AN APPROXIMATION TO THE SAMPLING VARIANCE OF AN ESTIMATED MAXIMUM VALUE OF GIVEN FREQUENCY BASED ON FIT OF DOUBLY EXPONENTIAL DISTRIBUTION OF MAXIMUM VALUES[1]

By Bradford F. Kimball

N. Y. State Department of Public Service

1. Introduction. Given the doubly exponential distribution of maximum values

$$F(x) = \exp(-e^{-y}), \qquad y = \alpha(x - u), \tag{1}$$

where α and u are unknown parameters, with a prescribed frequency F_0 the "reduced variate" y is fixed, say at $y = y_0$. Thus with

$$F_0 = .99, \qquad y_0 = 4.60015 \cdots .$$

Given a sample of n maximum values x_i, we are interested in the sampling variance of

$$\hat{x} = g(\hat{u}, \hat{\alpha}) = \hat{u} + y_0/\hat{\alpha} \tag{2}$$

due to sampling variations of the estimates $\hat{u}$ and $\hat{\alpha}$.

H. Fairfield Smith has recently pointed out to me that the examples of applications of sufficient statistical estimation functions to this problem given in a previous paper (see [1, pp. 307–309]) give too large a range for $\hat{x} = g(\hat{u}, \hat{\alpha})$ because the sample points $(\hat{u}, \hat{\alpha})$ within the confidence region of the constant probability ellipse apply to optimum estimates of $(\hat{u}, \hat{\alpha})$ rather than to that of $g = g(\hat{u}, \hat{\alpha})$. What the problem calls for is the determination of the positions of curves $\bar{g}(u, \alpha)$ and $\underline{g}(u, \alpha)$ such that the integral of the *pdf* of the estimation functions over all sample values $(\hat{u}, \hat{\alpha})$ which lie between these two curves is equal to the confidence level (taken as .95 in previous paper). Further considerations of this being the shortest interval $\bar{g} - \underline{g}$, also come into play.

As so often happens in research, the previous analysis, although not giving the final answer, suggests the next step. If we change our parameters to

$$g = g(u, \alpha) = u + y_0/\alpha, \qquad \alpha' = \alpha \tag{3}$$

and are able to carry through the inverse of the maximum likelihood solution for fitting of (1) to n sample values x_i, then we shall be in a position to find the asymptotic marginal distribution of $\sqrt{n}(\hat{g} - g)$, which will give the answer to our problem (see [2]).

The Jacobian of this transformation of parameters is

$$\partial(u, \alpha)/\partial(g, \alpha') = \begin{vmatrix} 1 & y_0^2/\alpha'^2 \\ 0 & 1 \end{vmatrix} = 1,$$

and hence for $\alpha' > 0$ no new singularities are introduced.

[1] This involves a correction of a previous paper [1].

2. The equations of the maximum likelihood solution. For a sample of size n, the *pdf* of the sampling distribution in terms of the old parameters is given by

$$P[u, \alpha, O_n(x_i)] = \alpha^n \exp [-\Sigma e^{-\alpha(x_i-u)}] \exp [-\Sigma\alpha(x_i - u)],$$

and

$$\log P = n \log \alpha - \Sigma e^{-\alpha(x_i-u)} - \alpha\Sigma x_i + n\alpha u;$$

$$= n[\log \alpha - e^{\alpha u}(\Sigma e^{-\alpha x_i}/n) - \alpha\bar{x} + \alpha u].$$

Now change to the new parameters and use the substitutions:

$$z_i = e^{-\alpha x_i}, \qquad \bar{z} = (\Sigma z_i)/n, \qquad z_0 = e^{-\alpha u} = e^{y_0}\cdot e^{-\alpha' g}.$$

Thus

$$\partial z_0/\partial g = -\alpha' z_0, \qquad \partial z_0/\partial\alpha' = -g z_0,$$

and denoting $\log P$ by L we write

$$L = n[\log \alpha' - \bar{z}/z_0 - \alpha'\bar{x} + \alpha' g - y_0].$$

Hence

$$L_g = -n\alpha'[\bar{z}/z_0 - 1]; \tag{4}$$

$$L_{\alpha'} = n[1/\alpha' - \partial(\bar{z}/z_0)/\partial\alpha' - \bar{x} + g]. \tag{5}$$

3. Derivation of expected values needed. Recall that

$$\bar{z}/z_0 = e^{-y_0}\Sigma e^{-\alpha'(x_i-g)}/n = \Sigma e^{-\alpha(x_i-u)}/n.$$

Hence

$$\partial(\bar{z}/z_0)/\partial\alpha' = -e^{-y_0}\Sigma(x_i - g)e^{-\alpha'(x_i-g)}/n, \tag{6}$$

$$\partial(\bar{z}/z_0)/\partial\alpha = -\Sigma(x_i - u)e^{-\alpha(x_i-u)}/n;$$

$$\partial^2(\bar{z}/z_0)/\partial\alpha'^2 = e^{-y_0}\Sigma(x_i - g)^2 e^{-\alpha'(x_i-g)}, \tag{7}$$

$$\partial^2(\bar{z}/z_0)/\partial\alpha^2 = \Sigma(x_i - u)^2 e^{-\alpha(x_i-u)}/n.$$

By investigation of the generating function

$$G(t) = E[\Sigma(z_i/z_0)^{1-t}], \qquad z_i = e^{-\alpha x_i},$$

it can be shown that

$$E[\Sigma e^{-\alpha(x_i-u)}/n] = 1,$$

$$E[\Sigma(x_i - u)e^{-\alpha(x_i-u)}/n] = -(1/\alpha)\Gamma'(2) = -(1/\alpha)(1 - C),$$

where C denotes Euler's constant, .577216 $\cdots$, and

$$E[\Sigma(x_i - u)^2 e^{-\alpha(x_i-u)}/n] = (1/\alpha^2)\Gamma''(2) = (1/\alpha^2)(\pi^2/6 + C^2 - 2C).$$

Hence to find expected values of (6) and (7) we note that

$$-e^{-y_0}\Sigma(x_i - g)e^{-\alpha'(x_i-g)}/n = -\Sigma(x_i - g)e^{-\alpha(x_i-u)}/n;$$
$$= -\Sigma(x_i - u)e^{-\alpha(x_i-u)}/n + (y_0/\alpha)\Sigma e^{-\alpha(x_i-u)}/n,$$

and therefore

$$(8) \qquad E[\partial(\bar{z}/z_0)/\partial\alpha'] = E[\partial(\bar{z}/z_0)/\partial\alpha] + (y_0/\alpha)\, E(\bar{z}/z_0).$$

Similar analysis shows that

$$(9) \quad E[\partial^2(\bar{z}/z_0)/\partial\alpha'^2] = E[\partial^2(\bar{z}/z_0)/\partial\alpha^2] + (2y_0/\alpha)E[\partial(\bar{z}/z_0)/\partial\alpha] + (y_0^2/\alpha^2)E[\bar{z}/z_0].$$

4. The inverse of the maximum likelihood solution. It will first be noted that the maximum likelihood equations (4) and (5) for determining best estimates of g and α' become identical to those for determining best estimates of old parameters u and α, when the transformation of parameters (3) is applied to them. This is easily verified by applying relations developed above.[2]

This means that *the best estimates* $\hat{g}$ *and* $\hat{\alpha}'$ *obtained from* (4) *and* (5) *are related to the best estimates of old parameters* $\hat{u}$ *and* $\hat{\alpha}$ *by*

$$(10) \qquad \hat{g} = \hat{u} + y_0/\hat{\alpha}, \qquad \hat{\alpha}' = \hat{\alpha}.$$

We now proceed to set up the inverse of the maximum likelihood solution. In order to do this we first need the variance-covariance matrix of the direct solution. This is (see [2])

$$\begin{Vmatrix} E[-L_{gg}] & E[-L_{g\alpha'}] \\ E[-L_{\alpha' g}] & E[-L_{\alpha'\alpha'}] \end{Vmatrix}.$$

Now

$$L_{gg} = -n\alpha'^2(\bar{z}/z_0), \qquad E[-L_{gg}] = n\alpha'^2,$$
$$L_{g\alpha'} = -n[\bar{z}/z_0 - 1 + \alpha'\partial(\bar{z}/z_0)/\partial\alpha'], \qquad E[L_{g\alpha'}] = n(1 - C + y_0),$$
$$L_{\alpha'\alpha'} = -n[1/\alpha'^2 - \partial^2(\bar{z}/z_0)/\partial\alpha'^2],$$
$$E[-L_{\alpha'\alpha'}] = (n/\alpha'^2)[\pi^2/6 + (1 - C + y_0)^2].$$

Thus the variance-covariance matrix of the estimation functions (4) and (5) is

$$\begin{Vmatrix} n\alpha'^2 & n(1 - C + y_0) \\ n(1 - C + y_0) & (n/\alpha'^2)[\pi^2/6 + (1 - C + y_0)^2] \end{Vmatrix}.$$

The asymptotic form of the inverse solution for $\sqrt{n}\,(\hat{g} - g)$ and $\sqrt{n}\,(\hat{\alpha}' - \alpha')$ will have the variance-covariance matrix which is the reciprocal of the above matrix, multiplied by n. The determinant value of the above matrix reduces to $n^2(\pi^2/6)$. Thus the reciprocal matrix, adjusted by multiplying by n, is

[2] See equations (5.2) of [1] and note $+\partial(\bar{z}/z_0)/\partial\alpha$ in second equation of (5.2) should read $-\partial(\bar{z}/z_0)/\partial\alpha$.

$$\text{(11)} \qquad \left\| \begin{matrix} (1/\alpha'^2)[1 + (1 - C + y_0)^2/(\pi^2/6)] & -(1 - C + y_0)/(\pi^2/6) \\ -(1 - C + y_0)/\pi^2/6) & \alpha'^2/(\pi^2/6) \end{matrix} \right\|.$$

This gives the solution sought. From the general theory of the maximum likelihood solution (see [2]) the distribution of $[\sqrt{n}(\hat{g} - g), \sqrt{n}(\hat{\alpha}' - \alpha')]$ is asymptotically normal. Hence *the marginal distribution of $\sqrt{n}(\hat{g} - g)$ will be asymptotically normal, and for finite n, the standard deviation may be approximated by*

$$\text{(12)} \qquad \sigma(\hat{g} - g) = [1/(\sqrt{n}\alpha')]\sqrt{1 + (1 - C + y_0)^2/(\pi^2/6)}.$$

Now the correlation coefficient for the asymptotic bivariate normal distribution is seen to be

$$r = -(1 - C + y_0)/\sqrt{\pi^2/6 + (1 - C + y_0)^2}.$$

If α' were known, we should have the standard deviation of $\sqrt{n}(\hat{g} - g)$ reduced by factor $\sqrt{1 - r^2}$. This is found to be equal to the reciprocal of the second factor in the equation (12). Hence we conclude that *if α' be known, the standard deviation of $(\hat{g} - g)$, for finite n, is given approximately by*

$$\text{(13)} \qquad \sigma(\hat{g} - g) = 1/(\sqrt{n}\alpha').$$

5. An example. Using same example outlined in previous paper (see [1, pp. 307–309]), we have $n = 57$, $\hat{\alpha}' = .01924$, $1 - C = .422784$, $y_0 = 4.60015$. This gives $\sigma = 27.826$. For 95% confidence interval we take $(1.96)\sigma = 54.54$, and with $\hat{u} = 180.6$,

$$\hat{g} = \hat{u} + y_0/\hat{\alpha} = 419.7,$$

and the interval is approximated by

$$| \hat{g} - g | < 54.5,$$

which as an approximation gives the symmetrical interval

$$365.2 < g < 474.2.$$

Method 4 used in previous paper gave the longer interval (see Introduction) which was not symmetrical about $\hat{g}$;

$$362.8 < g < 507.4.$$

REFERENCES

[1] B. F. Kimball, "Sufficient statistical estimation functions for the parameters of the distribution of maximum values," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 299–309.

[2] S. S. Wilks, *Mathematical Statistics*, Princeton Univ. Press, 1943, p. 139.