Frequently Asked Questions on Flood Statistics

1. What is 100-yr flood?
   Answer: 100-yr flood has a probability of 1/100 = .01, or 1%, of being **equaled or exceeded** in any single year.

2. What is Return Period?
   Answer: An annual maximum event has a return period (or recurrence interval) of T years if its magnitude is equaled or exceeded once, on the average, every T years. The reciprocal of T is the exceedance probability of that event, that is, the probability that the event is equaled or exceeded in any one year.

3. What is the probability that at least one 100-yr flood will occur during the 50-yr lifetime of a flood control project? Ans: 0.395. So there is a 39.5 % chance of occurrence of the 100-yr or larger event over the 50-yr lifetime of the project. This is just risk of failure. Please see 6 also.
   $$\text{Risk} = 1- (1-1/T)^{n} \quad = 1 - (1-1/100)^{50} \quad = 1 - .6050 \quad = 0.395$$

4. What is the probability that 100-yr flood or greater will not occur in 50-yr?
   Answer: Probability of no occurrence
   $$= (1-p)^{n} = (1-1/T)^{n} = (1 - 1/100)^{50} = .6050$$
   Thus, there is a 60.5% chance (60.5% reliability) that the 100-yr or larger flood will not occur during a sequence of 50 –yr.

5. (a) On average, how many times will a 10-yr flood occur in 50-yr period? Ans: 5
   (b) What is the probability that exactly 5 and only 5 10-yr or greater independent floods will occur in a 50-yr period? Ans: 0.185 or 18.5%
   (c) What is the probability that exactly 4 and only 4 10-yr or greater independent floods will occur in a 50-yr period? Ans: 0.181 or 18.1%
   (d) What is the probability that exactly 6 and only 6 10-yr or greater independent floods will occur in a 50-yr period? Ans: 0.154 or 15.4%

6. What is Risk?
   Answer: Over a sequence of n yr, the probability that the T-yr event or larger will occur at least once is called the **risk**. Thus risk is the sum of the probabilities of 1 T-yr or larger flood, 2 T-yr or larger floods, 3 T-yr or larger floods, …., n floods occurring during n-year period, but it is easier to calculate 1- prob(0 T-yr floods):
   Risk or probability of at least one occurrence
   $$= 1 - P(0) \quad = 1- \text{Prob (no occurrence in n yr)} \quad = 1- (1-p)^{n}$$
   $$= 1- (1-1/T)^{n} \quad \text{where } T = \text{return period}$$

Table 1.0 shows the return periods for various degrees of risk and expected design life. For example, for 98.4-yr event (return period) and 50-yr expected design life, the risk or the probability of occurrence of a98.4-yr event or larger in 50-yr design life is 40% and the reliability or the probability of no occurrence of 98.4 yr event in 50-yr design life is 60%.

Question 5 Equations
Answer:
(a) A 10-yr flood has p = 1/10 = 0.1
Expected probability = n*p = 50*.1 = 5
Thus, on average a 10-yr flood will occur 5 times in a 50-yr period.

(b) Binomial distribution: fx = (x: n, p) = $\binom{n}{x}$ p^x q^{n-x} where q = 1-p

$$= \binom{50}{5} (0.1)^5 (0.9)^{50-5}$$

$$= (50!/(5!45!))(0.1)^5 (0.9)^{50-5}$$
$$= 0.1849$$

Thus, there is 18% chance that 5 and only 5 number 10-yr events in 50-yr periods will occur.

(c) Probability of 4 and only 4 number 10-yr flood in a 50-yr period

$$= \binom{50}{4} (0.1)^4 (0.9)^{50-4}$$
$$= (50!/(4!46!))(0.1)^4 (0.9)^{46}$$

$$= .1809$$

Thus, there is 18% chance that 4 and only 4 number10-yr events in 50-yr periods will occur.

(d) Probability of 6 and only 6 number 10-yr flood in a 50-yr period

$$= \binom{50}{6} (0.1)^6 (0.9)^{50-6}$$
$$= (50!/(6!44!))(0.1)^6 (0.9)^{44}$$

$$= .1541$$

Thus, there is 15% chance that 6 and only 6 number 10-yr events in 50-yr periods will occur.

7. What is Reliability?
Answer: Reliability may be defined as 1- risk. Thus,

$$\text{Reliability} = (1-p)^n$$
$$= (1-1/T)^n$$

References:

1. Frequency Analysis
2. Statistical Method in Hydrology – Charles T. Haan

Table 1.0

Return Periods for Various Degrees of Risk and Expected Design Life

		Expected Design Life							
RISK	RELIABILITY	2	5	10	15	20	25	50	100
75	25	2	4.1	7.7	11.3	14.9	18.5	36.6	72.6
63	37	2.6	5.5	10.6	15.6	20.6	25.6	50.8	101.1
50	50	3.4	7.7	14.9	22.1	29.4	36.6	72.6	144.8
40	60	4.4	10.3	20.1	29.9	39.7	49.4	98.4	196.3
30	70	6.1	14.5	28.5	42.6	56.6	70.6	140.7	280.9
25	75	7.5	17.9	35.3	52.6	70	87.4	174.3	348.1
20	80	9.5	22.9	45.3	67.7	90.1	112.5	224.6	448.6
15	85	12.8	31.3	62	92.8	123.6	154.3	308.2	615.8
10	90	19.5	48	95.4	142.9	190.3	237.8	475.1	949.6
5	95	39.5	98	195.5	292.9	390.4	487.9	975.3	1950.1
2	98	99.5	248	495.5	743	990.5	1238	2475.4	49503
1	99	199.5	498	995.5	1493	1990.5	2488	4975.5	9950.4
0.5	99.5	399.5	998	1995.5	2993	3990.5	4988	9975.5	19950.5

Life of Project

Return Period of Design Storm

Risk that design storm or larger will occur during project life

$Prob(2 aces) = f_X(2; 52, 5, 4) = \binom{4}{2}\binom{48}{3} / \binom{52}{5} = 0.040$

Acceptance Sampling Problem

$Prob(5 def) = f_X(5; 50, 20, 12) = \binom{12}{5}\binom{38}{15} / \binom{50}{20} = 0.26$

BERNOULLI PROCESSES

Binomial Distribution

Consider a discrete time scale. At each point on this time scale an event may either occur or not occur. Let the probability of the event occurring be p for every point on the time scale. Thus the occurrence of the event at any point on the time scale is independent of the history of any prior occurrences or nonoccurrences. The probability of an occurrence at the i^{th} point on the time scale is p for i = 1, 2, ... A process having these properties is said to be a Bernoulli process.

As an example of a Bernoulli process consider that during any year the probability of the maximum flow exceeding 10,000 cfs on a particular stream is p. Common terminology for a flow exceeding a given value is an exceedance. Further consider that the peak flow in any year is independent from year to year (a necessary condition for the process to be a Bernoulli process). Let q = 1-p be the probability of not exceeding 10,000 cfs. We can neglect the probability of a peak of exactly 10,000 cfs since the peak flow rates would be a continuous process so the probability of a peak of exactly 10,000 cfs would be zero. In this example the time scale is discrete with the points being nominally 1 year in time apart. We can now make certain probabilistic statements about the occurrence of a peak flow in excess of 10,000 cfs (an exceedance).

For example the probability of an exceedance occurring in year 3 and not in years 1 or 2 can be evaluated from equation 2.9 as qqp since the process is independent from year to year. The probability of (exactly) one exceedance in any 3-year period is pqq + qpq + qqp since the exceedance could occur in either the first, second or third year. Thus the probability of (exactly) one exceedance in three years is $3pq^2$.

In a similar manner the probability of 2 exceedances in 5 years can be found from the summation of the terms ppqqq, pqpqq, pqqpq, ..., qqqpp. It can be seen that each of these terms is equivalent to p^2q^3 and that the number of terms is equal to the number of ways of arranging 2 items (the p's) among 5 items (the p's and q's). Therefore the total number of terms is $\binom{5}{2}$ or 10 so that the probability of exactly 2 exceedances in 5 years is $10\, p^2q^3$.

This result can be generalized so that the probability of X=x exceedances in n years is $\binom{n}{x} p^x q^{n-x}$. The result is applicable to any Bernoulli process so that the probability of X=x occurrences of an event in n independent trials if p is the probability of an occurrence in a single trial is given by

$$f_X(x; n, p) = \binom{n}{x} p^x q^{n-x} \qquad x = 0, 1, 2, ..., n \qquad (4.5)$$

Equation 4.5 is known as the binomial distribution.

The binomial distribution and the Bernoulli process are not limited to a time scale. Any process that may occur with probability p at discrete points in time or space or in individual trials may be a Bernoulli process and follow the binomial distribution.

The cumulative binomial distribution is

$$F_X(x; n, p) = \sum_{i=0}^{x} \binom{n}{i} p^i q^{n-i} \qquad x = 0, 1, 2, ..., n \qquad (4.6)$$

and gives the probability of x or fewer occurrences of an event in n independent trials if the probability of an occurrence in any trial is p.

Continuing the above example, the probability of less than 3 exceedances in 5 years is

$$F_X(2; 5, p) = \sum_{i=0}^{2} \binom{5}{i} p^i q^{5-i}$$

$$= f_X(0; 5, p) + f_X(1; 5, p) + f_X(2; 5, p)$$

The mean and variance of the binomial distribution are

$$E(X) = np \qquad (4.7)$$

$$Var(X) = npq \qquad (4.8)$$

The coefficient of skew is $(q-p)/\sqrt{npq}$ so that the distribution is symmetrical for p = q, skewed to the right for q>p and skewed to the left for q<p.

Example 4.4. On the average, how many times will a 10-year flood occur in a 40-year period? What is the probability that exactly this number of 10-year floods will occur in a 40-year period?

$f_X(1, 100, .01)$
$= \binom{100}{1}(.01)^1(.99)^{99}$
$= 100(.01)(.99)^{99}$
$\frac{100!}{1!\,99!}$
$= 37\%$

Solution: A 10-year flood has p = 1/10 = 0.1

$$E(X) = np = 40(0.1) = 4$$

$$f_X(4; 40, 0.1) = \binom{40}{4} (0.1)^4 (0.9)^{36} = 0.2059$$

Comment: This problem illustrates the difficulty of explaining the concept of return period to laymen. We have said that on the average a 10-year event occurs once every 10 years and that in a 40-year period we expect it to occur 4 times. Yet we have also shown that in about 80% (100(1-0.2059)) of all possible independent 40-year periods the 10-year event will not occur exactly 4 times. As a matter of fact the probability that it will occur 3 times is nearly identical to the probability it will occur 4 times (0.2003 vs. 0.2059). The number of occurrences, X, is truly a random variable (with a binomial distribution).

The individual and cumulative terms of the binomial distribution are tabled in many references (see for instance Beyer (1968) or Selby (1970)). The highest value of p given in most tables is 0.5. For values of p in excess of 0.5 the roles of p and q and x and n-x can be reversed since $f_X(x; n, p) = f_X(n-x; n, q)$.

The binomial distribution has an additive property (Gibra 1973). That is if X has a binomial distribution with parameters n_1 and p and Y has a binomial distribution with parameters n_2 and p, then Z=X+Y has a binomial distribution with parameters $n=n_1 + n_2$ and p.

The binomial distribution can be used to approximate the hypergeometric distribution if the sample selected is small in comparison to the number of items N from

Source :- Statistical Method in Hydrology p 70-71 - Charles T. Haan.

purpose :- FAQ Statistical Questions.

Source: - Frequency Analysis - Training book from Mark.

Fitting a Distribution to Data

An intuitive use for moment estimates is to fit probability distributions to data by equating the moment estimates obtained from the data to the functional form for the distribution. For example, the normal distribution has parameters mean μ and variance σ^2. The **method of moments** fit for the normal distribution is simply to use the estimates from the data for the mean and variance. As another example, the parameter λ of the exponential distribution is equal to the reciprocal of the mean of the distribution. Hence, the method of moments fit is $\hat{\lambda} = 1/\hat{\mu}$.

Graphical methods and the method of maximum likelihood are two alternative procedures to the method of moments for fitting distributions to data. Graphical methods will be discussed in Section 3.6. Although the method of maximum likelihood is superior to the method of moments by some statistical measures, it is generally much more computationally complex than the method of moments and is beyond the scope of this text. Maximum likelihood methods for several distributions used in hydrology are presented by Kite (1977).

The main objective of determining the parameters of a distribution is usually to evaluate its CDF. In some cases, however, the same purpose may be accomplished without calculating the actual distribution parameters. Instead, the distribution is evaluated using frequency factors K, defined as

$$K = \frac{x - \bar{x}}{S_x}. \tag{3.43}$$

The value of K is a function of the desired value for the CDF and may also be a function of the skewness. Hence, if K is known for the calculated skewness and desired CDF value, the corresponding value of x can be calculated. Frequency factors will be illustrated subsequently for several of the distributions to be discussed.

3.4 RETURN PERIOD OR RECURRENCE INTERVAL

The most common means used in hydrology to indicate the probability of an event is to assign a **return period** or **recurrence interval** to the event. The return period is defined as follows:

> An annual maximum event has a return period (or recurrence interval) of T years if its magnitude is equaled or exceeded once, *on the average*, every T years. The reciprocal of T is the exceedance probability of the event, that is, the probability that the event is equaled or exceeded in any one year.

purpose: Backup for FAQ statistical question.

Thus, the 50 yr flood has a probability of 0.02, or 2%, of being equaled or exceeded in any single year. It is imperative to realize that the return period implies nothing about the actual time sequence of an event. The 50-yr flood does not occur like clockwork once every 50 yr. Rather, one expects that on the average, about twenty 50-yr floods will be experienced during a 1000-yr period. There could in fact be two 50-yr floods in a row (with probability $0.02 \times 0.02 = 0.0004$ for independent events).

The concept of a return period implies independent events and is usually found by analyzing the series of maximum annual floods (or rainfalls, etc.). The largest event in one year is assumed to be independent of the largest event in any other year. But it is also possible to apply such an analysis to the n largest independent events from an n-yr period, regardless of the year in which they occur. In this case, if the second largest event in one year was greater than the largest event in another year, it could be included in the frequency analysis. This series of n largest (independent) values is called the series of **annual exceedances**, as opposed to an annual maximum series. Both series are used in hydrology, with little difference at high return periods (rare events). There are likely to be more problems of ensuring independence when using annual exceedances, but for low return periods annual exceedances give a more realistic lower return period for the same magnitude than do annual maxima. The relationship between return period based on annual exceedances T_e and annual maxima T_m is (Chow, 1964)

$$T_e = \frac{1}{\ln T_m - \ln(T_m - 1)}. \tag{3.44}$$

This relationship is shown in Fig. 3.10.

EXAMPLE 3.4

ANNUAL EXCEEDANCES AND ANNUAL MAXIMA

Consider the following hypothetical sequence of river flows.

YEAR	THREE HIGHEST INDEPENDENT FLOWS (cfs)
1	700, 300, 150
2	900, 600, 100
3	550, 400, 200
4	850, 650, 350
5	500, 350, 100

The sequences of ranked annual maxima and annual exceedances are as follows.

$p^x(1-p)^{n-x}$. But this represents just one possible sequence for x successes and $n-x$ failures; all possible sequences must be considered, including those in which the successes do not occur consecutively. The number of possible ways (combinations) of choosing x events out of n possible events is given by the binomial coefficient (Parzen, 1960)

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}. \tag{3.45}$$

Thus, the desired probability is the product of the probability of any one sequence and the number of ways in which such a sequence can occur:

$$P(x) = \binom{n}{x} p^x(1-p)^{n-x}, \qquad x = 0, 1, 2, 3, \ldots, n. \tag{3.46}$$

The notation $B(n, p)$ indicates the binomial distribution with parameters n and p; example PMFs are shown in Fig. 3.11. From Eqs. (3.19) and (3.28), the mean and variance of x are

$$E(x) = np, \tag{3.47}$$

$$\mathrm{Var}(x) = np(1-p). \tag{3.48}$$

The skewness is

$$g = \frac{1-2p}{[np(1-p)]^{0.5}}. \tag{3.49}$$

Clearly, the skewness is zero and the distribution is symmetric if $p = 0.5$.

The cumulative distribution function is

$$F(x) = \sum_{i=0}^{x} \binom{n}{i} p^i(1-p)^{n-i}. \tag{3.50}$$

Evaluation of the CDF can get very cumbersome for large values of n and intermediate values of x. It is tabulated by the Chemical Rubber Company (n.d.) and by the National Bureau of Standards (1950). For large values of n, the relationship between the binomial and beta distribution may be used (Abramowitz and Stegun, 1964; Benjamin and Cornell, 1970), or it may be approximated by the normal when $p \approx 0.5$.

EXAMPLE 3.5

RISK AND RELIABILITY

Over a sequence of n yr, what is the probability that the T-yr event will occur at least once? The probability of occurrence in any one year (event) is $p = 1/T$, and the number of occurrences is $B(n, p)$. The Prob (at least one occurrence in n events) is called the **risk**. Thus, the risk is the sum of the probabilities of 1 flood, 2 floods, 3 floods, . . . , n floods occurring

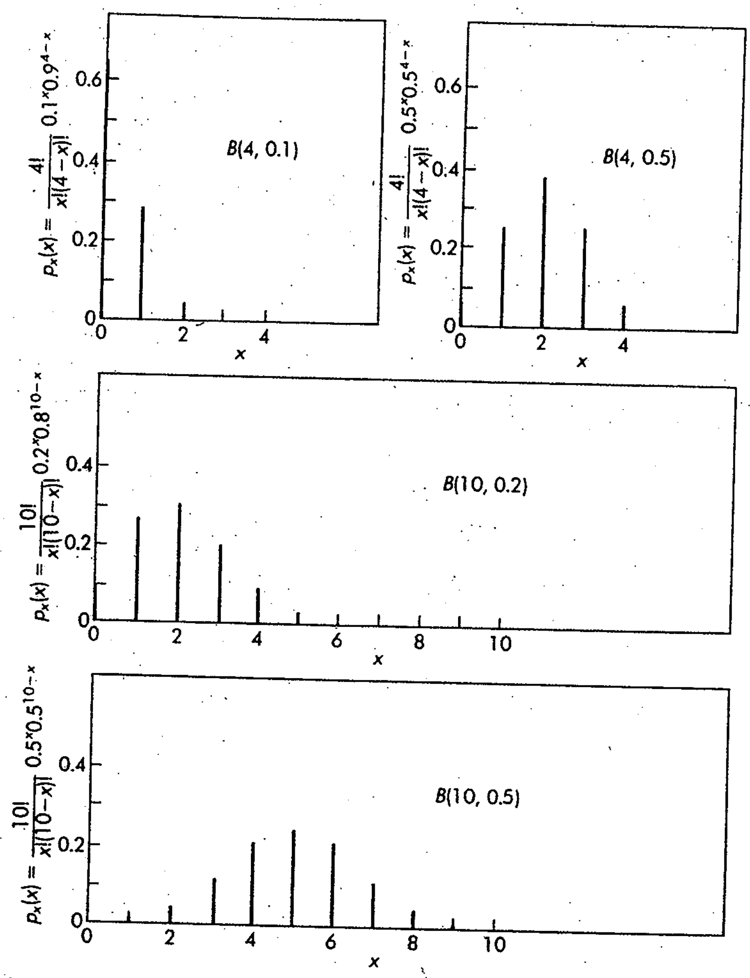

FIGURE 3.11

Binomial probability mass function (PMF). (From Benjamin and Cornell, 1970, Fig. 3.3.1.)

during the n-yr period, but this would be a very tiresome way in which to compute it. Instead,

$$\begin{aligned} \text{Risk} &= 1 - P(0) \\ &= 1 - \text{Prob (no occurrence in } n \text{ yr)} \\ &= 1 - (1 - p)^n \\ &= 1 - (1 - 1/T)^n. \end{aligned} \tag{3.51}$$

Reliability is defined as $1 - \text{risk}$. Thus,

$$\begin{aligned}\text{Reliability} &= (1-p)^n \\ &= (1-1/T)^n. \end{aligned} \tag{3.52}$$

This concept of risk and reliability is very important for hydrologic design. Equation (3.51) can be used to determine the return period required for a given design life and level of risk. Values are shown in Table 3.3 that illustrate the very high return periods required for low risk for a long design life.

EXAMPLE 3.6

CRITICAL FLOOD DESIGN

Consider the 50-yr flood ($p = 0.02$).

a) What is the probability that at least one 50-yr flood will occur during the 30-yr lifetime of a flood control project? This is just the risk of failure discussed above, and the distribution of the number of failures is $B(30, 0.02)$. Thus, from Eq. (3.51),

$$\begin{aligned}\text{Risk} &= 1-(1-0.02)^{30} \\ &= 1-0.98^{30} \\ &= 1-0.545 \\ &= 0.455.\end{aligned}$$

If this risk is too great, the engineer might design for the 100-yr event for which the risk is

$$\text{Risk} = 1 - 0.99^{30} = 0.26$$

and the reliability is 0.74. For this latter circumstance, there is a 26% chance of occurrence of the 100-yr event over the 30-yr lifetime of the project.

b) What is the probability that the 100-yr flood will not occur in 10 yr? In 100 yr?

For $n = 10$, $P(x=0) = (1-p)^{10} = 0.99^{10} = 0.92$,
For $n = 100$, $P(x=0) = (1-p)^{100} = 0.99^{100} = 0.37$.

Thus, there is a 37% chance (37% reliability) that the 100-yr flood will not occur during a sequence of 100 yr.

c) In general, what is the probability of having no floods greater than the T-yr flood during a sequence of ~~T~~ n yr?

$P(x=0) = (1-1/T)^{\cancel{T}\,n}$

TABLE 3.3

Return Periods for Various Degrees of Risk and Expected Design Life (Eq. 3.51)

RISK (%)	RELIABILITY (%)	EXPECTED DESIGN LIFE, n (yr)							
		2	5	10	15	20	25	50	100
75	25	2.0	4.1	7.7	11.3	14.9	18.5	36.6	72.6
63	37	2.6	5.5	10.6	15.6	20.6	25.6	50.8	101.1
50	50	3.4	7.7	14.9	22.1	29.4	36.6	72.6	144.8
40	60	4.4	10.3	20.1	29.9	39.7	49.4	98.4	196.3
30	70	6.1	14.5	28.5	42.6	56.6	70.6	140.7	280.9
25	75	7.5	17.9	35.3	52.6	70.0	87.4	174.3	348.1
20	80	9.5	22.9	45.3	67.7	90.1	112.5	224.6	448.6
15	85	12.8	31.3	62.0	92.8	123.6	154.3	308.2	615.8
10	90	19.5	48.0	95.4	142.9	190.3	237.8	475.1	949.6
5	95	39.5	98.0	195.5	292.9	390.4	487.9	975.3	1950.1
2	98	99.5	248.0	495.5	743.0	990.5	1238.0	2475.4	4950.3
1	99	199.5	498.0	995.5	1493.0	1990.5	2488.0	4975.5	9950.4
0.5	99.5	399.5	998.0	1995.5	2993.0	3990.5	4988.0	9975.5	19950.5