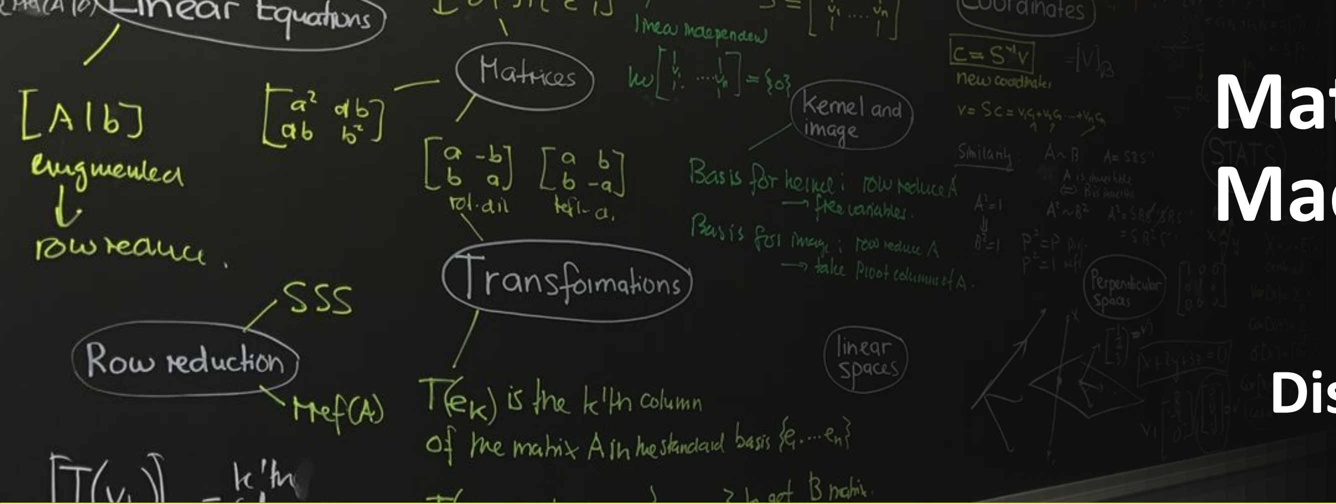


Mathematical Foundatio
Machine Learning and Da

Discrete and Continuous Rando

Dr. Zubair Khalid

Department of Electrical Engin
School of Science and Engineer
Lahore University of Manageme

https://www.zubairkhalid.org/ee

Discrete Random Variable

- **Random Variable (Definition):**

  Random variable is a **function** which **maps** elements from the **sample space** to the **real line**.

- Random Variables are denoted by upper case letters (*X* or *Y*).

- Individual outcomes for RV are denoted by lower case letters (*x* or *y*).

- Mathematically, $X(\omega)$ is a real-valued function defined for $\omega \in \Omega$.

- For each element of an experiment's sample space, the random variable can take on exactly one value.
- **Discrete Random Variable: A RV that can take on only a finite or countably infinite set of outcomes.**

LUMS
A Not-for-Profit University

Example 1

A random variable $X(\omega)$ = number of heads if three coins are tossed at the same time

Sampe space: $\Omega := \{\text{TTT, TTH, THT, HTT, THH, HTH, HHT, HHH}\}$

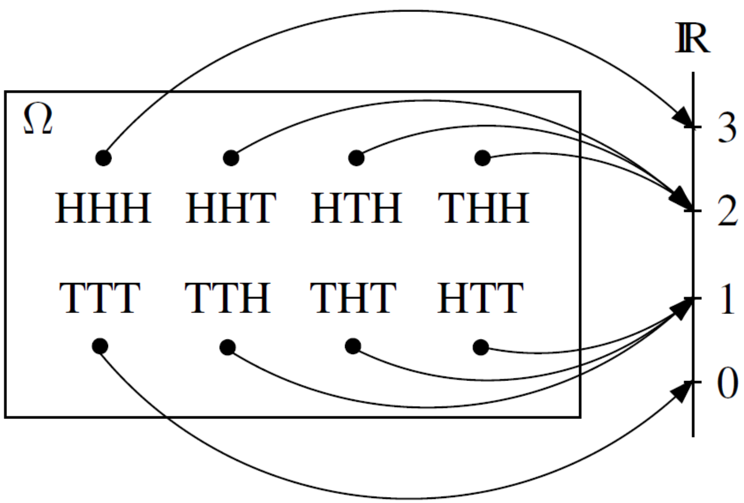


Fig 1: Illustration of RV mapping

$$X(\omega) := \begin{cases} 0, & \omega = \text{TTT}, \\ 1, & \omega \in \{\text{TTH, THT, HTT}\}, \\ 2, & \omega \in \{\text{THH, HTH, HHT}\}, \\ 3, & \omega = \text{HHH}. \end{cases}$$

Example 2

A random variable $X(\omega)$ = number of girls in a family of 4 kids.

Lower case x is a particular value of $X(\omega)$.

ω	**Random Variable *X***
BBBB	$x=0$
GBBB	$x=1$
BGBB	$x=1$
BBGB	$x=1$
BBBG	$x=1$
GGBB	$x=2$
GBGB	$x=2$
GBBG	$x=2$
BGGB	$x=2$
BGBG	$x=2$
BBGG	$x=2$
BGGG	$x=3$
GBGG	$x=3$
GGBG	$x=3$
GGGB	$x=3$
GGGG	$x=4$

Example 3

Random variable, *Y* = Sum of the up faces of the two die.

Die 1

Die 2

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

y
2
3
4
5
6
7
8
9
10
11
12

Probability Mass Function (pmf)

- **Probability Mass Function:** Assigns probabilities (masses) to the individual outcomes. (Also referred as probability density function.)

- For a random variable X, its pmf is given by

$$\boxed{p_X(x_i) \;:=\; \mathsf{P}(X = x_i)}$$

- By axioms of probability;
  - pmf is between 0 and 1 $\quad 0 \;\leq\; p_X(x_i) \;\leq\; 1$
  - sum of all probabilities equal to 1 $\quad \sum_i p_X(x_i) \;=\; 1$

Example 1

A random variable $X(\omega)$ = number of heads if three coins are tossed at the same time

$$p_X(0) = \mathsf{P}(X=0) = \mathsf{P}(\{\text{TTT}\}) = \frac{|\{\text{TTT}\}|}{|\Omega|} = \frac{1}{8}$$

$$p_X(1) = \mathsf{P}(X=1) = \mathsf{P}(\{\text{HTT,THT,TTH}\}) = \frac{3}{8}$$

$$p_X(2) = \mathsf{P}(X=2) = \mathsf{P}(\{\text{HHT,HTH,HHT}\}) = \frac{3}{8}$$

$$p_X(3) = \mathsf{P}(X=3) = \mathsf{P}(\{\text{HHH}\}) = \frac{1}{8}$$

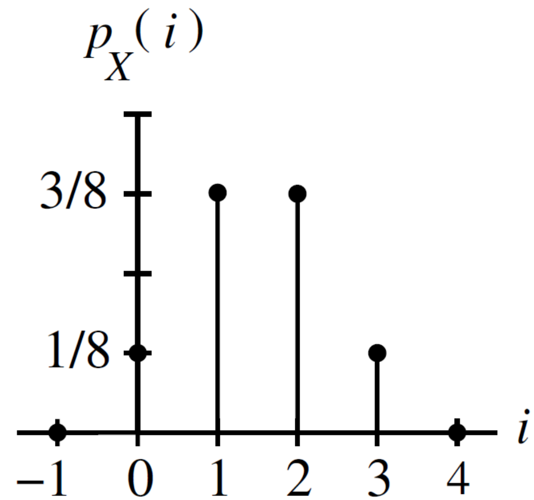


Fig 2: pmf of RV X

Example 2

A random variable X = number of girls in a family of 4 kids

Number of Girls, x	Probability, $p_X(x)$
0	1/16
1	4/16
2	6/16
3	4/16
4	1/16
Total	16/16=1.00

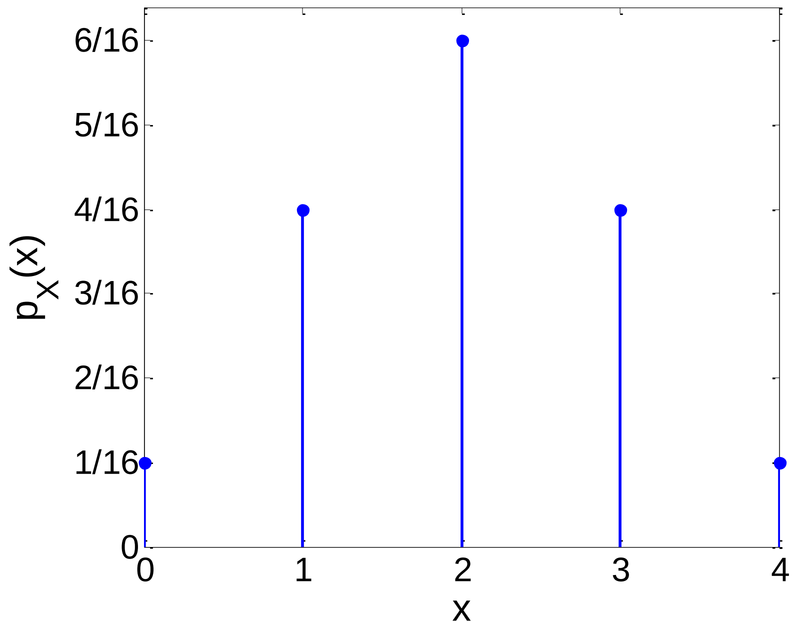


Fig 3: pmf of RV

What is the probability of exactly 3 girls in 4 kids?

What is the probability of at least 3 girls in 4 kids?

Example 3

Random variable, Y = Sum of the up faces of the two die.

y	p_Y (y)
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

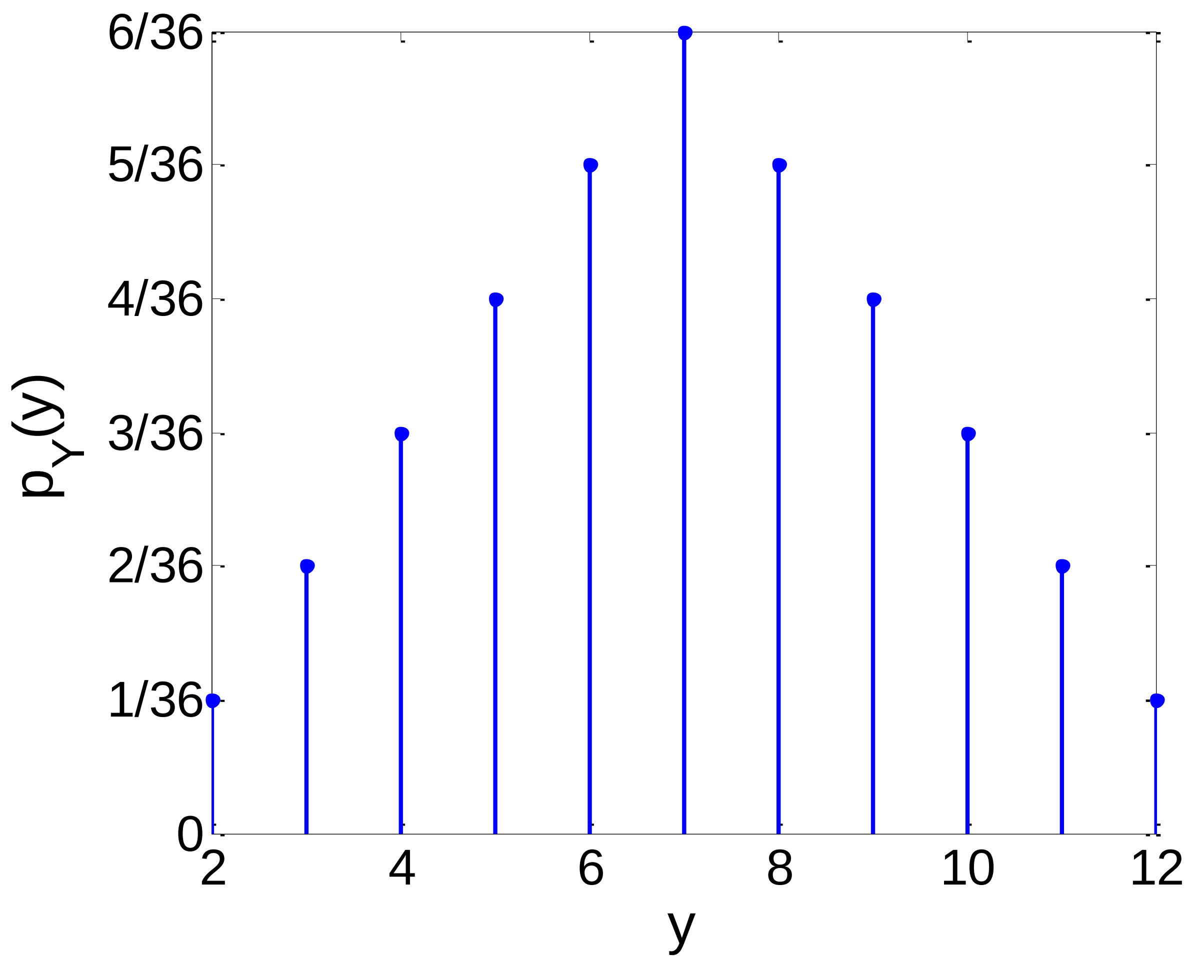


Fig 4: pmf of RV

Important Random Variables

1. **Bernoulli Random Variable**

   If there are only two outcomes of an experiment, the experiment is modeled with uniform random variable. For example, the tossing of coin is modeled with Bernoulli random variable.

   - It is most common to associate {0,1} to the outcomes of an experiment.
   - pmf is given by,

$$p_X(0) = p$$

$$p_X(1) = 1 - p$$

Important Random Variables

2. Uniform Random Variable:

If the outcomes of an experiment are finite, and are equally likely, the experiment is modeled with uniform random variable.

- If there are n outcomes of an experiment, probability of each outcome = $\dfrac{1}{n}$.
- If outcomes are indexed, $k=1, 2, \ldots, n$, $\mathsf{P}(X = k) = \dfrac{1}{n}, \quad k = 1, \ldots, n$
- pmf is given by,

$$p_X(k) = \begin{cases} 1/n, & k = 1, \ldots, n, \\ 0, & \text{otherwise.} \end{cases}$$

Important Random Variables

3. Poisson Variable:

A random variable X is said to have a Poisson probability mass function with parameter $\lambda > 0$, denoted by $X \sim \text{Poisson}(\lambda)$, if

$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots.$$

- Parameter λ fully characterizes the distribution.

- Used in modelling of physical phenomenon arising in different applications:
- arrival of photons at a telescope
- distribution of nodes in wireless sensor networks
- telephone calls arriving in a system
- arrival of network messages in a queue for transmission

Important Random Variables - Examples

***Example* 2.6.** Ten neighbors each have a cordless phone. The number of people using their cordless phones at the same time is totally random. Find the probability that more than half of the phones are in use at the same time.

***Example* 2.7.** The number of hits to a popular website during a 1-minute interval is given by a Poisson(λ) random variable. Find the probability that there is at least one hit between 3:00 am and 3:01 am if $\lambda = 2$. Then find the probability that there are at least 2 hits during this time interval.

***Example* 2.16.** A light sensor uses a photodetector whose output is modeled as a Poisson(λ) random variable X. The sensor triggers an alarm if $X > 15$. If $\lambda = 10$, compute $\mathsf{P}(X > 15)$.

Multiple Random Variable

- When events are defined by more than one random variable.

- Let X represent one variable and Y represent another random variable, which maps elements of sample space to real line, but can be different, then the event involving both X and Y is described as

$$\{X \in B, Y \in C\} := \{\omega \in \Omega : X(\omega) \in B \text{ and } Y(\omega) \in C\}$$

- This is taken as an event that X belongs to B and Y belongs to C.

- **Very important to understand the concept:** the event above is a function of two random variable and is comprised of only those points on the real line which are common between B and C, that is,

$$\{X \in B, Y \in C\} = \{X \in B\} \cap \{Y \in C\}$$

LUMS
A Not-for-Profit University

Multiple Random Variable – Probability mass function

- The joint probability involving two random variables is given by the probability of the joint event

$$\begin{aligned} \mathsf{P}(X \in B, Y \in C) &:= \mathsf{P}(\{X \in B, Y \in C\}) \\ &= \mathsf{P}(\{X \in B\} \cap \{Y \in C\}) \end{aligned}$$

- taking $B = \{x_i\}$ and $C = \{y_j\}$, define joint probability mass function,

$$\boxed{p_{XY}(x_i, y_j) := \mathsf{P}(X = x_i, Y = y_j)}$$

- **Interpretation:** $p_{XY}(x_i, y_j)$ gives the probability that the RV $X = x_i$ and RV $Y = y_i$ at the same time.

- **Marginal probability mass function:** We can obtain $p_X(x_i)$ and $p_Y(y_j)$

$$\boxed{p_X(x_i) = \sum_j p_{XY}(x_i, y_j)} \qquad \boxed{p_Y(y_j) = \sum_i p_{XY}(x_i, y_j)}$$

Multiple Random Variable – Concept of Independence

- When RVs X and Y are independent events, we can write the joint probability as

$$\mathsf{P}(X \in B, Y \in C) = \mathsf{P}(X \in B)\mathsf{P}(Y \in C)$$

$$\mathsf{P}(X = x_i, Y = y_j) = \mathsf{P}(X = x_i)\,\mathsf{P}(Y = y_j)$$

- Equivalently, we can write in terms of joint pmf and individual pms of RVs:

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

- The concepts presented for two random variables are also valid for more than two random variables.

Multiple Random Variable – Examples

***Example* 2.8.** On a certain aircraft, the main control circuit on an autopilot fails with probability p. A redundant backup circuit fails independently with probability q. The aircraft can fly if at least one of the circuits is functioning. Find the probability that the aircraft cannot fly.

***Example* 2.9.** Let X, Y, and Z be the number of hits at a website on three consecutive days. Assuming they are i.i.d. Poisson(λ) random variables, find the probability that on each day the number of hits is at most n.

Expectation of a Random Variable

- Expectation of a random variable is defined as;

$$\mathsf{E}[X] := \sum_i x_i \, \mathsf{P}(X = x_i)$$

$$\boxed{\mathsf{E}[X] = \sum_i x_i \, p_X(x_i)}$$

- Expectation of a random variable gives an average value of the values x_1, x_2, $\ldots$, a random variable can take with probabilities $\mathsf{P}(X = x_1)$, $\mathsf{P}(X = x_2)$, $\ldots$.

- Expectation is a linear operator: $\mathsf{E}[aX + bY] = \mathsf{E}[aX] + \mathsf{E}[bY] = a\mathsf{E}[X] + b\mathsf{E}[Y]$

- **law of the unconscious statistician (LOTUS):** If another RV Y is a function of RV X given by, $Y = f(X)$, the expected value of RV Y is given in terms of pmf of the RV X as

$$\boxed{\mathsf{E}[Y] = \mathsf{E}[f(X)] = \sum_i f(x_i) p_X(x_i)}$$

Expected Values of Discrete RV's

- Mean : Long-run average value a RV.

- Variance – Average squared deviation between a realization of a RV and its mean. Quantifies the spread around

- Standard Deviation – Positive square root of variance , measure of spread.

- Notation:
  - Mean: $\mathsf{E}[X] = m$
  - Variance: $\mathrm{var}(X) = \mathsf{E}[(X - m)^2] = \sigma^2$
  - Standard Deviation: σ

Moments of Random Variable

Moments:

- $n-$th momemt of a RV X is defined as $\mathsf{E}[X^n]$.
- Mean, $x = \mathsf{E}[X]$ is the first moment.

Central Moments - Moments around center (mean):

- $n-$th central momemt of a RV X is defined as $\mathsf{E}[(X-m)^n]$.
- Variance, $\operatorname{var}(X) = \mathsf{E}[(X-m)^2] = \sigma^2$ is the second central moment.
- Skewness: $\mathsf{E}[(X-m)^3]/\sigma^3$
- Kurtosis: $\mathsf{E}[(X-m)^4]/\sigma^4$

Variance of Random Variable

- Variance, $\mathrm{var}(X) = \mathsf{E}[(X-m)^2] = \sigma^2$, is often computed as

$$\mathrm{var}(X) = \mathsf{E}[X^2] - (\mathsf{E}[X])^2$$

Derivation:

$$\begin{aligned}
\mathrm{var}(X) &:= \mathsf{E}[(X-m)^2] \\
&= \mathsf{E}[X^2 - 2mX + m^2] \\
&= \mathsf{E}[X^2] - 2m\mathsf{E}[X] + m^2, \\
&= \mathsf{E}[X^2] - m^2 \\
&= \mathsf{E}[X^2] - (\mathsf{E}[X])^2.
\end{aligned}$$

LUMS
A Not-for-Profit University

Continuous Random Variable

Continuous Random Variable

- A continuous random variable is one for which the outcome can be any value in an interval of the real number line.

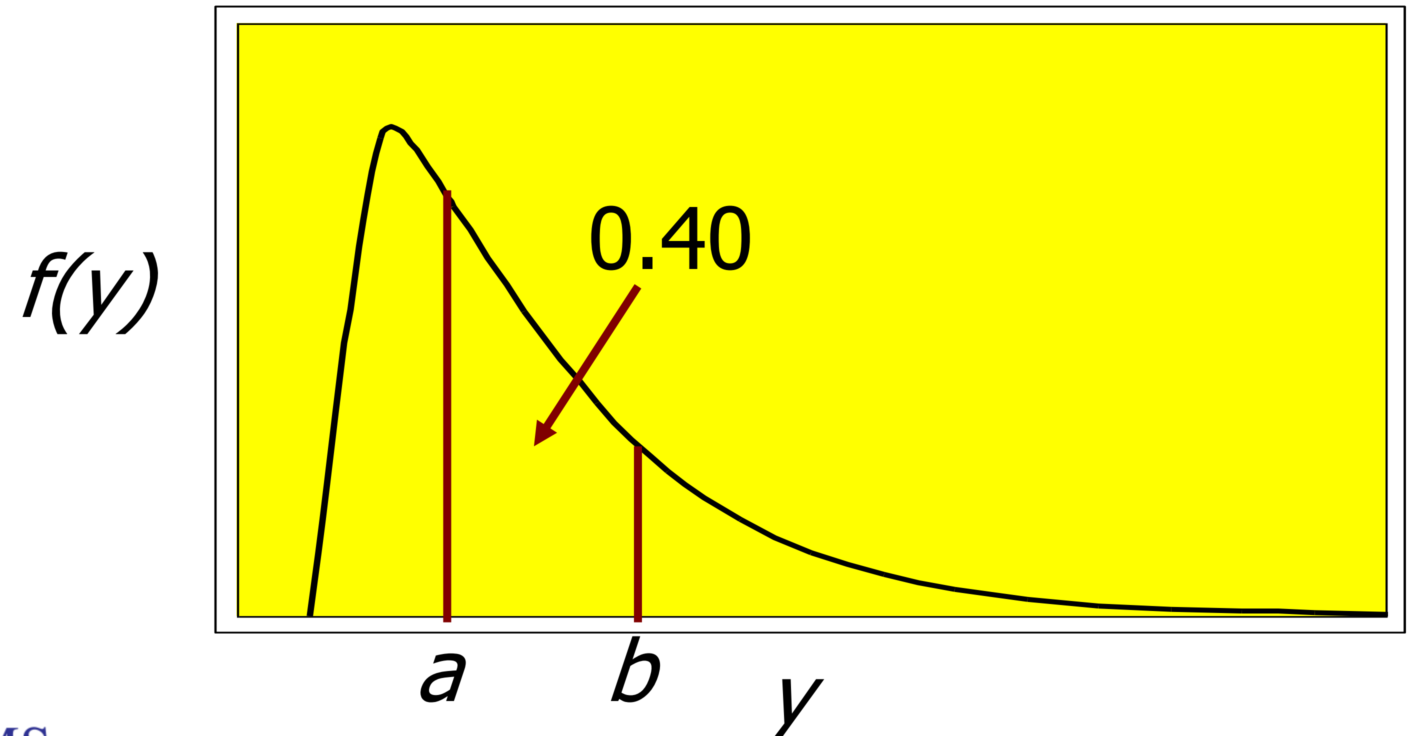
- There are always infinitely many sample points in the sample space.

- For **discrete** random variables, only the value listed in the **pmf** have positive probabilities, all other values have probability zero.

- For continuous random variables, the probability of every specific value is zero. Probability only exists for an interval of values for continuous RV., that is, for continuous RV Y,
  - We don't calculate $P(Y = y)$, we calculate $P(a < Y < b)$, where a and b are real numbers.
  - For a continuous random variable $P(Y = y) = 0$.

Probability density function

- The **probability density function (pdf)** denotes a curve against the possible values of random variable and the area under an interval of the curve is equal to the probability that random variable is in that interval.
- For example if $f(y)$ denotes the pdf of RV Y, we calculate $P(a < Y < b)$,



pmf versus pdf

- For a discrete random variable, we have probability mass function (pmf).
- The pmf looks like a bunch of spikes, and probabilities are represented by the heights of the spikes.
- For a continuous random variable, we have a probability density function (pdf).
- The pdf looks like a curve, and probabilities are represented by areas under the curve.

Characteristics of pdf

- Given Y is a continuous random variable with pdf is $f(x)$.
- By axioms of probability, $f(x)$ must satisfy the following conditions:

1. $f(x) \geq 0 \text{ for all } x \in R$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$

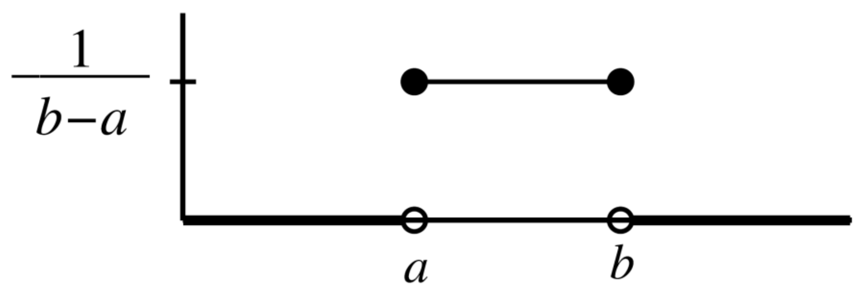
Important Continuous RVs

- **Uniform random variable:** used to model the experiments in which outcome is constrained to lie in a known interval, say [*a*,*b*] and all possible outcomes are equally likely.

- Define uniform random variable *f* ~ uniform[a,b] for *a*<*b* with pdf

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b, \\ 0, & \text{otherwise.} \end{cases}$$

- Plot of pdf

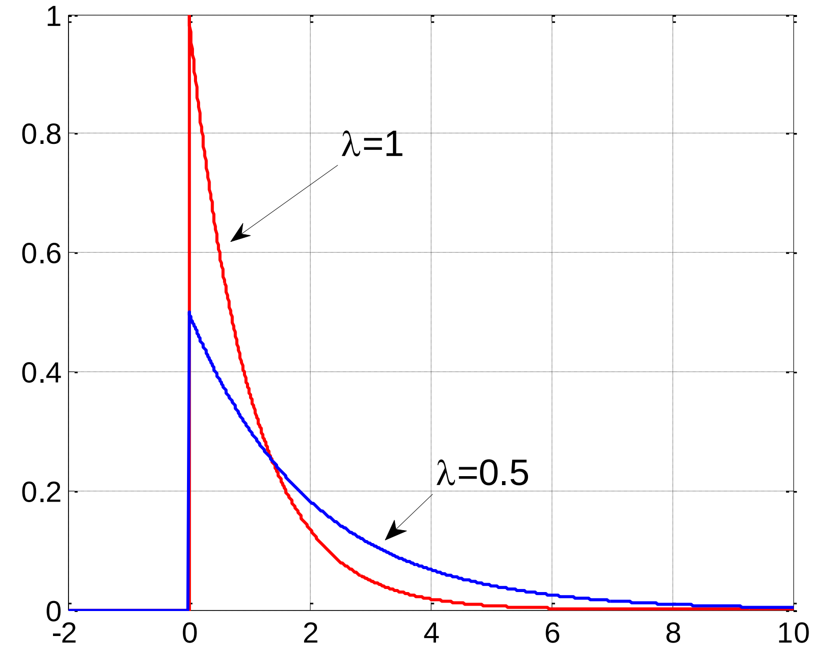


Important Continuous RVs

- **Exponential random variable:** used to model lifetimes, such as
  - how long it takes before next phone call arrivrs
  - how long it takes a computer network to transmit a message
  - how long it takes a radioactive particle to decay

- Define $f \sim \exp(\lambda)$ for $\lambda > 0$ with pdf given by:

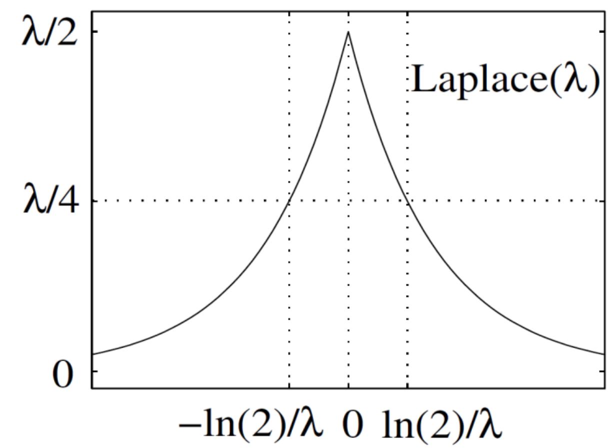
$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$



Important Continuous RVs

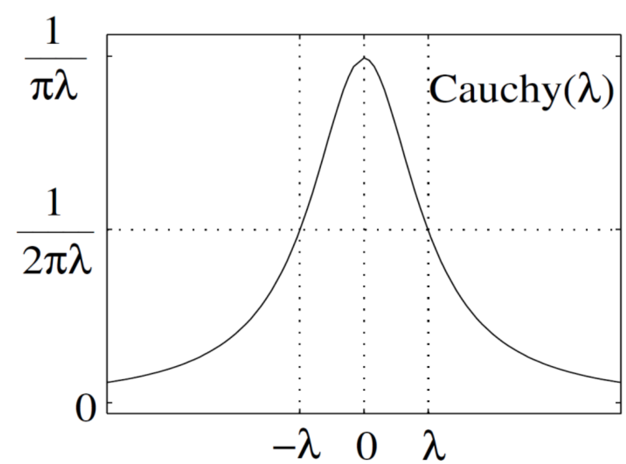
- **Laplace (double sided exponential) random variable :**
- Denoted by $f \sim$ Laplace(λ) for $\lambda > 0$:

$$f(x) = \frac{\lambda}{2} e^{-\lambda |x|}$$



- **Cauchy random variable:**
- Denoted by $f \sim$ Cauchy(λ) for $\lambda > 0$:

$$f(x) = \frac{\lambda / \pi}{\lambda^2 + x^2}$$

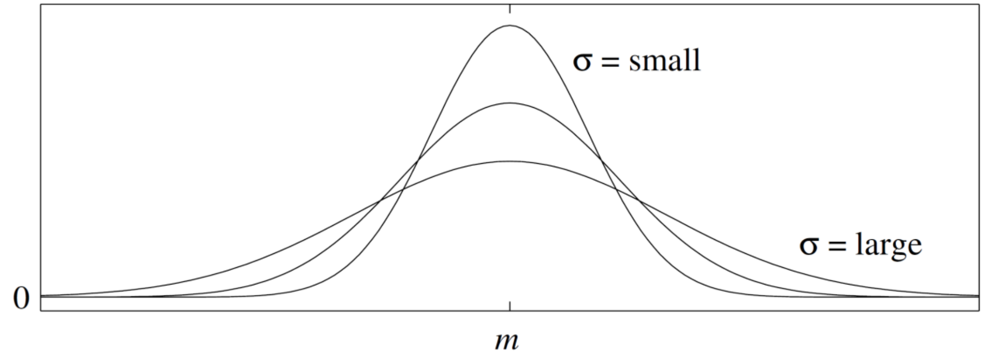


Important Continuous RVs

- **Gaussian (Normal) random variable:**
- Define Gaussian RV $f \sim N(m, \sigma^2)$

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

- center $m \in R$
- standard deviation, $\sigma^2 \in R^+$, quantifies the spread of the pdf
- $N(0, 1)$ is called standard normal density



Some Examples

***Example* 4.1.** In coherent radio communications, the phase difference between the transmitter and the receiver, denoted by Θ, is modeled as having a density $f \sim \text{uniform}[-\pi, \pi]$. Find $\mathsf{P}(\Theta \leq 0)$ and $\mathsf{P}(\Theta \leq \pi/2)$.

Some Examples

***Example* 4.4.** An Internet router can send packets via route 1 or route 2. The packet delays on each route are independent $\exp(\lambda)$ random variables, and so the difference in delay between route 1 and route 2, denoted by X, has a Laplace(λ) density. Find

$$\mathsf{P}(-3 \leq X \leq -2 \text{ or } 0 \leq X \leq 3).$$

Some Examples

Problem. For Gaussian RV $f \sim N(m, \sigma^2)$, show that $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

You may use the following information to derive the result :

$$\int_{-\infty}^{\infty} e^{-y^2}\,dy = \sqrt{\pi}$$

Expectation of Random Variable

- Law of the unconscious statistician (LOTUS) version for continuous random variable X :

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

- Recall, mean or average, $m = \mathsf{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx$

- $\mathsf{var}(X) = \mathsf{E}[X^2] - (\mathsf{E}[X])^2$

Expectation of Random Variable

***Example* 4.7.** If X is a uniform$[a,b]$ random variable, find $\mathsf{E}[X]$, $\mathsf{E}[X^2]$, and $\mathsf{var}(X)$.

Expectation of Random Variable

***Example* 4.9.** If X is an exponential random variable with parameter $\lambda = 1$, find all moments of X.

Expectation of Random Variable

***Example* 4.7.** If X is a uniform$[a,b]$ random variable, find $\mathsf{E}[X]$, $\mathsf{E}[X^2]$, and $\mathsf{var}(X)$.

***Example* 4.9.** If X is an exponential random variable with parameter $\lambda = 1$, find all moments of X.