CPSC 531: System Modeling and Simulation

Carey Williamson
Department of Computer Science
University of Calgary
Fall 2017

Overview

- The world a model-builder sees is probabilistic rather than deterministic:
  - Some probability model might well describe the variations

Goals:

- Review the fundamental concepts of probability
- Understand the difference between discrete and continuous random variable
- Review the most common probability models

UNIVERSITY OF CALGARY

Outline

- Probability and random variables
  - Random experiment and random variable
  - Probability mass/density functions
  - Expectation, variance, covariance, correlation
- Probability distributions
  - Discrete probability distributions
  - Continuous probability distributions
  - Empirical probability distributions

Outline

- Probability and random variables
  - Random experiment and random variable
  - Probability mass/density functions
  - Expectation, variance, covariance, correlation
- Probability distributions
  - Discrete probability distributions
  - Continuous probability distributions
  - Empirical probability distribution

Probability

Is widely used in mathematics, science, engineering, finance and philosophy to draw conclusions about the likelihood of potential events and the underlying mechanics of complex systems

- Probability is a measure of how likely it is for an event to happen
- We measure probability with a number between 0 and 1
- If an event is certain to happen, then the probability of the event is 1
- If an event is certain not to happen, then the probability of the event is 0

Random Experiment

- An experiment is called *random* if the outcome of the experiment is uncertain
- For a random experiment:
  - The set of all possible outcomes is known before the experiment
  - The outcome of the experiment is not known in advance
- *Sample space* Ω of an experiment is the set of all possible outcomes of the experiment
- Example: Consider random experiment of tossing a coin twice. Sample space is:

$$\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$$

Probability of Events

- An *event* is a subset of sample space

Example 1: in tossing a coin twice, $E=\{(H,H)\}$ is the event of having two heads

Example 2: in tossing a coin twice, $E=\{(H,H), (H,T)\}$ is the event of having a head in the first toss

- *Probability* of an event E is a numerical measure of the likelihood that event E will occur, expressed as a number between 0 and 1,

$$0 \leq \mathbb{P}(E) \leq 1$$

  - If all possible outcomes are equally likely: $\mathbb{P}(E) = |E|/|\Omega|$
  - Probability of the sample space is 1: $\mathbb{P}(\Omega) = 1$

Joint Probability

- Probability that two events A and B occur in a single experiment:

$$\mathbb{P}(A \textbf{ and } B) = \mathbb{P}(A \cap B)$$

- Example: drawing a single card at random from a regular deck of cards, probability of getting a red king
  - A: getting a red card
  - B: getting a king
  - $\mathbb{P}(A \cap B) = \frac{2}{52}$

Independent Events

- Two events A and B are independent if the occurrence of one does not affect the occurrence of the other:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

- Example: drawing a single card at random from a regular deck of cards, probability of getting a red king
  – A: getting a red card $\Rightarrow \mathbb{P}(A) = 26/52$
  – B: getting a king $\Rightarrow \mathbb{P}(B) = 4/52$
  – $\mathbb{P}(A \cap B) = \frac{2}{52} = \mathbb{P}(A)\mathbb{P}(B) \Rightarrow A$ and B are independent

Mutually Exclusive Events

- Events A and B are mutually exclusive if the occurrence of one implies the non-occurrence of the other, i.e., $A \cap B = \phi$:
$$\mathbb{P}(A \cap B) = 0$$

- Example: drawing a single card at random from a regular deck of cards, probability of getting a red club
  - A: getting a red card
  - B: getting a club
  - $\mathbb{P}(A \cap B) = 0$

- Complementary event of event A is event $[not\ A]$, i.e., the event that A does not occur, denoted by $\bar{A}$
  - Events A and $\bar{A}$ are mutually exclusive
  - $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$

Union Probability

- Union of events A and B:

$$\mathbb{P}(A \textbf{ or } B) = \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

- If A and B are mutually exclusive:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$

- Example: drawing a single card at random from a regular deck of cards, probability of getting a red card or a king
  - A: getting a red card $\Rightarrow \mathbb{P}(A) = 26/52$
  - B: getting a king $\Rightarrow \mathbb{P}(B) = 4/52$
  - $\mathbb{P}(A \cap B) = \frac{2}{52}$
  - $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = \frac{26}{52} + \frac{4}{52} - \frac{2}{52} = \frac{28}{52}$

Conditional Probability

- Probability of event A given the occurrence of some event B:

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

- If events A and B are independent:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = P(A)$$

- Example: drawing a single card at random from a regular deck of cards, probability of getting a king given that the card is red
  - A: getting a red card $\Rightarrow \mathbb{P}(A) = 26/52$
  - B: getting a king $\Rightarrow \mathbb{P}(B) = 4/52$
  - $\mathbb{P}(A \cap B) = \frac{2}{52}$
  - $\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{2}{26} = \mathbb{P}(B)$

Random Variable

- A numerical value can be associated with each outcome of an experiment
- A *random variable* X is a function from the sample space Ω to the real line that assigns a real number $X(s)$ to each element s of Ω

$$X: \Omega \rightarrow R$$

- Random variable takes on its values with some probability

Random Variable

- Example: Consider random experiment of tossing a coin twice. Sample space is:

$$\Omega = \{(H,H), (H,T), (T,H), (T,T)\}$$

Define random variable X as the number of heads in the experiment:

$$X((T,T)) = 0,\; X((H,T))=1,$$
$$X((T,H)) = 1, X((H,H))=2$$

- Example: Rolling a die.
  Sample space $\Omega = \{1,2,3,4,5,6)$.

  Define random variable X as the number rolled:

$$X(j) = j, \qquad 1 \le j \le 6$$

- Example: roll two fair dice and observe the outcome

  Sample space = $\{(i,j) \mid 1 \leq i \leq 6,\ 1 \leq j \leq 6\}$

  i: integer from the first die

  j: integer from the second die

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Possible outcomes

- Random variable X: sum of the two faces of the dice

$$X(i,j) = i+j$$

  - $\mathbb{P}(X = 12) = \mathbb{P}(\ (6,6)\ ) = 1/36$
  - $\mathbb{P}(X = 10) = \mathbb{P}(\ (5,5), (4,6), (6,4)\ ) = 3/36$

- Random variable Y: value of the first die
  - $\mathbb{P}(Y = 1) = 1/6$
  - $\mathbb{P}(\ Y = i) = 1/6, \quad 1 \leq i \leq 6$

Types of Random Variables

- Discrete
  - Random variables whose set of possible values can be written as a finite or infinite sequence
  - Example: number of requests sent to a web server
- Continuous
  - Random variables that take a continuum of possible values
  - Example: time between requests sent to a web server

Probability Mass Function (PMF)

- X: discrete random variable
- $p(x_i)$: probability mass function of X, where

$$p(x_i) = \mathbb{P}(X = x_i)$$

- Properties:

$$0 \leq p(x_i) \leq 1$$

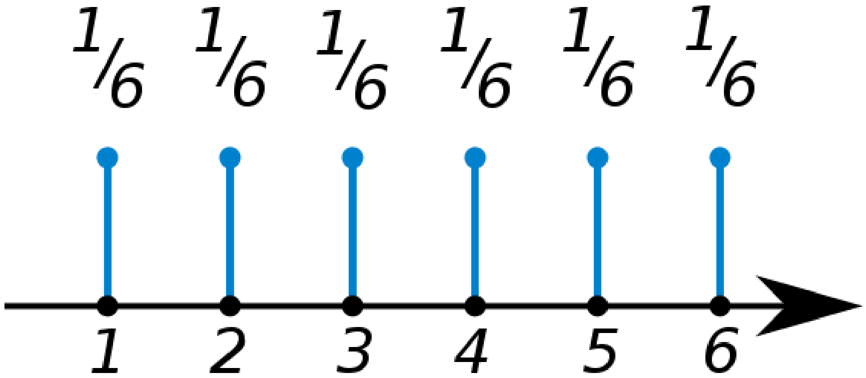
$$\sum_{x_i} p(x_i) = 1$$

- Number of heads in tossing three coins

x_i	$p(x_i)$
0	1/8
1	3/8
2	3/8
3	1/8

$$\sum_{x_i} p(x_i) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$$

- Number rolled in rolling a fair die



$$\sum_{x_i} p(x_i) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$$

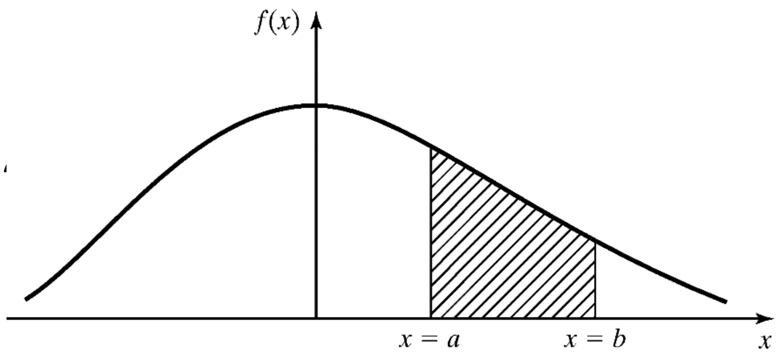
Probability Density Function (PDF)

- X: continuous random variable
- $f(x)$: probability density function of X

$$f(x) = \frac{d}{dx} F(x)$$

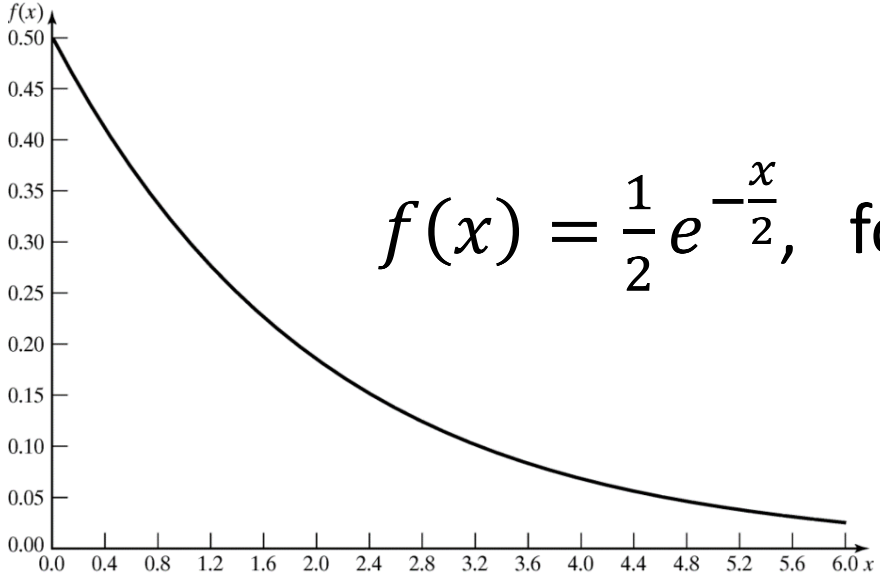
CDF of X

- Note:
  - $\mathbb{P}(X = x) = 0$!!
  - $f(x) \neq \mathbb{P}(X = x)$
  - $\mathbb{P}(x \leq X \leq x + \Delta x) \approx f(x)$
- Properties:
  - $\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx$
  - $\int_{-\infty}^{+\infty} f(x)dx = 1$



Probability Density Function

Example: Life of an inspection device is given by X, a continuous random variable with PDF:



$$f(x) = \frac{1}{2} e^{-\frac{x}{2}}, \quad \text{for } x \geq 0$$

- X has an exponential distribution with mean 2 years
- Probability that the device's life is between 2 and 3 years:

$$\mathbb{P}(2 \leq X \leq 3) = \frac{1}{2} \int_{2}^{3} e^{-\frac{x}{2}} dx = 0.14$$

Cumulative Distribution Function (CDF)

- X: discrete or continuous random variable
- $F(x)$: cumulative probability distribution function of X, or simply, probability distribution function of X

$$F(x) = \mathbb{P}(X \leq x)$$

  - If X is discrete, then $F(x) = \sum_{x_i \leq x} p(x_i)$
  - If X is continuous, then $F(x) = \int_{-\infty}^{x} f(t)dt$

- Properties
  - $F(x)$ is a non-decreasing function, i.e., if $a < b$, then $F(a) \leq F(b)$
  - $\lim_{x \to +\infty} F(x) = 1$, and $\lim_{x \to -\infty} F(x) = 0$

- All probability questions about X can be answered in terms of the CDF, e.g.:

$$\mathbb{P}(a < X \leq b) = F(b) - F(a), \text{ for all } a \leq b$$

Discrete random variable example.

- Rolling a die, X is the number rolled
  - $p(i) = \mathbb{P}(X = i) = 1/6, \qquad 1 \leq i \leq 6$
  - $F(i) = \mathbb{P}(X \leq i)$

$$= p(1) + \cdots + p(i)$$

$$= i/6$$

Cumulative Distribution Function

Continuous random variable example.

- The inspection device has CDF:

$$F(x) = \frac{1}{2}\int_0^x e^{-t/2}dt = 1 - e^{-x/2}$$

  - The probability that the device lasts for less than 2 years:

$$\mathbb{P}(X \leq 2) = F(2) = 1 - e^{-1} = 0.632$$

  - The probability that it lasts between 2 and 3 years:

$$\mathbb{P}(2 \leq X \leq 3) = F(3) - F(2) = (1 - e^{-\frac{3}{2}}) - (1 - e^{-1}) = 0.145$$

Joint Probability Distribution

- Joint probability distribution of random variables X and Y is defined as

$$F(x, y) = \mathbb{P}(X \leq x, Y \leq y)$$

- X and Y are independent random variables if

$$F(x, y) = F_X(x) \cdot F_Y(y)$$

  - Discrete: $p(x, y) = p_X(x) \cdot p_Y(y)$
  - Continuous: $f(x, y) = f_X(x) \cdot f_Y(y)$

Expectation of a Random Variable

- **Mean** or **Expected Value**:

$$\mu = E[X] = \begin{cases} \sum_{i=1}^{n} x_i p(x_i) & \text{discrete } X \\ \int_{-\infty}^{\infty} x f(x) dx & \text{continuous } X \end{cases}$$

- Example: number of heads in tossing three coins

$$\begin{aligned} E[X] &= 0 \cdot p(0) + 1 \cdot p(1) + 2 \cdot p(2) + 3 \cdot p(3) \\ &= 1 \cdot 3/8 + 2 \cdot 3/8 + 3 \cdot 1/8 \\ &= 12/8 \\ &= 1.5 \end{aligned}$$

Expectation of a Function

- $g(X)$: a real-valued function of random variable X
- How to compute $E[g(X)]$?
  - If X is discrete with PMF $p(x)$:

$$E[g(X)] = \sum_{x} g(x)p(x)$$

  - If X is continuous with PDF $f(x)$:

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx$$

- Example: X is the number rolled when rolling a die
  - PMF: $p(x) = 1/6$, for $x = 1,2, \ldots, 6$

$$E[X^2] = \sum_{x=1}^{6} x^2 p(x) = \frac{1}{6}(1 + 2^2 + \cdots + 6^2) = \frac{91}{6} = 15.17$$

Properties of Expectation

- X, Y: two random variables
- a, b: two constants

$$E[aX] = aE[X]$$

$$E[X + b] = E[X] + b$$

$$E[X + Y] = E[X] + E[Y]$$

Misuses of Expectations

- ***Multiplying means to get the mean of a product***

$$E[XY] \neq E[X]E[Y]$$

- Example: tossing three coins
  - X: number of heads
  - Y: number of tails
  - $E[X] = E[Y] = 3/2 \Rightarrow E[X]E[Y] = 9/4$
  - $E[XY] = 3/2$

$$\Rightarrow E[XY] \neq E[X]E[Y]$$

- ***Dividing means to get the mean of a ratio***

$$E\left[\frac{X}{Y}\right] \neq \frac{E[X]}{E[Y]}$$

Variance of a Random Variable

- The variance is a measure of the *spread* of a distribution around its mean value
- Variance is symbolized by $V[X]$ or $Var[X]$ or σ^2:
  - Mean is a way to describe the *location* of a distribution
  - Variance is a way to capture its *scale or degree* of being spread out
  - The unit of variance is the square of the unit of the original variable
- σ: standard deviation
  - Defined as the square root of variance $V[X]$
  - Expressed in the same units as the mean

Variance of a Random Variable

- **Variance**: The expected value of the square of distance between a random variable and its mean

$$\sigma^2 = V[X]$$

$$= E[(X-\mu)^2] = \begin{cases} \sum_{i=1}^{n} (x_i - \mu)^2 p(x_i) & \text{discrete } X \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx & \text{continuous } X \end{cases}$$

where, $\mu = E[X]$

- Equivalently:

$$\sigma^2 = E[X^2] - (E[X])^2$$

- Example: number of heads in tossing three coins

$$E[X] = 1.5$$

$$\begin{aligned}\sigma^2 &= (0 - 1.5)^2 \cdot p(0) + (1 - 1.5)^2 \cdot p(1) \\ &\quad + (2 - 1.5)^2 \cdot p(2) + (3 - 1.5)^2 \cdot p(3) \\ &= 9/4 \cdot 1/8 + 1/4 \cdot 3/8 + 1/4 \cdot 3/8 + 9/4 \cdot 1/8 \\ &= 24/32 \\ &= 3/4\end{aligned}$$

Variance of a Random Variable

- Example: The mean of life of the previous inspection device is:

$$E[X]=\frac{1}{2}\int_0^\infty xe^{-x/2}dx = -xe^{-x/2}\Big|_0^\infty + \int_0^\infty e^{-x/2}dx = 2$$

- To compute variance of X, we first compute $E[X^2]$:

$$E[X^2]=\frac{1}{2}\int_0^\infty x^2e^{-x/2}dx = -x^2e^{-x/2}\Big|_0^\infty + 2\int_0^\infty xe^{-x/2}dx = 8$$

- Hence, the variance and standard deviation of the device's life are:

$$V[X]=8-2^2=4$$

$$\sigma=\sqrt{V[X]}=2$$

Properties of Variance

- X, Y: two random variables
- a, b: two constants

$$V[X] \geq 0$$

$$V[aX] = a^2 V[X]$$

$$V[X + b] = V[X]$$

- If X and Y are independent:

$$V[X + Y] = V[X] + V[Y]$$

Coefficient of Variation

- **Coefficient of Variation**:

$$\text{CV} = \frac{\text{Standard Deviation}}{\text{Mean}} = \frac{\sigma}{\mu}$$

- Example: number of heads in tossing three coins

$$\text{CV} = \frac{\sqrt{3/4}}{3/2} = \frac{1}{\sqrt{3}}$$

- Example: inspection device

$$\left.\begin{aligned} E[X] &= 2 \\ \sigma &= 2 \end{aligned}\right\} \Rightarrow \text{CV} = 1$$

Covariance

- Covariance between random variables X and Y denoted by $Cov(X,Y)$ or $\sigma^2_{X,Y}$ is a measure of how much X and Y change together

$$\sigma^2_{X,Y} = E[(X - E[X])(Y - E[Y])]$$
$$= E[XY] - E[X]E[Y]$$

- For independent variables, the covariance is zero:

$$E[XY] = E[X]E[Y]$$

- Note: Although independence always implies zero covariance, the reverse is **not** true

Covariance

- Example: tossing three coins
  - X: number of heads
  - Y: number of tails
  - $E[X] = E[Y] = 3/2$
- $E[XY]$?
  - X and Y depend on each other
  - $Y = 3 - X$
  - $E[XY] = 0 \times P(0) + 2 \times P(2)$
    $= 3/2$
- $\sigma^2_{X,Y} = E[XY] - E[X]E[Y]$
  $= 3/2 - 3/2 \times 3/2$
  $= -3/4$

x	*y*	*xy*	*p(x)*
0	3	0	1/8
1	2	2	3/8
2	1	2	3/8
3	0	0	1/8

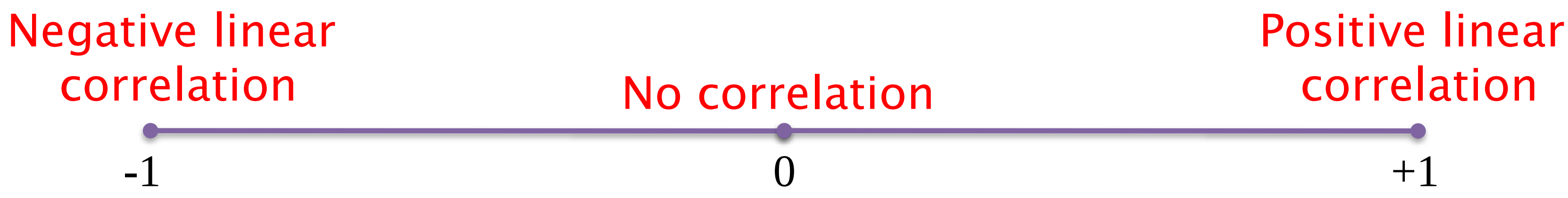
xy	*p(xy)*
0	2/8
2	6/8

Correlation

- Correlation Coefficient between random variables X and Y, denoted by $\rho_{X,Y}$, is the normalized value of their covariance:

$$\rho_{X,Y} = \frac{\sigma^2_{X,Y}}{\sigma_X \sigma_Y}$$

- Indicates the strength and direction of a linear relationship between two random variables
- The correlation always lies between -1 and +1



- Example: tossing three coins

$$\rho_{X,Y} = \frac{-3/4}{\sqrt{3/4}\sqrt{3/4}} = -1$$

Outline

- Probability and random variables
  - Random experiment and random variable
  - Probability mass/density functions
  - Expectation, variance, covariance, correlation
- Probability distributions
  - Discrete probability distributions
  - Continuous probability distributions
  - Empirical probability distribution

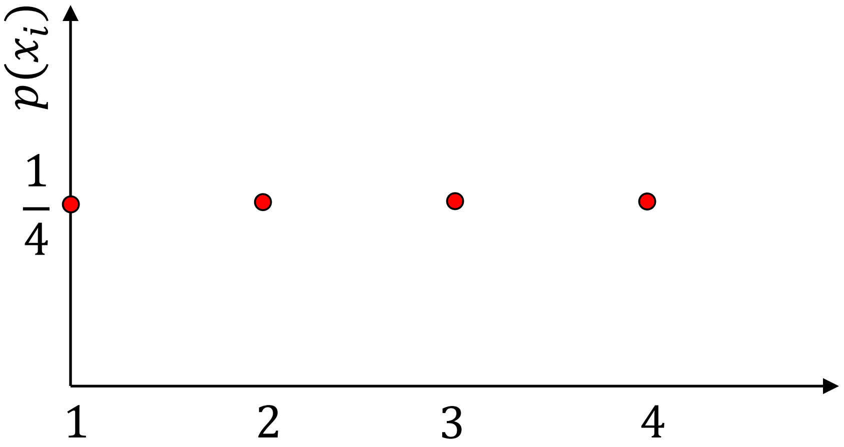
Outline

- Probability and random variables
  - Random experiment and random variable
  - Probability mass/density functions
  - Expectation, variance, covariance, correlation
- Probability distributions
  - Discrete probability distributions
  - Continuous probability distributions
  - Empirical probability distribution

Discrete Uniform Distribution

- A random variable X has discrete uniform distributed if each of the n values in its range, say $x_1, x_2, \dots, x_n$, has equal probability.

- PMF: $p(x_i) = \mathbb{P}(X = x_i) = \frac{1}{n}$



Discrete Uniform Distribution

- Consider a discrete uniform random variable X on the consecutive integers $a, a+1, a+2, \ldots, b$, for $a \leq b$. Then:

$$E[X] = \frac{b+a}{2}$$

$$V[X] = \frac{(b-a+1)^2 - 1}{12}$$

Bernoulli Trial

- Consider an experiment whose outcome can be a success with probability p or a failure with probability $1 - p$:
  - $X = 1$ if the outcome is a success
  - $X = 0$ if the outcome is a failure
- X is a Bernoulli random variable with parameter p
  - where $0 \leq p \leq 1$ is the success probability
- PMF:

$$\mathrm{p}(1) = \mathbb{P}(X = 1) = p$$
$$\mathrm{p}(0) = \mathbb{P}(X = 0) = 1 - p$$

- Properties:
  - $E[X] = p$ and $V[X] = p(1 - p)$

UNIVERSITY OF CALGARY

Binomial Distribution

- X: number of successes in n $(n = 1,2, \ldots)$ independent Bernoulli trials with success probability p

- X is a binomial random variable with parameters (n, p)

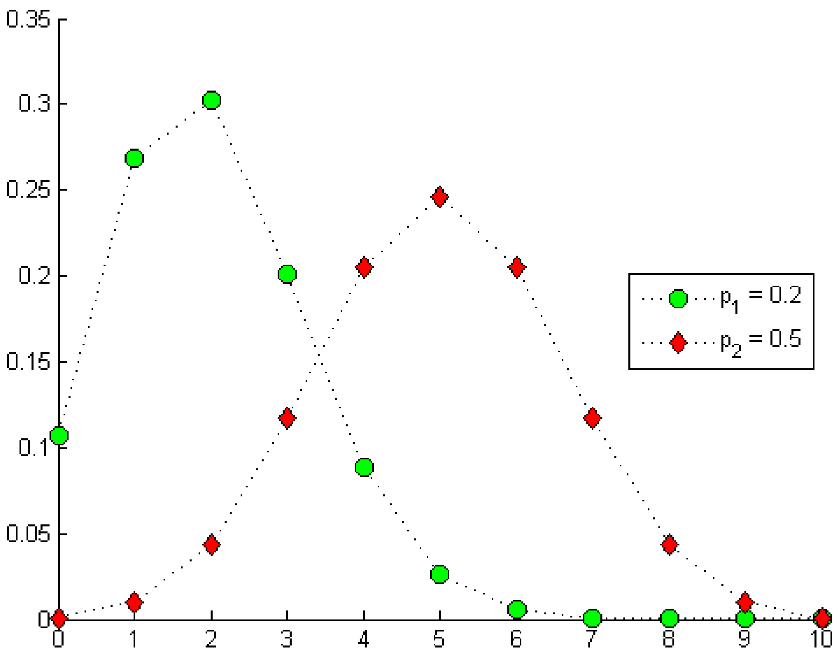
- PMF: Probability of having k $(k = 0,1,2, \ldots, n)$ successes in n trials

$$p(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

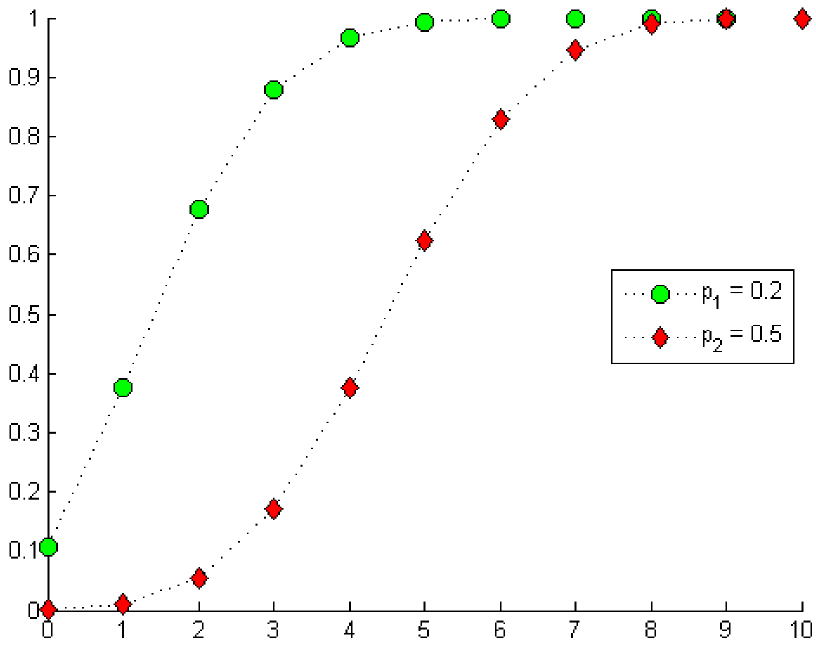
where, $$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- Properties:
  - $E[X] = np$ and $V[X] = np(1-p)$

Example: Binomial Distribution



Binomial distribution PMF
$(n = 10)$



Binomial distribution CDF
$(n = 10)$

Geometric Distribution

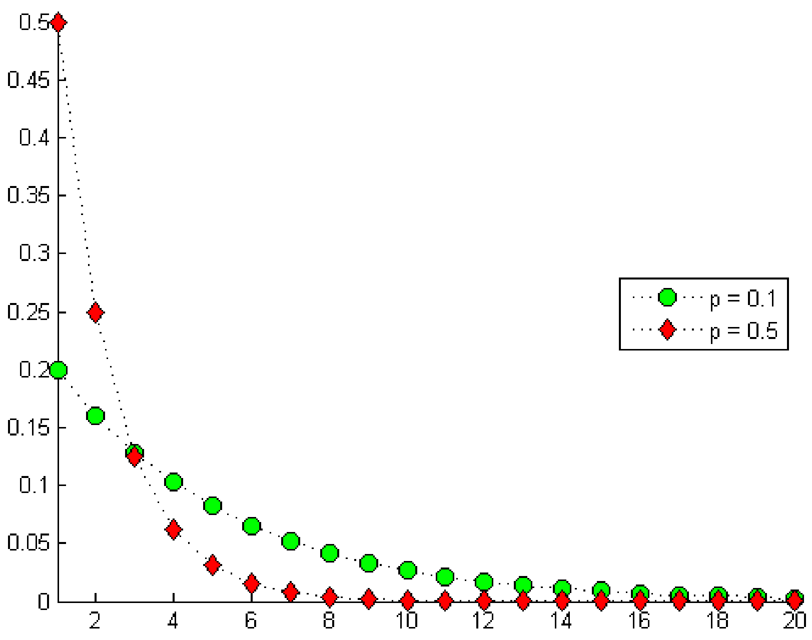
- X: number of Bernoulli trials until achieving the first success
- X is a geometric random variable with success probability p
- PMF: probability of k $(\mathrm{k} = 1,2,3, \ldots)$ trials until the first success

$$p(k) = p(1-p)^{k-1}$$

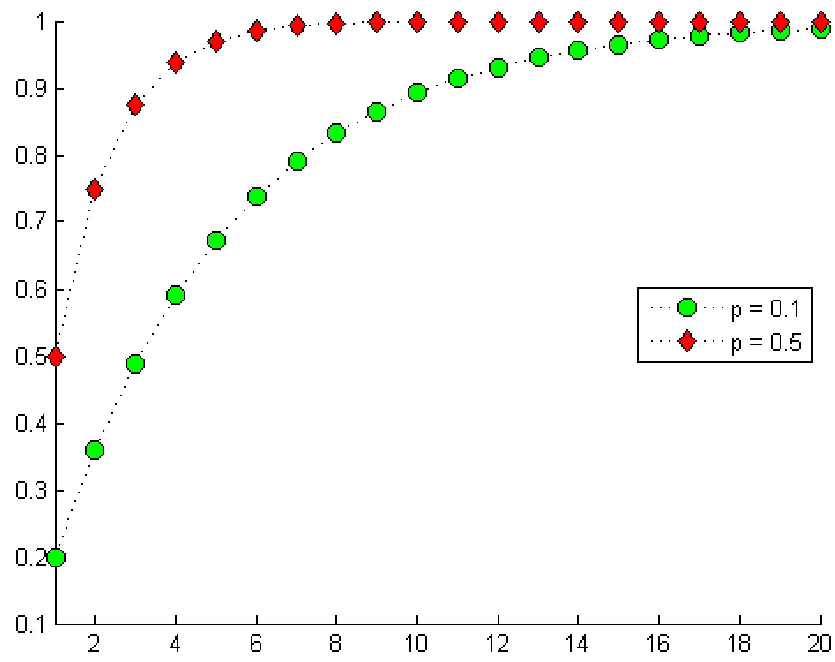
- CDF: $F(k) = 1 - (1-p)^k$
- Properties:

$$E[X] = \frac{1}{p}, \text{ and } V[X] = \frac{1-p}{p^2}$$

Example: Geometric Distribution



Geometric distribution PMF



Geometric distribution CDF

Poisson Distribution

- Number of events occurring in a fixed time interval
  - Events occur with a known rate and are independent
- Poisson distribution is characterized by the rate λ
  - Rate: the average number of event occurrences in a fixed time interval
- Examples
  - The number of calls received by a switchboard per minute
  - The number of packets coming to a router per second
  - The number of travelers arriving to the airport for flight registration per hour

Poisson Distribution

Random variable X is Poisson distributed with rate parameter λ

- PMF: the probability that there are exactly k events in a time interval

$$p(k) = \mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda},\ k = 0,1,2, \ldots$$

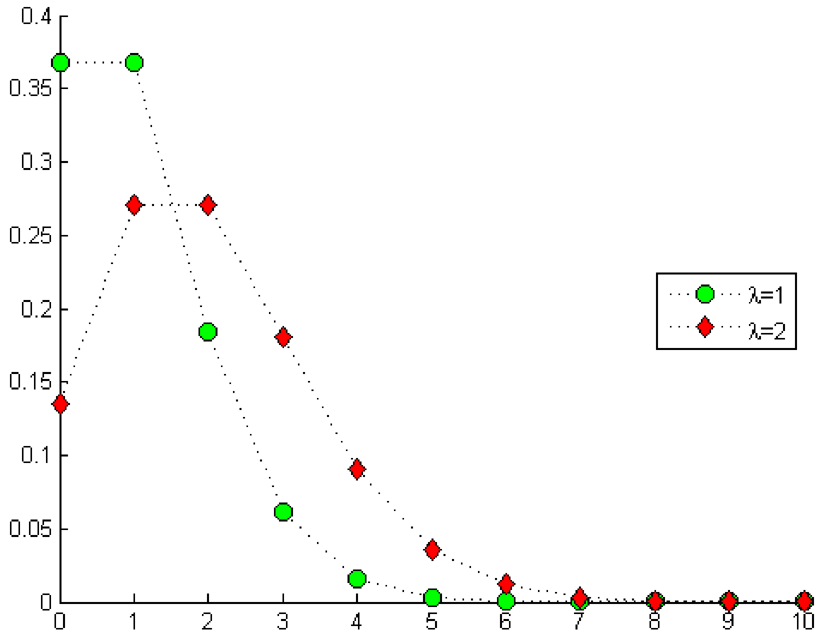
- CDF: the probability of at least k events in a time interval

$$F(k) = \mathbb{P}(X \leq k) = \sum_{i=0}^{k} \frac{\lambda^i}{i!} e^{-\lambda}$$

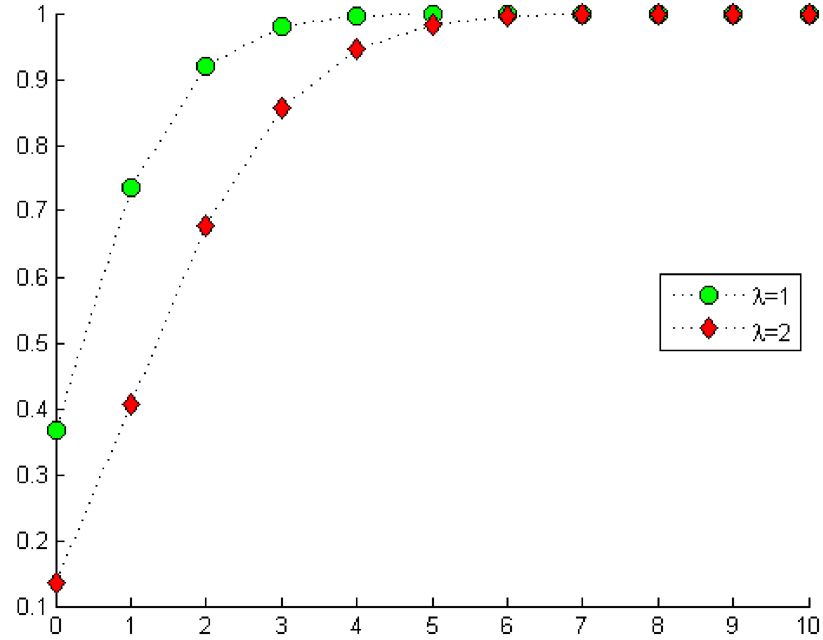
- Properties:

$E[X] = \lambda$
$V[X] = \lambda$

Example: Poisson Distribution



Poisson distribution PMF



Poisson distribution CDF

Example: Poisson Distribution

The number of cars that enter a parking lot follows a Poisson distribution with a rate equal to $\lambda = 20$ cars/hour

- The probability of having exactly 15 cars entering the parking lot in one hour:

$$\mathrm{p}(15) = \frac{20^{15}}{15!} e^{-20} = 0.051649$$

- The probability of having more than 3 cars entering the parking lot in one hour:

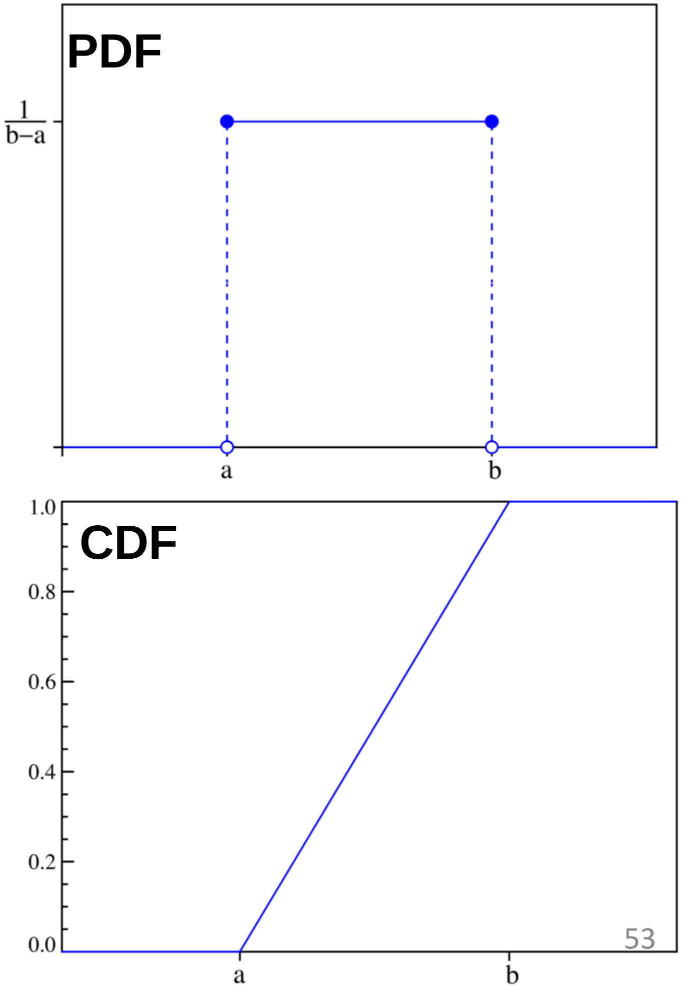
$$\mathbb{P}(X > 3) = 1 - \mathbb{P}(X \leq 3)$$
$$= 1 - [P(0) + P(1) + P(2) + P(3)]$$
$$= 0.9999967$$

Outline

- Probability and random variables
  - Random experiment and random variable
  - Probability mass/density functions
  - Expectation, variance, covariance, correlation
- Probability distributions
  - Discrete probability distributions
  - Continuous probability distributions
  - Empirical probability distribution

Uniform Distribution

- A random variable X has **continuous uniform distribution** on the interval $[a,b]$, if its PDF and CDF are:
  - PDF: $f(x) = \frac{1}{b-a}$, for $a \le x \le b$
  - CDF: $F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x \le \mathrm{b} \\ 1 & x > b \end{cases}$



- Properties:

$$E[X] = \frac{a+b}{2}, \text{ and } V[X] = \frac{(a-b)^2}{12}$$

Uniform Distribution Properties

- $\mathbb{P}(x_1 < X < x_2)$ is proportional to the length of the interval $x_2 - x_1$

$$\mathbb{P}(x_1 < X < x_2) = \frac{x_2 - x_1}{b - a}$$

- Special case: standard uniform distribution denoted by $X \sim U(0,1)$

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

  - Very useful for random number generation in simulations

Exponential Distribution

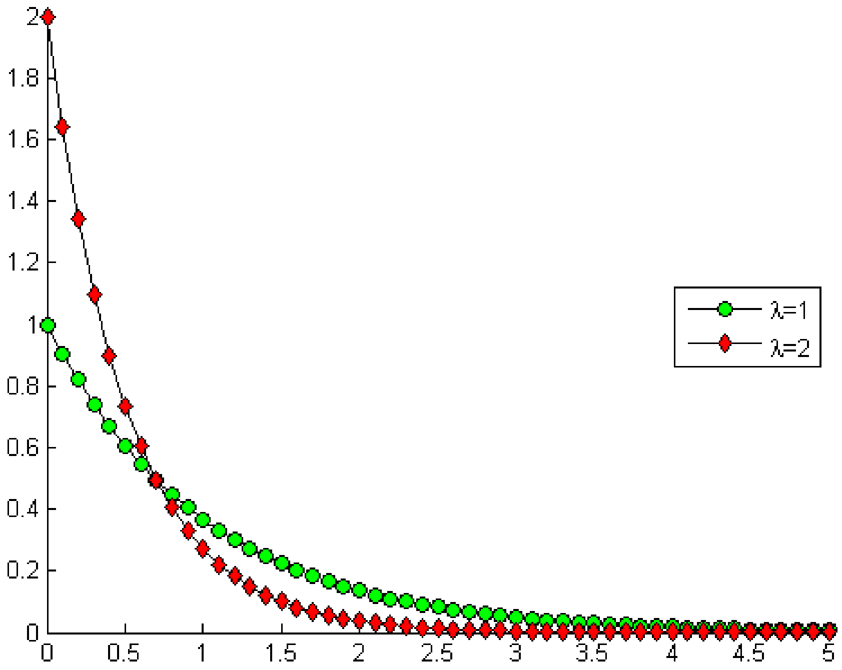
- A random variable X is exponentially distributed with parameter λ if its PDF and CDF are:
  - PDF: $f(x) = \lambda\, e^{-\lambda x}$, for $x \geq 0$
  - CDF: $F(x) = 1 - e^{-\lambda x}$, for $x \geq 0$

- Properties:

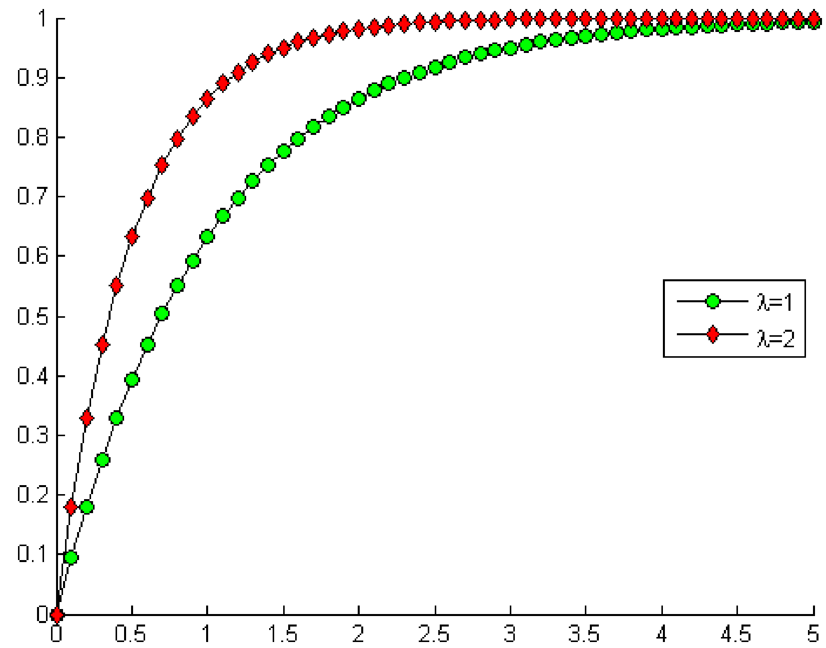
$E[X] = \frac{1}{\lambda}$, and $V[X] = \frac{1}{\lambda^2}$

- The exponential distribution describes the time between consecutive events in a Poisson process of rate λ

Example: Exponential Distribution



Exponential distribution PDF



Exponential distribution CDF

Memoryless Property

- Memoryless is a property of certain probability distributions such as exponential distribution and geometric distribution
  - future events do not depend on the past events, but only on the present event

- Formally: random variable X has a memoryless distribution if

$$\mathbb{P}(X > t + s \mid X > s) = \mathbb{P}(X > t), \quad \text{for } s, t \geq 0$$

- Example: The probability that you will wait t more minutes given that you have already been waiting s minutes is the same as the probability that you wait for more than t minutes from the beginning!

Example: Exponential Distribution

- The time needed to repair the engine of a car is **exponentially distributed** with a mean time equal to 3 hours.
  - The probability that the car spends more than the average wait time in repair:

$$\mathbb{P}(X > 3) = 1 - F(3) = e^{-\frac{3}{3}} = 0.368$$

  - The probability that the car repair time lasts between 2 to 3 hours is:

$$\mathbb{P}(2 \leq X \leq 3) = F(3) - F(2) = 0.145$$

  - The probability that the repair time lasts for another hour given that it has already lasted for 2.5 hours:

Using the memoryless property of the exponential distribution,

$$\mathbb{P}(X > (1 + 2.5) | X > 2.5) = \mathbb{P}(X > 1) = 1 - F(1) = e^{-\frac{1}{3}} = 0.717$$

Normal Distribution

- The **Normal distribution**, also called the **Gaussian distribution**, is an important continuous probability distribution applicable in many fields

- It is specified by two parameters: **mean** (μ) and **variance** (σ^2)

- The importance of the normal distribution as a statistical model in **natural** and **behavioral** sciences is due in part to the **Central Limit Theorem**

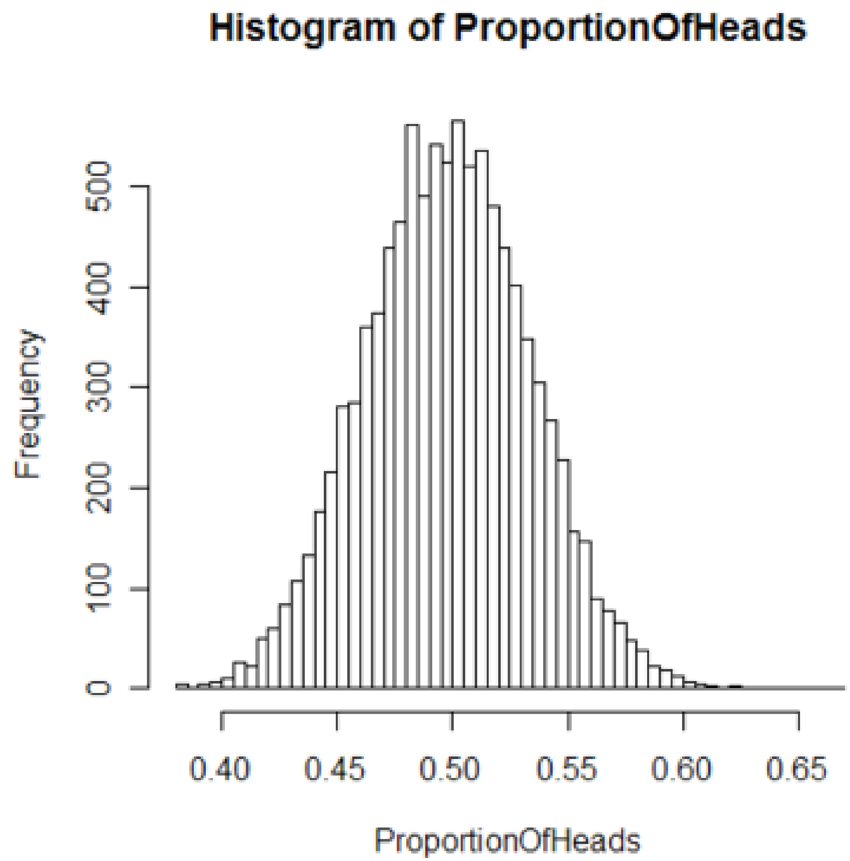
- It is usually used to model system error (e.g. channel error), the distribution of natural phenomena, height, weight, etc.

- There are two main reasons for the popularity of the normal distribution:

1. ***The sum of n independent normal variables is a normal variable.*** If, $X_i \sim N(\mu_i, \sigma_i)$ then $X = \sum_{i=1}^{n} a_i X_i$ has a normal distribution with mean $\mu = \sum_{i=1}^{n} a_i \mu_i$ and variance $\sigma^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2$

2. ***The mean of a large number of independent observations from any distribution tends to have a normal distribution.*** This result, which is called **central limit theorem**, is true for observations from all distributions
=> Experimental errors caused by many factors are normal

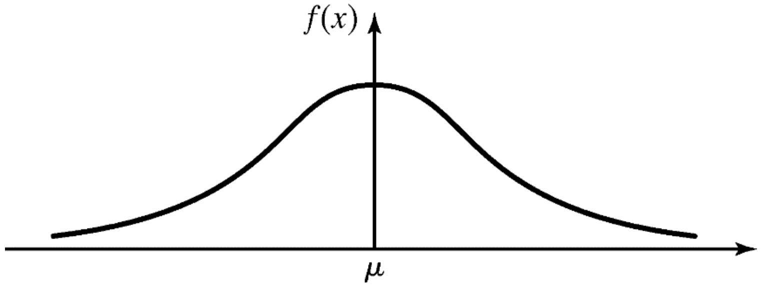
Central Limit Theorem



Histogram plot of average proportion of heads in a fair coin toss, over a large number of sequences of coin tosses.

Normal Distribution

- Random variable X is normally distribution with parameters (μ, σ^2), i.e., $X \sim N(\mu, \sigma^2)$:
  - PDF: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, for $-\infty \leq x \leq +\infty$
  - CDF: does not have any closed form!
  - $E[X] = \mu$, and $V[X] = \sigma^2$

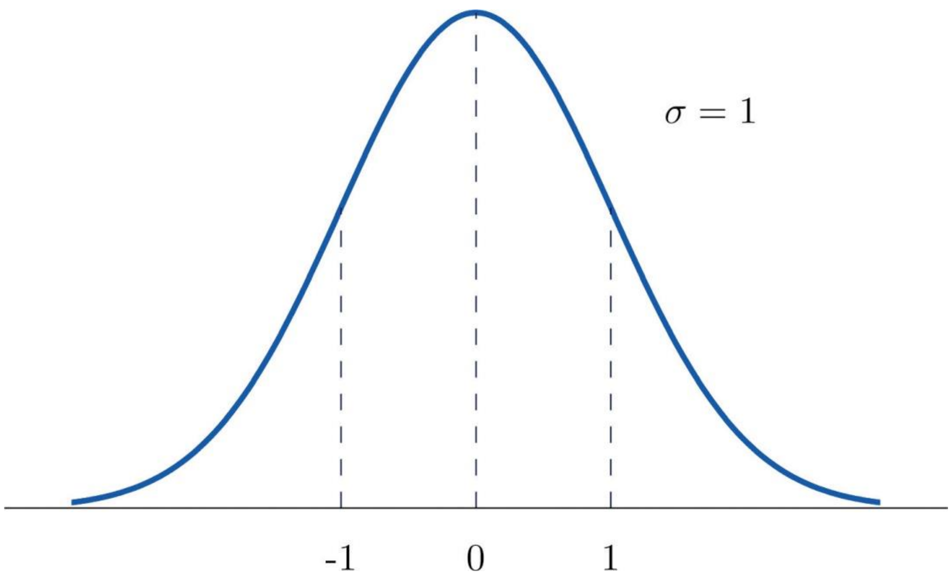


- Properties:
  - $\lim_{x \to \pm\infty} f(x) = 0$
  - Normal PDF is a symmetrical, bell-shaped curve centered at its expected value μ
  - Maximum value of PDF occurs at $x = \mu$

Standard Normal Distribution

- Random variable Z has Standard Normal Distribution if it is normally distributed with parameters $(0, 1)$, i.e., $Z \sim N(0, 1)$:
  - PDF: $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, for $-\infty \leq x \leq +\infty$
  - CDF: commonly denoted by $\Phi(z)$:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}x^2}\, dx$$



- Evaluating the distribution $X \sim N(\mu, \sigma^2)$:
  - $F(x) = \mathbb{P}(X \leq x)$?

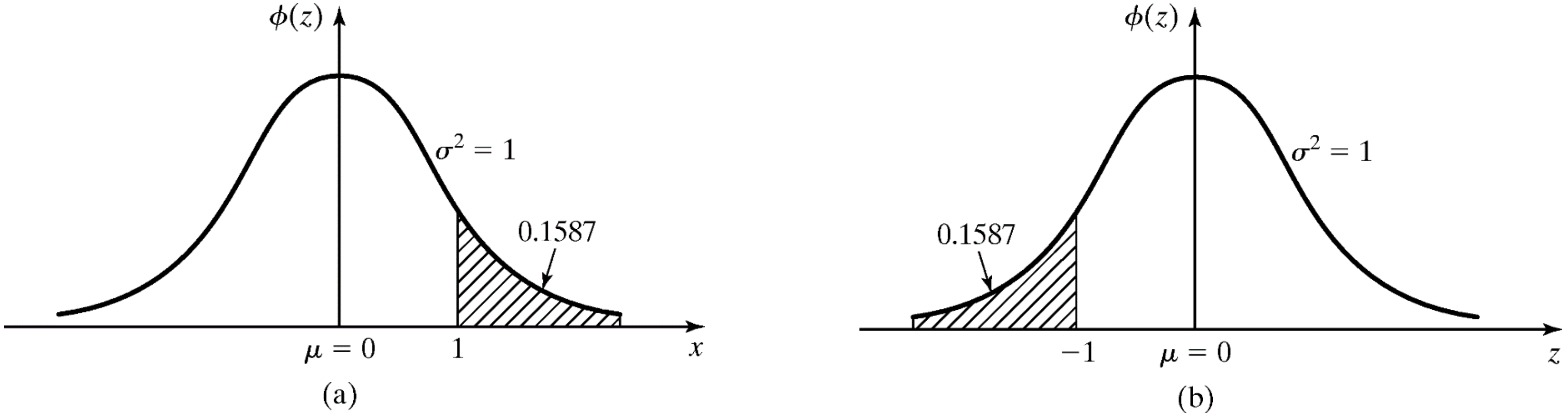
Two techniques:

1. Use numerical methods (no closed form)
2. Use the standard normal distribution
   - $\Phi(z)$ is widely tabulated
   - Use the transformation $Z = \frac{X-\mu}{\sigma}$
   - If $X \sim N(\mu, \sigma^2)$ then $Z \sim N(0, 1)$, i.e., standard normal distribution:

$$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right)$$
$$= \mathbb{P}\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

- Example: The time required to load an oceangoing vessel, X, is distributed as $N(12, 4)$
  - The probability that the vessel is loaded in less than 10 hours:

$$F(10) = \Phi\left(\frac{10-12}{2}\right) = \Phi(-1) = 0.1587$$



  - Using the symmetry property, $\Phi(1)$ is the complement of Φ (-1)

Stochastic Process:

Collection of random variables indexed over time

- **Example:**
  - $N(t)$: number of jobs at the CPU of a computer system over time
  - Take several identical systems and observe $N(t)$
  - The number $N(t)$ at any time t is a random variable
  - Can find the probability distribution functions for $N(t)$ at each possible value of t
- Notation: $\{N(t): t \geq 0\}$

Poisson Process

- Counting Process:
  A stochastic process that represents the total number of events occurred in the time interval $[0, t]$

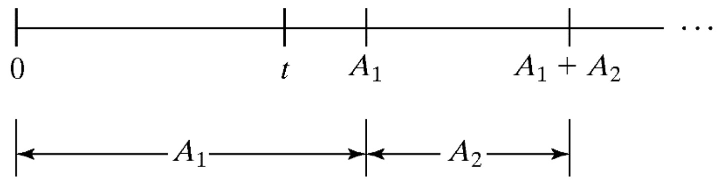
- Poisson Process:
  The counting process $\{N(t), t \geq 0\}$ is a Poisson process with rate λ, if:
  - $N(0) = 0$
  - The process has independent increments
  - The number of events in any interval of length t is Poisson distributed with mean λt. That is, for all $s, t \geq 0$

$$\mathbb{P}(N(t+s) - N(s) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

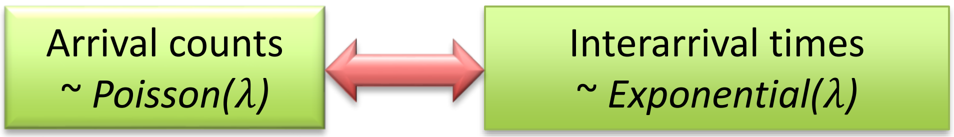
Property: equal mean and variance: $E[N(t)] = V[N(t)] = \lambda t$

Interarrival Times

- Consider the interarrival times of a Poisson process with rate λ, denoted by $A_1, A_2, \ldots$, where A_i is the elapsed time between arrival i and arrival $i+1$



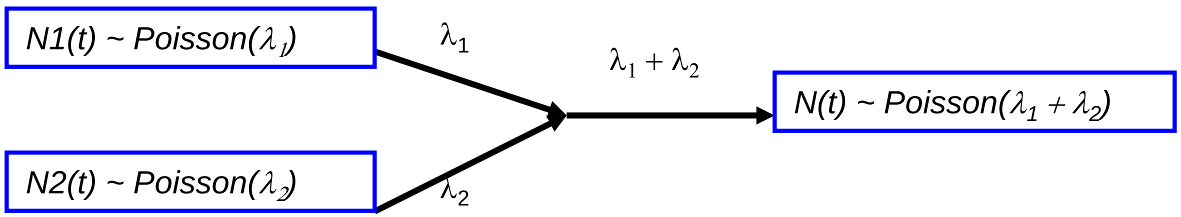
- Interarrival times, $A_1, A_2, \ldots$ are independent identically distributed exponential random variables having the mean $1/\lambda$



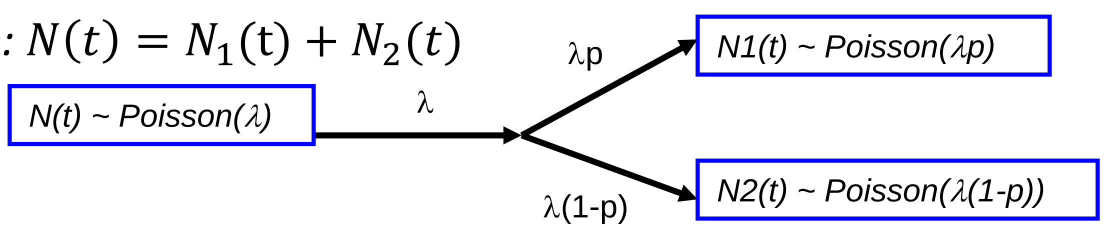
- Proof?

Splitting and Pooling

- Pooling:
  - $N_1(t)$: Poisson process with rate λ_1
  - $N_2(t)$: Poisson process with rate λ_2
  - $N(t) = N_1(t) + N_2(t)$: Poisson process with rate $\lambda_1 + \lambda_2$



- Splitting:
  - $N(t)$: Poisson process with rate λ
  - Each event is classified as Type I, with probability p and Type II, with probability $1 - p$
  - $N_1(t)$: The number of type I events is a Poisson process with rate $p\lambda$
  - $N_2(t)$: The number of type II events is a Poisson process with rate $(1 - p)\lambda$
  - *Note:* $N(t) = N_1(\mathrm{t}) + N_2(t)$



More on Poisson Distribution

- $\{N(t), t \geq 0\}$: a Poisson process with arrival rate λ

- Probability of no arrivals in a small time interval h:

$$\mathbb{P}(N(h) = 0) = e^{-\lambda h} \approx 1 - \lambda h$$

- Probability of one arrivals in a small time interval h:

$$\mathbb{P}(N(h) = 1) = \lambda h \cdot e^{-\lambda h} \approx \lambda h$$

- Probability of two or more arrivals in a small time interval h:

$$\mathbb{P}(N(h) \geq 2) = 1 - \big(\mathbb{P}(N(h) = 0) + \mathbb{P}(N(t) = 1)\big)$$

$$\approx 0$$

Outline

- Probability and random variables
  - Random experiment and random variable
  - Probability mass/density functions
  - Expectation, variance, covariance, correlation
- Probability distributions
  - Discrete probability distributions
  - Continuous probability distributions
  - Empirical probability distribution

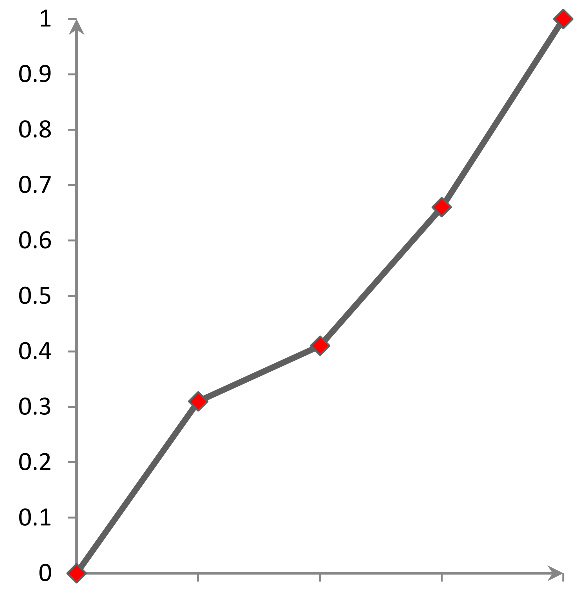
UNIVERSITY OF CALGARY

Empirical Distribution

- A distribution whose parameters are the observed values in a sample of data:
  - Could be used if no theoretical distributions fit the data adequately
  - Advantage: no assumption beyond the observed values in the sample
  - Disadvantage: sample might not cover the entire range of possible values

Empirical Distribution

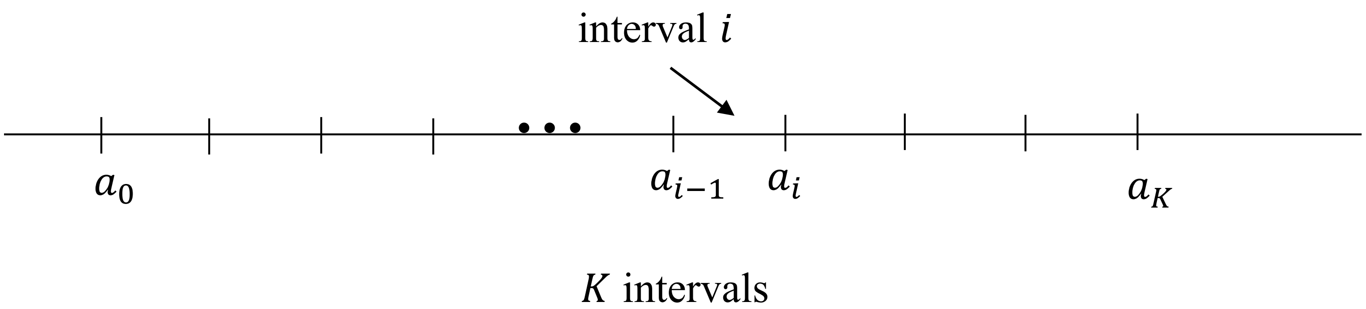
- “Piecewise Linear” empirical distribution
  - Used for continuous data
  - Appropriate when a large sample data is available
  - Empirical CDF is approximated by a piecewise linear function:
    - the ‘jump points’ connected by linear functions



Piecewise Linear Empirical CDF

Empirical Distribution

- Piecewise Linear empirical distribution
  - Organize X-axis into K intervals
  - Interval i is from a_{i-1} to a_i for $i = 1,2, \dots, K$
  - p_i: relative frequency of interval i
  - c_i: relative cumulative frequency of interval i, i.e., $c_i = p_1 + \cdots + p_i$



  - Empirical CDF:
    - If x is in interval i, i.e., $a_{i-1} < x \leq a_i$, then:

$$F(x) = c_{i-1} + \alpha_i (x - a_{i-1})$$

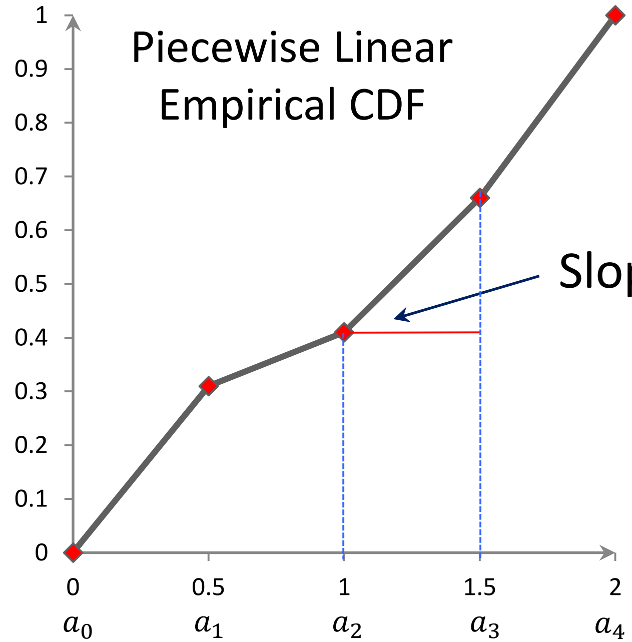
where, slope α_i is given by

$$\alpha_i = \frac{c_i - c_{i-1}}{a_i - a_{i-1}}$$

Example Empirical Distribution

- Suppose the data collected for 100 broken machine repair times are:

i	*Interval (Hours)*	*Frequency*	*Relative Frequency*	*Cumulative Frequency*	*Slope*
1	0.0 < x ≤ 0.5	31	0.31	0.31	0.62
2	0.5 < x ≤ 1.0	10	0.10	0.41	0.2
3	1.0 < x ≤ 1.5	25	0.25	0.66	0.5
4	1.5 < x ≤ 2.0	34	0.34	1.00	0.68



Slope $\alpha_3 = \frac{c_3 - c_2}{a_3 - a_2} = 0.5$