CE 513: Statistical Methods in Civil Engineering

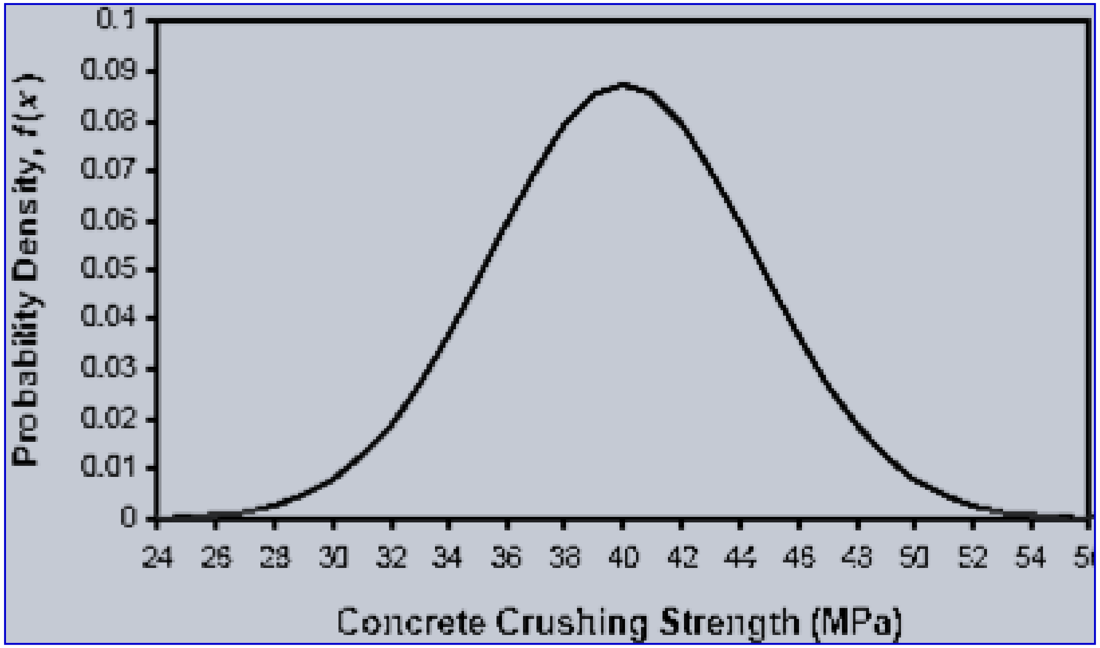
Lecture- 4: Continuous RV

Dr. Budhaditya Hazra
Room: N-307
Department of Civil Engineering

Continuous RVs

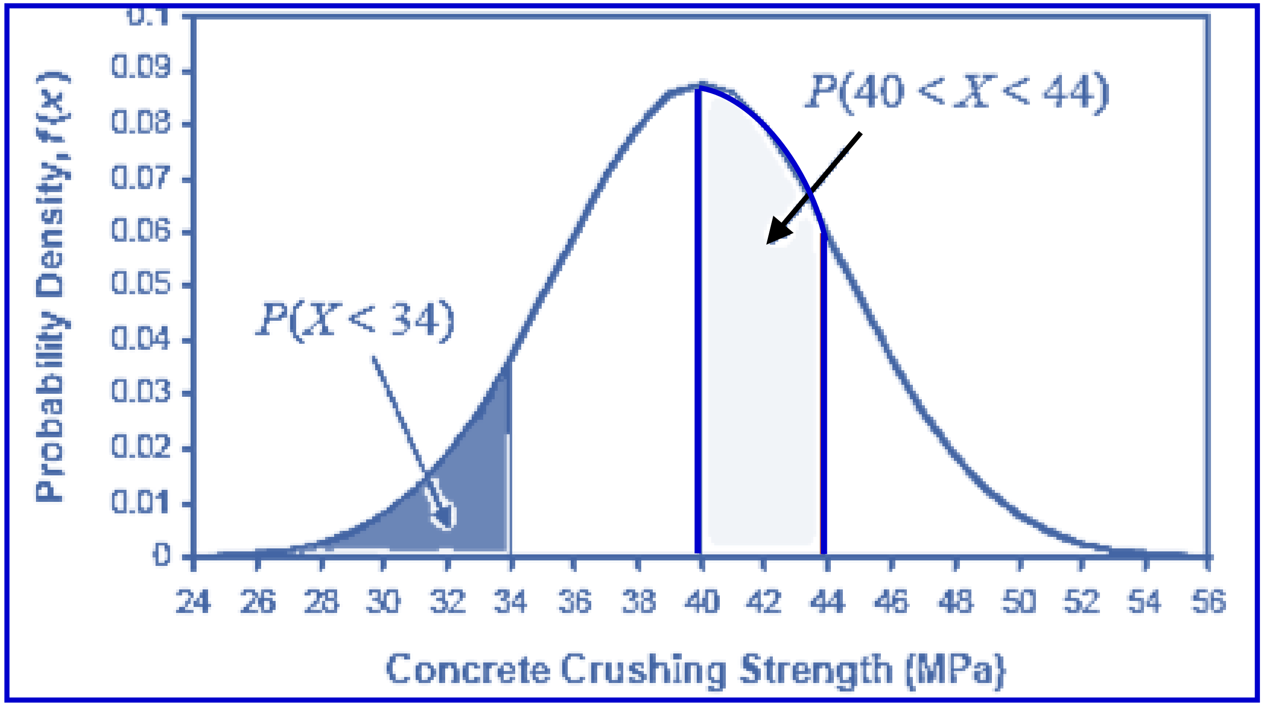
- A continuous random variable can assume any value within a given range e.g. Concrete crushing strength

- The probability content of a continuous random variable is described by the probability density function(PDF)



Continuous RVs

- The **probability** associated with the random variable in a given **range** is represented by the **area under the PDF**



Total area = 1.0

CDF

The **cumulative distribution function (CDF)**

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

- The CDF is equal to cumulative probability (ranges between 0 and 1)
- It is apparent from above that the PDF is the first derivative of the CDF

$$f_X(u) = \left.\frac{dF_X(x)}{dx}\right|_{x=u}$$

Properties of $f_X(x)$

1. $f_X(x) \geq 0$
2. $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$
3. $f_X(x)$ is piecewise continuous.
4. $P(a < X \leq b) = \int_a^b f_X(x)\, dx$

If X is a continuous r.v., then

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b)$$

$$= \int_a^b f_X(x)\, dx = F_X(b) - F_X(a)$$

CDF & Quantile function

- In some cases, we may be interested in finding out what is the value of the random variable for a given probability
- Probabilistic bounds that are important for design purposes
  - The result is called the **percentile** or quantile value
  - For example, the value of the random variable associated with 95 % (cumulative) probability is the 95th percentile value

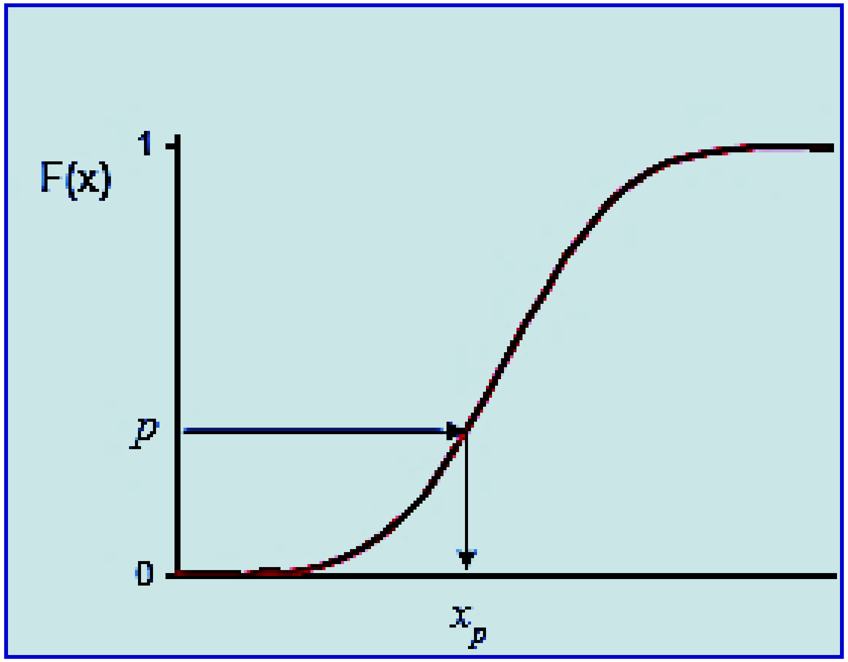
CDF & Quantile function

To estimate the percentile values, we must **invert** the CDF as :

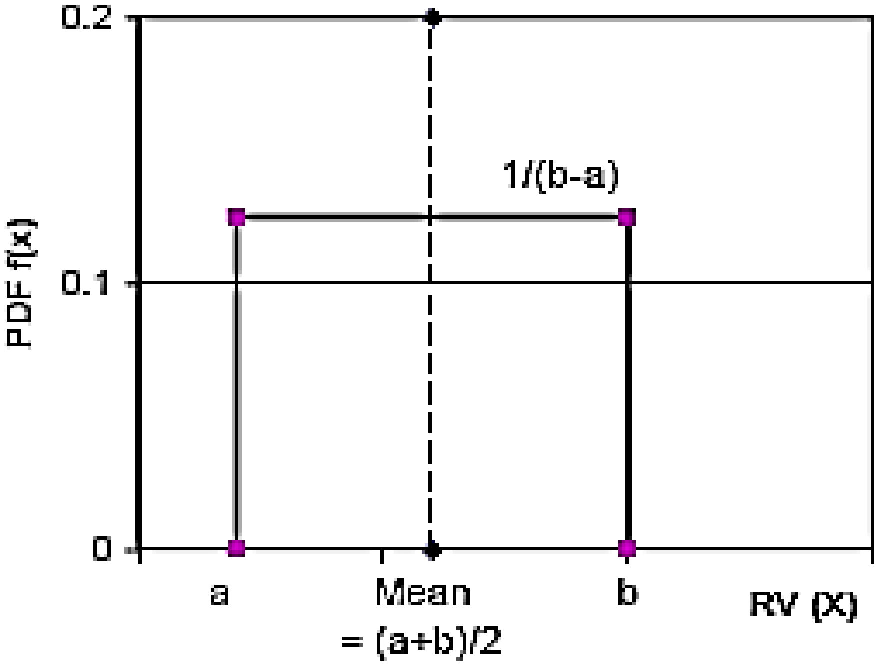
$$F_X(x) = p$$

$$\Downarrow$$

$$x_p = F_X^{-1}(p)$$



Uniform distribution



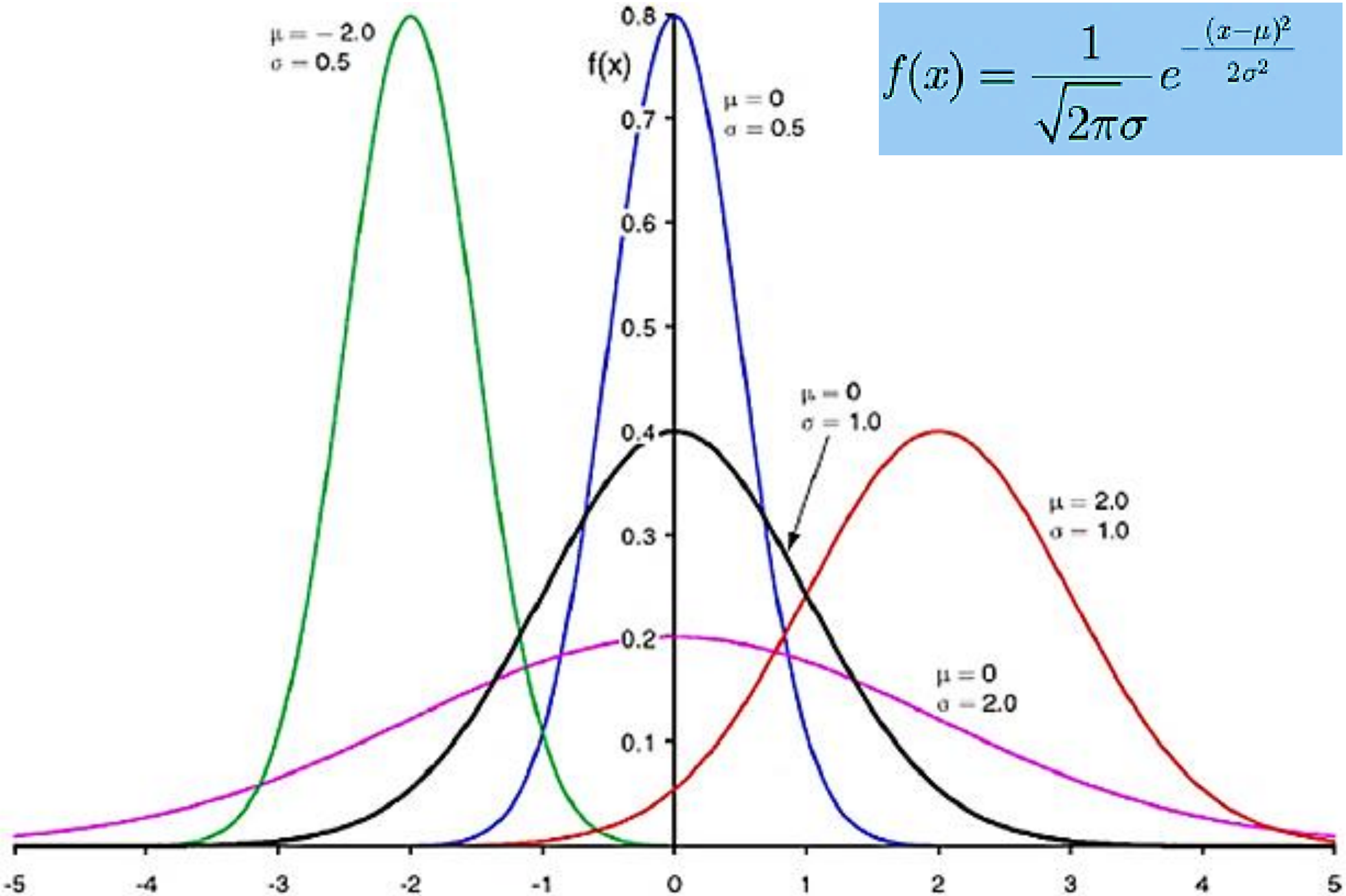
$$\text{Mean}: \quad \mu = \frac{(a+b)}{2}$$

$$\text{Variance}: \; \sigma^2 = \frac{(b-a)^2}{12}$$

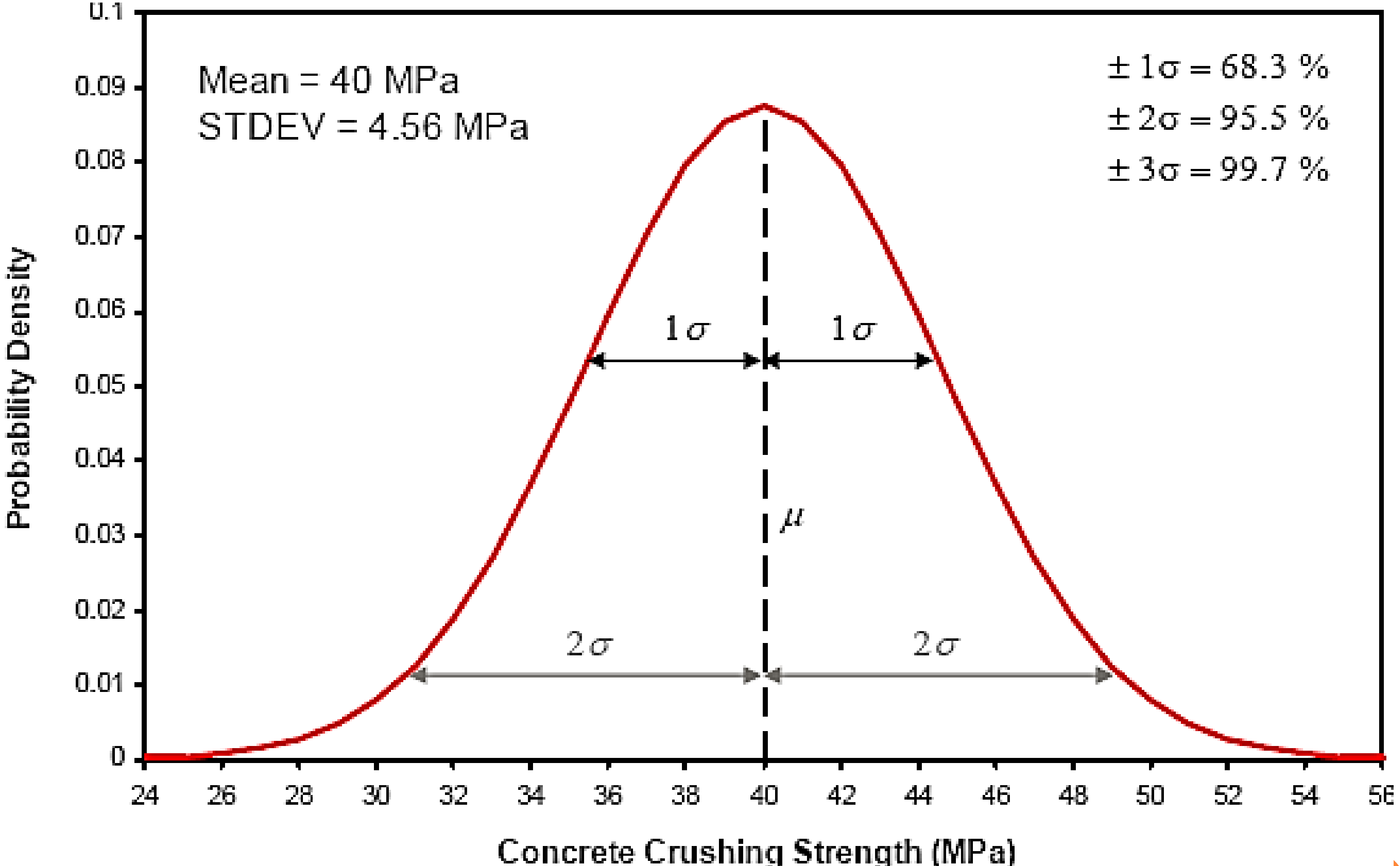
- It is the simplest distribution
- It is the most uncertain distribution between a & b

Normal distribution

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Normal distribution



Standard normal distribution

The **Standard Normal** variate is used to transform the original random variable *x* into standard format as

$$s = \frac{x - \mu}{\sigma}$$

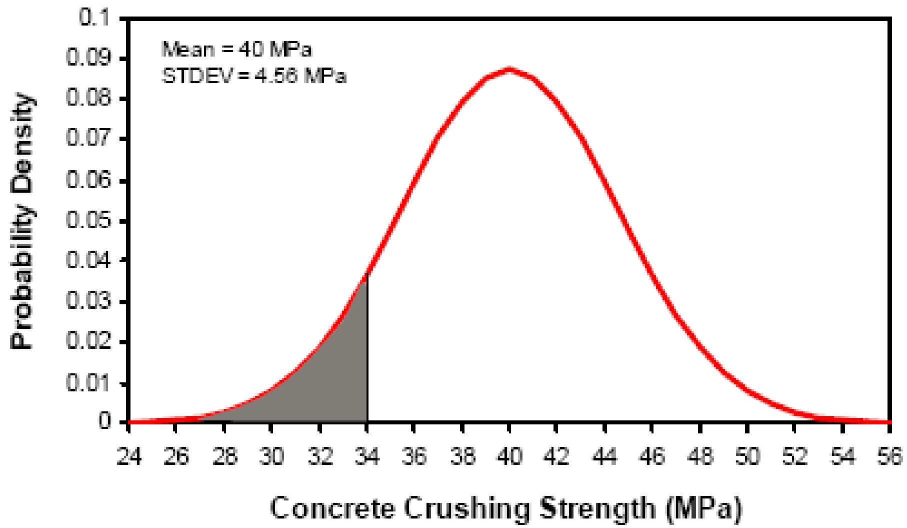
- The Standard Normal distribution is denoted as *N*(0,1)and has a **mean** of **zero** and **standard deviation** equal to **one**
- Because of its wide use, the CDF of the Standard Normal variate is denoted as $\Phi(s)$

Example: A reliability problem

A concrete column is expected to support a stress of 34 MPa.

- Assuming the Normal distribution for concrete strength, what is the probability of failure?
- The sample mean and standard deviation computed from tests are equal to 40 Mpa and 4.56 MPa

Soln: Probability of failure is the area under the Normal PDF



- The probability that the concrete strength is less than or equal to the applied stress (34 MPa) is obtained using the Standard Normal CDF as

$$P(X \leq 34) = \Phi\left(\frac{34-40}{4.56}\right) = \Phi(-1.316) = 0.094$$

- Therefore, given an estimated average value of 40 Mpa from the 35 laboratory tests with a standard deviation of 4.56 MPa, the probability of failure is **9.4 %**

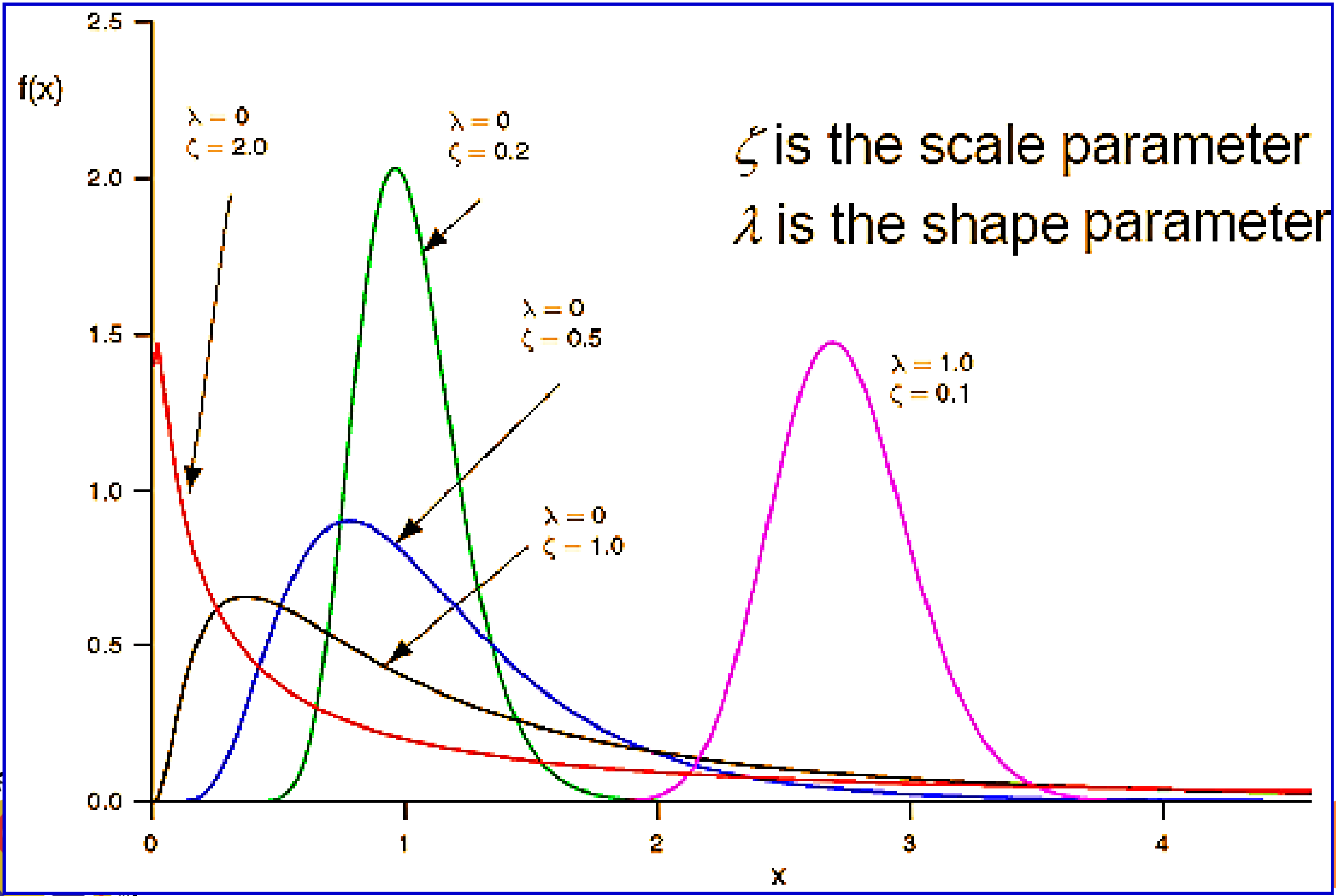
Log-Normal distribution

- The logarithmic or **Log-Normal distribution** is used when the random variable **cannot** take on a **negative** value
- A random variable follows the **Log-Normal** distribution if the **logarithm** of the random variable is **Normally** distributed
- **ln (*X*)** follows the Normal distribution; ***X*** =>follows the Lognormal distribution

$$f(x) = \frac{1}{\sqrt{2\pi}x\zeta} e^{-\frac{(\ln x - \lambda)^2}{2\zeta^2}} \qquad x \geq 0; \zeta > 0$$

Indian Institute of Technology Guwahati • भारतीय प्रौद्योगिकी संस्थान गुवाहाटी •

Log-Normal distribution



Log-Normal distribution

- The Log-Normal distribution is related to the Normal distribution, and can be evaluated using the **Standard Normal** distribution as

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx = \Phi\left(\frac{\ln x - \lambda}{\zeta}\right)$$

- The distribution parameters are related to the Normal distribution parameters as

$$\lambda = \ln(\mu) - \frac{1}{2}\zeta^2$$

$$\zeta = \sqrt{\ln(1+\delta^2)}$$

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

Log-Normal distribution

$$\lambda = \ln(\bar{x}) - \frac{1}{2}\zeta^2$$

$$\zeta = \sqrt{\ln\left(1 + \frac{s^2}{\bar{x}^2}\right)}$$

The distribution parameters are :

- Shape parameter λ= Mean of ln(x)
- Scale parameter ζ= STDEV of ln(x)

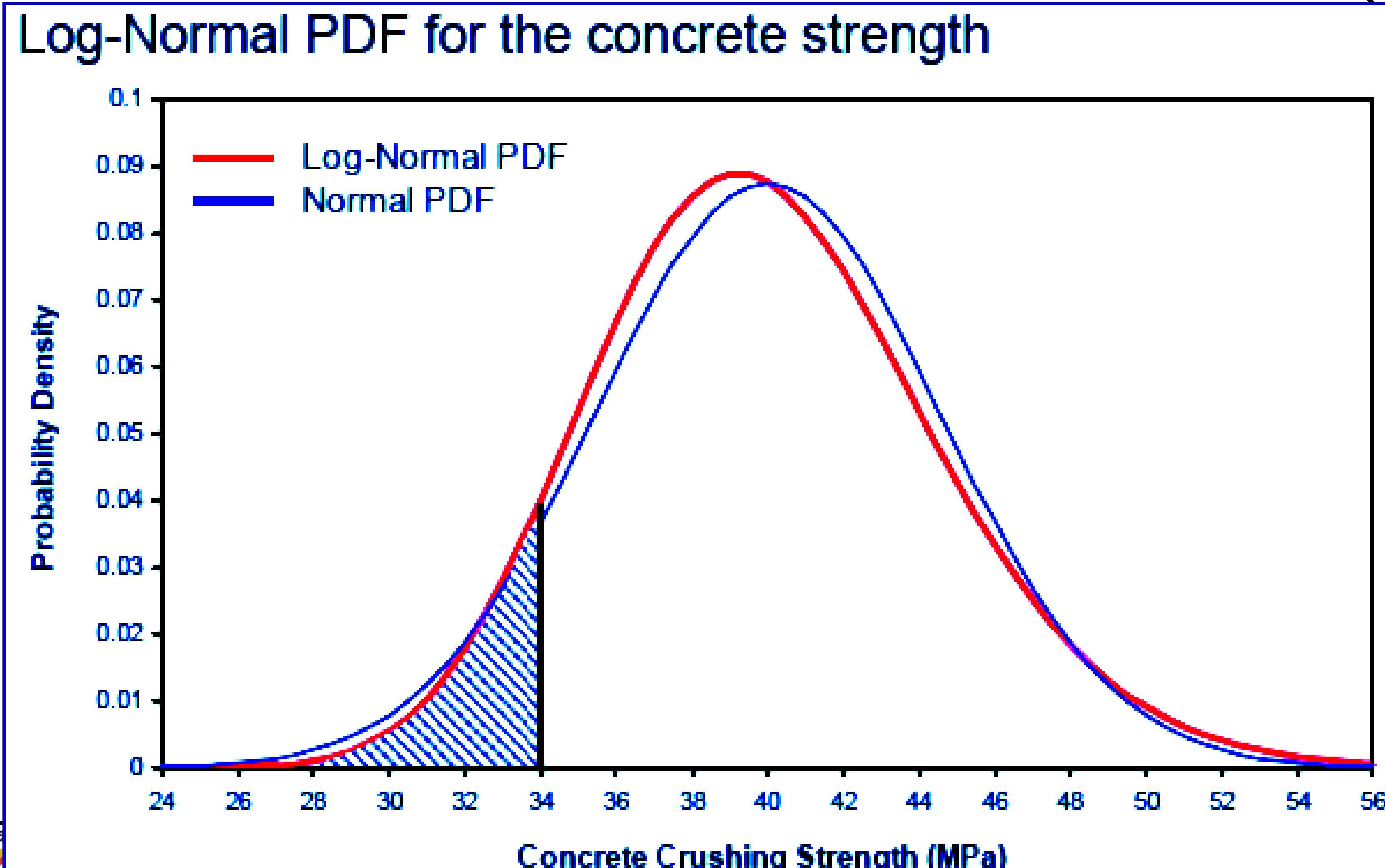
Log-Normal distribution

Assuming the concrete strength is described by the Log-Normal distribution, what is the probability that the concrete strength is less than or equal to 34 MPa?

Soln: The lognormal distribution parameters are :

$$\zeta = \sqrt{\ln\left(1 + \frac{s^2}{\bar{x}^2}\right)} = \sqrt{\ln\left(1 + \frac{(4.56)^2}{(40.0)^2}\right)} = 0.114$$

$$\lambda = \ln(\bar{x}) - \frac{1}{2}\zeta^2 = \ln(40.0) - \frac{1}{2}(0.114)^2 = 3.682$$

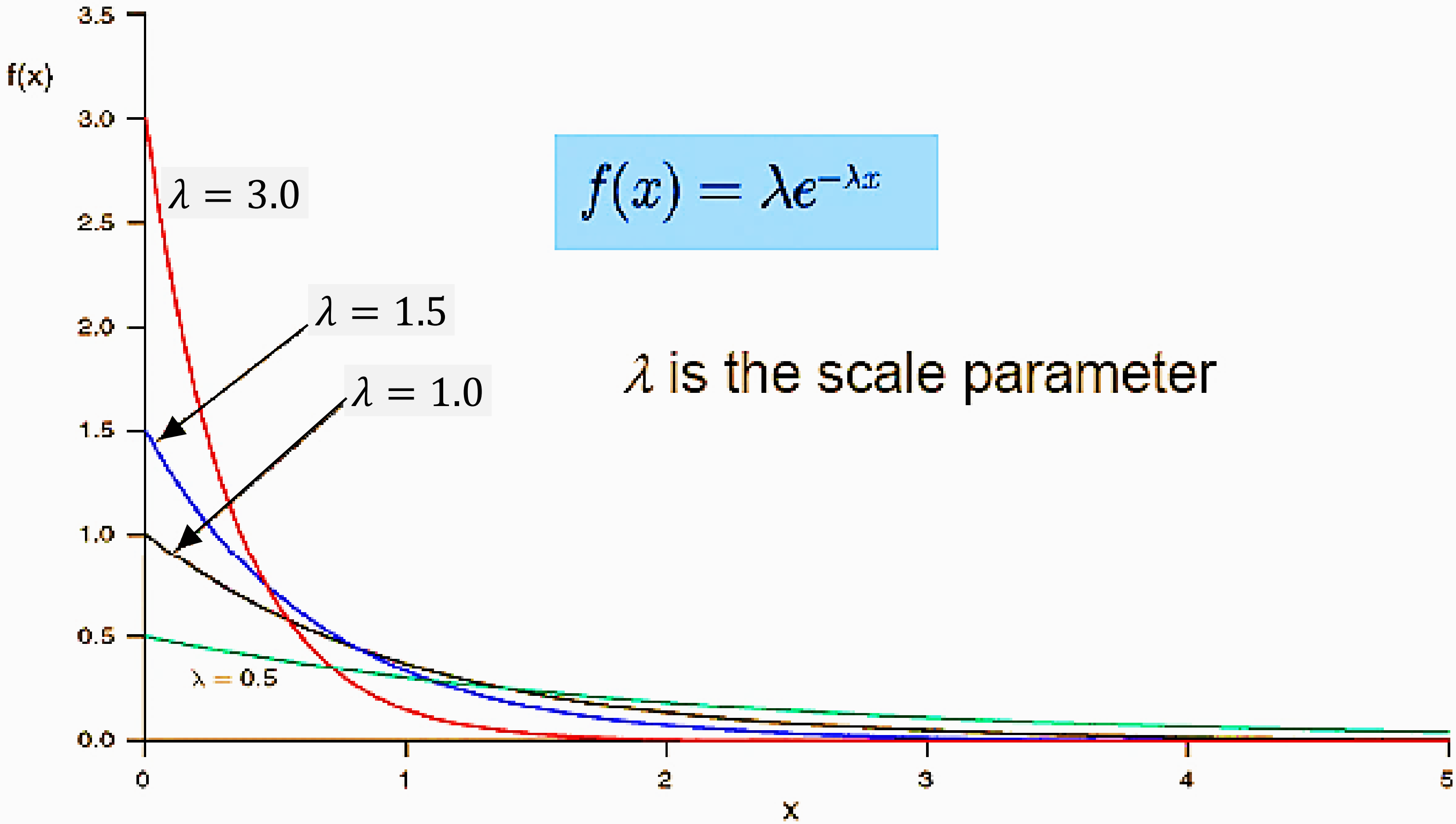

Log-Normal PDF for the concrete strength
Log-Normal PDF
Normal PDF
Probability Density
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56
Concrete Crushing Strength (MPa)

- The probability that the concrete strength is less than or equal to 34 Mpa is obtained using the Standard Normal CDF as

$$P(X \leq 34) = \Phi\left(\frac{\ln(34) - \lambda}{\zeta}\right) = \Phi\left(\frac{\ln(34) - 3.682}{0.114}\right) = \mathbf{0.085}$$

- Assuming the concrete strength follows the Log-Normal distribution (i.e., the LOG of the concrete strength follows the Normal distribution), there is a **8.5 %**chance that the concrete strength is less than or equal to 34 MPa

Exponential distribution



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

λ is the scale parameter

Exponential distribution

The cumulative distribution function (CDF) of the Exponential distribution is given by:

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

- The distribution parameters can be estimated using the sample data (i.e. sample statistics)
- The scale parameter λ is equal to or simply the reciprocal of the sample average

Exponential distribution

Assuming the concrete strength is described by the exponential distribution, what is the probability that the concrete strength is less than or equal to 34 MPa?

$$\lambda = \frac{1}{\bar{x}} = \frac{1}{40} = 0.025$$

$$P(X \leq 34) = F(34) = 1 - e^{-0.025(34)} = 0.573$$

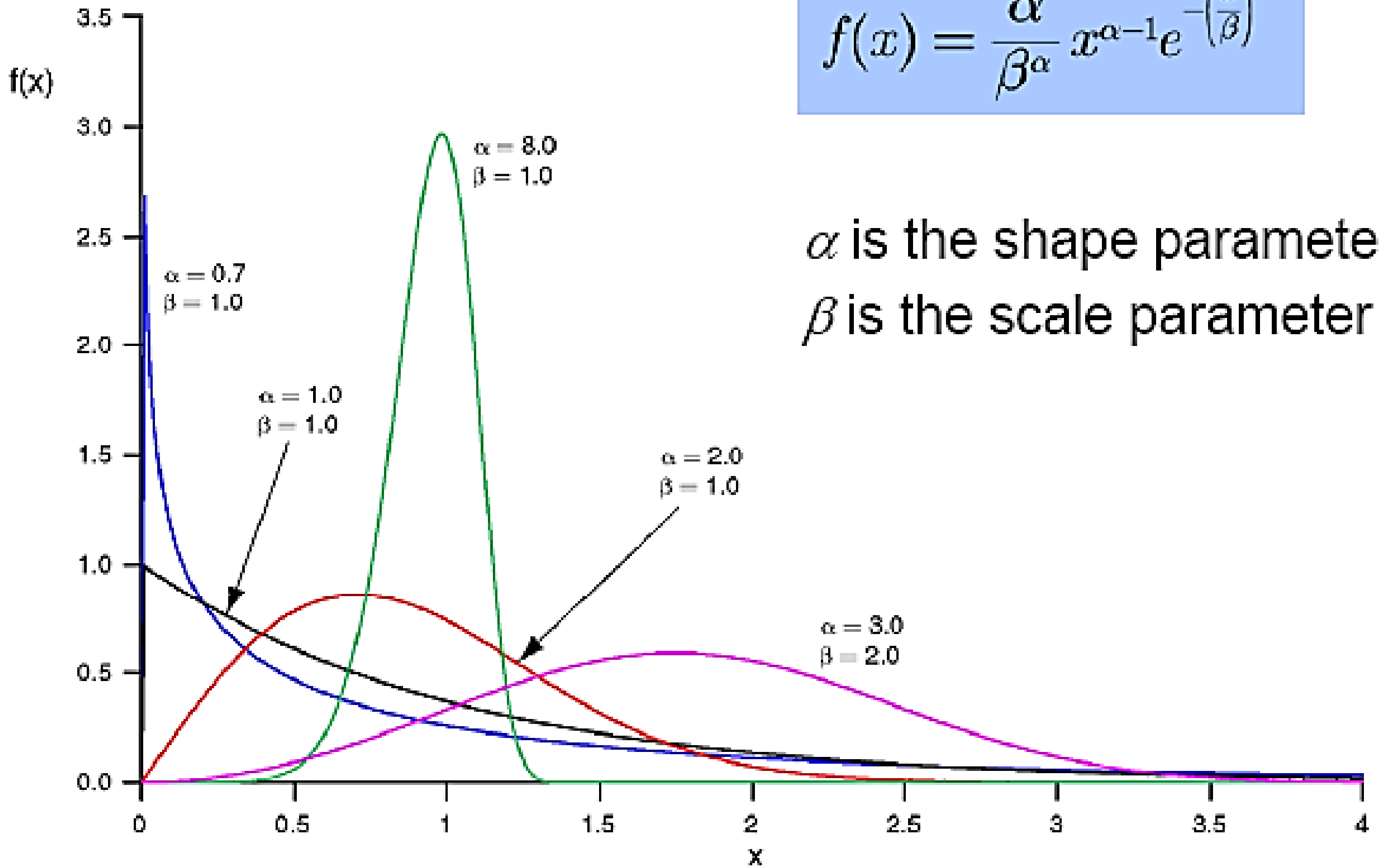
Weibull distribution

- The Weibull probability distribution is a very flexible distribution
  - Due to the shape parameter
- It is used extensively in modeling the time to failure distribution analysis
- The Weibull distribution is derived theoretically as a form of an Extreme Value Distribution
- It is also used to model extreme events like strong winds, hurricanes, typhoons etc

Weibull distribution

The **probability density function (PDF)** of the Weibull distribution is

$$f(x) = \frac{\alpha}{\beta^{\alpha}} x^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}$$



α is the shape parameter

β is the scale parameter

Weibull distribution

- The **cumulative distribution function (CDF)** of the Weibull distribution is

$$F(x) = 1 - e^{-\left(\frac{x}{\beta}\right)^{\alpha}}$$

- The distribution parameters can be estimated from the sample statistics using the Method of Moments as
  - Sample Average = $\bar{x} = \beta\Gamma\left(1 + \frac{1}{\alpha}\right)$
  - Sample STDEV = $s = \beta\sqrt{\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma\left(1 + \frac{1}{\alpha}\right)^{2}}$

Reliability problem using Weibull distribution

Assuming the concrete strength is described by the Weibull distribution, what is the probability that the concrete strength is less than or equal to 34 MPa?

Reliability problem using Weibull distribution

Solution:

- From before, the sample mean and standard deviation were equal to 40 MPa and 4.56 MPa, respectively
- The Weibull distribution parameters are obtained from

$$\beta\Gamma\left(1+\frac{1}{\alpha}\right)=40 \quad \text{and} \quad \beta\sqrt{\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma\left(1+\frac{1}{\alpha}\right)^2}=4.56$$

- Solving 2 equations and 2 unknowns (using the SOLVER function in Excel) results in α = 10.59 and β = 41.95

Alternate approach: Solve for α and β using nonlinear equation solution techniques

$$1 + s^2/\bar{x}^2 = \frac{\Gamma\left(1+\frac{2}{\alpha}\right)}{\Gamma^2\left(1+\frac{1}{\alpha}\right)}$$

→ Main equation to be solved

Use bisection method to solve for α

Task: Solve the above problem in MATLAB and verify using Excel goal-seek solver

Submit the assignment solution by Monday aug-14

The probability that the concrete strength is less than or equal to 34 MPa is therefore

$$P(X \leq 34) = F(34) = 1 - e^{-\left(\frac{34}{41.95}\right)^{10.59}} = 0.103$$

Using MATLAB command:

p = wblcdf (34, 41.95, **10.59**) = 0.1024

Inverse Weibull distribution

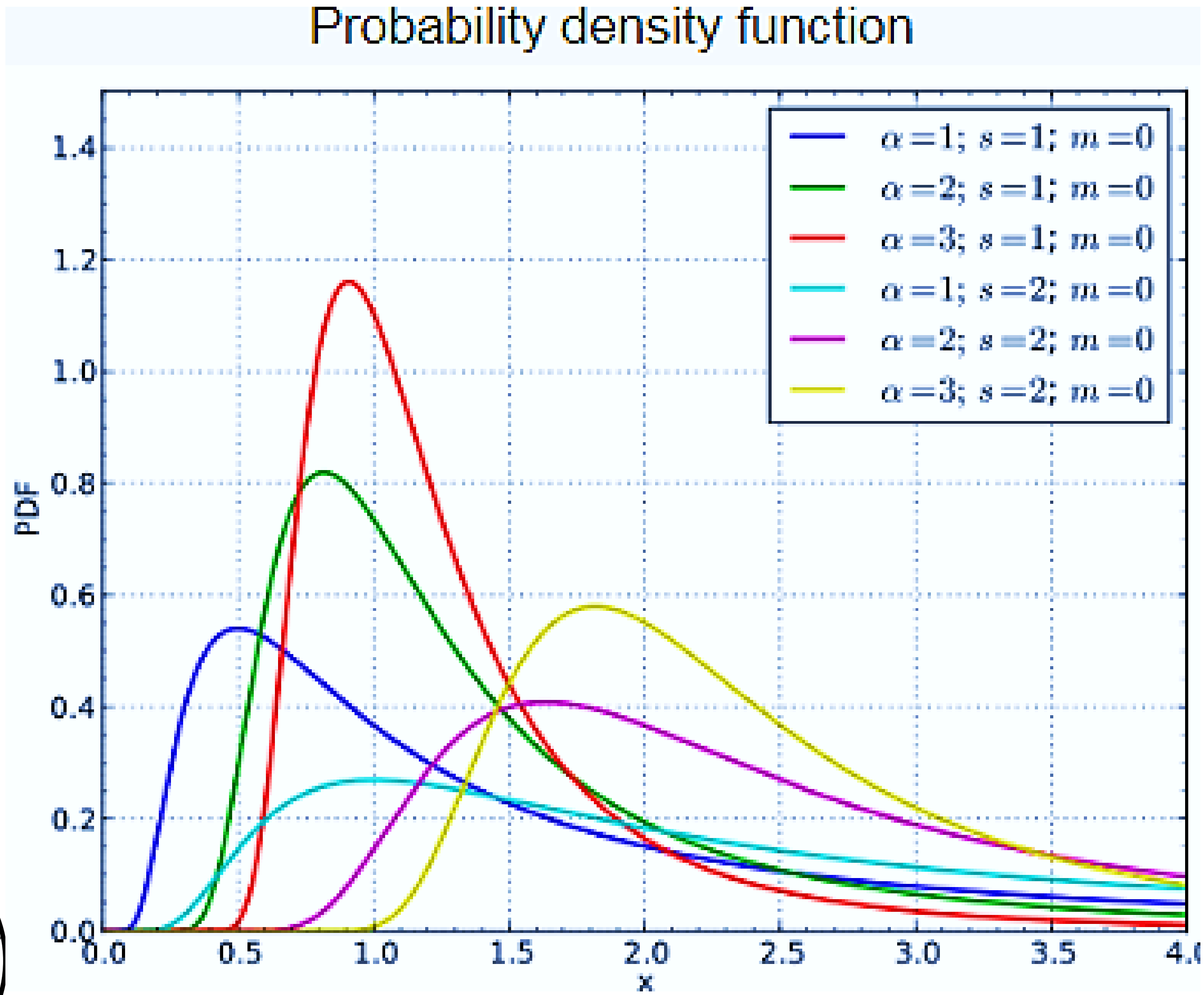
The **Fréchet distribution**, also known as inverse Weibull distribution, is a special case of the generalized extreme value distribution. It has the cumulative distribution function

$$\Pr(X \le x) = e^{-x^{-\alpha}} \text{ if } x > 0.$$

where $\alpha > 0$ is a shape parameter. It can be generalised to include a location parameter m (the minimum) and a scale parameter $s > 0$ with the cumulative distribution function

$$\Pr(X \le x) = e^{-\left(\frac{x-m}{s}\right)^{-\alpha}} \text{ if } x > m.$$

Inverse Weibull distribution



Gamma distribution

- The Gamma distribution is another flexible probability distribution that may offer a good model to some sets of failure data
- The Gamma distribution arises theoretically as the time to first fail distribution for a system with standby Exponentially distributed backups
- The Gamma distribution is commonly used in Bayesian reliability applications e.g. using prior information to update the constant (Exponential) repair rate for a system following a homogeneous Poisson process (HPP) model

Gamma distribution

Similar to the Weibull distribution, there are many different variations of writing the Gamma distribution

(alternative format)

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

- The **probability density function (PDF)** is

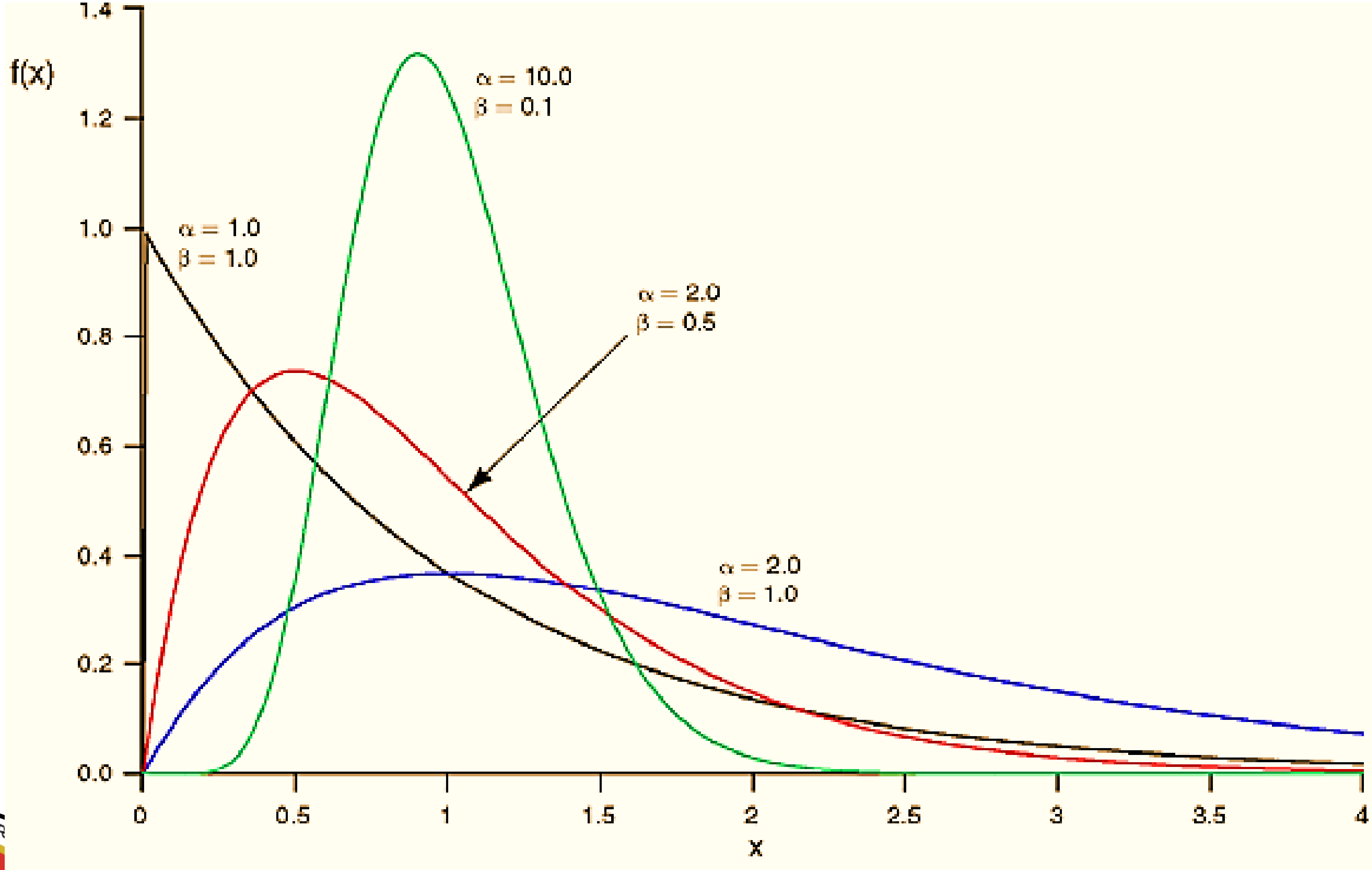
$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \qquad 0 \le x \le \infty; \;\; \alpha, \beta > 0$$

α is the shape parameter; **β** is the scale parameter

- When **α = 1** the Gamma distribution reduces to the Exponential distribution with **1/β= λ**

CDF: $F(x) = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)}$

Gamma distribution



Gamma distribution

Task: Find out the mean and the variance for the gamma distributed random variable, using the form of $f(x)$ given underneath

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$$

$$0 \leq x \leq \infty; \ \ \alpha, \beta > 0$$

CE 513: Statistical Methods in Civil Engineering

Lecture- 6: Bivariate RV

Dr. Budhaditya Hazra
Room: N-307
Department of Civil Engineering

Multiple RVs

- Consider 2 RVs X and Y
- If the RVs are discrete, then the joint probability distribution is described by the joint probability mass function(PMF)
- $p_{X,Y}(x,y) = P[(X = x) \cap (Y = y)]$

CDF:

$$F_{X,Y}(x,y) = \sum_{x_i < x} \sum_{y_i < y} p_{X,Y} = P[(X \leq x) \cap (Y \leq y)]$$

Continuous RVs

- Consider 2 continuous RVs X and Y

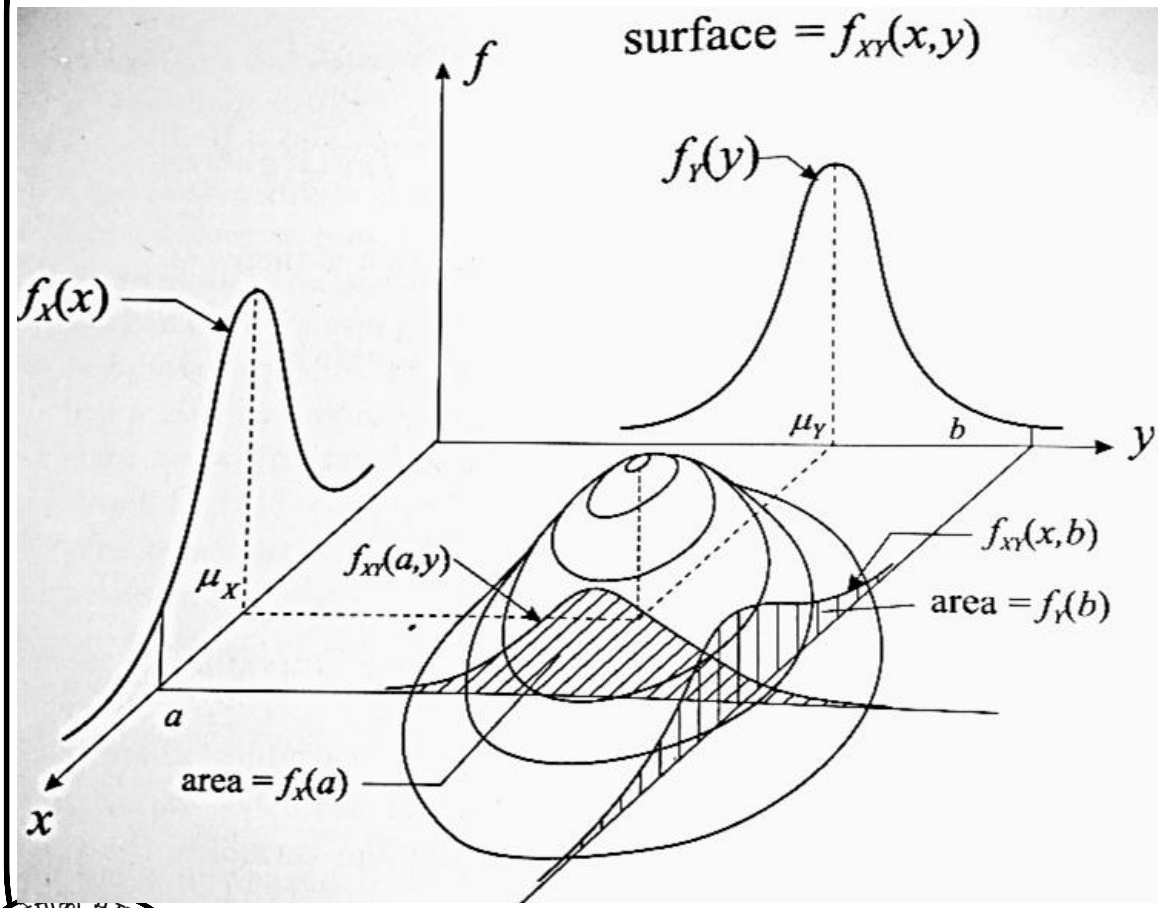
$$f_{XY}(x,y)dx\,dy \simeq \Pr\left(x < X \leq x+dx, y < Y \leq y+dy\right),$$

$$\Pr\left(a < X \leq b, c < Y \leq d\right) = \int_c^d \int_a^b f_{XY}(x,y)dx\,dy.$$

$$F_{XY}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{XY}(u,v)du\,dv.$$

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x\,\partial y}.$$

Continuous RV



CDF

$$F(x,\,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} p(u,\,v)\,dv\,du$$

Marginal PDF

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$$

Moments of continuous RV

$$E[XY] = \iint_{-\infty}^{\infty} xy\, f_{X,Y}(x,y)dxdy$$

Cov (X,Y)= $\sigma_{xy} = E[(X-\mu_x)(Y-\mu_Y)]$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_x)(y-\mu_y) f_{X,Y}(x,y)dx\, dy$$

$$\rho_{xy} = \frac{Cov(X,Y)}{\sigma_x \sigma_y} = \frac{E[(X-\mu_x)(Y-\mu_y)]}{\sigma_x \sigma_y}$$

Properties of moments

- $E[aX + b] = a\,E[X] + b$
- $Var[X] = E[X^2] - (E[X])^2$
- $Var[aX + b] = a^2 Var(X)$
- Cov(X,Y)= $E[XY] - E[X]E[Y]$
- Var(X+Y)=Var(X)+Var (Y) + 2Cov(X,Y)

Independence

Recall

$$P(A|B)=\frac{P(A\cap B)}{P(B)};P(B)\neq 0.$$

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

Define $A=\{X\leq x\}$ and $B=\{Y\leq y\}$

$$X\perp Y\Rightarrow P(X\leq x\cap Y\leq y)=P(X\leq x)P(Y\leq y)$$

$$\Rightarrow P_{XY}(x,y)=P_X(x)P_Y(y)$$

$$\Rightarrow p_{XY}(x,y)=p_X(x)p_Y(y)$$

Bi-variate Gaussian distribution

$$p(x, y) = \frac{1}{2\pi} \cdot \frac{1}{|\mathbf{S}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{\nu} - \boldsymbol{\mu}_\nu)^T \mathbf{S}^{-1}(\boldsymbol{\nu} - \boldsymbol{\mu}_\nu)\right]$$

$$\mathbf{S} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}, \quad \boldsymbol{\nu} = \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{and} \quad \boldsymbol{\mu}_\nu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$$

Bivariate Gaussian distribution Alternate Form

X and Y are said to be jointly Gaussian if

$$p_{XY}(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{(1-r_{XY}^2)}}\exp\left[-\frac{1}{2(1-r_{XY}^2)}\left\{\left(\frac{x-\eta_X}{\sigma_X}\right)^2+\left(\frac{y-\eta_Y}{\sigma_Y}\right)^2-\frac{2r_{XY}(x-\eta_X)(y-\eta_Y)}{\sigma_X\sigma_Y}\right\}\right]$$

$$-\infty<x<\infty;-\infty<y<\infty$$

$$\textbf{Notes:}\ \begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left[\begin{pmatrix} \eta_X \\ \eta_Y \end{pmatrix}\ \begin{pmatrix} \sigma_X^2 & r_{XY}\sigma_X\sigma_Y \\ r_{XY}\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right]$$

$\begin{pmatrix} \sigma_X^2 & r_{XY}\sigma_X\sigma_Y \\ r_{XY}\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$ is known as the covariance matrix.

Example-1

The joint pdf of a bivariate r.v. (X, Y) is given by

$$f_{XY}(x, y) = \begin{cases} kxy & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

where k is a constant.

(*a*) Find the value of k.

(*b*) Are X and Y independent?

(*c*) Find $P(X + Y < 1)$.

Solution

How will you find k ?

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = k\int_0^1\int_0^1 xy\,dx\,dy = k\int_0^1 y\left(\left.\frac{x^2}{2}\right|_0^1\right)dy$$

$$= k\int_0^1 \frac{y}{2}\,dy = \frac{k}{4} = 1$$

Solution

How will you find marginal pdfs

$$f_X(x) = \begin{cases} \int_0^1 4xy\, dy = 2x & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = \begin{cases} 2y & 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Is $f_{XY}(x, y) = f_X(x) f_Y(y)$?

Solution

$$f_{XY}(x,y)=\begin{cases}4xy & 0<x<1,\ 0<y<1\\ 0 & \text{otherwise}\end{cases}$$

$$f_X(x)=2x \qquad 0<x<1$$

$$f_Y(y)=2y \qquad 0<y<1$$

Conditional densities

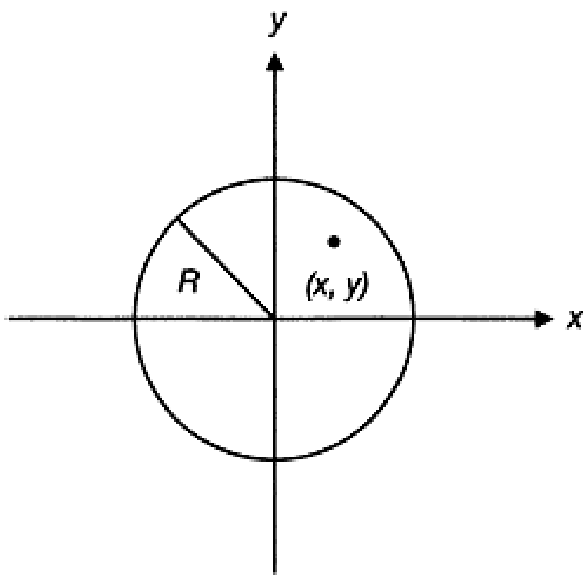
$$f_{Y|X}(y|x)=\frac{4xy}{2x}=2y \qquad 0<y<1, 0<x<1$$

$$f_{X|Y}(x|y)=\frac{4xy}{2y}=2x \qquad 0<x<1, 0<y<1$$

Example-2 new

Suppose we select one point at random from within the circle with radius R. If we let the center of the circle denote the origin and define X and Y to be the coordinates of the point chosen (Fig), then (X, Y) is a uniform bivariate r.v. with joint pdf given by

$$f_{XY}(x,y)=\begin{cases} k & x^2+y^2 \le R^2 \\ 0 & x^2+y^2 > R^2 \end{cases}$$



(a)

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = k \iint_{x^2+y^2\le R^2} dx\,dy = k(\pi R^2) = 1$$

Thus, $k = 1/\pi R^2$.

(b) the marginal pdf of X is

$$f_X(x) = \frac{1}{\pi R^2}\int_{-\sqrt{(R^2-x^2)}}^{\sqrt{(R^2-x^2)}} dy = \frac{2}{\pi R^2}\sqrt{R^2-x^2} \qquad x^2 \le R^2$$

Hence,

$$f_X(x) = \begin{cases} \dfrac{2}{\pi R^2}\sqrt{R^2-x^2} & |x| \le R \\ 0 & |x| > R \end{cases}$$

By symmetry, the marginal pdf of Y is

$$f_Y(y) = \begin{cases} \dfrac{2}{\pi R^2}\sqrt{R^2-y^2} & |y| \le R \\ 0 & |y| > R \end{cases}$$

Example-3 new

Let (X, Y) be a bivariate r.v. with the joint pdf

$$f_{XY}(x, y) = \frac{x^2 + y^2}{4\pi} e^{-(x^2+y^2)/2} \qquad -\infty < x < \infty, -\infty < y < \infty$$

Show that X and Y are not independent but are uncorrelated.

Example-3 new

$$f_X(x) = \frac{1}{4\pi}\int_{-\infty}^{\infty}(x^2 + y^2)\, e^{-(x^2+y^2)/2}\, dy$$

$$= \frac{e^{-x^2/2}}{2\sqrt{2\pi}}\left(x^2\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}dy + \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\, y^2 e^{-y^2/2}dy\right)$$

Noting that the integrand of the first integral in the above expression is the pdf of $N(0;\, 1)$ and the second integral in the above expression is the variance of $N(0;\, 1)$, we have

$$f_X(x) = \frac{1}{2\sqrt{2\pi}}(x^2 + 1)\, e^{-x^2/2} \qquad -\infty < x < \infty$$

Since $f_{XY}(x, y)$ is symmetric in x and y, we have

$$f_Y(y) = \frac{1}{2\sqrt{2\pi}}(y^2 + 1)\, e^{-y^2/2} \qquad -\infty < y < \infty$$

Now $f_{XY}(x, y) \neq f_X(x)\, f_Y(y)$, and hence X and Y are not independent.

Check Uncorrelated-ness

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx = 0$$

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y)\, dy = 0$$

since for each integral the integrand is an odd function.

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{XY}(x, y)\, dx\, dy = 0$$

The integral vanishes because the contributions of the second and the fourth quadrants cancel those of the first and the third. Thus, $E(XY) = E(X)E(Y)$, and so X and Y are uncorrelated.