Appendix 11

EXPECTED PROBABILITY

The principle of gambling based upon estimated probabilities can be applied to water resources development decisions. However, because probabilities must be inferred from random sample data, they are uncertain and mathematical expectation cannot be computed exactly as errors due to uncertainty do not necessarily compensate. For example, if the estimate based on sample data is that a certain flood magnitude will be exceeded on the average once in 100 years, it is possible that the true exceedance could be three or four more times per hundred years, but it can never be less than zero times per hundred years. The impact of errors in one direction due to uncertainty can be quite different from the impact of errors in the other direction. Thus, it is not adequate to simply be too high half the time and too low the other half. It is necessary to consider the relative impacts of being too high or too low.

It is possible to delineate uncertainty with considerable accuracy when dealing with samples from a normal distribution. Therefore, when flood flow frequency curves conform fairly closely to the logarithmic normal distribution, it is possible to delineate uncertainty of frequency or probability estimates of flood flows.

Figure 11-1 is a generalized representation of the range of uncertainty in probability estimates based on samples drawn from a normal population. The vertical scale can represent the logarithm of streamflow. The curves show the likelihood that the true frequency of any flood magnitude exceeds the value shown on the frequency scale. The curve labeled .50 is the curve that would be used for the best frequency estimate of a log-normal population. From this curve a magnitude of 2 would be exceeded on the average 30 times per thousand events. The figure also shows a 5 percent chance that the true frequency is 150 or more times per thousand or a 5 percent chance that the true frequency is two times or less per thousand events.

If a magnitude of 2.0 were selected at 20 independent locations, the best estimate for the frequency is 3 exceedances per hundred years for each location. The estimated total exceedance for all 20 locations would be 60 per 100 years. However, due to sampling uncertainties, true frequencies for a magnitude of 2.0 would differ at each location and total exceedances per 100 years at the 20 locations might be represented by the following tabulation.

Exceedances Per 100 Years at Each of 20 Locations*

20	5	3	.9	
12	5	2	.8	
10	4	2	.5	Total Exceedances = Approximately 90
8	4	2	.3	
7	3	1	.1	

*Determined from Figure 11-1 using 0.05 parameter value increments from .025 through .975.

The total of these exceedances is about 90 per 100 years or 30 more than obtained using the best probability estimate as the true probability at each location. If, however, the mathematically derived expected probability function were used instead of the traditional "best" estimate we could read the expected probability curve of Figure 11-1 to obtain the value of about 4.5 exceedances per 100 events. This value when applied to each of the 20 locations would give an estimate of 90 exceedances per 100 years at all 20 locations. Thus, while the expected probability estimate would be wrong in the high direction more frequently than in the low direction, the heavier impacts of being wrong in the low direction would compensate for this. It can be noted, at this point, that expected probability is the average of all estimated true probabilities.

If a flood frequency estimate could be accurately known--that is, the parent population could be defined--the frequency distribution of observed flood events would approach the parent population as the number of observations approaches infinity. This is not the case where probabilities are not accurately known. However, if the expected probabilities as illustrated in Figure 11-1 can be computed, observed flood frequency for a large number of independent locations will approach the estimated flood frequency as the number of observations approaches infinity and the number of locations approaches infinity.

It appears that the answer to the question as to whether expected probability should be used at a single location would be identical to the answer to the question, "What is a fair wager for a single gamble?" If the gamble must be undertaken, and ordinarily it must, then the answer to the above question is that the wager should be proportional to the expected return. In determining whether the expected probability concepts should apply for a single location, the same line of reasoning would indicate that it should.

It has been shown (21) that for the normal distribution the expected probability P_N can be obtained from the formula

$$P_N = \text{Prob}\left[t_{N-1} > K_n \left(\frac{N}{N+1}\right)^{1/2}\right] \qquad (11\text{-}1)$$

where K_n is the standard normal variate of the desired probability of exceedance, N is the sample size, and t_{N-1} is the Student's t-statistic with N-1 degrees of freedom.

The actual calculations can be carried out using tables of the t-statistic, or the modified values shown in Table 11-1 (31). To use Table 11-1, enter with the sample size minus 1 and read across to the column with the desired exceedance probability. The value read from the table is the corrected plotting position.

The expected probability correction may also be calculated from the following equations (34) which are based on Table 11-1. For selected exceedance probabilities greater than 0.50, and a given sample size, the appropriate P_N value equals 1 minus the value in Table 11-1 or the equations 11-2.

Exceedance Probability	Expected Probability, P_N	
.0001	$.0001\ (1.0 + 1600/N^{1.72})$	(11-2a)
.001	$.001\ (1.0 + 280/N^{1.55})$	(11-2b)
.01	$.01\ (1.0 + 26/N^{1.16})$	(11-2c)
.05	$.05\ (1.0 + 6/N^{1.04})$	(11-2d)
.10	$.1\ (1.0 + 3/N^{1.04})$	(11-2e)
.30	$.3\ (1.0 + 0.46/N^{0.925})$	(11-2f)

For floods with an exceedance probability of 0.01 based on samples of 20 annual peaks, for example, the expected probability of exceedance from equation 11-2c is (.01) (1.0 + 26/32.3) or 0.018. Use of Table 11-1 gives 0.0174. Comparable equations for adjusting the computed discharge upward to give a discharge for which the expected probability equals the exceedance probability are available (22).

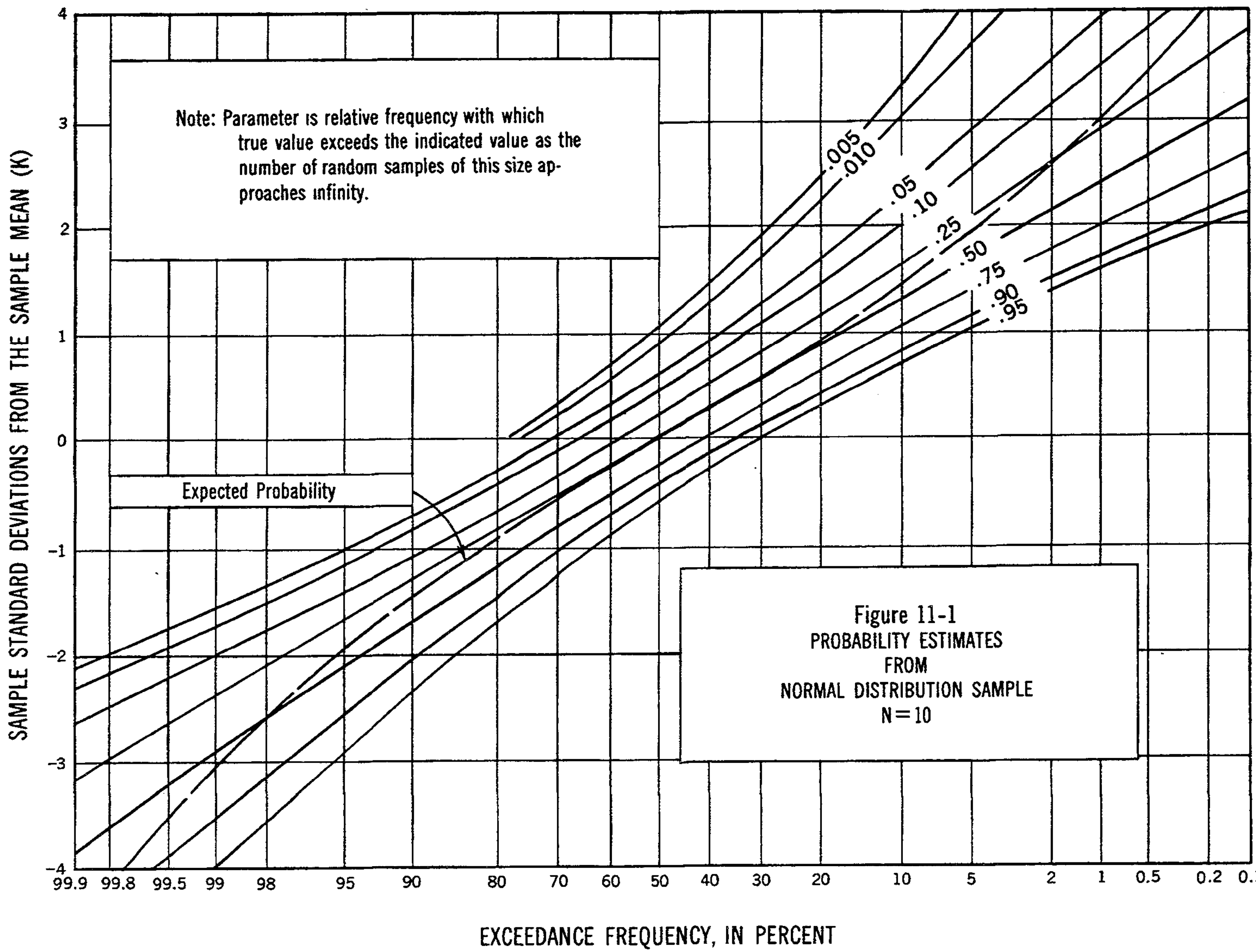

Figure 11-1
PROBABILITY ESTIMATES
FROM
NORMAL DISTRIBUTION SAMPLE
N=10

Table 11-1

TABLE OF P_N VERSUS P_∞

For use with samples drawn from a normal population

N-1 \ P_∞	.50	.30	.10	.05	.01	.001	.0001
1	.500	.372	.243	.204	.154	.121	.102
2	.500	.347	.193	.146	.090	.057	.043
3	.500	.336	.169	.119	.064	.035	.023
4	.500	.330	.154	.104	.050	.024	.0137
5	.500	.325	.146	.094	.042	.0179	.0092
6	.500	.322	.138	.088	.036	.0138	.0066
7	.500	.319	.135	.083	.032	.0113	.0050
8	.500	.317	.131	.079	.029	.0094	.0039
9	.500	.316	.127	.076	.027	.0082	.0031
10	.500	.315	.125	.073	.025	.0072	.0025
11	.500	.314	.123	.071	.023	.0064	.0021
12	.500	.313	.121	.069	.022	.0058	.0018
13	.500	.312	.119	.068	.021	.0052	.0016
14	.500	.311	.118	.067	.020	.0048	.0014
15	.500	.311	.117	.066	.0196	.0045	.0013
16	.500	.310	.116	.065	.0190	.0042	.0012
17	.500	.310	.115	.064	.0184	.0040	.0011
18	.500	.309	.114	.063	.0179	.0038	.0010
19	.500	.309	.113	.062	.0174	.0036	.00091
20	.500	.308	.113	.062	.0170	.0034	.00084
21	.500	.308	.112	.061	.0167	.0033	.00078
22	.500	.308	.111	.061	.0163	.0031	.00073
23	.500	.307	.111	.060	.0161	.0030	.00068
24	.500	.307	.110	.060	.0158	.0029	.00064
25	.500	.307	.110	.059	.0155	.0028	.00060
26	.500	.306	.109	.059	.0153	.0027	.00057
27	.500	.306	.109	.059	.0151	.0026	.00054
28	.500	.306	.109	.058	.0149	.0026	.00051
29	.500	.306	.108	.058	.0147	.0025	.00049
30	.500	.306	.108	.058	.0145	.0024	.00046
40	.500	.304	.106	.056	.0133	.0020	.00034
60	.500	.303	.104	.054	.0122	.0016	.00025
120	.500	.302	.102	.052	.0111	.0013	.00017
∞	.500	.300	.100	.050	.0100	.0010	.00010

NOTE: P_N values above are usable approximately with Pearson Type III distributions having small skew coefficients.