SHORT COMMUNICATIONS

ESTIMATION OF GUMBEL DISTRIBUTION PARAMETERS

N. P. SINGH
Unit of the Project Coordinator,
(Rapeseed-Mustard), Haryana Agricultural University,
Hisar 125 004, India.

IN order to apply statistical tools successfully for making inferences, the knowledge of the structure of original population is essential. However, in practice this is rarely known and one needs to estimate the parameters involved with the available data.

In particular, for characters of continuous nature, the location and scale parameters of a distribution function are related with the corresponding parameters of extreme order statistics distribution function through sample size. If the attention is confined to the location, scale and shape parameters, the three asymptotic distribution functions of extreme order statistics given by Fisher and Tippett[1]

$$\begin{aligned} G(x) &= \exp(-\exp(-x)) \quad \infty \leqslant x \leqslant \infty && \text{(Gumbel distribution)},\\ H(x) &= \exp(-x^{-k}) && \text{if } x > 0: k > 0,\\ &= 0 && \text{if } x \leqslant 0 \\ & && \text{(Fréchet distribution)},\\ W(x) &= \exp(-(-x)^k) && \text{if } x \leqslant 0,\ k > 0,\\ &= 1 && \text{if } x > 0\\ & && \text{(Weibull distribution)}, \quad (1) \end{aligned}$$

may be used with reasonable precision irrespective of the nature of population.

Consequently, to estimate the location and scale parameters, it is essential to know the domain of attraction to which the population distribution belongs. In particular, if one has reasons to believe that population distribution belongs to the domain of attraction of $G(x)$, the parameters may be estimated through asymptote

$$G(x) = \exp\left(-\exp\left(-\left(\frac{x-u}{b}\right)\right)\right) \text{ (for largest values)},$$

$$G(x) = 1 - \exp\left(-\exp\left(\frac{x-u}{b}\right)\right) \quad \text{(for smallest values)}.$$

Generally, Gumbel distribution is applied to solve important problems related to engineering technology, resource management and weather phenomenon such as estimation and forecasting of pressures, precipitation, snowfall, rainfall, temperature, flood etc. In all such problems, the data are recorded on a time scale. Let us consider the case of flood records. Suppose x_1, $x_2, \ldots, x_{365}$ be the daily maximum discharges of a river at a particular point. This can be rearranged in order of their magnitudes, *i.e.*, $z_1, z_2, \ldots z_{365}$ where z_{365} is the annual maximum flood. The studies on forecasting of different phenomenon distributed as Gumbel distribution and their subsequent estimation are based on z_{365} only. However, the values of lower order statistics are necessarily recorded to arrive at z_{365}. Keeping this in mind, the Gumbel distribution was estimated in the note using mth largest (smallest) order statistics of n samples by using the method of maximum likelihood estimation. The variance-covariance of the estimates was also calculated for different values of m for a comparative study[2].

ESTIMATES BASED ON *m*th ORDER STATISTICS

The asymptotic density function of the mth order statistics is given by

$$dG = \frac{1}{b\Gamma(m)} \exp(\pm mz) \exp[\pm \exp(-z)], \quad (2)$$

where $z = (x-u)/b$, the upper (lower) signs being those for the distribution of smallest (largest) values and u and b are location and scale parameters. The maximum likelihood estimates of u and b are

$$\hat{u} = \pm \hat{b} \log\left(\sum_{i=1}^{n} \exp(\pm x_i/\hat{b})/mn\right), \quad (3)$$

$$\hat{b} = \mp m\left[\bar{x}_i - \sum_{i=1}^{n} x_i \exp(\pm x_i/\hat{b}) \Big/ \sum_{i=1}^{n} \exp(\pm x_i/\hat{b})\right], \quad (4)$$

and variance covariance matrix is

$$\begin{bmatrix} \dfrac{mn}{b^2} & \mp \dfrac{mn\Gamma'(m+1)}{b^2\Gamma(m+1)} \\ \mp \dfrac{mn\Gamma'(m+1)}{b^2\Gamma(m+1)} & \dfrac{n}{b^2} + \dfrac{mn\Gamma''(m+1)}{b^2\Gamma(m+1)} \end{bmatrix}^{-1}. \quad (5)$$

It is clear that (3) involves $\hat{u}$ and $\hat{b}$ and (4) is function of only $\hat{b}$. Although in theory, it should be possible to find out the value of $\hat{b}$ from (4) but in practice, this is not so due to the involvement of summations and $\hat{b}$ as a power of an exponential function. Equation (4) can only be solved by iteration method. The first approximation for $\hat{b}$ was obtained by the method of moments and given as

$$\hat{b}_1 = \frac{S}{\left[\frac{\Gamma''(m)}{\Gamma(\bar{m})} - \left[\frac{\Gamma'(m)}{\Gamma(m)}\right]^2\right]^{1/2}}, \qquad (6)$$

where S is the standard deviation of the sample observations and Γ' and Γ'' are first and second derivatives of gamma function. The estimates for different value of m can be obtained from (3) and (4).

To make a comparative study the variances are calculated for $m = 1, 2, 3, ---, 30$. It has been observed that the efficiency, as indicated by variances, increased significantly: (n var$(\hat{u})/\hat{b}^2$ decreased from 1.108665 to 0.975751 and n var$(\hat{b})/\hat{b}^2$ decreased from 0.607927 to 0.558701), when instead of first maximum (minimum), 2nd maximum (minimum) order statistics are used to estimate the parameters of Gumbel distribution. However, when third maximum (minimum) order statistics are used to estimate the parameter of Gumbel distribution, the efficiency as indicated by variance, reduced significantly in case of location parameter and increased insignificantly in case of scale parameter; (n var$(\hat{u})/\hat{b}^2$ increased from 0.975751 to 1.185539 and n var$(\hat{b})/\hat{b}^2$ decreased from 0.558701 to 0.540112). If 4th maximum (minimum) order statistics are used instead of third, n var$(\hat{u})/\hat{b}^2$ increased from 1.185539 to 1.453204 and n var$(\hat{b})/\hat{b}^2$ decreased from 0.540112 to 0.530422. Similarly for other lower order statistics, n var$(\hat{u})/\hat{b}^2$ increases significantly but on the contrary the extent of reduction in n var$(\hat{b})/\hat{b}^2$ is insignificant. When the 30th minimum (maximum) order statistics are used for estimation, n var$(\hat{u})/\hat{b}^2$ and n var$(\hat{b})/\hat{b}^2$ were observed as 5.922497 and 0.504159 respectively which clearly indicated the rapid increase in the values of n var$(\hat{u})/\hat{b}^2$ and slow reduction in the values of n var$(\hat{b})/\hat{b}^2$. On the basis of variances it can be deduced that to obtain efficient maximum likelihood estimates of u and b the second maximum (minimum) can be fruitfully utilized. The correlation coefficients reach ± 1 as m increases. Here again, the change is maximum when second maximum (minimum) order statistics are used in place of first maximum (minimum) order statistics. Later on it reduces to zero.

1 January 1985

1. Fisher, R. A. and Tippett, L. H. C., *Proc. Camb. Phil. Soc.*, 1928, **24**, 180.
2. Singh, N. P., *Theory of extreme values and its application of flood control.* Unpublished doctoral dissertation, Haryana Agricultural University, Hisar, India, 1982.

ON THE ESTIMATION OF MEAN USING SUPPLEMENTARY INFORMATION ON TWO AUXILIARY CHARACTERS IN DOUBLE SAMPLING

H. P. SINGH

Department of Mathematics and Statistics,
College of Agricultural Engineering,
Jawaharlal Nehru Krishi Vishwa Vidyalaya,
Jabalpur 482004, India.

CONSIDER finite population of size N (N is assumed large here). Let y_1 and y_2 denote the two auxiliary characters and let y_0 denote character under study. Now when two auxiliary characters are used in a survey there may be three possible sampling schemes[1-4].

1. A preliminary sample of size n' is selected by simple random sampling without replacement (SRSWOR) for observing the auxiliary characters y_1, y_2 and smaller subsample of size n is selected from the n' observations.
2. The preliminary sample is selected as in above but the smaller sample of size n is selected independently.
3. Many times the auxiliary information may be collected by two different agencies and hence two independent preliminary samples of sizes n'_1 and n'_2 are selected for observing y_1 and y_2 and the smaller sample of size n is also selected independently from the population by SRSWOR. To facilitate comparison we take $n'_1 = n'_2 = n'$.

Further, in what follows we shall use the following notations throughout the investigation:

$\bar{y}_i$: sample means for y_i ($i = 0, 1, 2$) based on sample of size n,
$\bar{y}'_1$ and $\bar{y}'_2$: sample means of y_1 and y_2 respectively based on the larger sample of size n',
$\bar{Y}_i$: population mean of y_i ($i = 0, 1, 2$),
σ_i: standard deviation of y_i ($i = 0, 1, 2$),
C_i ($= \sigma_i/\bar{Y}_i$): population coefficient of variation of y_i ($i = 0, 1, 2$),