Probability Distribution

Prof. (Dr.) Rajib Kumar Bhattacharjya
Indian Institute of Technology Guwahati
Guwahati, Assam
Email: rkbc@iitg.ernet.in Web: www.iitg.ernet.in/rkbc
Visiting Faculty NIT Meghalaya

Probabilistic Approach:

A random variable X is a variable described by a probability distribution.

- The distribution specifies the chance that an observation x of the variable will fall in a specific range of X.
- A set of observations $n_1, n_2, n_3, \ldots\ldots\ldots\ldots, n_n$ of the random variable is called a sample.
- It is assumed that samples are drawn from a hypothetical infinite population possessing constant statistical properties.
- Properties of sample may vary from one sample to others.

The probability P (A) of an event is the chance that it will occur when an observation of the random variable is made.

$$P(A) = \lim_{n \to \infty} \left(\frac{n_A}{n} \right)$$

n_A --- number in range of event A.

n----- Total observations

1. Total Probability

$$P(A_1) + P(A_2) + P(A_3) + \ldots \ldots \ldots \ldots \ldots \ldots + P(A_n) = P(\Omega) = 1$$

1. Complementarity

$$P(\overline{A}) = 1 - P(A)$$

1. Conditional Probability

$$P\left(\frac{B}{A}\right) = \frac{P(A \cap B)}{P(A)}$$

B will occur provided A has already occurred.

Joint Probability

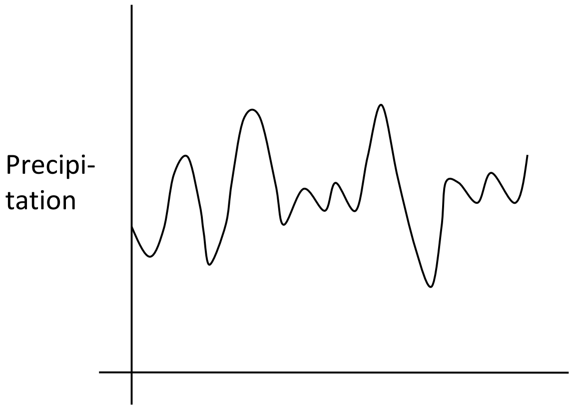
$$P(A \cap B) = P(A)P(B)$$

Example:

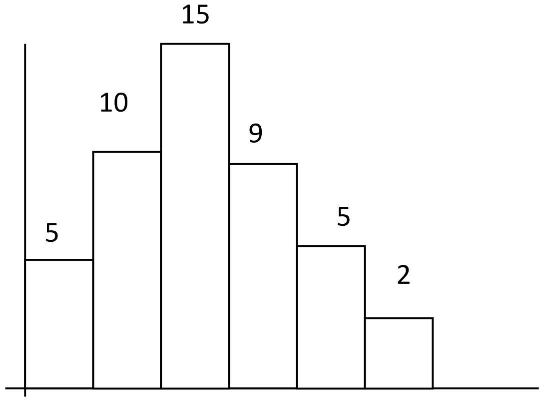
The probability that annual precipitation will be less than 120 mm is 0.333. What is the probability that there will be two successive year of precipitation less than 120 mm.

$P(R<35) = 0.333$

$P(C) = 0.333^2 = 0.111$



Year



Frequency Histogram

Relative Frequency:

$$f_s(x_i) = \frac{n_i}{n}$$

Which is equivalent to $P(x_i - \Delta x \leq X \leq x_i)$, the probability that the random variable X will lie in the interval $[x_i - \Delta x, x_i]$

Cumulative Frequency Function:

$$F_s(x_i) = \sum_{j=1}^{i} f_s\left(x_j\right)$$

This is estimated as $P(X \leq x_i)$, the cumulative probability of x_i.
This is estimated for sample data, corresponding function for population will be

Probability density Function:

$$f(x) = \lim_{\substack{n\to\infty \\ \Delta x\to 0}} \frac{f_s(x)}{\Delta x}$$

Probability Distribution Function:

$$F(x) = \lim_{\substack{n\to\infty \\ \Delta x\to 0}} F_s(x)$$

Whose derivative is the probability density function

$$f(\gamma) = \frac{dF(\gamma)}{dx}$$

The cumulative probability as $P(X \leq x)$, can be expressed as

$$P(X \leq x) = F(x) = \int_{-\infty}^{x} f(u)du$$

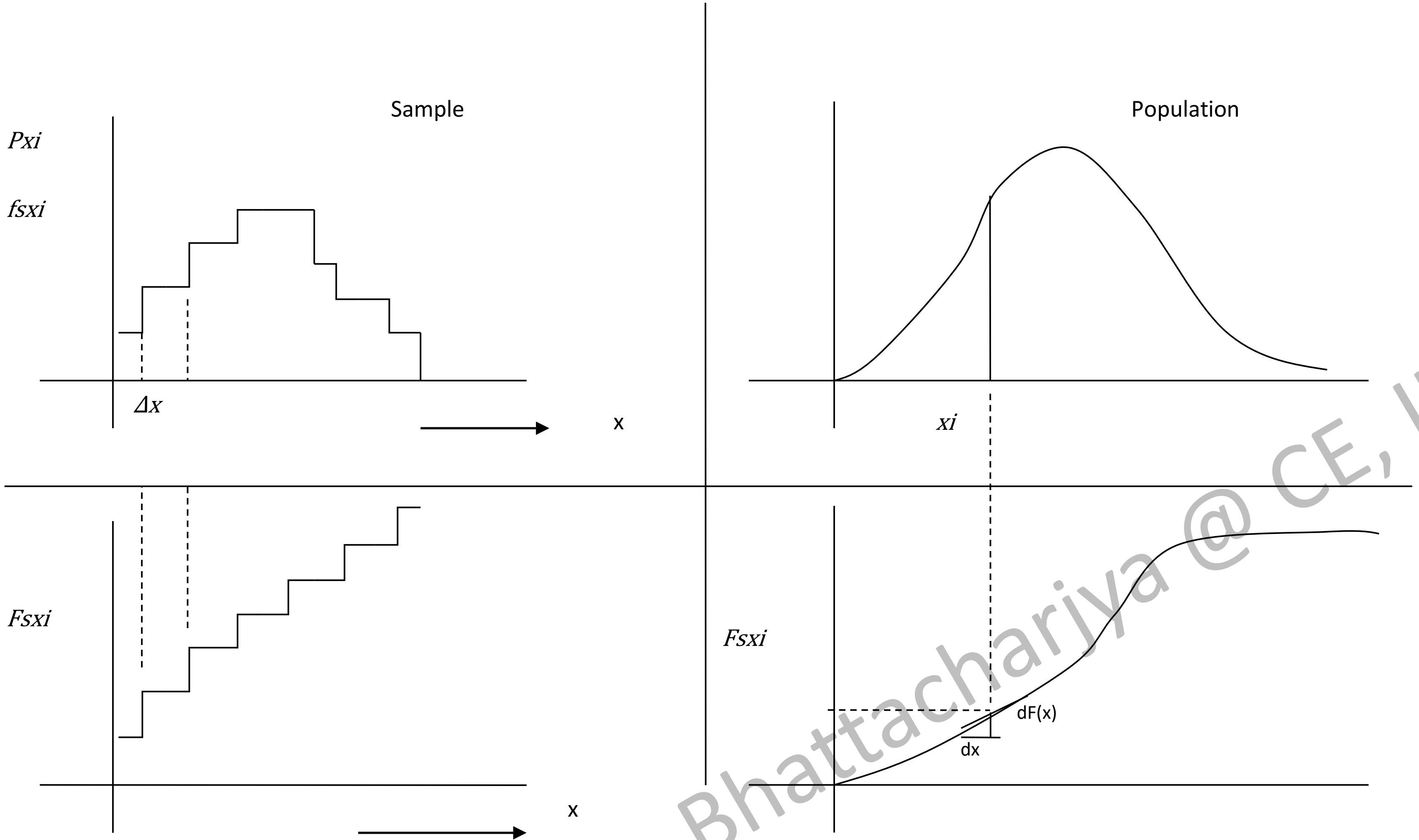
$$P(x_i) = P(x_i - \Delta x \leq X \leq x_i)$$

$$= \int_{x_i-\Delta x}^{x_i} f(x)dx$$

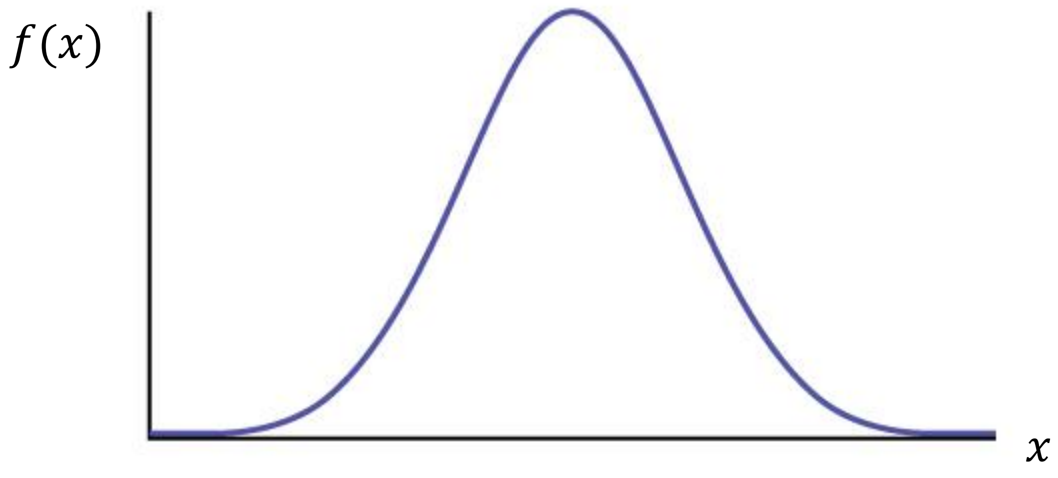
$$= \int_{-\infty}^{x_i} f(x)dx - \int_{-\infty}^{x_i-\Delta x} f(x)dx$$

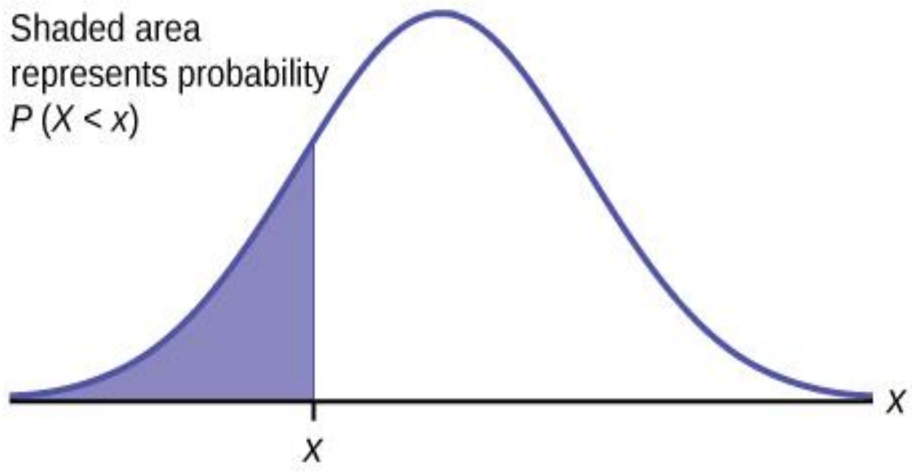
$$=F(x_i) - F(x_i - \Delta x)$$

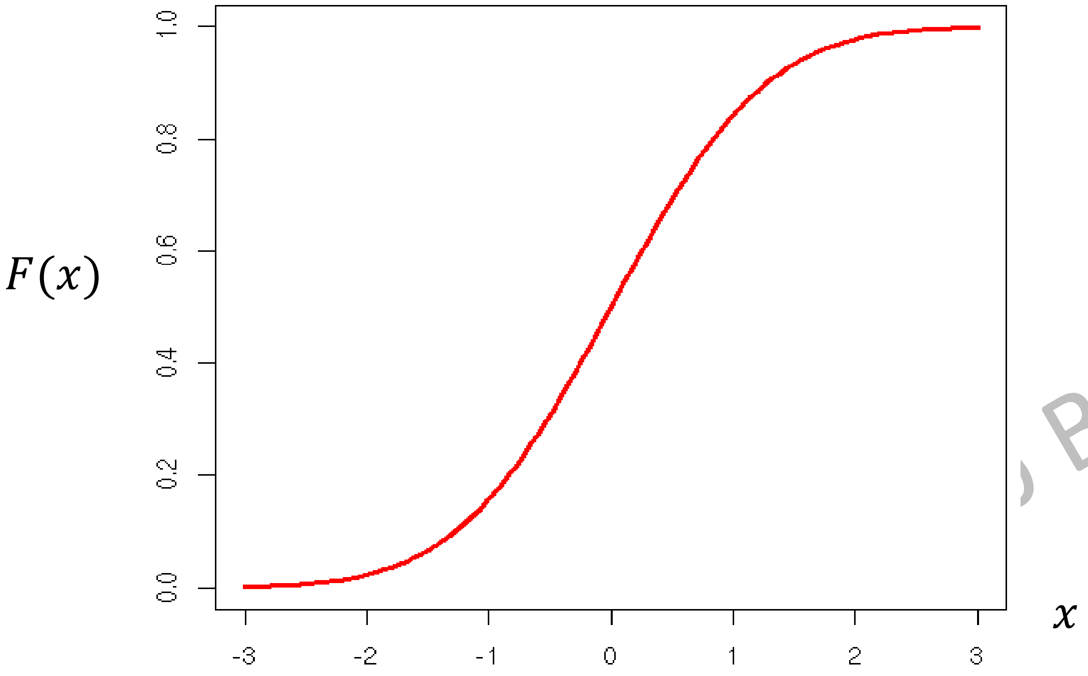
$$=F(x_i) - F(x_{i-1})$$


Sample
Population
Pxi
fsxi
Δx
x
xi
Fsxi
Fsxi
dF(x)
dx
x

Probability density function







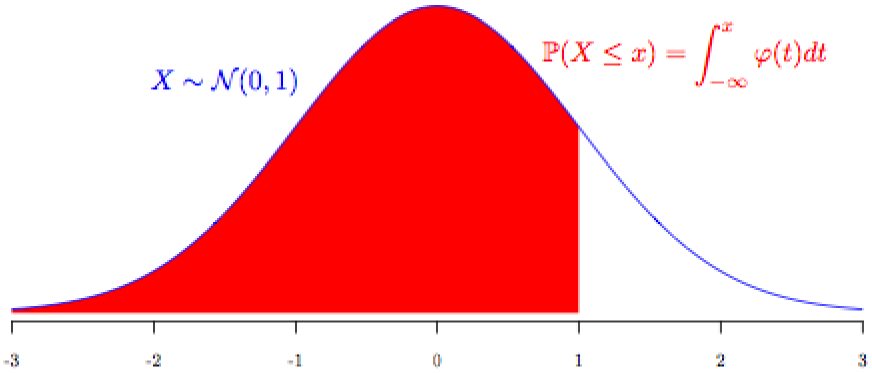
$$f(x) = \frac{1}{\sqrt{2\pi}} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

$$f(z) = \frac{1}{\sqrt{2\pi}} exp\left[-\frac{z^2}{2}\right] \qquad z = \frac{x-\mu}{\sigma} \qquad -\infty \leq z \leq \infty$$

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

Cumulative probability of standard normal distribution

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990



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

$$F(z) = B \qquad \text{for } z < 0$$

$$F(z) = 1 - B \qquad \text{for } z \geq 0$$

$$B = \frac{1}{2}\left[1 + 0.196854|z| + 0.115194|z|^2 + 0.000344|z|^3 + 0.019527|z|^4\right]^{-4}$$

Ex. What is the probability that the standard normal random variable z will be less than -2? Less than 1? What is $P(-2 < Z < 1)$?

Solution

$$P(Z < -2) = F(-2) = 1 - F(2) = 1 - 0.9772 = 0.228$$

$$P(Z < 1) = F(1) = 0.8413$$

$$P(-2 < Z < 1) = F(1) - F(2) = 0.841 - 0.023 = 0.818$$

Example 2:

The annual runoff of a stream is modeled by a normal distribution with mean and standard deviation of 5000 and 1000 ha-m respectively.

i. Find the probability that the annual runoff in any year is more than 6500 ha-m.

ii. Find the probability that it would be between 3800 and 5800 ha-m.

Solution:

Let X is the random variable denoting the annual runoff. Then z is given by

$z = \frac{(X-5000)}{1000}$ N (0, 1)

(i) P (X≥6500) = P (z $\geq\frac{(6500-5000)}{1000}$)

= P (z≥1.5)

= 1-P (z≤1.5)

= 1- F (1.5)

F (1.5) = 0.9332

P (X≥6500) =1-0.9332 = 0.0688

P (X≥6500) =0.0688

(ii) P (3800≤X≤5800)

$= \text{P} [\frac{(3800-5000)}{1000} \leq z \leq \text{P} \ \frac{(5800-5000)}{1000}]$

= P [-1.2≤z≤0.8]

= F (0.8)-F (-1.2)

= F (0.8)-[1-F (1.2)]

=0.7881-(1-0.8849)

=0.673

Statistical parameter

Expected value $$E(x) = \mu = \int_{-\infty}^{\infty} x f(x) dx$$

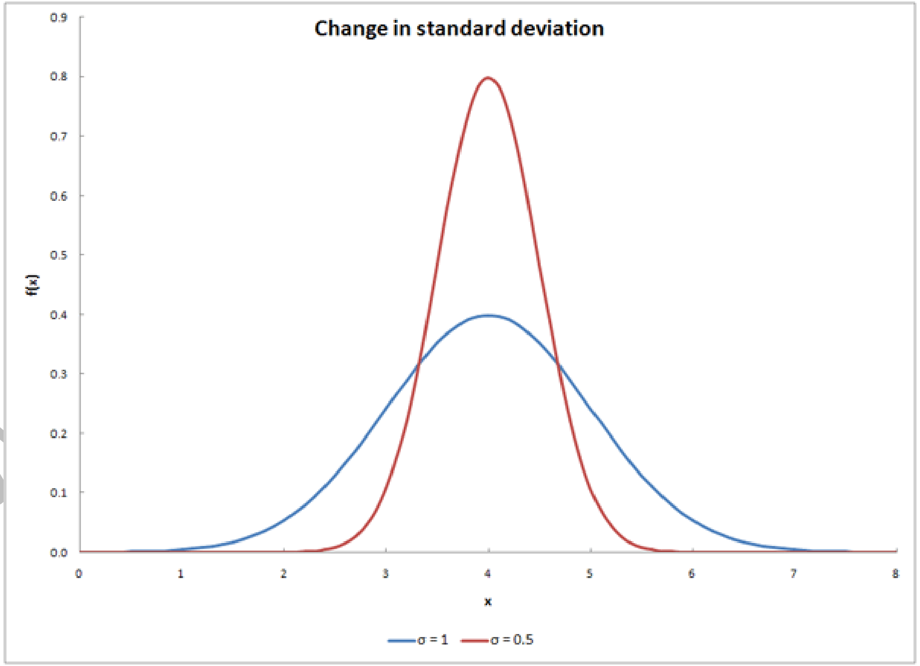
It is the first moment about the origin of the random variable, a measure of the midpoint or central tendency of the distribution

The sample estimate of the mean is the average $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

The variability of data is measured by the variance σ^2

$$E[(x-\mu)^2] = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$$

The sample estimate of the variance is given by $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$



A symmetry of distribution about the mean is a measured by the skewness. This is the third moment about the mean

$$E[(x-\mu)^3] = \int_{-\infty}^{\infty} (x-\mu)^3 f(x)dx$$

The coefficient of skewness γ is defined as

$$\gamma = \frac{1}{\sigma^3} E[(x-\mu)^3]$$

A sample estimate for γ is given by

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

Fitting a probability distribution

A probability distribution is a function representing the probability of occurrence of a random variable.

By fitting a distribution function, we can extract the probabilistic information of the random variable

Fitting distribution can be achieved by the method of moments and the method of maximum likelihood

Method of moments

Developed by Karl Pearson in 1902

He considered that good estimate of the parameters of a probability distribution are those for which moments of the probability density function about the origin are equal to the corresponding moments of the sample data

Karl Pearson (27 March 1857 – 27 April 1936) was an English mathematician

The moments are

$$E(x) = \mu = \int_{-\infty}^{\infty} xf(x)dx \ = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$E[(x-\mu)^2] = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \quad = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$E[(x-\mu)^3] = \int_{-\infty}^{\infty} (x-\mu)^3 f(x)dx \quad = \frac{n\sum_{i=1}^{n}(x_i - \bar{x})^3}{(n-1)(n-2)}$$

Method of moments

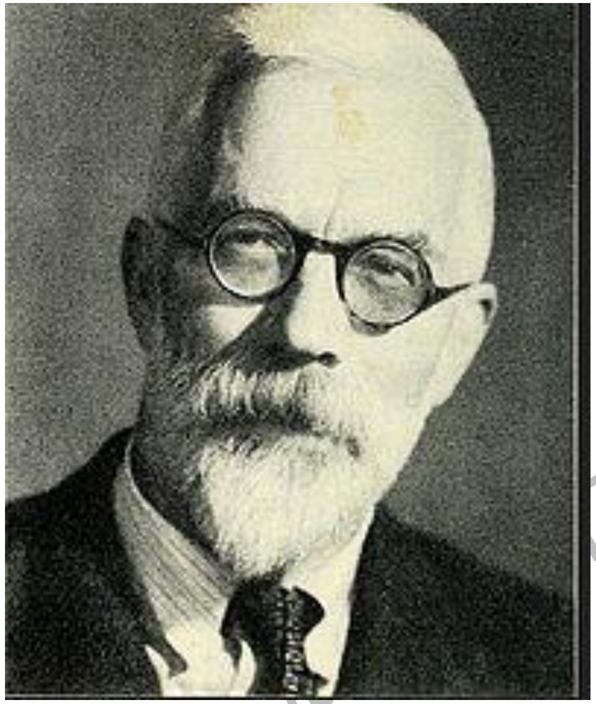
Developed by R A Fisher in 1922

He reasoned that the best value of a parameter of a probability distribution should be that value which maximizes the likelihood of joint probability of occurrence of the observed sample

Suppose that the sample space is divided into intervals of length dx and that a sample of independent and identically distributed observations $x_1, x_2, x_3, \ldots, x_n$ is taken

The value of the probability density for $X = x_i$ is $f(x_i)$

The probability that the random variable will occur in that interval including x_i is $f(x_i)dx$



Sir Ronald Aylmer Fisher (17 February 1890 – 29 July 1962), known as R.A. Fisher, was an English statistician, evolutionary biologist, mathematician, geneticist, and eugenicist.

Since the observation is independent, their joint probability of occurrence is

$$f(x_1)dxf(x_2)dxf(x_3)dx \dots f(x_n)dx = \prod_{i=1}^{n} f(x_i)dx^n$$

Since dx is fixed, we can maximize

$$Max\ L = \prod_{i=1}^{n} f(x_i)$$

$$Max\ lnL = \sum_{i=1}^{n} lnf(x_i)$$

Example: The exponential distribution can be used to describe various kinds of hydrological data, such as inter arrival times of rainfall events. Its probability density function is $f(x) = \lambda e^{-\lambda x}$ for $x > 0$. Determine the relationship between the parameter λ and the first moment about the origin μ.

Solution:

Method of moments

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

$$= \int_{-\infty}^{\infty} x \lambda e^{-\lambda x} dx$$

$$\mu = \frac{1}{\lambda}$$

Method of maximum likelihood

$$lnL = \sum_{i=1}^{n} lnf(x_i)$$

$$lnL = \sum_{i=1}^{n} ln\left(\lambda e^{-\lambda x_i}\right)$$

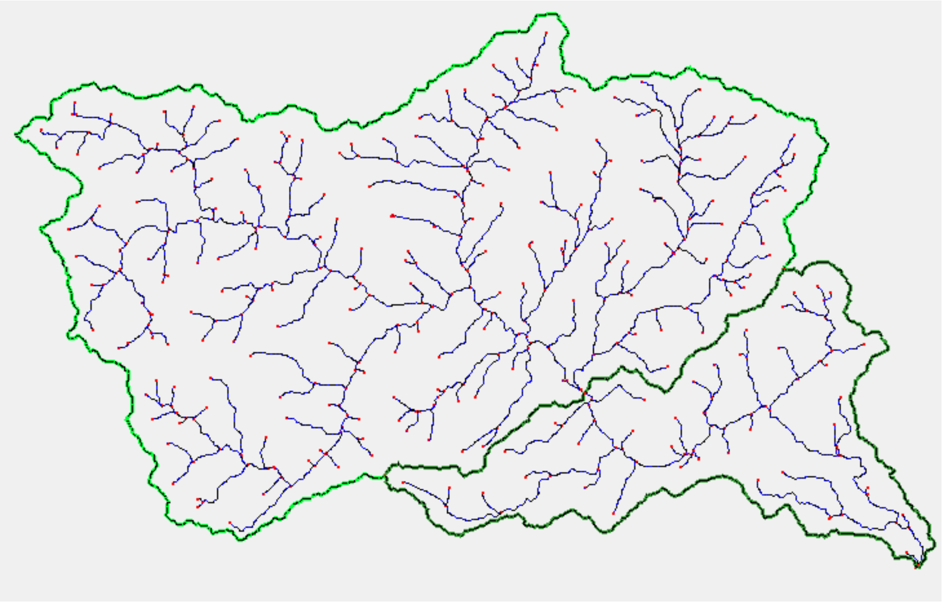
$$lnL = \sum_{i=1}^{n} (ln\lambda - \lambda x_i)$$

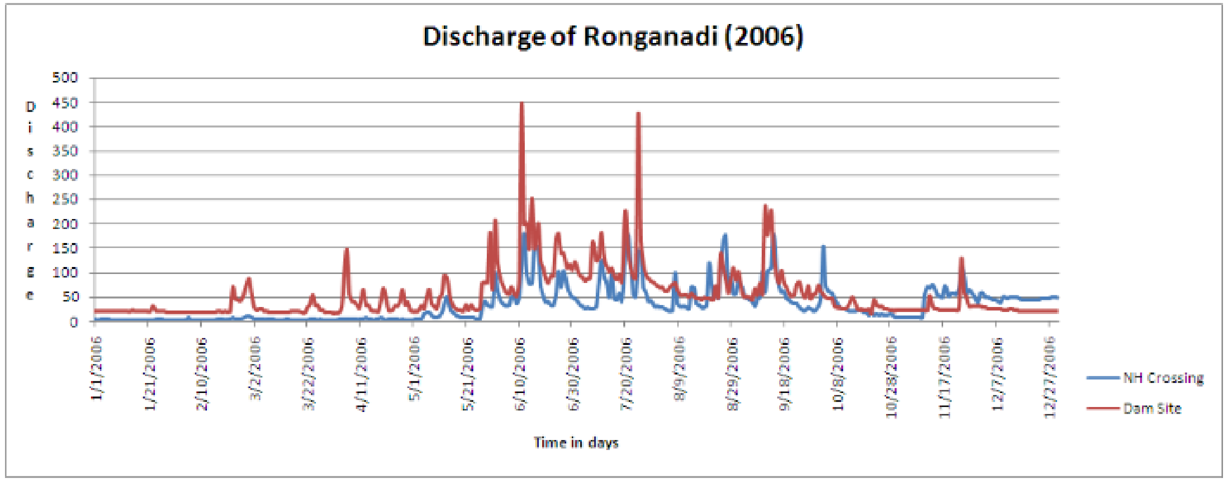
$$lnL = nln\lambda - \lambda \sum_{i=1}^{n} x_i$$

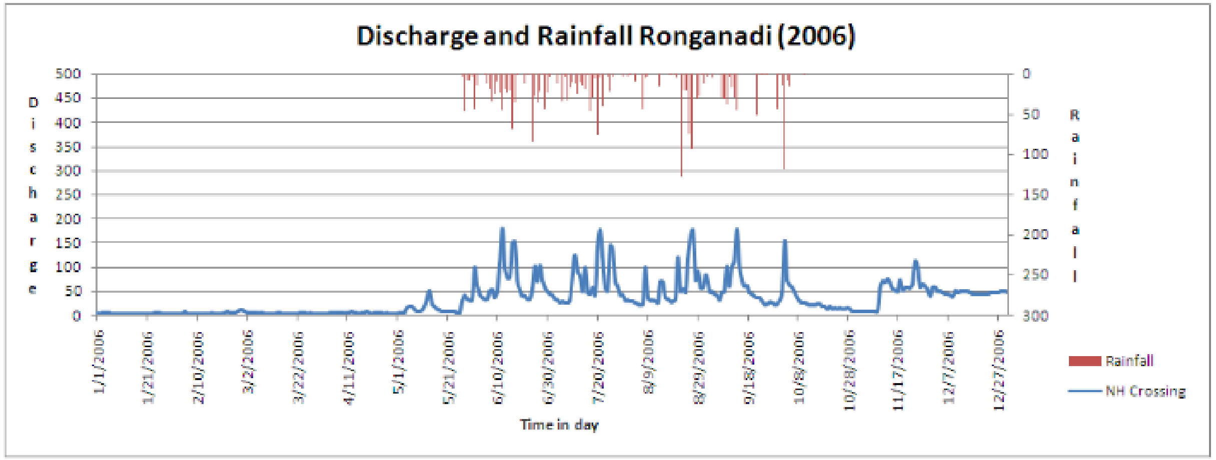
$$\frac{\partial lnL}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

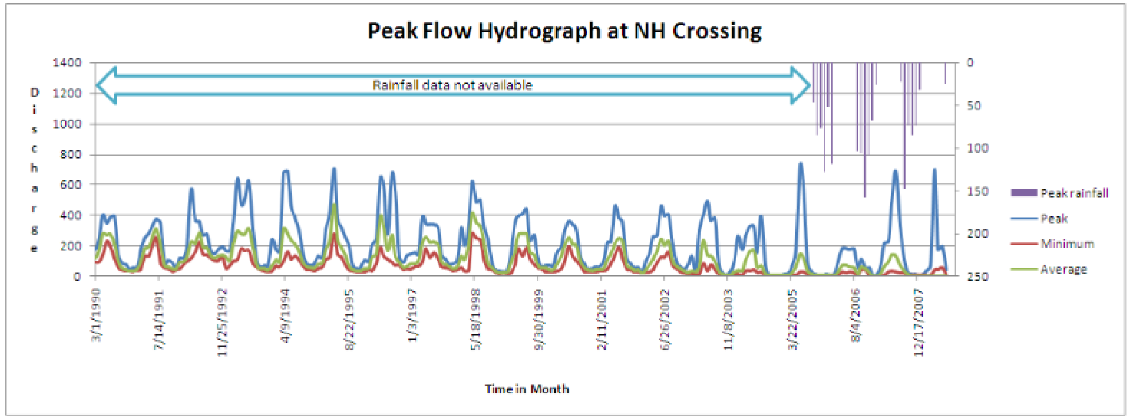
$$\frac{1}{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

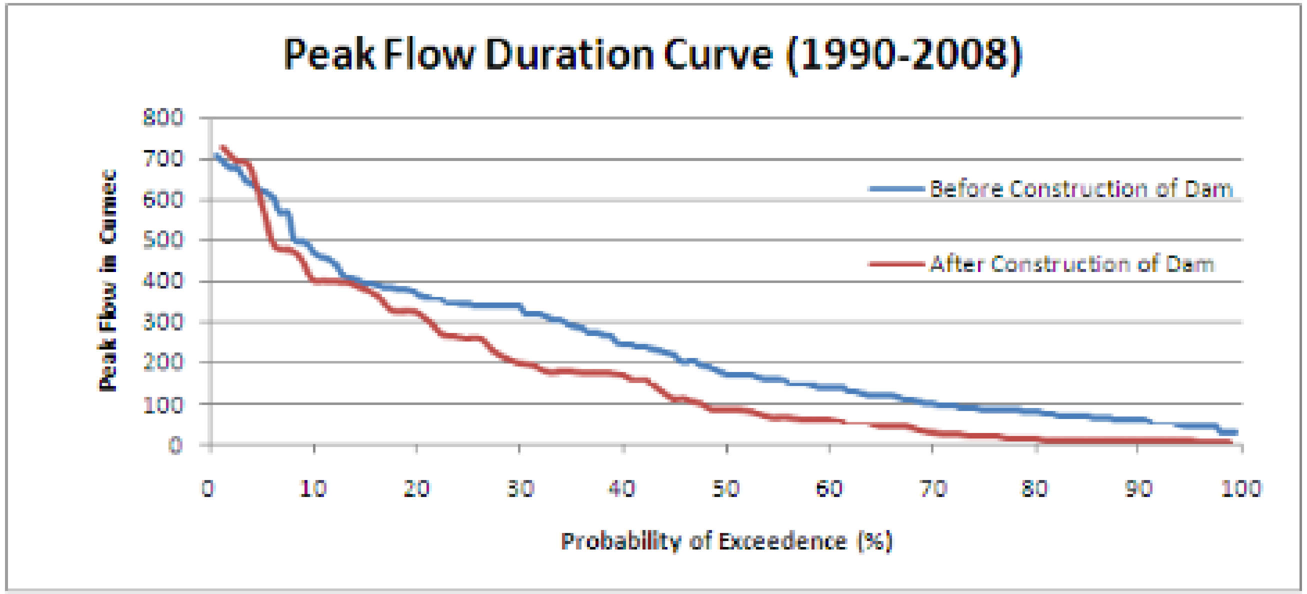
$$\boxed{\lambda = \frac{1}{\bar{x}}}$$

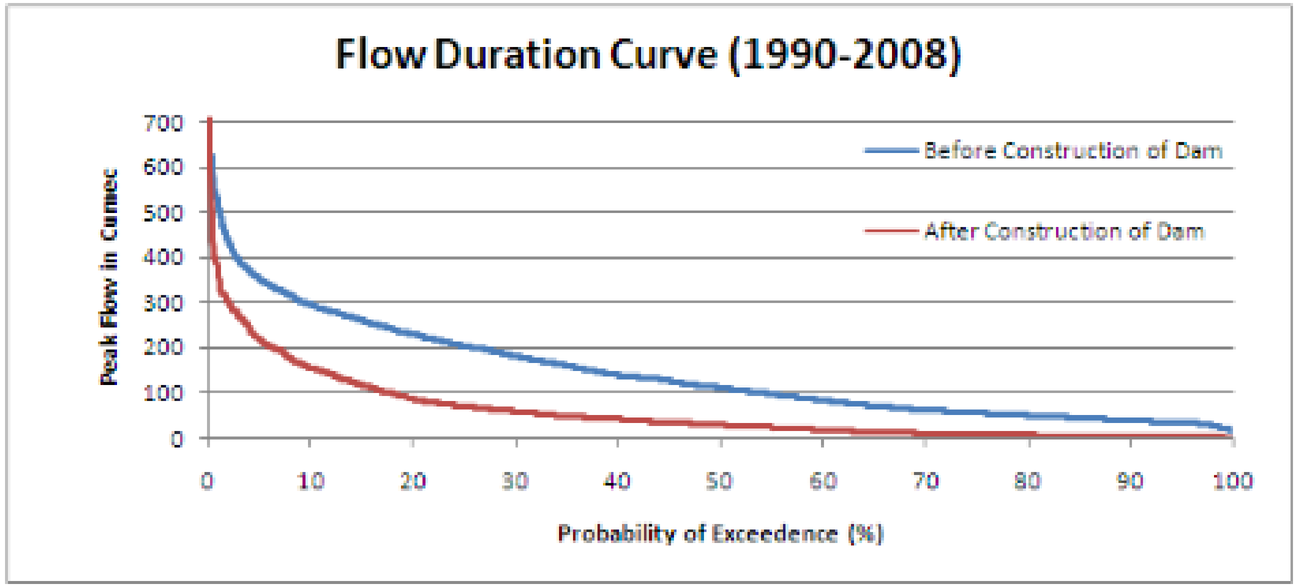



Discharge of Ronganadi (2006)
Discharge
500
450
400
350
300
250
200
150
100
50
0
1/1/2006
1/21/2006
2/10/2006
3/2/2006
3/22/2006
4/11/2006
5/1/2006
5/21/2006
6/10/2006
6/30/2006
7/20/2006
8/9/2006
8/29/2006
9/18/2006
10/8/2006
10/28/2006
11/17/2006
12/7/2006
12/27/2006
Time in days
NH Crossing
Dam Site


Discharge and Rainfall Ronganadi (2006)
Discharge
500
450
400
350
300
250
200
150
100
50
0
Rainfall
0
50
100
150
200
250
300
1/1/2006
1/21/2006
2/10/2006
3/2/2006
3/22/2006
4/11/2006
5/1/2006
5/21/2006
6/10/2006
6/30/2006
7/20/2006
8/9/2006
8/29/2006
9/18/2006
10/8/2006
10/28/2006
11/17/2006
12/7/2006
12/27/2006
Time in day
Rainfall
NH Crossing


Peak Flow Hydrograph at NH Crossing
Rainfall data not available
Discharge
1400
1200
1000
800
600
400
200
0
0
50
100
150
200
250
3/1/1990
7/14/1991
11/25/1992
4/9/1994
8/22/1995
1/3/1997
5/18/1998
9/30/1999
2/11/2001
6/26/2002
11/8/2003
3/22/2005
8/4/2006
12/17/2007
Time in Month
Peak rainfall
Peak
Minimum
Average

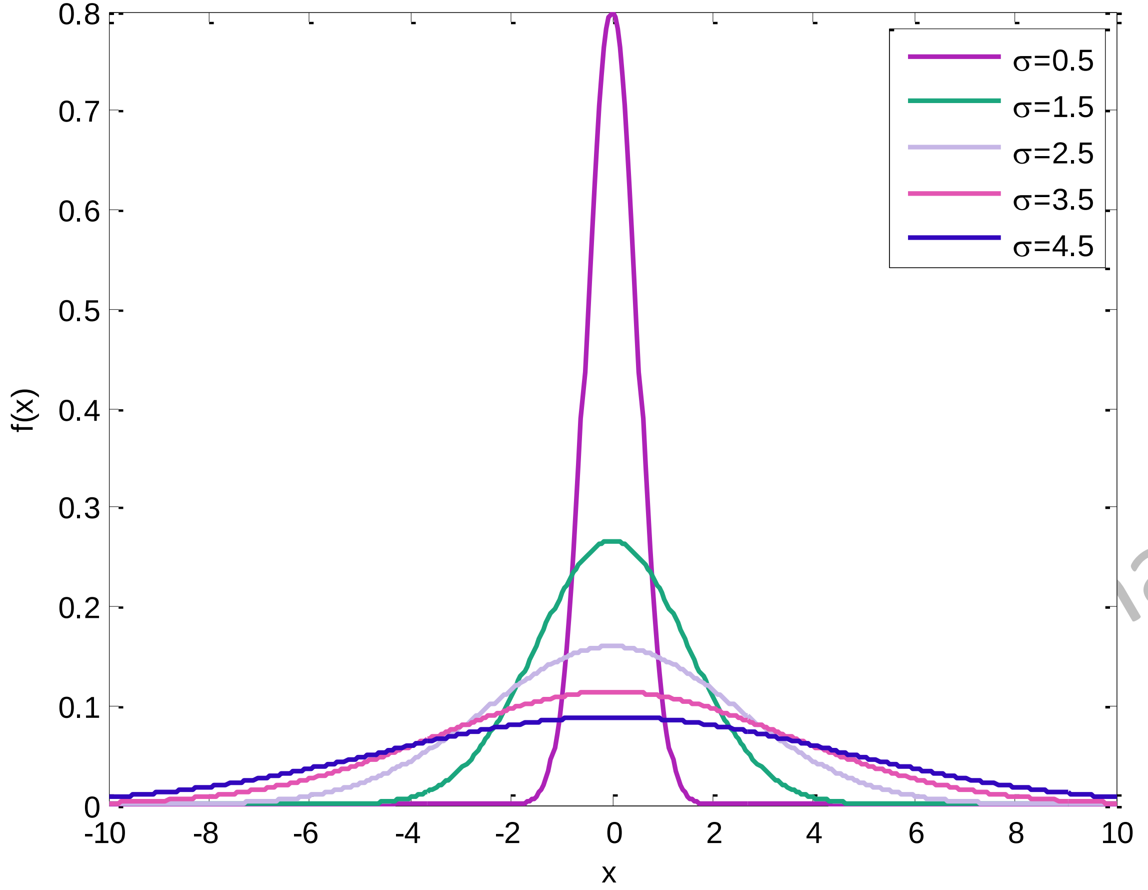

Peak Flow Duration Curve (1990-2008)
800
700
600
500
400
300
200
100
0
Peak Flow in Cumec
Before Construction of Dam
After Construction of Dam
0
10
20
30
40
50
60
70
80
90
100
Probability of Exceedence (%)

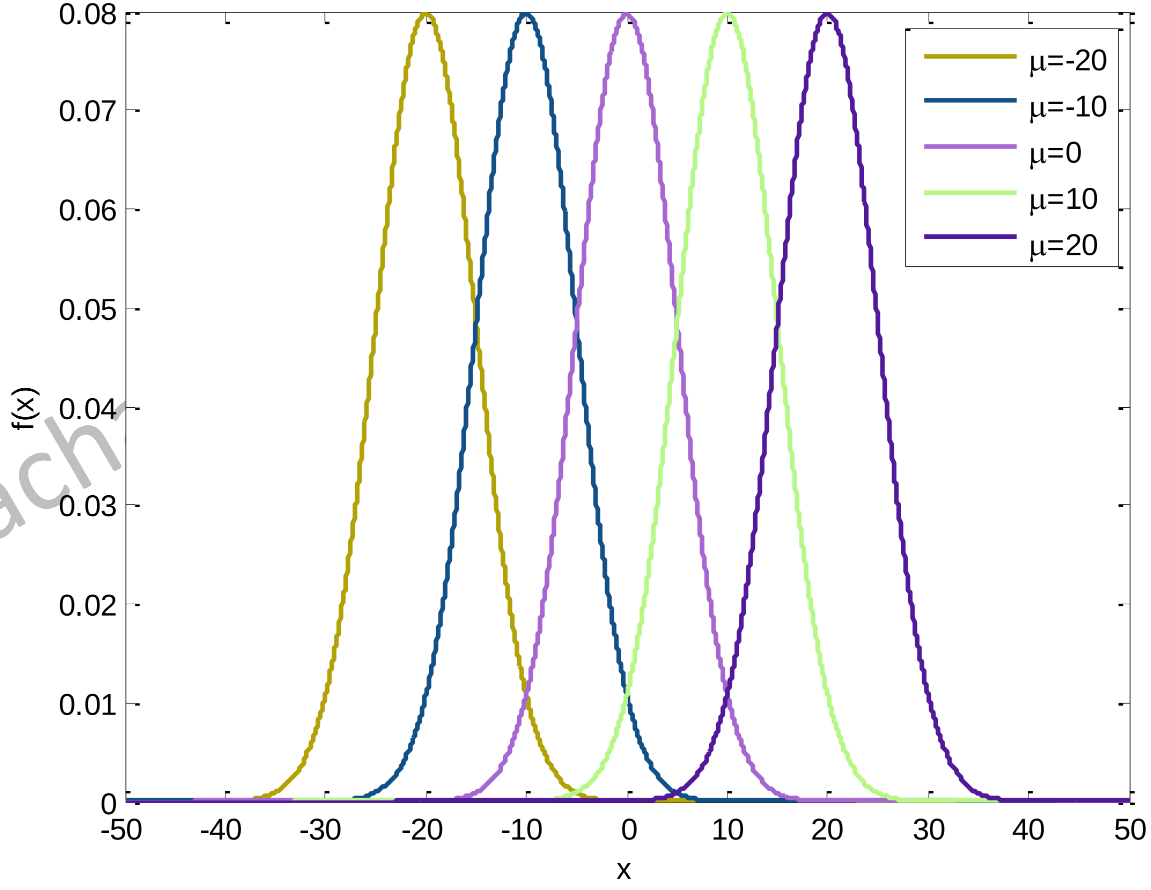

Flow Duration Curve (1990-2008)
700
600
500
400
300
200
100
0
Peak Flow in Cumec
Before Construction of Dam
After Construction of Dam
0
10
20
30
40
50
60
70
80
90
100
Probability of Exceedence (%)

Normal family	Normal, lognormal, lognormal-III
Generalized extreme value family	EV1 (Gumbel), GEV, and EVIII (Weibull)
Exponential/Pearson type family	Exponential, Pearson type III, Log-Pearson type III

Normal Distribution

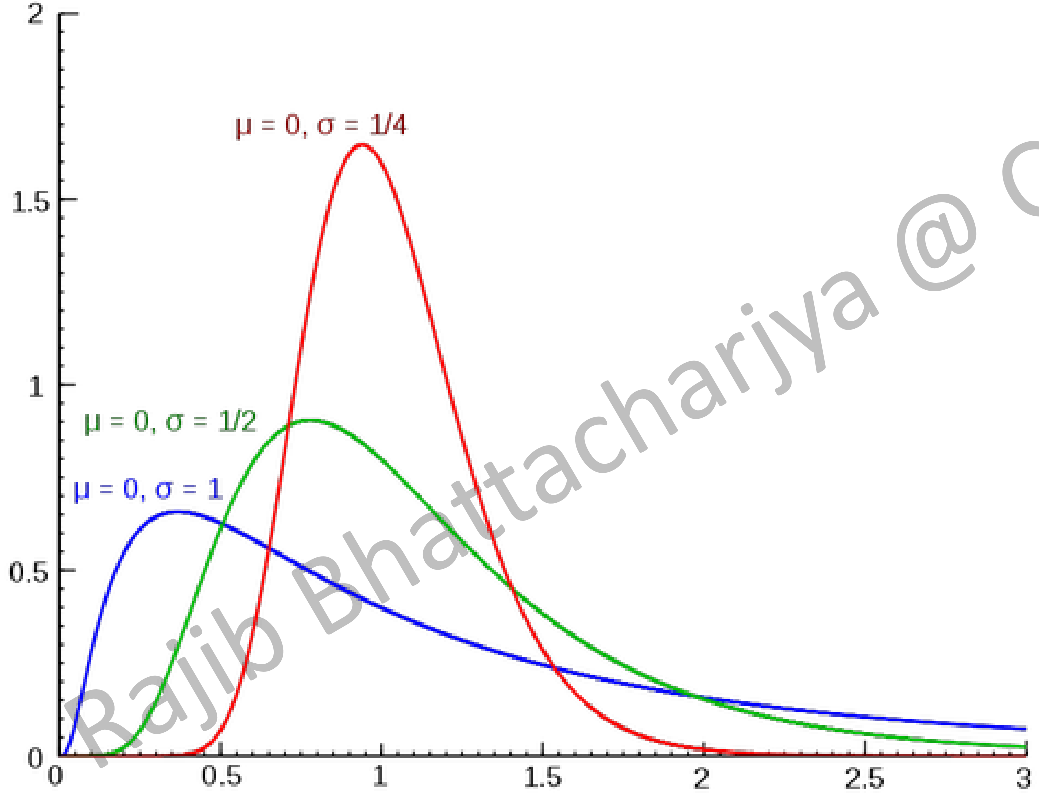
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty \le x \le +\infty$$





Log-Normal Distribution

$$f(y) = \frac{1}{\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2} \qquad x > 0 \text{ and } y = \ln(x)$$

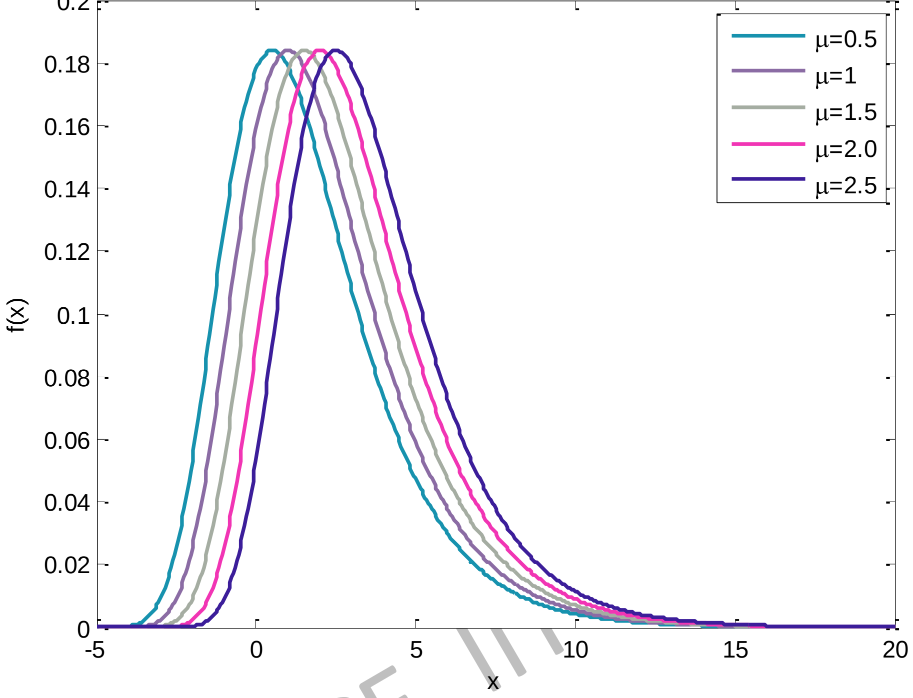


Extreme value (EV) distributions

- Extreme values – maximum or minimum values of sets of data
- Annual maximum discharge, annual minimum discharge
- When the number of selected extreme values is large, the distribution converges to one of the three forms of EV distributions called Type I, II and III

EV type I distribution

If M1, M2…, Mn be a set of daily rainfall or streamflow, and let X = max(Mi) be the maximum for the year. If Mi are independent and identically distributed, then for large n, X has an extreme value type I or Gumbel distribution.

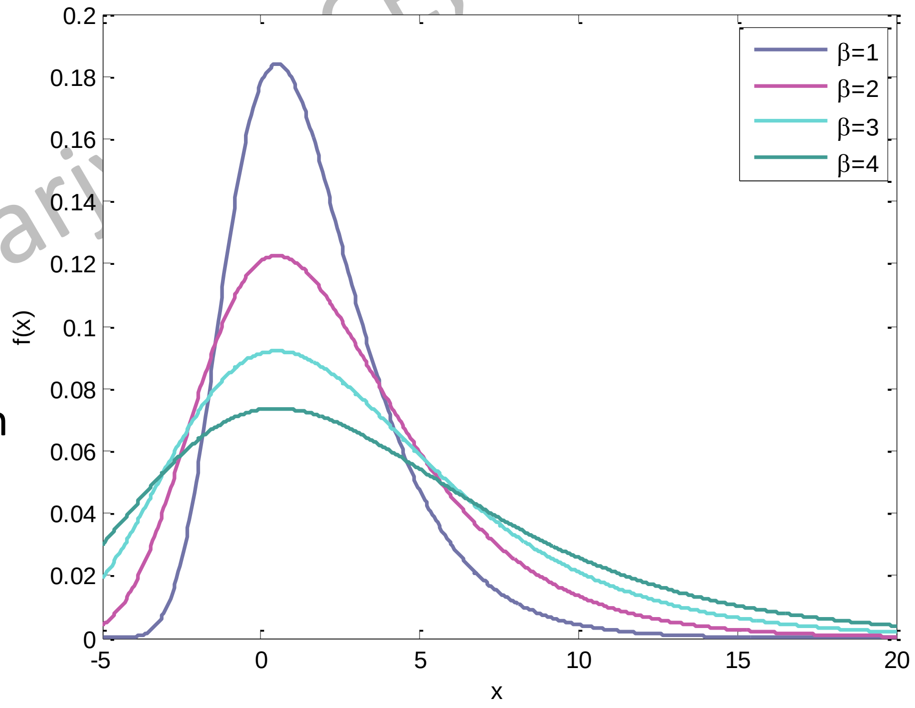


$$f(x) = \frac{1}{\beta} e^{-(y+e^{-y})}$$

$$y = \frac{x-\mu}{\beta}$$

$$\mu = \bar{x} - 0.5772\beta \qquad \beta = \frac{\sqrt{6}S_x}{\pi}$$

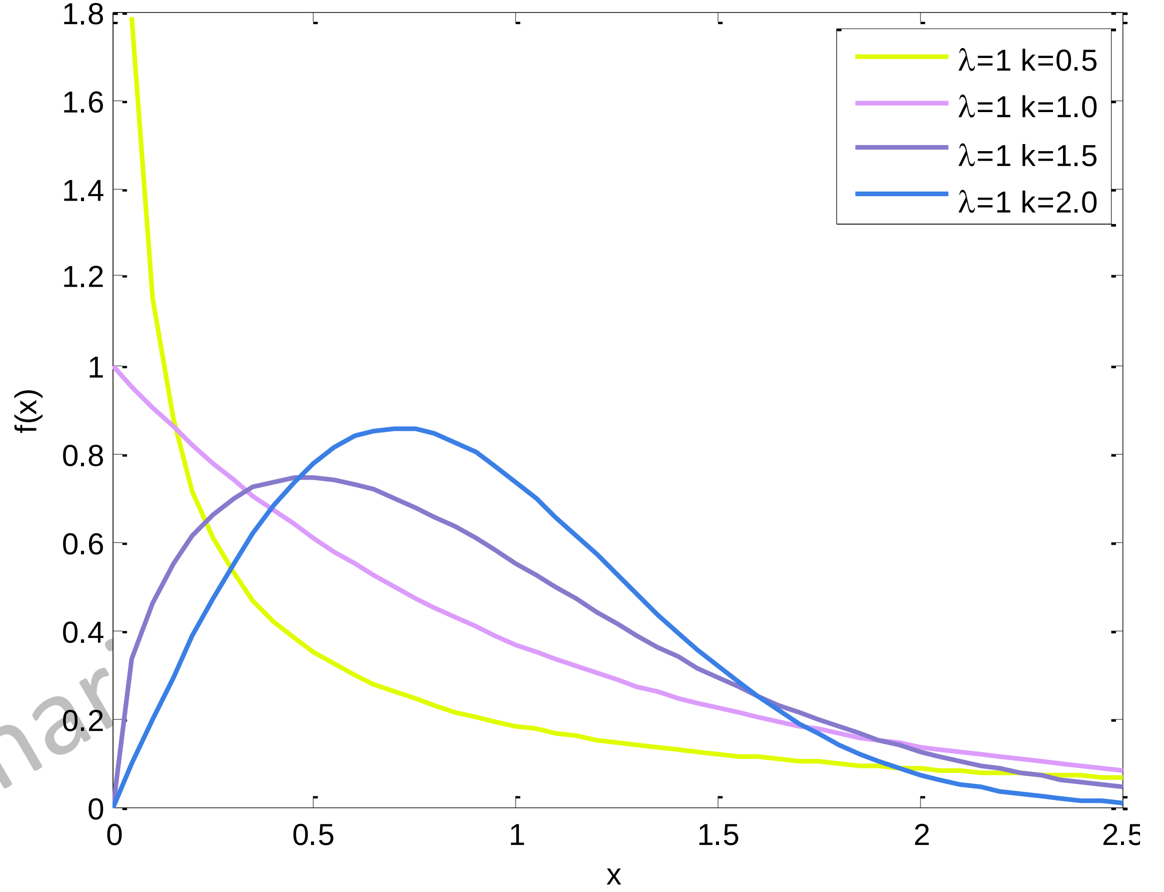
Distribution of annual maximum stream flow follows an EV1 distribution



EV type III distribution

If Wi are the minimum stream flows in different days of the year, let X = min(Wi) be the smallest. X can be described by the EV type III or Weibull distribution.

$$f(x) = \left(\frac{k}{\lambda}\right)\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^{k}} \qquad x > 0, \lambda, k > 0$$

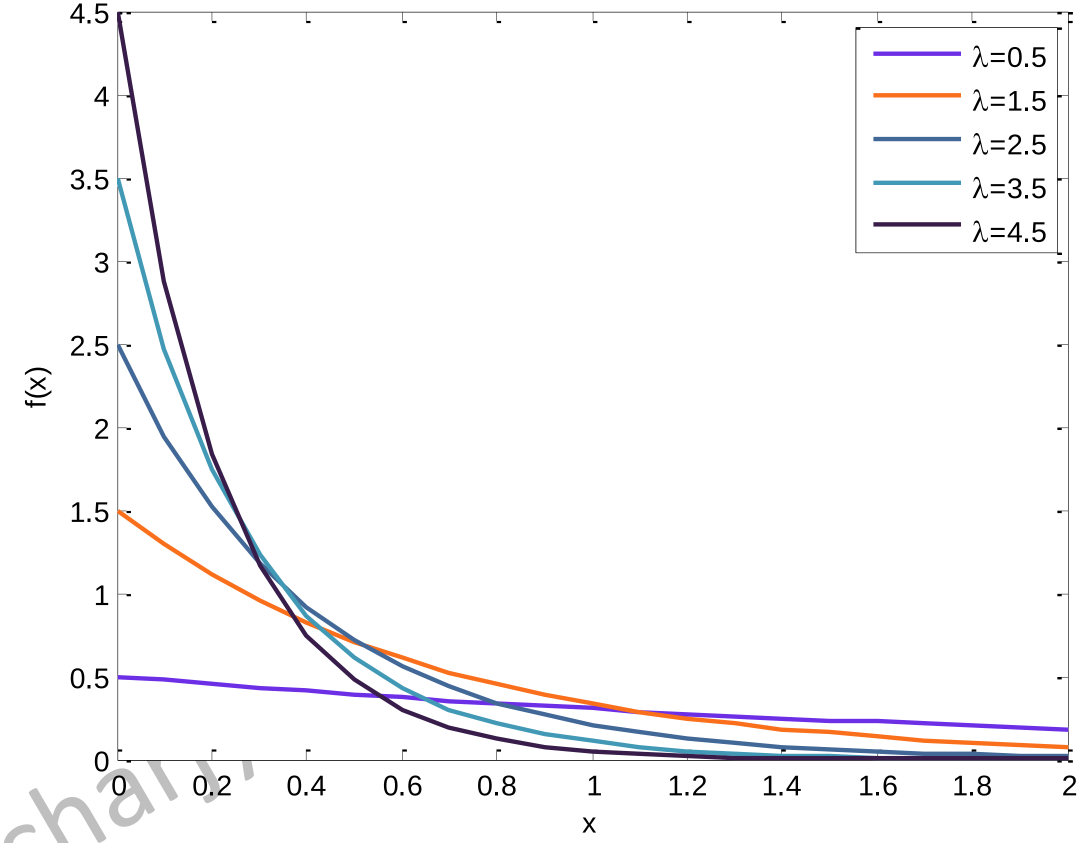


Distribution of low flows (eg. 7-day min flow) follows EV3 distribution.

Exponential distribution

Poisson process – a stochastic process in which the number of events occurring in two disjoint subintervals are independent random variables.

In hydrology, the interarrival time (time between stochastic hydrologic events) is described by exponential distribution



$$f(x) = \lambda e^{-\lambda x} \qquad x \geq 0; \lambda = \frac{1}{x}$$

Interarrival times of polluted runoffs, rainfall intensities, etc are described by exponential distribution.

Gamma Distribution

- The time taken for a number of events (b) in a Poisson process is described by the gamma distribution
- Gamma distribution – a distribution of sum of b independent and identical exponentially distributed random variables.

$$f(x) = \frac{\lambda^{\beta} x^{\beta-1} e^{-\lambda x}}{\Gamma(\beta)} \qquad x \geq 0; \Gamma = gamma\ \ function$$

Skewed distributions (eg. hydraulic conductivity) can be represented using gamma without log transformation.

Pearson Type III

Named after the statistician Pearson, it is also called three-parameter gamma distribution. A lower bound is introduced through the third parameter (ε)

$$f(x) = \frac{\lambda^{\beta}(x-\varepsilon)^{\beta-1}e^{-\lambda(x-\varepsilon)}}{\Gamma(\beta)} \qquad x \geq \varepsilon; \Gamma = gamma\ \ function$$

It is also a skewed distribution first applied in hydrology for describing the pdf of annual maximum flows.

Log-Pearson Type III

If log X follows a Person Type III distribution, then X is said to have a log-Pearson Type III distribution

$$f(x) = \frac{\lambda^{\beta}(y-\varepsilon)^{\beta-1}e^{-\lambda(y-\varepsilon)}}{\Gamma(\beta)} \qquad y = log\ x \geq \varepsilon$$

Frequency analysis for extreme events

Q. Find a flow (or any other event) that has a return period of T years

$$f(x) = \frac{1}{\alpha}\exp\left[-\frac{x-u}{\alpha} - \exp\left(-\frac{x-u}{\alpha}\right)\right] \qquad F(x) = \exp\left[-\exp\left(-\frac{x-u}{\alpha}\right)\right]$$

EV1 pdf and cdf

$$\alpha = \frac{\sqrt{6}s_x}{\pi} \qquad u = \bar{x} - 0.5772\alpha$$

Define a reduced variable y $\quad y = \dfrac{x-u}{\alpha}$

$$F(x) = \exp[-\exp(-y)]$$

$$y = -\ln[-\ln(F(x))] = -\ln[-\ln(1-p)] \quad where\ p = P(x \geq x_T)$$

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

If you know T, you can find y_T, and once y_T is know, x_T can be computed by

$$x_T = u + \alpha y_T$$

Example 01

Given annual maxima for 10-minute storms. The mean and standard deviation are 0.649 in and 0.177 in.

Find 5 & 50 year return period 10-minute storms

$$\bar{x} = 0.649\, in$$

$$s = 0.177\, in$$

$$\alpha = \frac{\sqrt{6}s}{\pi} = \frac{\sqrt{6} * 0.177}{\pi} = 0.138 \qquad u = \bar{x} - 0.5772\alpha = 0.649 - 0.5772 * 0.138 = 0.569$$

$$y_5 = -\ln\left[\ln\left(\frac{T}{T-1}\right)\right] = -\ln\left[\ln\left(\frac{5}{5-1}\right)\right] = 1.5$$

$$\boxed{x_5 = u + \alpha y_5 = 0.569 + 0.138 * 1.5 = 0.78\, in}$$

$$\boxed{x_{50} = 1.11\, in}$$

Normal Distribution

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

$$K_T = \frac{x_T - \bar{x}}{s} = z_T$$

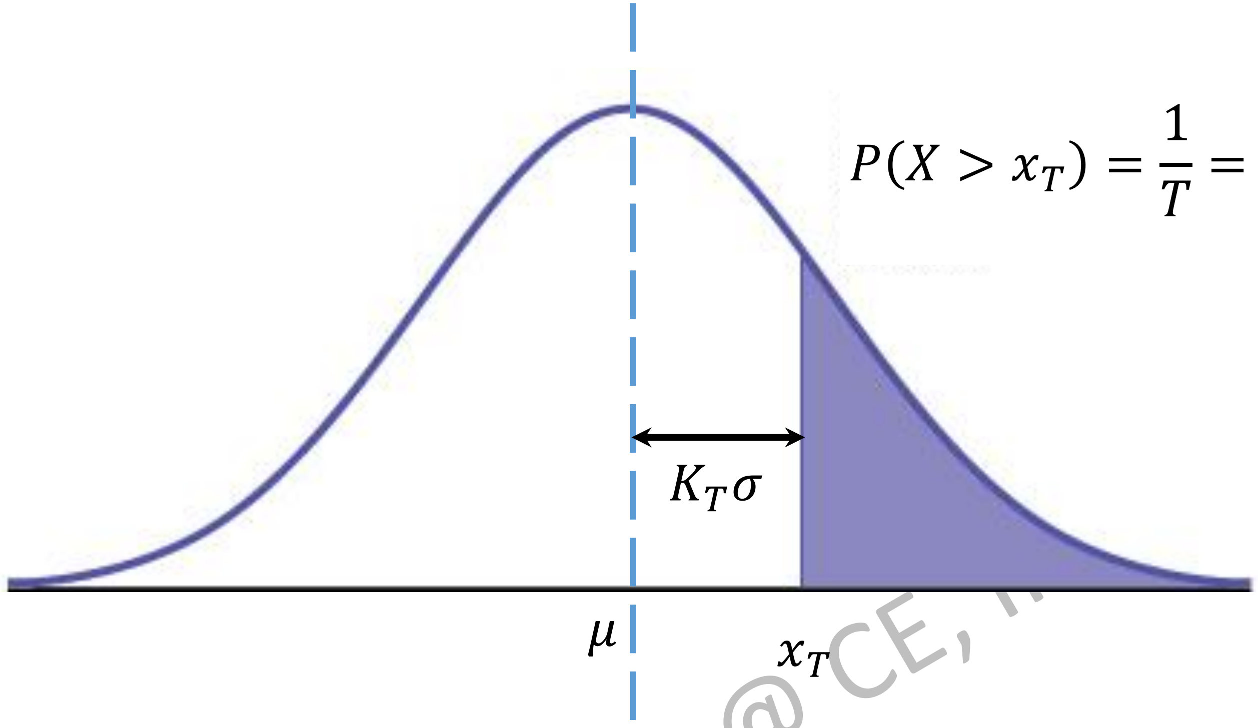
So the frequency factor for the Normal Distribution is the standard normal variate

$$x_T = \bar{x} + K_T s = \bar{x} + z_T s$$

Example: 50 year return period

$$T = 50;\ p = \frac{1}{50} = 0.02;\ K_{50} = z_{50} = 2.054$$

$$w = \left[ln\left(\frac{1}{p^2}\right)\right]^{1/2} \qquad 0 < p \leq 0.5$$



$$P(X > x_T) = \frac{1}{T} = \int_{x_T}^{\infty} f(x)dx$$

$$z = w - \frac{2.515517 + 0.802853w + 0.010328w^2}{1 + 1.432788w + 0.189269w^2 + 0.001308w^3}$$

EV-I (Gumbel) Distribution

$$F(x) = \exp\left[-\exp\left(-\frac{x-u}{\alpha}\right)\right] \qquad \alpha = \frac{\sqrt{6}s}{\pi} \qquad u = \bar{x} - 0.5772\alpha \qquad y_T = -\ln\left[\ln\left(\frac{T}{T-1}\right)\right]$$

$$\begin{aligned} x_T &= u + \alpha y_T \\ &= \bar{x} - 0.5772\,\frac{\sqrt{6}}{\pi}s + \frac{\sqrt{6}}{\pi}s\left\{-\ln\left[\ln\left(\frac{T}{T-1}\right)\right]\right\} \\ &= \bar{x} - \frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{T}{T-1}\right)\right]\right\}s \end{aligned}$$

$$x_T = \bar{x} + K_T s \qquad \boxed{K_T = -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{T}{T-1}\right)\right]\right\}}$$

Example 02

Given annual maximum rainfall, calculate 5-yr storm using frequency factor

$$K_T = -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{T}{T-1}\right)\right]\right\}$$

$$K_T = -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{5}{5-1}\right)\right]\right\} = 0.719$$

$$x_T = \bar{x} + K_T s$$

$$= 0.649 + 0.719 \times 0.177$$

$$= 0.78\ in$$

Distribution	PDF	CDF	Range	Mean
Normal	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ Or $\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}, \quad z=\frac{x-\mu}{\sigma}$	$\int_{-\infty}^{x}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx$	$-\infty \le x \le +\infty$	μ and σ
Lognormal $y=\ln(x)$	$\frac{1}{\sigma_y\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu_y}{\sigma_y}\right)^2}$	$\int_{-\infty}^{y}\frac{1}{\sigma_y\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu_y}{\sigma_y}\right)^2}dy$	$-\infty \le y \le +\infty$ $0 \le x \le +\infty$	μ_y and σ_y
Extreme value Type I $y=(x-\beta)/\alpha$	$\frac{1}{\alpha}e^{-(y+e^{-y})}$	e^{-e-y}	$-\infty \le x \le +\infty$	$\beta=\bar{x}-0.5772\alpha$ $\alpha=\frac{\sqrt{6}S_x}{\pi}$
Extreme value Type III	$\alpha x^{\alpha-1}\beta^{-\alpha}e^{-(x/\beta)\alpha}$	$1-e^{-(x/\beta)\alpha}$	$x \ge 0$	$\beta\Gamma(1+1/\alpha)$ $\beta[\Gamma(1+2/\alpha)$

Ex. Annual maximum values of 10 min duration rainfall at some place from 1913 to 1947 are presented in Table below. Develop a model for storm rainfall frequency analysis using Extreme Value Type I distribution and calculate the 5, 10, and 50 year return period maximum values of 10 min rainfall of the area.

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	0.36	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	
Mean	0.649		SD	0.177

Solution

$$\alpha = \frac{\sqrt{6}s}{\pi} = \frac{\sqrt{6} \times 0.177}{\pi} = 0.138$$

$$u = \bar{x} - 0.5772\alpha = 0.649 - 0.5772 \times 0.138 = 0.569$$

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

$$y_T = -ln\left[ln\left(1 - \frac{1}{5}\right)\right] = 1.5$$

$$x_T = u + \alpha y_T = 0.569 + 0.138 \times 1.5 = 0.78$$

$$\boxed{K_T = -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{T}{T-1}\right)\right]\right\}}$$

$$K_T = -\frac{\sqrt{6}}{\pi}\left\{0.5772 + ln\left(ln\left(\frac{5}{5-1}\right)\right)\right\} = 0.719$$

$$x_T = \bar{x} + K_T s = 0.649 + 0.719 \times 0.177 = 0.78$$

LOG – PEARSON TYPE III DISTRIBUTION

$$y = \log(x)$$

Calculate mean $\bar{y}$, standard deviation s_y and the coefficient of skewness C_s

$$K_T = z + (z^2 - 1)k + \frac{1}{3}(z^3 - 6z)k^2 - (z^2 - 1)k^3 + zk^4 + \frac{1}{3}k^5 \qquad k = \frac{C_s}{6}$$

The value of z can be calculated using the procedure described earlier.

TABLE D.6 FREQUENCY FACTOR FOR THE PEARSON TYPE III DISTRIBUTION WITH POSITIVE SKEW COEFFICIENTS (AFTER WATER RESOURCES COUNCIL, 1967)

Skew coef.	Recurrence interval (years)							
	1.0101	2	5	10	25	50	100	200
	Percent chance (≥)							
c_S	99	50	20	10	4	2	1	0.5
3.0	−0.667	−0.396	0.420	1.180	2.278	3.152	4.051	4.970
2.9	−0.690	−0.390	0.440	1.195	2.277	3.134	4.013	4.904
2.8	−0.714	−0.384	0.460	1.210	2.275	3.114	3.973	4.847
2.7	−0.740	−0.376	0.479	1.224	2.272	3.093	3.932	4.783
2.6	−0.769	−0.368	0.499	1.238	2.267	3.071	3.889	4.718
2.5	−0.799	−0.360	0.518	1.250	2.262	3.048	3.845	4.652
2.4	−0.832	−0.351	0.537	1.262	2.256	3.023	3.800	4.584
2.3	−0.867	−0.341	0.555	1.274	2.248	2.997	3.753	4.515
2.2	−0.905	−0.330	0.574	1.284	2.240	2.970	3.705	4.444
2.1	−0.946	−0.319	0.592	1.294	2.230	2.942	3.656	4.372
2.0	−0.990	−0.307	0.609	1.302	2.219	2.912	3.605	4.298
1.9	−1.037	−0.294	0.627	1.310	2.207	2.881	3.553	4.223
1.8	−1.087	−0.282	0.643	1.318	2.193	2.848	3.499	4.147
1.7	−1.140	−0.268	0.660	1.324	2.179	2.815	3.444	4.069
1.6	−1.197	−0.254	0.675	1.329	2.163	2.780	3.388	3.990
1.5	−1.256	−0.240	0.690	1.333	2.146	2.743	3.330	3.910
1.4	−1.318	−0.225	0.705	1.337	2.128	2.706	3.271	3.828
1.3	−1.383	−0.210	0.719	1.339	2.108	2.666	3.211	3.745
1.2	−1.449	−0.195	0.732	1.340	2.087	2.626	3.149	3.661
1.1	−1.518	−0.180	0.745	1.341	2.066	2.585	3.087	3.575
1.0	−1.588	−0.164	0.758	1.340	2.043	2.542	3.022	3.489
0.9	−1.660	−0.148	0.769	1.339	2.018	2.498	2.957	3.401
0.8	−1.733	−0.132	0.780	1.336	1.993	2.453	2.891	3.312
0.7	−1.806	−0.116	0.790	1.333	1.967	2.407	2.824	3.223
0.6	−1.880	−0.099	0.800	1.328	1.939	2.359	2.755	3.132
0.5	−1.955	−0.083	0.808	1.323	1.910	2.311	2.686	3.041
0.4	−2.029	−0.066	0.816	1.317	1.880	2.261	2.615	2.949
0.3	−2.104	−0.050	0.824	1.309	1.849	2.211	2.544	2.856
0.2	−2.178	−0.033	0.830	1.301	1.818	2.159	2.472	2.763
0.1	−2.252	−0.017	0.836	1.292	1.785	2.107	2.400	2.670
0	−2.326	0	0.842	1.282	1.751	2.054	2.326	2.576

Ex. Calculate the 5 and 50 year return period maximum discharge of the Guadalupe River, Texas using lognormal and 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	20600	10100	15200	

SOLUTION

Lognormal Distribution

$\bar{y} = 4.2743$ $\quad s_y = 0.4027$ $\quad C_s = -0.0696$

K_{50} for $C_s = 0$ is 2.054

$$y_{50} = \bar{y} + K_{50} s_y$$
$$= 4.2743 + 2.054 \times 0.4027 = 5.101$$

$$x_{50} = (10)^{5.101} = 126300\ cfs$$

$$x_5 = 41060\ cfs$$

Log-Pearson Type III Distribution

$$K_{50} = 2.054 + \frac{2.0 - 2.054}{-0.1 - 0}(-0.0696 - 0) = 2.016$$

$$y_{50} = \bar{y} + K_{50}s_y = 4.2743 + 2.016 \times 0.4027 = 5.0863$$

$$x_{50} = (10)^{5.0863} = 121990\ cfs$$

$$x_5 = 41170\ cfs$$

PLOTING POSITIONS

Plotting position refers to the probability value assigned to each piece of data to be plotted

$$P(X \geq x_m) = \frac{m}{n}$$ California's formula

$$P(X \geq x_m) = \frac{m-1}{n}$$ Modified formula

$$P(X \geq x_m) = \frac{m-0.5}{n}$$ Hazen (1930) formula

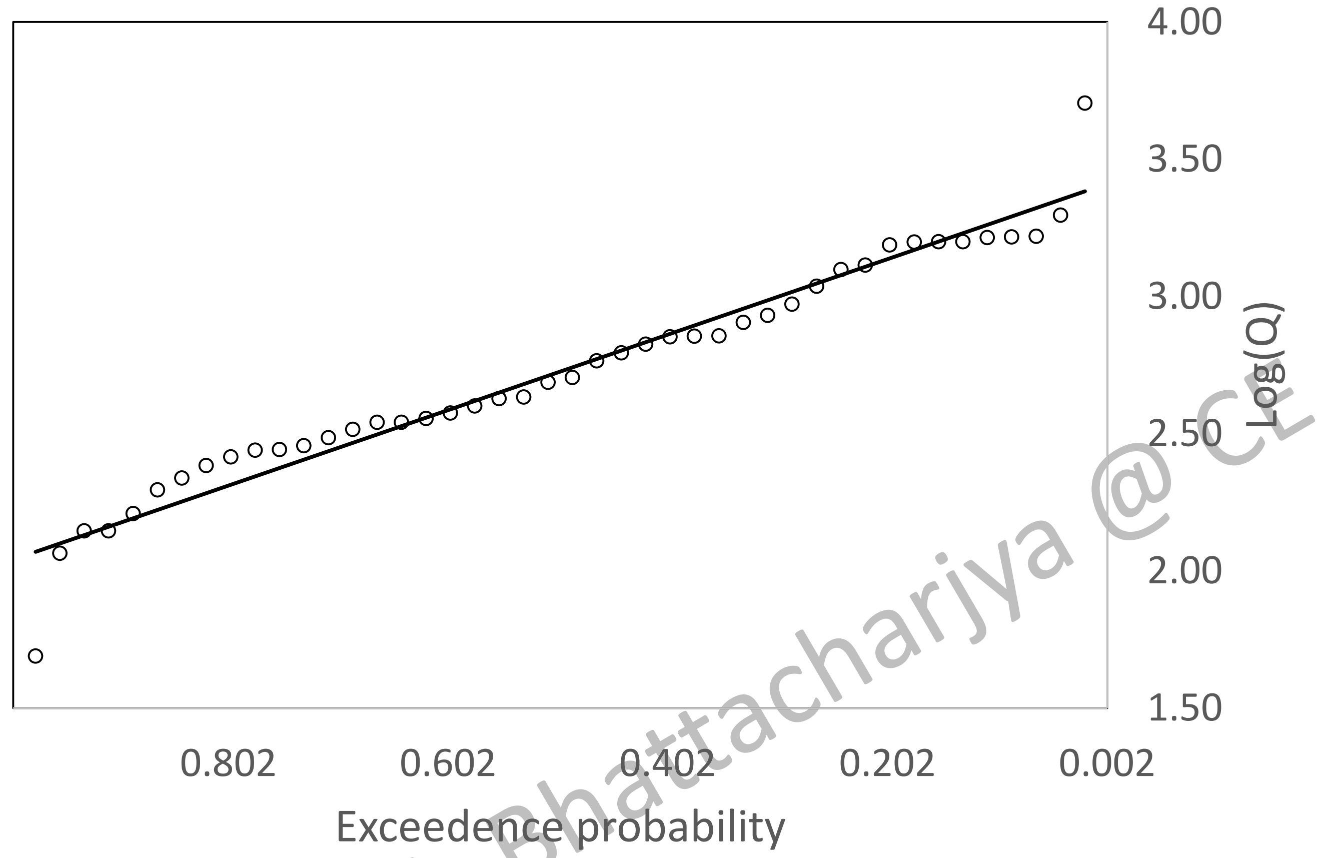
$$P(X \geq x_m) = \frac{m-0.3}{n+0.4}$$ Chegodayev's formula

$$P(X \geq x_m) = \frac{m}{n+1}$$ Weibull formula

$$\boxed{P(X \geq x_m) = \frac{m-b}{n+1-2b}}$$

$b = 0.5$	For Hazen (1930) formula
$b = 0.3$	For Chegodayev's formula
$b = 0$	For Weibull formula
$b = 3/8$	For Blom's formula
$b = 1/3$	For Tukey's formula
$b = 0.44$	For Gringorten's formula

Year	Maimum Discharge (m3/s)	Rank	Exceedence probability	w	z	log(Q)	Log Q from Lognormal distribution
1936	5069	1	0.022	2.76	2.01	3.70	3.54
1967	1982	2	0.044	2.50	1.70	3.30	3.41
1972	1657	3	0.067	2.33	1.50	3.22	3.33
1958	1651	4	0.089	2.20	1.35	3.22	3.27
1941	1642	5	0.111	2.10	1.22	3.22	3.22
1942	1586	6	0.133	2.01	1.11	3.20	3.17
1940	1583	7	0.156	1.93	1.01	3.20	3.13
1961	1580	8	0.178	1.86	0.92	3.20	3.10
1977	1543	9	0.200	1.79	0.84	3.19	3.07
1947	1303	10	0.222	1.73	0.76	3.11	3.03
1968	1254	11	0.244	1.68	0.69	3.10	3.00
1935	1090	12	0.267	1.63	0.62	3.04	2.98
1973	937	13	0.289	1.58	0.56	2.97	2.95
1975	855	14	0.311	1.53	0.49	2.93	2.92
1952	804	15	0.333	1.48	0.43	2.91	2.90
1938	719	16	0.356	1.44	0.37	2.86	2.88
1957	716	17	0.378	1.40	0.31	2.86	2.85
1974	714	18	0.400	1.35	0.25	2.85	2.83
1960	671	19	0.422	1.31	0.20	2.83	2.81
1945	623	20	0.444	1.27	0.14	2.79	2.78
1949	583	21	0.467	1.23	0.08	2.77	2.76
1946	507	22	0.489	1.20	0.03	2.70	2.74
1937	487	23	0.511	1.16	-0.03	2.69	2.72
1969	430	24	0.533	1.12	-0.08	2.63	2.69
1965	425	25	0.556	1.08	-0.14	2.63	2.67
1976	399	26	0.578	1.05	-0.20	2.60	2.65
1950	377	27	0.600	1.01	-0.25	2.58	2.62
1978	360	28	0.622	0.97	-0.31	2.56	2.60
1944	348	29	0.644	0.94	-0.37	2.54	2.58
1951	348	30	0.667	0.90	-0.43	2.54	2.55
1953	328	31	0.689	0.86	-0.49	2.52	2.53
1962	306	32	0.711	0.83	-0.55	2.49	2.50
1959	286	33	0.733	0.79	-0.62	2.46	2.48
1966	277	34	0.756	0.75	-0.68	2.44	2.45
1971	276	35	0.778	0.71	-0.75	2.44	2.42
1970	260	36	0.800	0.67	-0.83	2.42	2.39
1954	242	37	0.822	0.63	-0.91	2.38	2.36
1943	218	38	0.844	0.58	-0.99	2.34	2.33
1948	197	39	0.867	0.53	-1.08	2.30	2.29
1964	162	40	0.889	0.49	-1.19	2.21	2.25
1955	140	41	0.911	0.43	-1.30	2.15	2.20
1939	140	42	0.933	0.37	-1.43	2.15	2.15
1963	116	43	0.956	0.30	-1.60	2.06	2.08
1956	49	44	0.978	0.21	-1.83	1.69	1.99


4.00
3.50
3.00
2.50
2.00
1.50
Log(Q)
0.802
0.602
0.402
0.202
0.002
Exceedence probability

RELIABILITY OF ANALYSIS

Confidence Limits: Statistical estimate often presented with a range, or confidence interval within which the true value can be expected to lie.

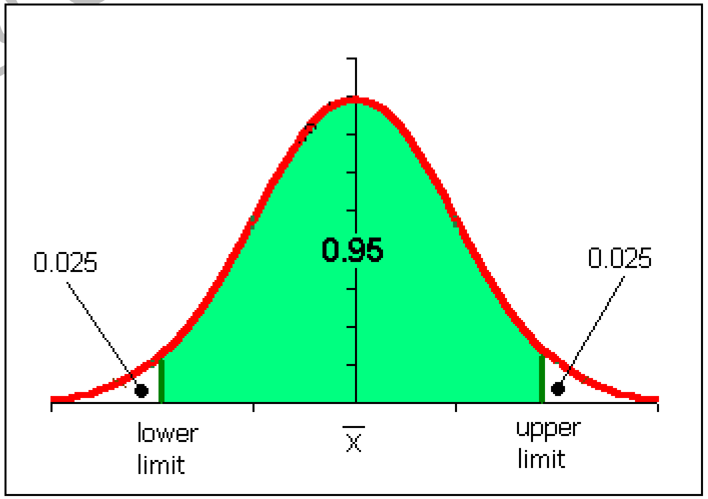
Confidence Level β

Significant Level α $\qquad \alpha = \frac{1-\beta}{2}$ $\qquad$ i.e. is $\beta = 90\%$, then $\alpha = 0.05$

For a return period T, the upper limit $U_{T,\alpha}$ and lower limit $L_{T,\alpha}$ may be specified as

$$U_{T,\alpha} = \bar{y} + s_y K_{T,\alpha}^U$$

$$U_{T,\alpha} = \bar{y} + s_y K_{T,\alpha}^L$$



$$K_{T,\alpha}^U = \frac{K_T + \sqrt{K_T^2 - ab}}{a} \qquad K_{T,\alpha}^L = \frac{K_T - \sqrt{K_T^2 - ab}}{a}$$

$$a = 1 - \frac{z_\alpha^2}{2(n-1)} \qquad b = K_T^2 - \frac{z_\alpha^2}{n}$$

z_α is the standard normal variable with exceedence probability α

Problem: Determine 90% confidence limits for the 100 year discharge using the following data

Logarithmic mean =3.639, standard deviation = 0.4439, coefficient of skewedness = -0.64 for 16 years of data

Solution:

$\beta = 0.9 \qquad \alpha = 0.05$

z_α has the exceedence probability of 0.05. This cumulative probability is 0.95.

$z_\alpha = 1.645$

Considering log-normal distribution, the $K_{100} = 1.843$

$$K_{T,\alpha}^{U} = \frac{K_T + \sqrt{K_T^2 - ab}}{a} = \frac{1.843 + \sqrt{1.843^2 - ab}}{a} = 2.7714$$

$$a = 1 - \frac{z_\alpha^2}{2(n-1)} = 1 - \frac{1.645^2}{2(16-1)} = 0.909799$$

$$K_{T,\alpha}^{L} = \frac{K_T - \sqrt{K_T^2 - ab}}{a} = \frac{1.843 - \sqrt{1.843^2 - ab}}{a} = 1.2804$$

$$b = K_T^2 - \frac{z_\alpha^2}{n} = 1.843^2 - \frac{1.645^2}{16} = 3.227522$$

$$U_{T,\alpha} = \bar{y} + s_y K_{T,\alpha}^{U} = 3.639 + 0.4439 \times 2.7714 = 4.869225$$

$$L_{T,\alpha} = \bar{y} + s_y K_{T,\alpha}^{L} = 3.639 + 0.4439 \times 1.2804 = 4.207211$$

Testing the Goodness of Fit

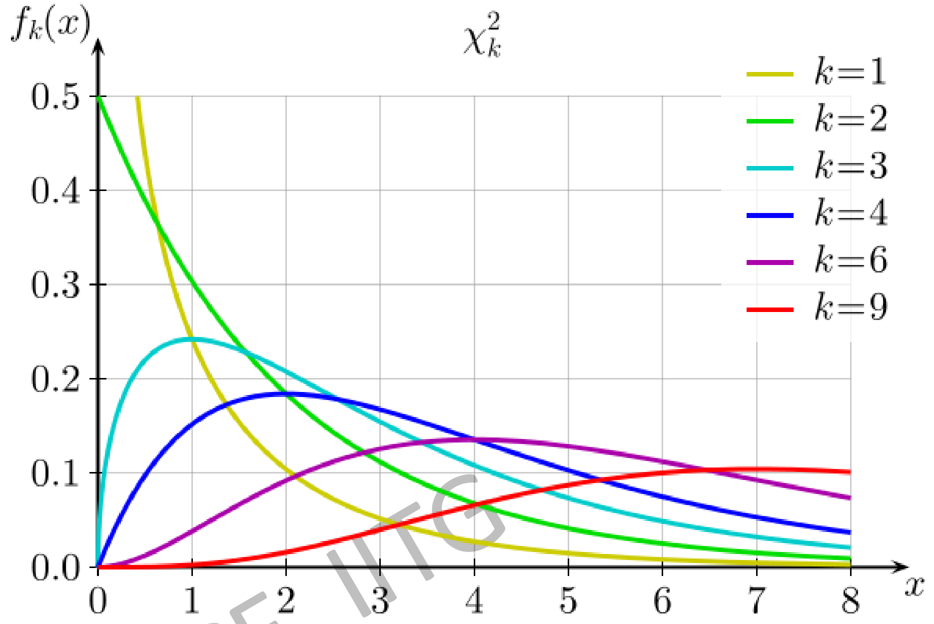
χ^2 test is used to test the goodness of fitting

$$\chi^2 = \sum_{i=1}^{m} \frac{n[f_s(x_i) - p(x_i)]^2}{p(x_i)}$$

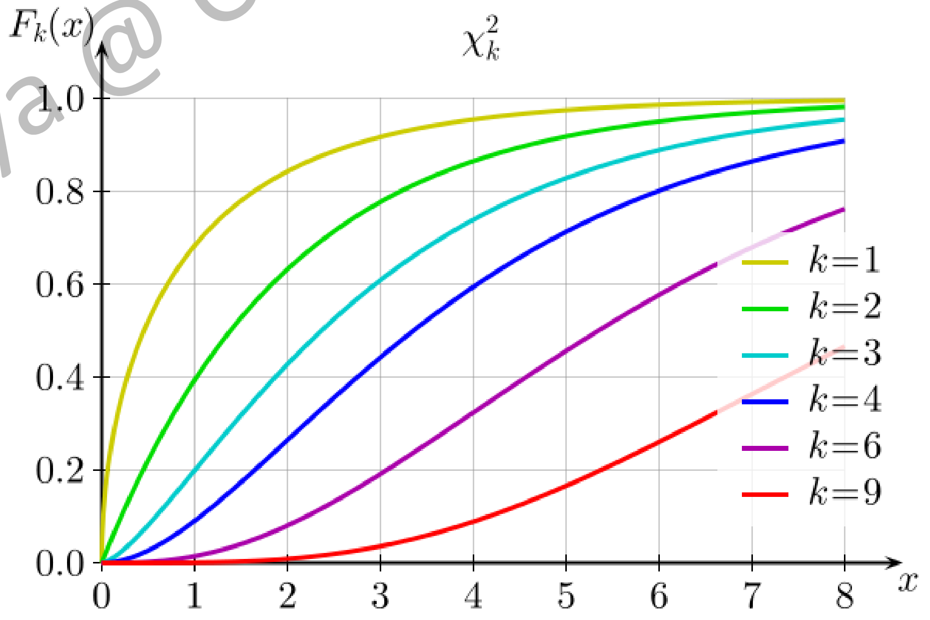
Where m is the number of interval

$nf_s(x_i)$ is the observed number of occurrence and $np(x_i)$ is the corresponding expected number of occurrence in the interval i.

For describing χ^2 test, χ^2 probability distribution must be defined



$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{\frac{k}{2}-1} e^{-x/2}$$



$$\mathrm{F}(x) = \frac{1}{\Gamma(k/2)} \gamma(k/2, x/2)$$

Degree of freedom $v = m - p - 1$

TABLE IV

Chi-Square (χ^2) Distribution

Degrees of Freedom	Area to the Right of Critical Value							
	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
1	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892

Check the goodness of fit on Normal Distribution

$$F(z) = B \qquad \text{for } z < 0$$

$$F(z) = 1 - B \qquad \text{for } z \geq 0$$

$$B = \frac{1}{2}[1 + 0.196854|z| + 0.115194|z|^2 + 0.000344|z|^3 + 0.019527|z|^4]^{-4}$$

Interval	Range	n_i
1	20	1
2	25	2
3	30	6
4	35	14
5	40	11
6	45	16
7	50	10
8	55	5
9	60	3
10	65	1
		69
Mean	39.77	
SD	9.17	

Interval	Range	n_i	$f_s(x_i)$	$F_s(x_i)$	z_i	B	$F(x_i)$	$p(x_i)$	χ^2
1	20	1	0.0145	0.0145	-2.1559	0.0154	0.0154	0.0154	0.004084
2	25	2	0.0290	0.0435	-1.6107	0.0535	0.0535	0.0380	0.147864
3	30	6	0.0870	0.1304	-1.0654	0.1436	0.1436	0.0901	0.007632
4	35	14	0.2029	0.3333	-0.5202	0.3012	0.3012	0.1577	0.895245
5	40	11	0.1594	0.4928	0.0251	0.4901	0.5099	0.2087	0.801553
6	45	16	0.2319	0.7246	0.5703	0.2840	0.7160	0.2061	0.22287
7	50	10	0.1449	0.8696	1.1156	0.1325	0.8675	0.1515	0.01965
8	55	5	0.0725	0.9420	1.6609	0.0482	0.9518	0.0843	0.115501
9	60	3	0.0435	0.9855	2.2061	0.0136	0.9864	0.0346	0.159163
10	65	1	0.0145	1.0000	2.7514	0.0032	0.9968	0.0104	0.108361
		69	1						2.481923
Mean	39.77								
SD	9.17								

$$\boxed{z_i = \frac{x_i - \pi}{\sigma}}$$

Degree of freedom $v = m - p - 1 = 10 - 2 - 1 = 7$ $\qquad$ $\chi^2_{7,1-0.95} = 14.067$

Since $\chi^2_{7,1-0.95}$ is greater than χ^2_c, the null hypothesis can not be rejected