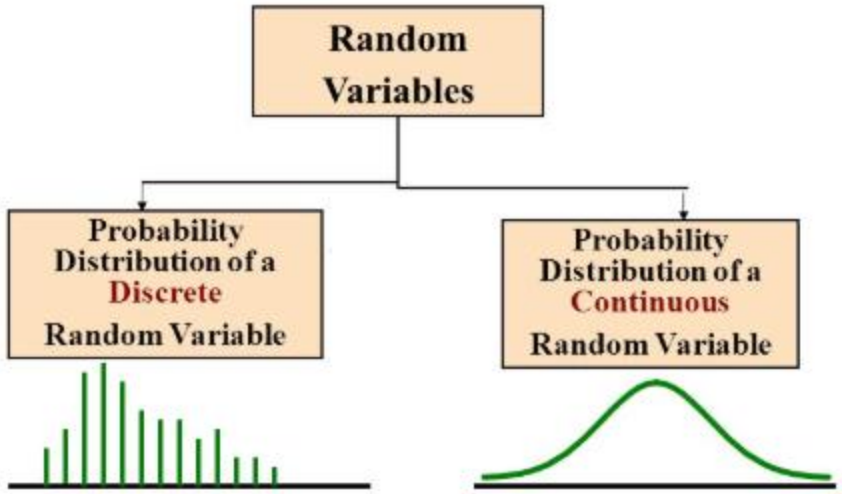
Random Variables

- In an experiment, a measurement is usually denoted by a variable such as X.

- In a random experiment, a variable whose measured value can change (from one replicate of the experiment to another) is referred to as a **random variable**.

Definition of a Random Variable

- A numerical value to each outcome of a particular experiment
- Example 1 : Machine Breakdowns
  - Sample space : $S = \{electrical, mechanical, misuse\}$
  - Each of these failures may be associated with a repair cost
  - State space : $\{50, 200, 350\}$
  - Cost is a random variable : 50, 200, and 350



A **discrete** random variable is a random variable with a finite (or countably infinite) set of real numbers for its range.
A **continuous** random variable is a random variable with an interval (either finite or infinite) of real numbers for its range.

Examples of **continuous** random variables:

electrical current, length, pressure, temperature, time, voltage, weight

Examples of **discrete** random variables:

number of scratches on a surface, proportion of defective parts among 1,000 tested, number of transmitted bits received in error

Probability

- Used to quantify likelihood or chance

- Used to represent risk or uncertainty in engineering applications

- Can be interpreted as our **degree of belief** or **relative frequency**

Example: Your team has won 9 games from a total of 12 games played:
the **Frequency** of winning is 9. the **Relative Frequency** of winning is 9/12 = 75%

Probability

- Probability statements describe the likelihood that particular values occur.

- The likelihood is quantified by assigning a number from the interval [0, 1] to the set of values (or a percentage from 0 to 100%).

- Higher numbers indicate that the set of values is more likely.

Probability

- A probability is usually expressed in terms of a random variable.
- For the part length example, X denotes the part length and the probability statement can be written in either of the following forms

$$P(X \in [10.8, 11.2]) = 0.25 \qquad \text{or} \qquad P(10.8 \leq X \leq 11.2) = 0.25$$

- Both equations state that the probability that the random variable X assumes a value in [10.8, 11.2] is 0.25.

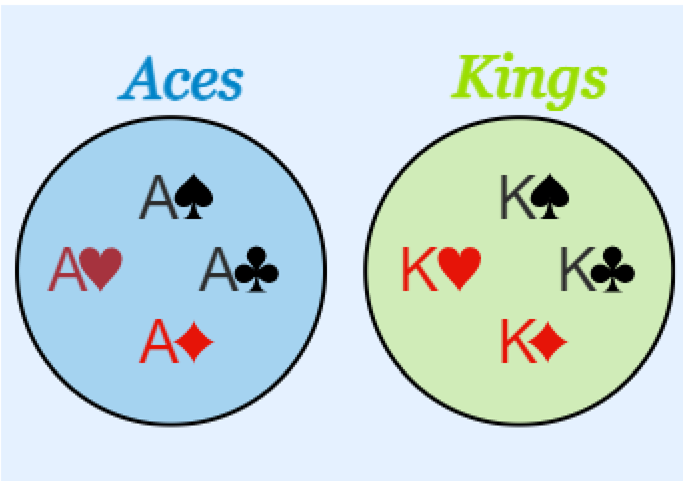
Probability

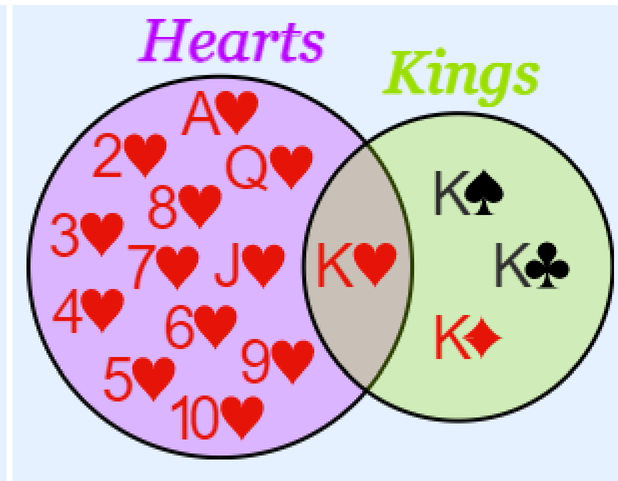
<u>Complement of an Event</u>

- Given a set E, the complement of E is the set of elements that are not in E. The **complement** is denoted as E'.

<u>Mutually Exclusive Events</u>

- The sets E_1 , E_2 ,…,E_k are **mutually exclusive** if the intersection of any pair is empty. That is, each element is in one and only one of the sets E_1 , E_2 ,…,E_k .





Aces and Kings are
Mutually Exclusive
(can't be both)

Hearts and Kings are
not Mutually Exclusive
(can be both)

Probability

Probability Properties

1. $P(X \in R) = 1$, where R is the set of real numbers.
2. $0 \leq P(X \in E) \leq 1$ for any set E. (3-1)
3. If $E_1, E_2, \ldots, E_k$ are mutually exclusive sets,
   $P(X \in E_1 \cup E_2 \cup \ldots \cup E_k) = P(X \in E_1) + \cdots + P(X \in E_k)$.

Probability

Events

- A measured value is not always obtained from an experiment. Sometimes, the result is only classified (into one of several possible categories).
- These categories are often referred to as **events**.

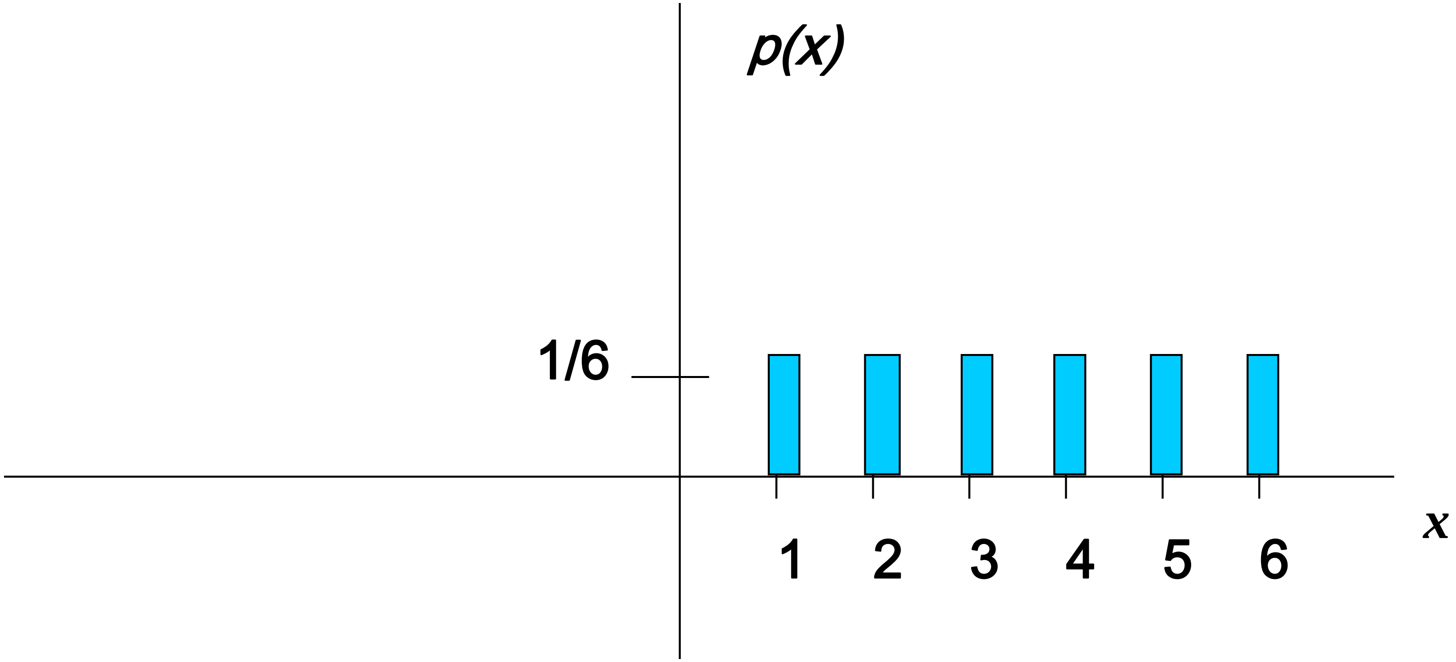
<u>Illustrations</u>

- The current measurement might only be recorded as *low, medium,* or *high*; a manufactured electronic component might be classified only as defective or not; and either a message is sent through a network or not.

Probability functions

- A probability function maps the possible values of x against their respective probabilities of occurrence, $p(x)$
- $p(x)$ is a number from 0 to 1.0.
- The area under a probability function is always 1.

Discrete example: roll of a die

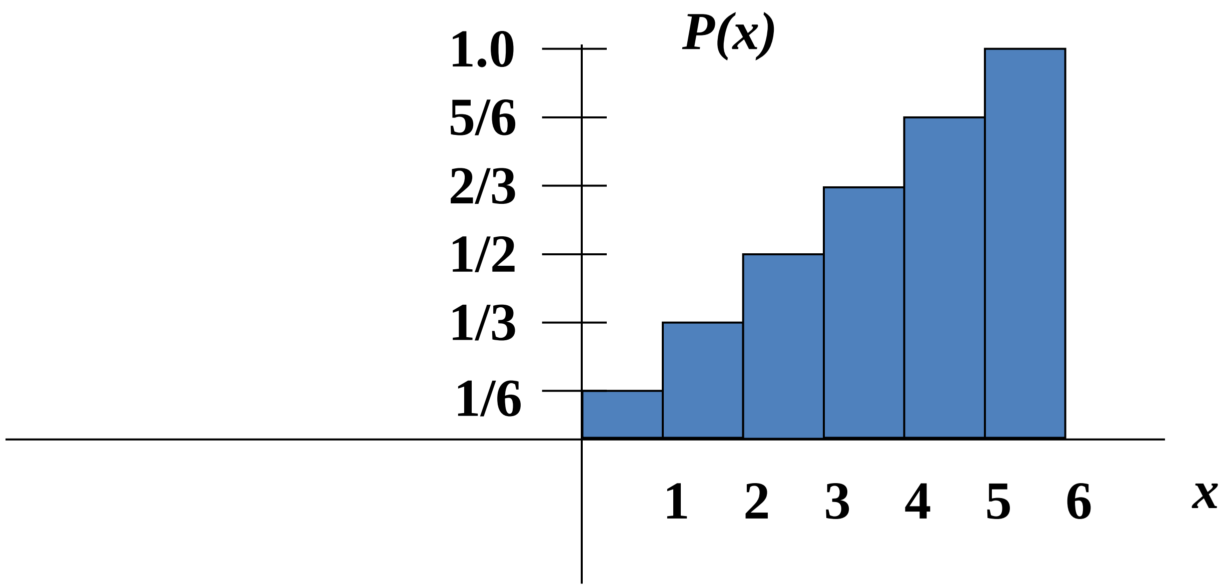


$$\sum_{\text{all } x} P(x) = 1$$

Probability mass function (pmf)

x	$p(x)$
1	$p(x=1)=1/6$
2	$p(x=2)=1/6$
3	$p(x=3)=1/6$
4	$p(x=4)=1/6$
5	$p(x=5)=1/6$
6	$\underline{p(x=6)=1/6}$
	1.0

Cumulative distribution function (CDF)



Cumulative distribution function

x	$P(x \le A)$
1	$P(x \le 1)$=1/6
2	$P(x \le 2)$=2/6
3	$P(x \le 3)$=3/6
4	$P(x \le 4)$=4/6
5	$P(x \le 5)$=5/6
6	$P(x \le 6)$=6/6

Examples

1. What's the probability that you roll a 3 or less?

$P(x \leq 3) = 1/2$

2. What's the probability that you roll a 5 or higher?

$P(x \geq 5) = 1 - P(x \leq 4) = 1\text{-}2/3 = 1/3$

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)

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

x	***f(x)***
9	**.25**
10	**.25**
11	**.25**
12	**<u>.25</u>**
	1.0

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	**<u>13/25</u>**

24/25

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

Practice Problem:

- The number of ships to arrive at a harbor on any given day is a random variable represented by *x*. The probability distribution for *x* is:

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

Find the probability that on a given day:

a. exactly 14 ships arrive $p(x=14) = .1$

b. At least 12 ships arrive $p(x \geq 12) = (.2 + .1 + .1) = .4$

c. At most 11 ships arrive $p(x \leq 11) = (.4 + .2) = .6$

Practice Problem:

You are lecturing to a group of 1000 students. You ask them to each randomly pick an integer between 1 and 10. Assuming, their picks are truly random:

- What's your best guess for how many students picked the number 9?

  Since $p(x=9)$ = 1/10, we'd expect about 1/10th of the 1000 students to pick 9. 100 students.

- What percentage of the students would you expect picked a number less than or equal to 6?

  Since $p(x \leq 6)$ = 1/10 + 1/10 + 1/10 + 1/10 + 1/10 + 1/10 =.6 60%

Probability Density Function

Probability Density Function (PDF)

- used to specify the probability of the random variable falling within a particular range of values
  - Probabilistic properties of a continuous random variable

The **probability density function** (or pdf) $f(x)$ of a continuous random variable is used to determine probabilities from areas as follows:

$$P(a < X < b) = \int_a^b f(x)\,dx \tag{3-2}$$

The properties of the pdf are

(1) $f(x) \geq 0$

(2) $\int_{-\infty}^{\infty} f(x) = 1$

Example of Continuous Random Variables

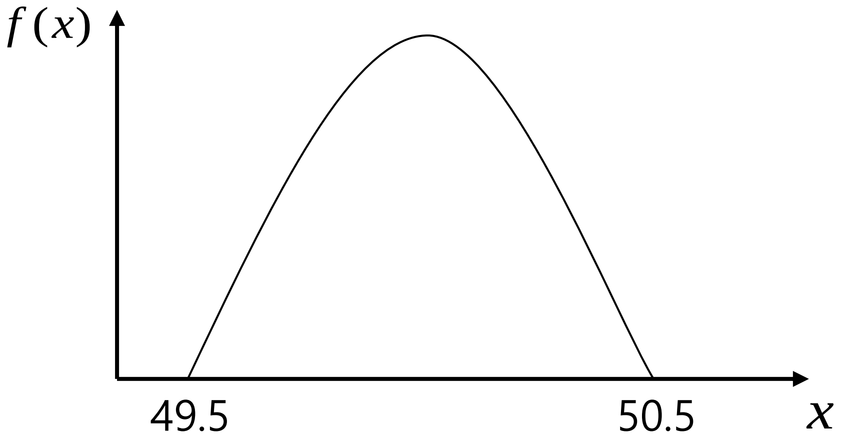
- Example: Metal Cylinder Production
  - Suppose that the random variable X is the diameter of a randomly chosen cylinder manufactured by the company. Since this random variable can take **any value between** 49.5 and 50.5, it is a **continuous** random variable.

Probability Density Function

- Example
  - Suppose that the diameter of a metal cylinder has a p.d.f

$$f(x) = 1.5 - 6(x - 50.0)^2 \quad \text{for} \quad 49.5 \le x \le 50.5$$

$$f(x) = 0, \quad \text{elsewhere}$$



Probability Density Function

- This is a valid p.d.f.

$$\int_{49.5}^{50.5} (1.5 - 6(x - 50.0)^2)dx = [1.5x - 2(x - 50.0)^3]_{49.5}^{50.5}$$

$$= [1.5 \times 50.5 - 2(50.5 - 50.0)^3]$$

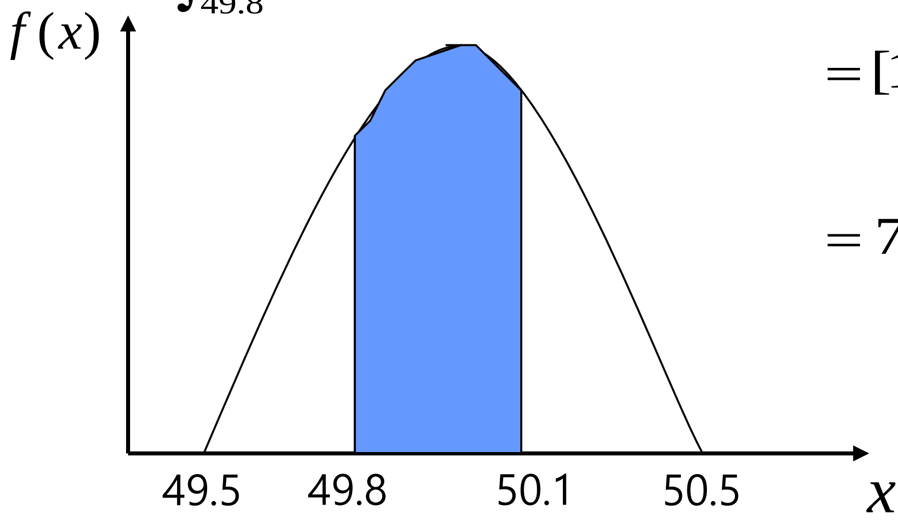
$$-[1.5 \times 49.5 - 2(49.5 - 50.0)^3]$$

$$= 75.5 - 74.5 = 1.0$$

Probability Density Function

- The probability that a metal cylinder has a diameter between 49.8 and 50.1 mm can be calculated to be

$$\int_{49.8}^{50.1} (1.5 - 6(x-50.0)^2)dx = [1.5x - 2(x-50.0)^3]_{49.8}^{50.1}$$



$$\begin{aligned} &= [1.5 \times 50.1 - 2(50.1-50.0)^3] \\ &\quad - [1.5 \times 49.8 - 2(49.8-50.0)^3] \\ &= 75.148 - 74.716 = 0.432 \end{aligned}$$

Problem

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 Distribution Function

The **cumulative distribution function** (or cdf) of a continuous random variable X with probability density function $f(x)$ is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\, du$$

for $-\infty < x < \infty$.

- $f(x) = \dfrac{dF(x)}{dx}$
- $P(a < X \leq b) = P(X \leq b) - P(X \leq a)$
  $= F(b) - F(a)$
- $P(a \leq X \leq b) = P(a < X \leq b)$

Probability Density Function

- Example

$$F(x) = P(X \leq x) = \int_{49.5}^{x} (1.5 - 6(y - 50.0)^2)\,dy$$

$$= [1.5y - 2(y - 50.0)^3]_{49.5}^{x}$$

$$= [1.5x - 2(x - 50.0)^3] - [1.5 \times 49.5 - 2(49.5 - 50.0)^3]$$

$$= 1.5x - 2(x - 50.0)^3 - 74.5$$

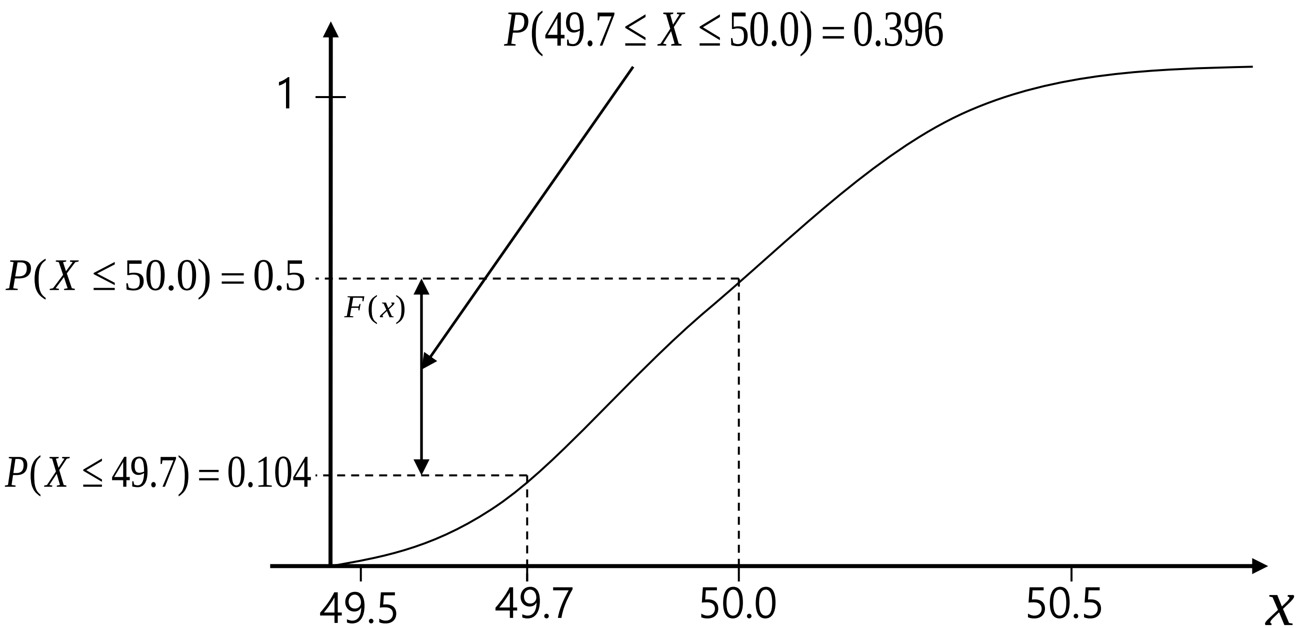
$$P(49.7 \leq X \leq 50.0) = F(50.0) - F(49.7)$$

$$= (1.5 \times 50.0 - 2(50.0 - 50.0)^3 - 74.5)$$

$$-(1.5 \times 49.7 - 2(49.7 - 50.0)^3 - 74.5)$$

$$= 0.5 - 0.104 = 0.396$$

Probability Density Function



Cumulative Distribution Function

Let X be a discrete random variable with range $R_X = \{1, 2, 3, \ldots\}$. Suppose the PMF of X is given by

$$P_X(k) = \frac{1}{2^k} \text{ for } k = 1, 2, 3, \ldots$$

Find $P(2 < X \leq 5)$.
Find $P(X > 4)$.

b. To find $P(2 < X \leq 5)$, we can write

$$P(2 < X \leq 5) = F_X(5) - F_X(2) = \frac{31}{32} - \frac{3}{4} = \frac{7}{32}.$$

Or equivalently, we can write

$$P(2 < X \leq 5) = P_X(3) + P_X(4) + P_X(5) = \frac{1}{8} + \frac{1}{16} + \frac{1}{32} = \frac{7}{32},$$

which gives the same answer.

c. To find $P(X > 4)$, we can write

$$P(X > 4) = 1 - P(X \leq 4) = 1 - F_X(4) = 1 - \frac{15}{16} = \frac{1}{16}.$$

Expectations of Discrete Random Variables

- Expectation of a discrete random variable with p.m.f

$$P(X = x_i) = p_i$$

$$E(X) = \sum_i p_i x_i$$

- Expectation of a continuous random variable with p.d.f $f(x)$

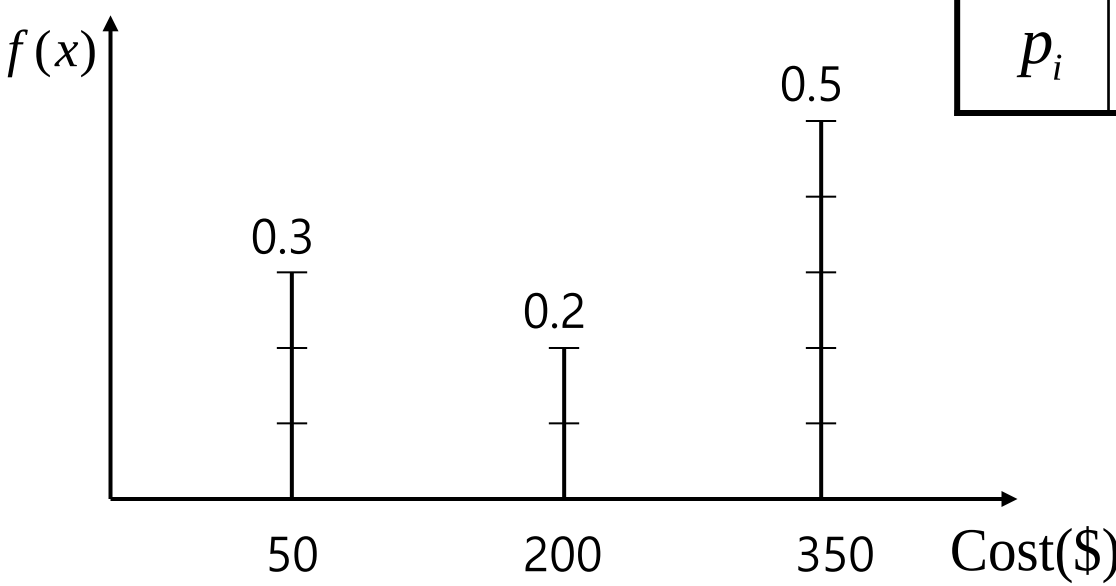
$$E(X) = \int_{\text{state space}} x f(x) dx$$

- The expected value of a random variable is also called the **mean of the random variable**

Probability Mass Function

- Example 1 : Machine Breakdowns
  - P (cost=50)=0.3, P (cost=200)=0.2, P (cost=350)=0.5
  - 0.3 + 0.2 + 0.5 =1

x_i	50	200	350
p_i	0.3	0.2	0.5



Expectations of Discrete Random Variables

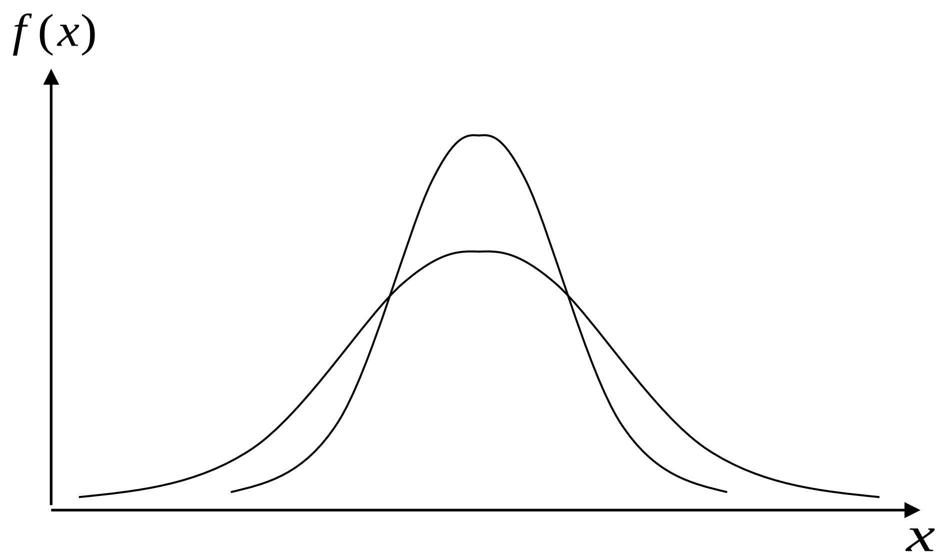
- Example 1 (discrete random variable)
  - The expected repair cost is

$$E(\text{cost}) = (\$50 \times 0.3) + (\$200 \times 0.2) + (\$350 \times 0.5) = \$230$$

Definition and Interpretation of Variance

$$\begin{aligned}\mathrm{Var}(X) &= E((X-E(X))^2) \\ &= E(X^2 - 2XE(X) + (E(X))^2) \\ &= E(X^2) - 2E(X)E(X) + (E(X))^2 \\ &= E(X^2) - (E(X))^2\end{aligned}$$

Two distribution with identical mean values but different variances



Examples of Variance Calculations

- Example

$$\mathrm{Var}(X) = E((X - E(X))^2) = \sum_i p_i (x_i - E(X))^2$$

$$= 0.3(50 - 230)^2 + 0.2(200 - 230)^2 + 0.5(350 - 230)^2$$

$$= 17{,}100 = \sigma^2$$

$$\sigma = \sqrt{17{,}100} = 130.77$$

Binomial Experiments

A **binomial experiment** is a probability experiment that satisfies the following conditions.

1. The experiment is repeated for a fixed number of trials, where each trial is independent of other trials.
2. There are only two possible outcomes of interest for each trial. The outcomes can be classified as a success (S) or as a failure (F).
3. The probability of a success $P(S)$ is the same for each trial.
4. The random variable x counts the number of successful trials.

Notation for Binomial Experiments

Symbol	*Description*
n	The number of times a trial is repeated.
$p = P(S)$	The probability of success in a single trial.
$q = P(F)$	The probability of failure in a single trial. ($q = 1 - p$)
x	The random variable represents a count of the number of successes in n trials: $x = 0, 1, 2, 3, \ldots, n$.

Binomial Experiments

Example:
Decide whether the experiment is a binomial experiment. If it is, specify the values of *n*, *p*, and *q*, and list the possible values of the random variable *x*. If it is not a binomial experiment, explain why.

- You randomly select a card from a deck of cards, and note if the card is an Ace. You then put the card back and repeat this process 8 times.

This is a binomial experiment. Each of the 8 selections represent an independent trial because the card is replaced before the next one is drawn. There are only two possible outcomes: either the card is an Ace or not.

$$n = 8 \qquad p = \frac{4}{52} = \frac{1}{13} \qquad q = 1 - \frac{1}{13} = \frac{12}{13} \qquad x = 0, 1, 2, 3, 4, 5, 6, 7, 8$$

Binomial Experiments

Example:
Decide whether the experiment is a binomial experiment. If it is, specify the values of n, p, and q, and list the possible values of the random variable x. If it is not a binomial experiment, explain why.

- You roll a die 10 times and note the number the die lands on.

This is not a binomial experiment. While each trial (roll) is independent, there are more than two possible outcomes: 1, 2, 3, 4, 5, and 6.

Binomial Probability Formula

In a binomial experiment, the probability of exactly *x* successes in *n* trials is

$$P(x) = {}_nC_x p^x q^{n-x} = \frac{n!}{(n-x)!\,x!} p^x q^{n-x}.$$

Example:

A bag contains 10 chips. 3 of the chips are red, 5 of the chips are white, and 2 of the chips are blue. Three chips are selected, with replacement. Find the probability that you select exactly one red chip.

p = the probability of selecting a red chip $= \frac{3}{10} = 0.3$

q = 1 – *p* = 0.7

n = 3

x = 1

$$P(1) = {}_3C_1(0.3)^1(0.7)^2$$

$$= 3(0.3)(0.49)$$

$$= 0.441$$

Binomial Probability Distribution

Example:

A bag contains 10 chips. 3 of the chips are red, 5 of the chips are white, and 2 of the chips are blue. Four chips are selected, with replacement. Create a probability distribution for the number of red chips selected.

p = the probability of selecting a red chip $= \dfrac{3}{10} = 0.3$

$q = 1 - p = 0.7$

$n = 4$

$x = 0, 1, 2, 3, 4$

x	$P(x)$
0	0.240
1	0.412
2	0.265
3	0.076
4	0.008

The binomial probability formula is used to find each probability.

Finding Probabilities

Example:
The following probability distribution represents the probability of selecting 0, 1, 2, 3, or 4 red chips when 4 chips are selected.

x	$P(x)$
0	0.24
1	0.412
2	0.265
3	0.076
4	0.008

a.) Find the probability of selecting no more than 3 red chips.

b.) Find the probability of selecting at least 1 red chip.

a.) P (no more than 3) = $P\,(x \leq 3) = P\,(0) + P\,(1) + P\,(2) + P\,(3)$

$= 0.24 + 0.412 + 0.265 + 0.076 = 0.993$

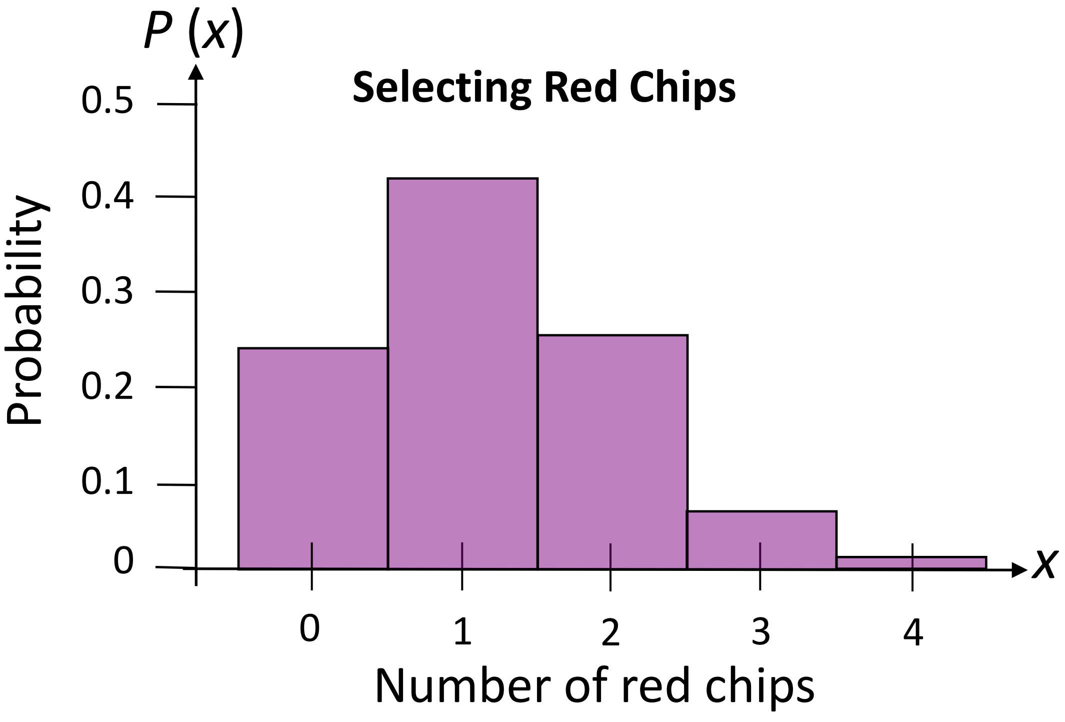
b.) P (at least 1) = $P\,(x \geq 1) = 1 - P\,(0) = \underbrace{1 -}_{\text{Complement}} 0.24 = 0.76$

Graphing Binomial Probabilities

Example:

The following probability distribution represents the probability of selecting 0, 1, 2, 3, or 4 red chips when 4 chips are selected. Graph the distribution using a histogram.

x	$P(x)$
0	0.24
1	0.412
2	0.265
3	0.076
4	0.008



Mean, Variance and Standard Deviation

Population Parameters of a Binomial Distribution

Mean: $\mu = np$

Variance: $\sigma^2 = npq$

Standard deviation: $\sigma = \sqrt{npq}$

Example:

One out of 5 students at a local college say that they skip breakfast in the morning. Find the mean, variance and standard deviation if 10 students are randomly selected.

$$n = 10 \qquad p = \frac{1}{5} = 0.2 \qquad q = 0.8$$

$$\mu = np = 10(0.2) = 2$$

$$\sigma^2 = npq = (10)(0.2)(0.8) = 1.6$$

$$\sigma = \sqrt{npq} = \sqrt{1.6} \approx 1.3$$

Geometric Distribution

A **geometric distribution** is a discrete probability distribution of a random variable x that satisfies the following conditions.

1. A trial is repeated until a success occurs.
2. The repeated trials are independent of each other.
3. The probability of a success p is constant for each trial.

The **probability that the first success will occur on trial *x*** is

$$P(x) = p(q)^{x-1}, \text{ where } q = 1 - p.$$

A bag contains six blue balls and four red balls. Balls are randomly drawn from the bag, one at a time, until a red ball is obtained. If we assume that each drawn ball is replaced before the next one is drawn, what is the probability that the experiment stops after exactly five balls have been drawn?

Success with Probability = 4/10=0.4

The **probability that the success will occur on trial *x*** is, here x=5

$$P\,(x) = p(q)^{x-1}, \text{ where } q = 1 - p.$$

$$P[X=5] = p_X(5) = p(1-p)^{5-1} = p(1-p)^4 = (0.4)(0.6)^4 = 0.05184$$

Geometric Distribution

Example:

A fast food chain puts a winning game piece on every fifth package of French fries. Find the probability that you will win a prize,

a.) with your third purchase of French fries,

b.) with your third or fourth purchase of French fries.

$p = 0.20$ $q = 0.80$

a.) $x = 3$

$$P(3) = (0.2)(0.8)^{3-1}$$
$$= (0.2)(0.8)^2$$
$$= (0.2)(0.64)$$
$$= 0.128$$

b.) $x = 3, 4$

$$P(3 \text{ or } 4) = P(3) + P(4)$$
$$\approx 0.128 + 0.102$$
$$\approx 0.230$$

Geometric Distribution

Example:

A fast food chain puts a winning game piece on every fifth package of French fries. Find the probability that you will win a prize,

a.) with your third purchase of French fries,

b.) with your third or fourth purchase of French fries.

$p = 0.20$ $q = 0.80$

a.) $x = 3$

$$P(3) = (0.2)(0.8)^{3-1}$$
$$= (0.2)(0.8)^2$$
$$= (0.2)(0.64)$$
$$= 0.128$$

b.) $x = 3, 4$

$$P(3 \text{ or } 4) = P(3) + P(4)$$
$$\approx 0.128 + 0.102$$
$$\approx 0.230$$

Poisson Distribution

The **Poisson distribution** is a discrete probability distribution of a random variable x that satisfies the following conditions.

1. The experiment consists of counting the number of times an event, x, occurs in a given interval. The interval can be an interval of time, area, or volume.
2. The probability of the event occurring is the same for each interval.
3. The number of occurrences in one interval is independent of the number of occurrences in other intervals.

The probability of exactly *x* occurrences in an interval is

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

where $e \approx 2.71828$ and μ is the mean number of occurrences.

The **probability** of an **event** is the **number** of favorable outcomes divided by the total **number** of outcomes possible.

For example:

Rolling a 3 on a die, the **number of events** is 1 (there's only a single 3 on each die), and the **number** of outcomes is 6.

Message arrive at a switchboard in a Poisson manner at an average rate of six per hour. Find the probability for each of the following events.

a) Exactly two messages arrive within one hour.
b) No message arrives within one hour.
c) At least three messages arrive within one hour.

$a)\ u = 6, x = 2$

$$P(x = 2) = \frac{6^2 e^{-6}}{2!} = 0.0446$$

b) $u = 6, x = 0$

$$P(x = 0) = \frac{6^0 e^{-6}}{0!} = 0.0024$$

$C)\ u = 6, x = 2$

$P(x \geq 3) = 1 - P(x < 3)$

$P(x \geq 3) = 1 - \{P(0) + P(1) + P(2)\}$

$$P(x \geq 3) = 1 - \{\frac{6^0 e^{-6}}{0!} + \frac{6^1 e^{-6}}{1!} + \frac{6^2 e^{-6}}{2!}\}$$

$P(x \geq 3) = 0.9380$

Poisson Distribution

Example:

The mean number of power outages in the city of Brunswick is 4 per year. Find the probability that in a given year,

a.) there are exactly 3 outages,

b.) there are more than 3 outages.

a.) $\mu = 4, \quad x = 3$

$$P(3) = \frac{4^3(2.71828)^{-4}}{3!}$$

$$\approx 0.195$$

b.) $P(\text{more than } 3)$

$$= 1 - P(x \leq 3)$$

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

$$= 1 - (0.195 + 0.147 + 0.073 + 0.018)$$

$$\approx 0.567$$

Two or Multi random variable

- Single random variable defined on a given sample space.
- Two or multi random variable
- Consider two random variable X and Y defined on the same sample space. For example, X denotes the grade of a students and Y denotes the height of the same student.