# Probability Estimates Based on Small Normal-Distribution Samples

LEO R. BEARD

*6133 Welty Way*
*Sacramento 24, California*

*Abstract.* Although it is impossible to establish a probability from sample data for a single enterprise so that the ratio of favorable to total future events in the enterprise approaches that probability as the number of events approaches infinity, it is possible to compute the expectation of the probability or *expected probability* from sample data for a single enterprise so that the ratio of favorable to total future events approaches that quantity as the number of such *enterprises* approaches infinity. Because of this property of the expected probability, it can serve as a basis for computing the expected return on an investment, with confidence that the actual return will approach the expected return as the number of such investments increases. Computation of expected probability $P_N$ based on a random sample from a normal parent population and its applicability to flood-control work are discussed.

*Introduction.* In flood studies, it has been demonstrated [*Corps of Engrs.*, 1955] that the logarithms of annual maximum flows at any location on a stream are ordinarily distributed in reasonable accord with the Gaussian normal distribution. In many locations, large expenditures are being contemplated to protect a community against a particular extreme magnitude of flow. The anticipated frequencies of flows in excess of various magnitudes are of primary concern in determining the anticipated future benefits of the project and therefore in establishing the degree of protection that can be afforded. These anticipated frequencies must often be based on stream-flow records less than 25 years in length. Thus, large expenditures can hinge on probability estimates derived from fitting a normal distribution to less than 25 events.

Although it is recognized that large uncertainties exist in the estimation of flood probabilities, the manner in which they are ordinarily used in flood-control work precludes the direct application of tolerance limits or safety factors. Annual flood-control benefits are computed by the use of mathematical expectation, that is, the sum of the cross products of estimated flood frequencies per year and corresponding damages per flood for all pertinent ranges of flood magnitude. If the frequencies are purposely overestimated in order to make sure that the structure will be safe, benefits will be overestimated and uneconomical projects will result. If they are purposely underestimated in order to be sure that computed benefits will be obtained, unsafe structures will result. If frequencies are underestimated for the purpose of computing benefits and overestimated for design purposes, many worth-while structures would be declared economically unsound. Thus, it is essential for ordinary flood-control purposes that the best possible estimate of flood probabilities be used.

To define what is wanted in the best possible estimate of a probability is not simple. If the true exceedance probability is known, it can be expected that the ratio of the number of future events in excess of the specified magnitude to the total number of events will approach that probability as events occur indefinitely. There would be considerable uncertainty, however, in what will happen in the next 50 or 100 years. On the other hand, if the true exceedance probability is not known and must be estimated, there is a second degree of uncertainty involved, and there apparently can be no exact statement regarding relative future frequencies as events occur indefinitely. In the first case, we can at least compute the expected return on an investment directly from known probabilities, but not in the second case. However, it is possible to estimate probabilities for each individual enterprise so that the estimated and true probabilities will average out properly over a large number of enterprises. Then a true probability statement can be made, but it would be contingent on the number of *enterprises* increasing toward infinity. The probability estimate having this quality is herein called the *expected probability*.

In its relation to flood-control work, this would mean that, although it is not possible to evaluate flood probabilities accurately for a single project, probabilities for individual projects can be evaluated so that the aggregate benefits attained at a number of independent projects will approach the expected aggregate benefits as the number of projects increases. Investment in a single project on the basis of expected return would bear some similarity to investment in a single spin of a roulette wheel. Just as the mathematical worth of such a spin before it occurs can be computed, the mathematical worth of a single flood-control investment could be computed.

*Expected exceedance probability* $P_N$. The expected exceedance probability $P_N$ of a given magnitude that can be expressed as a function of sample statistics (such as the sum of the sample mean and twice the sample standard deviation) can be defined as the average of the true exceedance probabilities of an infinite number of magnitudes that might be determined in the same manner from random samples of the same size ($N$) derived from the same parent population. For example, if an infinite number of 10-event samples were obtained from a normal parent population, the sample mean plus 2 times the sample standard deviation (which sum would represent a different magnitude in every sample) of some extreme samples would be exceeded by more than 25 per cent of the parent population, and that sum for some opposite-extreme samples would be exceeded by less than 0.1 per cent. However, the average exceedance probability of $M + 2S$ values based on an infinite number of random samples can be computed mathematically or determined experimentally, and this average is herein designated $P_N$.

If any magnitude having an expected exceedance probability $P_N$ is designated as $X'$, then, by definition, the average of the true exceedance probabilities of all $X'$ values will approach $P_N$. Since this is true regardless of the nature of the parent population, it follows that the $X'$ values can be obtained from different parent populations (and even by different methods) and the statement will remain true as long as all $X'$ values correspond to the same fixed value of $P_N$. This will be demonstrated by tests based on actual stream-flow data from different locations (different parent populations).

*Application to a normal distribution. Proschan* [1953] proved that the expected exceedance probability of a statistic based on the mean and standard deviation of a sample from a normal population is a simple function of Student's $t$-distribution. In terms of $P_N$, the relationship is as follows:

$$P_N = \text{Expected prob } [X > (M + kS)] = \text{Prob}\left[t_{N-1} > k\left(\frac{N}{N+1}\right)^{1/2}\right] \quad (1)$$

in which $X$ is an unknown future event, $k$ is any constant, and $t_{N-1}$ is Student's t-statistic with $N - 1$ degrees of freedom. The mean ($M$) and standard deviation ($S$) as used above are determined from sample data by use of the following equations:

$$M = \sum X/N \quad (2)$$

$$S^2 = \sum (X - M)^2/N - 1 \quad (3)$$

in which $X$ is a magnitude of a single event and $N$ is the number of events in the sample. Table 1 gives $P_N$ values for use with samples drawn from a normal population and is based on equation 1 and available tables of $t$.

In order to illustrate the applicability of Table 1, we selected 1200 events by use of random numbers from a normal population with a mean of zero and a standard deviation of 1. The events were grouped into samples of 10 each, and the mean and standard deviations for each sample were computed in accordance with equations 2 and 3. For each sample, various multiples of the standard deviations were added in turn to the mean, and the true exceedance probability of each sum was obtained from knowledge of the parent population. These probabilities from all samples and for a given number of standard deviations from the mean were averaged. The results are shown in Table 2. It is apparent that the average true probability observed in the tests is much closer to the theo-

TABLE 1. Distribution of Expected Probability $P_N$
For use with samples drawn from a normal population.

| $N-1$ | $P_N$ .500 | .450 | .400 | .350 | .300 | .250 | .200 | .150 | .125 | .100 | .050 | .025 | .0125 | .0100 | .005 | .0025 | .0005 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | .00 | .19 | .40 | .62 | .89 | 1.22 | 1.69 | 2.40 | 2.96 | 3.77 | 7.73 | 15.56 | 31.17 | 38.97 | 77.96 | 155.93 | 779.70 |
| 2 | .00 | .16 | .33 | .51 | .71 | .94 | 1.23 | 1.60 | 1.85 | 2.18 | 3.37 | 4.97 | 7.16 | 8.04 | 11.46 | 16.27 | 36.49 |
| 3 | .00 | .15 | .31 | .47 | .65 | .86 | 1.09 | 1.40 | 1.59 | 1.83 | 2.63 | 3.56 | 4.67 | 5.08 | 6.53 | 8.33 | 14.47 |
| 4 | .00 | .15 | .30 | .45 | .62 | .81 | 1.03 | 1.30 | 1.47 | 1.68 | 2.34 | 3.04 | 3.83 | 4.10 | 5.04 | 6.13 | 9.43 |
| 5 | .00 | .14 | .29 | .44 | .60 | .79 | .99 | 1.25 | 1.41 | 1.59 | 2.18 | 2.78 | 3.42 | 3.63 | 4.36 | 5.16 | 7.41 |
| 6 | .00 | .14 | .28 | .43 | .59 | .77 | .97 | 1.21 | 1.36 | 1.54 | 2.08 | 2.62 | 3.17 | 3.36 | 3.96 | 4.62 | 6.37 |
| 7 | .00 | .14 | .28 | .43 | .58 | .75 | .95 | 1.19 | 1.33 | 1.50 | 2.01 | 2.51 | 3.01 | 3.18 | 3.71 | 4.27 | 5.73 |
| 8 | .00 | .14 | .28 | .42 | .58 | .74 | .94 | 1.17 | 1.31 | 1.47 | 1.96 | 2.43 | 2.90 | 3.05 | 3.54 | 4.04 | 5.31 |
| 9 | .00 | .14 | .27 | .42 | .57 | .74 | .93 | 1.15 | 1.29 | 1.45 | 1.92 | 2.37 | 2.82 | 2.96 | 3.41 | 3.87 | 5.01 |
| 10 | .00 | .13 | .27 | .41 | .57 | .73 | .92 | 1.14 | 1.28 | 1.43 | 1.89 | 2.33 | 2.75 | 2.89 | 3.31 | 3.74 | 4.79 |
| 11 | .00 | .13 | .27 | .41 | .56 | .73 | .91 | 1.13 | 1.26 | 1.42 | 1.87 | 2.29 | 2.70 | 2.83 | 3.23 | 3.64 | 4.62 |
| 12 | .00 | .13 | .27 | .41 | .56 | .72 | .91 | 1.12 | 1.25 | 1.41 | 1.85 | 2.26 | 2.66 | 2.78 | 3.17 | 3.56 | 4.48 |
| 13 | .00 | .13 | .27 | .41 | .56 | .72 | .90 | 1.12 | 1.25 | 1.40 | 1.83 | 2.24 | 2.62 | 2.74 | 3.12 | 3.49 | 4.37 |
| 14 | .00 | .13 | .27 | .41 | .55 | .71 | .90 | 1.11 | 1.24 | 1.39 | 1.82 | 2.22 | 2.59 | 2.71 | 3.07 | 3.44 | 4.28 |
| 15 | .00 | .13 | .27 | .41 | .55 | .71 | .89 | 1.11 | 1.23 | 1.38 | 1.81 | 2.20 | 2.57 | 2.68 | 3.04 | 3.39 | 4.20 |
| 16 | .00 | .13 | .27 | .40 | .55 | .71 | .89 | 1.10 | 1.23 | 1.38 | 1.80 | 2.18 | 2.54 | 2.66 | 3.01 | 3.35 | 4.13 |
| 17 | .00 | .13 | .26 | .40 | .55 | .71 | .89 | 1.10 | 1.22 | 1.37 | 1.79 | 2.17 | 2.53 | 2.64 | 2.98 | 3.31 | 4.07 |
| 18 | .00 | .13 | .26 | .40 | .55 | .71 | .88 | 1.09 | 1.22 | 1.36 | 1.78 | 2.16 | 2.51 | 2.62 | 2.95 | 3.28 | 4.02 |
| 19 | .00 | .13 | .26 | .40 | .55 | .70 | .88 | 1.09 | 1.22 | 1.36 | 1.77 | 2.14 | 2.49 | 2.60 | 2.93 | 3.25 | 3.98 |
| 20 | .00 | .13 | .26 | .40 | .55 | .70 | .88 | 1.09 | 1.21 | 1.36 | 1.77 | 2.14 | 2.48 | 2.59 | 2.91 | 3.23 | 3.94 |
| 21 | .00 | .13 | .26 | .40 | .54 | .70 | .88 | 1.09 | 1.21 | 1.35 | 1.76 | 2.13 | 2.47 | 2.57 | 2.89 | 3.21 | 3.90 |
| 22 | .00 | .13 | .26 | .40 | .54 | .70 | .88 | 1.08 | 1.21 | 1.35 | 1.75 | 2.12 | 2.46 | 2.56 | 2.88 | 3.19 | 3.87 |
| 23 | .00 | .13 | .26 | .40 | .54 | .70 | .88 | 1.08 | 1.20 | 1.35 | 1.75 | 2.11 | 2.45 | 2.55 | 2.86 | 3.17 | 3.84 |
| 24 | .00 | .13 | .26 | .40 | .54 | .70 | .87 | 1.08 | 1.20 | 1.34 | 1.74 | 2.10 | 2.44 | 2.54 | 2.85 | 3.15 | 3.82 |
| 25 | .00 | .13 | .26 | .40 | .54 | .70 | .87 | 1.08 | 1.20 | 1.34 | 1.74 | 2.10 | 2.43 | 2.53 | 2.84 | 3.14 | 3.80 |
| 26 | .00 | .13 | .26 | .40 | .54 | .70 | .87 | 1.08 | 1.20 | 1.34 | 1.74 | 2.09 | 2.42 | 2.52 | 2.83 | 3.12 | 3.78 |
| 27 | .00 | .13 | .26 | .40 | .54 | .70 | .87 | 1.08 | 1.20 | 1.34 | 1.73 | 2.09 | 2.42 | 2.52 | 2.82 | 3.11 | 3.76 |
| 28 | .00 | .13 | .26 | .40 | .54 | .69 | .87 | 1.07 | 1.20 | 1.34 | 1.73 | 2.08 | 2.41 | 2.51 | 2.81 | 3.10 | 3.74 |
| 29 | .00 | .13 | .26 | .40 | .54 | .69 | .87 | 1.07 | 1.19 | 1.33 | 1.73 | 2.08 | 2.40 | 2.50 | 2.80 | 3.09 | 3.72 |
| 30 | .00 | .13 | .26 | .40 | .54 | .69 | .87 | 1.07 | 1.19 | 1.33 | 1.72 | 2.07 | 2.40 | 2.50 | 2.79 | 3.08 | 3.70 |
| 40 | .00 | .13 | .26 | .39 | .54 | .69 | .86 | 1.06 | 1.18 | 1.32 | 1.70 | 2.05 | 2.36 | 2.45 | 2.74 | 3.01 | 3.59 |
| 60 | .00 | .13 | .26 | .39 | .53 | .68 | .85 | 1.05 | 1.17 | 1.31 | 1.68 | 2.02 | 2.32 | 2.41 | 2.68 | 2.94 | 3.49 |
| 120 | .00 | .13 | .26 | .39 | .53 | .68 | .85 | 1.05 | 1.16 | 1.29 | 1.66 | 1.99 | 2.28 | 2.37 | 2.63 | 2.87 | 3.39 |
| ∞ | .00 | .13 | .25 | .38 | .52 | .67 | .84 | 1.04 | 1.15 | 1.28 | 1.64 | 1.96 | 2.24 | 2.33 | 2.58 | 2.81 | 3.29 |

TABLE 2. Average Probabilities

| Statistic | Average True Probability | Expected Probability | Normal-Curve Area |
|---|---|---|---|
| | Exceedance Probability | | |
| $M + 4S$ | .0039 | .0021 | .000032 |
| $M + 3S$ | .0131 | .0094 | .0014 |
| $M + 2S$ | .052 | .045 | .023 |
| $M + S$ | .195 | .184 | .159 |
| $M$ | .510 | .500 | .500 |
| | Nonexceedance Probability | | |
| $M - S$ | .177 | .184 | .159 |
| $M - 2S$ | .041 | .045 | .023 |
| $M - 3S$ | .0080 | .0094 | .0014 |
| $M - 4S$ | .0022 | .0021 | .000032 |

retical expected probability (from Table 1) than to the probability obtained in the usual manner from a table of normal-curve areas.

*Applicability to hydrologic frequencies.* For the purpose of testing the applicability of Table 1 to samples drawn from various different normal parent populations, 70 long records of stream flow at stations throughout the United States were selected for study. As noted above, the logarithms of annual maximum flows are ordinarily distributed in reasonable accord with the normal distribution. In this test, the logarithms of annual maximum daily flows at each station were divided into three groups, group 1 consisting of every third year beginning with the first, group 2 consisting of every third year

**This table shows k for values of $P_N$ as illustrated:**

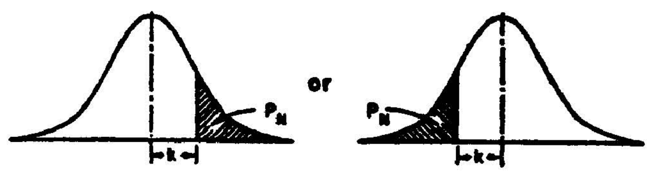


TABLE 3. Frequency Test—Annual Maximum Daily Runoff Using 210 Samples with Average Length of 25 Years

| $k$ | Independent Data: Observed No. | Independent Data: Observed Ratio | Expected Probability* | Data within Samples: Observed No. | Data within Samples: Observed Ratio | Normal-Curve Area |
|---|---|---|---|---|---|---|
| | | | Exceedance Frequency | | | |
| 2.7 | 19 | .0073 | .0070 | 11 | .0021 | .0035 |
| 2.4 | 30 | .0115 | .0137 | 31 | .0060 | .0082 |
| 2.1 | 62 | .0238 | .0253 | 83 | .0160 | .0179 |
| 1.8 | 99 | .038 | .045 | 176 | .034 | .036 |
| 1.5 | 165 | .063 | .077 | 322 | .062 | .067 |
| 1.2 | 295 | .113 | .125 | 581 | .112 | .115 |
| 0.9 | 498 | .192 | .194 | 956 | .184 | .184 |
| 0.6 | 752 | .289 | .281 | 1452 | .279 | .274 |
| 0.3 | 1036 | .398 | .386 | 2056 | .395 | .382 |
| 0.0 | 1329 | .511 | .500 | 2657 | .511 | .500 |
| | | | Nonexceedance Frequency | | | |
| 0.0 | 1271 | .489 | .500 | 2543 | .489 | .500 |
| −0.3 | 993 | .382 | .386 | 1974 | .380 | .382 |
| −0.6 | 711 | .273 | .281 | 1391 | .267 | .274 |
| −0.9 | 489 | .188 | .194 | 924 | .178 | .184 |
| −1.2 | 311 | .120 | .125 | 582 | .112 | .115 |
| −1.5 | 204 | .078 | .077 | 358 | .069 | .067 |
| −1.8 | 139 | .053 | .045 | 185 | .036 | .036 |
| −2.1 | 72 | .0277 | .0253 | 87 | .0167 | .0179 |
| −2.4 | 45 | .0173 | .0137 | 45 | .0087 | .0082 |
| −2.7 | 30 | .0115 | .0070 | 14 | .0027 | .0035 |

* Interpolated approximately.

**TABLE 4. Frequency Test—Annual Maximum Daily Runoff Using 210 Samples with Average Length of 12 Years**

| | Independent Data | | | Data within Samples | | |
|---|---|---|---|---|---|---|
| k | Observed No. | Observed Ratio | Expected Probability* | Observed No. | Observed Ratio | Normal-Curve Area |
| | | | Exceedance Frequency | | | |
| 2.7 | 52 | .0100 | .0126 | 3 | .0012 | .0035 |
| 2.4 | 86 | .0165 | .0209 | 7 | .0027 | .0082 |
| 2.1 | 150 | .0288 | .0342 | 29 | .0112 | .0179 |
| 1.8 | 254 | .049 | .056 | 65 | .025 | .036 |
| 1.5 | 402 | .077 | .088 | 155 | .060 | .067 |
| 1.2 | 666 | .128 | .137 | 282 | .108 | .115 |
| 0.9 | 1031 | .198 | .204 | 477 | .183 | .184 |
| 0.6 | 1526 | .293 | .289 | 723 | .278 | .274 |
| 0.3 | 2075 | .399 | .390 | 1015 | .390 | .382 |
| 0.0 | 2648 | .509 | .500 | 1309 | .503 | .500 |
| | | | Nonexceedance Frequency | | | |
| 0.0 | 2552 | .491 | .500 | 1291 | .497 | .500 |
| −0.3 | 2013 | .387 | .390 | 997 | .383 | .382 |
| −0.6 | 1493 | .287 | .289 | 722 | .278 | .274 |
| −0.9 | 1040 | .200 | .204 | 485 | .187 | .184 |
| −1.2 | 712 | .137 | .137 | 298 | .115 | .115 |
| −1.5 | 492 | .095 | .088 | 163 | .063 | .067 |
| −1.8 | 324 | .062 | .056 | 80 | .031 | .036 |
| −2.1 | 205 | .0394 | .0342 | 32 | .0123 | .0179 |
| −2.4 | 134 | .0258 | .0209 | 3 | .0012 | .0082 |
| −2.7 | 96 | .0185 | .0126 | 1 | .0004 | .0035 |

* Interpolated approximately.

beginning with the second, and group 3 consisting of every third year beginning with the third. Then a normal-distribution curve was fitted to two-thirds of each record, and the indicated frequencies were compared with (*a*) frequencies of the magnitudes upon which the curve was based (events within samples) and (*b*) frequencies of the magnitudes in the remaining third of the record (independent events). Portions of the record were then interchanged, and this procedure was carried out three times for each station. Observed frequencies were combined for all 70 stations, and the results are shown in Table 3. It will be noted that, whereas the values within the samples are distributed very nearly normally[1] on the average, the independent values observed under the same conditions are distributed very nearly in accord with $P_N$ values.

The test described above was repeated by fitting a normal distribution curve to each third of each record and comparing frequencies with the remaining two-thirds. The results, shown in Table 4, agree with those in Table 3.

A similar test was based on the logarithms of annual maximum 60-minute precipitation measured at 121 weather stations throughout the United States. The expected probabilities of statistics computed from the last 25 odd-numbered years were compared with corresponding exceedance frequencies observed in the last 25 even-numbered-years, and vice versa. Results, shown in Table 5, are similar to those of Tables 3 and 4, all of which appear to support the expected probability theory satisfactorily.

*Conclusion.* When the mathematical expectation of chance events must be computed from

[1] There is a small systematic difference due to the adjustment of the standard deviation to the universe, that is, dividing by $N - 1$ instead of $N$ in equation 3.

TABLE 5. Frequency Test—Annual Maximum 60-Minute Precipitation Using 242 Samples with Length of 25 Years

| k | Independent Data | | | Data within Samples | | |
|---|---|---|---|---|---|---|
| | Observed | | Expected Probability* | Observed | | Normal-Curve Area |
| | No. | Ratio | | No. | Ratio | |
| | | | Exceedance Frequency | | | |
| 2.7 | 59 | .0098 | .0070 | 18 | .0030 | .0035 |
| 2.4 | 109 | .0180 | .0137 | 45 | .0074 | .0082 |
| 2.1 | 175 | .0289 | .0253 | 119 | .0197 | .0179 |
| 1.8 | 304 | .050 | .045 | 225 | .037 | .036 |
| 1.5 | 478 | .079 | .077 | 433 | .072 | .067 |
| 1.2 | 773 | .128 | .125 | 698 | .115 | .115 |
| 0.9 | 1137 | .188 | .194 | 1120 | .185 | .184 |
| 0.6 | 1637 | .271 | .281 | 1608 | .266 | .274 |
| 0.3 | 2226 | .368 | .386 | 2205 | .364 | .382 |
| 0.0 | 2941 | .486 | .500 | 2911 | .481 | .500 |
| | | | Nonexceedance Frequency | | | |
| 0.0 | 3109 | .514 | .500 | 3139 | .519 | .500 |
| −.03 | 2368 | .391 | .386 | 2336 | .386 | .382 |
| −0.6 | 1695 | .280 | .281 | 1690 | .279 | .274 |
| −0.9 | 1124 | .186 | .194 | 1097 | .181 | .184 |
| −1.2 | 709 | .117 | .125 | 649 | .107 | .115 |
| −1.5 | 417 | .069 | .077 | 342 | .057 | .067 |
| −1.8 | 239 | .040 | .045 | 163 | .027 | .036 |
| −2.1 | 136 | .0225 | .0253 | 75 | .0124 | .0179 |
| −2.4 | 58 | .0096 | .0137 | 30 | .0050 | .0082 |
| −2.7 | 26 | .0043 | .0070 | 8 | .0013 | .0035 |

* Interpolated approximately.

random-sample data, as in estimating future flood damages, it is believed that probabilities should be estimated by use of the *expected probability* concept and that the associated expectation is an exact mathematical quantity subject only to assumptions as to the form of the parent population and the randomness of events.

*Acknowledgment.* This paper was developed in the U. S. Army Engineer District, Sacramento, California, under the Civil Works Investigations program of the Corps of Engineers. Runoff data used for test purposes were assembled in the Washington District of the Corps of Engineers and published by the U. S. Geological Survey. Precipitation data used for test purposes were assembled by the U. S. Weather Bureau, Hydrologic Services Division, in Washington, D. C.

## References

Corps of Engineers, Washington District, *Stream Flow Volume-Duration-Frequency Studies*, June 1955.

Proschan, F., Confidence and tolerance intervals for the normal distribution, *J. Am. Statist. Assoc.*, Sept. 1953.

(Manuscript received August 31, 1959; revised March 9, 1960.)