1. Hydrologic Statistics

Hydrologic processes evolve in space and time in a manner that is partly predictable , or deterministic, and partly random or *stochastic process*.This section describes hydrologic data from pure random processes using statistical parameters and functions. Consequently we will focus on the observations themselves rather than on the physical processes which produced them. Statistics is a science of description, not causality.

A random variable X is a variable described by a probability distribution, the distribution specifies that chance that an observation x of the variable will fall in a specified range of X. A set of observations $x_1, x_2, ..., x_n$ of the random variable is called a sample that is drawn from a hypothetical infinite population possessing constant statistical properties.

The *probability* of an event $P(A)$ is the chance that it will occur when and observations of the random variable is made. If a sample of n observations has n_A values in the range of event A, then the relative frequency of A is n_A/n. As the sample size increases, the relative frequency becomes a progressively better estimate of the probability of the event.

$$P(A) = lim_{n \to \infty} \frac{n_A}{n} \tag{1}$$

If the outcomes $A, B, C...$ of an experiment are mutually exclusive

$$P(\text{A or B or C or ...}) = P(A) + P(B) + P(C) + ... \tag{2}$$

and the sum of the probabilities of all possible outcomes is unity. If the outcomes of successive or separate trials are independent (the result of one doesn't affect outcomes of other trials:

$$P(\text{A and B and C and ...}) = P(A) \times P(B) \times P(C) \times ... \tag{3}$$

See Example 1.

1a. *Discrete Random Variables*

If a random variable can take on only specific exact numerical values, it is called a discrete random variable (days with rainfall greater than 25 mm). If a particular random variable can take on any numerical value within some interval, it is called a continuous random variable (streamflow). We can convert continuous time series into discrete.

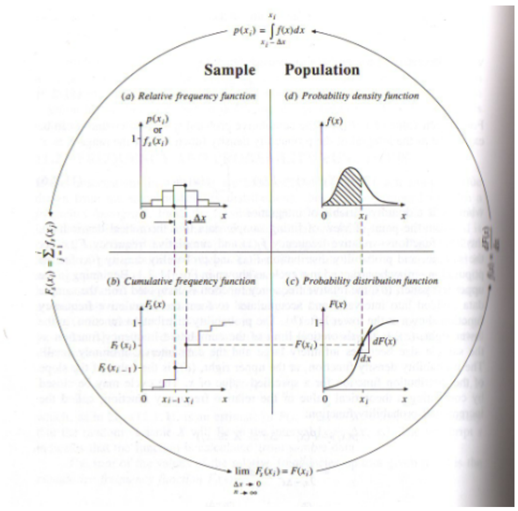


Figure 1: Discrete and Continuous Variables

The probability distribution for a discrete random variable X is the probability that X takes on a particular value x_i:

$$p_X(x_i) = P(X = x_i) \tag{4}$$

$p_X(x_i)$ is the probability function.

$$\sum_{\text{all X}} p_X(x_i) = 1 \tag{5}$$

The cumulative probability function ($P_X(x_i)$) is the probability that X takes on a value less than or equal to x_i.

$$P_X(x_i) = P(X \leq x_i) = \sum_{allX \leq x_i} p_X(x_i) \tag{6}$$

The probability that X takes on values between x_j and x_k

$$P(x_j \leq X \leq x_k) = \sum_{x_i=x_j}^{x_k} p_X(x_i) \tag{7}$$

1b. *Continuous Random Variables*

The probability of occurrence of any particular exact value is zero. To define these probabilities we define a *cumulative distribution function* of a continuous variable X as

$$F_X(x) = P(X \leq x) \tag{8}$$

The relative probability that a continuous random variable X takes on a particular value x is expressed by the *probability density function*

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{9}$$

$$f_X(x) = \int_{-\infty}^{x} f_X(x)dx \tag{10}$$

The probability that X takes on a value between $x = a$ and $x = b$ is given by

$$P(x \leq X \leq b) = \int_a^b f_X(x)dx = F_X(b) - F_X(a) \tag{11}$$

$$\int_{-\infty}^{\infty} f_X(x)dx = 1 \tag{12}$$

The cdf is a complete specification of the statistical properties of a random variable. There are many distributions commonly used for hydrologic variables including *uniform, normal, lognormal (hydraulic conductivity in porous media, raindrop sizes in a sorm), exponential (interarrival time of storms), gumbel, generalized extreme value, generalized pareto*

Statistics are numbers calculated from a sample which summarize its important characteristics. A statistical parameter is the *expected value E* of some function of a random variable. Denote an arbitraty function of X as $\psi(X)$ then the expected value of $\psi(X)$ is

$$E[\psi(X)] = \int_{-\infty}^{\infty} \psi(x) f_X(x) dx \tag{13}$$

1c. *Quantiles*

One of the simples ways of describing the distribution of a random variable is to give the value of several quantiles of the distribution. The q^{th} quantile of the variable X is the value x_q that is larger than $100q$ percent of all values.

$$F_X(x_q) = q \tag{14}$$

The most commonly reported quantiles are the median $x_{.50}$, the lower quartile $x_{.25}$ and the upper quartile $x_{.75}$. The interval [$x_{.25}$,$x_{.75}$] is the interquartile range. If we have a sample of N measured values of a random variable X, we must first rank (sort) the values. then we estimate x_q as $x_(i)$ where i indicates rank. According to the Weibull plotting-position formula

$$q = \frac{i}{N+1} \tag{15}$$

We must interpolate to obtain x_q for the desired q

2. Design Storm

Precipitation pattern defined for use in the design of a hydrologic system. Design storm serves as the system input to a rainfall-runoff model, like the rational method for determining peak flow rates in storm sewers and highway culverts, or to determine storm hyetographs to input into rainfall-runoff analysis for spillway design in large reservoir projects.

2a. *Design Precipitation Depth*

1. For point precipitation analysis (as opposed to areal precipitation), the annual maximum precipitation (depth or intensity) for a given time interval in a storm is found by computing a series of running totals of rainfall depth for that time interval starting at various points in the storm, then selecting the maximum value of this series. Notice that as the time period increases, the average intensity sustained by he storm decreases. This method is useful to compare how severe a particular storm is, compared to other storms and give useful data for design of flow control

structures. We choose the annual maximum precipitation for a given duration, and the process is repeated for several durations.

Time (min)	Rainfall (in)	30 min	1 hr	2 hr
0				
5	0.02			
10	0.34			
15	0.1			
20	0.04			
25	0.19			
30	0.48	1.17		
35	0.5	1.65		
40	0.5	1.81		
45	0.51	2.22		
50	0.16	2.34		
55	0.31	2.46		
60	0.66	2.64	3.81	
65	0.36	2.5	4.15	
70	0.39	2.39	4.2	
75	0.36	2.24	4.46	
80	0.54	2.62	4.96	
85	0.76	**3.07**	5.53	
90	0.51	2.92	**5.56**	
95	0.44	3	5.5	
100	0.25	2.86	5.25	
105	0.25	2.75	4.99	
110	0.22	2.43	5.05	
115	0.15	1.82	4.89	
120	0.09	1.4	4.32	8.13
125	0.09	1.05	4.05	**8.2**
130	0.12	0.92	3.78	7.98
135	0.03	0.7	3.45	7.91
140	0.01	0.49	2.92	7.88
145	0.02	0.36	2.18	7.71
150	0.01	0.28	1.68	7.24
Max depth		3.07	5.56	8.2
Max intensity		6.14	5.56	4.1

2. Then, we rank the values for each duration from highest to lowest and compute the estimated quantile for all ranks.

3. Then we can interpolate to determine the depths associated with the return periods of interest. Suppose that an extreme event is defined to have occurred if a random variable X is greater than or equal to some level x_T. The *recurrence interval* τ is the time between occurrences of $X > x_T$. The **return period** is the expected value of τ, or the average recurrence interval between events equalling or exceeding a specified magnitude. The probability of $P(X \geq x_T$ may be related to the return period

$$E(\tau) = T = \frac{1}{p} \tag{16}$$

where p is the probability of $X \geq x_T$. So the probability of occurrence of an event in any observation is the inverse of its return period. For a 2, 5, 10 and 25 year rainfalls are read from the 50, 20, 10 and 4% exceedence probabilities $P(X > x) = 1 - F_X(x)$.

Example 4.6 from Dingman.

i. Extreme Value Distributions The plotting-position approach is not appropriate for estimating return periods greater than the length of the record, so in this cases more sophisticated statistical analysis may be required Since observations of extreme events are located in the extreme tail of the probability distribution of all observations from which they are drawn (the parent population) it is not surprising that their probability distribution is different from that of the parent population. There are three asymptotic forms of the distributions of extreme values, Type I, Type II and Type III.

$$F(x) = exp\left(-\left(1 - k\frac{x-u}{\alpha}\right)^{1/k}\right) \tag{17}$$

The three cases are:

Extreme Value Type I k=0. Used for storm rainfalls

Extreme Value Type II k¡0

Extreme Value Type III k¿0. Used for drought flows applied to -x

The Extreme Value Type I (EVI) is (Storm rainfalls are most commonly modeled by the extreme value type I distribution:

$$F(x) = exp\left(-exp\left(-\frac{x-u}{\alpha}\right)\right) \tag{18}$$

The parameters are $\alpha = 6^{0.5}s/\pi$ and $u = \bar{x} - 0.5772\alpha$ where s is the estimate of the standard deviation and $\bar{x}$ is the estimate of the mean. If we define $y = \frac{x-u}{\alpha}$ then,

$$F(x) = exp\left(-exp\left(-y\right)\right) \tag{19}$$

and

$$y = -ln\left(ln\left(\frac{1}{F(x)}\right)\right) \tag{20}$$

The return period T is given by:

$$\frac{1}{T} = P(x \geq x_T) = 1 - P(x < x_T) = 1 - F(x_T) \tag{21}$$

$$F(x_T) = \frac{T-1}{T} \tag{22}$$

and

$$y_T = -ln\left(ln\left(\frac{T}{T-1}\right)\right) \tag{23}$$

2b. *Intensity Duration Frequency analysis*

The most common approach to determine the rainfall event to be used in hydrologic design involves a relationship between intensity (or depth), duration and frequency or return period. Hydrologists use standard intensity-duration-curves. We follow a procedure very similar to that explained in the examples above to create a curve with duration plotted on the horizontal axis, intensity on the vertical axis and a series of curves, one for each return period. The average intensity is expressed as $i = P/T_d$ where P is the rainfall depth and T_d is the duration.

FIGURE 4-49
Logarithmic-probability plots of depths of 1-, 6-, and 24-hr rainfall data from Tables 4-11 and 4-12.

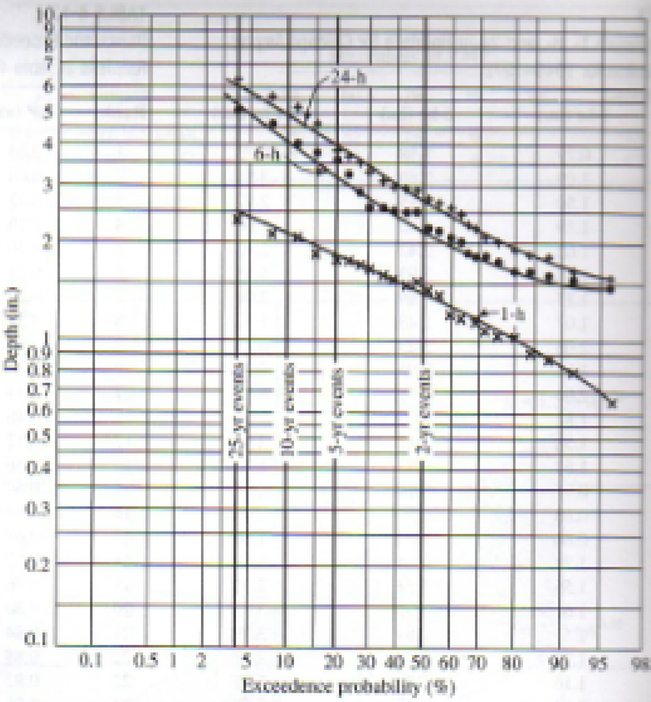
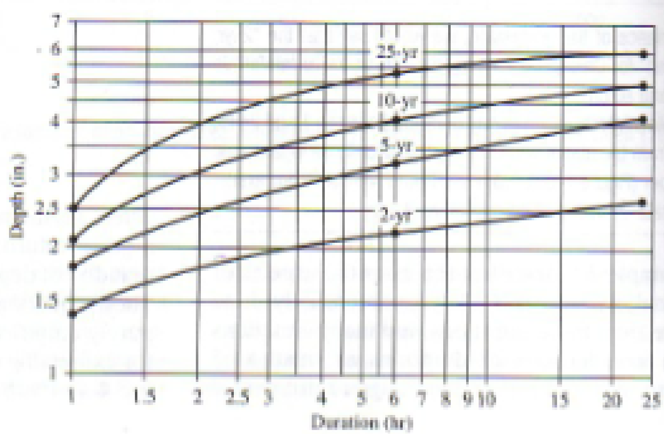
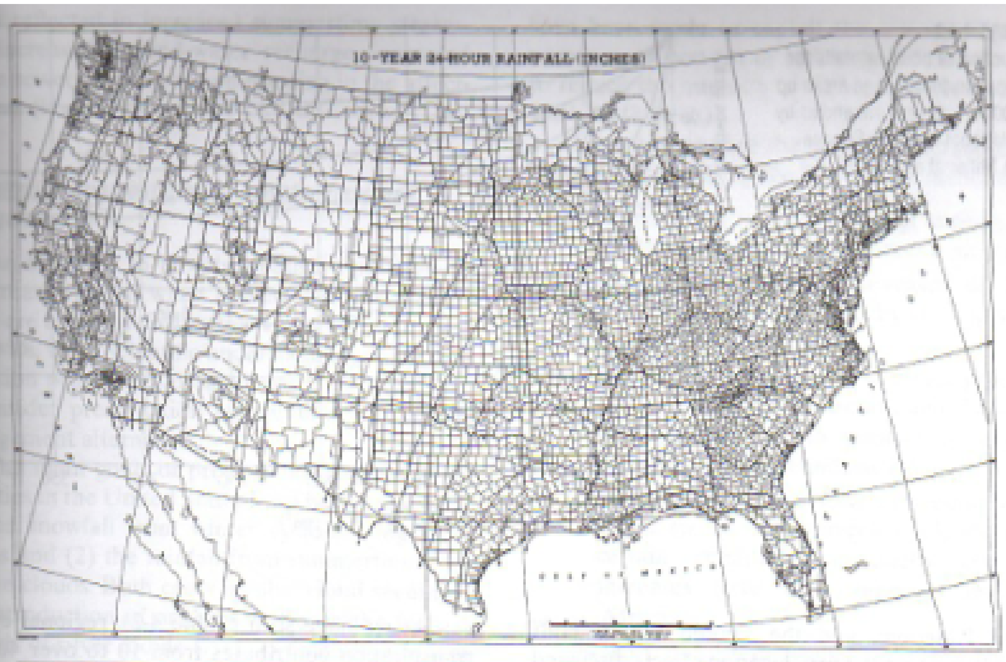


FIGURE 4-50
Depth–duration–frequency plot for Chicago Airport for durations from 1 to 24 hr and return periods from 2 to 25 yr (Example 4-6).




10-YEAR 1-HOUR RAINFALL (INCHES)


100-YEAR 1-HOUR RAINFALL (INCHES)


10-YEAR 24-HOUR RAINFALL (INCHES)


100-YEAR 24-HOUR RAINFALL (INCHES)

FIGURE 4.54