21DS636-Statistical Modelling

Probability Theory

Random Variable & Probability Distribution

Dr. Deepa Gupta, Professor
Human Language Technology Lab
Department of Computer Science & Engineering
ASE-Bangalore

Recap: Summary of the Rules of Probability

The additive rule

$$\boxed{P[A \cup B] = P[A] + P[B] - P[A \cap B]}$$

and

$$P[A \cup B] = P[A] + P[B] \quad \text{if } A \cap B = \phi$$

The Rule for complements

for any event E

$$\boxed{P[\bar{E}] = 1 - P[E]}$$

Conditional probability

$$P[A|B] = \frac{P[A \cap B]}{P[B]}$$

The multiplicative rule of probability

$$P[A \cap B] = \begin{cases} P[A]P[B|A] & \text{if } P[A] \neq 0 \\ P[B]P[A|B] & \text{if } P[B] \neq 0 \end{cases}$$

Independent Events

and

$$P[A \cap B] = P[A]P[B]$$

if A and B are independent.

$P(A \cap B | C)) = P(A|c) \cdot P(B|C)$

Bayes' Theorem

- If $A_1, A_2, \ldots, \ldots\ldots A_n$ is a partition of a sample space, then the **posterior probabilities** of the event A_i conditional on an event ***B*** can be obtained from the probabilities $P(A_i)$ and $P(B|A_i)$ using the formula

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)}$$

Rule 1

$$n(A_1 \cup A_2 \cup A_3 \cup \ldots.) = n(A_1) + n(A_2) + n(A_3) + \ldots$$

if the sets $A_1, A_2, A_3, \ldots$ are pairwise mutually exclusive (i.e. $A_i \cap A_j = \phi$)

Basic counting formulae

a, b, c

Counting problems are not easy. The more practice better the techniques

3
$= 3 \cdot 2 \cdot 1 = 6$

1. Orderings

$n! =$ the number of ways you can order n objects

i.e abc ≠ cba

2. Permutations

$$_{\cancel{n}}^{n}P_k = \frac{n!}{(n-k)!} =$$

The number of ways that you can choose k objects from n in a specific order

$^{n}P_{k}$

3. Combinations

abc = bca = acb =

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

The number of ways that you can choose k objects from n (order of selection irrelevant)

Random Variables

Numerical Quantities whose values are determine by the outcome of a random experiment

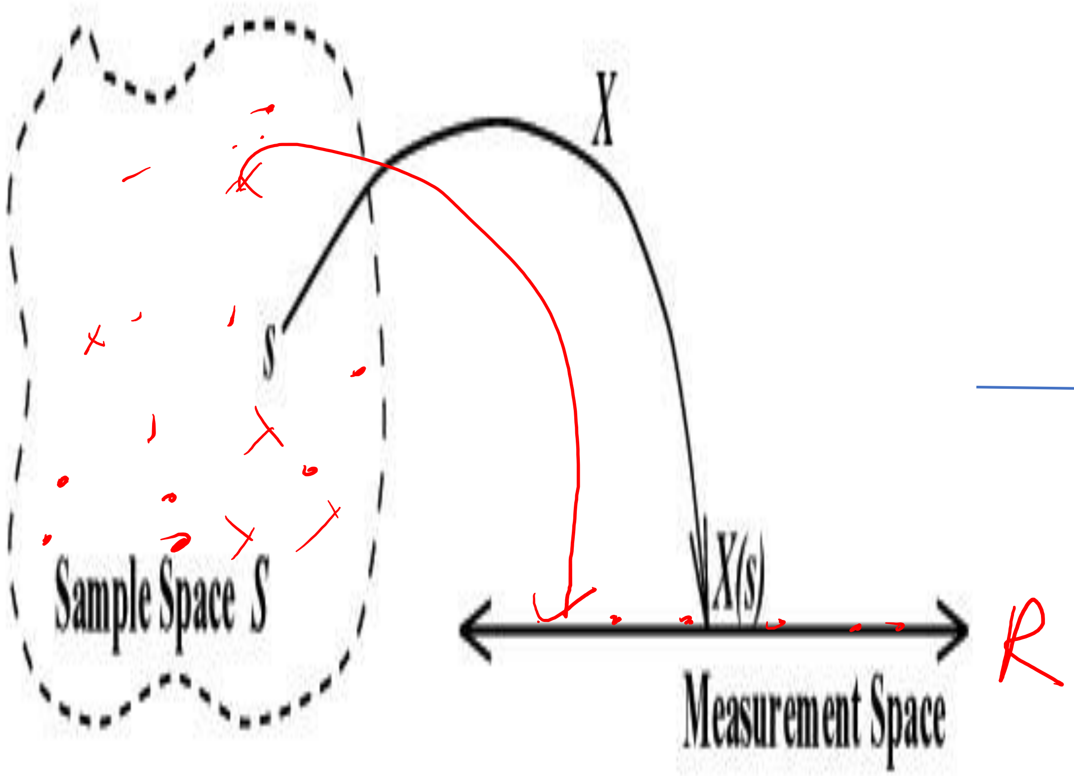
Definition of a Random Variable

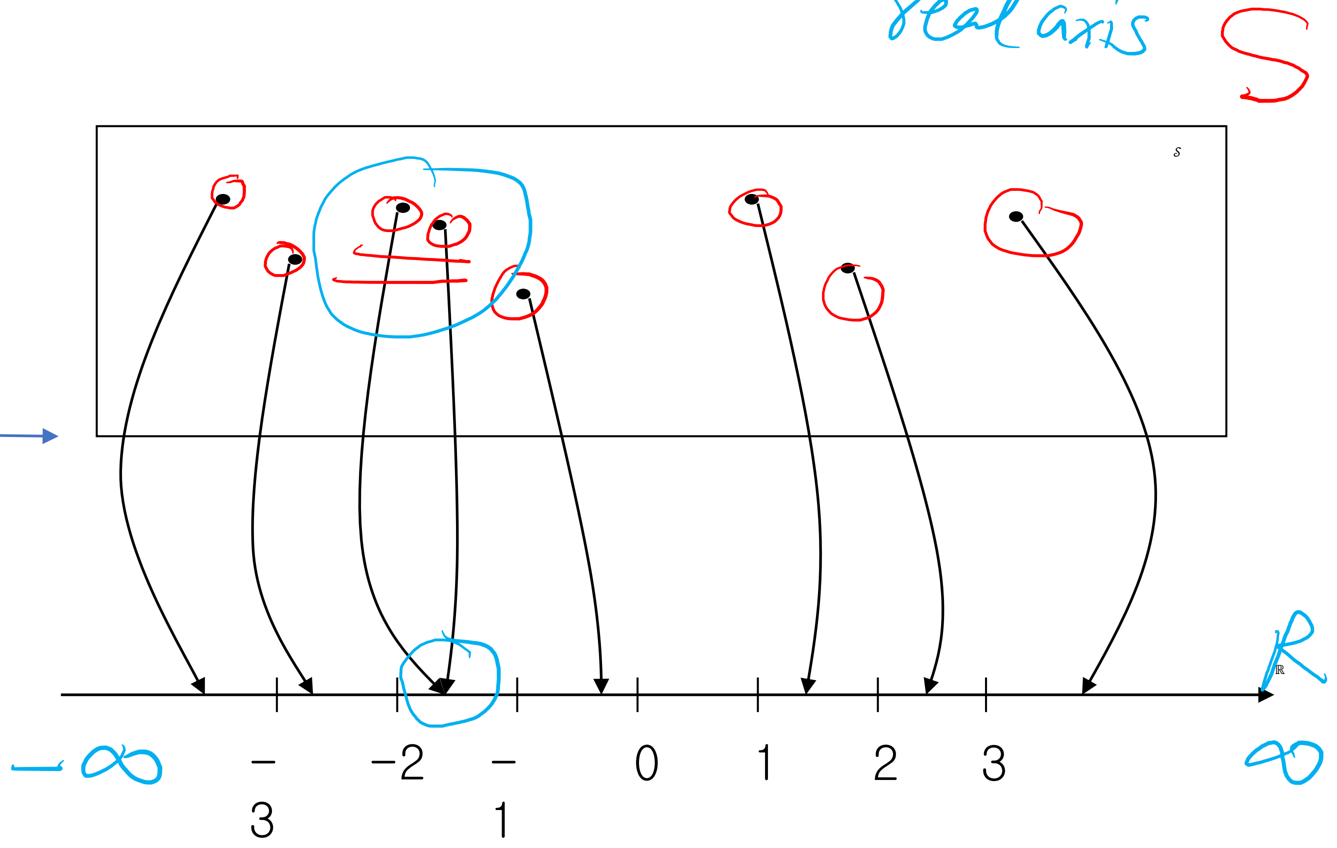
$X : S \to R$

sample space

real axis

- **Random Variable as a Measurement**
  - A numerical value to each outcome of a particular experiment





Formal definition of Random Variable

$P(x, y)$ $\quad \frac{P(g(x))}{P(g(x), g(y))}$

- A numerical value can be associated with each outcome of an experiment

Y W R U V $\quad f(x) = x^2$

- A *random variable* X is a function from the sample space S to the real line that assigns a real number $X(s)$ to each element *s* of *S*

$$\boxed{X: S \rightarrow R}$$

$X(s) = \alpha, \quad \alpha \in R$

$$\boxed{\begin{aligned} f &: X \rightarrow Y \\ f(x) &= y \end{aligned}}$$

Why Random Variable?

- Venn diagram/tree diagram is having limited scope
- Dealing with random variable is easier than working with sample space **(why?)**
  - Mapping a set theoretic operations (sample space) to real number

    → real valued fn
  - Working with point function is easier than set functions
  - One can manipulate these functions like any other real valued function

    - integration, differentiation, inequalities etc.

$f(x)$

Types of Random Variable

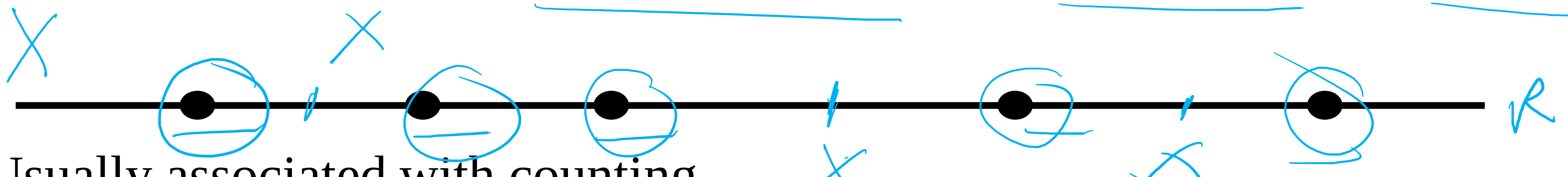
- **Discrete R.V.**
  - Takes on one of a finite (or at least countable) number of different values
  - Integer valued

- **Continuous R.V.**
  - The set of possible values for X are all real numbers i.e. infinite uncountable
  - Continuous random variables can take any value in an interval.
  - They are used to model physical characteristics such as time, length, position, etc.

Discrete Random Variables

<u>Discrete Random Variable</u>: A random variable usually assuming an integer value.

- a discrete random variable assumes values that are isolated points along the real line. That is neighboring values are not "possible values" for a discrete random variable



Note: Usually associated with counting

- The number of times a head occurs in 10 tosses of a coin
- The number of auto accidents occurring on a weekend
- The size of a family

Examples: Discrete Random Variable

$S=\{1,2,3,4,5,6\}$

- A die is rolled and X = number of spots showing on the upper face.

  $X: 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6$

- Two dice are rolled and X = Total number of spots showing on the two upper faces.

  $X: 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \quad 10 \quad 11 \quad 12$

- A coin is tossed n = 100 times and X = number of times the coin toss resulted in a *head*.

  $X: 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad - \quad — \quad 100$

- We observe X, the number of hurricanes in the Carribean from April 1 to September 30 for a given year

  $X: 0 \quad 1 \quad 2 \quad 3 \quad - \quad - \quad - \quad -$

- Family Size, X= no. of dependents

  $X: 0 \quad 1 \quad 2 \quad 3 \quad — \quad — \quad —$

- X= No. of students present in a class from July 29th 2020 to August 30th 2020 for the given Odd semester July 2020 to November 2020

  CSE(AI) = 75

  $X: 0 \quad 1 \quad 2 \quad 3 \quad — \quad — \quad 75$

Examples: Continuous Random Variable

C.R.V

- A student is selected at random from a class and *X* = weight of that individual.
- Let X be the length of a randomly selected telephone call.
- X= height of students in class
- X=time it takes to get to school from home
- X=distance traveled between home to office
- V = hours until light bulb fails

$X: S \rightarrow R \rightarrow X: 1, 2, 3 - - - 18$

The Probability distribution of Random Variable

$\sum P_1 \; P_2 \; P_3 \; , \; P_{18} = 1$

- A Mathematical description of the possible values of the random variable together with the probabilities of those values
  - The probability distribution is a complete probabilistic description of a random variable.
- It is represented broadly in two ways:
  - Probability distribution or **probability mass function(pmf, Discrete R.V.)** or **probability density function(pdf, continuous R.V.)**

D.r.V → pmf → Pdf

  - Cumulative distribution function (CDF) or distribution function

Probability Mass Function(pmf)

- The probability distribution of a discrete random variable is described by its : ***probability mass function*** $p(x)$ i.e. pmf p(x)

r.v   X: 1 2 3 4 5 6

- $p(X=x)$ = the probability that X takes on the value x.

X   P(X=3) = 1/6

- This can be given in either a **tabular form** or in the form of **an equation.**

- It can also be displayed in a **graph**.

Example 1

- A die is rolled and *X* = number of spots showing on the upper face.

$P(X=x) \longrightarrow$ PMF

Tabular

~~x~~	1	2	3	4	5	6
$p(x)$	1/6	1/6	1/6	1/6	1/6	1/6

$P(X=x)$

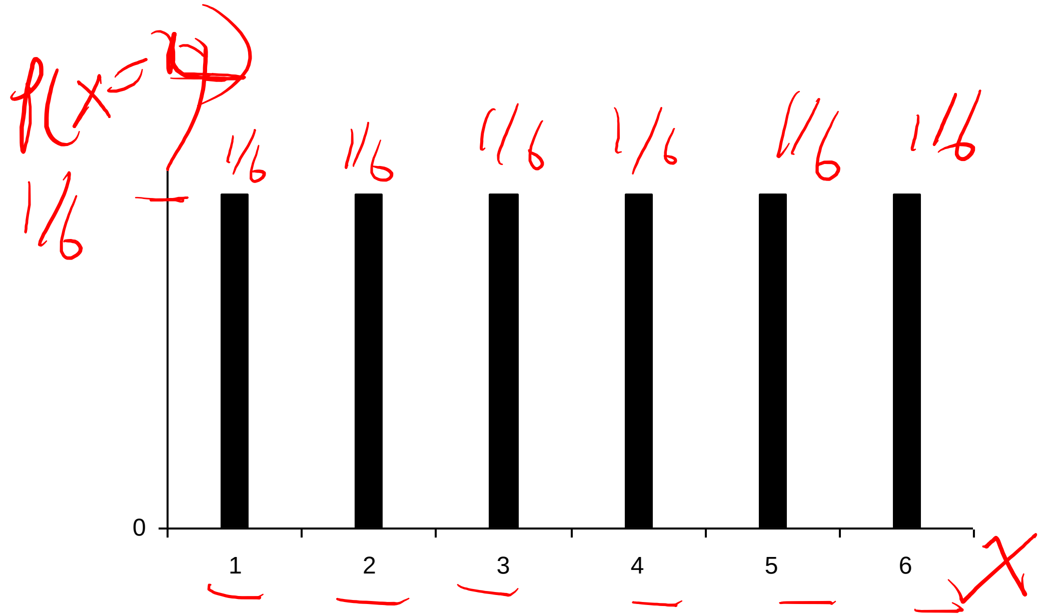
formula

– $p(x) = 1/6$ if $x = 1, 2, 3, 4, 5, 6$

Graphs

To plot a graph of *p(x)*, draw bars of height *p(x)* above each value of *x*.

Rolling a die

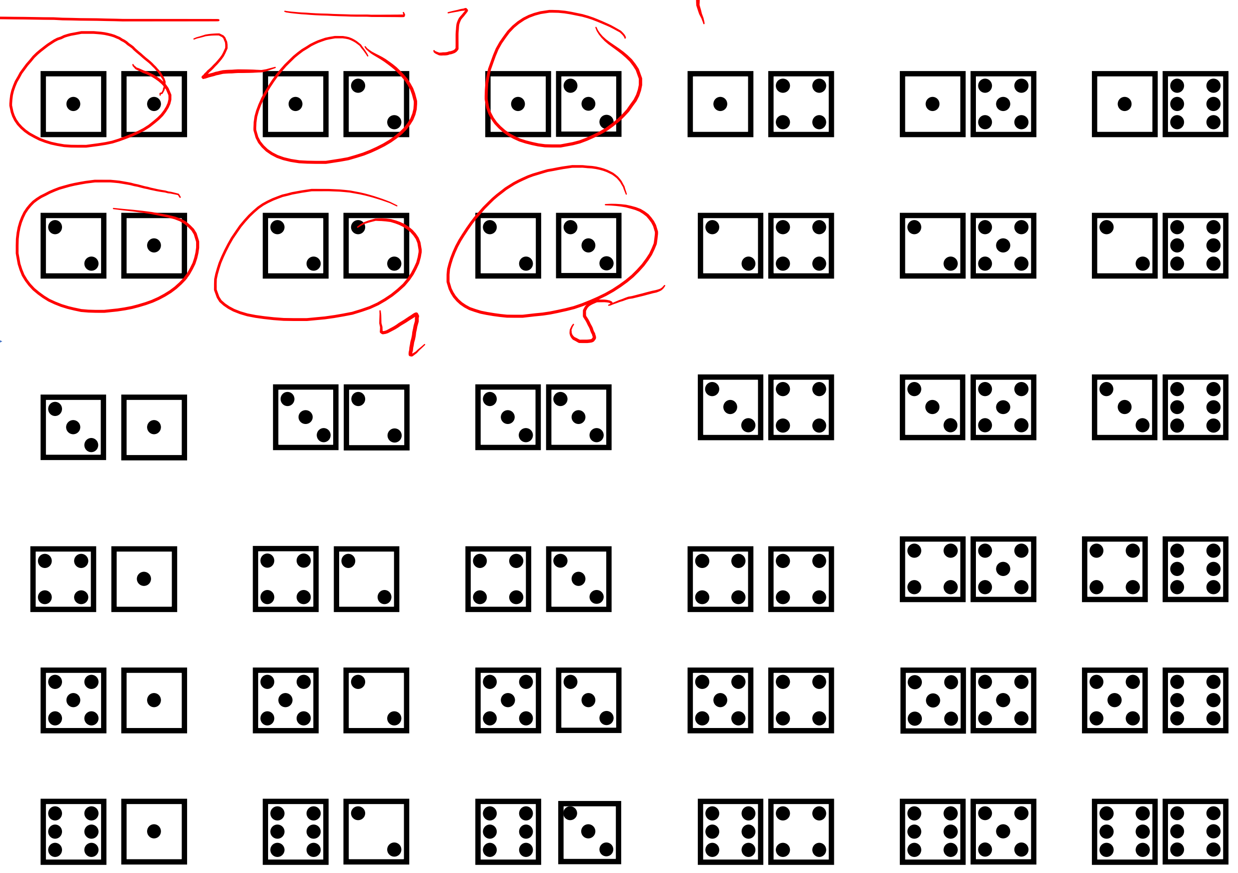


Example 2

X : 2 3 4 5 6 7 8 9 10 11 12

Two dice are rolled and *X* = Total number of spots showing on the two upper faces.

36 possible outcome for rolling two dice



Example 2

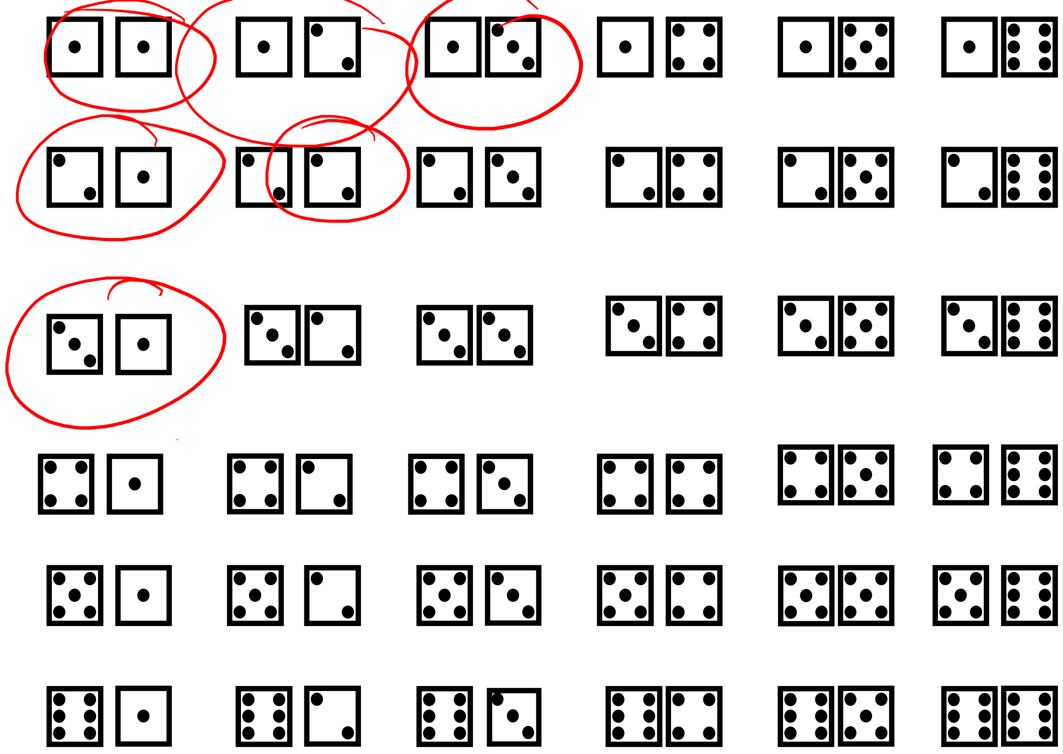
- Two dice are rolled and X = Total number of spots showing on the two upper faces.

$\frac{1}{36} + \frac{1}{36} = \frac{2}{36}$

$P(X=3) = P(\ \ \text{OR}\ \ )$

$\cancel{X}$	2	3	4	5	6	7	8	9	10	11	12
$p(x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

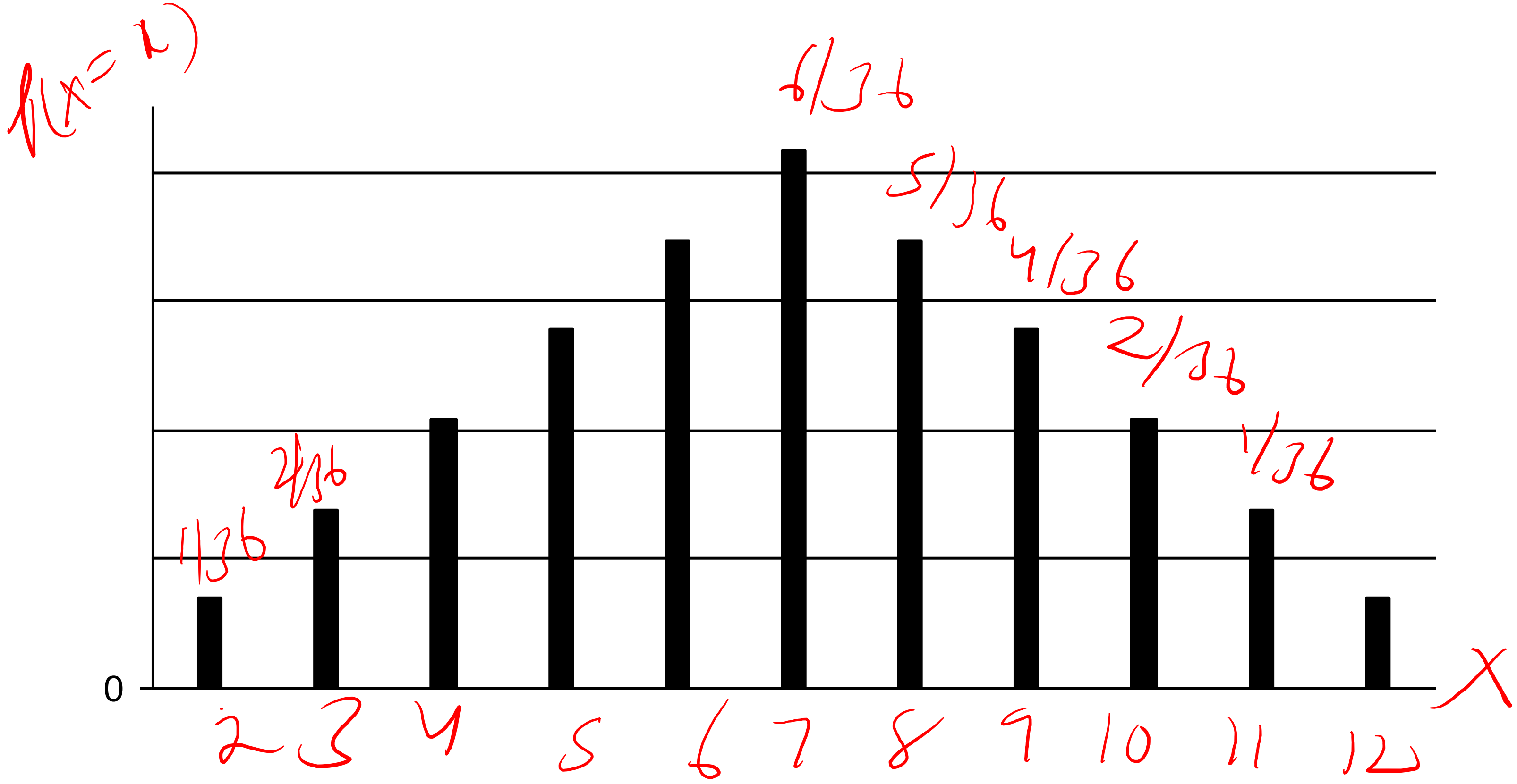
$P(X=2) = P(\ \ ) = \frac{1}{36}$



Formula: $p(x) = \begin{cases} \dfrac{x-1}{36} & x = 2,3,4,5,6 \\ \dfrac{13-x}{36} & x = 7,8,9,10,11,12 \end{cases}$

Example 2 :Graphs

- Two dice are rolled and X = Total number of spots showing on the two upper faces.



Pmf Properties

(d.r.v)

Every probability function must satisfy:

1. The probability assigned to each value of the random variable must be between 0 and 1, inclusive:

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

A event

$0 \leq P(A) \leq 1$

2. The sum of the probabilities assigned to all the values of the random variable must equal 1:

$$\sum_x p(x) = 1$$

$A_1\ A_2 - - A_n \subseteq S$

$A_1, 2 - - A_n$ ME

$P(A_1 \cup A_2 \cup A_3 \cup A_n) = P(S)$

$\Rightarrow \sum_{i=1}^{n} P(A_i) = 1$

$P(X \geq c)$

3. $$P[a \leq X \leq b] = \sum_{x=a}^{b} p(x)$$

$P(a \geq X)$

$P(X < b$

$$= p(a) + p(a+1) + \cdots + p(b)$$

Practice Problem

Which of the following are probability functions?

a) $f(x)=.25$ for $x=9,10,11,12$

b) $f(x)=(3-x)/2$ for $x=1,2,3,4$

c) $f(x)=(x^2+x+1)/25$ for $x=0,1,2,3$

Answer (a)

$P(X=x)$

a. $f(x)$=.25 for x=9,10,11,12

x	*f(x)*
9	.25
10	.25
11	.25
12	.25

1.0

X

x	9	10	11	12
$P(x)$	0.25	0.25	0.25	0.25
	≥ 0	>0	>0	>0

$= 1$

Yes, probability function!

Answer (b)

b. $f(x) = (3-x)/2$ for $x=1,2,3,4$

x	$f(x)$
1	(3-1)/2=1.0
2	(3-2)/2=.5
3	(3-3)/2=0
4	(3-4)/2=-.5

Though this sums to 1, you can't have a negative probability; therefore, it's not a probability function.

Answer (c)

c. $f(x) = (x^2+x+1)/25$ for x=0,1,2,3

x	f(x)
0	1/25
1	3/25
2	7/25
3	13/25

24/25

Doesn't sum to 1. Thus, it's not a probability function.

Example 4

R W

- A bag contains 3 Red and 4 white balls. Find the probability distribution of number of red balls in 3 draws with replacement from the bag.

Sol

X : no. of red balls in 3 draws with replacement from a bag.

X : 0 1 2 3

X	0	1	2	3
$P(X=x)$	$\left(\frac{4}{7}\right)^3$	$3\cdot\frac{3}{7}\cdot\left(\frac{4}{7}\right)^2$	$3\left(\frac{3}{7}\right)^2\frac{4}{7}$	$\left(\frac{3}{7}\right)^3$

$P(X=x) = P(X=0)$

$= P(\text{no red ball})$

$= P(WWW)$

$= P(W)\cdot P(W)P(W)$

$= 4/7 \cdot 4/7 \cdot 4/7$

$P(X=3) = P(RRR)$

$= 3/7 \cdot 3/7 \cdot 3/7$

$P(X=1)$

$= P(RWW \text{ or } WRW \text{ or } WWR)$

$= 3/7 \cdot 4/7 \cdot 4/7 + 4/7 \cdot 3/7 \cdot 4/7 + 4/7 \cdot 4/7 \cdot 3/7$

$= 3 \cdot \frac{3}{7}\left(\frac{4}{7}\right)^2$

Example 5

- A mail order company has six telephone lines. Let X denote number of lines in use at a specified time. Suppose that Pmf of X is given below:

X	0	1	2	3	4	5	6
P(X=x)	0.10	0.15	0.20	0.25	0.20	0.06	0.04

- Calculate the probability of each of the following events:
  a) At most 3 lines are in use
  b) At least 3 lines are in use
  c) Less than 3 lines are in use
  d) Between 2 and 5 lines, inclusive, are in use
  e) At least 4 lines are not in use
  f) Less than three lines are in use given that at most 3 lines are in use
  g) Between 2 and 4 lines, inclusive, are not in use.

AMRITA VISHWA VIDYAPEETHAM

Example 5 : Solution

X	0	1	2	3	4	5	6
P(X=x)	0.10	0.15	0.20	0.25	0.20	0.06	0.04

a) At most 3 lines are in use
b) At least 3 lines are in use
c) Less than 3 lines are in use
d) Between 2 and 5 lines, inclusive, are in use

$P(X \le 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$

$P(X < 3) = 0.10 + 0.15 + 0.20 + 0.25$

b) $P(X \ge 3) = 0.78$

$= P(X=3) + P(X=4) + P(X=5) + P(X=6)$

$= 0.25 + 0.20 + 0.06 + 0.04$

$= \boxed{}$

$1 - P(X<3) = 1 - [P(X=0) + P(X=1) + P(X=2)]$

$P(2 \le X \le 5) = P(X=2) + P(X=3) + P(X=4) + P(X=5)$

Example 5 : Solution

X	0	1	2	3	4	5	6
P(X=x)	0.10	0.15	0.20	0.25	0.20	0.06	0.04

e) At least 4 lines are not in use
f) Less than three lines are in use given that at most 3 lines are in use
g) Between 2 and 4 lines, inclusive, are not in use.

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

(e) 4 → 2 lines are in use; 5 → 1 line in use; 6 → 0 line are in use

P(At least 4 lines are not in use) $= P(X \le 2)$

$= P(X=2) + P(X=1) + P(X=0)$

$= 0.20 + 0.15 + 0.10$

$= 0.58$

(f) $P(X<3 \mid X \le 3) = \frac{P(X<3 \cap X \le 3)}{P(X \le 3)} = \frac{P(X<3)}{P(X \le 3)}$

(g) $1 - ($ $P(X=4) + P(X=3) + P(X=2)$

Example 6:

$P(|X-1| \le 2) = P(-2 \le (X-1) \le +2)$

$= P(-2+1 \le X-1+1 \le +2+1)$

$= P(-1 \le X \le +3)$

$P(X=-1) + P(X=0) + P(X=1) + P(X=2)$

$|X| < 1$

$\Rightarrow -1 < X < 1$

- Pmf of X is as follows:

X	-2	-1	0	1	2
P(X=x)	1/8	2/8	a/8	a/8	1/8

It give pmf is a valid pmf then

- Calculate the probability of each of the following events:
  - i. The value of a → W.K.T $\sum_x P(x) = 1$
  - ii. P(X≤2)
  - iii. P(-1 ≤ X ≤ 1)
  - iv. P(X ≤-1 OR X=2) $= P(X=-2) + P(X=-1) + P(X=2)$
  - v. P(X≥ 1|0 ≤X ≤2)
  - vi. P(|X-1| ≤2)
  - vii. P(X=0)

$\Rightarrow \frac{1}{8} + \frac{2}{8} + \frac{a}{8} + \frac{a}{8} + \frac{1}{8} = 1$

$\Rightarrow \frac{2a+4}{8} = 1$

$\Rightarrow 2a+4 = 8$

$\Rightarrow 2a = 8-4$

$\Rightarrow a = 4/2 = 2$

$= \frac{P(X \ge 1 \cap 0 \le X \le 2)}{P(0 \le X \le 2)}$

$\frac{P(1 \le X \le 2)}{P(0 \le X \le 2)}$

$= a/8 = 2/8$

Continuous Random Variables

<u>Continuous Random Variable</u>: The random variable is continuous if the range of X is uncountably infinite

- A continuous random variable can assume any value along a line interval, including every possible value between any two points on the line

Note: Usually associated with a measurement

- Blood Pressure
- Weight gain
- Height

Probability Distributions of Continuous Random Variables

Probability Density Function

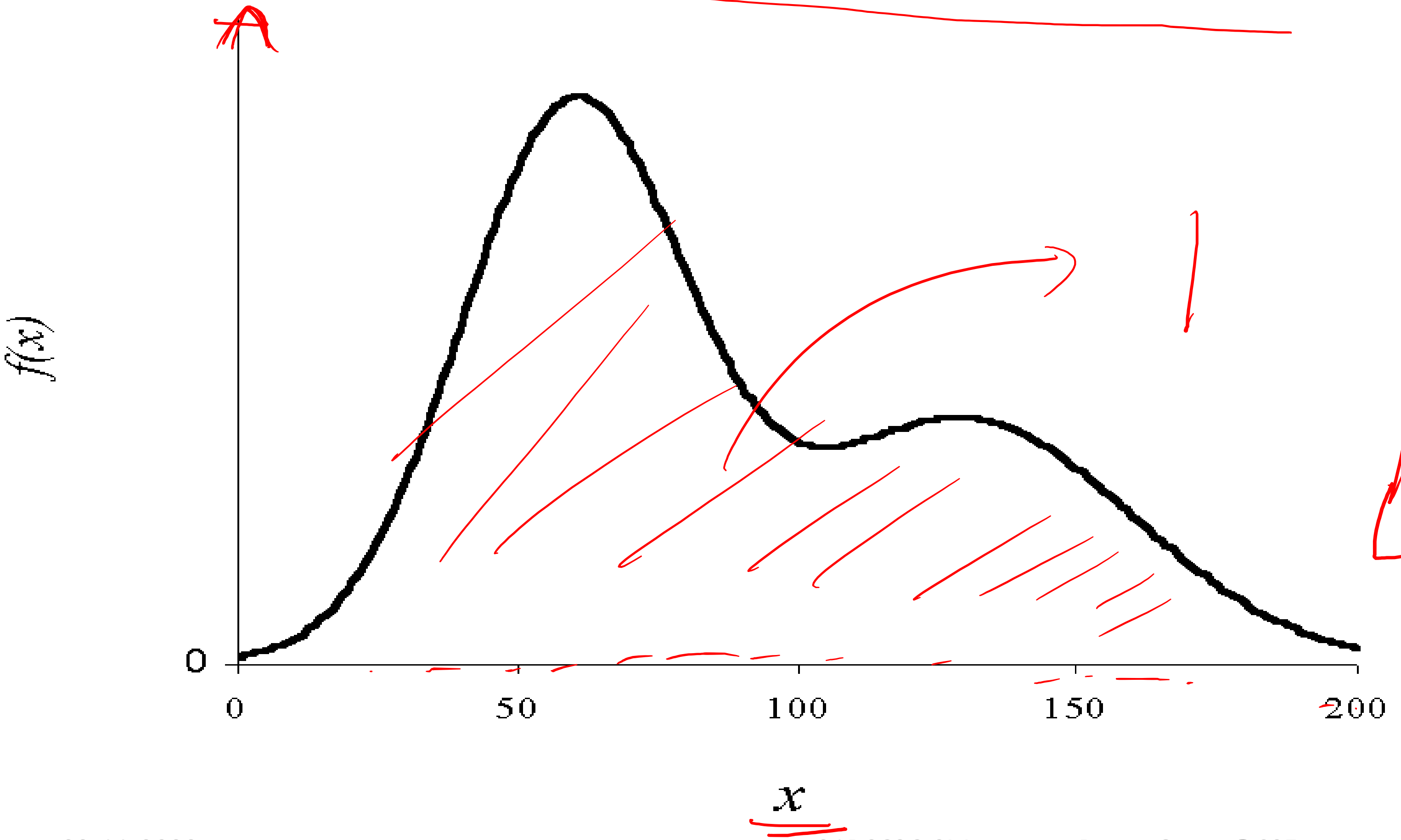
$\Sigma \rightarrow \int$

The **probability distribution** of a ***continuous*** random variable is described by ***probability density curve f(x)***.

$P(X=x) = \sum_{x_i} P(X=x_i)$

$f(x) \geq 0$

$= 1$



$$\boxed{\int_{-\infty}^{\infty} f(x)\,dx = 1}$$

PDF of a Continuous R.V. : Basic Properties

- The pdf of a continuous random variable X must satisfy three conditions.

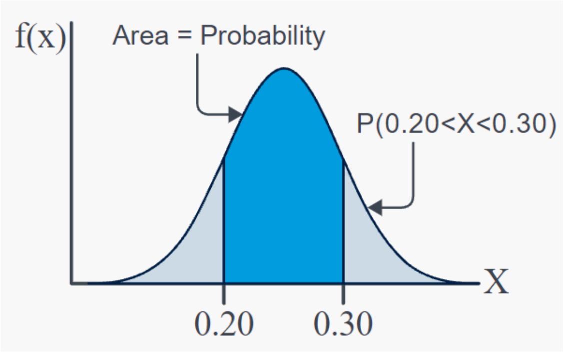
  ✓ It is a nonnegative function **(but unlike in the discrete case it may take on values exceeding 1).**

  ✓ **Its definite integral over the whole real line equals one.** That is $\int_{-\infty}^{\infty} f(x)dx = 1$

  ✓ Its definite integral over a subset B of the real numbers gives the probability that X takes a value in B. That is,

$\int_0^2 f(x)dx$ $$\int_B f(x) = P(X \in B)$$ $B = [0,2]$



for “every” subset B of the real numbers.

As a special case (the usual case) for all real numbers a and b

$$P(a<X<b) = \int_a^b f(x)dx = P(a \leq X \leq b)$$

Put simply, the probability is simply the area under the pdf curve over the interval [a,b].

PDF of a Continuous R.V. : Basic Properties

- If X has uncountable range and such a pdf, then X is a continuous random variable. In this case we often refer to f as a continuous pdf.
  - Note that this means f is the pdf of a continuous random variable. It does not necessarily mean that f is a continuous function.

$(y = |x|)$

- Note that by this definition the probability of X taking on a single value **a** is always 0.

$$P(X = a) = P(a \le X \le a) = \int_a^a f(x)dx = 0$$

$\delta \approx 0.00001$

This is, of course, quite different from the situation for discrete random variables.

$P(X = a) = ?$

- Consequently, we can be sloppy about inequalities. That is

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

Note: **This is blatantly false for discrete random variables.**

$\int_a^b f^*(x)\,dx = g(x)\Big|_a^b = g(b) - g(a)$ valid pdf

Example 1:

→ pdf

$\int_3^5 f(x)\,dx = \int_3^5 x/8\,dx = \frac{x^2}{2}\cdot\frac{1}{8}\Big|_3^5 = \frac{1}{16}[25-9] = 1$

- Suppose that f(x) is f(x)=x/8 3<x<5

Determine the following :

a) P(X<4) $= \int_3^4 f(x)\,dx \quad \int_3^4 x/8\,dx = \frac{x^2}{16}\Big|_3^4 = \frac{16-9}{16} = \frac{7}{16}$

b) P(X>3.5) $> \int_{3.5}^5 x/8\,dx = \frac{x^2}{16}\Big|_{3.5}^5 =$

c) P(4<X<5) $\int_4^5 x/8\,dx = \quad -\infty \quad \infty$

integ

$\int x^n\,dx = \frac{x^{n+1}}{n+1} + C$

0 3 4 5 ∞

Example 1:

- Suppose that f(x) is f(x)=x/8 3<x<5

Determine the following :

a) P(X<4.5) $= \int_{3}^{4.5} x/8 \, dx$

b) P(X<3.5 OR X>4.5)

c) P(X<4.5|4<X<5)

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

$P(A \text{ OR } B) = P(A) + P(B) - P(A \cap B)$

$P(X<3.5) + P(X>4.5)$

$= \int_{3}^{3.5} x/8 \, dx + \int_{4.5}^{5} x/8 \, dx$

$= \square + \square$

d) $P(3.5 < X < 7)$

$= \int_{3.5}^{5} x/8 \, dx + \int_{5}^{7} 0 \, dx =$

$\frac{P(X<4.5 \cap 4<X<5)}{P(4<X<5)}$

$\frac{P(4<X<4.5)}{P(4<X<5)}$

$\int_{4}^{4.5} x/8 \, dx \Big/ \int_{4}^{5} x/8 \, dx$

Example 2:

Let X be a continuous random variable whose probability density function is:

$$f(x) = \frac{x^3}{4} \geq 0$$

for an interval $0 < x < c$. What is the value of the constant c that makes $f(x)$ a valid probability density function?

$$\int_a^b x^n \, dx = \frac{x^{n+1}}{n+1}\Big|_a^b$$

Sol: valid Pdf $\Rightarrow \int_{-\infty}^{\infty} f(x)\,dx = 1$ also $f(x) \geq 0$

$$\therefore \int_0^c x^3/4 \, dx = \frac{1}{4}\frac{x^4}{4}\Big|_0^c = 1 \Rightarrow \frac{1}{16} c^4 = 1$$

$$\Rightarrow \left(\frac{c}{2}\right)^4 = 1$$

$$\Rightarrow \left(\frac{c}{2}\right)^4 = (1)^4$$

$$\Rightarrow \boxed{c = 2}$$

Example 3:

- X be a continuous random variable and its pdf

$$f(x) = \begin{cases} ax & 0 \le x \le 1 \\ a & 1 \le x \le 2 \\ 3a - ax & 2 \le x \le 3 \end{cases}$$

w.k.t $\int_{-\infty}^{\infty} f(x)\,dx = 1$

Find the following:

if $f(x)$ is pdf $\Rightarrow \int_{-\infty}^{\infty} f(x)\,dx = 1$

- a
- P(0<X<1.7)

$\Rightarrow \int_0^1 ax\,dx + \int_1^2 a\,dx + \int_2^3 (3a - ax)\,dx$

$\Rightarrow \frac{ax^2}{2}\Big|_0^1 + ax\Big|_1^2 + \left(3ax - \frac{ax^2}{2}\right)\Big|_2^3 = 1$

$\Rightarrow \frac{a}{2}(1-0) + a(2-1) + \left(9a - \frac{9a}{2} - 6a + 2a\right) = 1$

Example:

Let the continuous random variable X denote the distance in micrometers from the start of a track on a magnetic disk until the first flaw. Historical data show that the distribution of X can be modeled by a pdf $f(x) = \frac{1}{2000} e^{-x/2000}$, $x \geq 0$. For what proportion of disks is the distance to the first flaw greater than 1000 micrometers?

Solution. The density function and the requested probability are shown in Fig. 3-9.
Now,

$$P(X > 1000) = \int_{1000}^{\infty} f(x)\, dx = \int_{1000}^{\infty} \frac{e^{-x/2000}}{2000}\, dx = -e^{-x/2000} \Big|_{1000}^{\infty} = e^{-1/2} = 0.607$$

What proportion of parts is between 1000 and 2000 micrometers?

Solution. Now,

$$P(1000 < X < 2000) = \int_{1000}^{2000} f(x)\, dx = -e^{-x/2000} \Big|_{1000}^{2000} = e^{-1/2} - e^{-1} = 0.239$$

Because the total area under $f(x)$ equals 1, we can also calculate $P(X < 1000) = 1 - P(X > 1000) = 1 - 0.607 = 0.393$.

Cumulative Probability Distribution

F

- The cumulative distribution function (cdf) $F(x)$ for a discrete rv X gives, for any specified number x, the probability $P(X \leq x)$.i.e.

X is d.r.v $P(X=x) \rightarrow$ Pmf

$P(X \leq 9.5)$

$$F(x) = F_X(x) = P(X \leq x) = \sum_{y \leq x} P_X(X = y)$$

$X \rightarrow$ 0 7, 8, 8.5 9.1 0 9 9.5 - - - 100

- It is obtained by summing the pmf p(Y=y) over all possible values y satisfying y ≤ x.

- The cdf F(x) of a continuous random variable has the same definition as that for a discrete random variable. That is,

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

P

d.r.v cdf pdf c.r.v cdf

The Probability distribution of Random Variable

- The probability distribution is a complete probabilistic description of a random variable.

  - Once we know the probability distribution of a random variable, we know everything we can learn about it from statistics.

  - All statistical concepts (expectation, variance, etc.) are derived from it.

Mean(Expectation) and Variance

- Two numbers are often used to summaries the probability distribution for a random variable X.
  - The **mean/expectation/expected value/average/weighted average** is the measure of the **center** of the probability distribution.
  - The variance is the measure of the **dispersion or variability** in the probability distribution

Expected value, or mean

- If we understand the underlying probability function of a certain phenomenon, then we can make informed decisions based on how we expect *X* to behave on-average over the long-run…
  - (so called "frequentist" theory of probability).

- Expected value is just the weighted average or mean of random variable *X*.
  - Imagine placing the masses $p(x)$ at the points X on a beam; the balance point of the beam is the expected value of x.

Definition: Expectation of R.V. X

- Let X denote a **discrete** random variable with probability function $p(x)$ then the expected value of X, $E(X)$ is defined to be:

$$\boxed{\mu = E(X) = \sum_x xp(x) = \sum_i x_i p(x_i)}$$

- and if X is **continuous** with probability density function $f(x)$

$$\boxed{E(X) = \int_{-\infty}^{\infty} xf(x)dx}$$

- **$E(X) = \mu$ → these symbols are used interchangeably**
- **Expected value is an extremely useful concept for good decision-making!**

Notes:

- The mean is a *weighted average* of the values of X.
- The mean is the *long-run average* value of the random variable.
- The mean is *center of gravity* of the probability distribution of the random variable

Empirical Mean is a special case of Expected Value…

Sample mean, for a sample of n subjects: =

$$\bar{x} = \quad \bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \sum_{i=1}^{n} x_i \left(\frac{1}{n}\right)$$

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

1 2 3 4 5

The probability (frequency) of each person in the sample is 1/n.

Example

If X is a random integer between 1 and 10, what's the expected value of X?

$$\mu = E(x) = \sum_{i=1}^{10} i\left(\frac{1}{10}\right) = \frac{1}{10}\sum_{i}^{10} i = (.1)\frac{10(10+1)}{2} = 55(.1) = 5.5$$

X	1	2	3	4	5	6	7	8	9	10
P(x)	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$			$\frac{1}{10}$				$\frac{1}{10}$

$E[X] = \sum x P(x) = 1\frac{1}{10} + 2\frac{1}{10} + \ldots$

$= \sum x_i / 10 \quad 10 \cdot \frac{1}{10}$

AMRITA VISHWA VIDYAPEETHAM

Example: the lottery

The Lottery

- A certain lottery works by picking 6 numbers from 1 to 49. It costs $1.00 to play the lottery, and if you win, you win $2 million after taxes.

- If you play the lottery once, what are your expected winnings or losses?

X : -1 $+2$ million

$P(X=x)$:

Lottery

Calculate the probability of winning in 1 try:

$$\frac{1}{(C_6^{49})} = \frac{1}{\frac{49!}{43!\,6!}} = \frac{1}{13{,}983{,}816}$$

$$= 7.2 \text{ x } 10^{-8}$$

“49 choose 6”

Out of 49 numbers, this is the number of distinct combinations of 6.

The probability function (note, sums to 1.0):

x$	*p(x)*
-1	.999999928
+ 2 million	*7.2 x 10^{-8}*

$\Rightarrow 1 - 7.2 \times 10^{-8}$

$1 - 0.000000072$

Expected Value

The probability function

$x\$$	$p(x)$
-1	.999999928
+ 2 million	7.2×10^{-8}

Question: If you play the lottery every week for 10 years, what are your expected winnings or losses?

520 x (-.86) = -$447.20

Expected Value

E(X) = P(win)*$2,000,000 + P(lose)*-$1.00

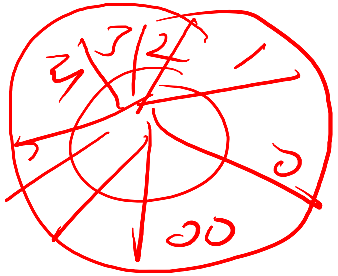
$= 2.0 \times 10^6 * 7.2 \times 10^{-8} + .999999928\ (-1) = .144 - .999999928 = -\$.86$

- Negative expected value is never good!
- You shouldn't play if you expect to lose money!

Gambling (or how casinos can afford to give so many free drinks…)

38

- A roulette wheel has the numbers 1 through 36, as well as 0 and 00. If you bet \$1 that an odd number comes up, you win or lose \$1 according to whether or not that event occurs. If random variable X denotes your net gain, X=1 with probability 18/38 and X= -1 with probability 20/38.



- Solution: E(X) = 1(18/38) – 1 (20/38) = -\$.053

On average, the casino wins (and the player loses) 5 cents per game.

X	-1	+1
P(X=x)	20/38	18/38

The casino rakes in even more if the stakes are higher:

E(X) = 10(18/38) – 10 (20/38) = -\$.53

If the cost is \$10 per game, the casino wins an average of 53 cents per game. If 10,000 games are played in a night, that's a cool \$5300.

Example:

$Y = X^2 + 1$

$Y = X^3 + X^2 + 2$

- X be a continuous random variable and its pdf

$$f(x) = \begin{cases} 0.5x & 0 \le x \le 1 \\ 0.5 & 1 \le x \le 2 \\ 3/2 - 0.5x & 2 \le x \le 3 \end{cases}$$

- Find the expected value of X.

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

$$\int_0^1 0.5x\, dx + \int_1^2 0.5\, dx + \int_2^3 (3/2 - 0.5x)\, dx$$

$$E[X] = \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^1 x(0.5x)\, dx + \int_1^2 0.5 \times x\, dx + \int_2^3 x(3/2 - 0.5x)\, dx$$

$$= 0.5 \frac{x^3}{3}\Big|_0^1 + 0.5 \frac{x^2}{2}\Big|_1^2 + \int_2^3 x(3/2 - 0.5x)\, dx$$

$$= (0.5/3)(1^3 - 0) + \frac{1}{4}(4-1) + \left(\frac{3}{2}\frac{x^2}{2} - 0.5\frac{x^3}{3}\right)\Big|_2^3$$

$$E(X^2) = \int_0^1 x^2 (0.5x)\, dx + \ldots$$

Definition: Expected Value of g(X), X is a R.V.

X

Let X denote a **discrete** random variable with probability function $p(x)$ then the expected value of $g(X)$, $E[g(X)]$ is defined to be:

pmf

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

X

g(X)

$y = X^2$

g(x)

and if X is **continuous** with probability density function $f(x)$

of X

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

Expected Value as a mathematical operator

$X = 6$

If c= a constant number (i.e., not a variable) and *X* and *Y* are any random variables…

$Y = cX$

- E(c) = c
- E(c*X*)=cE(*X*)
  - **Example:** If the casino charges \$10 per game instead of \$1, then the casino expects to make 10 times as much on average from the game (See roulette example in previous slides!)
- E(c + *X*)=c + E(*X*)
  - **Example:** if the casino throws in a free drink worth exactly \$5.00 every time you play a game, you always expect to (and do) gain an extra \$5.00 regardless of the outcome of the game.
- E(*X*+*Y*)= E(*X*) + E(*Y*)
  - **Example:** If you play the lottery twice, you expect to lose: -\$.86 + -\$.86.
  - NOTE: This works even if X and Y are dependent!! Does not require independence!!

$$E[g(X)] = \begin{cases} \sum_x g(x)p(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} g(x)f(x)dx & \text{if } X \text{ is continuous} \end{cases}$$

Rules: $E[c] = c$ where c is a constant

Proof if $g(X) \equiv c$ then $E[g(X)] = E[c] = \int_{-\infty}^{\infty} cf(x)dx$

$$= c \int_{-\infty}^{\infty} f(x)dx = c$$

The proof for discrete random variables is similar.

$E[aX + b] = aE[X] + b$ where a, b are constants

$E(cX)$

$= cE[X]$

$E[c+X]$

$= c + E(X)$

Proof

if $g(X) \equiv aX + b$ then $E[aX + b] = \int_{-\infty}^{\infty} (ax + b)f(x)dx$

$X \longrightarrow f(x)$

$$= a \int_{-\infty}^{\infty} xf(x)dx + b \int_{-\infty}^{\infty} f(x)dx$$

$Y = g(X) = X^2 + 1$

$$= aE(X) + b$$

The proof for discrete random variables is similar.

Example:

- At a particular times, a chemical supply company keeps a slot of 200 lb of methylene chloride. It takes the substance 10lb/bottle. Assume that the no. of bottles X is random variable with pmf

X	1	2	3	4
P(X=x)	0.4	0.2	0.1	0.3

- Compute mean and mean of the left substance

1) $E[X] = \sum_x x P(x) = (1 \times 0.4) + (2 \times 0.2) + (3 \times 0.1) + (4 \times 0.3)$
$= 0.4 + 0.4 + 0.3 + 1.2 = 2.3$

$\boxed{Y = 200 - 10X}$

w.r.t
$E(aX+b) = aE(X) + b$

$\Rightarrow E[Y] = E[200 - 10X]$
$= 200 - 10E[X]$
$= 200 - 10 \times (2.3)$
$= 200 - 23$
$= 177$

Example:

Suppose X has a uniform distribution from 0 to b.

valid pdf

Then:

$$\boxed{f(x) = \begin{cases} \frac{1}{b} & 0 \le x \le b \\ 0 & x < 0, x > b \end{cases}}$$

Find the **expected value** of $A = X^2$.

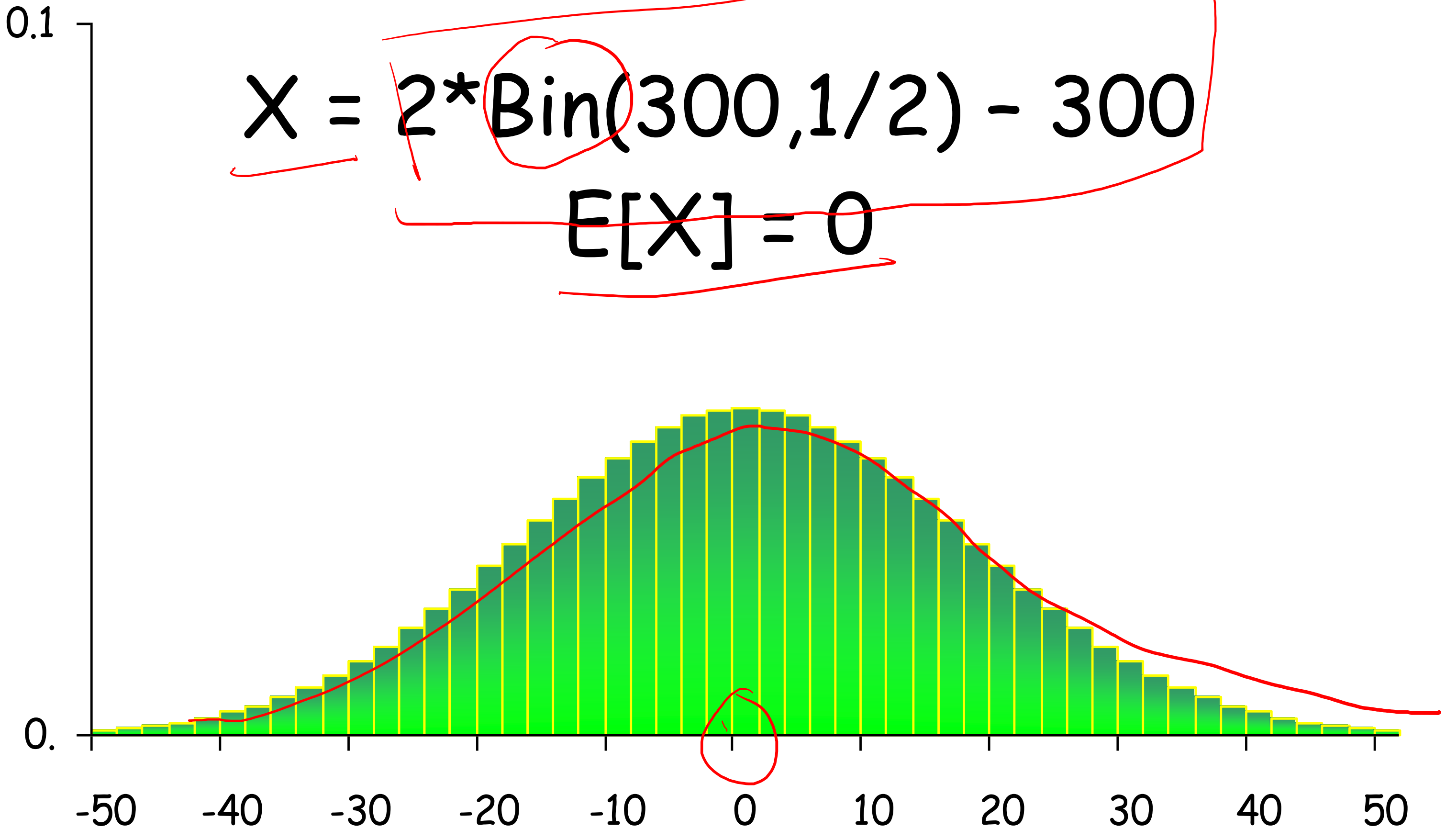
If X is the length of a side of a square (chosen at random form 0 to b) then A is the area of the square

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^b x^2 \frac{1}{b-0} dx = \left[\frac{1}{b}\frac{x^3}{3}\right]_0^b = \frac{b^3 - 0^3}{3(b)} = \frac{b^2}{3} = b^2\left(\frac{1}{3}\right)$$

= 1/3 the maximum area of the square

Expected value isn't everything though…

Histogram 1



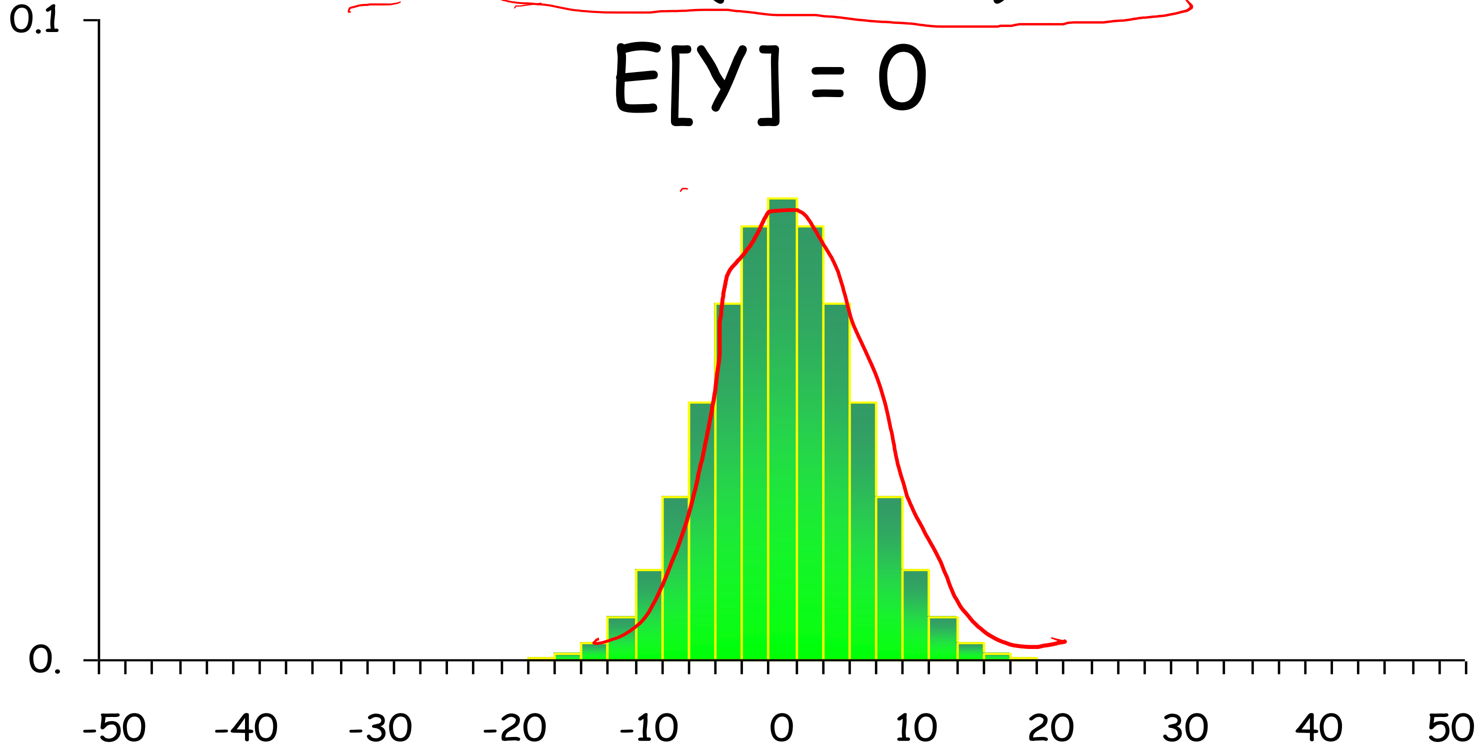
$$X = 2 * Bin(300, 1/2) - 300$$

$$E[X] = 0$$

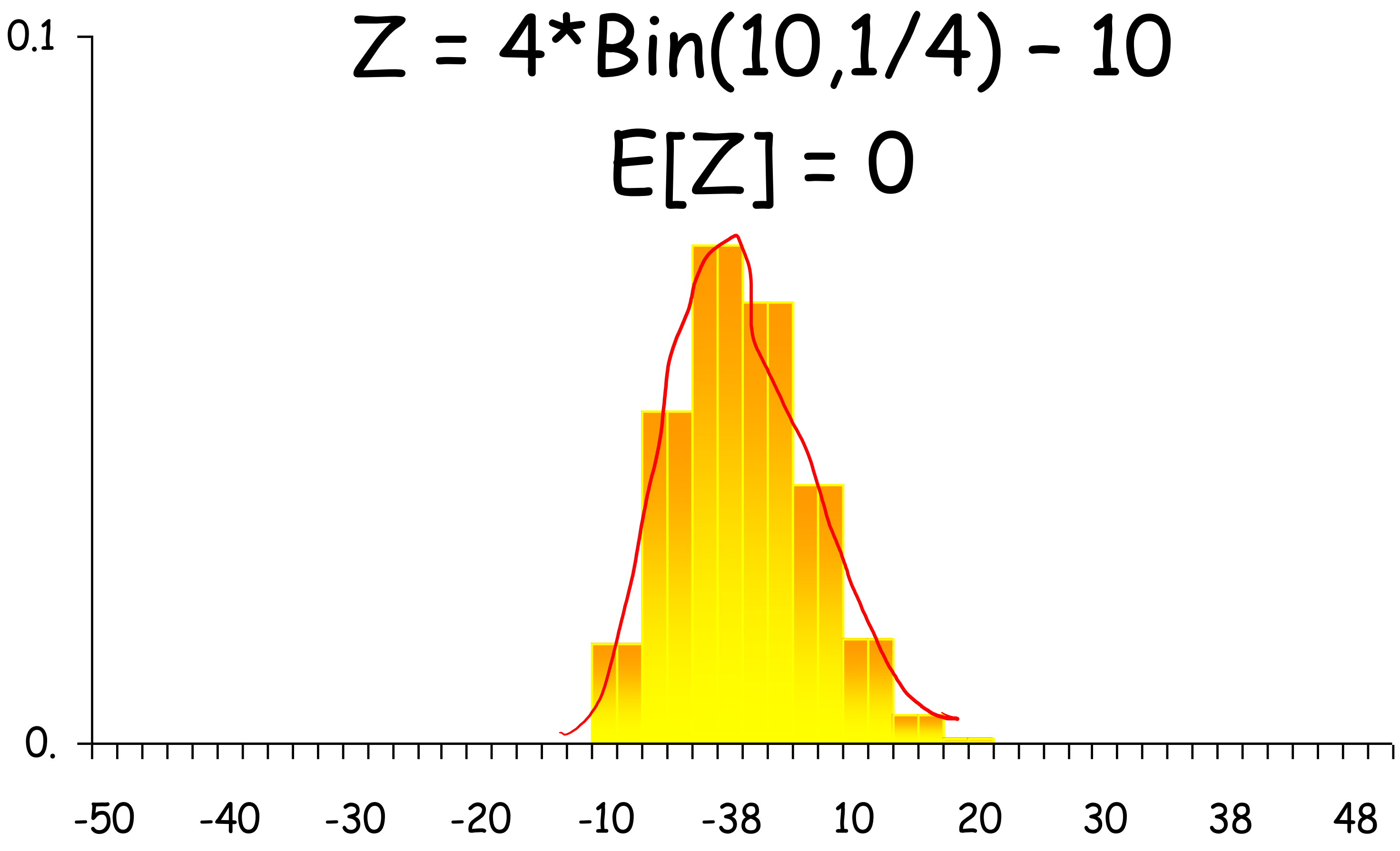
Histogram 2

$$Y = 2*Bin(30,1/2) - 30$$

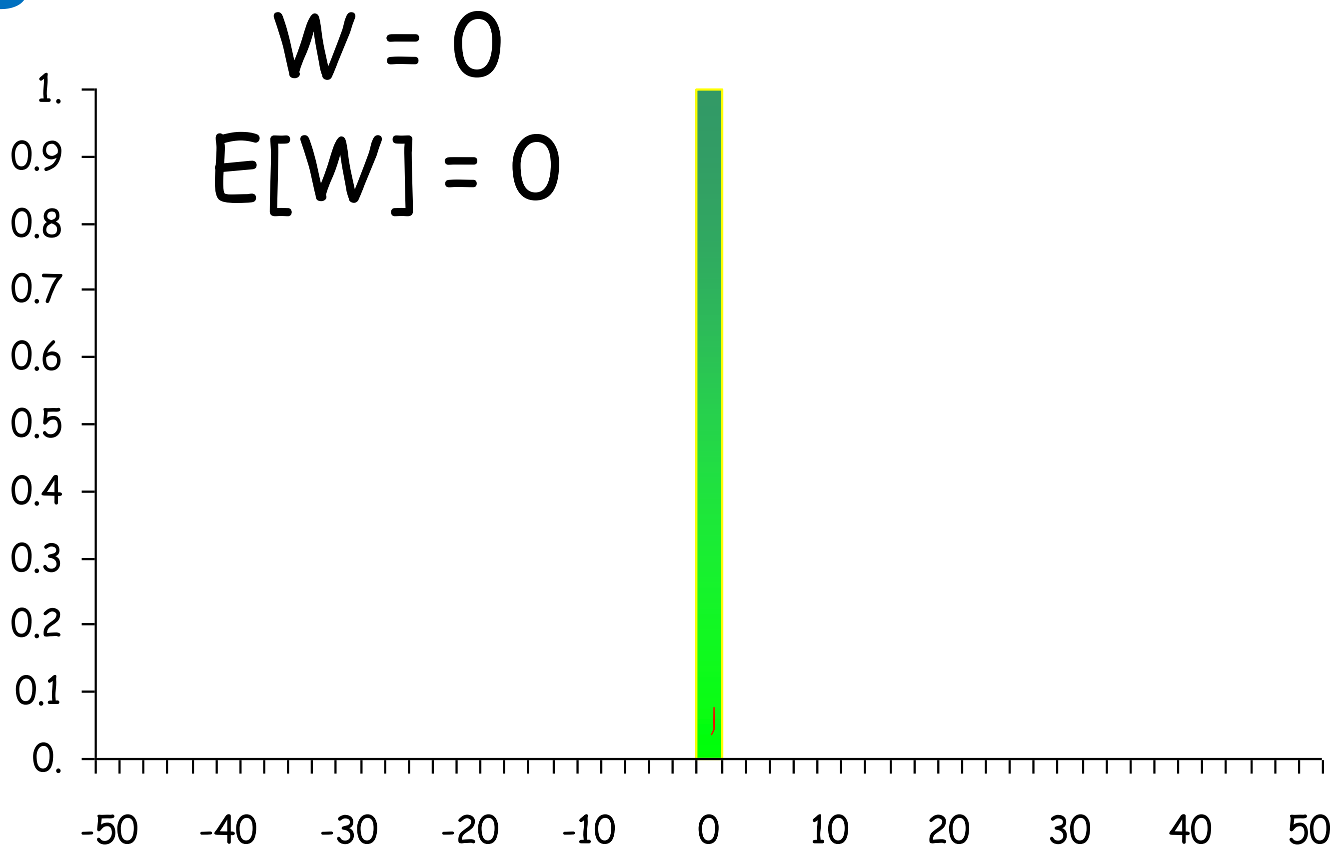
$$E[Y] = 0$$



Histogram 3



Histogram 4



A natural question:

- Is there a good parameter that allow to distinguish between these distributions?

- Is there a way to measure the spread?

Answer: Variance

Variance and Standard Deviation

μ

σ^2

- The variance of X, denoted by Var(X) is the mean squared deviation of X from its expected value $\mu = E(X)$:

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

- The standard deviation of X, denoted by SD(X) is the square root of the variance of X: **SD(X) =** $\sqrt{\boldsymbol{Var(X)}}$

- ***Note:*** We square because squaring has better properties than absolute value. Take square root to get back linear average distance from the mean (="standard deviation").

Discrete case:

$E((X-\mu)^2)$

$= E(g(X))$

$\sum_x (x-\mu)^2 f(x)$

$$\boxed{Var(X) = \sigma^2 = \sum_{\text{all } x} (x_i - \mu)^2 p(x_i)}$$

Continuous case:

$E[(X-\mu)^2]$

$$\boxed{Var(X) = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx}$$

Computational Formula for Variance

Claim: $\mathrm{Var}(X) = E(X^2) - E(X)^2.$

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

Proof:

$$V(X) - E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2] = \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) f(x)\,dx$$

$(a+b)^2 = a^2 + b^2 + 2ab$

$$= E[X^2] - 2\mu \underbrace{E[X]}_{=\mu} + \mu^2$$

$$= \int_{-\infty}^{\infty} x^2 f(x)\,dx - 2\mu \int_{-\infty}^{\infty} x f(x)\,dx$$

$$= E[X^2] - 2\mu^2 + \mu^2$$

$$= E[x^2] + \mu^2 \int_{-\infty}^{\infty} f(x)\,dx - 2\mu E[x] + \mu^2 \cdot 1$$

$$\boxed{V(X) = E[X^2] - (E[X])^2}$$

Example

Find the variance and standard deviation for the number of ships to arrive at the harbor.

x	10	11	12	13	14
P(x)	.4	.2	.2	.1	.1

Find → $Var(X) = E(X^2) - [E(X)]^2$

$E(X) = \sum_{i=1}^{5} x_i \ \ p(x_i) = (10)(.4) + (11)(.2) + 12(.2) + 13(.1) + 14(.1) = 11.3$

$$E(X^2) = \sum_{i=1}^{5} {x_i}^2 p(x_i) = (100)(.4) + (121)(.2) + 144(.2) + 169(.1) + 196(.1) = 129.5$$

$Var(X) = E(X^2) - [E(X)]^2 = 129.5 - 11.3^2 = 1.81$

$stddev(X) = \sqrt{1.81} = 1.35$

Interpretation: **On an average day, we expect 11.3 ships to arrive in the harbor, plus or minus 1.35. This gives you a feel for what would be considered a usual day!**

Variance as a Mathematical Operator

If c= a constant number (i.e., not a variable) and *X* and *Y* are random variables, then

1. Var(c) = 0
2. Var (c+X)= Var(X)
3. Var(cX)= c^2Var(X)
4. Var(X+Y)= Var(X) + Var(Y) **ONLY IF X and Y are independent!!!!**
5. Var(X+Y)= Var(X) + Var(Y)+2Cov(X,Y) **IF X and Y are not independent**

$$E((cX)^2) - (E[cX])^2 \quad cE[X]$$
$$= E[c^2X^2] - (E(X))^2$$
$$= c^2E[X^2] - c^2(E(X))^2$$
$$= c^2[E[X^2] - (E[X])^2]$$
$$= c^2V(X)$$

Example:

- If X be a random variable with pdf

$$f(x) = \begin{cases} \frac{x^2}{3} & -1 < x < 2 \\ 0 & otherwise \end{cases}$$

- Find expected value and variance. Also obtain expected value and variance of 4X+3.

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

$V[X] = E[X^2] - (E[X])^2$

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

$V(aX+b) = a^2 V(X)$

$E[X] = \int_{-1}^{2} x \cdot \frac{x^2}{3} dx = \int_{-1}^{2} \frac{x^3}{3} dx = \frac{x^4}{3 \times 4}\Big|_{-1}^{2} = \frac{1}{12}[16-1] = \frac{15}{12} = \frac{5}{4}$

$V[X] = E[X^2] - (E[X])^2 = \int_{-1}^{2} x^2 \cdot \frac{x^2}{3} dx - \left(\frac{5}{4}\right)^2 = \frac{1}{3 \times 5} x^5 \Big|_{-1}^{2} - \frac{25}{16} = \frac{1}{15}(32+1) - \frac{25}{16} = \frac{33}{15} - \frac{25}{16} = \boxed{\ }$

Why to Study Moments of a R.V. --Motivation

- X and Y two discrete random variables with pmfs

X=x	-6	6
P(Y=y)	1/2	1/2

Y	-12	0	12
P(Y=y)	1/8	3/4	1/8

$E[X]$

$E[X^2]$

$E[X^3]$

$E[X^4]$