Topic

Review of Probability and Statistics

IE 630

Discrete Event System Simulation

Jayendran Venkateswaran

Veeraruna Kavitha

Probability Overview

Text books:

J. Banks, J. S. Carson, B. L. Nelson and D. M. Nicol (2012),
“Discrete Event System Simulation”,
Pearson Education International Series

S. Ross, “INTRODUCTION TO PROBABILITY AND STATISTICS FOR ENGINEERS AND SCIENTISTS”

Experiment, Sample Space & Events

- Experiment is a process whose outcome is not know with certainty
- Sample space is a set of all possible outcomes of an experiment
- Event is any subset of the sample space
- Probabilities are numbers assigned to events that indicate how likely it is that the event will occur when an experiment is performed

Experiment, Sample Space & Events (2)

Example 2a

- Experiment… → Rolling of a die
- Sample space… → 1, 2, 3, 4, 5, 6
- Event… → Outcome is 1

  Outcome is ≥5
- Probabilities indicate how likely it is that the event will occur when experiment is performed

Prob{Outcome is 1} = 1/6

Prob{Outcome is ≥5} = 2/6

Experiment, Sample Space & Events (3)

Example 2b

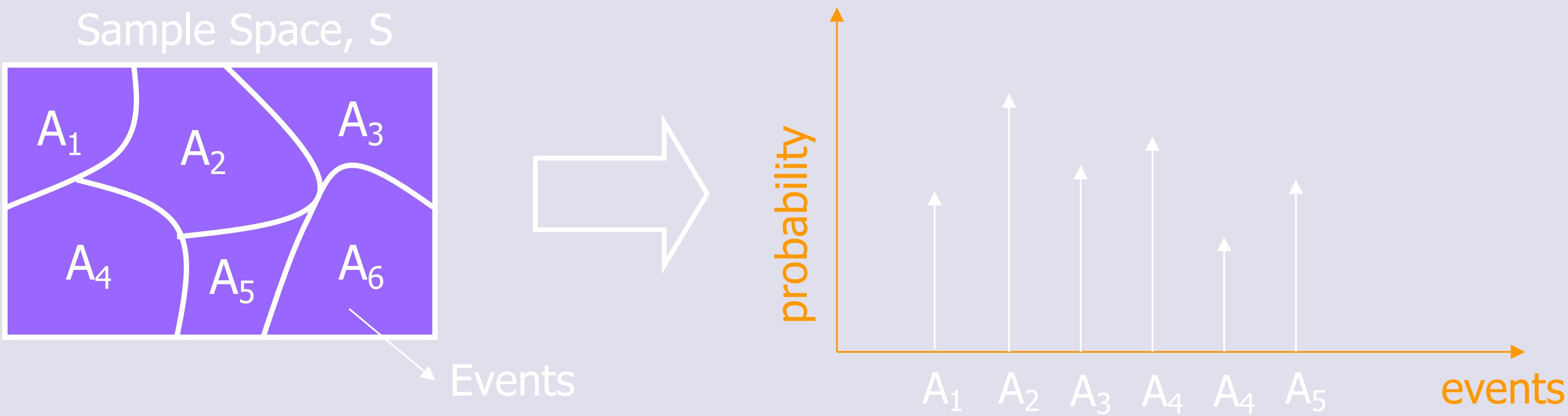
- Experiment… → Toss two fair coins
- Sample space… → {H, H} {H,T} {T, H} {T,T}
- Event… → Outcome is at most one H

  Outcome is at least one H
- Probabilities indicate how likely it is that the event will occur when experiment is performed

Prob{Outcome is at most one H} = 3/4

Prob{Outcome is at least one H} = 3/4

Axioms of Probability



- Axiom 1: $0 \leq P(A) \leq 1$
- Axiom 2: $P(S) = 1$
- Axiom 3: If $A_i \cap A_j = \emptyset$, then

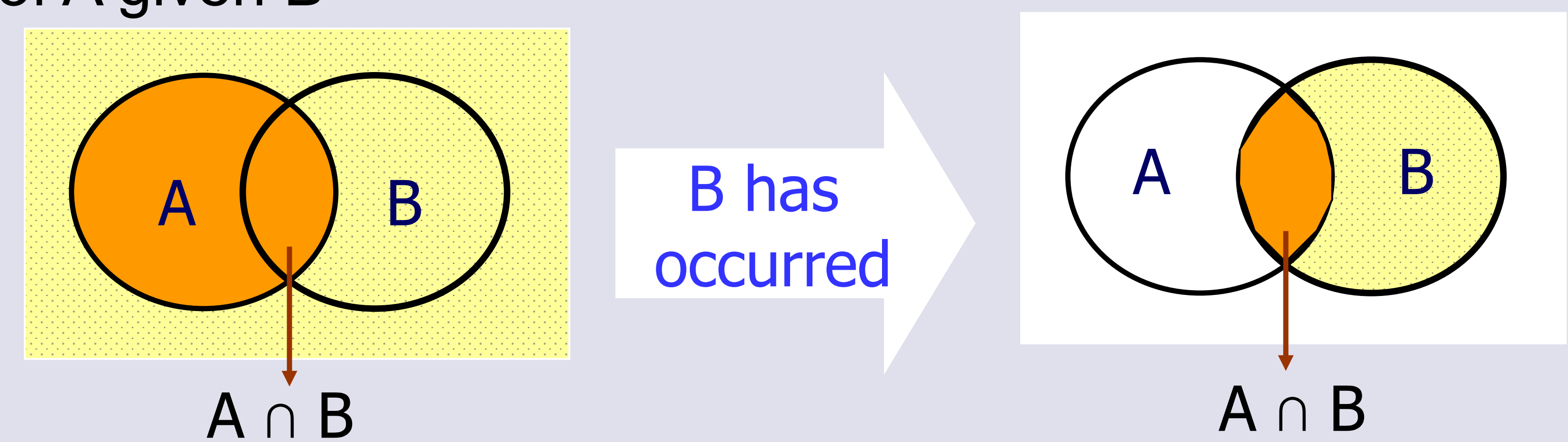
$$P(A_i \cup A_j) = P(A_i) + P(A_j)$$

Conditional Probability

- Suppose A and B are two events in an experiment.
- Probability that event *A* will occur given that *B* has occurred is given by

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad \text{for } P(B) > 0$$

Probability of A given B



Random variable & Distribution function

- Random variable is a function that assigns a real number to each point in sample space
  - Example 2c: Rolling a pair of dice
- Cumulative distribution function $F(x)$ of a random variable X is defined as probability of event $\{X \leq x\}$
- Properties of cdf

1. $0 \leq F(x) \leq 1$ for all x
2. $F(x)$ is non-decreasing
3.

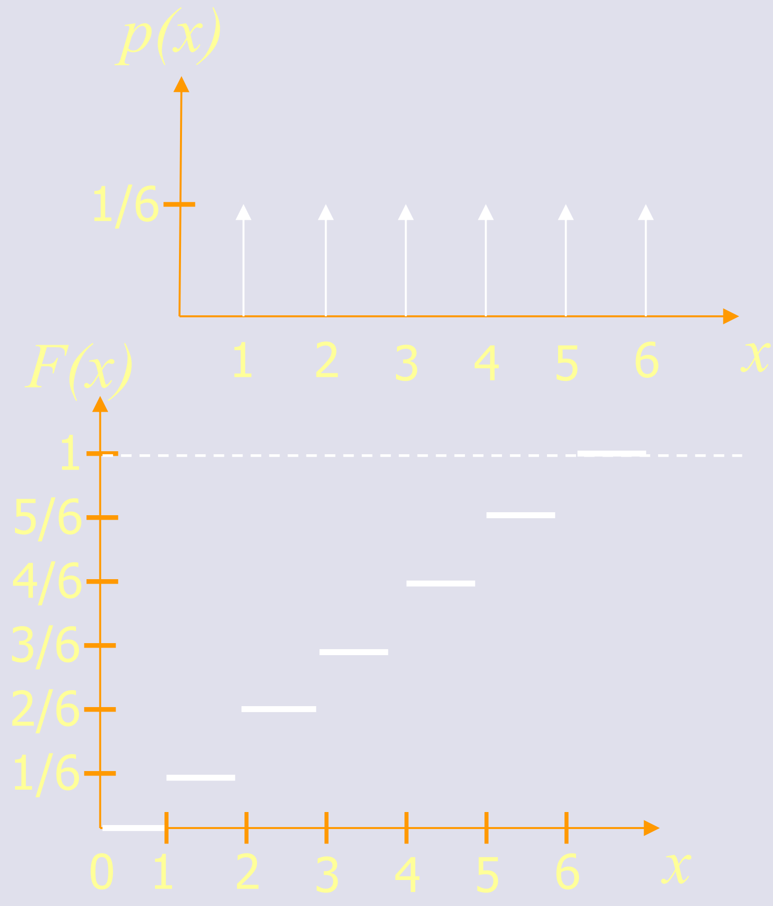
$$\lim_{x \to \infty} F(x) = 1 \quad \text{and} \quad \lim_{x \to -\infty} F(x) = 0$$

Discrete random variable

- X is a discrete random variable if it can take on at most countable number of values
- Probability mass function
- Cumulative distribution function

Example 2d

Number appearing on rolling a die



Continuous random variable

- *X* is continuous if there exists a nonnegative function $f(x)$ such that for any set of real numbers *C*:
- Probability density function $P\{X \in C\} = \int_C f(x)dx$
- Interpretation of density function
  - Probability associated with each value x is zero
  - Density: probability that *X* will be very near x
- Cumulative Distribution function

$F_X(x)$

Percentile and Median

- α percentile: smallest value of x at which the CDF takes on a value of at least α

$$P\{X \le x\} \ge \alpha$$

- Median is the 50-percentile
  - Smallest value of x such that $F(x) \ge 0.5$

Mean or Expectation

- Expectation of *X* or Expected value of *X* or Mean of *X* <u>is weighted average of the possible values that X can take on</u>

- When *X* is discrete random variable
  - Example 2e: Suppose demand is a random variable that takes on the value 1, 2, 3, 4 with respective probabilities 1/8, 1/2, 1/8, 1/4

- When *X* is continuous random variable
  - Example 2f: Suppose pdf of *X* is given by

$$f(x) = 4x^3 \text{ if } 0<x<1, \text{ and } 0 \text{ otherwise}$$

Properties of Mean

- What is $E[g(X)]$? $\qquad$ g is some function

1. $E[aX + b] =$ $aEx + b$
2. $E[X_1 + X_2] =$
3. $E[X_1 - X_2] =$
4. $E[E[X]] =$
5. $E[X - E[X]] =$ $\longrightarrow$
6. $E[(aX)^2] =$

$$\sum_{x \in \Omega_x} g(x)\, p_x(x)$$

$$\int g(x)\, f_x(x)\, dx$$

$\left.\begin{array}{l} a \rightarrow \\ b \rightarrow \end{array}\right\}$ deterministic

$$Y = X - EX$$

$$EY$$

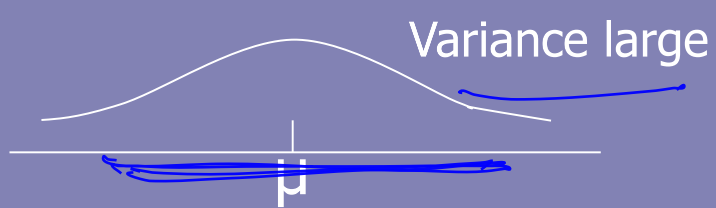
Variance

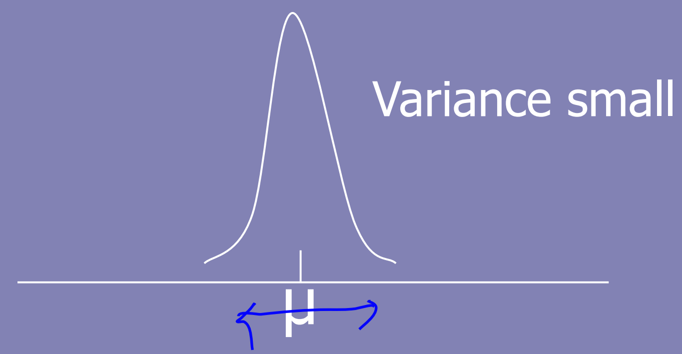
By defn expected

- Variance is the average value of the square of the difference between X and $E[X]$

$$Var(X) = E\left[(X - EX)^2\right]$$

- Variance is the measure of dispersion





- Standard deviation
- Coefficient of variation
  - Normalized measure of dispersion

Covariance

$E[(X-EX)(Y-EY)]$

- Covariance is the measure of dependence between random variables, say, *X* and *Y*
  - Cov(*X*, *Y*) = $E[XY]$ $\quad EX = 0,\ EY = 0$
  - If *X* & *Y* increase or decrease together, Cov(X,Y)>0
  - If *X* tends to increase when *Y* decreases (or vise versa), Cov(X,Y)<0

- Proposition: If *X* and *Y* are independent, then Cov(*X*, *Y*) = 0

  $\hookrightarrow EXY = EX\,EY$

  $E[(X-EX)(Y-EY)] = E[X-EX]\,E[Y-EY]$

- Correlation measure

Properties of Variance & Covariance

1. $\text{Var}(X) \geq$ 0
2. $\text{Var}(aX + b) =$
3. $\text{Var}(X + Y) =$
4. $\text{Var}(X - Y) =$

$\text{Var}(b) = 0$

5. If X and Y are independent
   1. $\text{Var}(X+Y) =$
   2. $E[XY] =$

Special Random Variables

DISCRETE RANDOM VARIABLES

- Discrete
- Binomial
- Poisson
- Geometric

CONTINUOUS RANDOM VARIABLES

Uniform

Exponential

Normal

Triangular

Beta

Gamma

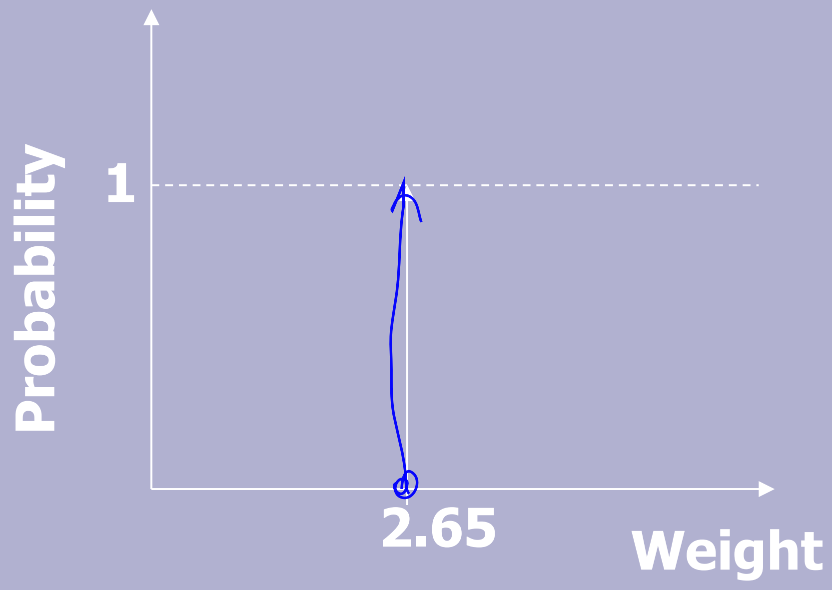
Erlang

LogNormal

Probability Distributions

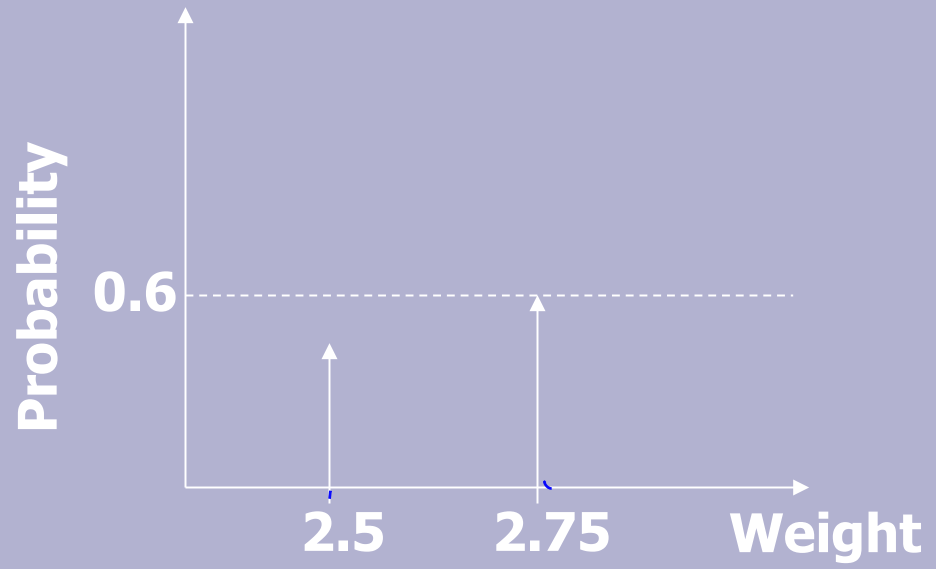
Deterministic

- No distribution at all!
- If deterministic value is used then it is assumed NO uncertainty exists
- Example: Weight = 2.65 Kgs



Discrete Random Variable

- Discrete
  - Random variable takes on only certain outcomes, with associated probabilities
  - Example: Weight = 2.5 with probability 0.4,
    = 2.75 with probability 0.6



Poisson random variable

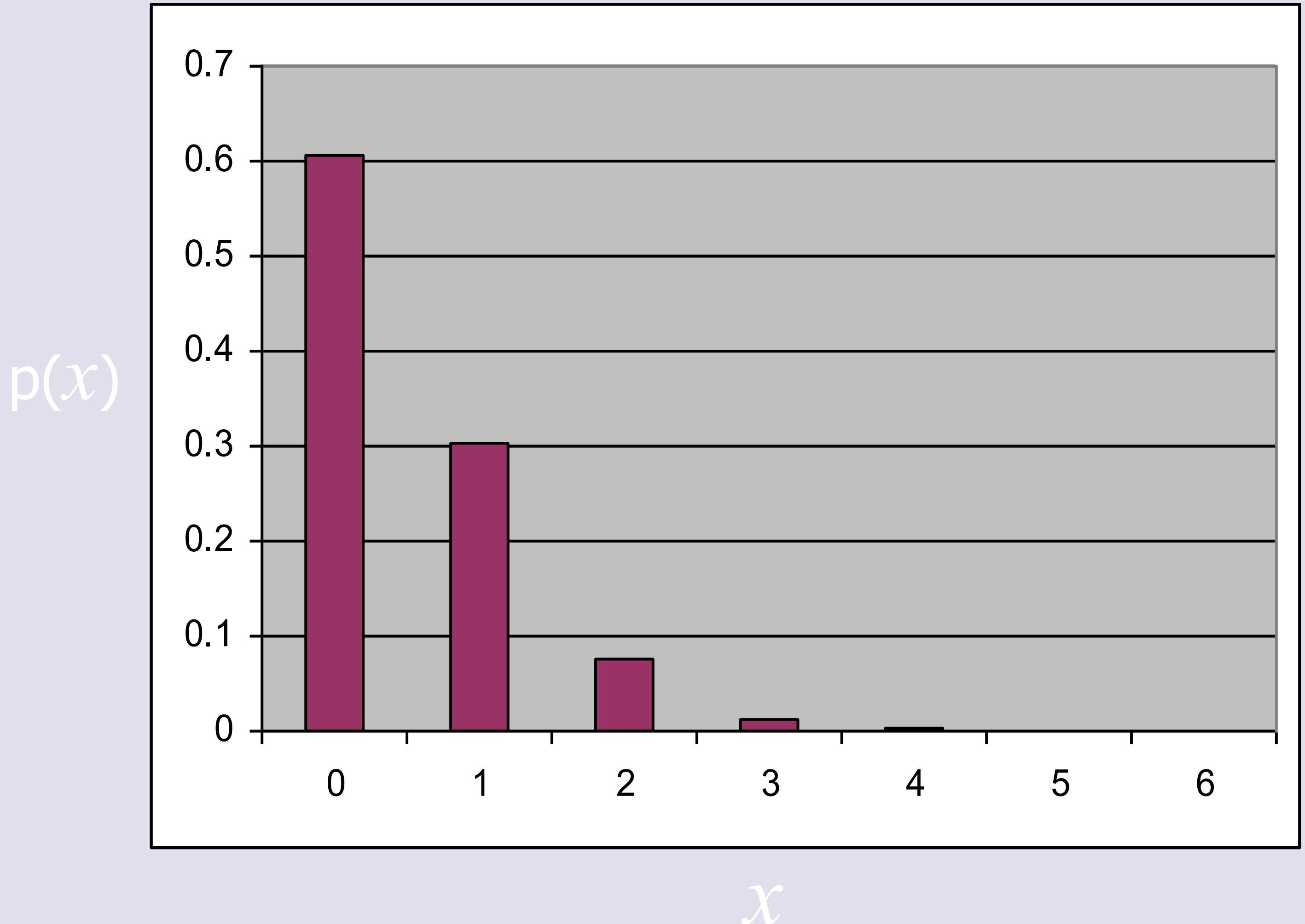
- Discrete random variable takes on one of the values 0, 1, 2, … is said to be Poisson random variable with parameter λ
  - Number of events on an interval of time
  - Number of items demanded from inventory

- Probability mass function

$\{ p_X(x) \}_{x \in \Omega_X}$

- Mean =
- Variance =

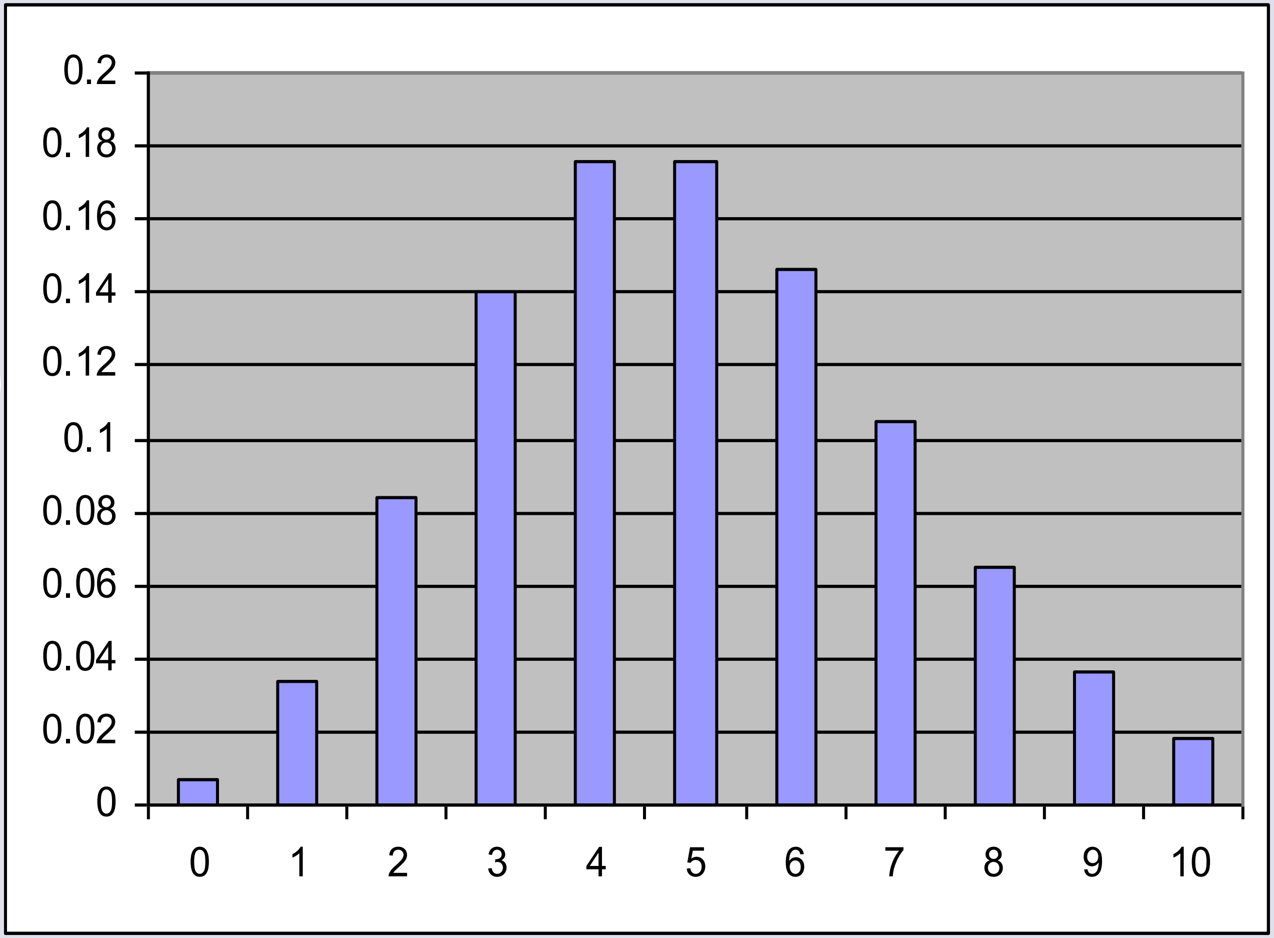
Poisson random variable (2)



With
λ=0.5

Poisson random variable (3)

p(x)



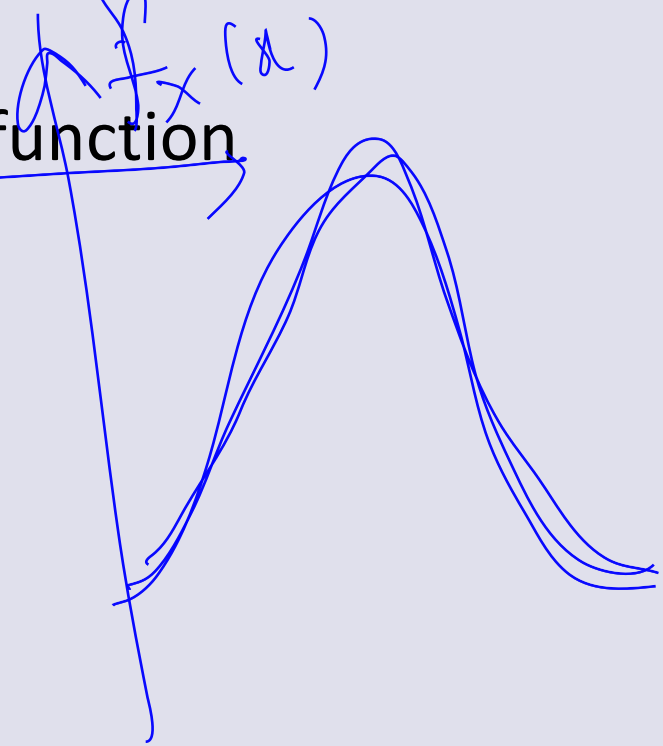
x

With
λ=5

Normal Distribution

- Many natural phenomena can be approximated by Normal distribution
  - Two parameter distribution
  - Bell shaped curve symmetric about the mean

- Probability density function



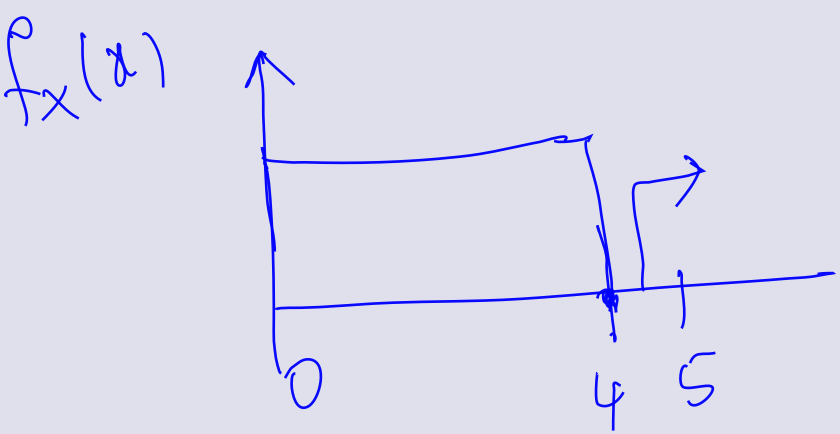
- Mean = 0
- Variance =

Standard Normal Distribution

- N(0, 1) distribution is called standard normal distribution
- Distribution function
- $X \sim N(\mu, \sigma^2)$ can be mapped to $Z \sim N(0,1)$ as follows:

$$Z = (X - \mu) / \sigma$$

- Now can evaluate all probabilities concerning X in terms of Φ



$P(X > 4 + b \mid X > 4)$

Exponential random variable

inter event times

Memoryless

- Continuous random variable
- PDF $\lambda e^{-\lambda x}$
- CDF
- Mean & Variance

$$P(X > a+b \mid X > a) = P(X > b)$$

Geometric

- Exponential random variables are the only continuous random variable with the "memoryless" property

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

show that

$$P(X > x) = e^{-\lambda x}$$

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

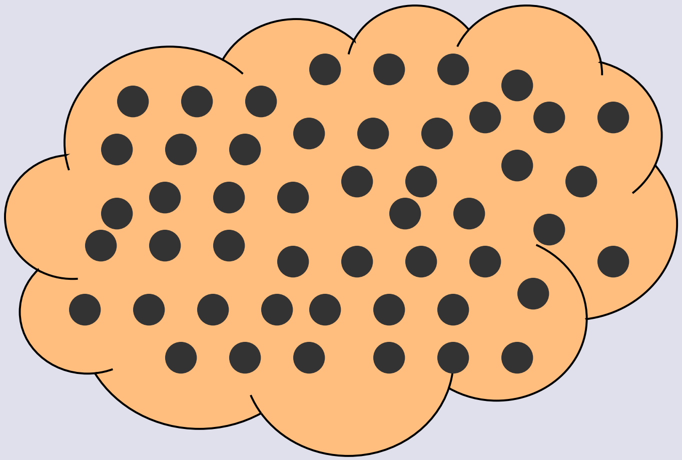
Poisson Process

- Represents arrival process in many systems
- arrivals in banks, call centers, drive in restaurants etc
- If An is the inter-arrival between n-th and (n+1)-th arrival, then An is exponentially distributed with parameter λ.
- A1 is actual arrival time of first one.
- λ – arrival rate ~ number of arrivals / unit time
- N(t) = # of arrivals in time t – Poisson distributed with parameter λt.

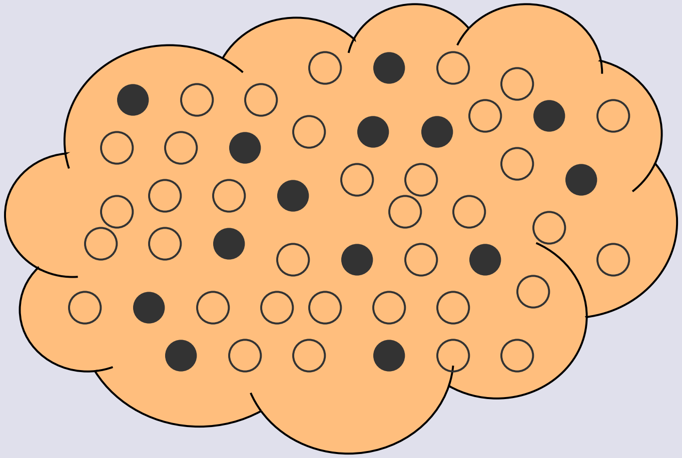
Statistics Overview

Population vs. Sample

- **Population** is the universal set of all things under study
  - Example: all students of IITBombay



- **Sample** is the portion of the population selected for analysis
  - Example: all 2nd year MTech students with CPI > 8.5



Parameter vs. Statistic

- Parameter describes a characteristic of full population
  - Example: 65% of all IITBombay students are males
- Statistic describes a characteristic of the sample
  - Example: 83% of students sampled are males

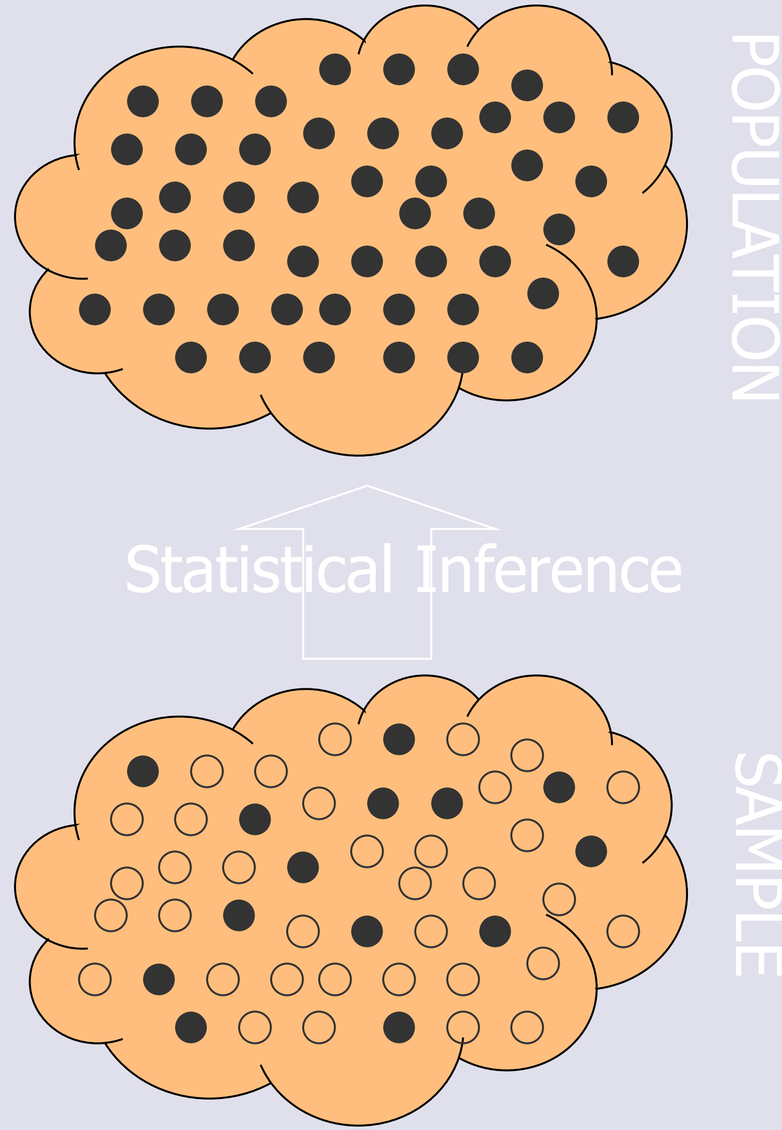
Generally

- Greek letters represent population parameter
- Roman letters represents sample statistic

Descriptive & Inferential Statistics

- Descriptive statistics are methods used to summarize & interpret collected data

- Inferential statistics are methods used to estimate unknown population characteristics based on sample results

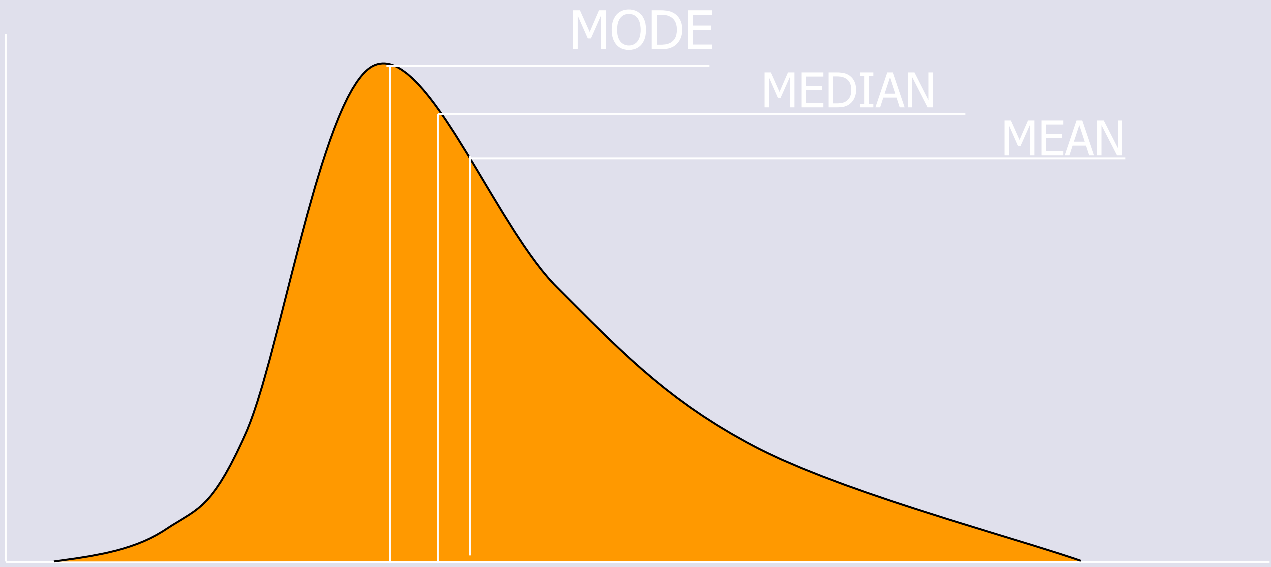


Independent Identically Distributed (i.i.d)

- Independence
  - Two events are independent if occurrence of one does <u>not</u> influence the occurrence of the other
  - Two random variables X_1 and X_2 are independent <u>if and only</u> if for any numbers a_1 and a_2, the events $\{X_1 \leq a_1\}$ and $\{X_2 \leq a_2\}$ are independent
- Identically distributed
  - Random variables are said to be identically distributed if all have the same probability distribution
- *i.i.d random variables simplifies statistical inference*

Measures of Central Tendency

- Central tendency is the measure of “middle” of a data set
- Mean:	Arithmetic average of data set
- Median: Middle value of ordered data set
- Mode:	Most frequent value(s) in data set



Sample Mean

- Sample Mean is the arithmetic average of the data set
- Suppose $X_1, \ldots, X_n$ are iid random variables drawn from a population with (unknown) mean μ and (unknown) variance σ^2
- Sample mean is used to estimate the population mean (μ)
  - Sample mean is centered about population mean but the spread (variance) reduces as sample size increases

Sample Median

- Median is the middle observation of ordered data set
- Example 2g
  1. Median of 9, 3, 6, 4, 2 is
  2. Median of 2, 3, 4, 6, 7, 9 is

- Median is resistant to outliers (unlike mean)
- Example 2h
  - Median of 2, 3, 4, 6, 7, 840 is

Sample Mode

- Mode is value of data set the occurs most frequently
  - Example 2i: Mode of 2, 4, 4, 6, 7, 8 is
- Mode need not be unique
  - Example 2j: Mode of 2, 2, 4, 6, 7, 7, 8, 11 is
- Data can have no MODE
  - Example: 3, 5, 6, 8, 9

Dispersion Statistics

- Mean, Median & Mode insufficient to characterize data sets. For example:
  - Example 2k:
    - Data Set 1: 98, 99, 100, 101, 102
    - Data Set 2: 20, 50, 100, 130, 200
  - Mean & Median are the same (no mode)
  - But, doesn’t the data sets look different?
- Need to quantify the spread of the data set
  - Range
  - Variance
  - Standard Deviation

Sample Range

- Range is the difference between the largest and smallest observations made

- Example
  - Data Set 1: 98, 99, 100, 101, 102 -> Range: 4
  - Data Set 2: 20, 50, 100, 130, 200 -> Range: 180

Sample Variance & Standard Deviation

- Suppose $X_1,\ldots, X_n$ are iid random variables drawn from a population with (unknown) mean μ and (unknown) variance σ^2
- Sample Variance is defined as:
- Sample variance is used to estimate the population variance (σ^2)
  - Note: (n-1) is used to compute S^2 to obtain an unbiased estimator of σ^2 since $E[S^2] = \sigma^2$
- Standard deviation is square root of variance
  - Same units as mean

Central Limit Theorem

- Let $X_1, X_2, \ldots, X_n$ be a sequence of IID random variables each with mean μ and variance σ^2. Then for large *n*, the distribution of $(X_1 + X_2 + \ldots + X_n)$ is approximately Normal with mean $n\mu$ and variance $n\sigma^2$

- How large a sample is required for Normal approximation to be valid?

Sampling from Normal Distribution

- Suppose X_1, X_2, …, X_n be IID sample from Normal population having mean μ and variance σ^2

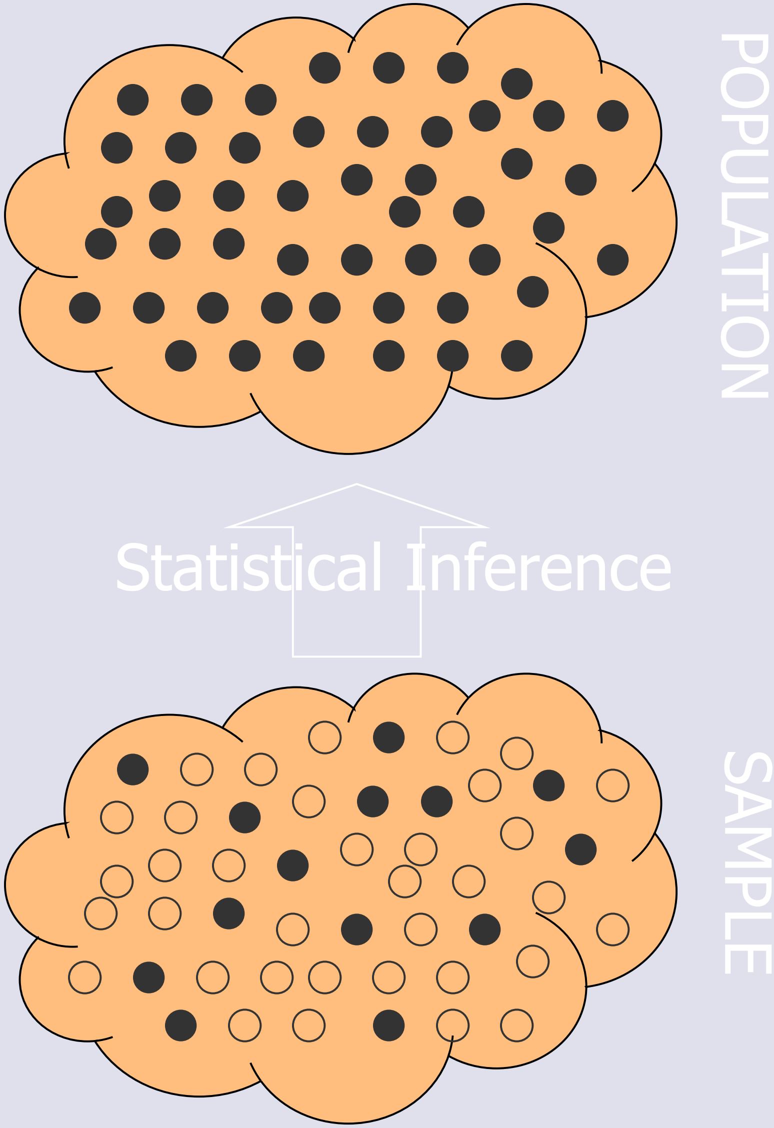
  then, X and S^2 are independent random variables,

  distribution of sample mean is $N(\mu, \sigma^2/n)$,

  and distribution of $(n-1)S^2/\sigma^2$ is chi-square with $n-1$ degrees of freedom

Statistical Inference

- How to use observed data (sample) to make inferences about the unknown population parameters?
- Let $X_1, X_2, \ldots, X_n$ be random sample from population distribution F_{ϑ} where ϑ is vector of unknown parameters



Estimators

- Any statistic (based on sample) used to estimate the value of unknown parameter ϑ (of population) is called an *estimator of ϑ*

- Maximum Likelihood Estimation (MLE)
  - Popular method of parameter estimation
  - *"If I assume that the underlying population distribution is ___________ then my best guess for the parameters will be so-and-so"*

- Point Estimates

Confidence Interval Estimates

- Interval within which we have certain level of confidence that μ falls.
  - 100(1 – α)% CI
  - CI attempts to bound the error between sample mean and population mean
  - Can construct one-sided CI or two-sided CI
- $X_1, \ldots, X_n$ be samples from Normal population
  1. 100(1 – α)% CI when σ^2 known
  2. 100(1 – α)% CI when σ^2 unknown

Statistical Hypothesis Testing

- Hypotheses: Are claims about population characteristics
  - It can be about the nature of distribution, or parameters of population distribution

- Hypothesis testing
  - "Can the difference between observed result and expected result be attributed to chance?"
  - Testing by contradiction
  - Remember: We take a decision about the population characteristics based on samples

Statistical Hypothesis Testing (2)

- Null hypothesis, H_0
  - Samples are purely by chance
- Alternate Hypothesis, H_A
  - Typically, Not-H_0
  - Samples are result of some real effect

- Define a test statistic
  - If test statistic lies in some predefined region C then 'reject' null hypothesis, else 'accept' it.
  - Many types of test statistics

Statistical Hypothesis Testing (3)

- Types of errors
  - Type I error: Rejecting H_0 when it is correct
  - Type II error: Accepting H_0 when it is false

- Classical way:
  - Specify a value α (level of significance of test)
  - Now, we require that the probability of type I error < α

Statistical Hypothesis Testing (4)

- *p*-value
  - It is the probability of getting a result at least as extreme as the observed result.
  - Lower the p-value, more ‘significant’ the result
  - Statistical significant : it is unlikely to have occurred by chance
  - H_0 is accepted if α < p-value, else rejected