_m, where $r_m = n - m - r_1 - r_2 - \cdots - r_{m-1}$ are removed from the test**.** The likelihood function of the progressive type-II is defined as follows: [Balakrishnan & Aggarwala (2000)] [1]

$$L(X; \theta|R) = \prod_{i=1}^{m} C_i\, f(x_{(i)}; \theta)[1 - F(x_{(i)}; \theta)]^{r_i}, \qquad\qquad (1)$$

where $C_i = \prod_{i=0}^{m-1}(n - \sum_{j=0}^{i} r_j - i)$, $i = 1, \ldots, m$ and $r_0 = 0$. x_i is the lifetime of the *ith* order statistics and $f(x; \theta)$ and $F(x; \theta)$ are the density function and the cumulative function of our concerning distribution, respectively. If

$r_1 = r_2 = \cdots = r_{m-1} = 0$, then $r_m = n - m$ which corresponds to the type II censoring. If $r_1 = r_2 = \cdots = r_m = 0$, then $n = m$ which corresponds to the complete sample.

In this paper, section 2 will describe the model of the generalized extreme value (GEV) distribution, section 3; parameters estimation under progressive type-II is discussed, Section 4; a numerical illustration for different values of sample size and parameters are conducted.

2. MODEL

Let X be a generalized extreme value random variable, with probability density function is given by

$$f(x) = \frac{1}{\sigma} t(x)^{k+1} e^{-t(x)}; x \in \begin{cases} (-\infty, \mu - \frac{\sigma}{k}] & \text{if } k < 0 \\ (-\infty, \infty) & \text{if } k = 0 \\ [\mu - \frac{\sigma}{k}, \infty) & \text{if } k > 0 \end{cases} \tag{2}$$

$\sigma > 0$, $\mu, k \in R$; R is the set of real numbers,
where

$$t(x) = \begin{cases} (1 + k\frac{x-\mu}{\sigma})^{-\frac{1}{k}}, & \text{if } k \neq 0 \\ e^{-\frac{(x-\mu)}{\sigma}}, & \text{if } k = 0 \end{cases},$$

with cumulative distribution
$$F(x) = e^{-t(x)}. \tag{3}$$

Special cases, for $k = 0$, the distribution (1) becomes the Gumble distribution. For $k > 0$, the distribution becomes the Frechet distribution and for $k < 0$ the distribution reduces to reserved Weibull distribution.

The generalized extreme value (GEV) distribution have mean in the form

$$E(x) = \begin{cases} \mu + \frac{\sigma}{k}[\Gamma(1 - k) - 1] & ; k \neq 0, k < 1 \\ \mu + \sigma v & ; k = 0 \\ not\ exists & ; k \geq 1 \end{cases} ; v \text{ is Euler's Constant}, \tag{4}$$

and variance

$$Var(x) = \begin{cases} \frac{\sigma^2}{k^2}[\Gamma(1 - 2k) - \Gamma^2(1 - k)]; k \neq 0, k < \frac{1}{2} \\ \sigma^2 \frac{\pi^2}{6} & ; k = 0 \\ not\ exists & ; k \geq \frac{1}{2} \end{cases} ; \pi \text{ is constant}. \tag{5}$$

Also, the moment generating function of (GEV) is defined as

$$M(t) = E(e^{tx}) = e^{(\mu - \frac{\sigma}{k})t} \int_0^\infty e^{-y + \frac{\sigma t}{k} y^{-k}} dy ; \quad x = \mu - \frac{\sigma}{k} + \frac{\sigma}{k} y^{-k}.$$

Then

$$M(t) = e^{(\mu - \frac{\sigma}{k})t} \int_0^\infty \left[\sum_{n=0}^\infty \frac{(\frac{\sigma t}{k} y^{-k})^n}{n!} \right] e^{-y} dy,$$

$$M(t) = e^{(\mu - \frac{\sigma}{k})t} \left[\int_0^\infty e^{-y} dy + \int_0^\infty \left(\frac{\sigma t}{k} y^{-k} \right) e^{-y} dy + \int_0^\infty \frac{((\frac{\sigma t}{k})^2 y^{-2k})}{2!} e^{-y} dy + \cdots \right],$$

$$M(t) = e^{(\mu - \frac{\sigma}{k})t} \left[1 + \frac{\sigma t}{1! k} \Gamma(1 - k) + \frac{(\sigma t)^2}{2! k^2} \Gamma(1 - 2k) + \cdots \right],$$

$$M(t) = e^{(\mu - \frac{\sigma}{k})t} \left[1 + \sum_{n=1}^\infty \frac{(\sigma t)^n \Gamma(1 - nk)}{n! k^n} \right]. \tag{6}$$

3. PARMETERS ESTIMATION UNDER PROGRESSIVE TYPE-II CENSORED SAMPLES

For the generalized extreme value (GEV) distribution and under progressive type-II censored in (1), the likelihood function will be as follows:

$$L(X; \mu, \sigma, k|R) = \prod_{i=1}^{m} C_i \left[\frac{1}{\sigma} y_i^{k+1} e^{-y_i}\right][1 - e^{-y_i}]^{r_i}, \tag{7}$$

where $y_i = [1 + \frac{k}{\sigma}(x_i - \mu)]^{-\frac{1}{k}}$

The logarithm of the likelihood function in (7) becomes:

$$\ln L = \sum_{i=1}^{m} C_i - m \ln \sigma + (k+1) \sum_{i=1}^{m} \ln y_i - \sum_{i=1}^{m} y_i + \sum_{i=1}^{m} r_i \ln[1 - e^{-y_i}]. \tag{8}$$

To find the MLEs for the unknown parameters μ, σ and k, we find the first derivatives of the equation (8) with respect to μ, σ and k and equating it by 0, the system of non linear equations to get $\hat{\mu}$, $\hat{\sigma}$ and $\hat{k}$ are obtained as follows:

$$(\hat{k}+1) \sum_{i=1}^{m} y_i^k - \sum_{i=1}^{m} y_i^{k+1} + \sum_{j=1}^{m} r_j \frac{y_i^{k+1}}{(e^{y_i}-1)} = 0,$$

$$m - \sum_{i=1}^{m} y_i + \sum_{j=1}^{m} r_j \frac{y_i}{(e^{y_i}-1)} = 0,$$

and

$$m + \sum_{i=1}^{m} \ln y_i - \sum_{i=1}^{m} y_i \ln y_i + \sum_{j=1}^{m} r_j \frac{y_i \ln y_i}{(e^{y_i}-1)} = 0, \tag{9}$$

where $y_i = [1 + \frac{k}{\sigma}(x_i - \mu)]^{-\frac{1}{k}}$.

No explicit solution for this system was found, so that a numerical method such as Mathcad package and computer facilities is needed to obtain the value of estimators.

The approximate asymptotic variance-covariance matrix for μ, σ and k can be obtained by inverting the information matrix with the elements that are negative of the expected values of the second order derivatives of logarithms of the likelihood functions. Cohen (1965) [2] concluded that the approximate variance covariance matrix may be obtained by replacing expected values by their MLE's. Now the Fisher information matrix associated with μ, σ and k is defined as:

$$I(\mu, \sigma, k) = -E \begin{bmatrix} \frac{\partial^2 \ln L}{\partial \mu^2} & \frac{\partial^2 \ln L}{\partial \mu \partial \sigma} & \frac{\partial^2 \ln L}{\partial \mu \partial k} \\ \frac{\partial^2 \ln L}{\partial \mu \partial \sigma} & \frac{\partial^2 \ln L}{\partial \sigma^2} & \frac{\partial^2 \ln L}{\partial \sigma \partial k} \\ \frac{\partial^2 \ln L}{\partial \mu \partial k} & \frac{\partial^2 \ln L}{\partial \sigma \partial k} & \frac{\partial^2 \ln L}{\partial k^2} \end{bmatrix}. \tag{10}$$

Therefore, the elements of the information matrix approximately are given by:

$$\frac{\partial^2 \ln L}{\partial \mu^2} = \frac{\hat{k}(\hat{k}+1)}{\hat{\sigma}^2} \sum_{i=1}^{m} y_i^{2\hat{k}} - \frac{(\hat{k}+1)}{\hat{\sigma}^2} A - \frac{1}{\hat{\sigma}^2} \sum_{j=1}^{m} r_j \frac{y_i^{2\hat{k}+2} e^{y_i}}{(e^{y_i}-1)^2},$$

$$\frac{\partial^2 \ln L}{\partial \sigma^2} = -\frac{m}{\hat{\sigma}^2 \hat{k}} + \frac{(\hat{k}+1)}{\hat{\sigma}^2 \hat{k}} \sum_{i=1}^{m} y_i^{2\hat{k}} + \frac{(\hat{k}-1)}{\hat{\sigma}^2 \hat{k}^2} B + \frac{2}{\hat{\sigma}^2 \hat{k}^2} C - \frac{(\hat{k}+1)}{\hat{\sigma}^2 \hat{k}^2} A - \frac{1}{\hat{\sigma}^2 \hat{k}^2} \sum_{i=1}^{m} r_i \frac{y_i^2 e^{y_i}}{(e^{y_i}-1)^2} + \frac{2}{\hat{\sigma}^2 \hat{k}^2} \sum_{i=1}^{m} r_i \frac{y_i^{\hat{k}+2} e^{y_i}}{(e^{y_i}-1)^2} - \frac{1}{\hat{\sigma}^2 \hat{k}^2} \sum_{i=1}^{m} r_i \frac{y_i^{2\hat{k}+2} e^{y_i}}{(e^{y_i}-1)^2}$$

,

$$\frac{\partial^2 \ln L}{\partial k^2} = \frac{(\hat{k}+3)}{\hat{k}^3} m - \frac{(3\hat{k}+1)}{\hat{k}^4} B + \frac{2(2\hat{k}+1)}{\hat{k}^4} C - \frac{2(\hat{k}+1)}{\hat{k}^3} E + \frac{2}{\hat{k}^3} D - \frac{(\hat{k}+1)}{\hat{k}^4} A - \frac{1}{\hat{k}^2} F + \frac{2}{\hat{k}^2} \sum_{i=1}^{m} \ln y_i - \frac{2(\hat{k}+2)}{\hat{k}^3} \sum_{i=1}^{m} y_i^{\hat{k}} + \frac{(\hat{k}+1)}{\hat{k}^3} \sum_{i=1}^{m} y_i^{2\hat{k}} - \frac{1}{\hat{k}^4} \sum_{i=1}^{m} r_i \frac{y_i^2 e^{y_i}}{(e^{y_i}-1)^2} + \frac{2}{\hat{k}^4} \sum_{i=1}^{m} r_i \frac{y_i^{\hat{k}+2} e^{y_i}}{(e^{y_i}-1)^2} - \frac{2}{\hat{k}^3} \sum_{i=1}^{m} r_i \frac{y_i^2 \ln y_i e^{y_i}}{(e^{y_i}-1)^2} - \frac{1}{\hat{k}^4} \sum_{i=1}^{m} r_i \frac{y_i^{2\hat{k}+2} e^{y_i}}{(e^{y_i}-1)^2} + \frac{2}{\hat{k}^3} \sum_{i=1}^{m} r_i \frac{y_i^{\hat{k}+2} \ln y_i e^{y_i}}{(e^{y_i}-1)^2} - \frac{1}{\hat{k}^2} \sum_{i=1}^{m} r_i \frac{(y_i \ln y_i)^2 e^{y_i}}{(e^{y_i}-1)^2}$$

,

$$\frac{\partial^2 \ln L}{\partial \mu \partial \sigma} = -\frac{(\hat{k}+1)}{\hat{\sigma}^2} \sum_{i=1}^{m} y_i^{2\hat{k}} - \frac{1}{\hat{\sigma}^2 \hat{k}} C + \frac{(\hat{k}+1)}{\hat{\sigma}^2 \hat{k}} A - \frac{1}{\hat{\sigma}^2 \hat{k}} \sum_{i=1}^{m} r_i \frac{y_i^{\hat{k}+2} e^{y_i}}{(e^{y_i}-1)^2} + \frac{1}{\hat{\sigma}^2 \hat{k}} \sum_{i=1}^{m} r_i \frac{y_i^{2\hat{k}+2} e^{y_i}}{(e^{y_i}-1)^2},$$

$$\frac{\partial^2 \ln L}{\partial \mu \partial k} = -\frac{1}{\hat{\sigma}\hat{k}}\sum_{i=1}^{m} y_i^{\hat{k}} + \frac{(\hat{k}+1)}{\hat{\sigma}\hat{k}}\sum_{i=1}^{m} y_i^{2\hat{k}} + \frac{(\hat{k}+1)}{\hat{\sigma}\hat{k}^2}C - \frac{(\hat{k}+1)}{\hat{\sigma}\hat{k}^2}A + \frac{1}{\hat{\sigma}\hat{k}}D + \frac{1}{\hat{\sigma}\hat{k}^2}\sum_{i=1}^{m}r_i\frac{y_i^{\hat{k}+2}e^{y_i}}{(e^{y_i}-1)^2} - \frac{1}{\hat{\sigma}\hat{k}^2}\sum_{i=1}^{m}r_i\frac{y_i^{2\hat{k}+2}e^{y_i}}{(e^{y_i}-1)^2} + \frac{1}{\hat{\sigma}\hat{k}}\sum_{i=1}^{m}r_i\frac{y_i^{\hat{k}+2}\ln y_i e^{y_i}}{(e^{y_i}-1)^2}$$

and

$$\frac{\partial^2 \ln L}{\partial \sigma \partial k} = -\frac{m}{\hat{\sigma}\hat{k}^2} + \frac{(\hat{k}+2)}{\hat{\sigma}\hat{k}^2}\sum_{i=1}^{m} y_i^{\hat{k}} - \frac{(\hat{k}+1)}{\hat{\sigma}\hat{k}^2}\sum_{i=1}^{m} y_i^{2\hat{k}} + \frac{(\hat{k}+1)}{\hat{\sigma}\hat{k}^3}B - \frac{2(\hat{k}+1)}{\hat{\sigma}\hat{k}^3}C + \frac{1}{\hat{\sigma}\hat{k}^2}E - \frac{1}{\hat{\sigma}\hat{k}^2}D + \frac{(\hat{k}+1)}{\hat{\sigma}\hat{k}^3}A + \frac{1}{\hat{\sigma}\hat{k}^3}\sum_{i=1}^{m}r_i\frac{y_i^2 e^{y_i}}{(e^{y_i}-1)^2} -$$
$$\frac{2}{\hat{\sigma}\hat{k}^3}\sum_{i=1}^{m}r_i\frac{y_i^{\hat{k}+2}e^{y_i}}{(e^{y_i}-1)^2} + \frac{1}{\hat{\sigma}\hat{k}^2}\sum_{i=1}^{m}r_i\frac{y_i^2\ln y_i e^{y_i}}{(e^{y_i}-1)^2} + \frac{1}{\hat{\sigma}\hat{k}^3}\sum_{i=1}^{m}r_i\frac{y_i^{2\hat{k}+2}e^{y_i}}{(e^{y_i}-1)^2} - \frac{1}{\hat{\sigma}\hat{k}^2}\sum_{i=1}^{m}r_i\frac{y_i^{\hat{k}+2}\ln y_i e^{y_i}}{(e^{y_i}-1)^2}$$

$$(11)$$

where $A = \sum_{i=1}^{m} y_i^{2\hat{k}+1} - \sum_{j=1}^{m} r_j \frac{y_i^{2\hat{k}+1}}{(e^{y_i}-1)}$,

$B = \sum_{i=1}^{m} y_i - \sum_{i=1}^{m} r_i \frac{y_i}{(e^{y_i}-1)}$,

$C = \sum_{i=1}^{m} y_i^{\hat{k}+1} - \sum_{i=1}^{m} r_i \frac{y_i^{\hat{k}+1}}{(e^{y_i}-1)}$,

$D = \sum_{i=1}^{m} y_i^{\hat{k}+1}\ln y_i - \sum_{i=1}^{m} r_i \frac{y_i^{\hat{k}+1}\ln y_i}{(e^{y_i}-1)}$,

$E = \sum_{i=1}^{m} y_i \ln y_i - \sum_{i=1}^{m} r_i \frac{y_i \ln y_i}{(e^{y_i}-1)}$,

and $F = \sum_{i=1}^{m} y_i (\ln y_i)^2 - \sum_{i=1}^{m} r_i \frac{y_i(\ln y_i)^2}{(e^{y_i}-1)}$.

The asymptotic distributions of MLE parameters are

$$\sqrt{n}\left(\hat{\theta}_i - \theta_i\right) \approx N_3\left(0, I^{-1}(\hat{\theta}_i)\right); i = 1, 2, 3. \tag{12}$$

The approximation $100(1-\alpha)\%$ confidence intervals (C. I) for unknown parameters based on the asymptotic distribution of the GEV are determined as

$$\hat{\theta}_i \pm Z_{\alpha/2}\sqrt{I^{-1}(\hat{\theta}_i)}; i = 1,2,3, \tag{13}$$

where $Z_{\alpha/2}$ is the upper $\alpha/2$ percentile of a standard normal distribution.
.

4. NUMERICAL ILUSTERATION

In this section, a numerical example is presented to illustrate the computational methods detailed in the preceding sections. Using Mathcad package, we follow the following steps:

1. Using equation (1), generate random samples of size $n = 30, 50, \text{and } 100$ for the values of parameters $\mu = 1.7, \sigma = 0.6$ and $k = 0.3$.
2. Under progressive type-II censored samples, let we have $r_1 = 0$, $r_2 = 4$, $r_3 = 3$, $r_4 = 5$, $r_5 = 4$ and $r_6 = 10$, where r_i is the number of surviving units that are randomly selected and removed from the experiment at the ith stage. Using equations (9) and (11), calculate the estimators for the three parameters and its variances.
3. Using equation (13), calculate the confidence intervals (C. I) for each estimator at 95% level of significance.
4. Repeat the last steps to find the complete case by putting $r_1 = r_2 = r_3 = r_4 = r_5 = r_6 = 0$.

Table (1): Parameters Estimation of GEV Under Progressive Type-II Censored and Confidence Interval C. I ($\mu = 1.7$, $\sigma = 0.6$ and $k = 0.3$)

Scheme		Progressive type-II censored		Complete	
Sample size		Estimator $C.I.(\alpha = 0.05)$	Variance	Estimator $C.I.(\alpha = 0.05)$	Variance
$n = 100$	$\hat{\sigma}$	0.283 (-0.479, 1.045)	0.151	0.637 (0.526, 0.748)	0.003
	$\hat{k}$	-0.002 (-3.871, 3.867)	3.896	0.218 (0.074, 0.362)	0.005
$n = 50$	$\hat{\mu}$	1.411 (0.779, 2.043)	0.104	1.703 (1.497, 1.909)	0.011
	$\hat{\sigma}$	0.403 (-1.531, 2.337)	0.974	0.653 (0.49, 0.816)	0,006
$n = 30$	$\hat{\mu}$	1.498 (0.615, 2.381)	0.203	1.829 (1.452, 2.206)	0.037
	$\hat{\sigma}$	0.434 (-1.801, 2.669)	1.300	0.966 (0.519, 1.413)	0.052
	$\hat{k}$	0.001 (-5.237, 5.241)	7.144	0.0002 (-0.123, 0.124)	0.003

A comprehensive numerical investigation is needed to study the properties of defined estimators numerically.

5. REFERENCES

[1] Balakrishan, N. and Aggarwala. R. "*Progressive Censoring- Theory, Methods and Applications*". Birkhauser, Boston, (2000).

[2] Cohen, A.C. "Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and Censored Samples". Technometrics, 7, 579-588, 1965.

[3] Coles, S.G. *"An Introduction to the Statistical Modeling of Extreme Values"*. Springer Series in Statistics, 2001.

[4] Embrechts, P., Klüppelberg, C. and Mikosch, T."*Modelling Extremal Events: for Insurance and Finance*". Springer, 1997.

[5] Katz, R.W., Parlange, M.B. and Naveau, P."Statistics of Extremes in Hydrology. Advances in Water Resources, 25, 1287-1304, 2002.

[6] McNeil, A.J., Frey, R. and Embrechts, P. "*Quantitative Risk Management*:*Concepts, Techniques, and Tools*" Princeton University Press, 2005.