JOURNAL OF GEOPHYSICAL RESEARCH VOL. 68, NO. 3 FEBRUARY 1, 1963

A Plotting Rule for Extreme Probability Paper

IRVING I. GRINGORTEN

Air Force Cambridge Research Laboratories
Bedford, Massachusetts

Abstract. A primary purpose in plotting a set of ordered observations on probability paper is to simplify the inspection of their distribution. With this limited objective a plotting rule for the ith of n ordered observations from a double exponential distribution is obtained as $\hat{P}_i = (i - 0.44)/(n + 0.12)$, where n is the number of observations, i is the order of the observation from the smallest, and $\hat{P}_i$ is the estimate of the cumulative frequency of the ith term.

Probability paper for a given cumulative distribution $P(x)$ of a random variable x is so designed that a plot of $P(x)$ against x or a function of x is a straight line. But for a finite, ordered sample of observations, $x_1, \cdots, x_i, \cdots, x_n$, we must first decide how to estimate $P(x_i)$ before we can plot the n finite values on probability paper.

Kimball [1960], after a careful examination of the problem, concludes that the choice of plotting rule should be governed by the purpose of plotting the data. He lists three different kinds of purposes, of which the first purpose is the adopted objective of this paper, 'to test whether the sample data indicate that the universe is of the prescribed type.'

This study is further limited to a graphic test for a double exponential distribution of a variable x,

$$P(x) = \exp(-e^{-y}) \quad y = \gamma + (x - m)/\beta \quad (1)$$

where $m = E(x)$, $\beta = \sqrt{6}/\pi$ times the standard deviation of x, and $\gamma = 0.5772 \cdots$ (Euler's constant). The corresponding probability paper (Figure 1) is known as extreme probability paper, first suggested by *Powell* [1942], prepared by *Gumbel* [1943], and modified by *Court* [1952].

Two plotting formulas have frequently been used, for specified purposes, for the ith of n ordered observations [*Gumbel*, 1958] on any probability paper. These are

$$\hat{P}_i = (i - 1/2)/n \quad (2)$$

$$\hat{P}_i = i/(n + 1) \quad (3)$$

The latter formula, as shown by Gumbel, utilizes the expected probability for each ordered observation regardless of the basic distribution of the variate. *Blom* [1958] has suggested another formula for the Gaussian distribution on normal probability paper which both he and Kimball have shown to have special merit:

$$\hat{P}_i = (i - 3/8)/(n + 1/4) \quad (4)$$

All three formulas (2, 3, 4) are of the general form

$$\hat{P}_i = (i - a)/(n + 1 - 2a) \quad 0 \leq a < 1 \quad (5)$$

differing only in the value of a. All n observations can be plotted on a finite piece of probability paper since $0 < P_1 < P_n < 1$.

If (5) is accepted, the problem of selecting a suitable plotting formula becomes the problem of finding a suitable value for a. Gumbel has shown that the return period of the largest value approaches the sample size in plotting position when (3) is used. In this article, also, the largest value is at the core of the problem. But its plotting position $\hat{P}_n$ is made the same as that for the value of x from which it is expected to depart the least (in the mean square sense). Since it is well known that this value is the true mean, we make

$$\hat{P}_n = P[E(x_n)]$$

This condition should place the largest observed value as close as possible to the ideal straight line where it is most important to have a good fit.

For the double exponential distribution [*Gumbel*, 1958, p. 116] the mean of the highest of n reduced variates is

$$E(y_n) = \gamma + \ln n$$

which gives $P(x_n)$ from (1) above. If P_n is made equal to $P(\gamma + \ln n) = P$ then the value of a is obtained from (5),

$$a = (P - n + nP)/(2P - 1) \quad (6)$$

Values of $E(y_i)$ have been given by *Lieblein and Salzer* [1957]. For $i = n$, the values of P corresponding to $E(y_n)$, when inserted in (6), give values of a against n as follows:

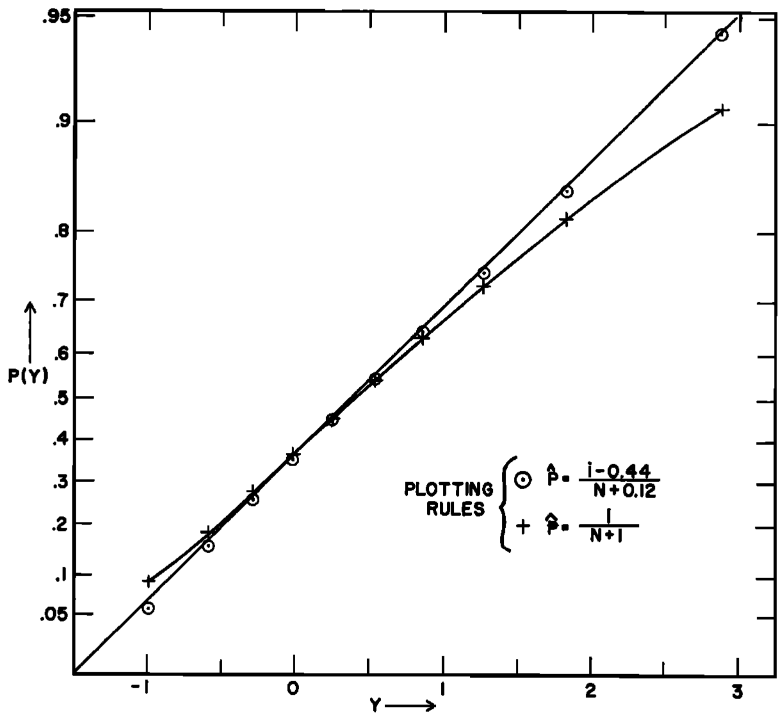

Fig. 1. Sample of extreme probability paper. The ten small circles are at the plotted positions of the expected values of ten random numbers, after ordering, by the plotting formula of this paper. The ten crosses are plotted in accordance with a commonly accepted plotting formula (3).

n	10	20	30	40	50
	60	70	80	90	100
a	0.448	0.443	0.442	0.441	0.440
	0.440	0.440	0.440	0.439	0.439

For $n > 100$ the probability of the largest value is $P = \exp[-e^{-\gamma}/n + O(1/n^2)]$, where $O(1/n^2)$ is a term of order $1/n^2$. Hence $a = [1 - e^{-\gamma} + O(1/n)]/[1 - O(1/n)]$ which approaches the value 0.439 as n becomes very large.

Hence, for $n \geq 20$, to two decimal places, $a = 0.44$, and (5) becomes

$$\hat{P}_i = (i - 0.44)/(n + 0.12) \qquad (7)$$

This last equation, then, is the recommended plotting formula for the ith of n ordered observations, if the n observations are plotted to facilitate their inspection. To illustrate the effectiveness of this plotting formula when the distribution is of type 1, even for sample sizes less than 20, Figure 1 has circles around the expected values of ten ordered observations, plotted by (7). For contrast, the crosses in Figure 1 show the positions plotted by (3), which has been favored in the geophysical literature for its simplicity. The latter points become separated from the ideal straight line, particularly at the largest values, hampering visual inspection of the data.

References

Blom, Gunnar, *Statistical Estimates and Transformed Beta-Variables,* John Wiley & Sons, New York, 1958.

Court, Arnold, Some new statistical techniques in geophysics, *Advan. Geophys., 1,* 45–85, 1952.

Gumbel, E. J., On the plotting of flood discharges, *Trans. Am. Geophys. Union,* part II, 699–716, 1943.

Gumbel, E. J., *Statistics of Extremes,* Columbia University Press, New York, 1958.

Kimball, Bradford F., On the choice of plotting positions on probability paper, *J. Am. Statist. Assoc., 55,* 546–560, 1960.

Lieblein, J., and H. E. Salzer, Table of the first moment of ranked extremes, *J. Res. Natl. Bur. Std., 59,* 203–206, 1957.

Powell, R. W., A simple method of estimating flood frequencies, *Civil Engrs., N. Y., 14,* 105, 1942.

(Manuscript received May 21, 1962; revised November 9, 1962.)