Module 8: Probability and Statistical Methods in Water Resources Engineering

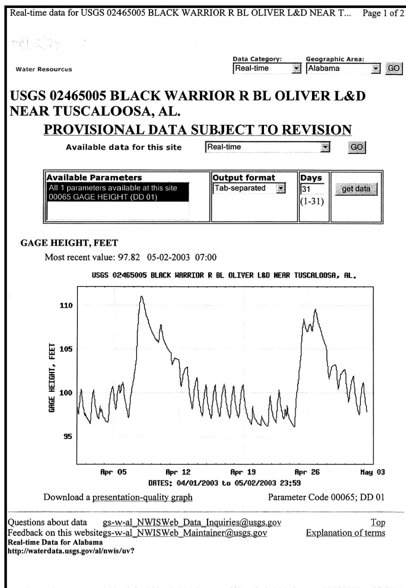
Bob Pitt

University of Alabama

Tuscaloosa, AL

Probability and Statistics

- “Design” storms are frequently used to size hydrologic structures.
- The main question is: What is the possibility of occurrence of a larger storm than our culvert or bridge is barely capable of handling? (what is the likelihood of failure)
- We must be able to evaluate the probability of hydrologic events.
- Frequent statistical goal is therefore to fit a standard probability distribution to observed precipitation and runoff data.



Flow data are available from numerous USGS operated flow recording stations. Data is usually available on a real-time basis. Besides the current flow conditions, summaries of extreme flows are also tabulated and available. These data can be used to predict the frequency of extreme flows of most interest for design.

Probability and Return Periods

- Concepts of probability and statistics are closely associated, particularly when dealing with a set of data.
- The probability of the occurrence of a particular event is simply the chance that the event will occur.
- If there is a finite number of events, a not-necessarily equal probability may be assigned to each.
- If the possible outcomes cover a continuous range of values, the probability can only be expressed by a mathematical function.

- What is the probability of rolling a 1 with a single roll of a six-faced die?
  - Since there are six equally likely outcomes, the probability of obtaining a 1 is 1/6.
- What is the probability of rolling either a 1 or a 4?
  - There are 2 possible outcomes, so the probability of obtaining either a 1 or a 4 is 2/6.
- The probability of equaling or exceeding a particular value is a cumulative probability.
  - Our interest centers on the probability that an event will be equaled or exceeded within a given time frame.
  - "p" is the probability that an event of a specific magnitude is equaled or exceeded in that year.
  - Therefore, 1-p is the probability that the event is ***not*** equaled or exceeded in that year.

- If a discharge of 15,000 cfs has a probability, p = 0.02, then the probability is 2 percent that this discharge will be equaled or exceeded in any one year (and 98% that it will not).
- The reciprocal of p is the return period, or recurrence interval, "t_p". This is the ***average*** time interval between discharges that are equal to, or greater than, a specified discharge:

$$t_p = \frac{1}{p}$$

- The return period expresses the average interval between events, but it does not give specific information concerning the likelihood of occurrence during the design life of the project.
- What is the probability that the discharge of 15,000 cfs with a return period of 50 years will occur during the life of a dam?
  - Will it occur at all?
  - Can it occur more than once?
- The probability that the discharge is not equaled or exceeded in two years is: (1-p)(1-p), or $(1-p)^N$ where N is the period of interest.
- The probability "J" that the event will occur at least once during N years is:

$$J = 1 - (1-p)^N$$

Table 5–1 PROBABILITY THAT AN EVENT WITH RETURN PERIOD t_p WILL BE EQUALED OR EXCEEDED DURING A TIME INTERVAL OF LENGTH *N*

	Time *N* (years)							
	1	5	10	25	50	100	200	500
t_p, year	Probability *J*							
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.500	0.969	0.999	1.000	1.000	1.000	1.000	1.000
5	0.200	0.672	0.893	0.996	1.000	1.000	1.000	1.000
10	0.100	0.410	0.651	0.928	0.995	1.000	1.000	1.000
20	0.050	0.226	0.401	0.723	0.923	0.994	1.000	1.000
30	0.033	0.156	0.288	0.572	0.816	0.966	0.999	1.000
40	0.025	0.119	0.224	0.469	0.718	0.920	0.994	1.000
50	0.020	0.096	0.183	0.397	0.636	0.867	0.982	1.000
100	0.010	0.049	0.096	0.222	0.395	0.634	0.866	0.993
150	0.007	0.033	0.065	0.154	0.284	0.488	0.738	0.965
200	0.005	0.025	0.049	0.118	0.222	0.394	0.633	0.918
250	0.004	0.020	0.039	0.095	0.182	0.330	0.551	0.865
300	0.003	0.017	0.033	0.080	0.154	0.284	0.487	0.812
350	0.003	0.014	0.028	0.069	0.133	0.249	0.436	0.761
400	0.002	0.012	0.025	0.061	0.118	0.221	0.394	0.714
450	0.002	0.011	0.022	0.054	0.105	0.199	0.359	0.671
500	0.002	0.010	0.020	0.049	0.095	0.181	0.330	0.632
1000	0.001	0.005	0.010	0.025	0.049	0.095	0.181	0.394

Prasuhn 1987

- If a dam has a design life of 50 years (N=50 years), the 15,000 cfs flow (having a recurrence interval, t_p, of 50 years), will occur, or be exceeded with a probability of 63.6 percent (***not*** 100%!).
- What is the probability that this flow will occur (or be exceeded) during the 5-year construction period? (N=5 years). (Answer: 9.6%).
- What is the probability that a "100 year" storm will occur at least once during a 200 year period? During a 50 year period? During 1 year?

Probability Distributions

- A probability density function (PDF) is a continuous mathematical expression that determines the probability of a specific event.
- The distribution that best fits the set of data is expected to give the best estimate of the probability of an event that has not been observed.
- Actual discharges or precipitation values over a period of years form a continuous function, because any value is possible, within a broad range.
- We will only examine a few possible probability distributions.

Uniform distributions are of little interest in hydrology, but are a simple place to start, and are similar to the die problem.

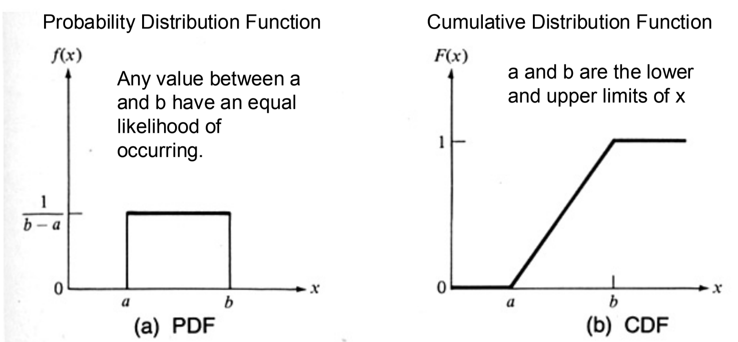


Figure 5–1 Uniform probability distribution.

The probability that an outcome will be equal to, or greater than, a particular value of x is:

$$p = 1 - F(x) = 1 - \int_{-\infty}^{x} f(x)dx$$

Prasuhn 1987

Normal distributions are familiar bell-shaped curves.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\bar{x})^2}{2\sigma^2}\right]$$

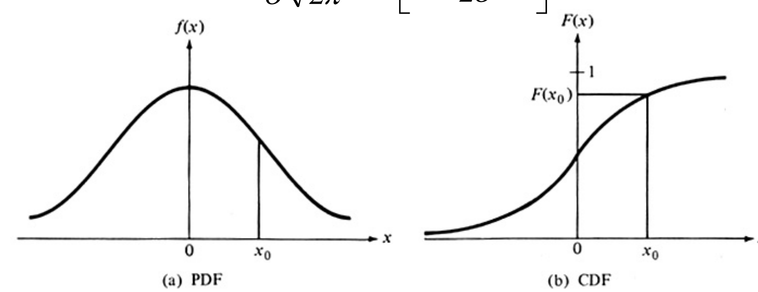


Figure 5–2 Normal probability distribution.

Prasuhn 1987

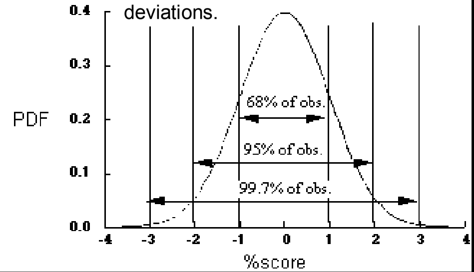
The normal distribution PDF is defined by two distribution parameters, the mean and the standard deviation. The mean is the average of all of the observations:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

and is at the center of the distribution. The standard deviation describes the width of the distribution (the spread of the data):

$$\sigma = \left[\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}\right]^{1/2}$$

About 68% of the data is within +/- 1 standard deviations of the mean, while about 98% is within +/- 3 standard deviations.



The normal distribution does not provide a satisfactory fit to flood discharges and other hydrologic data. The distribution extends from negative to positive infinity and therefore assigns a probability to negative flows. A specific event can be related to the probability of exceedance p by:

$$x = \bar{x} + K\sigma$$

where K is the frequency factor (given in the following table for specific values of p). Using actual data, K can be calculated:

$$K = \frac{(x - \bar{x})}{\sigma}$$

This is the number of standard deviations the data point is located from the mean

If actual hydrologic data are to be expressed with respect to p, the probability of exceedance in a year, the data set will often be based on the single peak value observed for each year (the annual series).

Table 5–2 FREQUENCY FACTOR FOR THE NORMAL DISTRIBUTION

Exceedance probability *p*	Frequency factor *K*	Exceedance probability *p*	Frequency factor *K*
0.0001	3.719	0.550	−0.126
0.0005	3.291	0.600	−0.253
0.001	3.090	0.650	−0.385
0.005	2.576	0.700	−0.524
0.010	2.326	0.750	−0.674
0.020	2.054	0.800	−0.842
0.030	1.881	0.850	−1.036
0.040	1.751	0.900	−1.282
0.050	1.645	0.925	−1.440
0.075	1.440	0.950	−1.645
0.100	1.282	0.960	−1.751
0.150	1.036	0.970	−1.881
0.200	0.842	0.980	−2.054
0.250	0.674	0.990	−2.326
0.300	0.524	0.995	−2.576
0.350	0.385	0.999	−3.090
0.400	0.253	0.9995	−3.291
0.450	0.126	0.9999	−3.719
0.500	0.000		

Prasuhn 1987

Example 5-1 (Prasuhn 1987)

Determine the probability that a discharge of 20,000 cfs will be equaled or exceed in any one year, if the mean of the annual series of river discharges is 10,000 cfs and the standard deviation is (1) 3,000 cfs, or (2) 6,000 cfs. What is the return period in each case?

Solution:

(1)

$$K = \frac{20{,}000cfs - 10{,}000cfs}{3{,}000cfs} = 3.33$$

Looking up this value of K on the previous table yields a p of about 0.0005, and the corresponding recurrence interval:

$$t_p = \frac{1}{p} = \frac{1}{0.0005} = 2{,}000\,years$$

(2)

$$K = \frac{20{,}000cfs - 10{,}000cfs}{6{,}000cfs} = 1.67$$

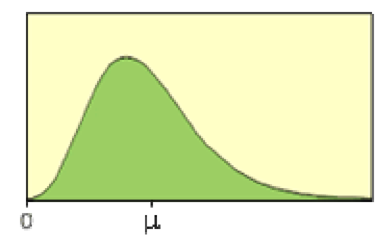
The corresponding p value on the table is about 0.05

$$t_p = \frac{1}{p} = \frac{1}{0.05} = 20 years$$

Therefore, the same value can have vastly different recurrence intervals, even with the same average flow rates, as the standard deviation changes.

Log-Normal Distributions

The log-normal distribution is closely related to the normal distribution. The values are transformed by taking the $\log_{10}$ of the data. This distribution is much more useful than the basic normal distribution as no negative numbers are allowed, while large positive values are acceptable. The following plot shows this distorted distribution in "real" (non-transformed) space:



This distribution assumes that the logarithms of the discharges are normally distributed. The prior equations can be used describing the mean and standard deviations, if the following transformation is used: $y_i = \log x_i$

The mean of the logarithms themselves can be expressed directly:

$$\log \bar{x} = \frac{1}{N}\sum_{i=1}^{N} \log x_i$$

As an alternative, the mean can be found by taking the logarithm of the geometric mean of the set of values:

$$\log \bar{x} = \log\left(x_1 x_2 x_3 \ldots x_N\right)^{1/N}$$

The standard deviation can also be directly calculated where the values are both based on the logarithms of the actual data:

$$\log \sigma = \left[\frac{\sum_{i=1}^{N}\left(\log x_i - \log \bar{x}\right)^2}{N-1}\right]^{1/2}$$

The probability of exceedance can be related to the occurrence using:

$$\log x = \log \bar{x} + K\sigma_{\log x}$$

The frequency factors may be determined from the prior table for normal distributions.

Log Pearson Type III Distribution

The problem with most hydrologic data is that an equal data spread does not occur above and below the mean. The lower side is limited to the range from the mean to zero, while there is no limit to the upper range. This results in a skewed distribution that may not be completely corrected by the log-normal distribution. The coefficient of skew, a, is defined by:

$$a = \frac{N\sum_{i=1}^{N}(x_i - \bar{x})^3}{(N-1)(N-2)\sigma^3}$$

The use of log values in the log-normal distribution tends to reduce the distribution distortion. However, some skewness usually remains.

To determine the skewness when using log values, the following can be used:

$$a_{\log x} = \frac{N\sum_{i=1}^{N}(\log x_i - \log\bar{x})^3}{(N-1)(N-2)\sigma^3_{\log x}}$$

The normal and log-normal distributions assume zero skew. If some skew exists in the data, these distributions result in errors. The log Pearson Type III distribution was developed in 1967 to improve the fit of hydrologic data. This method uses a third parameter, the skew coefficient, in addition to the mean and standard deviation. The following table gives the frequency factors for this distribution. The zero skew values correspond to the log-normal distribution.

Table 5–3 FREQUENCY FACTOR *K* FOR THE LOG PEARSON TYPE III DISTRIBUTION

Skew	Recurrence interval (years)								
	1.010	1.111	2	5	10	25	50	100	200
	Probability of exceedance (%)								
a	99	90	50	20	10	4	2	1	0.5
				Positive Skew					
3.0	−0.667	−0.660	−0.396	0.420	1.180	2.278	3.152	4.051	4.970
2.5	−0.799	−0.771	−0.360	0.518	1.250	2.262	3.048	3.845	4.652
2.0	−0.990	−0.895	−0.307	0.609	1.302	2.219	2.912	3.605	4.298
1.8	−1.078	−0.945	−0.282	0.643	1.318	2.193	2.848	3.499	4.147
1.6	−1.197	−0.994	−0.254	0.675	1.329	2.163	2.780	3.388	3.990
1.4	−1.318	−1.041	−0.225	0.705	1.337	2.128	2.706	3.271	3.828
1.2	−1.449	−1.086	−0.195	0.732	1.340	2.087	2.626	3.149	3.661
1.0	−1.588	−1.128	−0.164	0.758	1.340	2.043	2.542	3.022	3.489
0.9	−1.660	−1.147	−0.148	0.769	1.339	2.018	2.498	2.957	3.401
0.8	−1.733	−1.166	−0.132	0.780	1.336	1.993	2.453	2.891	3.312
0.7	−1.806	−1.183	−0.116	0.790	1.333	1.967	2.407	2.824	3.223
0.6	−1.880	−1.200	−0.099	0.800	1.328	1.939	2.359	2.755	3.132
0.5	−1.955	−1.216	−0.083	0.808	1.323	1.910	2.311	2.686	3.041
0.4	−2.029	−1.231	−0.066	0.816	1.317	1.880	2.261	2.615	2.949
0.3	−2.104	−1.245	−0.050	0.824	1.309	1.849	2.211	2.544	2.856
0.2	−2.178	−1.258	−0.033	0.830	1.301	1.818	2.159	2.472	2.763
0.1	−2.252	−1.270	−0.017	0.836	1.292	1.785	2.107	2.400	2.670
0.0	−2.326	−1.282	0.000	0.842	1.282	1.751	2.054	2.326	2.576
				Negative Skew					
−0.1	−2.400	−1.292	0.017	0.846	1.270	1.716	2.000	2.252	2.482
−0.2	−2.472	−1.301	0.033	0.850	1.258	1.680	1.945	2.178	2.388
−0.3	−2.544	−1.309	0.050	0.853	1.245	1.643	1.890	2.104	2.294
−0.4	−2.615	−1.317	0.066	0.855	1.231	1.606	1.834	2.029	2.201
−0.5	−2.686	−1.323	0.083	0.856	1.216	1.567	1.777	1.955	2.108
−0.6	−2.755	−1.328	0.099	0.857	1.200	1.528	1.720	1.880	2.106
−0.7	−2.824	−1.333	0.116	0.857	1.183	1.488	1.663	1.806	1.926
−0.8	−2.891	−1.336	0.132	0.856	1.166	1.448	1.606	1.733	1.837
−0.9	−2.957	−1.339	0.148	0.854	1.147	1.407	1.549	1.660	1.749
−1.0	−3.022	−1.340	0.164	0.852	1.128	1.366	1.492	1.588	1.664
−1.2	−3.149	−1.340	0.195	0.844	1.086	1.282	1.379	1.449	1.501
−1.4	−3.271	−1.337	0.225	0.832	1.041	1.198	1.207	1.318	1.351
−1.6	−3.388	−1.329	0.254	0.817	0.994	1.116	1.166	1.197	1.216
−1.8	−3.499	−1.318	0.282	0.799	0.945	1.035	1.069	1.087	1.097
−2.0	−3.605	−1.302	0.307	0.777	0.895	0.959	0.980	0.990	0.995
−2.5	−3.845	−1.250	0.360	0.711	0.771	0.739	0.798	0.799	0.800
−3.0	−4.051	−1.180	0.396	0.636	0.660	0.666	0.666	0.667	0.667

Prasuhn 1987

Example 5-2 (Prasuhn 1987)

The mean of the logarithms of the annual series of river discharges is 2.700 (which corresponds to a geometric mean for the peak flows of 501 m^3/sec). The standard deviation of the same log values is 0.65. Determine the discharge with a 100-year period if the coefficient of skew is (1) -0.4, (2) 0, and (3) +0.4.

Solution:

Log x = 2.7 + 0.65 K

(1) K = 2.029; log Q = 4.019; Q = 10,400 m^3/sec
(2) K = 2.326; log Q = 4.212; Q = 16,300 m^3/sec (log-normal)
(3) K = 2.615; log Q = 4.400; Q = 25,100 m^3/sec

Statistical Analysis

An annual series uses the maximum values for each year. A drawback to the annual series is that some years may have several large peaks, while other years may have peaks that are much lower. However, only one value per year can be used, discarding some potentially valuable information. A partial duration series uses all values above a selected value. For rare events, the results are similar and the annual series is recommended because it is easier to obtain. However, if more frequent flows are of interest (such as recurrence intervals of less than one year), the partial duration series should be used.

The following table shows all annual peak flows, plus all others greater than 9,000 cfs. This data can therefore be used for either a partial duration series, or an annual series analysis.

Table 5–4 PEAK DISCHARGE DATA FOR THE BIG SIOUX RIVER AT AKRON, IOWA.

Year	Peak Q (cfs)	Year	Peak Q (cfs)	Year	Peak Q (cfs)	Year	Peak Q (cfs)
1929	20,800	1944	15,900	1954	21,700		13,100
1930	3740		11,600		15,600	1970	8580
1931	1390		11,100	1955	4940	1971	7310
1932	16,900		9840	1956	1840	1972	10,500
1933	14,200	1945	12,300	1957	19,400		10,200
1934	10,600		9820	1958	1120	1973	12,500
1935	3000	1946	8970	1959	8430	1974	3000
1936	18,000	1947	10,500	1960	49,500	1975	2920
1937	5760	1948	10,800	1961	9050	1976	3250
1938	12,700	1949	11,400	1962	54,300	1977	5270
	11,200		11,400		9010	1978	18,200
	9800		9170	1963	1650	1979	30,500
1939	6300	1950	5450	1964	2540		13,100
1940	11,700	1951	28,800	1965	21,000		12,600
1941	5820	1952	33,000	1966	16,500	1980	8730
1942	21,400		9650	1967	5300	1981	3180
	16,600		16,500	1968	635		
1943	12,000	1953	21,800	1969	80,800		

Prasuhn 1987

The following table shows the last segment of an annual series analysis of this data. Included are the ranks for the smallest 12 annual peak flows, the corresponding flows, and several columns of summary statistics for these flows. This analysis can be easily conducted on a spreadsheet program.

Each flow has its recurrence interval calculated for N years of record (53 years), where m is the rank:

$$t_p = \frac{N+1}{m}$$

The probability of exceedance is the reciprocal of the return period, or:

$$p = \frac{m}{N+1}$$

A rank of 1 in a period of 10 years leads to a probability of p = 1/11 = 0.0909. N+1 results in the best estimate for limited data sets. In a partial duration series, N still refers to the duration of the record and is different than the number of observations.

Table 5–5 STATISTICAL ANALYSIS OF THE ANNUAL SERIES FOR THE BIG SIOUX RIVER AT AKRON, IOWA. *(continued)*

Rank (m) (1)	Flow rate (cfs) (2)	t_p (yr) (3)	p (%≥) (4)	log x (5)	$(\log x - \overline{\log x})$ (6)	$(\log x - \overline{\log x})^2$ (7)	$(\log x - \overline{\log x})^3$ (8)
42	3740	1.29	77.78	3.573	−0.3763	0.1416	−0.0533
43	3250	1.26	79.63	3.512	−0.4373	0.1912	−0.0836
44	3180	1.23	81.48	3.502	−0.4467	0.1996	−0.0892
45	3000	1.20	83.33	3.477	−0.4721	0.2228	−0.1052
46	3000	1.17	85.19	3.477	−0.4721	0.2228	−0.1052
47	2920	1.15	87.04	3.465	−0.4838	0.2341	−0.1132
48	2540	1.12	88.89	3.405	−0.5443	0.2963	−0.1613
49	1840	1.10	90.74	3.265	−0.6844	0.4683	−0.3205
50	1650	1.08	92.59	3.217	−0.7317	0.5354	−0.3917
51	1390	1.06	94.44	3.143	−0.8062	0.6499	−0.5239
52	1120	1.04	96.30	3.049	−0.9000	0.8099	−0.7289
53	635	1.02	98.15	2.803	−1.1464	1.3142	−1.5066

Prasuhn 1987

Summations for Table 5–5

$$\Sigma x = 735{,}875 \text{ (col. 2)}$$
$$\Sigma(x - \bar{x}) = 1.144 \times 10^{-4} \text{ (col. not printed)}$$
$$\Sigma(x - \bar{x})^2 = 1.0940 \times 10^{10} \text{ (col. not printed)}$$
$$\Sigma \log x = 209.306 \text{ (col. 5)}$$
$$\Sigma(\log x - \overline{\log x}) = 1.565 \times 10^{-7} \text{ (col. 6)}$$
$$\Sigma(\log x - \overline{\log x})^2 = 9.9743 \text{ (col. 7)}$$
$$\Sigma(\log x - \overline{\log x})^3 = -1.5454 \text{ (col. 8)}$$

Statistical Results for Table 5–5

Mean of x = 13,844 cfs

Standard deviation of x = 14,505 cfs

Mean of log x = 3.949

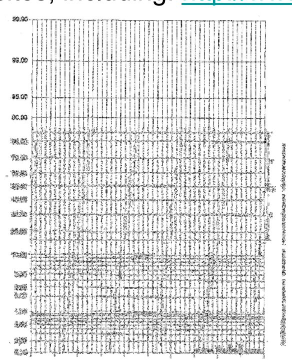
(Geometric mean of x = 8896 cfs)

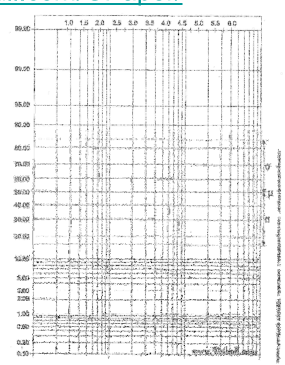
Standard deviation of log x = 0.4380

Skew of log x = −0.368

Prasuhn 1987

It is now possible to plot the data and the fitted equations for the different distributions to normal and log-normal probability paper to determine the best fitting distribution, and to use the plots to determine flows for different recurrence intervals. Probability paper can be easily downloaded at several web sites, including: http://www.weibull.com/GPaper/





The use of probability paper allows visual clues as to the best distribution (usually the one with the best fit for the data has a straight line, at least for normal and log-normal plots, or that fits the curved plotted line for log-Pearson type III plots). The first step is to create the annual series (or partial series) data and rank the observations, usually from the largest to the smallest. Then calculate the probability for each observation, using p = m/(N+1). Finally, just plot the flow values against the calculated p values. The following plots are examples using the Sioux River data.

Do an in-class example to plot the following 9 observations:
15 5 78 56 13 7 32 22 88

The following plot uses the Big Sioux River data on a normal plot. For this plot, the flow values are plotted on an arithmetic scale and the probabilities are plotted on scales that are distorted so that a normal distribution would plot as a straight line. Note that this is ***not*** a log scale.

Besides the data points (which are not along a straight line, an indication that this is not a suitable distribution), a straight line which corresponds to the best fit for this data is also plotted. The equation for this line is based on the data characteristics:

$$x = \bar{x} + K\sigma$$

$$x = 13{,}844 + 14{,}505K$$

To plot the straight line, values of K are obtained from the prior table of K values for normality probability for selected p values. The x flow values are then calculated corresponding to these p values, and then plotted to form the line.

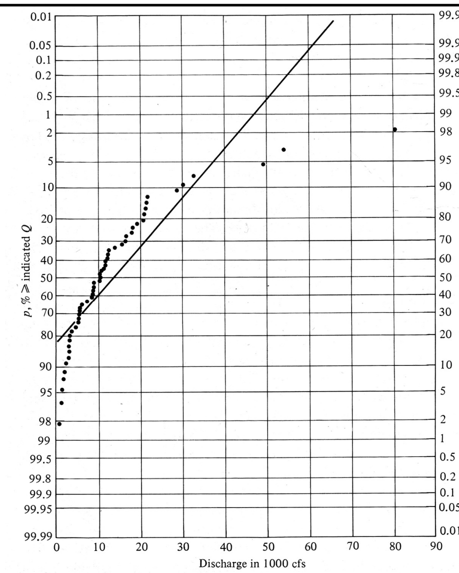


Figure 5–3 Normal distribution for the Big Sioux River at Akron, Iowa.

Prasuhn 1987

Not a very good fit, so the following plot for log-normal distributions are attempted, using the same data, but using log-normal probability paper.

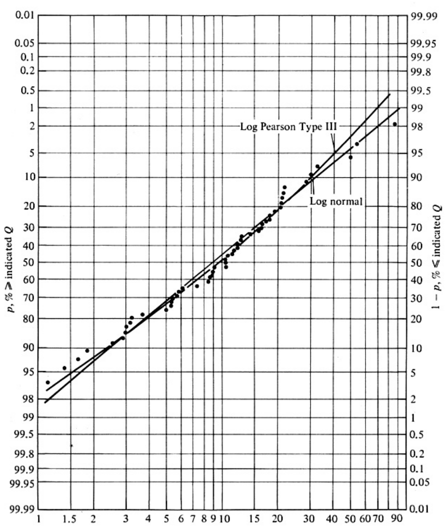


Figure 5–4 Log-normal and log Pearson Type III distributions for the Big Sioux River at Akron, Iowa.

Prasuhn 1987

The flow values are plotted on the log scale, and the same probability vs. flow values are used to plot the straight line. The actual equation for the straight line is:

$$\log x = \log \bar{x} + K\sigma_{\log x}$$

$$\log x = 3.949 + 0.4380K$$

This plot also shows the log Pearson type III plot, using the calculated skew.

The curved line for the log Pearson type III distribution is obtained the same way, except the calculated skew value is used to obtain the K parameter for the equation. In this example, the skew coefficient a is -0.368.

Both of these fitted lines are not perfect fits of the observed data, and lead to very different results when used to extrapolate to large recurrence interval flows. The Pearson curve fits the overall data range better, but the log-normal curve fits the 3 largest values (usually of most interest) better. The best choice is therefore sometimes difficult to determine.

What is the expected discharge having a 200 year recurrence interval (t_p = 200 years, p = 0.005)? The calculated lines could be extended to this value, or the equations can be directly used.

Log-normal:
p = 0.005 and a =0
K = 2.576 from the table

Therefore:
Log x = 3.949 + (0.4380) (2.576) = 5.0773
and x = $10^{5.0773}$ = 119,500 cfs

Log Pearson Type III:
p = 0.005 and a = -0.368
K = 2.231 from the table

Therefore:
Log x = 3.949 + (0.4380)(2.231) = 4.9262
and x = $10^{4.9262}$ = 84,400 cfs

There is considerable discrepancy between these two predicted values. The larger value is more conservative and is more consistent with the larger observed discharges. However, the lower value is probably the better estimate as it fits the complete data set better.

Measuring the large actual discharges is subject to considerable error, as they were likely occurring during flood stage conditions where the flow measurement station may have been submerged, or beyond calibration depths, requiring crude estimates of actual flows based on physical evidence. Also, the largest flow was not likely associated with an exact 54-year event. There is no way of knowing what size event it was; could have been associated with a much more rare event, such as the 100 year event that just happened to occur during the shorter period of record.

Normally, the log Pearson distribution is recommended as it considers the skew parameter, but caution is needed as unrealistic and excessive values of skew may occur for a particular river. Regional skew values should also be examined.

Extreme flow events are usually well outside of the normal channel and measurement accuracy suffers.

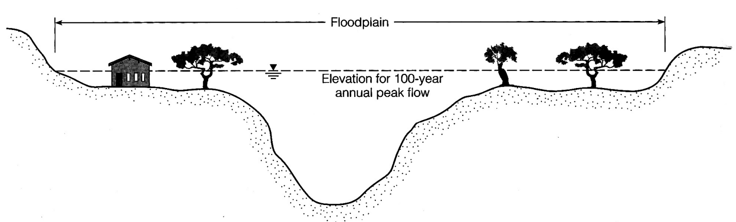


Figure 5.11 ■ Cross-Section of Floodplain

Chin 2000

Example 5-4 (Prasuhn 1987)

Use the three distribution methods to predict the 50-year flood on the Big Sioux River at Akron, Iowa.

Solution:

(1)Normal distribution:
t_p = 50 yrs, p = 0.02.
Therefore K = 2.054
x = 13,844 + (2.054)(14,505) = 43,600 cfs

(2) Log-normal distribution:
a = 0
K = 2.054
log x = 3.949 + (2.054)(0.4380) = 4.849
X = $10^{4.849}$ = 70,600 cfs

(3) Log Pearson type III
$a = -0.368$
$K = 1.852$

$\text{Log } x = 3.949 + (1.852)(0.4380) = 4.760$
$X = 10^{4.760} = 57{,}600$ cfs

Obviously, the application of statistical methods is not an exact science. The methods are extremely helpful in the interpretation of hydrologic data and the prediction of design tools, but the engineer must be aware of the limitations involved.

Not only are flood conditions important, but drought conditions can also be evaluated using the same methods. When dealing with rainfall, the intensity of the precipitation as well as the overall quantity is important.

Homework Problem:

Repeat the Big Sioux River analysis, but only use data from the last 10 years of record (1972 to 1981). Predict the 50 and 100 year flows using the log Pearson type III distribution. What are the limitations of using a short period of observations?

References

Chin, David, A. *Water-Resources Engineering*. Prentice Hall. 2000.

Prasuhn, Alan L. *Fundamentals of Hydraulic Engineering*. Holt, Rinehart and Winston. 1987.