CE 513: STATISTICAL METHODS IN CIVIL ENGINEERING

Lecture-1: Introduction & Overview

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

Schedule of Lectures

- Last class before puja break Sept-24 (Sunday): extra class
- Puja break: Sept-25 (Mon) to Oct-1 (Sun)
- 4 extra classes on weekends**:** **5 marks** bonus for full attendance
- Grading scheme: Midterm 30 %
- End term: 50%
- Surprise Quizzes : 20 %
- Lectures: **Mon** (8-9) **3102**; **Wed** (5 – 6:00 pm) **L4;**

  Tues (12-1:00) **3102**

INTROCUTION

- Principle aim of design: SAFETY
- Often this objective is non-trivial
- On occasions, structures fail to perform their intended function
- RISK is inherent
- Absolute safety can never be guaranteed for any engineering system; a probabilistic notion

A motivating example

F = 1 KN
EI = 10000000 Nm^2
L = 2 m

$$\Delta = \frac{FL^3}{3EI}$$

$$= 0.2667\ mm$$

Uncertainties

- Can we be always certain about EI ?
- For RCC, fixing a point or a single value of E

  is fraught with risks
- Are we always sure about I ? Or the dimensional

  properties ? Can be risky again
- In lot of practical applications, even F cannot be known

  for certain ?

Let's consider the cantilever beam example again, **now with some uncertainties**

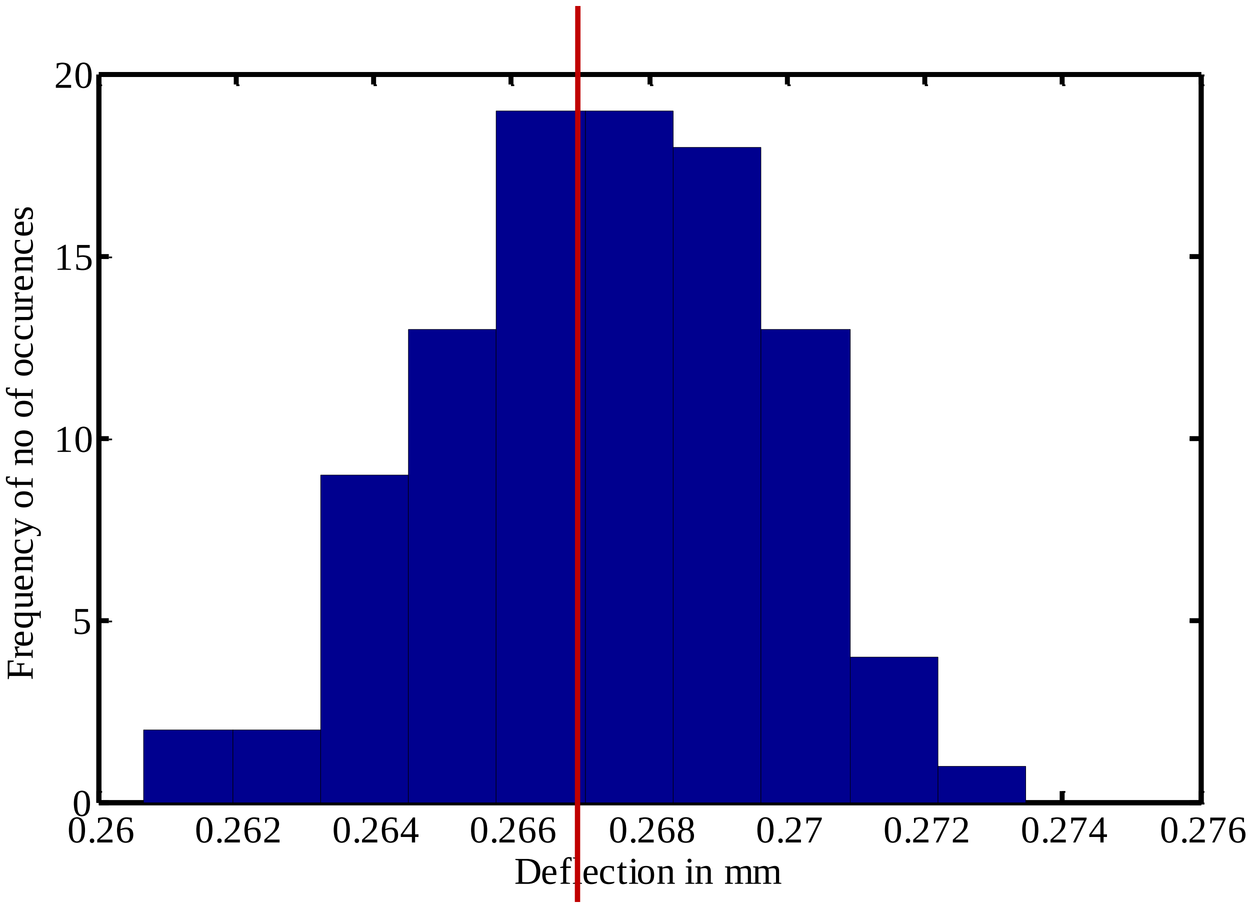
```
F=1+0.1*randn(100,1) ;    % F is normally distributed

EI=10^7+1000*randn(100,1);  % EI is normally distributed

L=2;

 for i=1:100
        delta(i)= F(i)*L^3/(3*EI(i));
end
```

The displacement becomes uncertain too

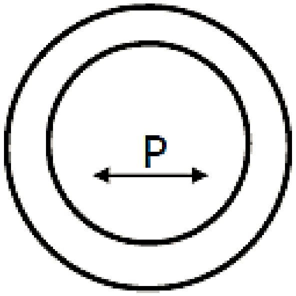


Mean value = 0.2674 mm

Practical example: Reliability based design

- Consider a pipe section with diameter D, thickness W, that is subjected to internal pressure P
- The hoop stress S in the pipe section is given as

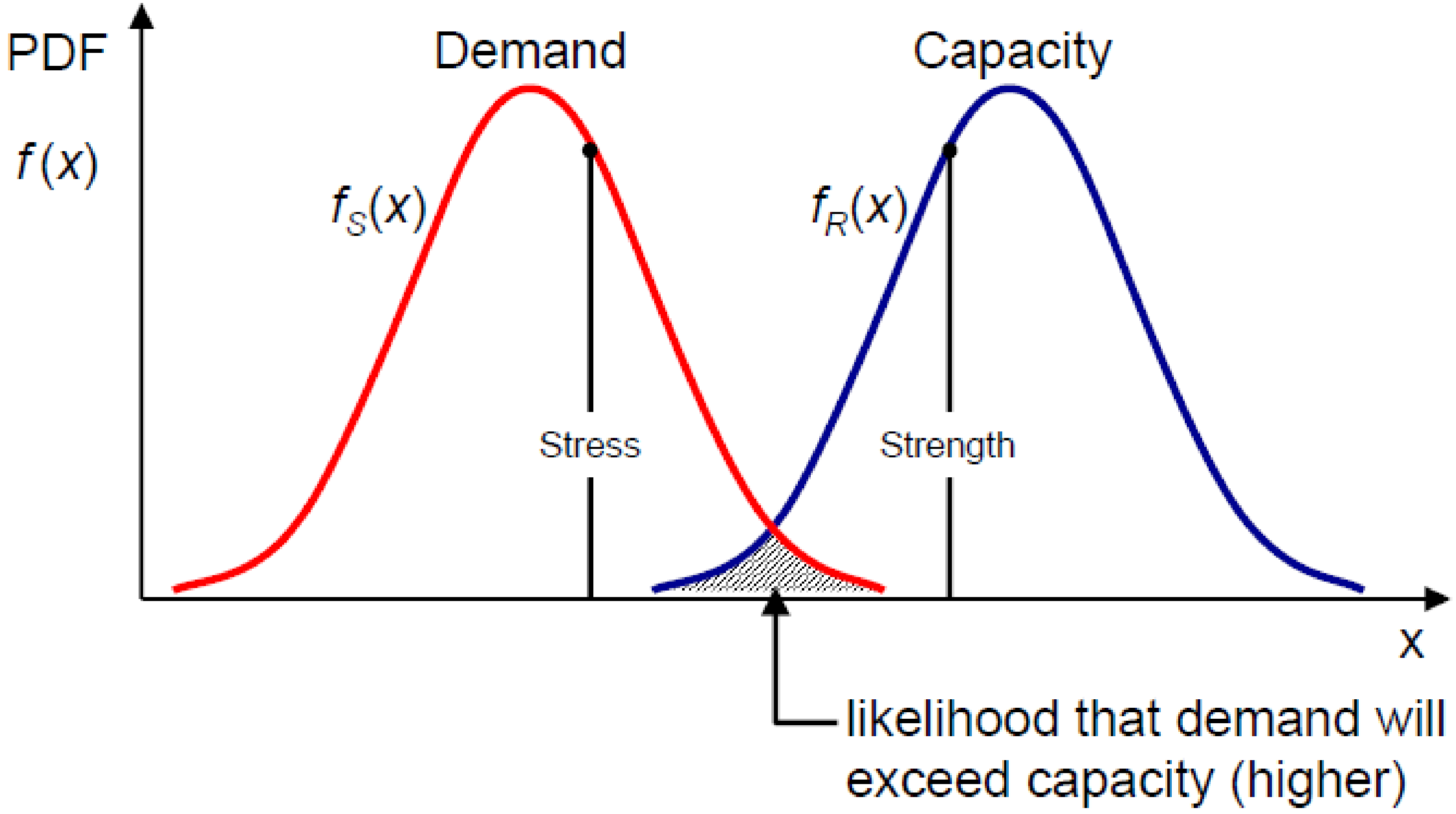
$$S = \frac{PD}{2W}$$



- The yield strength of the material is denoted by R
- Failure:

Stress (S) > Yield strength (R)

Reliability example



Probability

- Random variables are defined through the concept of probability
- Probability quantifies the variability in the **outcome** of an event, whose **exact** outcome cannot be predicted with certainty

Probability: *The likelihood of occurrence of an event relative to a set of alternative events.*

- Need to identify
  - A set (or range) of all possibilities (outcomes or sample space)
  - The event of interest

Classical definition

- Need to determine the likelihood associated with the occurrence of each event (= probability)

- Definition

For a game that has n equally likely outcomes, of which s outcomes correspond to "success" or winning, the probability of winning is given by s/n.

- All events have an equal likelihood of occurring
  - i.e. all outcomes are equally likely

- The classical probability concept is applicable to games of chance (i.e. gambling)

Sample space

- Consider a coin toss or rolling a die

Sample Space = 2 Events

"Heads"
or
"Tails"

Sample Space = 6 Events

{1,2,3,4,5,6}

- Or, tossing a quarter and a dime

H_Q T_Q H_D T_D

Sample Space = 4 Events

Example-1

When we roll a pair of balanced dice, what are the probabilities of getting (a) 3, (b) 2 or 12, or (c) 7?

Solution:

(a) There are two such outcomes (1,2) and (2,1). Thus, the probability is equal to 2/36 = **1/18**

(b) There are two such outcomes (1,1) and (6,6). Thus, the probability is equal to 2/36 = **1/18**

(c) There are six such outcomes (1,6), (2,5), (3,4), (6,1), (5,2) and (4,3). Thus, the probability is equal to 6/36 = **1/6**

Indian Institute of Technology Guwahati

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी

Issues with classical definition

- What is “equally likely”?
- What if not equally likely?

e.g.: what is the probability that sun would rise tomorrow ?

- No room for experimentation.

Frequency def

Frequency definition

If a random experiment has been performed *n* number of times and if *m* outcomes are favorable to event A, then the probability of event A is given by

$$P(A) = \lim_{n \to \infty} \frac{m}{n}$$

Issues

What is meant by limit here?

One cannot talk about probability without conducting an experiment

What is the probability that someone meets with an accident tomorrow?

Axioms of probability

Notions lacking definition

- Experiments
- Trials
- Outcomes

• An experiment is a physical phenomenon that is repeatable. A single performance of an experiment is called a trial. Observation made on a trial is called outcome.

• Axioms are statements commensurate with our experience. No formal proofs exist. All truths are relative to the accepted axioms.

Sample space

Sample space (Ω)

Set of all possible outcomes of a random experiment.

Examples

(1) Coin tossing: $\Omega=(h \quad t)$; Cardinality=2; finite sample space.

(2) Die tossing: $\Omega=(1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6)$; Cardinality=6; finite sample space.

(3) Die tossing till head appears for the first time:

$\Omega=(\mathrm{h} \quad \mathrm{th} \quad \mathrm{tth} \quad \mathrm{ttth} \quad \mathrm{tttth} \quad \cdots)$; Cardinality=$\infty$; countably infinite sample space.

(4) Maximum rainfall in a year: $\Omega=(0 \le X < \infty)$;

Cardinality=∞; uncountably infinite sample space.

Axioms of probability

1. The probability of an event is a real non-negative number

$$0 \le P(E) \le 1$$

2. The probability of a certain (or inevitable) event S is 1.0 (contains all the sample points in the sample space; i.e. is the sample space itself)

$$P(S) = 1.0$$

(conversely, the probability of an impossible event ϕ is zero)

3. Law of Addition

$$\sum_{k=1}^{N} P(E_k) = 1$$

Rigorously speaking the axioms of probability requires the knowledge of measure theory & sigma algebra

Problems that we will study

- **Random Processes:** Wind, earthquake & wave loads on engineering structures. Hydro-climatological processes like temperature & rainfall data
- **Stochastic Calculus:** If $F(x) = x^2\ dF \neq 2xdx$; where F(x) is a stochastic function of a random process **x**
- **Stochastic Differential Equations**: $\frac{dx}{dt} = f(x,t) + L(x,t)w(t)$; $w(t)$ is a realization of white noise
- **Monte Carlo Simulation**

Reference material

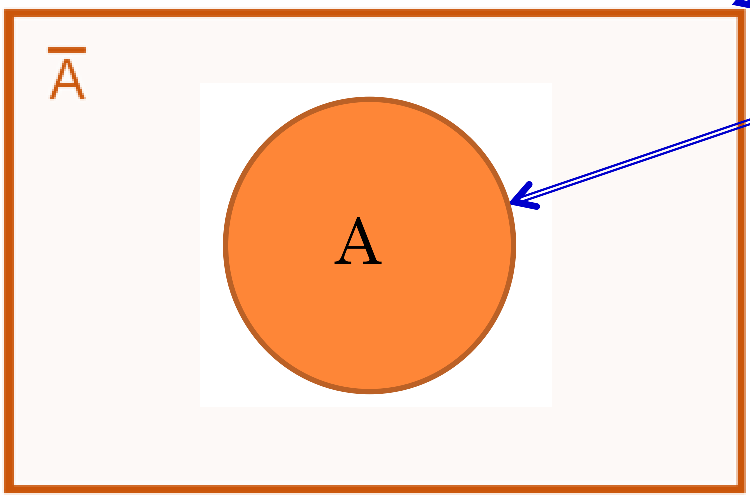
1. Probability, reliability and statistical methods in engineering design by **A. Halder & S. Mahadevan**
1. Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering **by** **Alfredo H-S. Ang & Wilson H. Tang**
2. Probability, Statistics, and Reliability for Engineers and Scientists by **Bilal M. Ayyub & Richard H. McCuen**
4. Probability, Random Variables, and Stochastic Processes: **Papoulis and Pillai**
5. From Elementary Probability to Stochastic Differential Equations with MAPLE® by **Sasha Cyganowski, Peter Kloeden** and **Jerzy Ombach**

CE 513: Statistical Methods in Civil Engineering

Lecture-2: Probability

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

Definitions & representations

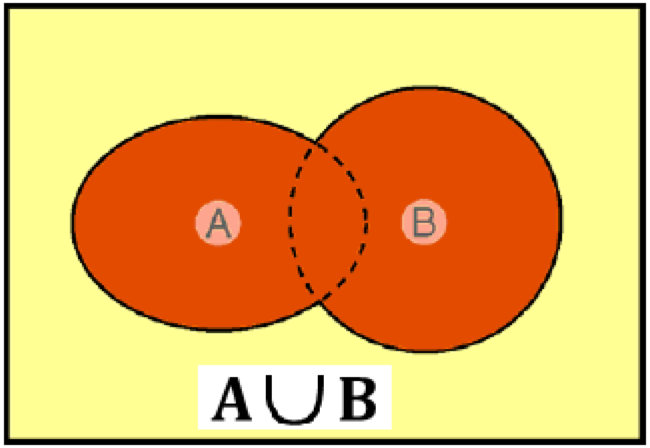


$$P(A) = \frac{area\ of\ A}{sample\ space}$$

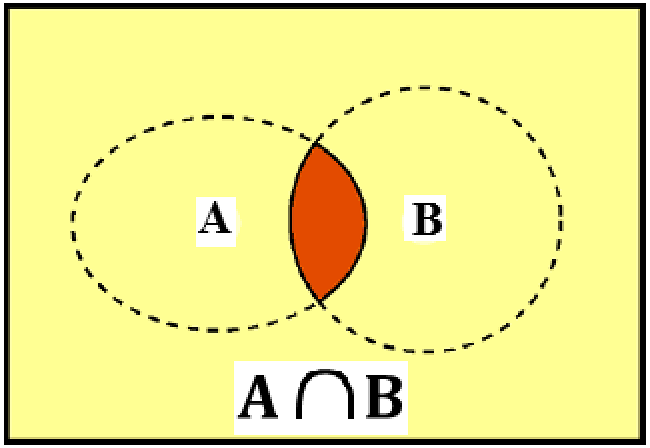
Complementary event $\bar{A}$

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

Intersection and Union



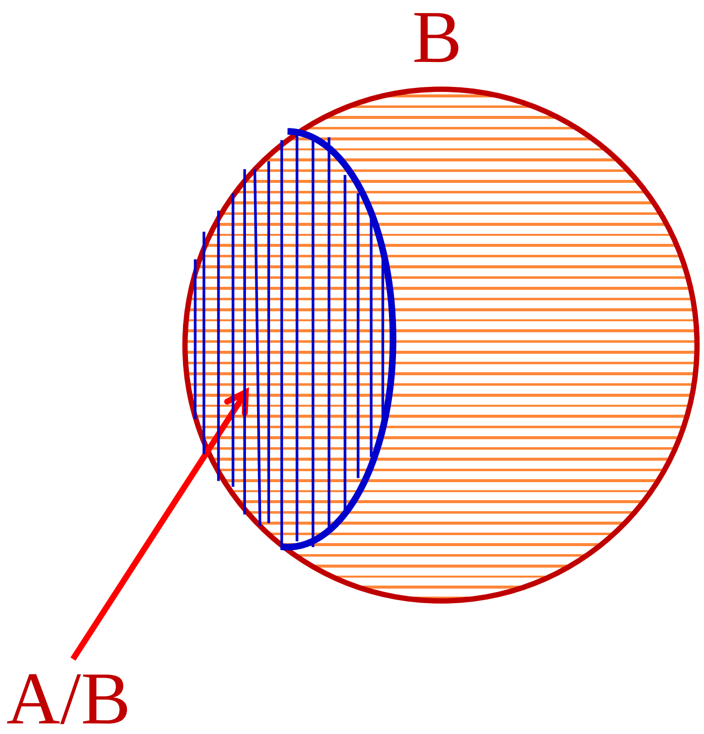
Either A or B or both occurs



Both A or B occur

What are mutually exclusive and mutually independent events ?

Conditional events

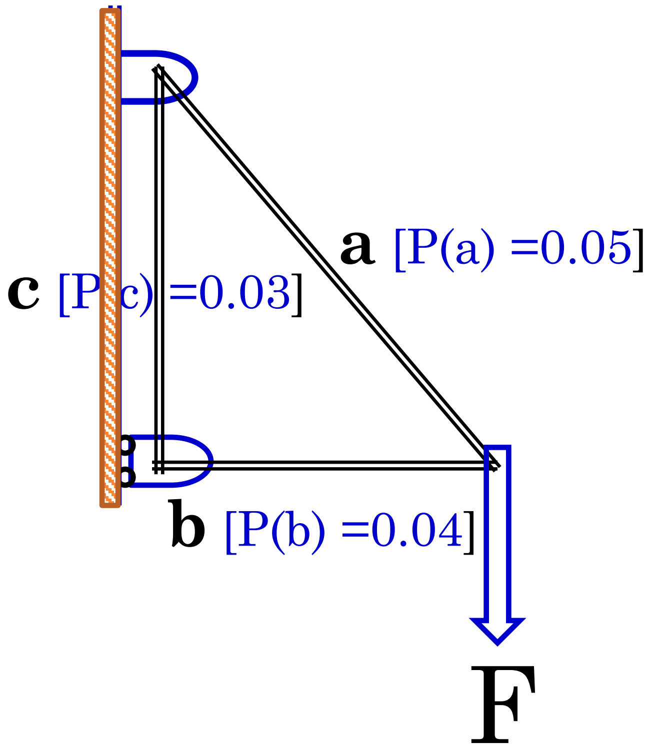


- Event B occurs first
- A occurs given that B has already occurred

$$P(A/B) = ?$$
$$= P(AB) / P(B)$$

Pay attention to the chalk-board notes for details

Example-2



Find the probability of failure of the truss ?

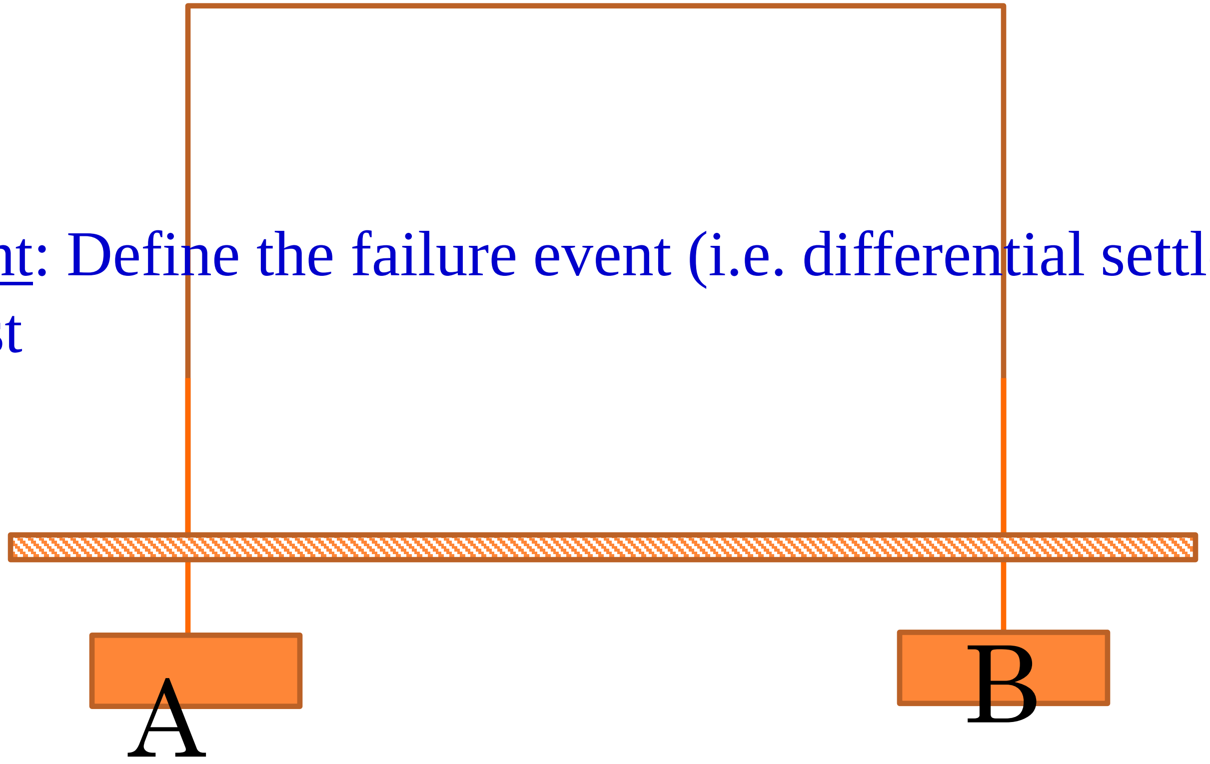
Assume: The failures of each of the members are mutually independent

Hint: Define the failure event first

Example-3

- Prob of settlement of each footing = 0.1
- Prob of settlement of each footing given the other one has settled = 0.8

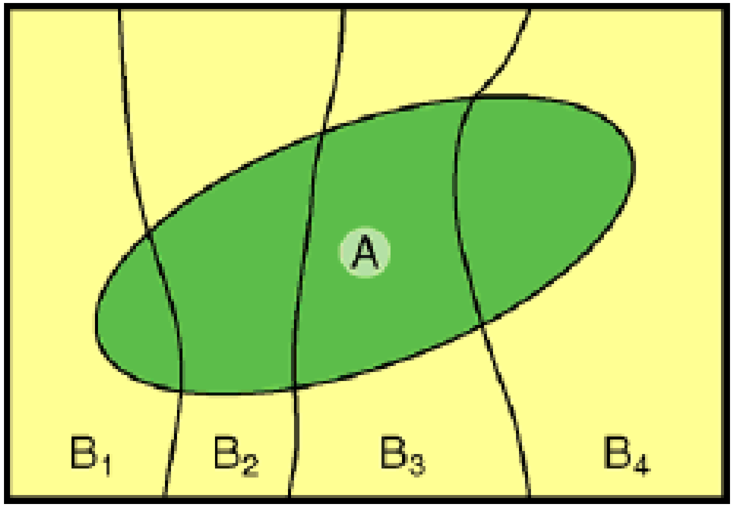
- **Find the prob of differential settlement**



Hint: Define the failure event (i.e. differential settlement) first

Theorem of total probability

- Sometimes probability of an event A cannot be assigned directly but can be assigned conditionally for a number of other events B_i
- B_i must be mutually exclusive and collectively exhaustive



$$P(A) = P(A \mid B_1)\cdot P(B_1) + P(A \mid B_2)\cdot P(B_2) + \ldots + P(A \mid B_n)\cdot P(B_n)$$
$$= \sum_{i=1}^{n} P(A \mid B_i)\cdot P(B_i)$$

Example-4

There is a possibility of rain or snow tomorrow but not both. The probability of rain is 40 % while the probability of snow is 60%. If it rains then the probability that I'll be late for my work is 20%. If it snows, however, the probability of being late for my work increases to 60%.

What is the probability that I will be late for work tomorrow?

Solution:

- Let R = event that it will rain tomorrow
  S = event that it will snow tomorrow
  L = event that I will be late for my work
- We are given
  $P(R) = 0.40$ and $P(S) = 0.60$
  $P(L \mid R) = 0.20$ and $P(L \mid S) = 0.60$

Using the rule of total probability

$$P(L) = P(L \mid R) \cdot P(R) + P(L \mid S) \cdot P(S)$$
$$= (0.20)(0.40) + (0.60)(0.60) = 0.44$$

Therefore, there is a 44 % chance that I will be late for work tomorrow...

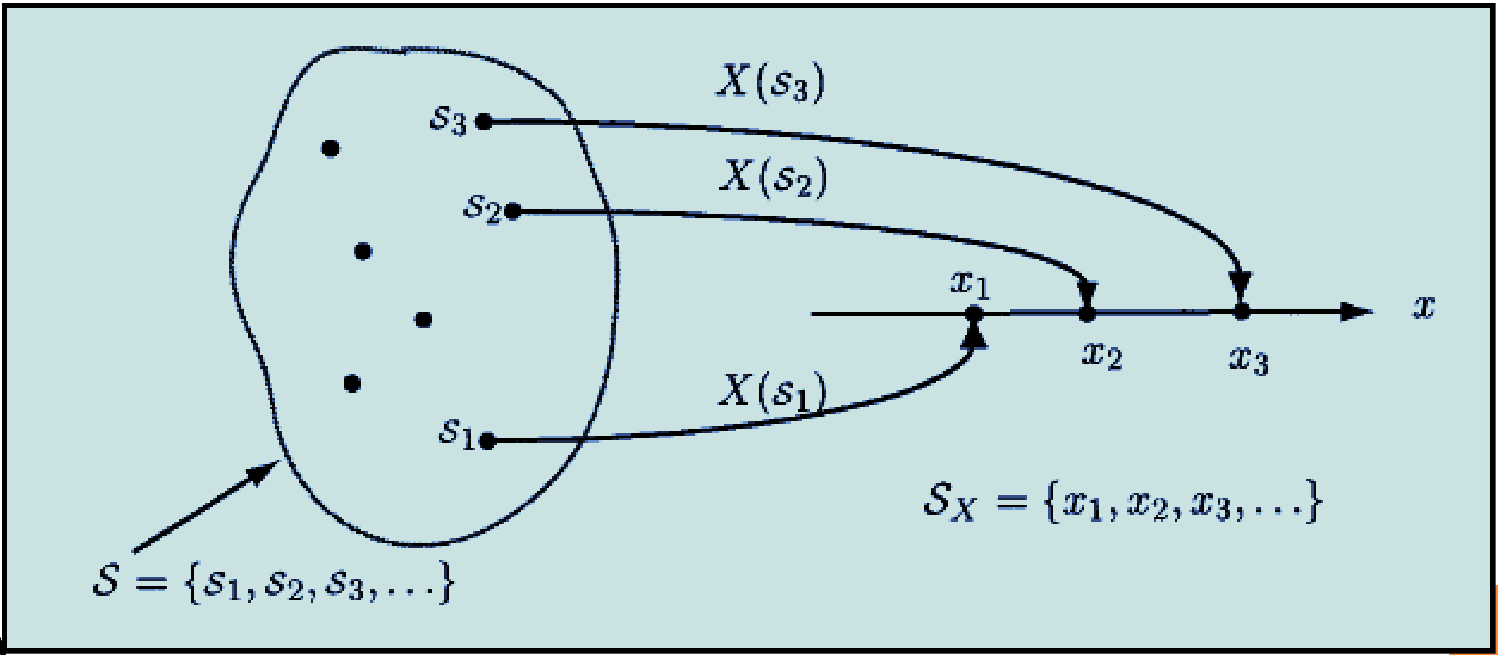
CE 513: STATISTICAL METHODS IN CIVIL ENGINEERING

Lecture-3: Random Variable

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

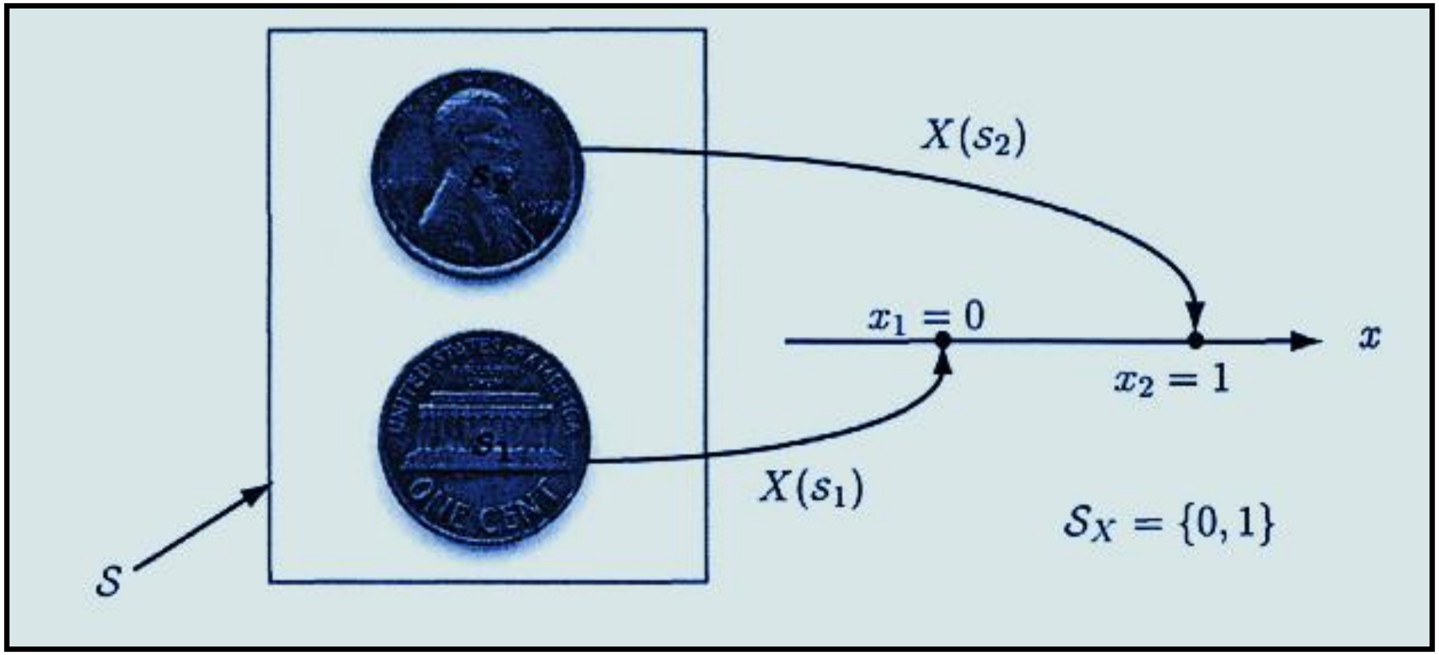
Random variable

Random Variable (RV): A finite single valued function that maps the set of all experimental outcomes in sample space S into the set of real numbers *R*, is said to be a RV



A random variable does *not* return a probability

Example: a coin toss



Discrete Random Variable

- Discrete random variables are generally used to describe events that are counted, for example: No of cars crossing the intersection
- Discrete random variables are expressed using **integers**
- The probability content of a discrete random variable is described using the **probability mass function**(PMF) and is denoted by $p_X(x)$

Discrete Random Variable

- The **cumulative distribution function**(CDF) is defined as a function of *x*, whose value is:

  The probability that *X* is less than or equal to *x*

- Because the events are **mutually exclusive**(i.e. *X* can only assume one value at a time) the CDF is obtained simply by **adding** the discrete probabilities as

$$F_X(x) = p_X(0) + p_X(1) + \cdots + p_X(x)$$

Example: PMF

Consider the problem of three nuclear reactors.

Assume that a reactor will be active and operating 90% of the time. What is the probability that at-least two reactors are operating at a given time?

Example: PMF

Let

X = no of reactors in operational at any given time

A = event that a reactor is active

O = event that a reactor is offline for service

Also let

0 = event that all reactors are offline

1 = event that 1 reactor is active and 2 are offline

2 = event that 2 reactors are active and 1 is offline

3 = event that all 3 reactors are active

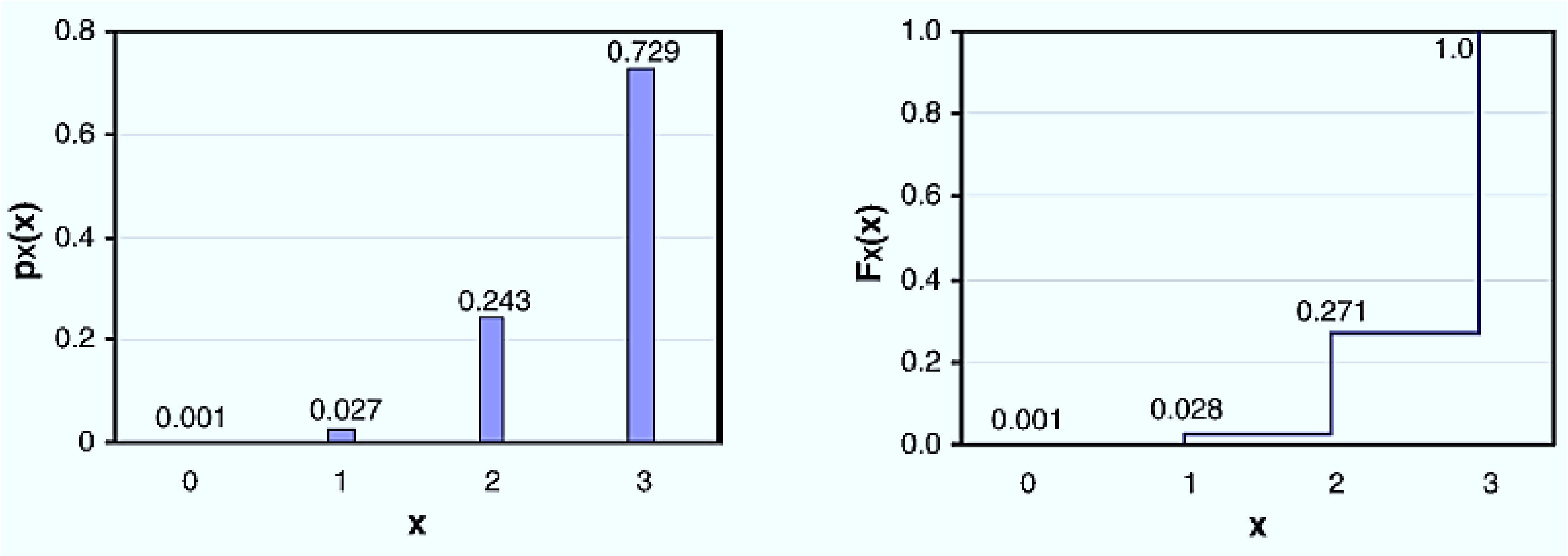
Example: PMF

We are given: $P(A) = 0.9$, $P(O) = 0.1$

Assuming the operation of the reactors is statistically independent, we can construct the PMF for the random variable X as

$p_X(0) = P(X = 0) = (0.1)(0.1)(0.1) = 0.001$
$p_X(1) = P(X = 1) = 3[(0.9)(0.1)(0.1)] = 0.027$
$p_X(2) = P(X = 2) = 3[(0.9)(0.9)(0.1)] = 0.243$
$p_X(3) = P(X = 3) = (0.9)(0.9)(0.9) = 0.729$

Example: PMF



Therefore, the probability that **at least** two reactors are operating is given by $X \geq 2$which is computed as

$$P(X \geq 2) = 1 - P(X < 2) = 1 - [p_X(0) + p_X(1)]$$
$$= 1 - [(0.001) + (0.027)] = \mathbf{0.972}$$

Properties of RV

A discrete random variable X can take m possible values X = $\{x_1, x_2, \ldots, x_m\}$ is the sample space

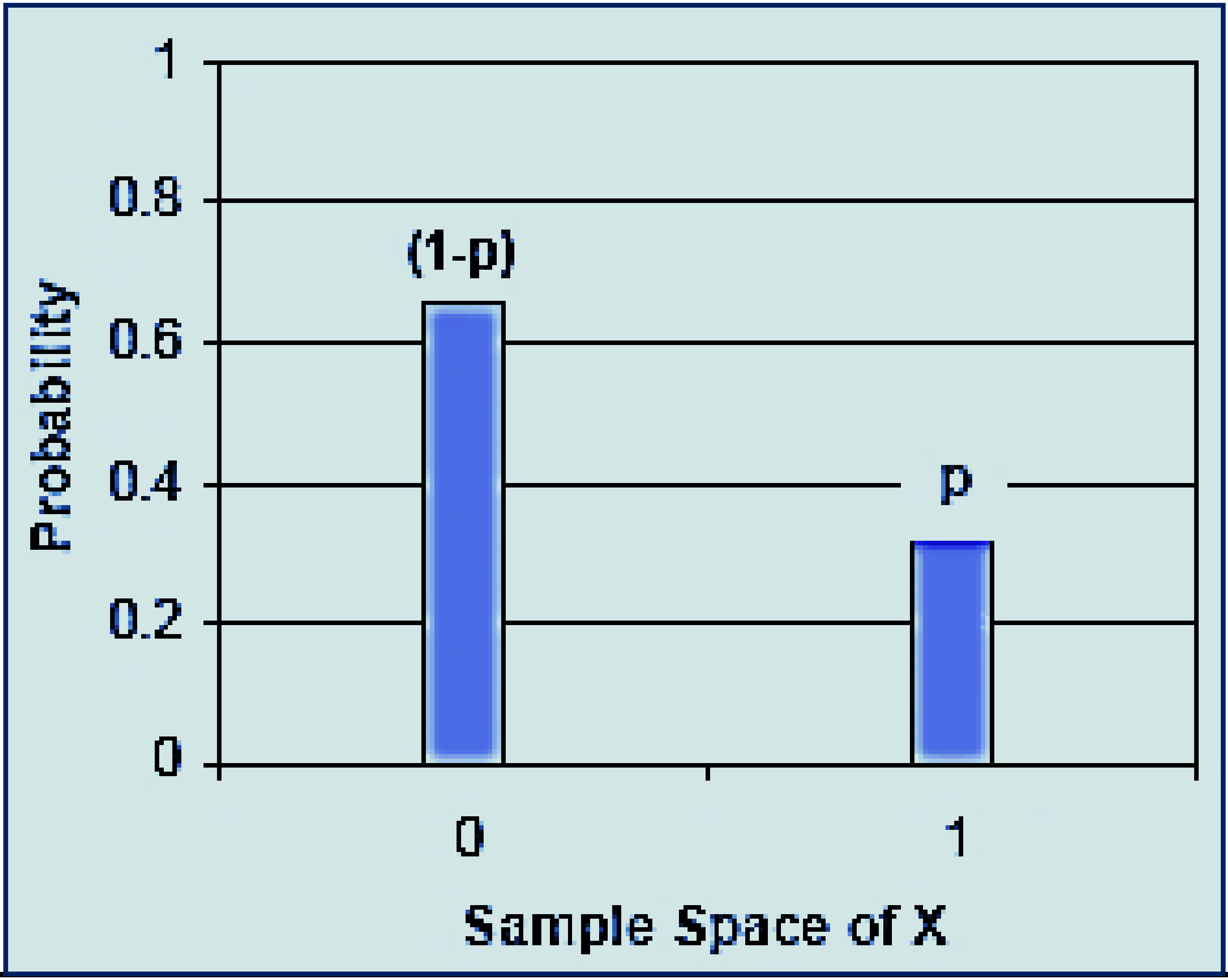
Rolling a die, X= {1, 2, 3, 4, 5, 6}

- $P(x_k)$ = Probability of taking a k^{th} value (= x_k) (from PMF)
- Expected Value or Mean = $\mu = \sum_{k=1}^{m} P(x_k) x_k$
- Variance of X = $\sigma^2 = \sum_{k=1}^{m} P(x_k)(x_k - \mu)^2$

Bernoulli trials

- Bernoulli random variable:
  - Takes only two values, X ≡ {0, 1}
- Occurrence of an event (i.e., X = 1) with probability = *p*
- No occurrence of event (i.e., X = 0) with probability = (1-*p*)

Bernoulli trials



Bernoulli trials example

- Suppose a system has 4 standby or backup units
  The probability of failure of each unit is ***p** per year*

- What is the probability that 1 unit will fail in the next year ?

Unit No.	1	2	3	4	Probability
Sequence					
1	**F**	S	S	S	$p(1-p)^3$
2	S	**F**	S	S	$(1-p)p(1-p)^2$
3	S	S	**F**	S	$(1-p)^2p(1-p)$
4	S	S	S	**F**	$(1-p)^3p$
				Total:	**4** $p(1-p)^3$

F = Fail; **S = Safe**

Binomial Distribution

- Suppose, the distribution of the number of failures X in a group of 4 machines is a RV
- The RV follows binomial distribution

$$P(X = k) = {}^{4}C_{k}\, p^{k}(1-p)^{(4-k)}$$

$$P(X=0) = {}^{4}C_{0}(=1)(1-p)^{4}$$

$$P(X=1) = {}^{4}C_{1}(=4)\,p(1-p)^{3}$$

$$P(X=2) = {}^{4}C_{2}(=6)\,p^{2}(1-p)^{2}$$

$$P(X=3) = {}^{4}C_{3}(=4)\,p^{3}(1-p)^{1}$$

$$P(X=4) = {}^{4}C_{4}(=1)\,p^{4}$$

Binomial Distribution

The number of trials (occurrence of transients or accidents = *m)*

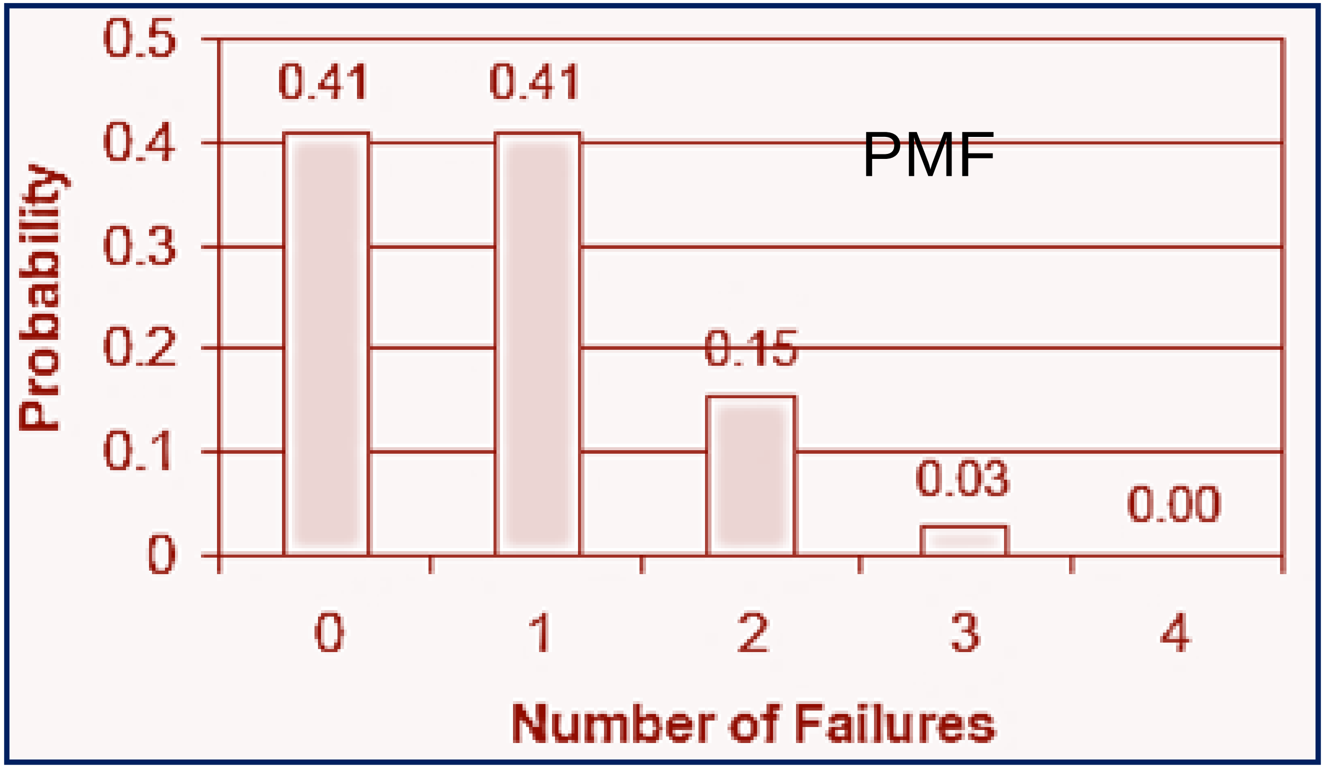
- The number of failures in *m* trials = *X*, a RV (*X* ≤ *m*)
- Probability of failure per transient/accident = p
- Binomial distribution (**Prob of exactly k occurrences in m trials**)

$$P(X = k) = {}^{m}C_{k}\, p^{k}(1-p)^{(m-k)} \; ; \quad k = 1, 2, 3, \ldots m$$

- Distribution parameters are = ***m*** and ***p***

Binomial Distribution

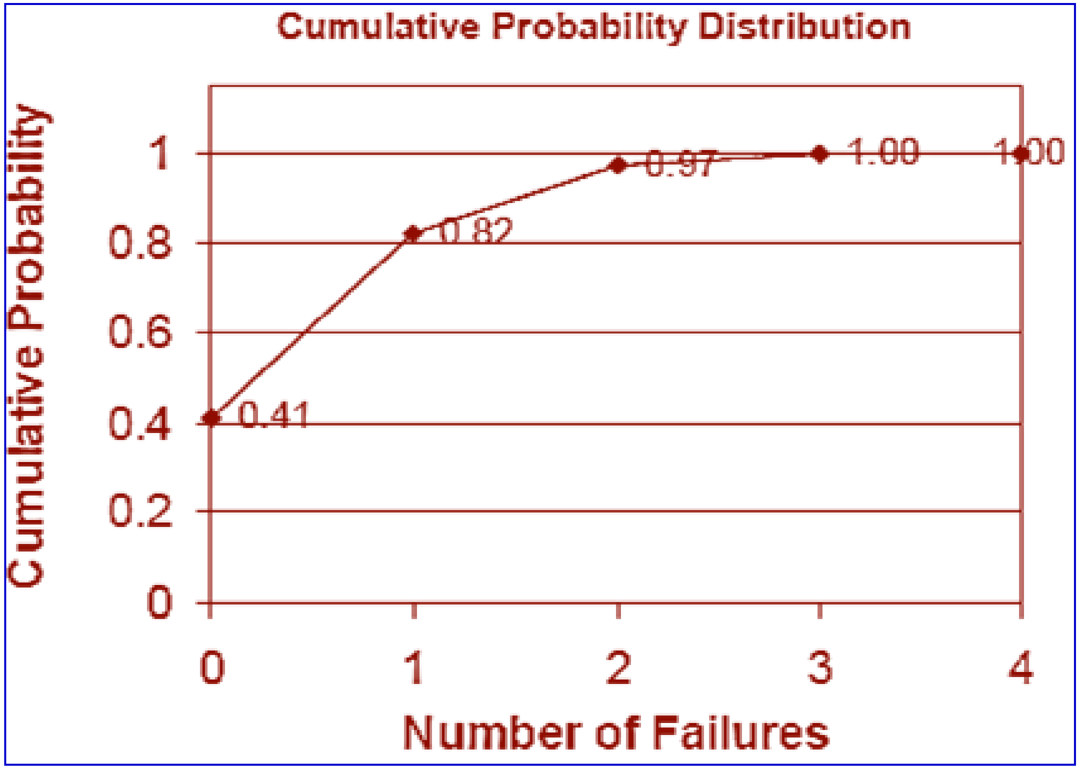
- Parameters: m = 4 machines and probability of failure p = 0.1
- The distribution of number of failures



Binomial Distribution

- What is the probability that there will be 2 or less failures? (Cumulative probability up to 2)

✓ Answer = P(X=0) + P(X=1) + P(X=2) = 0.97



Poisson Distribution

- Binomial distribution converges to the Poisson distribution
  - When probability of failure $p \rightarrow 0$ (very small)
  - And the population of component $m \rightarrow \infty$ (very large)
  - Such that $mp \rightarrow \mu$, constant called mean number of failures

- Poisson distribution gives the distribution of the number of failures (N)

$$P_N[k] = \frac{\mu^k e^{-\mu}}{k!} \quad (k = 0, 1, \ldots \infty)$$

Example: Poisson distribution

- Probability of failure of a component

  $p = 0.0025$ per year

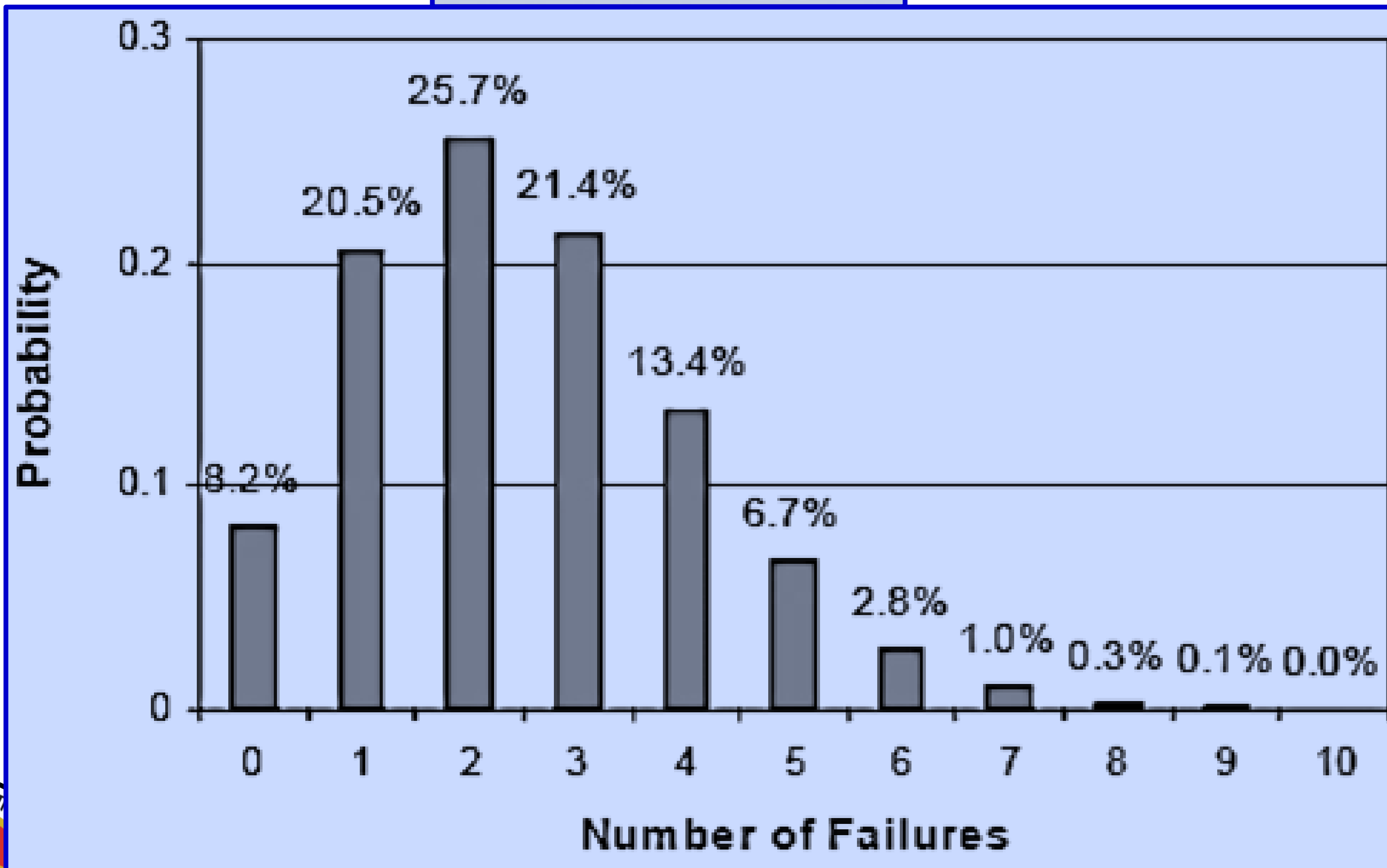
- The number of components in service

  $m = 1000$

- Mean number of failures

  $\mu = m\,p = 2.5$ failures per year

$$P_N[k] = \frac{\mu^k e^{-\mu}}{k!}$$



CE 513: STATISTICAL METHODS IN CIVIL ENGINEERING

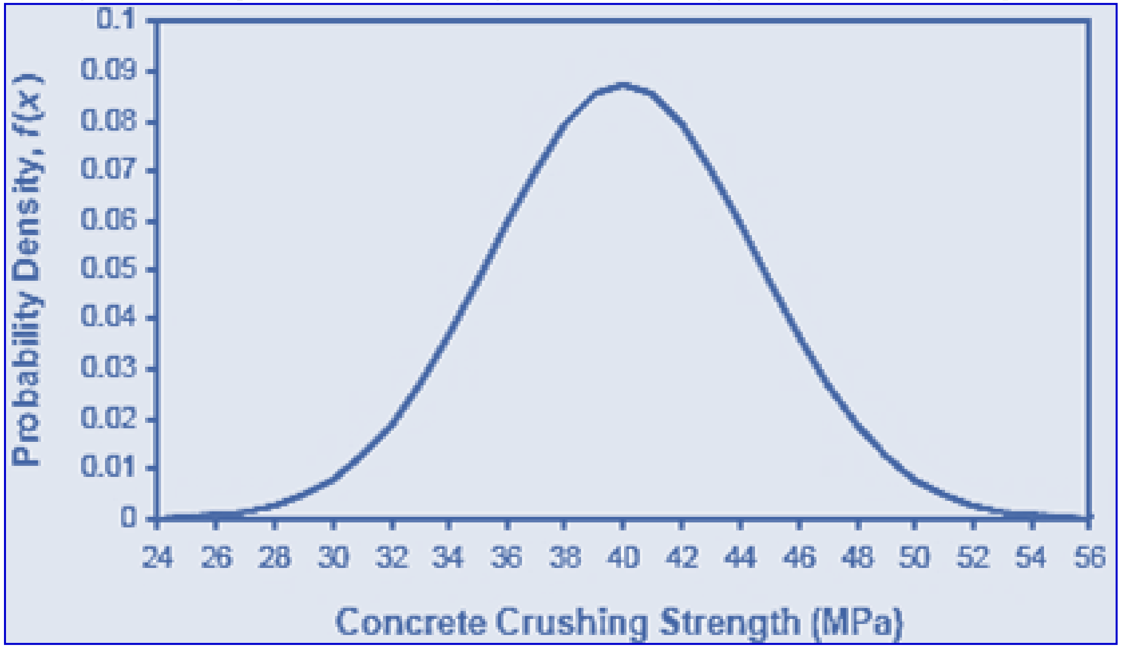
Lecture- 4: Continuous RV

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

Continuous RVs

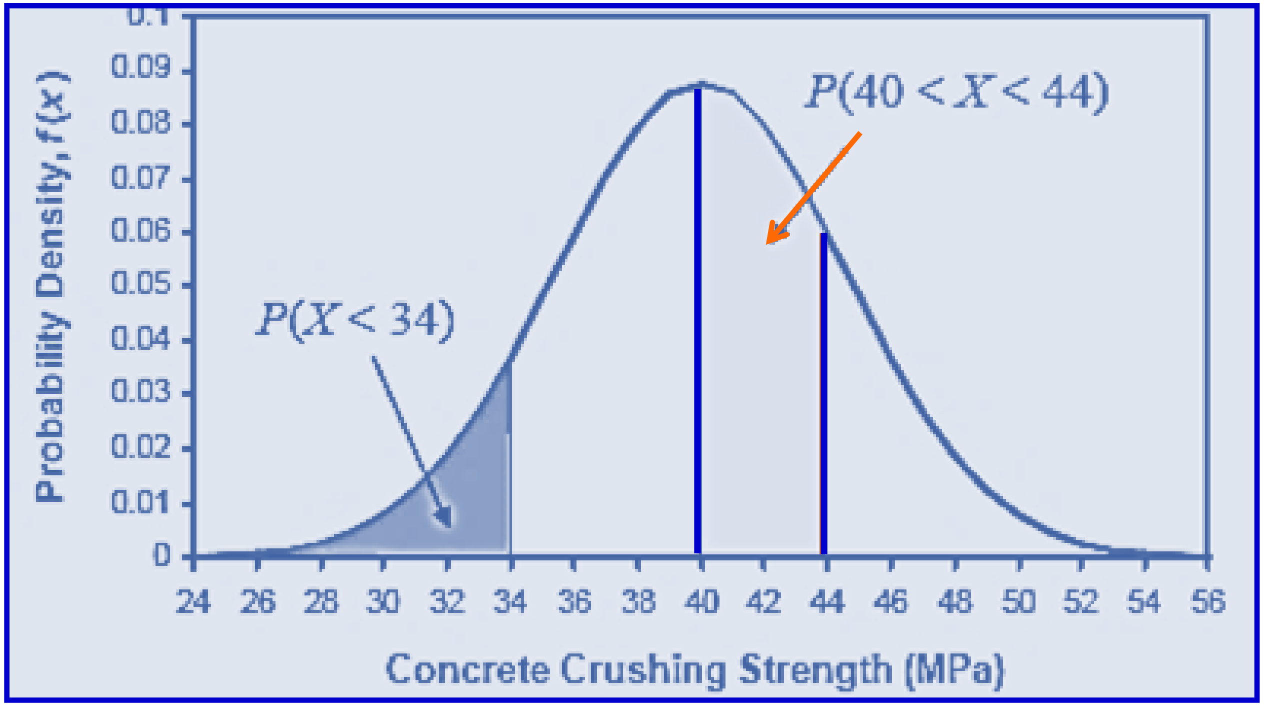
- A continuous random variable can assume any value within a given range e.g. Concrete crushing strength

- 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 95^{th} percentile value

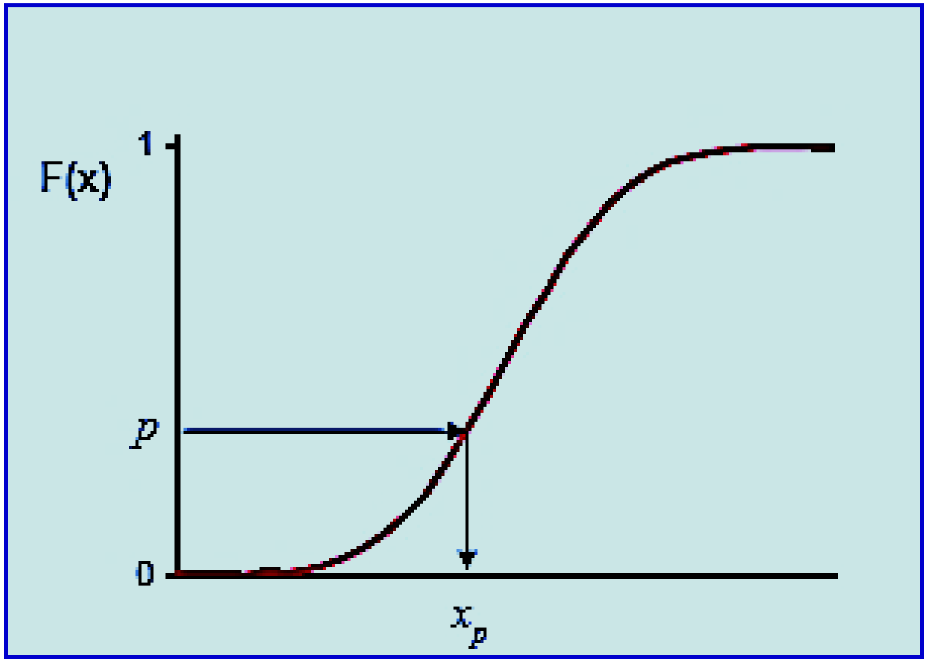
CDF & Quantile function

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

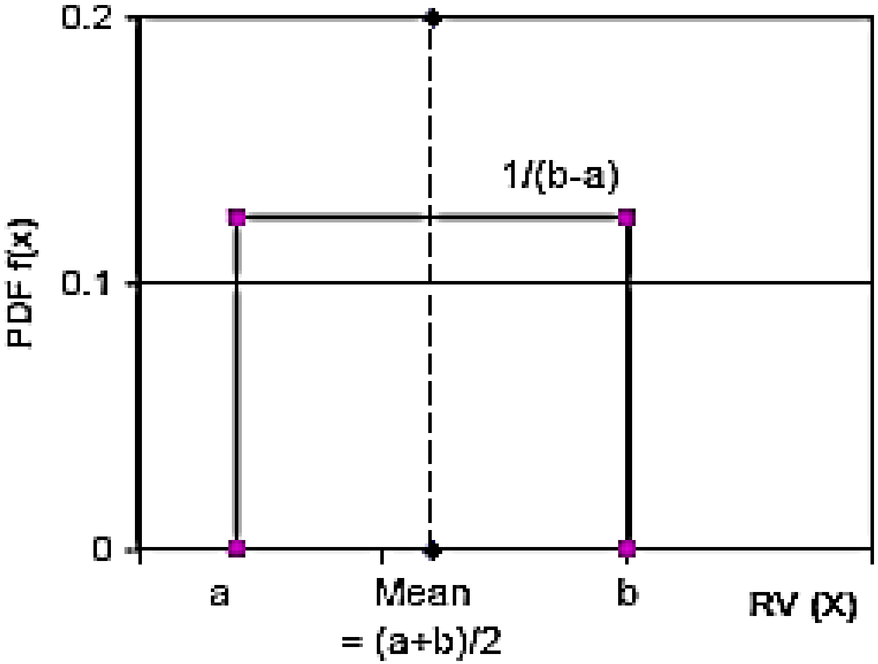
$$F_X(x) = p$$

⇓

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



Uniform distribution

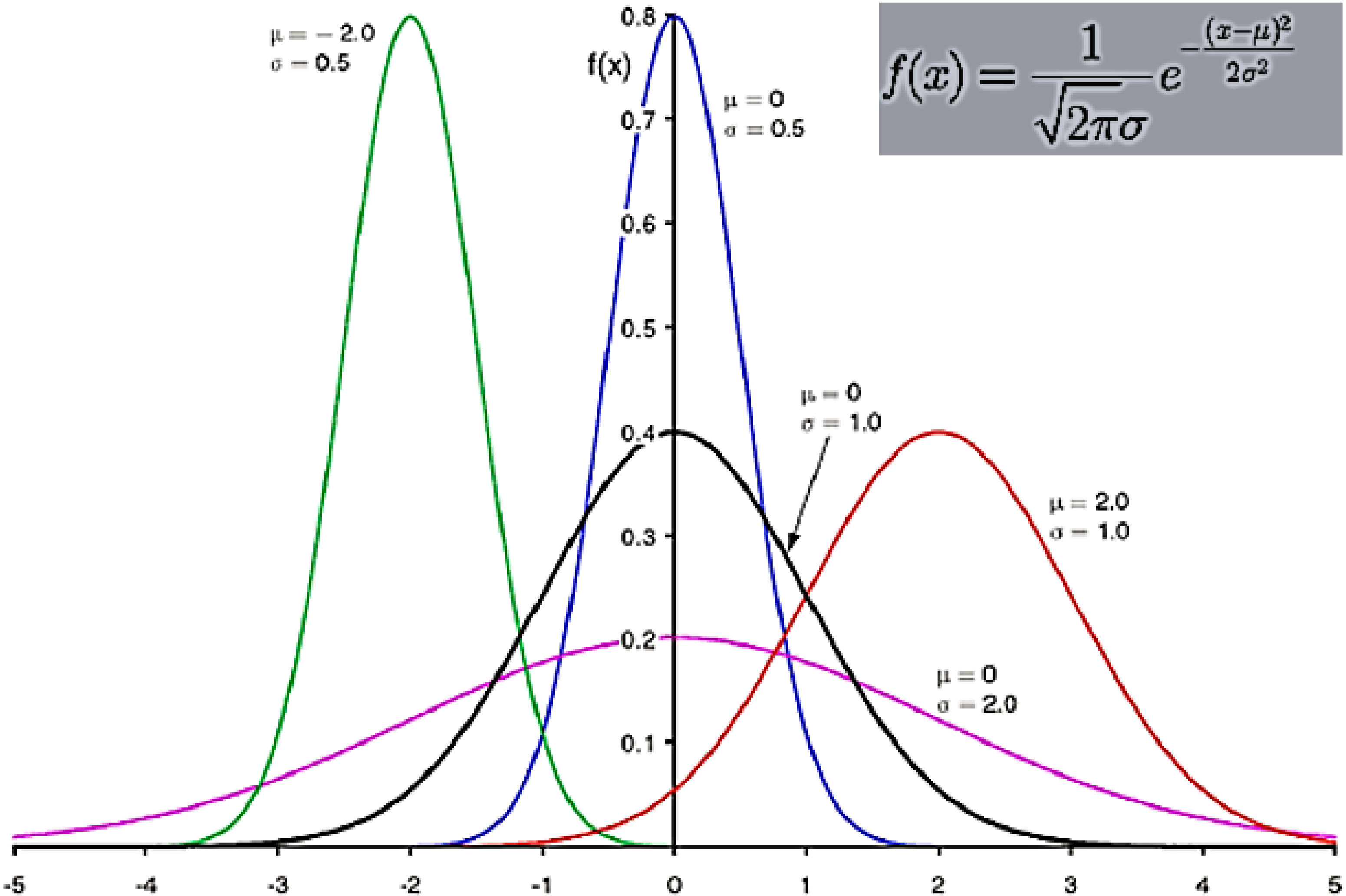


$$\text{Mean}: \quad \mu = \frac{(a+b)}{2}$$

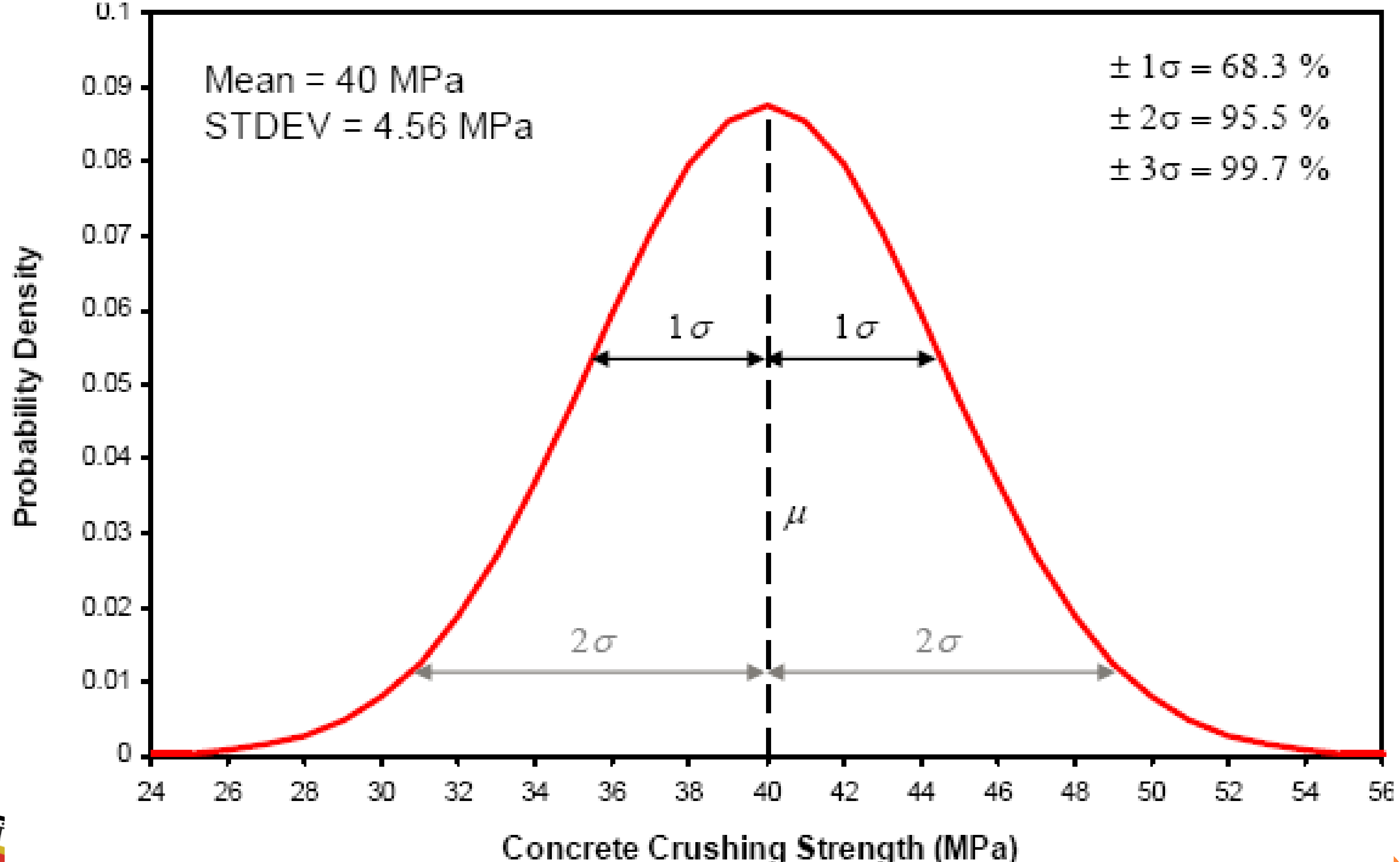
$$\text{Variance}: \quad \sigma^2 = \frac{(b-a)^2}{12}$$

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

Normal distribution



Normal distribution



Standard normal distribution

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

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

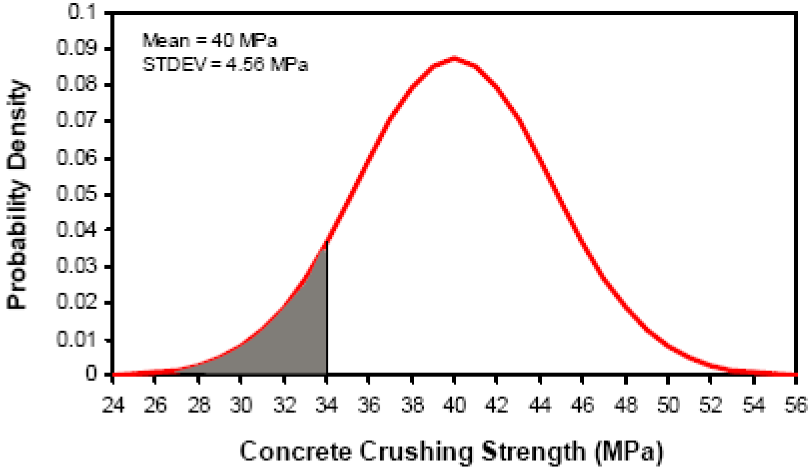
- The Standard Normal distribution is denoted as $N(0,1)$and has a **mean** of **zero** and **standard deviation** equal to **one**
- Because of its wide use, the CDF of the Standard Normal variate is denoted as $\Phi(s)$

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

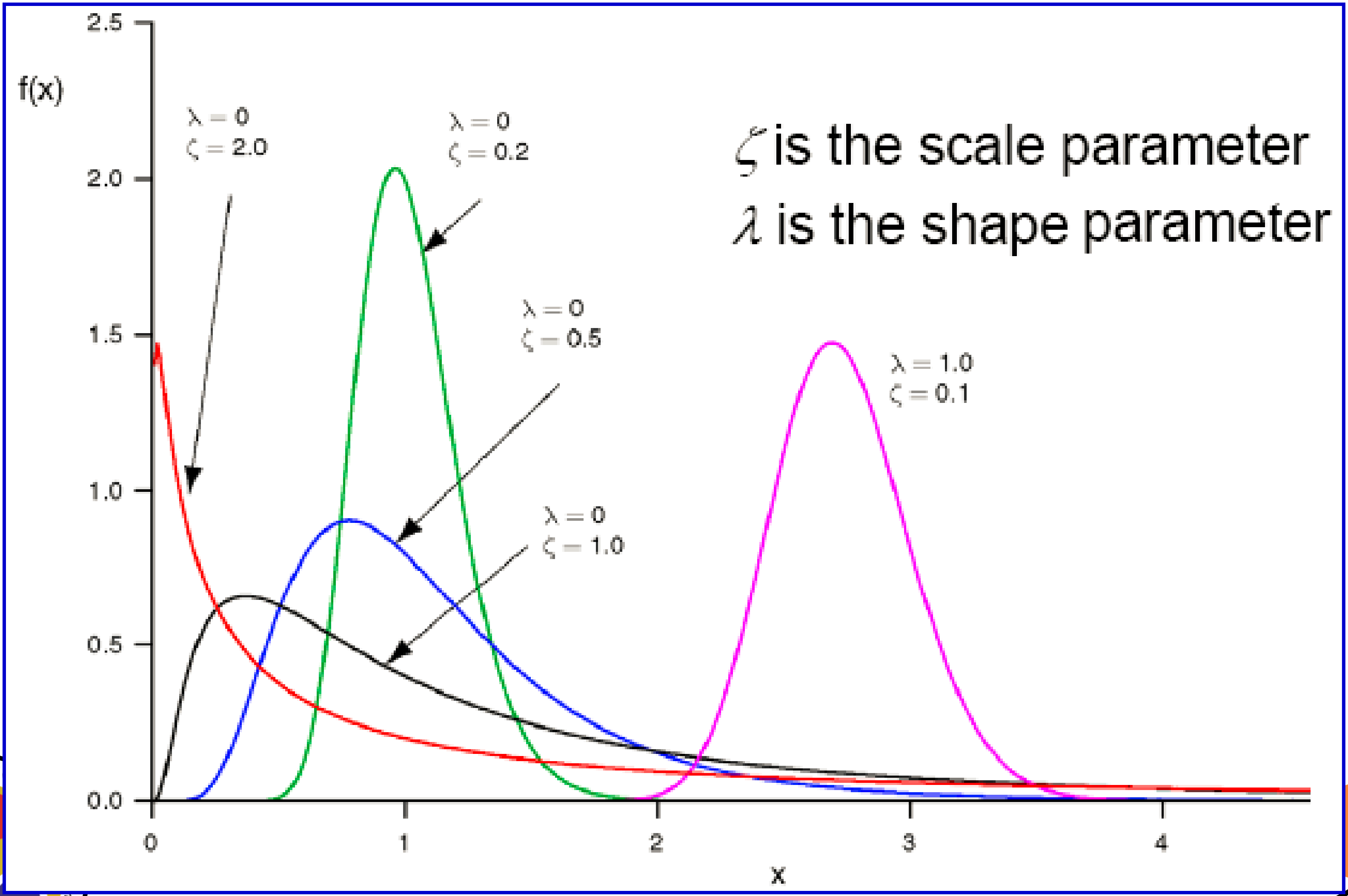
- The logarithmic or **Log-Normal distribution** is used when the random variable **cannot** take on a **negative** value

- 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

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

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

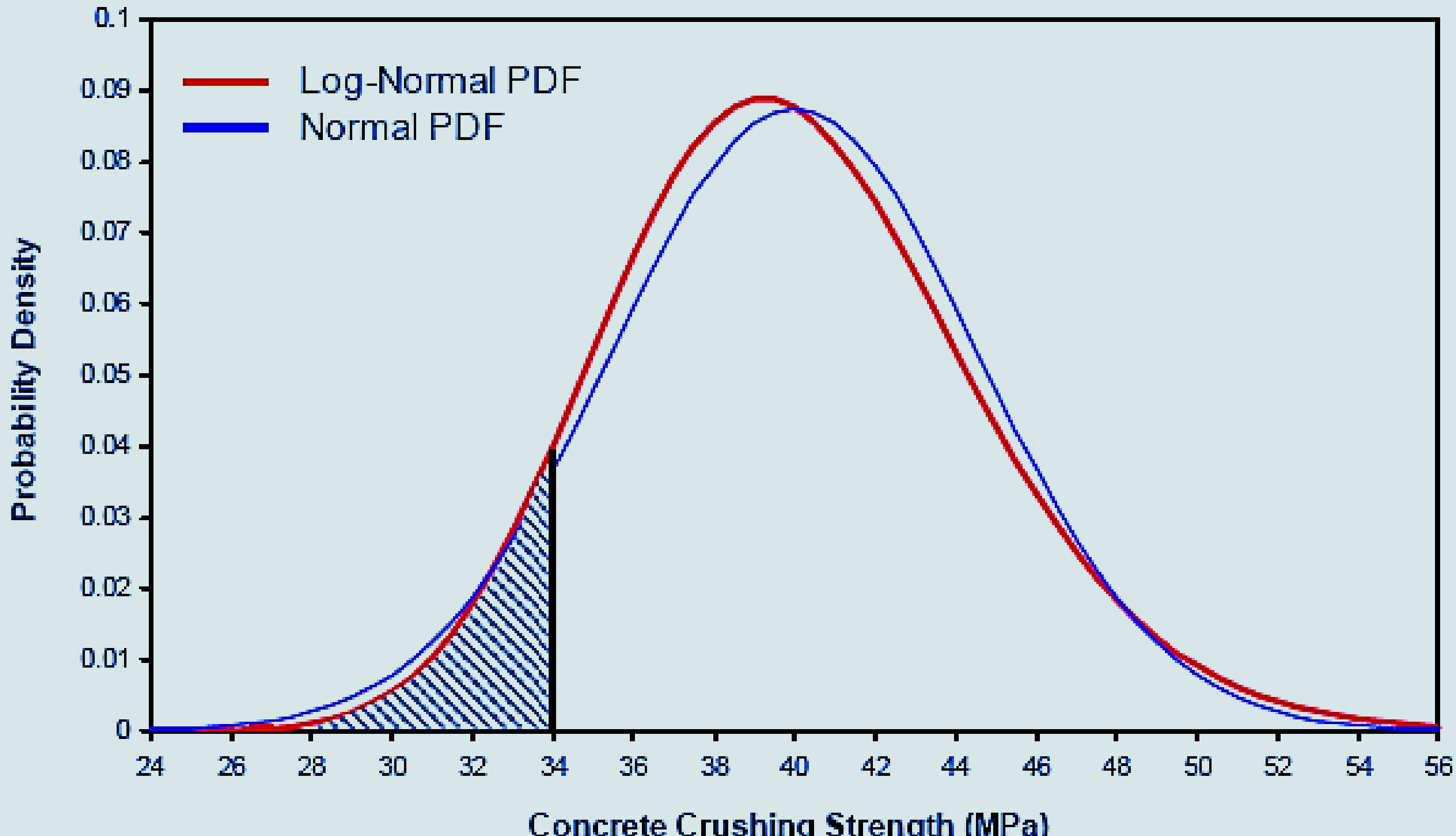
Assuming the concrete strength is described by the Log-Normal distribution, what is the probability that the concrete strength is less than or equal to 34 MPa?

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



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

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

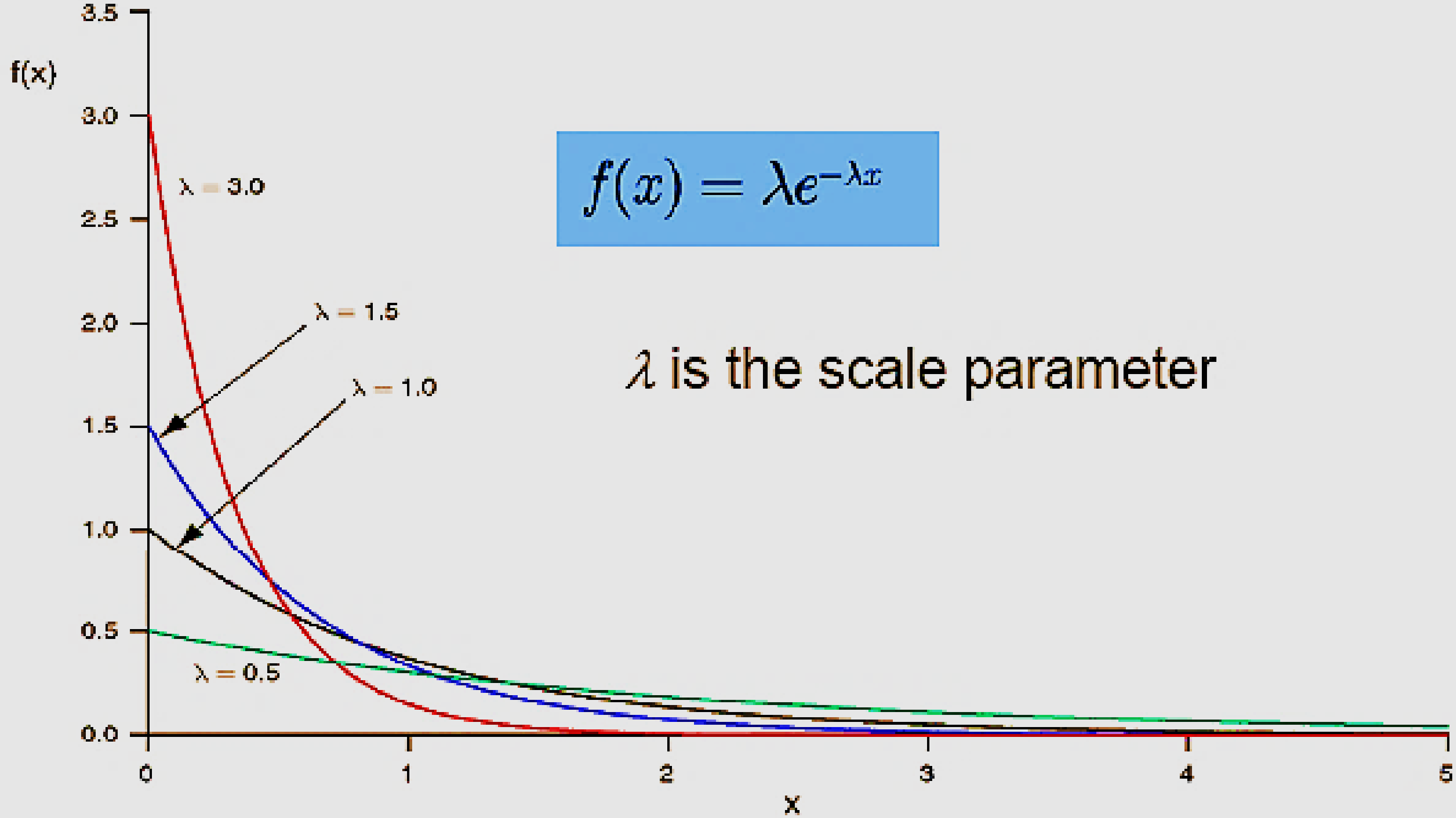
- Assuming the concrete strength follows the Log-Normal distribution (i.e., the LOG of the concrete strength follows the Normal distribution), there is a **8.5 %**chance that the concrete strength is less than or equal to 34 MPa

CE 513: STATISTICAL METHODS IN CIVIL ENGINEERING

Lecture- 5: Continuous RV

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

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?

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

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

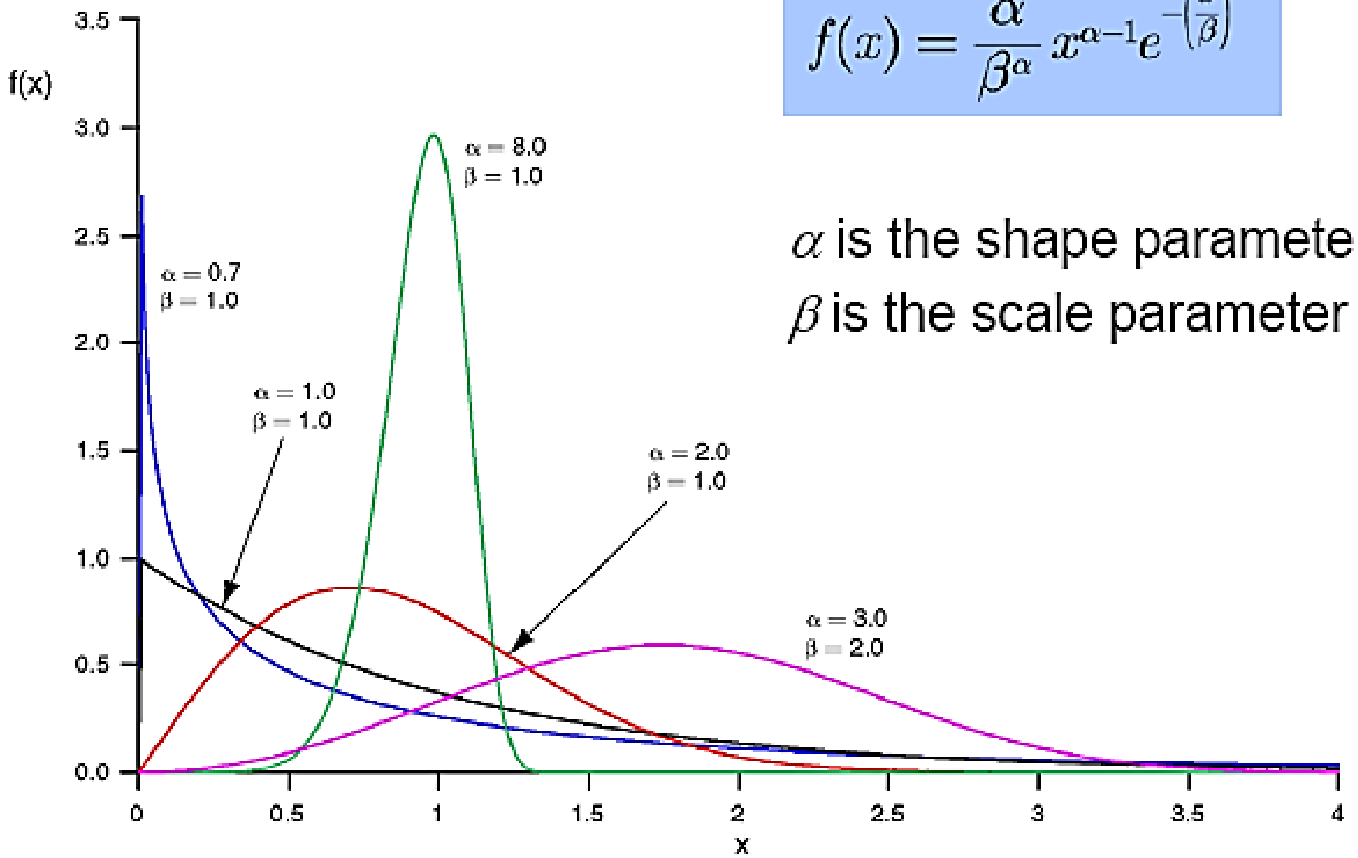
Weibull distribution

- 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

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

Using MATLAB command:

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

Inverse Weibull distribution

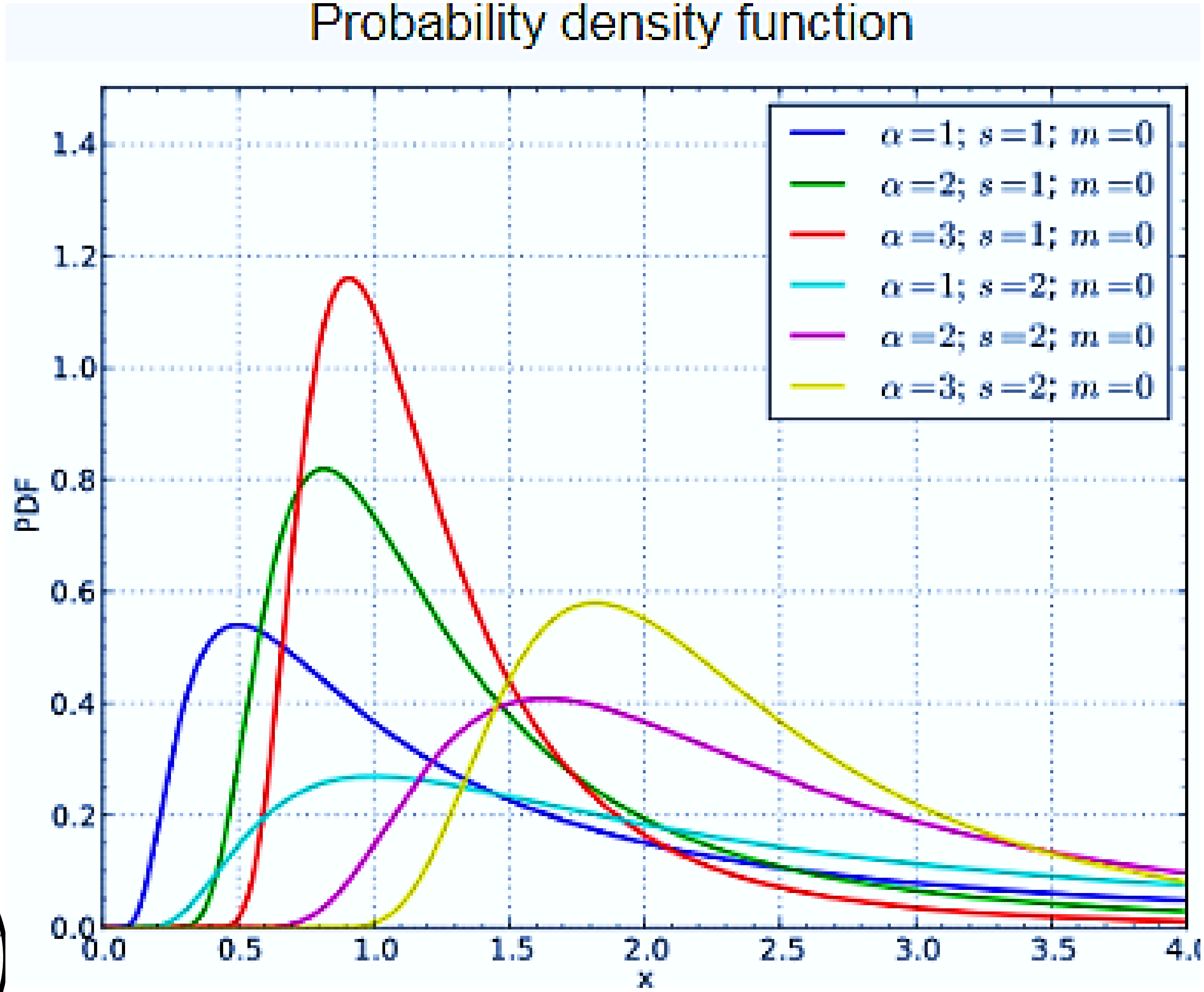
The **Fréchet distribution**, also known as inverse Weibull distribution, is a special case of the generalized extreme value distribution. It has the cumulative distribution function

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

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

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

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

(alternative format)

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

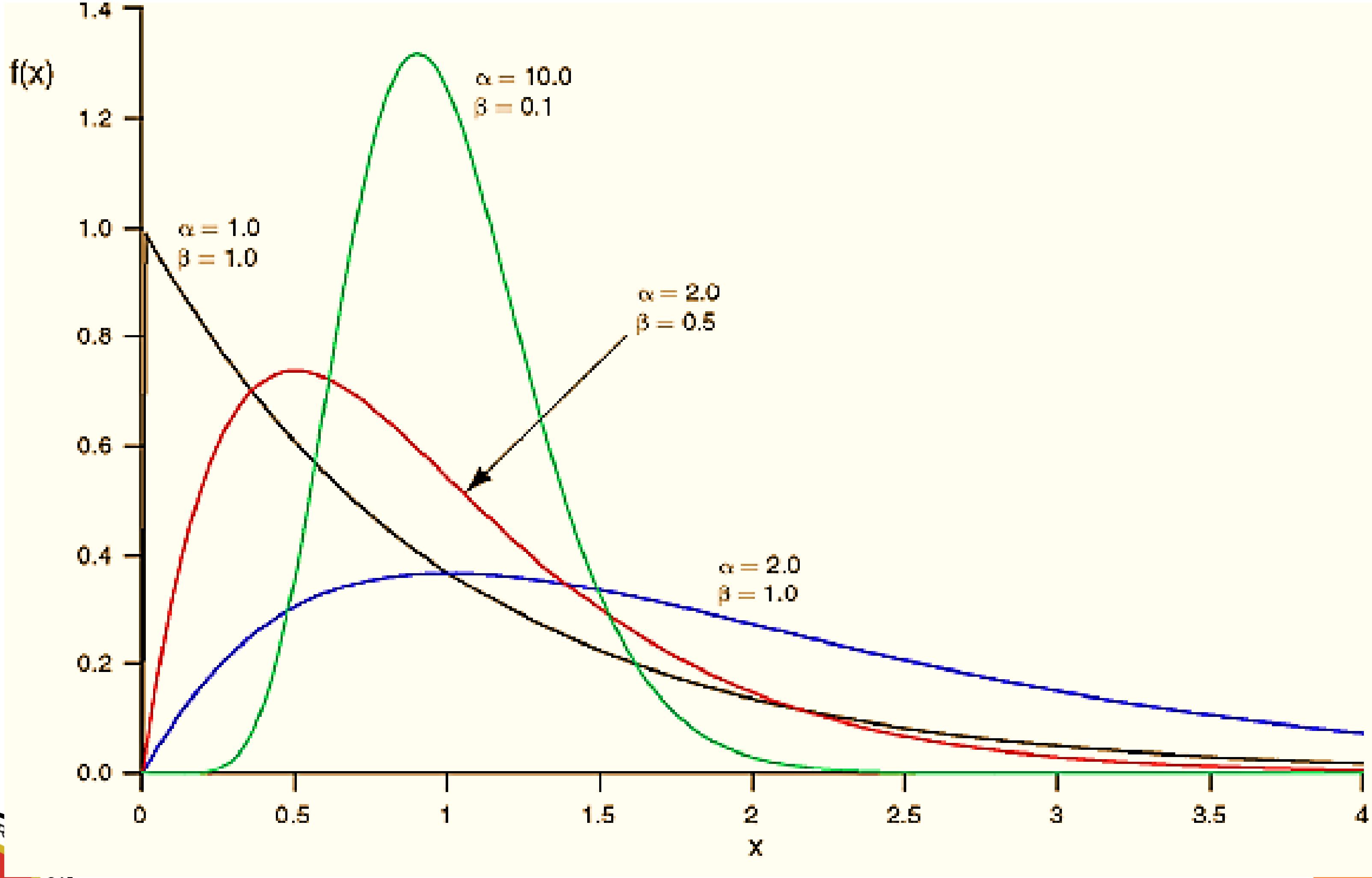
α is the shape parameter
β is the scale parameter

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

CDF:

$$F(x) = \frac{\Gamma_x(\alpha)}{\beta\Gamma(\alpha)}$$

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

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

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

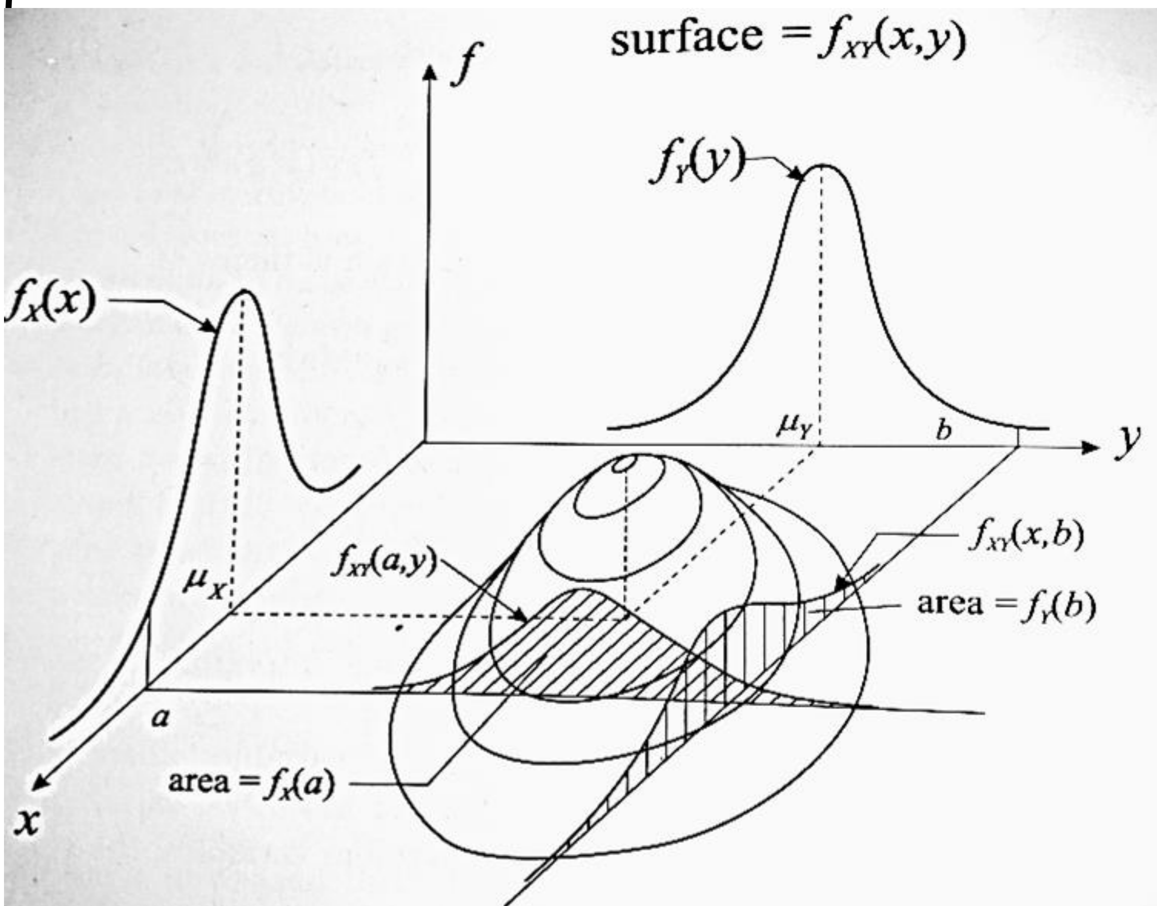
$$f_{XY}(x,y)dx\,dy \simeq \Pr\left(x < X \leq x + dx, y < Y \leq y + dy\right),$$

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

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

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

Continuous RV



CDF

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

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

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

$$\text{Cov}(X,Y) = \sigma_{xy} = E[(X-\mu_x)(Y-\mu_y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_x)(y-\mu_y)p(x,y)dxdy$$

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

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

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

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

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

Bi-variate Gaussian distribution

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

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

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 ?

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

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

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

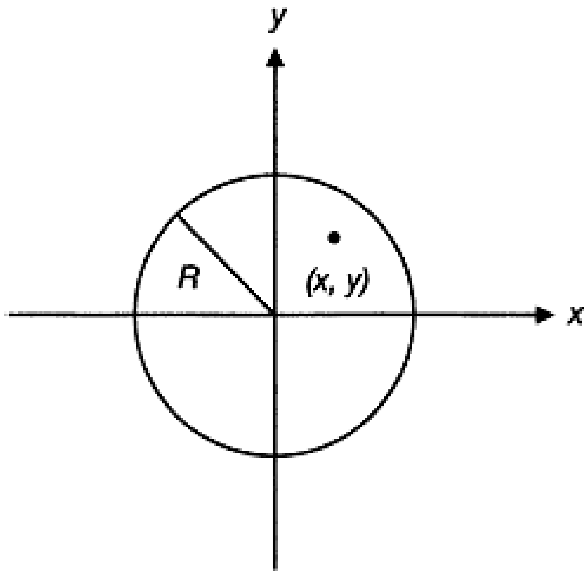
$$f_{Y|X}(y|x)=\frac{4xy}{2x}=2y \qquad 0<y<1, 0<x<1$$

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

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.

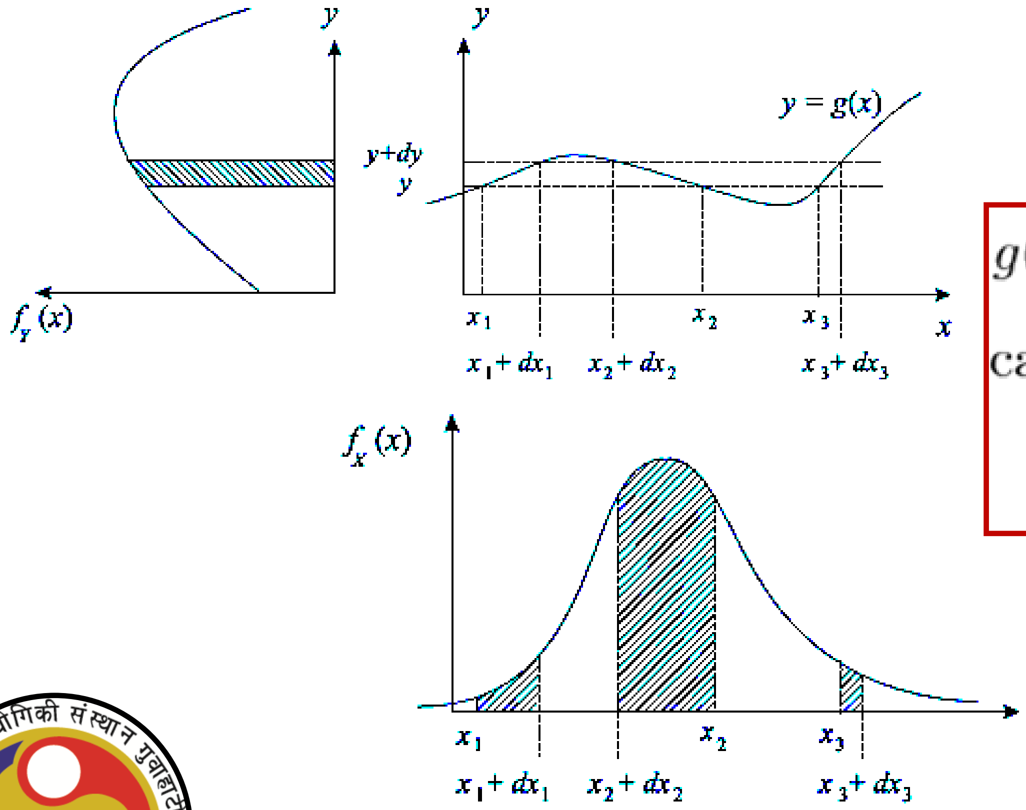
CE 513: STATISTICAL METHODS IN CIVIL ENGINEERING

Lectures- 8, 9: Functions of RVs

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

Function of random variables

Given $f_X(x)$ and $g(X)$, where $Y = g(X)$, there is an interest in finding $f_Y(y)$.



$g(X)$ is simple enough to allow calculation of the inverse

$$X = g^{-1}(Y)$$

Function of random variables

$$\{y < Y \le y + dy\} = \{x_1 < X \le x_1 + dx_1\} + \{x_2 + dx_2 < X \le x_2\} + \{x_3 < X \le x_3 + dx_3\},$$

$$\Pr(y < Y \le y + dy) = \Pr(x_1 < X \le x_1 + dx_1) + \Pr(x_2 + dx_2 < X \le x_2) + \Pr(x_3 < X \le x_3 + dx_3),$$

$$g'(X) \equiv \frac{dg}{dX} \equiv \frac{dy}{dX},$$

$$g'(x_i)\, dX|_{X=x_i} = dy,$$

$$f_Y(y) = \sum_{i=1}^{n} \frac{f_X(x_i)}{|g'(x_i)|}.$$

Example

$$X \text{ normally distributed, } f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right\}, \qquad -\infty < x < \infty,$$

$$Y = aX^2,\ a > 0$$

What is pdf of y ?

Solution:

Since only the real roots are needed, and there are no real solutions if $Y < 0$, then $f_Y(y) = 0$ for this domain

If $Y \geq 0$, there are two solutions,

$$x_1 = +\sqrt{\frac{y}{a}} \qquad x_2 = -\sqrt{\frac{y}{a}}.$$

Example

The functional relation is $g(X) = aX^2$, with its derivative

$$g'(X) = 2aX = 2a\sqrt{Y/a} = 2\sqrt{aY}$$

$$f_Y(y) = \sum_{i=1}^{2} \frac{f_X(x_i)}{|g'(x_i)|} = \frac{1}{2\sqrt{ay}} \left\{ f_X\left(\sqrt{\frac{y}{a}}\right) + f_X\left(-\sqrt{\frac{y}{a}}\right) \right\}, \qquad y \geq 0.$$

$$f_Y(y) = \frac{1}{\sigma_X \sqrt{2\pi a y}} \exp\left\{ -\frac{\left(\sqrt{y/a} - \mu_X\right)^2}{2a\sigma_X^2} \right\}, \qquad y > 0.$$

Exercise

Solve the following problem ?

The strain energy in a linearly elastic bar subjected to an axial force S is given by the equation

$$U = \frac{L}{2AE} S^2$$

where:

$L =$ length of the bar

$A =$ cross-sectional area of the bar

$E =$ modulus of elasticity of the material

Using $c = L/2AE$, we can rewrite

Now, if S is a lognormal variate with parameters λ and ζ, What is the pdf of U ?

Moments of functions of RVs

$$Y = a_1 X_1 + a_2 X_2$$

$$\text{Var}(Y) = a_1^2 \,\text{Var}(X_1) + a_2^2 \,\text{Var}(X_2) + 2 a_1 a_2 \rho_{X_1 X_2} \sigma_{X_1} \sigma_{X_2}$$

$$Y = \sum_{i=1}^{n} a_i X_i$$

$$E(Y) = \sum_{i=1}^{n} a_i E(X_i) = \sum_{i=1}^{n} a_i \mu_{X_i}$$

$$\text{Var}(Y) = \sum_{i=1}^{n} a_i^2 \text{Var}(X_i) + \sum_{i,j=1}^{n} \sum_{i \neq j}^{n} a_i a_j \text{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{n} a_i^2 \sigma_{X_i}^2 + \sum_{i,j=1}^{n} \sum_{i \neq j}^{n} a_i a_j \rho_{ij} \sigma_{X_i} \sigma_{X_j}$$

in which ρ_{ij} is the correlation coefficient between X_i and X_j

Moments of functions of RVs

In many cases derived probability distributions may be very difficult to evaluate for general nonlinear functions.

Either use Monte Carlo simulation to find the derived density

Or,

Estimate mean and variance using an approximate analysis which in Most of the practical applications is sufficient, although the Pdf may still be undermined.

Moments of general function of a single RV

For a general function of a single random variable X,

$$Y = g(X)$$

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

$$\text{Var}(Y) = \int_{-\infty}^{\infty} [g'(x) - \mu_X]^2 f_X(x)\, dx$$

To find the approximate expressions of mean and variance, we use Taylor's series to expand a function about its mean μ_X

$$g(X) = g(\mu_X) + (X - \mu_X)\frac{dg}{dX} + \frac{1}{2}(X - \mu_X)^2\frac{d^2g}{dX^2} + \cdots$$

Moments of general function of a single RV

First order approximation

$$g(X) \simeq g(\mu_X) + (X - \mu_X)\frac{dg}{dX}$$

$$E(Y) \simeq g(\mu_X)$$

$$\mathrm{Var}(Y) \simeq \mathrm{Var}(X - \mu_X)\left(\frac{dg}{dX}\right)^2 = \mathrm{Var}(X)\left(\frac{dg}{dX}\right)^2$$

Second order approx.

$$\mathrm{Var}(Y) \simeq \sigma_X^2\left(\frac{dg}{dX}\right)^2 - \frac{1}{4}\sigma_X^2\left(\frac{d^2g}{dX^2}\right)^2 + E(X - \mu_X)^3\frac{dg}{dX}\frac{d^2g}{dX^2} + \frac{1}{4}E(X - \mu_X)^4\left(\frac{d^2g}{dX^2}\right)^2$$

Example

The maximum impact pressure (in psf) of ocean waves on coastal structures may be determined by

$$p_m = 2.7\frac{\rho K U^2}{D}$$

where U is the random horizontal velocity of the advancing wave, with a mean of 4.5 fps and a c.o.v. of 20%. The other parameters are all constants as follows:

ρ = 1.96 slugs/cu ft, the density of sea water

K = length of hypothetical piston

D = thickness of air cushion

Assume a ratio of $K/D = 35$

Example

The first-order mean and standard deviation of p_m, are

$$E(p_m) \simeq 2.7(1.96)(35)(4.5)^2 = 3750.7 \text{ psf} = 26.05 \text{ psi}; \quad \text{and}$$

$$\text{Var}(p_m) \simeq \text{Var}(U)\left(2.7\rho\frac{K}{D}\right)^2(2\mu_U)^2 = (0.20 \times 4.5)^2(2.7 \times 1.96 \times 35)^2(2 \times 4.5)^2$$

Therefore, the standard deviation of p_m is

$$\sigma_{p_m} \simeq (0.20 \times 4.5)(2.7 \times 1.96 \times 35)(2 \times 4.5) = 1500.3 \text{ psf} = 10.42 \text{ psi}$$

For an improved mean value, we evaluate the second-order mean with Eq. 4.48 as follows:

$$E(Y) \simeq 3750.7 + \frac{1}{2}(0.20 \times 4.5)^2\left(2.7\rho\frac{K}{D}\right)(2)$$
$$= 3750.7 + \frac{1}{2}(0.20 \times 4.5)^2(2.7 \times 1.96 \times 35 \times 2)$$
$$= 3750.7 + 150.0 = 3900.7 \text{ psf} = 27.09 \text{ psi}$$

This shows that for this case the first-order mean is about 4% less than the second-order mean

Moments of general function of a multiple RVs

If Y is a function of several random variables,

$$Y = g(X_1, X_2, \ldots, X_n)$$

To find the approximate expressions of mean and variance, we use Taylor's series to expand a function about its mean μ_{X_i}

Expand the function $g(X_1, X_2, \ldots, X_n)$ in a Taylor series about the mean values $(\mu_{X_1}, \mu_{X_2}, \ldots, \mu_{X_n})$, yielding

$$Y = g(\mu_{X_1}, \mu_{X_2}, \ldots, \mu_{X_n}) + \sum_{i=1}^{n} (X_i - \mu_{X_i}) \frac{\partial g}{\partial X_i} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (X_i - \mu_{X_i})(X_j - \mu_{X_j}) \frac{\partial^2 g}{\partial X_i \partial X_j} + \cdots$$

where the derivatives are all evaluated at $\mu_{X_1}, \mu_{X_2}, \ldots, \mu_{X_n}$.

Moments of general function of a multiple RVs

First order approx.

$$\mathrm{Var}(Y) \simeq \sum_{i=1}^{n} \sigma_{X_i}^2 \left(\frac{\partial g}{\partial X_i}\right)^2 + \sum_{i,j=1}^{n} \sum_{i \neq j}^{n} \rho_{ij} \sigma_{X_i} \sigma_{X_j} \frac{\partial g}{\partial X_i} \frac{\partial g}{\partial X_j}$$

Second order approx of mean

$$E(Y) \simeq g(\mu_{X_1}, \mu_{X_2}, \ldots, \mu_{X_n}) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij} \sigma_{X_i} \sigma_{X_j} \left(\frac{\partial^2 g}{\partial X_i \partial X_j}\right)$$

What happens if X_i's are independent

$$E(Y) \simeq g(\mu_{X_1}, \mu_{X_2}, \ldots, \mu_{X_n}) + \frac{1}{2} \sum_{i=1}^{n} \sigma_{X_i}^2 \left(\frac{\partial^2 g}{\partial X_i^2}\right)$$

Example-1

According to the Manning equation, the velocity of uniform flow, in fps, in an open channel is

$$V = \frac{1.49}{n} R^{2/3} S^{1/2}$$

where:

S = slope of the energy line, in %

R = the hydraulic radius, in ft

n = the roughness coefficient of the channel

For a rectangular open channel with concrete surface, assume the following mean values and corresponding c.o.v.s:

Variable	*Mean Value*	*c.o.v.*
S	1%	0.10
R	2 ft	0.05
n	0.013	0.30

Example-1

Assuming that the above random variables are statistically independent, the first-order mean and variance of the velocity V are, respectively,

$$\mu_V \simeq \frac{1.49}{0.013}(2)^{2/3}(1)^{1/2} = 182 \text{ fps}; \quad \text{and}$$

$$\sigma_V^2 \simeq \sigma_S^2\left(\frac{1.49}{2\mu_n}\mu_R^{2/3}\mu_S^{-1/2}\right)^2 + \sigma_R^2\left(\frac{2\times 1.49}{3\mu_n}\mu_S^{1/2}\mu_R^{-1/3}\right)^2 + \sigma_n^2\left(-1.49\mu_R^{2/3}\mu_S^{1/2}\mu_n^{-2}\right)^2$$

$$= (0.10\times 1)^2\left(\frac{1.49}{2\times 0.013}(2)^{2/3}(1)^{-1/2}\right)^2 + (0.05\times 2)^2\left(\frac{2\times 1.49}{3\times 0.013}(1)^{1/2}(2)^{-1/3}\right)^2$$

$$+ (0.30\times 0.013)^2\left(-1.49(2)^{2/3}(1)^{1/2}(0.013)^{-2}\right)^2 = 82.79 + 36.80 + 0.21 = 119.80$$

yielding the standard deviation

$$\sigma_V = 10.94 \text{ fps}$$

Example-1

The corresponding second-order mean velocity would be, according to Eq.

$$\mu_V \simeq 182 + \frac{1}{2}\left[\sigma_S^2\left(-\frac{1.49}{4\mu_n}\mu_R^{2/3}\mu_S^{-3/2}\right) + \sigma_R^2\left(-\frac{2\times 1.49}{9\mu_n}\mu_S^{1/2}\mu_R^{-4/3}\right) + \sigma_n^2\left(\frac{2\times 1.49}{\mu_n^3}\mu_R^{2/3}\mu_S^{1/2}\right)\right]$$

$$= 182 + \frac{1}{2}\left[\begin{array}{l} -(0.1)^2\left(\frac{1.49}{4\times 0.013}(2)^{2/3}(1)^{-3/2}\right) - (0.05\times 2)^2\left(\frac{2\times 1.49}{9\times 0.013}\right)(1)^{1/2}(2)^{-4/3} \\ +(0.30\times 0.013)^2\left(\frac{2\times 1.49}{(0.013)^3}(2)^{2/3}(1)^{1/2}\right) \end{array}\right]$$

$$= 182 + \frac{1}{2}(-0.46 - 0.10 + 32.76) = 198.10 \text{ fps}$$

The first-order approximate mean velocity is about 8% lower than the corresponding second-order mean velocity. ◀

Example-2

The applied stress, S, in a beam is calculated as

$$S = \frac{M}{Z} + \frac{P}{A}$$

where:

$M =$ applied bending moment

$P =$ applied axial force

$A =$ cross-sectional area of the beam

$Z =$ section modulus of the beam

M, Z, and P are random variables with respective means and c.o.v.s as follows:

$$\mu_M = 45{,}000 \text{ in–lb}; \qquad \delta_M = 0.10$$
$$\mu_Z = 100 \text{ in}^3; \qquad \delta_Z = 0.20$$
$$\mu_P = 5000 \text{ lb}; \qquad \delta_P = 0.10$$
$$A = 50 \text{ in}^2$$

Example-2

Assume that M and P are correlated with a correlation coefficient of $\rho_{M,P} = 0.75$, whereas Z is statistically independent of M and P. We determine the mean and standard deviation of the applied stress S in the beam by first-order approximation as follows:

$$\text{Mean value:} \quad \mu_S \simeq \frac{\mu_M}{\mu_Z} + \frac{\mu_P}{A} = \frac{45{,}000}{100} + \frac{5000}{50} = 550 \text{ psi}$$

and variance:

$$\sigma_S^2 \simeq \sigma_M^2\left(\frac{1}{\mu_Z}\right)^2 + \sigma_Z^2\left(\frac{-\mu_M}{\mu_Z^2}\right)^2 + \sigma_P^2\left(\frac{1}{A}\right)^2 + 2\rho_{M,P}\sigma_M\sigma_P\left(\frac{1}{\mu_Z}\right)\left(\frac{1}{A}\right)$$

$$= 4500^2\left(\frac{1}{100}\right)^2 + 20^2\left(\frac{-45{,}000}{100^2}\right)^2 + 500^2\left(\frac{1}{50}\right)^2 + 2 \times 0.75 \times 4500 \times 500\left(\frac{1}{100}\right)\left(\frac{1}{50}\right)$$

$$= 10{,}900.00$$

from which we obtain the standard deviation of S, $\sigma_S = 104.40$ psi.

Example-2

Based on test data, the strength capacity of the beam, S_c, was estimated to have a mean strength of 800 psi and a standard deviation of 110 psi. Assuming that S and S_c are lognormal variates with the respective means and standard deviations determined above, we evaluate the parameters of the respective lognormal distributions as follows:

$$\zeta_S^2 = \ln\left(1 + \frac{104.40^2}{550^2}\right) = 0.0354; \qquad \zeta_{S_c}^2 = \left(\frac{110}{800}\right)^2 = (0.14)^2$$

and

$$\lambda_S = \ln 550 - \frac{1}{2}(0.0354) = 6.29; \qquad \lambda_{S_c} = \ln 800 - \frac{1}{2}(0.14)^2 = 6.67$$

The safety factor of the beam is defined as $\theta = S_c/S$. As S_c and S are both lognormal variates, the safety factor θ is also lognormal with the parameters

$$\lambda_\theta = \lambda_{S_c} - \lambda_S = 6.67 - 6.29 = 0.38$$

and

$$\zeta_\theta = \sqrt{\zeta_{S_c}^2 + \zeta_S^2} = \sqrt{(0.14)^2 + 0.0354} = 0.23$$

The beam will be overstressed when $\theta < 1.0$; therefore, the probability of this event is

$$P(\theta < 1.0) = \Phi\left(\frac{\ln 1.0 - \lambda_\theta}{\zeta_\theta}\right) = \Phi\left(\frac{0 - 0.38}{0.23}\right) = 1 - \Phi(1.65) = 1 - 0.950 = 0.050$$

That is, there is a 5% chance that the beam will be overstressed under the applied load.

CE 513: Statistical Methods in Civil Engineering

Lectures- 10: Parameter Estimation

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

Parameter Estimation

Probability papers

Let X be a random variable with PDF $P_X(x)$.

Let $\{x_i\}_{i=1}^{n}$ be a sample of X.

Probability paper is a special plotting device in which y-axis is scaled in such a way that the PDF function appears as a straight line.

Example

$P_X(x) = 1 - \exp(-\lambda x)$ $x \geq 0$

$1 - P_X(x) = G_X(x) = \exp(-\lambda x)$

$\log G_X(x) = -\lambda x$

The complement of the cumulative PDF appears as a straight line.

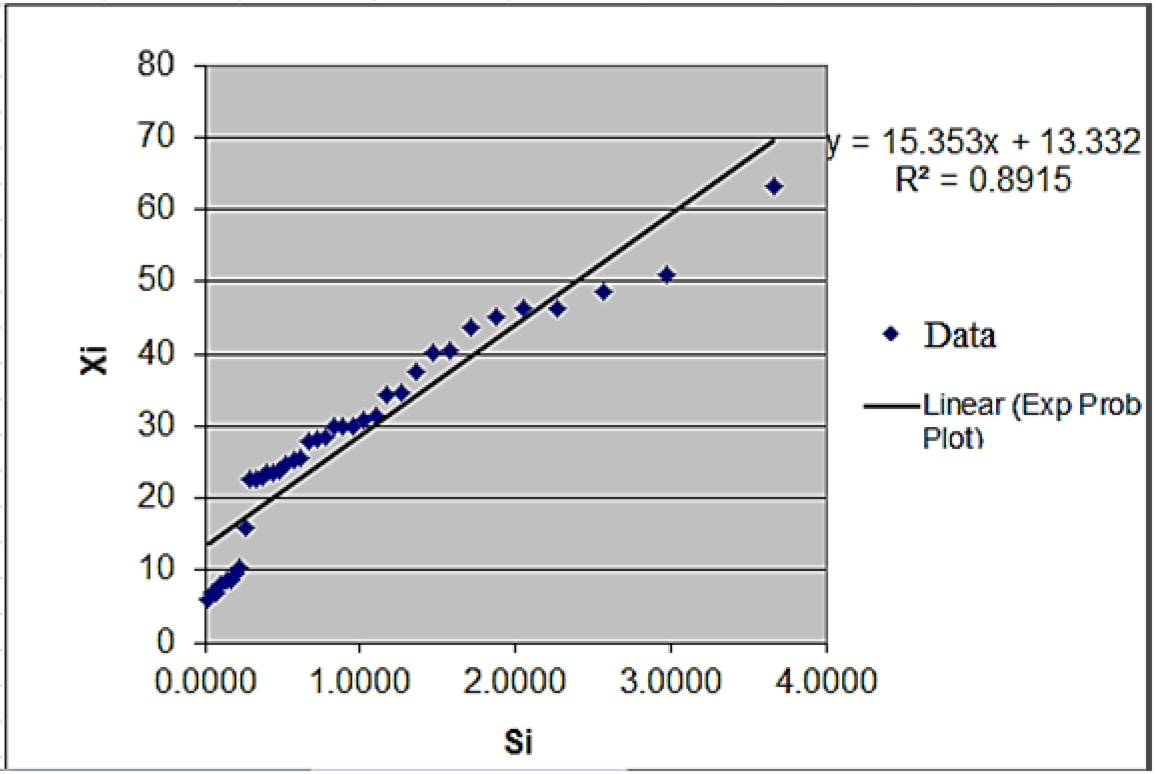
PP plot

Data (in increasing order)	Rank	Pi
5.96	1	0.03
6.83	2	0.05
6.84	3	0.08
8.17	4	0.10
8.68	5	0.13
8.74	6	0.15
9.41	7	0.18
10.36	8	0.21
15.9	9	0.23
22.5	10	0.26
22.7	11	0.28
23	12	0.31
23.509	13	0.33
23.6	14	0.36
23.7	15	0.38
24.7	16	0.41
25.3	17	0.44
25.407	18	0.46
28	19	0.49
28.2	20	0.51
28.5	21	0.54
30	22	0.56
30	23	0.59
30	24	0.62
[illegible]	[illegible]	[illegible]

$i/(N+1)$

PP plot

Si		
0.0260	mu=	46954
0.0526	lambda=	54398
0.0800		
0.1082		
0.1372		
0.1671		
0.1978		
0.2296		
0.2624		
0.2963		
0.3314		
0.3677		
0.4055		
0.4447		
0.4855		
0.5281		
0.5725		
0.6190		
0.6678		
0.7191		
0.7732		
0.8303		
0.8910		
0.9555		



PP plot-for practice

5.96	28
6.83	28.2
6.84	28.5
8.17	30
8.68	30
8.74	30
9.41	30.88
10.36	31.38
15.9	34.28
22.5	34.5
22.7	37.407
23	40.03
23.509	40.48
23.6	43.53
23.7	45
24.7	46.31
25.3	46.397
25.407	48.74
	50.888
	63.319

Maximum Likelihood Estimation

General Mathematical Statement of Estimation Problem:

For… Measured Data $\mathbf{x} = [\, x[0]\; x[1]\; \ldots\; x[N\text{-}1] \,]$

Unknown Parameter $\boldsymbol{\theta} = [\, \theta_1 \;\; \theta_2 \;\; \ldots \;\; \theta_p \,]$

$\boldsymbol{\theta}$ is Not Random

$\mathbf{x}$ is an N-dimensional random data vector

Q: What captures all the statistical information needed for an estimation problem ?

A: Need the N-dimensional PDF of the data, parameterized by θ

$$p(\mathbf{x};\boldsymbol{\theta})$$

We'll use $p(\mathbf{x};\boldsymbol{\theta})$ to find $\hat{\boldsymbol{\theta}} = g(\mathbf{x})$

Maximum Likelihood Estimation

Let $f(x;\theta)$ be the density function of population X

θ is the only parameter to be estimated

from a set of sample values $x_1, x_2, \ldots, x_n$

Joint density function of the sample

$$f(x_1, x_2, \ldots \ldots \ldots \ldots, x_n; \theta)$$

This is in general difficult to work with

- Simplify it by making independence assumption
- Each sample is sampled independently of the others
- Each sample belongs to the same parent distribution

Joint density simplifies to $f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)$

Maximum Likelihood Estimation

A better and somewhat well behaved function: *Likelihood*

We define the *likelihood function* L of a set of n sample values from the population by

$$L(x_1, x_2, \ldots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta).$$

In the case when X is discrete, we write

$$L(x_1, x_2, \ldots, x_n; \theta) = p(x_1; \theta) p(x_2; \theta) \cdots p(x_n; \theta).$$

- Likelihood function L is a function of a single variable θ
- Method of maximum likelihood: Comprises of choosing, as an estimate of θ, the particular value of that maximizes L

Maximum Likelihood Estimation

The maximum of $L(\theta)$ occurs at the value of θ where $\mathrm{d}L(\theta)/\mathrm{d}\theta$ is zero. Hence, in a large number of cases, the *maximum likelihood estimate* (MLE) $\hat{\theta}$ of θ based on sample values $x_1, x_2, \ldots,$ and x_n can be determined from

$$\frac{\mathrm{d}L(x_1, x_2, \ldots, x_n; \hat{\theta})}{\mathrm{d}\hat{\theta}} = 0.$$

Gaussian with known sigma

- the log-likelihood is:

$$\sum_{j=1}^{n} \ln p(\mathbf{x}_j|\theta) = \sum_{j=1}^{n} -\frac{1}{2}(\mathbf{x}_j - \mu)^t \Sigma^{-1}(\mathbf{x}_j - \mu) - \frac{1}{2}\ln (2\pi)^d |\Sigma|$$

- The gradient wrt to the mean is:

$$\nabla_\mu \sum_{j=1}^{n} \ln p(\mathbf{x}_j|\theta) = \sum_{j=1}^{n} \Sigma^{-1}(\mathbf{x}_j - \mu)$$

- Setting the gradient to zero gives:

$$\sum_{j=1}^{n} \Sigma^{-1}(\mathbf{x}_j - \mu^*) = \mathbf{0} \quad \Rightarrow \quad \mu^* = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j$$

Gaussian with unknown mean & sigma

- the log-likelihood is:

$$\mathcal{L} = \sum_{j=1}^{n} -\frac{1}{2\sigma^2}(x_j - \mu)^2 - \frac{1}{2}\ln 2\pi\sigma^2$$

- The gradient is:

$$\nabla_{\mu,\sigma^2}\mathcal{L} = \begin{bmatrix} \sum_{j=1}^{n} \frac{1}{\sigma^2}(x_j - \mu) \\ \sum_{j=1}^{n} -\frac{1}{2\sigma^2} + \frac{(x_j-\mu)^2}{2\sigma^4} \end{bmatrix} = 0$$

$$\mu^* = \frac{1}{n}\sum_{j=1}^{n} x_j \qquad \sigma^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \mu^*)^2$$

Question: Work out the case where sigma is known and varies at each point