THE GEOMETRIC RANGE FOR DISTRIBUTIONS OF CAUCHY'S TYPE

By E. J. Gumbel and R. D. Keeney

New York City and Metropolitan Life Insurance Company

1. Introduction. We consider large samples drawn from a symmetrical unlimited population whose distribution is of the Cauchy type, defined by the properties

$$(1) \qquad \lim_{x \to \infty} x^k[1 - F(x)] = A, \qquad \lim_{x \to -\infty} (-x)^k F(x) = A,$$

where k and A are positive and $F(x)$ stands for the probability function. This type of distribution has no moments of an order equal to or greater than k. We construct the distribution of a certain function of the extreme values, and require only the knowledge of the type of the initial distribution, not of the distribution itself.

From each sample we pick out the largest and smallest observations, x_n and x_1. If the median of the initial distribution is zero, and the sample size is large enough, the probability of any extreme x_n or $-x_1$ being negative can be neglected. If we draw N such samples, each of large size n, we obtain N pairs of extremes, $x_{n,\nu}$ and $x_{1,\nu}$ $(\nu = 1, 2, 3, \cdots, N)$. For each sample we can then compute the geometric mean, ρ, of these extremes:

$$(2) \qquad \rho = \sqrt{x_n(-x_1)},$$

which we henceforth call the *geometric* range.

The distribution of these geometric ranges can be obtained directly from the joint asymptotic distribution of the extremes. However, it is easier to obtain this distribution indirectly from the distribution of the reciprocal of the geometric range. This distribution of the reciprocal is of interest in itself: since it possesses all moments we can use it to estimate the parameters by the method of moments, whereas this problem seems to be very intricate if we start from the distribution of the geometric range itself.

2. The distribution of the reciprocal of the geometric range. The distribution of the reciprocal of the geometric range follows from a theorem of Elfving [1] which may be stated thus:

"Let x be a symmetrical unlimited variate with probability $F(x)$. Let ξ be defined by

$$(3) \qquad \xi = 2n\sqrt{F(x_1)[1 - F(x_n)]}.$$

Then the asymptotic density function $g(\xi)$ and the asymptotic probability $G(\xi)$ of ξ are:

$$(4) \qquad g(\xi) = \xi K_0(\xi); \qquad G(\xi) = 1 - \xi K_1(\xi),$$

where K_0 and K_1 are the modified Bessel functions of the second kind and of order zero and one."

Introducing instead of A the parameter u defined by $F(u) = 1 - 1/n$ we have, from (1), approximately for large n

$$(5) \quad F(x_1) = 1/n\left(\frac{u}{-x_1}\right)^k, \quad 1 - F(x_n) = 1/n\left(\frac{u}{x_n}\right)^k, \quad x_1 \le 0, x_n \ge 0, k > 0.$$

For the variable ξ in Elfving's theorem, we obtain asymptotically

$$(6) \qquad \xi_k/2 = u^k\rho^{-k}.$$

We attach a subscript k to ξ to show its dependence on k. The moments of ξ_k are obtained from a formula given by Watson ([3], p. 388) as

$$(7) \qquad \overline{\xi_k^l} = 2^l\Gamma^2(1 + l/2)$$

and all moments of this variate exist.

3. Estimate of parameters. From N sets, each of n observations, we pick out the largest and the smallest, $X_{n,\nu}$ and $X_{1,\nu}$. We subtract from each observed extreme the central value, m, of the $N \cdot n$ observations. If each $x_{n,\nu} = X_{n,\nu} - \text{m} \ge 0$ and $x_{1,\nu} = X_{1,\nu} - \text{m} \le 0$ the sample size is large enough.

Define $\eta = 1/\rho$. The first two moments of η are, from (7),

$$(8) \qquad \bar{\eta} = \frac{1}{u}\Gamma^2(1 + 1/2k), \qquad \overline{\eta^2} = \frac{1}{u^2}\Gamma^2(1 + 1/k).$$

Elimination of the parameter u from these two equations leads to

$$\frac{\overline{\eta^2}}{\bar{\eta}^2} = \frac{\Gamma^2(1 + 1/k)}{\Gamma^4(1 + 1/2k)}.$$

In terms of the coefficient of variation, V, this equation becomes

$$(9) \qquad \sqrt{1 + V^2} = \Gamma(1 + 1/k)/\Gamma^2(1 + 1/2k).$$

Substituting the value of V computed from the observations, we obtain an estimate of k, and hence can obtain an estimate of u from (8). This procedure is facilitated by Table 1.

4. The distribution of the geometric range. From a practical standpoint the geometric range itself is preferable to its reciprocal since it is easier to interpret and easier to calculate from the observed extremes. We want to establish its distribution $g_1(\rho)$. From the relation (6) of ρ to ξ_k and the knowledge of the distribution (4) of ξ_k we find

$$(10) \qquad G_1(\rho) = 1 - G(\xi_k) = 2u^k\rho^{-k}K_1(2u^k\rho^{-k})$$

and

$$(11) \qquad g_1(\rho) = \frac{2\xi_k\, ku^k}{\rho^{k+1}}K_0(\xi_k) = \frac{4k}{u}\left(\frac{u}{\rho}\right)^{2k+1}K_0\left(\frac{2u^k}{\rho^k}\right).$$

Since tables of these Bessel functions are available [2], the various probabilities and densities may be evaluated.

The simplest way to compare geometric ranges to the theory is the use of a probability paper (Figure 1). For its construction, consider the linear relation

$$\log \rho = \log u + (\log 2)/k - (\log \xi_k)/k \tag{12}$$

obtained from (6). Consequently we plot $-\log \xi_k$ on the abcissa and write the corresponding values $G_1(\rho)$, formula (10), on a horizontal axis. An upper parallel to the abscissa shows the return periods. The observed geometric ranges are plotted on the ordinate in a logarithmic scale. If the theory holds, the observed geometric ranges should be scattered about the straight line (12).

TABLE 1

The order k and the variation V of the reciprocal of the geometric range

Reciprocal Order $1/k$	Coefficient of variation V	Reciprocal Order $1/k$	Coefficient of variation V
.10	.088	.70	.556
.12	.104	.80	.632
.16	.138	.90	.709
.20	.171	.98	.772
.30	.251	1.00	.788
.40	.332	2.00	1.73
.50	.404	4.00	5.92
.60	.480	6.00	20.0

If less accurate estimates of u and k than those obtainable by the systematic methods (8) and (9), or the probability paper, will suffice, quick estimates can be obtained from the quantiles of the sample of geometric ranges. To the value $\rho = u$ corresponds, according to (6), $\xi_k = 2$ whence, from the tables [2], $G_1(u) = 2K_1(2) = .27973$. From N observed geometric ranges arranged in increasing magnitude we thus may pick out the mth, ρ_m, with the rank $m = .28\,N$ and use it as an estimate $u = \rho_m$. For the medians $\tilde{\xi}_k$ and $\tilde{\rho}$ we get $\tilde{\xi}_k = 1.257$ from the tables, and thus, by (6), $\tilde{\rho}^k = 1.591\, u^k$. This formula provides a quick estimate of k. We pick out the median $\tilde{\rho}$ of the N observed geometric ranges. Since we have an estimate of u, we obtain an estimate of k from

$$\frac{1}{k} = \frac{\log \tilde{\rho} - \log u}{\log 1.591} = 4.960 \log [\tilde{\rho}/\rho_m]. \tag{13}$$

5. Analogy between the geometric range and the range. A study of the various characteristics of the geometric range for distributions of Cauchy's type reveals structural similarities to the range for distributions of the exponential type.

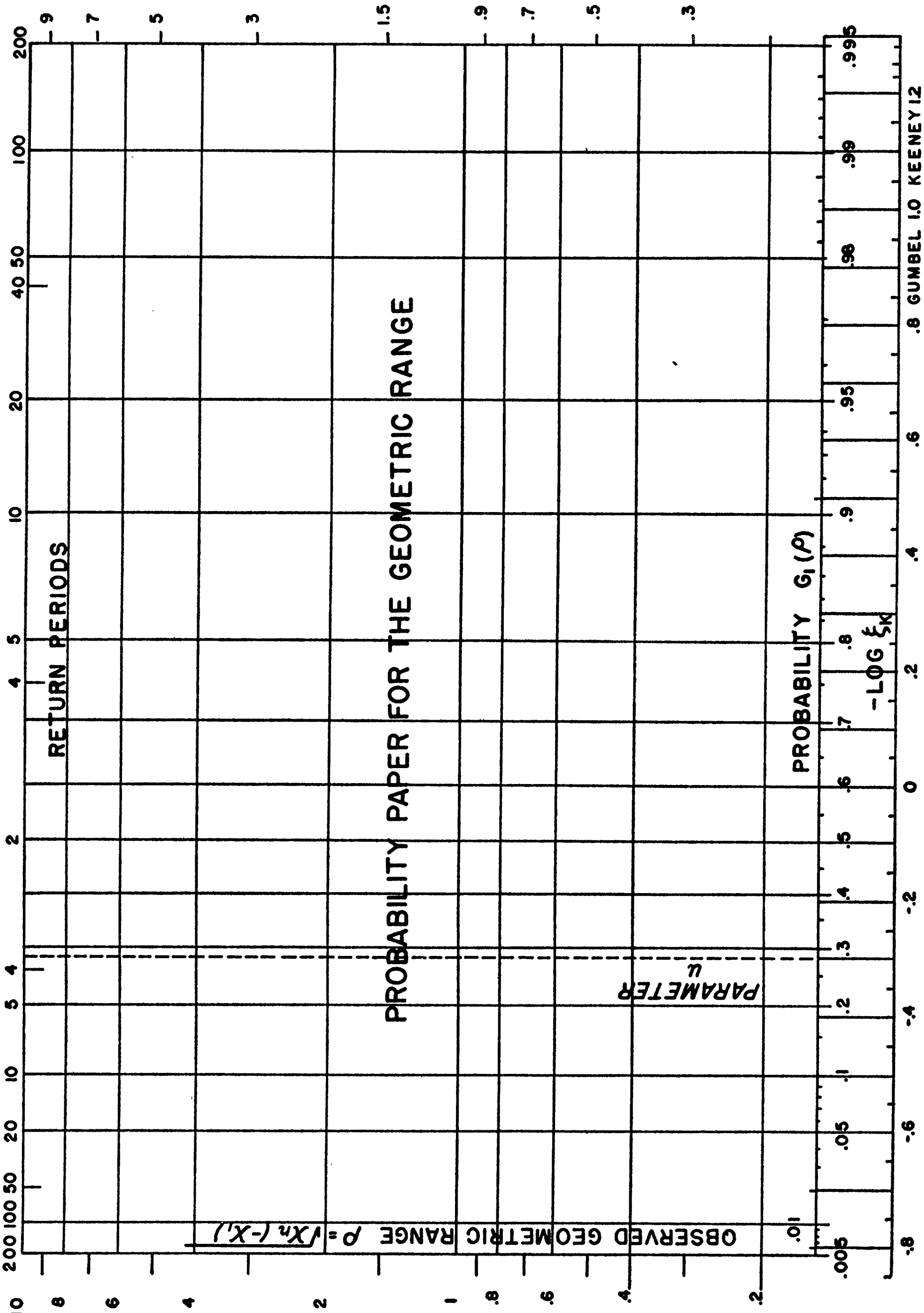
PROBABILITY PAPER FOR THE GEOMETRIC RANGE
RETURN PERIODS
PROBABILITY $G_1(\rho)$
$-\mathrm{LOG}\ \xi_K$
PARAMETER u
OBSERVED GEOMETRIC RANGE
GUMBEL 1.0 KEENEY 1.2

FIGURE 1

This is not altogether surprising, since (as shown in Table 2) after the appropriate transformations the probabilities of both are identical functions of the respective transformed variates.

Of course the two systems are mutually exclusive: if the observed ranges can be reproduced by the first system we conclude that all moments in the initial distribution exist. If on the other hand, the observed geometric ranges can be represented by the second sytem we conclude that no moments of an order greater than k exist.

TABLE 2
RANGES AND GEOMETRIC RANGES

Type of Initial Distribution	Exponential	Gauchy
Variate	Range	Geometric Range
Definition	$w = x_n + (-x_1)$	$\rho = \sqrt{x_n(-x_1)}$
Transformation	$z = 2\exp\left[-\frac{\alpha}{2}(x_n - x_1 - 2u)\right]$	$\xi_k = 2u^k \rho^{-k}$
Logarithm	$\lg z = \lg 2 - \frac{\alpha}{2}(x_n - x_1 - 2u)$	$\lg \xi_k = \lg 2 - \frac{k}{2}(\lg x_n + \lg(-x_1) - 2\lg u)$
Probability	$G(w) = z\,K_1(z)$	$G_1(\rho) = \xi_k\,K_1(\xi_k)$
Distribution	$g(w) = \frac{\alpha z^2}{2} K_0(z)$	$g_1(\rho) = \frac{4k}{u}\left(\frac{\xi_k}{2}\right)^{2k+1} K_0(\xi_k)$
Median	$\tilde{w} = 2u + .9286/\alpha$	$2\lg\tilde{\rho} = 2\lg u + .9286/k$
Mean	$\bar{w} = 2u + 2\gamma/\alpha$	$\lg\overline{\rho^{-1}} = -\lg u + 2\lg\Gamma(1 + \frac{1}{2}k)$

REFERENCES

[1] G. Elfving, "The asymptotical distribution of range in samples from a normal population," *Biometrika*, Vol. 35 (1947).
[2] *Tables of the Bessel-functions*, Vol. 6, British Association for the Advancement of Science, Cambridge, 1937.
[3] G. N. Watson, *Theory of Bessel-functions*, Cambridge University Press, 1944.

REMARK ON W. M. KINCAID'S "NOTE ON THE ERROR IN INTERPOLATION OF A FUNCTION OF TWO INDEPENDENT VARIABLES"

By T. N. E. Greville

Federal Security Agency

In a review of Dr. W. M. Kincaid's "Note on the Error in Interpolation of a Function of Two Independent Variables," (*Annals of Math. Stat.*, Vol. 19 (1948),