

CEGE 4501 Hydrologic Design

Chapter 9: Statistical Hydrology



UNIVERSITY OF MINNESOTA

Driven to DiscoverSM

Ardeshir Ebtehaj
Department of Civil, Environmental, and Geo-Engineering

December 13, 2017

Outline

Introduction

Common PDFs in Hydrology

Fitting a probability distribution

Goodness of Fit Test:

Frequency Analysis

Water Resources Council Method

Design Storms

Introduction I

Hydrologic processes can be partly explained by deterministic models, however some processes are not still well understood and shall be characterized through statistical models. Statistical models explain hydrologic processes based on their historical observations.

A random variable can be explained with a **probability distribution**, which is a parametric function that characterizes the probability of occurrence of that variable.

Random variable:

A random variable (X) is a variable whose possible values are outcomes of a random process.

There are two types of random variables:

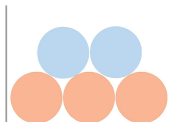
- A discrete random variable, which only takes countable number of values.
- A continuous random variable that takes infinite number of possible values.
 - A finite set of observations x_1, x_2, \dots, x_n of a random variable is called a *sample set*.
 - The space that all samples can be drawn is called the *sample space*.
 - A subset of the sample space is called an *event*.

For example $\Omega = \{x | x = 0, 1, 2, \dots, 10\}$ can represent a sample space, where $A_x = \{x | 1 \leq x \leq 6\}$ is an event.

The box in the right hand side has three oranges (o) and two blue (b) discs as the entire sample space.

$$p(b) = \frac{2}{5}$$

$$p(o) = \frac{3}{5}, \text{ where } p(o) + p(b) = 1$$

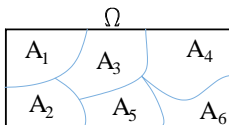


Introduction II

- **Total probability:**

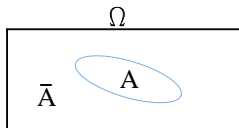
$$p(A_1) + p(A_2) + \dots + p(A_n) = p(\Omega) = 1$$

where A_1, \dots, A_n are disjoint events.



- **Complementarity:**

$$p(A) = 1 - p(\Omega - A) = 1 - p(\bar{A}) \quad \bar{A} : \text{compliment of } A$$

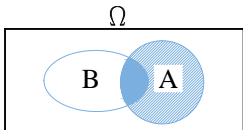


- **Conditional Probability:** Suppose we have two events A and B , the conditional probability $p(A|B)$ refers to the probability of the event A given that the event B has already occurred

$$p(A|B) = \frac{p(A \cap B)}{p(B)},$$

Introduction III

where $p(A \cap B)$ is the joint probability, which is shown with a solid hatch in the following sample space.



If two events are independent, we have,

$$p(A|B) = p(A) = \frac{p(A \cap B)}{p(B)}$$

and thus $p(A \cap B) = p(A) \cdot p(B)$.

- **Marginal probability:** The marginal probability of an event is a probability that does not account for probability of occurrence of other dependent random variables. For example, if $p(A \cap B) \neq \emptyset$ is not empty, then the marginal distribution of $p(A)$ is $p(A) = \sum_B p(A \cap B)$.

For example, if we assume that values of total annual precipitation amounts are independent random variables (X) and $p(X \leq 40'') = a$, then $p(X_1 \leq 40'' \cap X_2 \leq 40'') = a^2$, because these two events are assumed to be independent.

Moreover, if $p(X < 35'') = 0.333$ and $p(X > 45'') = 0.275$, then the complement probability is $p(35 \leq X \leq 45) = 1 - 0.333 - 0.275 = 0.392$.

Introduction IV

Frequency histograms vs probability distributions:

If we have a finite number of n independent and identically distributed samples of X , we can first determine the range of the random numbers and then divide it into discrete intervals with a size of Δx . Then we can count the number of values (n_i) that fall within $[x_i, x_i + \Delta x]$ and divide it by the total number of samples (n) to obtain the frequency of occurrence within each interval.

$$f_s(x_i \leq X \leq x_i + \Delta x) \simeq \frac{n_i}{n}$$

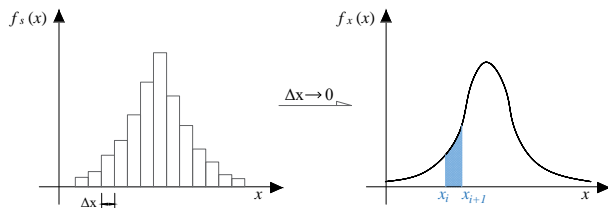


Figure 1: Probability histogram of discrete random numbers (left) versus probability density function of continuous random variables (right).

As $\Delta x \rightarrow 0$, we can say that the histogram approaches to the probability distribution function (PDF) $f_x(x)$, in which the probability of an interval is

$$p_r(x_i \leq X \leq x_i + \Delta x) = \int_{x_i}^{x_i + \Delta x} f_x(u) du.$$

Note that $f_s(\cdot)$ is the frequency function and $f_x(\cdot)$ refers to the probability density function (PDF).

Introduction V

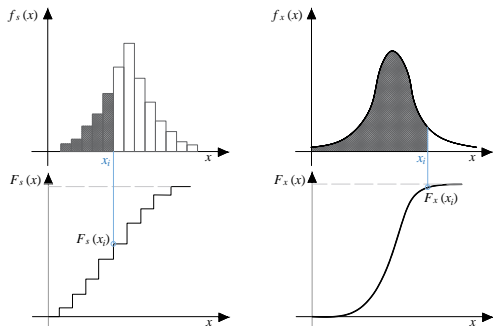


Figure 2: Cumulative frequency (left) and cumulative density functions (CDF) (right).

Clearly, we can define some of the frequency and probability density functions as follows to obtain the cumulative frequency and density functions,

$$F_s(x_i) = \sum_{j=1}^i f_s(x_j) \Rightarrow \text{cumulative frequency function}$$

$$F_X(x_i) = \int_{-\infty}^{x_i} f_X(u) du \Rightarrow \text{cumulative density function.}$$

Introduction VI

As a result, we can have,

$$\text{prob}(x_i \leq X \leq x_i + \Delta x) = F_X(x_i + \Delta x) - F_X(x_i) = \int_{-\infty}^{x_i + \Delta x} f_X(u) du - \int_{-\infty}^{x_i} f_X(u) du$$

Moments of probability distributions:

Moments of a probability distribution are statistical parameters that can be used to extract essential information about the position and shape of a probability distribution.

First-order moment of $f_X(x)$:

$$\mu = \mathbb{E}(x) = \int_{-\infty}^{\infty} x f_X(x) dx$$

Second-order central moment of $f_X(x)$:

$$\sigma^2 = \mathbb{E}(x - \mu)^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx$$

Third-order central moment of $f_X(x)$:

$$\gamma = \mathbb{E}(x - \mu)^3 = \int_{-\infty}^{\infty} (x - \mu)^3 f_X(x) dx$$

A **percentile** of a distribution is a statistic that indicates the value below which a given percentage of the probability mass falls. For example, the 95th percentile is the value below which 95% of the probability mass of A distribution is located.

Introduction VII

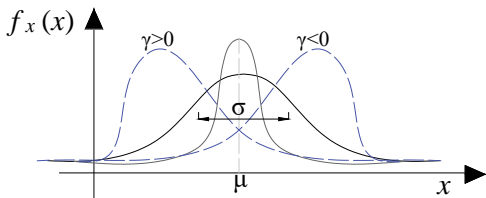


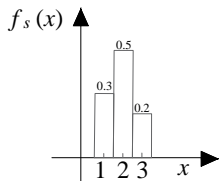
Figure 3: Mean (μ) of a PDF is a measure of its central tendency or location. Standard deviation (σ) is a measure of width or dispersion of the random variable around its mean. Densities with larger standard deviation are wider than those with smaller standard deviations. The third order central moment (γ) is a measure of symmetry or skewness of the random variable. The densities with $\gamma > 0$ are positively skewed and those with $\gamma < 0$ are negatively skewed.

As an example,

$$\mu = \bar{x} = \mathbb{E}(x) = 1 \times 0.3 + 2 \times 0.5 + 3 \times 0.2 = 1.9 = \sum_{i=1}^3 x_i f_s(x_i)$$

$$\sigma^2 = \mathbb{E}(x - \mu)^2 = (1 - 1.9)^2 \times 0.3 + (2 - 1.9)^2 \times 0.5 + (3 - 1.9)^2 \times 0.2 = 0.81 \times 0.3 + 0.01 \times 0.5 + 1.21 \times 0.2 = 0.49$$

$$\gamma = \mathbb{E}(x - \mu)^3 = (1 - 1.9)^3 \times 0.3 + (2 - 1.9)^3 \times 0.5 + (3 - 1.9)^3 \times 0.2 = 0.048$$



Introduction VIII

We often normalize or rescale the statistical moments to make them more meaningful.

Coefficient of variation:

$$CV = \frac{\sigma}{\mu}$$

Coefficient of skewness:

$$C_s = \frac{\mathbb{E}(x - \mu)^3}{\sigma^3}$$

Note: When sample size is small, discrete approximation of the moments may be biased. To obtain **unbiased** estimates, the following formulas shall be used:

$$\sigma_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2$$

$$C_s = \frac{n \sum_{i=1}^n (x_i - \mu)^3}{(n-1)(n-2)\sigma_s^3}, \text{ where } C_s < 0 \text{ (negative skewness)} \quad C_s > 0 \text{ (positive skewness).}$$

In the above unbiased sample statistics $\mu = \left(\sum_{i=1}^n x_i \right) / n$.

Common PDFs in Hydrology I

Normal Distribution:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

$$F_X(x) = \int_{-\infty}^x f_X(u) du$$

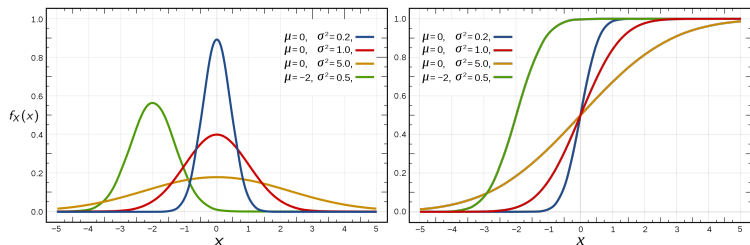


Figure 4: PDF (left) and CDF (right) of the Gaussian or Normal distribution.

$$\mathbb{E}(x) = \mu$$

$$\mathbb{E}(x - \mu)^2 = \sigma^2$$

$$\mathbb{E}(x - \mu)^3 = 0 \text{ and } \mathbb{E}(x - \mu)^4 = 3\sigma^4$$

Common PDFs in Hydrology II

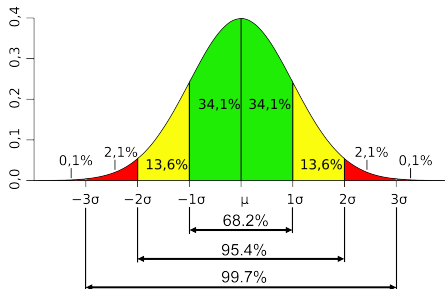


Figure 5: Distribution of the probability mass of a normal distribution based on different values of its standard deviation.

If we define the standard normal variable as $z = \frac{x-\mu}{\sigma}$, which is often called z-score, the distribution of z will have zero mean and a standard deviation equal to one. This distribution is called the **standard normal** distribution and has the following analytical form:

$$f_z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$$

The cumulative distribution function (CDF) of the standard normal is:

Common PDFs in Hydrology III

$$\Phi_z(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

In statistics, one often uses the so-called error function as follows:

$$\text{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^x e^{-u^2} du = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du$$

where, the complementary error function is defines as,

$$\text{erfc}(x) = 1 - \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-u^2} du$$

The values of the error function for different input values are given in the following table and can be obtained in MATALB using `erf(x)` function.

x	Hundredths digit of x									
	0	1	2	3	4	5	6	7	8	9
0.0	0.00000	0.01128	0.02256	0.03384	0.04511	0.05637	0.06762	0.07886	0.09008	0.10128
0.1	0.11246	0.12362	0.13476	0.14587	0.15695	0.16800	0.17901	0.18999	0.20094	0.21184
0.2	0.22270	0.23352	0.24430	0.25502	0.26570	0.27633	0.28690	0.29742	0.30788	0.31828
0.3	0.32863	0.33891	0.34913	0.35928	0.36936	0.37938	0.38933	0.39921	0.40901	0.41874
0.4	0.42839	0.43797	0.44747	0.45689	0.46623	0.47548	0.48466	0.49375	0.50275	0.51167
0.5	0.52050	0.52924	0.53790	0.54646	0.55494	0.56332	0.57162	0.57982	0.58792	0.59594
0.6	0.60386	0.61168	0.61941	0.62705	0.63459	0.64203	0.64938	0.65663	0.66378	0.67084

Figure 6: Values of the error function.

Common PDFs in Hydrology IV

We can use the error function to compute the cumulative distribution function of the standard normal density function $\Phi_z(z)$ as follows:

$$\Phi_z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{u^2}{2}} du, \quad \text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-u^2} du$$

If we do a change of variable as $u = \frac{t}{\sqrt{2}}$ and thus $du = \frac{dt}{\sqrt{2}}$, we have $u = 0 \Rightarrow t = 0$ and $u = z \Rightarrow t = \sqrt{2}z$. Applying this change of variable to the error function, we get

$$\begin{aligned} \text{erf}(z) &= \frac{2}{\sqrt{\pi}} \int_0^{\sqrt{2}z} e^{-\frac{t^2}{2}} \frac{dt}{\sqrt{2}} = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{2}z} e^{-\frac{t^2}{2}} dt \\ &= 2 \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\sqrt{2}z} e^{-\frac{t^2}{2}} dt - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 e^{-\frac{t^2}{2}} dt \right) \end{aligned}$$

As thus,

$$\text{erf}(z) = 2 \left(\Phi(\sqrt{2}z) - \phi(0) \right) = 2 \left(\Phi(\sqrt{2}z) - \frac{1}{2} \right)$$

which results in,

$$\Phi_z(\sqrt{2}z) = \frac{1 + \text{erf}(z)}{2}.$$

By another change of variable, we get

$$z \rightarrow \frac{z}{\sqrt{2}}$$

Common PDFs in Hydrology V

$$\Phi_z(z) = \frac{1 + \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)}{2}$$

Therefore if $x \sim \mathcal{N}(\mu, \sigma)$ and we are interested to compute $F_X(x) = p_r(X \leq x)$, we should follow the following steps:

$$z = \frac{x - \mu}{\sigma} \Rightarrow p_r(X \leq x) = \Phi_z(z) = \frac{1 + \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)}{2}$$

MATLAB: `normcdf(x, μ, σ)`

Example: Assume that X is from a normal distribution with $\mu = 2$ and $\sigma = 3$, what is the probability of $p_r(2.5 \leq x \leq 5)$?

$$x = 2.5 \Rightarrow z = \frac{2.5 - 2}{3} = 0.167$$

$$x = 5 \Rightarrow z = \frac{5 - 2}{3} = 1 \Rightarrow p_r(0.167 \leq z \leq 1) = \Phi(1) - \Phi(0.167)$$

$$\Phi(1) = \frac{1 + \operatorname{erf}\left(\frac{1}{\sqrt{2}}\right)}{2} = 0.8413$$

$$\Phi(0.167) = \frac{1 + \operatorname{erf}\left(\frac{0.167}{\sqrt{2}}\right)}{2} = 0.5663$$

$$p_r(0.167 \leq z \leq 1) = \Phi(1) - \Phi(0.167) \simeq 0.2752$$

Common PDFs in Hydrology VI

Log-normal Distribution:

Hydrologic variables are often skewed. A logarithm transformation often makes them more symmetric and allows a more robust estimation of their statistics. If a random variable $Y = \log X$ is normally distributed, then X has a log-normal distribution.

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu_y)^2}{2\sigma_y^2}\right) \quad x > 0$$

where μ_y and σ_y are the mean and standard deviation of Y .

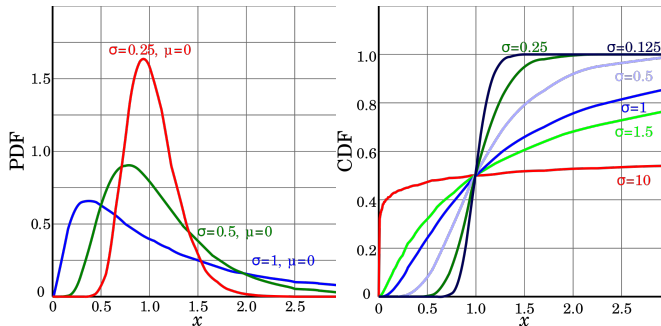


Figure 7: The PDF (left) and CDF (right) of the log-normal density function.

Common PDFs in Hydrology VII

Exponential distribution:

Some hydrologic process, such as the occurrence time between precipitation events can be explained by the an exponential distribution, which has the following expressions:

$$f_X(x) = \lambda e^{-\lambda x} \quad x \geq 0 \quad \lambda > 0$$

$$F_X(x) = 1 - e^{-\lambda x}$$

$$\mu_x = \frac{1}{\lambda} \quad \sigma_x^2 = \frac{1}{\lambda^2}.$$

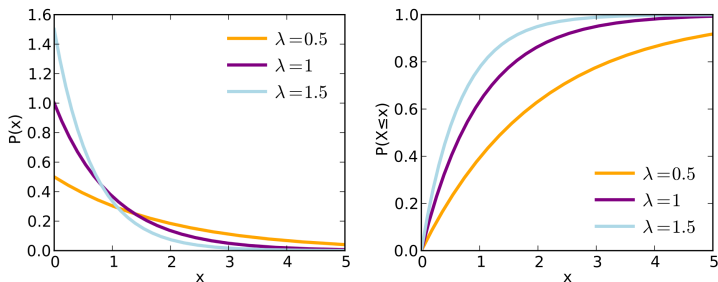


Figure 8: The PDF (left) and CDF (right) of an exponential density function.

Common PDFs in Hydrology VIII

Gamma distribution: A two parameter Gamma distribution is defined as follows:

$$f_X(x) = \frac{\lambda^\beta x^{\beta-1} e^{-\lambda x}}{\Gamma(\beta)} \quad \text{for } x \geq 0 \quad \text{and} \quad \lambda, \beta > 0$$

where λ is a width parameter and β catheterizes the shape. The following relationship holds between the parameters and the first and second order moments.

$$\lambda = \frac{\mu_x}{\sigma_x^2} \quad \text{and} \quad \beta = \frac{\mu_x^2}{\sigma_x^2}$$

The gamma function for positive integers and real numbers are $\Gamma(\alpha) = (\alpha - 1)!$ and $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$, respectively.

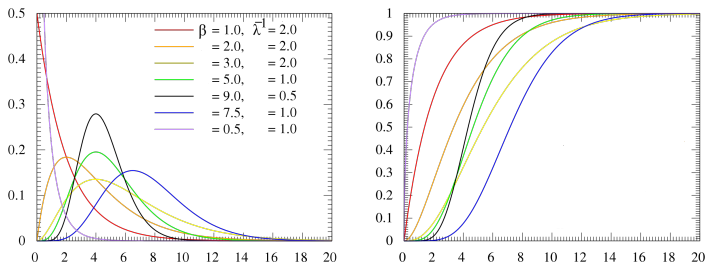


Figure 9: The PDF (left) and CDF (right) of the gamma density function.

Common PDFs in Hydrology IX

Gumbel distribution:

$$f(x, \beta) = \frac{1}{\beta} \exp \left[-\frac{x-u}{\beta} - \exp\left(-\frac{x-u}{\beta}\right) \right] \quad -\infty < x < \infty$$

$$u \simeq \mu - 0.5772\beta, \quad \beta = \frac{\sqrt{6}\sigma_x}{\pi}.$$

where μ and σ_x are the mean and standard deviation of x .

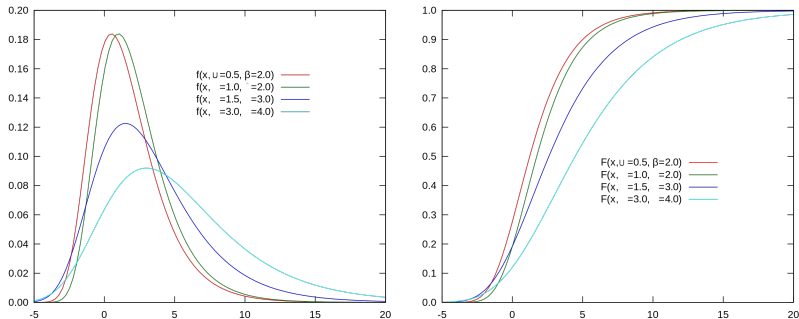


Figure 10: The PDF (left) and CDF (right) of the Gumbel distribution.

Common PDFs in Hydrology X

Pearson Type III:

This density is a three parameter gamma density function that has an extra location parameter ϵ .

$$f_X(x) = \frac{\lambda^\beta (x - \epsilon)^{\beta-1} e^{-\lambda(x-\epsilon)}}{\Gamma(\beta)} \quad x \geq \epsilon$$

$$\lambda = \frac{\sigma_x}{\sqrt{\beta}}, \quad \beta = \left(\frac{2}{C_s}\right)^2, \quad \epsilon = \mu - \sigma_x \sqrt{\beta}.$$

μ_x : mean of x

σ_x : standard deviation of x

C_s : skewness coefficient of x .

Log Pearson Type III:

$$f_X(x) = \frac{\lambda^\beta (y - \epsilon)^{\beta-1} e^{-\lambda(y-\epsilon)}}{x\Gamma(\beta)}; \quad \log x \geq \epsilon$$

$$\lambda = \frac{\sigma_y}{\sqrt{\beta}}; \quad \beta = \left(\frac{2}{C_s(y)}\right)^2; \quad \epsilon = \mu_y - \sigma_y \sqrt{\beta}$$

Fitting a probability distribution I

Method of Moments:

When we have a set of random samples, we are often interested to obtain a parametric representation for the distribution of those random numbers. To obtain that parametric representation, we first need to choose a PDF and then estimate its parameters such that the density represents well the observed samples. To that end, we can simply compute the sample statistics of the distribution and use them to estimate the parameters.

$$\bar{x} = \frac{\sum x_i}{n} \quad \sigma_x^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2 \quad C_s = \frac{n \sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)(n-2)\sigma_x^3}$$

For example for normal and exponential distribution we have:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{(x - \bar{x})^2}{2\sigma_x^2}\right) \quad \text{normal distribution}$$

$$f_X(x) = \lambda e^{-\lambda x} \quad \lambda = \frac{1}{\bar{x}}$$

Example: Assume that $x_i = \{2.4, 4.25, 0.77, 13.22, 3.55, 1.37\}$ are drawn from an exponential density function. What is the best estimate of the λ parameter based on the method of moment?

$$\bar{x} = \frac{2.4 + 4.25 + 0.77 + 13.22 + 3.55 + 1.37}{6} = 4.28$$

$\lambda = \frac{1}{\bar{x}} = 0.234$ and thus $f_X(x) = 0.234 e^{-0.234x}$.

However, the question is, [how can we choose the best probability density function?](#) In other words, we need a [goodness of fit test](#) to differentiate between different candidate PDFs and understand what is the best representative density function.

Fitting a probability distribution II

Chi-Squared test:

Let's assume that we have a set of n samples $\{x_j\}_{j=1}^n$ that are drawn from an unknown PDF. We can divide the observation domain into a set of intervals $[x_{i-1}, x_i]$, and compute the frequency of occurrence of x_i as follows $f_s(x_i) = \frac{n_i}{n}$, where

n : is the total number of observations and

n_i : is the number of observations that fall with $[x_{i-1}, x_i]$.

For a chosen probability model $f_X(x)$, we have $f(x_i) = F_X(x_i) - F_X(x_{i-1})$ where $F_X(\cdot)$ is the CDF function of the chosen probability model. If we assume that our histogram has m intervals, one may compute the following statistic that represents the deviation between the observed frequency of occurrence $nf_s(x_i)$ and expected ones $nf_X(x_i)$,

$$\chi_\kappa^2 = \sum_{i=1}^m \frac{[nf_s(x_i) - nf(x_i)]^2}{nf(x_i)} = \sum_{i=1}^m \frac{n[f_s(x_i) - f(x_i)]^2}{f(x_i)}$$

If we assume that the error $e_i = f(x_i) - p(x_i)$ is normally distributed then χ_κ^2 has a chi-squared distribution with degree of freedom $\kappa = m - p - 1$, where m is the number of intervals and p denotes the number fitted parameters for the chosen distribution.

Chi-squared distribution has the following density function:

$$f_\kappa(x) = \frac{1}{2^{\kappa/2} \Gamma(\frac{\kappa}{2})} x^{\frac{\kappa}{2}-1} e^{-\frac{x}{2}}$$
$$\mu = \kappa \quad \sigma_X^2 = 2\kappa \quad C_s = \sqrt{\frac{8}{\kappa}}$$

Fitting a probability distribution III

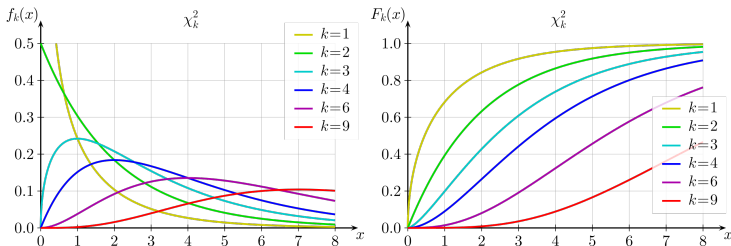


Figure 11: The PDF (left) and CDF (right) of the Chi-Squared distribution.

Example: We have a record of 69 years of annual precipitation data (inches) with the sample mean and standard deviation $\bar{x} = 39.77$ [in] and $\sigma = 9.17$ [in]. We have the null hypothesis that these annual precipitation data are not drawn from a Gaussian Distribution. Use the chi-squared test to accept or reject the null hypothesis.

Year	1910	1920	1930	1940	1950	1960	1970
0		48.7	44.8	49.3	31.2	46.0	33.9
1	39.9	44.1	34.0	44.2	27.0	44.3	31.7
2	31.0	42.8	45.6	41.7	37.0	37.8	31.5
3	42.3	48.4	37.3	30.8	46.8	29.6	59.6
4	42.1	34.2	43.7	53.6	26.9	35.1	50.5
5	41.1	32.4	41.8	34.5	25.4	49.7	38.6
6	28.7	46.4	41.1	50.3	23.0	36.6	43.4
7	16.8	38.9	31.2	43.8	56.5	32.5	28.7
8	34.1	37.3	35.2	21.6	43.4	61.7	32.0
9	56.4	50.6	35.1	47.1	41.3	47.4	51.8

Fitting a probability distribution IV

Interval (i)	Range (i)	n_i	$f_s(x_i)$	$F_s(x_i)$	z_i	$F(X < x_i)$	$f(x_i)$	χ^2_{res}
1	< 20	1	0.014	0.014	-2.157	0.015	0.015	0.004
2	20-25	2	0.029	0.043	-1.611	0.053	0.038	0.147
3	25-30	6	0.087	0.130	-1.065	0.144	0.090	0.008
4	30-35	14	0.203	0.333	-0.520	0.301	0.158	0.891
5	35-40	11	0.159	0.493	0.026	0.510	0.209	0.805
6	40-45	16	0.232	0.725	0.571	0.716	0.206	0.222
7	45-50	10	0.145	0.87	1.117	0.868	0.151	0.019
8	50-55	5	0.072	0.942	1.662	0.952	0.084	0.114
9	55-60	3	0.043	0.982	2.208	0.986	0.034	0.163
10	> 60	1	0.014	1.00	2.753	1.000	0.014	0.004
Total		69					1.00	2.377

For example in the above table, for the 4th interval, we have

$$f_s(x_4) = \frac{14}{69} = 0.203$$

$$F_s(x_4) = 0.014 + 0.029 + 0.087 + 0.203 = 0.333$$

Let's fit a Gaussian distribution with the following statistics:

$$\bar{x} = 39.77'' \quad \sigma = 9.17''$$

Having the above parameters for a Gaussian density, we can compute the probabilities of the third and fourth intervals as follows:

$$i = 4 \Rightarrow z_4 = \frac{X - \mu}{\sigma} = \frac{35 - 39.77}{9.17} = -0.52$$

$$\frac{1 + \text{erf}\left(\frac{z_4}{\sqrt{2}}\right)}{2} = \phi_z(z_4 = -0.52) = F(X_4 = 35) = 0.301$$

Fitting a probability distribution V

$$i = 3 \Rightarrow z_3 = \frac{30 - 29.77}{9.17} = -1.065$$

$$\phi_z(z_3 = -1.065) = F(X_3 = 30) = 0.144$$

$$f(x_4) = F(x_4) - F(x_3) = 0.158$$

And in the last column for $i = 4$, we have, $\frac{n[f_s(x_4) - f(x_4)]^2}{f(x_4)} = 0.81$ and the sum for all intervals is

$\chi_{\kappa}^2 = \sum_{i=1}^m \frac{n[f_s(x_i) - f(x_i)]^2}{f(x_i)} = 2.377$, where, $m = 10$, $p = 2$ and the degree of freedom $\kappa = m - p - 1 = 7$. The chi-squared statistic for significance level $\alpha = 0.95$ can be computed from available tables (e.g., <https://www.medcalc.org/manual/chi-square-table.php>) or existing software tools. In the problem at hand $x_{\alpha=0.95} = 14.1$ for $\kappa = 7$.

Note: if you have access to a computer program that can automatically compute $F(x_i)$, given the mean (29.77 in) and standard deviation ($\sigma=9.17$ in.) of the fitted normal distribution, you don't need to compute z_i . In MATLAB $f(x_i) = F(x_i) - F(x_{i-1}) = \text{normcdf}(x_i, \mu, \sigma) - \text{normcdf}(x_{i-1}, \mu, \sigma)$, where $\mu = 29.77$ [in] and $\sigma = 9.17$ [in] in the above example. Moreover the inverse of the Chi-Squared distribution can be obtained using the following command in MATLAB.

$$x_{\alpha} = 14.1 = \text{chi2inv}(0.95, 7)$$

In Excel the command is as follows:

$$14.1 = \text{chinv}(0.95, 7)$$

In MATLAB, the above Chi-squared test can be done with $\text{h} = \text{chi2gof}(x)$ in which for $h = 0$, we can reject the null hypothesis and assume that samples are drawn from the Gaussian density with significance level 95%.

Fitting a probability distribution VI

Here, the **null hypothesis** is that *the samples are not drawn from the normal distribution*. We have to check whether the hypothesis can be accepted or shall be rejected. It is important to note the null hypothesis is often referred to the case that there is no relationship between two measured phenomena, or no association exists between the sampled random variables and the chosen probability distribution.

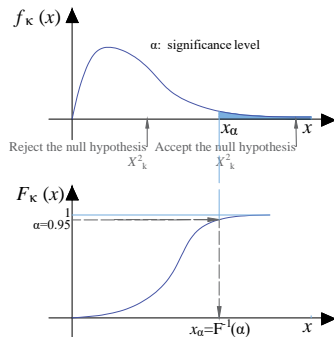


Figure 12: A schematic showing the Chi-Squared test for significance level $\alpha = 0.95$. When $\chi^2_{\kappa} \leq x_{\alpha}$, it is likely that the computed statistic is drawn from a chi-squared density. Thus, with significance level $\alpha = 0.95$, the samples are drawn from a Gaussian distribution and we can reject the null hypothesis. However, when $\chi^2_{\kappa} > x_{\alpha}$, the computed statistics may not belong to the chi-squared density and thus it is very likely that the samples are not drawn for the Gaussian density. Thus we can accept the null hypothesis with significance level $\alpha = 0.95$.

Frequency Analysis I

Frequency analysis in hydrology is for explaining extreme hydrologic events that can not be properly described via physical modeling. It is important to note that when we fit a probability density function to a finite number of samples, we assume that those samples are independent and identically distributed (i.i.d). For example, two samples of streamflow rate from today and tomorrow can be considered as two samples from the same distribution but it is very likely that they are not independent. **However, for example maximum annual streamflow rates can be considered as independent random variables.**

Extreme events: An extreme event is an event that has very low probability of occurrence and a long return period.

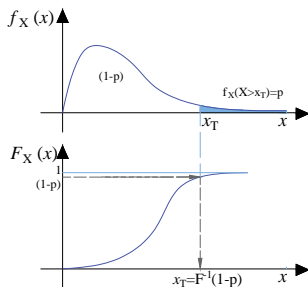


Figure 13: The extreme value x_T and return period T of an extreme event can be inferred from probability distribution of independent and identically distributed annual maxima of the event (e.g., streamflows) or block maxima of other variables of interest. Note that x_T is the $100(1 - p)$ th percentile of $f_X(x)$, where p denotes the probability of exceedance.

Frequency Analysis II

The following table represents the **annual daily maximum discharge in cfs** of a river station for almost five decades.

Year	1930	1940	1950	1960	1970
0		55900	13300	23700	9190
1		58000	12300	55800	9740
2		56000	28400	10800	58500
3		7710	11600	4100	33100
4		12300	8560	5720	25200
5	38500	22000	4950	15000	30200
6	179000	17900	1730	9790	14100
7	17200	46000	25300	70000	54500
8	25400	6970	58300	44300	12700
9	4940	26600	10100	15200	

For example assuming $x_T=50,000$ cfs, we can see that nine times the maximum annual flow exceeds this extreme event with the following recurrence intervals:

$$t_1 = 4, t_2 = 1, t_3 = 1, t_4 = 16, t_5 = 3, t_6 = 6, t_7 = 5, t_8 = 5 \text{ yr}$$

Therefore, the return period is $T = \mathbb{E}(t) = \Sigma t_i/n = \frac{41}{8} = 5.125$ year.

Let us assume that the probability of **exceedance** of an extreme event is denoted by p . If an extreme event occurs after t years, it means that there were $t - 1$ years without any extreme event. Therefore, the probability distribution of the return period maybe explained by a specific form of the binomial distribution as follows:

$$p_T(t) = (1 - p)^{t-1} p.$$

Frequency Analysis III

The expected values of t or the return period can be calculated as

$$\begin{aligned}T &= \mathbb{E}(t) = \sum_{t=1}^{\infty} t(1-p)^{t-1}p \\&= p + 2(1-p)p + 3(1-p)^2p + 4(1-p)^3p + \dots \\&= p[1 + 2(1-p) + 3(1-p)^2 + 4(1-p)^3 + \dots] \\&= \frac{p}{[1 - (1-p)]^2} = \frac{1}{p}\end{aligned}$$

Therefore, $T = \mathbb{E}(t) = \frac{1}{p}$, and thus the return period is equal to the inverse of the probability of exceedence of x_T ,

$$\frac{1}{T} = 1 - F(x_T) = p \quad \text{and thus} \quad x_T = F^{-1}(1-p).$$

Example: For an $x_T = 50000$ [cfs], the return period can be obtained as $T = \mathbb{E}(t) = 5.1$ yr, which results in $p = \frac{1}{T} = 0.195$.

Question: What is the probability that a T -year return period event will occur at least once in N years?

$$P_r(X < x_T \text{ for } N \text{ consecutive years}) = (1-p)(1-p) \dots = (1-p)^N$$

$$P_r(X \geq x_T \text{ at least once in } N \text{ years}) = 1 - (1-p)^N = 1 - \left(1 - \frac{1}{T}\right)^N$$

Frequency Analysis IV

Example: Estimate the probability that annual maximum discharge Q will exceed 50000 cfs at least once during the next 5 years.

$$p = \frac{1}{T} = \frac{1}{5.1} = 0.195$$

$$p(Q \geq 50000) \text{ cfs at least once in the next five years} = 1 - (1 - 0.195)^5 \simeq 0.66.$$

Extreme Value Distributions:

Two classes of distributions are commonly used in hydrologic extreme value analyses:

- (1) Extreme value type I,II,III
- (2) Log-Pearson Type III

Extreme value type I or the Gumbel distribution is often used for rainfall frequency analysis as follows:

$$f_X(x) = \frac{1}{\beta} \exp\left[-\frac{x-u}{\beta} - \exp\left(-\frac{x-u}{\beta}\right)\right] \quad -\infty < u < \infty$$

$$u = \mu - 0.5772\beta$$

$$\beta = \frac{\sqrt{6}\sigma_x}{\pi}$$

$$F_X(x) = \exp\left[-\exp\left(-\frac{x-u}{\beta}\right)\right] \quad -\infty < u < \infty$$

if we define a reduced variable as

$$y = \frac{x-u}{\beta},$$

Frequency Analysis V

we have:

$$F(X) = \exp[-\exp(-y)] \Rightarrow y = -\ln\left[\ln\left(\frac{1}{F_X(x)}\right)\right] \quad (1),$$

we showed that,

$$\begin{aligned} \frac{1}{T} = p &= \int_{x_T}^{\infty} f_X(x) dx = 1 - \int_{\infty}^{x_T} f_X(x) dx \\ &= 1 - F(x_T) \Rightarrow F(x_T) = \frac{T-1}{T}, \quad (2) \end{aligned}$$

from (1) & (2) we have,

$$\begin{aligned} y_T &= -\ln\left[\ln\left(\frac{T}{T-1}\right)\right] \\ x_T &= u + \beta y_T. \end{aligned}$$

Therefore, given the return period T , we can compute y_T and knowing the distribution parameters from the method of moment, we can obtain the associated extreme event x_T .

Frequency Analysis VI

Example: Annual maximum value of 10-minute-duration rainfall in a specific location is presented below. Calculate the 10-minute-duration maximum rainfall for 5- and 10-year return period.

Year	1910	1920	1930	1940
0		0.53	0.33	0.34
1		0.76	0.96	0.70
2		0.57	0.94	0.57
3	0.49	0.80	0.80	0.92
4	0.66	0.66	0.62	0.66
5	0.58	0.68	0.71	0.65
6	0.58	0.68	1.11	0.63
7	0.41	0.61	0.64	0.60
8	0.47	0.88	0.52	
9	0.74	0.49	0.64	

$$\mu = 0.649 \quad \sigma_x = 0.177 \text{ in} \quad \beta = \frac{\sqrt{6}\sigma_x}{\pi} = 0.138$$

$$u = \mu - 0.5772\beta = 0.569$$

$$y_T = -\ln \left[\ln \left(\frac{T}{T-1} \right) \right]$$

$$y_T = -\ln \left[\ln \left(\frac{5}{4} \right) \right] = 1.5$$

$$x_T = u + \beta y_T = 0.569 + 0.138 \times 1.50 = 0.78''$$

Frequency Analysis VII

Log-Pearson Type III is used for flood frequency in the United States. Recall that as the Gumbel distribution was invertible, we have closed form expression for its percentile and thus could find an analytical expression for the extreme values of annual rainfall maxima (x_T for return period T as follows:

$$f(X \geq x_T) = p \Rightarrow F_X^{-1}(1 - p) = x_T$$

However, the CDF of many distributions is not invertible. For computational convenience, we use a **frequency factor** K_T and express the percentile of the distribution for return period T as follows:

$$x_T = \mu_x + K_T \sigma_x,$$

where μ_x and σ_x are the mean and standard deviation of the random variable x . When the random variable has significant positive skewness, we take $y = \log x$ and then we have

$$y_T = \mu_y + K_T \sigma_y.$$

Finding y_T , then one can obtain $x_T = 10^{y_T}$. For complex distributions such as the log-Pearson type III, the relationship between K_T and T is given in some pre-calculated tables for the log-transformed variable. In the tables shown in the following slides, the frequency factor K_T is given for return period $T = 2, 5, 10, 25, 50, 100, 200$ yrs as a function of the sample skewness coefficient of the data (C_s) in the logarithm scale.

Frequency Analysis VIII

	K_T values for Pearson Type III distribution (positive skew)						
	Return period in years						
	2	5	10	25	50	100	200
Skew Coefficient C_S or C_W	Exceedence Probability						
	0.50	0.20	0.10	0.04	0.02	0.01	0.005
3.0	-0.396	0.420	1.180	2.278	3.152	4.051	4.970
2.9	-0.390	0.440	1.195	2.277	3.134	4.013	4.909
2.8	-0.384	0.460	1.210	2.275	3.114	3.973	4.847
2.7	-0.376	0.479	1.224	2.272	3.093	3.932	4.783
2.6	-0.368	0.499	1.238	2.267	3.071	3.889	4.718
2.5	-0.360	0.518	1.250	2.262	3.048	3.845	4.652
2.4	-0.351	0.537	1.262	2.256	3.023	3.800	4.584
2.3	-0.341	0.555	1.274	2.248	2.997	3.753	4.515
2.2	-0.330	0.574	1.284	2.240	2.970	3.705	4.444
2.1	-0.319	0.592	1.294	2.230	2.942	3.656	4.372
2.0	-0.307	0.609	1.302	2.219	2.912	3.605	4.298
1.9	-0.294	0.627	1.310	2.207	2.881	3.553	4.223
1.8	-0.282	0.643	1.318	2.193	2.848	3.499	4.147
1.7	-0.268	0.660	1.324	2.179	2.815	3.444	4.069
1.6	-0.254	0.675	1.329	2.163	2.780	3.388	3.990
1.5	-0.240	0.690	1.333	2.146	2.743	3.330	3.910
1.4	-0.225	0.705	1.337	2.128	2.706	3.271	3.828
1.3	-0.210	0.719	1.339	2.108	2.666	3.211	3.745
1.2	-0.195	0.732	1.340	2.087	2.626	3.149	3.661
1.1	-0.180	0.745	1.341	2.066	2.585	3.087	3.575
1.0	-0.164	0.758	1.340	2.043	2.542	3.022	3.489
0.9	-0.148	0.769	1.339	2.018	2.498	2.957	3.401
0.8	-0.132	0.780	1.336	1.993	2.453	2.891	3.312
0.7	-0.116	0.790	1.333	1.967	2.407	2.824	3.223
0.6	-0.099	0.800	1.328	1.939	2.359	2.755	3.132
0.5	-0.083	0.808	1.323	1.910	2.311	2.686	3.041
0.4	-0.066	0.816	1.317	1.880	2.261	2.615	2.949
0.3	-0.05	0.824	1.309	1.849	2.211	2.544	2.856
0.2	-0.033	0.830	1.301	1.818	2.159	2.472	2.763
0.1	-0.017	0.836	1.292	1.785	2.017	2.400	2.670
0.0	0	0.842	1.282	1.751	2.054	2.326	2.576

Frequency Analysis IX

	K_T values for Pearson Type III distribution (positive skew)						
	Return period in years						
	2	5	10	25	50	100	200
Skew Coefficient C_S or C_W	Exceedence Probability						
	0.50	0.20	0.10	0.04	0.02	0.01	0.005
-0.1	0.017	0.846	1.270	1.716	2.000	2.252	2.482
-0.2	0.033	0.850	1.258	1.680	1.945	2.178	2.388
-0.3	0.050	0.853	1.245	1.643	1.890	2.104	2.294
-0.4	0.066	0.855	1.231	1.606	1.834	2.029	2.201
-0.5	0.083	0.856	1.216	1.567	1.777	1.955	2.108
-0.6	0.099	0.857	1.200	1.528	1.720	1.880	2.106
-0.7	0.116	0.857	1.183	1.488	1.663	1.806	1.926
-0.8	0.132	0.856	1.166	1.448	1.606	1.733	1.837
-0.9	0.148	0.854	1.147	1.407	1.549	1.660	1.749
-1.0	0.164	0.852	1.128	1.366	1.492	1.588	1.664
-1.1	0.180	0.848	1.107	1.324	1.435	1.518	1.581
-1.2	0.195	0.844	1.086	1.282	1.379	1.449	1.501
-1.3	0.210	0.838	1.064	1.240	1.324	1.383	1.424
-1.4	0.225	0.832	1.041	1.198	1.270	1.318	1.351
-1.5	0.240	0.825	1.018	1.157	1.217	1.256	1.282
-1.6	0.254	0.817	0.994	1.166	1.166	1.197	1.216
-1.7	0.268	0.808	0.970	1.075	1.116	1.140	1.155
-1.8	0.282	0.799	0.945	1.035	1.069	1.087	1.097
-1.9	0.294	0.788	0.920	0.996	1.023	1.037	1.044
-2.0	0.307	0.777	0.895	0.959	0.980	0.990	0.995
-2.1	0.319	0.765	0.869	0.923	0.939	0.946	0.949
-2.2	0.330	0.752	0.844	0.888	0.900	0.905	0.907
-2.3	0.341	0.739	0.819	0.855	0.864	0.867	0.869
-2.4	0.351	0.725	0.795	0.823	0.830	0.832	0.833
-2.5	0.360	0.711	0.771	0.793	0.798	0.799	0.800
-2.6	0.368	0.696	0.747	0.764	0.768	0.769	0.769
-2.7	0.376	0.681	0.724	0.738	0.740	0.740	0.741
-2.8	0.384	0.666	0.702	0.712	0.714	0.714	0.714
-2.9	0.390	0.651	0.681	0.683	0.689	0.690	0.690
-3.0	0.396	0.636	0.666	0.666	0.666	0.667	0.667

Frequency Analysis X

Example: Sixteen years of the annual maxima of a river flow is given in the following table. Calculate the 50-year return period of annual maximum discharge using the log-Pearson type III distribution.

Year	1930	1940	1950	1960	1970
0		55900	13300	23700	9190
1		58000	12300	55800	9740
2		56000	28400	10800	58500
3		7710	11600	4100	33100
4		12300	8560	5720	25200
5	38500	22000	4950	15000	30200
6	179000	17900	1730	9790	14100
7	17200	46000	25300	70000	54500
8	25400	6970	58300	44300	12700
9	4940	26600	10100	15200	

The statistics of the annual maxima in the log-scale ($y = \log x$) are as follows:

$$\mu_y = \frac{\sum y_i}{n} = 4.27 \quad \sigma_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}} = 0.4027 \quad C_s = \frac{n \sum_{i=1}^n (y_i - \bar{y})^3}{(n-1)(n-2)\sigma_y^3} = -0.0696$$

For $C_s = -0.0696$ from the table of the frequency factors, we have,

$$K_{T=50} \simeq 2.054 + \frac{2.00-2.054}{(-1.0-0)} \times (-0.0696 - 0) = 2.016$$

Therefore, $y_{50} = \bar{y} + K_{50}\sigma_y = 4.2743 + 2.016 \times 0.4027 = 5.0863$.

Thus, $x_{50} = 10^{5.0863} = 121,990$ [cfs].

Water Resources Council Method I

The values of frequency factors K_T are very sensitive to the coefficients of skewness C_s . To avoid over estimation or underestimation of floods and financial consequences the Water Resources Council recommends a method that leads to a more accurate and robust estimate of the coefficient of skewness (C_w) as follows:

$$C_w = wC_s + (1 - w)C_m$$

where

C_s : is the sample skewness

C_m denotes a map skewness,

C_m denotes a weighted skewness (a more robust estimate),

w is an optimal weight that interpolates between these two quantities

Assuming that C_s and C_m are two independent variables, the optimal weight is defined such that it minimizes the variance of estimate of C_w as follows:

$$\text{Var}(C_w) = w^2\text{Var}(C_s) + (1 - w)^2\text{Var}(C_m)$$

To obtain the minimizer of $\text{Var}(C_w)$ we set its derivative to zero $\frac{d}{dw}[\text{Var}(C_w)] = 0$ which leads to optimal value of

$$w = \frac{\text{Var}(C_m)}{\text{Var}(C_s) + \text{Var}(C_m)}.$$

This value of the w is a minimizer of the variance of C_w as we can show that $\frac{d^2}{dw^2}[\text{Var}(C_w)] \geq 0$.

Water Resources Council Method II

Therefore, a robust estimate of the skewness coefficient is:

$$C_w = \frac{\text{Var}(C_m)C_s + \text{Var}(C_s)C_m}{\text{Var}(C_m) + \text{Var}(C_s)}$$

The water resources council recommends to consider $\text{Var}(C_m) = 0.3025$ and proposes a regional map for C_m .

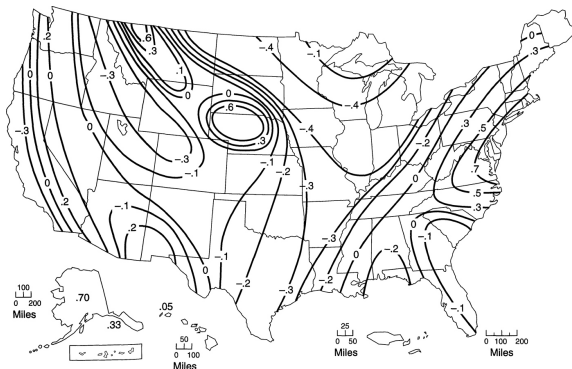


Figure 14: Generalized map skewness coefficients (C_m) of annual maximum streamflow from U.S. Water Resources Council (1981).

Water Resources Council Method III

For the Log-Pearson type III distribution and n years of annual maximum streamflow, the variance of the sample skewness can be obtained from the following formula:

$$\text{Var}(C_s) = 10^{A-B \log_{10}(\frac{n}{10})}$$

$$A = \begin{cases} -0.33 + 0.08|C_s| & |C_s| \leq 0.90 \\ -0.52 + 0.30|C_s| & |C_s| > 0.90 \end{cases} \quad B = \begin{cases} 0.94 - 0.26|C_s| & |C_s| \leq 1.50 \\ 0.55 & |C_s| > 1.50 \end{cases}$$

Example: Determine the flood with return period of 100 years from the following maximum annual streamflow data near Austin, TX, where $C_m = -0.3$.

Year	Annual maximum flow (x) [cfs]	y=log x
1967	303	2.48
1968	5640	3.75
1969	1050	3.02
1970	6020	3.78
1971	3740	3.57
1972	4580	3.66
1973	5140	3.71
1974	10560	4.02
1975	12840	4.10
1976	5140	3.71
1977	2520	3.40
1978	1730	3.23
1979	12400	4.09
1980	3400	3.53
1981	14300	4.15
1982	9540	3.97

Water Resources Council Method IV

Taking the logarithm of the data $y = \log x$, the statistics of y are,

$$\mu_y = 3.6388 \quad \sigma_y = \frac{\sum (y_i - \bar{y})^2}{n-1} = 0.4439 \quad C_s = \frac{n \sum_{i=1}^n (y_i - \bar{y})^3}{(n-1)(n-2)\sigma_y^3} = -1.244.$$

$$|C_s| > 0.9 \Rightarrow A = -0.52 + 0.3|C_s| = -0.52 + 0.3 \times 1.244 = -0.147.$$

$$|C_s| < 1.5 \Rightarrow B = 0.94 - 0.26|C_s| = 0.617.$$

Thus, we have $\text{Var}(C_s) = 10^{-0.147 - 0.617 \log_{10}(\frac{16}{10})} = 0.533$ and $\text{Var}(C_m) = 0.303$.

The optimal weight can be calculated as follows:

$$w = \frac{\text{Var}(C_m)}{\text{Var}(C_s) + \text{Var}(C_m)} = \frac{0.303}{0.303 + 0.533} = 0.362.$$

Then a robust estimate of the coefficient of skewness is

$$C_w = wC_s + (1-w)C_m = 0.362 \times (-1.244) + 0.638 \times (-0.3) = -0.64.$$

From the table of the frequency factors for the Log-Pearson Type III distribution, for negative skewness and $T = 100$ years, we have

$$C_w = -0.6 \rightarrow K_T = 1.88$$

$$C_w = -0.7 \rightarrow K_T = 1.806.$$

By linear interpolation, one can obtain $C_w = -0.64$ and calculate the frequency factor $K_T = 1.850$, which leads to

$$y_T = \mu_y + K_T \sigma_y = 3.639 + 1.85 \times 0.4439 = 4.46 \rightarrow Q_T = 10^{4.46} = 28,900 \text{ [cfs]}.$$

Design Storms I

As we discussed, we collect maximum precipitation for a specific duration (e.g., maximum of 10-minute duration rainfall) for each year and then use Extreme value type I distribution to obtain extreme precipitation events (x_T) for a specific return period.

This calculation is done by the National Weather Service and is published under the name of *Rainfall Frequency Atlas of the United States*", [Technical paper No.4](#). These maps are available for durations of 30 minutes to 24 hours and return period of 1 to 100 years (see Figure 15).

As a result based on the project design specifications, the design engineer shall interpolated between isohyets, which are lines with equal values of extreme precipitation. For shorter durations of 5 to 60 minutes.

There is another report called NOAA Technical document [Hydro-35](#). The maps of precipitation depths from 5-, 15- and 60-minute durations and return period of 2 and 100 years for 37 eastern states are available in this technical paper.

Currently, there is a Precipitation Frequency Data Server (PFDs), that can be used for obtaining the values of design precipitation for different return periods hdsc.nws.noaa.gov/hdsc/pfds almost all over the United States.

Design Storms II

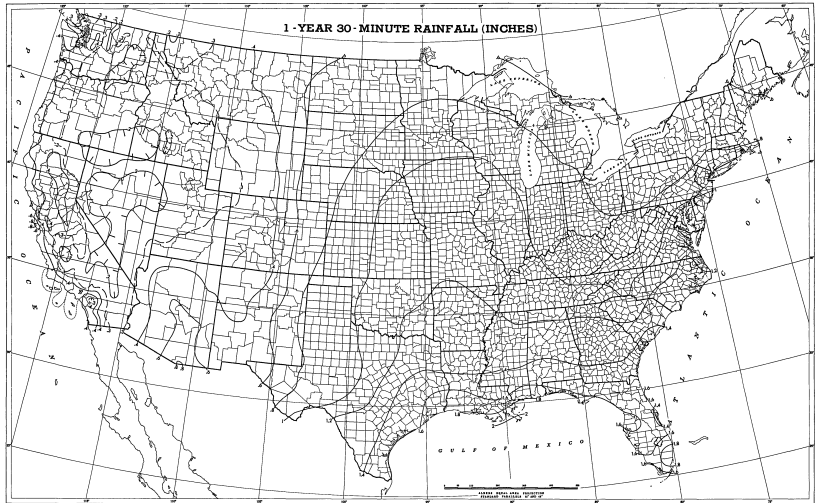


Figure 15: 1-year 30 minute rainfall (in) in the United States as presented in U.S. Weather Bureau technical paper 40.

Design Storms III

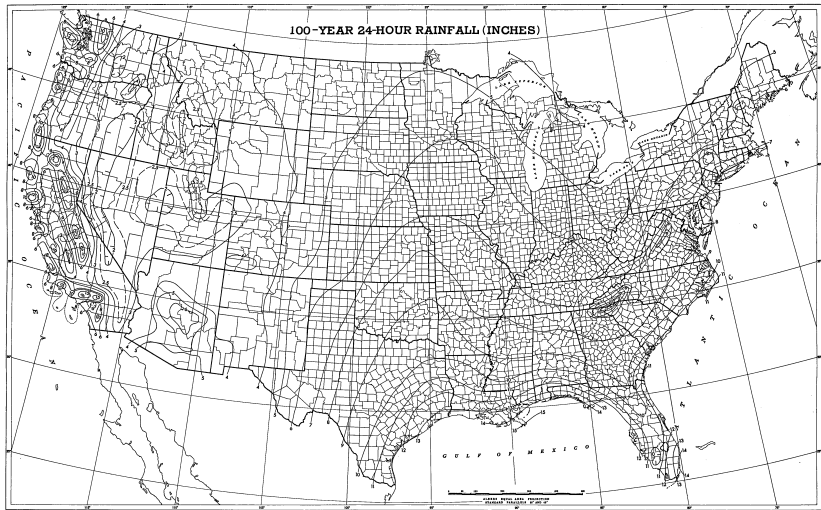


Figure 16: 100-year 24-hour rainfall (in) in the United States as presented in U.S. Weather Bureau technical paper 40.

Design Storms IV

Intensity-Duration-Frequency (IDF) curves

Intensity of precipitation is depth per unit time (mm/hr)

$$i = \frac{P}{T_d}$$

where p is the rainfall depth (mm, in) and T_d is the duration, usually in hours. The frequency is often expressed in terms of the return period T , which is the average length of time between precipitation events.

We discussed that using extreme value Type I distribution, we can compute extreme precipitation values for a specific duration (e.g. 10-minute) for different return periods (x_T). The explained calculation can be simply represented in a plot called Intensity-Duration-Frequency (IDF) curves.

For specific return periods, the IDF curves are typically represented by the following parametric equation:

$$i = \frac{c}{T_d^e + f} \quad (\text{in/hr}),$$

where T_d is the rainfall duration and e , c , f are empirical coefficients that can be obtained from historical observations of rainfall.

For example, for a 10-year return period and duration $T_d = 20$ minutes, these coefficients over Atlanta are $c = 97.5$, $e = 0.83$ and $f = 6.88$. Therefore, a 20-minute design rainfall with a 10-year return period in Atlanta is:

$$i = \frac{97.5}{20^{0.83} + 6.88} = 5.15 \quad [\text{in/hr}].$$

Design Storms V

Design precipitation hyetographs from IDF curves:

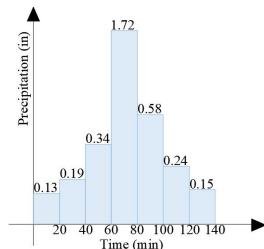
There are several methods to obtain the design hyetograph from IDF curves. Here, we explain the alternating block method through an example.

Example: Determine the design precipitation hyetograph, with 20 minutes intervals, for a $2\frac{1}{3}$ hour storm in Atlanta using the IDF curve $i = c/(T_d^e + f)$, where $c=97.5$, $e=0.83$ and $f=6.88$.

From this IDF curve, we can compute the rainfall intensities over 20 minutes intervals as follows:

Duration (min)	Intensity (in/hr)	Cumulative depth (in.)	Incremental depth (in.)
20	5.1591	$5.1591*20/60=1.72$	1.72
40	3.45	$3.45*40/60=2.3$	$0.58=2.3-1.72$
60	2.64	2.64	$0.34=2.64-2.30$
80	2.17	2.89	0.24
100	1.85	3.09	0.19
120	1.62	3.24	0.15
140	1.44	3.37	0.13

To shape the design hyetograph, we alternate incremental depths around their maximum in a descending order as shown in the right hand side. As is evident, this derivation is not unique depending on whether we start the alternation from the left or right hand side of the maximum incremental depth.



Design Storms VI

Rational Method:

Rational method, is a simple method for estimating the basin outflows based on different land use and rainfall return period. In this method, the rate of the outflow of a drainage basin is determined as follows:

$$Q = c i A$$

Q : outflow [cfs]

c : runoff coefficient $0 \leq c \leq 1$

i : rainfall intensity [in/hr]

A : area of the basin in acres [43560 ft^2]

We can divide the basin into m smaller sub-basins with more uniform land use and then obtain the outflow based on the following equation:

$$Q = i \sum_{j=1}^m c_j A_j$$

The runoff coefficient (c_j) for different land use and rainfall return period may be obtained from the following table.

Design Storms VII

Runoff Coefficients for Use in the Rational Method

Character of surface	Return period (years)						
	2	5	10	25	50	100	500
Developed							
Asphaltic	0.73	0.77	0.81	0.86	0.90	0.95	1.00
Concrete/roof	0.75	0.80	0.83	0.88	0.92	0.97	1.00
Grass areas (lawns, parks, etc.)							
<i>Poor condition</i> (grass cover less than 50% of the area)							
Flat, 0–2%	0.32	0.34	0.37	0.40	0.44	0.47	0.58
Average, 2–7%	0.37	0.40	0.43	0.46	0.49	0.53	0.61
Steep, over 7%	0.40	0.43	0.45	0.49	0.52	0.55	0.62
<i>Fair condition</i> (grass cover 50% to 75% of the area)							
Flat, 0–2%	0.25	0.28	0.30	0.34	0.37	0.41	0.53
Average, 2–7%	0.33	0.36	0.38	0.42	0.45	0.49	0.58
Steep, over 7%	0.37	0.40	0.42	0.46	0.49	0.53	0.60
<i>Good condition</i> (grass cover larger than 75% of the area)							
Flat, 0–2%	0.21	0.23	0.25	0.29	0.32	0.36	0.49
Average, 2–7%	0.29	0.32	0.35	0.39	0.42	0.46	0.56
Steep, over 7%	0.34	0.37	0.40	0.44	0.47	0.51	0.58
Undeveloped							
Cultivated land							
Flat, 0–2%	0.31	0.34	0.36	0.40	0.43	0.47	0.57
Average, 2–7%	0.35	0.38	0.41	0.44	0.48	0.51	0.60
Steep, over 7%	0.39	0.42	0.44	0.48	0.51	0.54	0.61
Pasture/range							
Flat, 0–2%	0.25	0.28	0.30	0.34	0.37	0.41	0.53
Average, 2–7%	0.33	0.36	0.38	0.42	0.45	0.49	0.58
Steep, over 7%	0.37	0.40	0.42	0.46	0.49	0.53	0.60
Forest/woodlands							
Flat, 0–2%	0.20	0.25	0.28	0.31	0.35	0.39	0.48
Average, 2–7%	0.31	0.34	0.36	0.40	0.43	0.47	0.56
Steep, over 7%	0.35	0.39	0.41	0.45	0.48	0.52	0.58

Note: The values in the table are the standards used by the City of Austin, Texas.

Source: Chow, Maidment, and Mays (1988).

Figure 17: Values of the runoff coefficients in the rational method.

Design Storms VIII

Rational method is widely used in the design of storm sewer systems. To that end, we first need to define the design storm in terms of its total depth and duration. The total depth of the design rainfall can be obtained from NOAA rainfall design maps (e.g., see Figure 14) or regional IDF curves for calculation the design intensity $i = p/T_d$. The design duration (T_d) is the time of the concentration of the basin that drains water into the sewer system,

$$T_d = t_c = \sum_{i=1}^n \frac{L_i}{V_i}.$$

Example: Determine the required pipe diameter for a storm sewer drainage system for a $T = 5$ yr return period, where $A = 5$ acres, $c = 0.6$ and $t_c = T_d = 10$ minutes. The design precipitation is give by following IDF equation,

$$i = \frac{100T^{0.2}}{T_d + 30} = 4.14 \text{ in/hr,}$$

$$Q = c i A = 0.6 \times 4.14 \times 5 = 12.42 \text{ cfs.}$$

Elevation of the basin changes 3 ft in 500 ft of length. Thus the slope is $S_0 = \frac{3}{500} = 0.006$ and the manning's coefficient is $n = 0.02$. Using the Manning equation, we have

$$Q = \frac{1.49}{n} S_0^{1/2} A R^{2/3}$$

$$Q = \frac{1.49}{n} S_0^{1/2} \frac{\pi D^2}{4} \left(\frac{D}{4}\right)^{2/3} = \frac{0.463}{n} S_0^{1/2} D^{8/3}$$

$$\text{and thus } D = \left(\frac{2.16 n Q}{\sqrt{S_0}}\right)^{3/8} = \left(\frac{2.16 \times 0.02 \times 12.42}{\sqrt{0.006}}\right)^{3/8} = 2 \text{ [ft].}$$