Chapter 4: Probability

Halima Bensmail

CS502

Monday 9-12pm

Room B013

RANDOMNESS AND PROBABILITY

We call a phenomenon **random** if individual outcomes are uncertain but there is nonetheless a regular distribution of outcomes in a large number of repetitions.

The **probability** of any outcome of a random phenomenon is the proportion of times the outcome would occur in a very long series of repetitions.

In other words:

Probability is a measure of how likely an event is to occur.

Example:
Match one of the probabilities that follow with each statement
Of likelihood given:

Probability: 0 0.01 0.3 0.99 1

(a) This event is impossible. It can never occur.

(b)This event is certain. It will occur on every trial.

(c)This event is very unlikely, but it will occur once in a while in a long sequence
of trials

Probability Models

PROBABILITY MODELS

The **sample space** of a random phenomenon is the set of all possible outcomes. S is used to denote sample space.

An **event** is an outcome or a set of outcomes of a random phenomenon. That is, an event is a subset of the sample space.

A **probability model** is a mathematical description of a random phenomenon consisting of two parts: a sample space S and a way of assigning probabilities to events.

PROBABILITY RULES

Rule 1. The probability $P(A)$ of any event A satisfies $0 \leq P(A) \leq 1$.

Rule 2. If S is the sample space in a probability model, then $P(S) = 1$.

Rule 3. The **complement** of any event A is the event that A does not occur, written as A^c. The **complement rule** states that

$$P(A \text{ does not occur}) = 1 - P(A)$$

Rule 4. Two events A and B are **disjoint** if they have no outcomes in common and so can never occur simultaneously. If A and B are disjoint,

$$P(A \text{ or } B) = P(A) + P(B)$$

This is the **addition rule for disjoint events.**

Random Variables and Probability Distributions

RANDOM VARIABLE

A random variable is a variable whose value is a numerical outcome of a random phenomenon.

Every random variable has some probability distribution associated with it.

PROBABILITY DISTRIBUTION

The **probability distribution** of a random variable X tells us what values X can take and how to assign probabilities to those values.

DISCRETE PROBABILITY DISTRIBUTIONS

The probability distribution of a discrete random variable X lists the possible values of X and their probabilities:

Value of X	x_1	x_2	x_3	$\cdots$	x_k
Probability	p_1	p_2	p_3	$\cdots$	p_k

The probabilities p_i must satisfy two requirements:

1. Every probability p_i is a number between 0 and 1, $0 \le p_i \le 1$.
2. The sum of the probabilities is exactly 1, $p_1 + p_2 + \cdots + p_k = 1$.

To find the probability of any event, add the probabilities p_i of the individual values x_i that make up the event.

Example

Buyers of a laptop computer model may choose to buy either a 10 GB (gigabyte), 20 GB, 30 GB, or 40 GB internal hard drive.

If we want to choose customers from the last 60 days at random to ask what influenced their choice of computer.

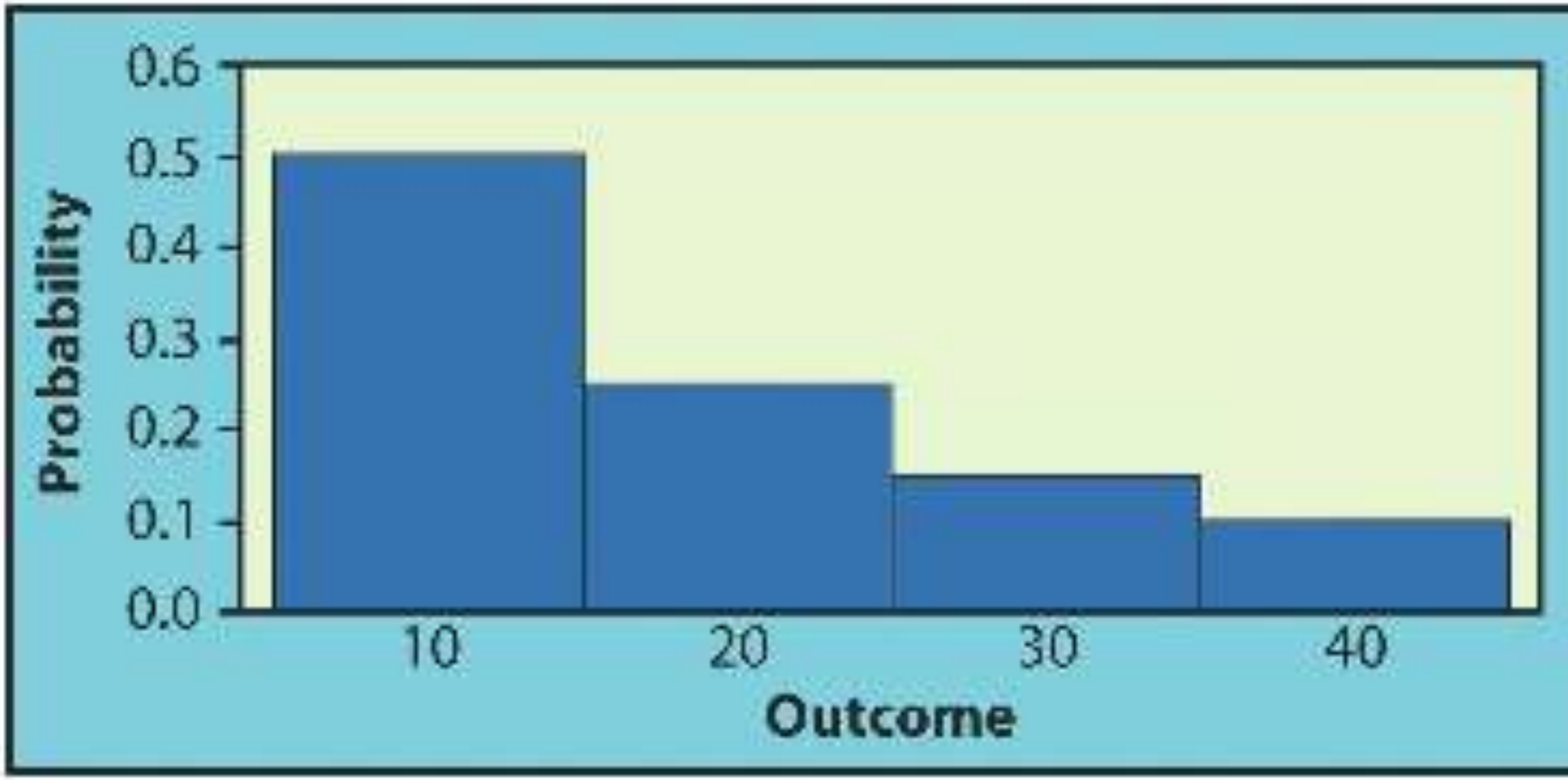
To "choose at random" means to give every customer of the last 60 days the same chance to be chosen.

The size (in gigabytes) of the internal hard drive chosen by a randomly selected customer is a random variable.

Supp

X: Hard drive size	**10**	**20**	**30**	**40**
prob	0.5	0.25	0.15	0.10

Probability distribution for the hard-drive sizes



Example:

All Possible Outcomes for **four tosses of a coin**,

$X = 0$	$X = 1$	$X = 2$	$X = 3$	$X = 4$
		HTTH		
		HTHT		
	HTTT	THTH	HHHT	
	THTT	HHTT	HHTH	
	TTHT	THHT	HTHH	
TTTT	TTTH	TTHH	THHH	HHHH

1- What is the random variable in this example?

2- Describe the distribution of X?

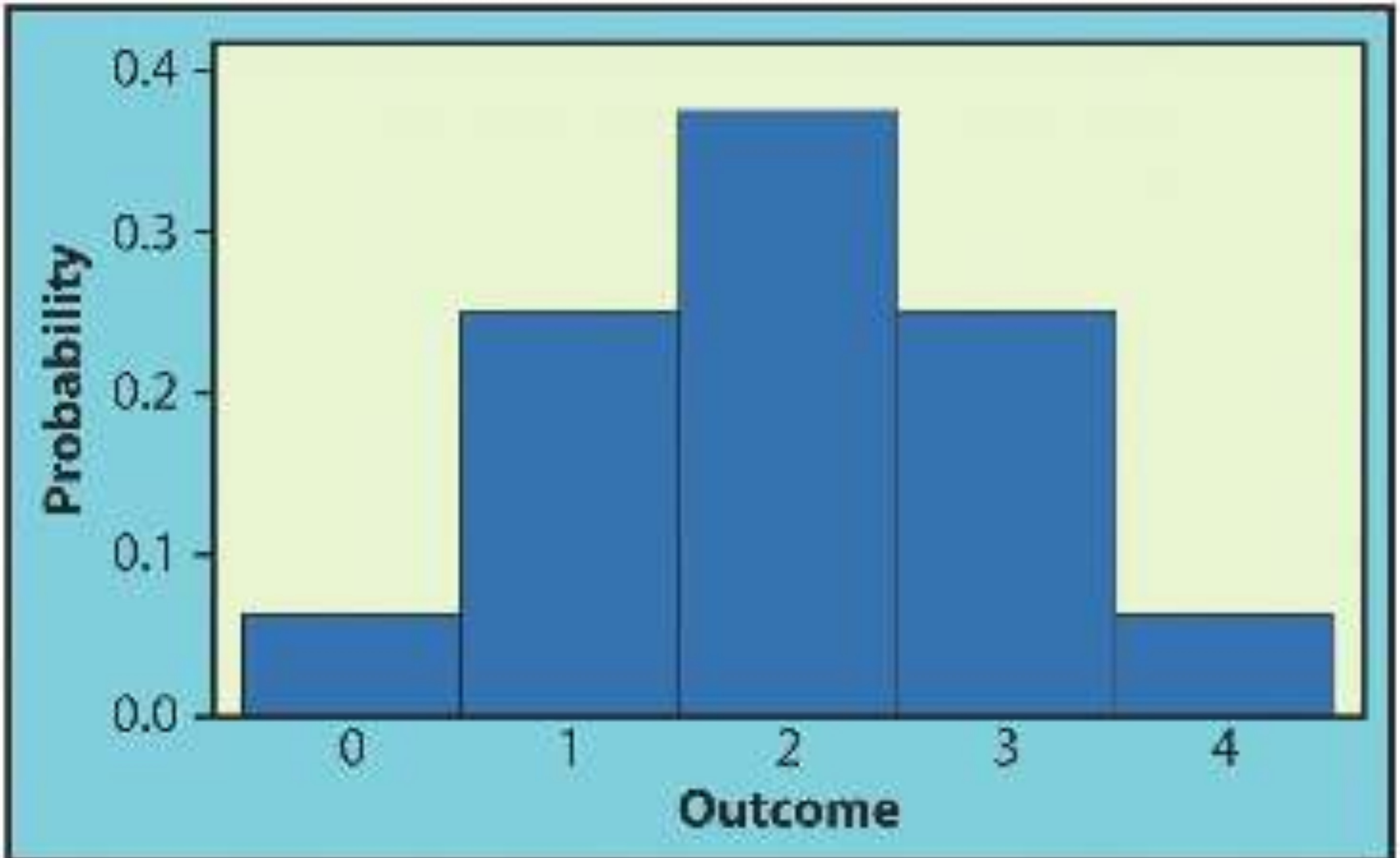
Example:

All Possible Outcomes for four tosses of a coin,

S={All possible outcomes}=Space
{TTTT, TTTH, TTHT, THTT, HTTT, TTHH, THHT, HHTT,THTH,HTHT,HTTH,THHH,HTHH,HHTH, HHHT,HHHH}
Total outcomes 16
P(X=??)=number of possible outcome/total outcomes
P(X=0)=1/16 = 0.0625
P(X=1)=4/16 = 0.25
P(X=2)=6/16 = 0.375
P(X=3)=4/16 = 0.25
P(X=4)=1/16 = 0.0625

X (n of heads Out of 4 tosses)	0	1	2	3	4
prob	1/16	4/16	6/16	4/16	1/16

A visual representation of the probability distribution of the number of heads in four tosses of a coin.



Let's do the following:

Color	Silver	White	Black	Dark green	Dark blue	Medium red
Prob	0.176	0.172	0.113	0.089	0.088	0.067

a- Plot the probability distribution of the color?

b- What is the probability that a car you choose has any color other than the six listed

c- What is the probability that a randomly chosen car has either a silver, white or black?

Example: Roll two dice

- X_1 - number on the first die
- X_2 - number on the second die
- $Y = X_1 + X_2$ - total number of points
  (a function of random variables is again a random variable)

Table of outcomes

Outcome	(X_1, X_2)	Y	Outcome	(X_1, X_2)	Y
⚀ ⚀	(1,1)	2	⚃ ⚀	(4,1)	5
⚀ ⚁	(1,2)	3	⚃ ⚁	(4,2)	6
⚀ ⚂	(1,3)	4	⚃ ⚂	(4,3)	7
⚀ ⚃	(1,4)	5	⚃ ⚃	(4,4)	8
⚀ ⚄	(1,5)	6	⚃ ⚄	(4,5)	9
⚀ ⚅	(1,6)	7	⚃ ⚅	(4,6)	10
⚁ ⚀	(2,1)	3	⚄ ⚀	(5,1)	6
⚁ ⚁	(2,2)	4	⚄ ⚁	(5,2)	7
⚁ ⚂	(2,3)	5	⚄ ⚂	(5,3)	8
⚁ ⚃	(2,4)	6	⚄ ⚃	(5,4)	9
⚁ ⚄	(2,5)	7	⚄ ⚄	(5,5)	10
⚁ ⚅	(2,6)	8	⚄ ⚅	(5,6)	11
⚂ ⚀	(3,1)	4	⚅ ⚀	(6,1)	7
⚂ ⚁	(3,2)	5	⚅ ⚁	(6,2)	8
⚂ ⚂	(3,3)	6	⚅ ⚂	(6,3)	9
⚂ ⚃	(3,4)	7	⚅ ⚃	(6,4)	10
⚂ ⚄	(3,5)	8	⚅ ⚄	(6,5)	11
⚂ ⚅	(3,6)	9	⚅ ⚅	(6,6)	12

Example: Roll two dice

$Y = X_1 + X_2$ total number of points

y	2	3	4	5	6	7	8	9	10	11	12
P(Y=y)											

MEAN OF A DISCRETE RANDOM VARIABLE

Suppose that X is a discrete random variable whose distribution is

Value of X	x_1	x_2	x_3	$\cdots$	x_k
Probability	p_1	p_2	p_3	$\cdots$	p_k

To find the mean of X, multiply each possible value by its probability, then add all the products:

$$\mu_X = x_1p_1 + x_2p_2 + \cdots + x_kp_k$$
$$= \sum x_ip_i$$

Expected value does not mean the value that you expect to see, it is what you would expect to be the average value of the outcomes on a large number of population.

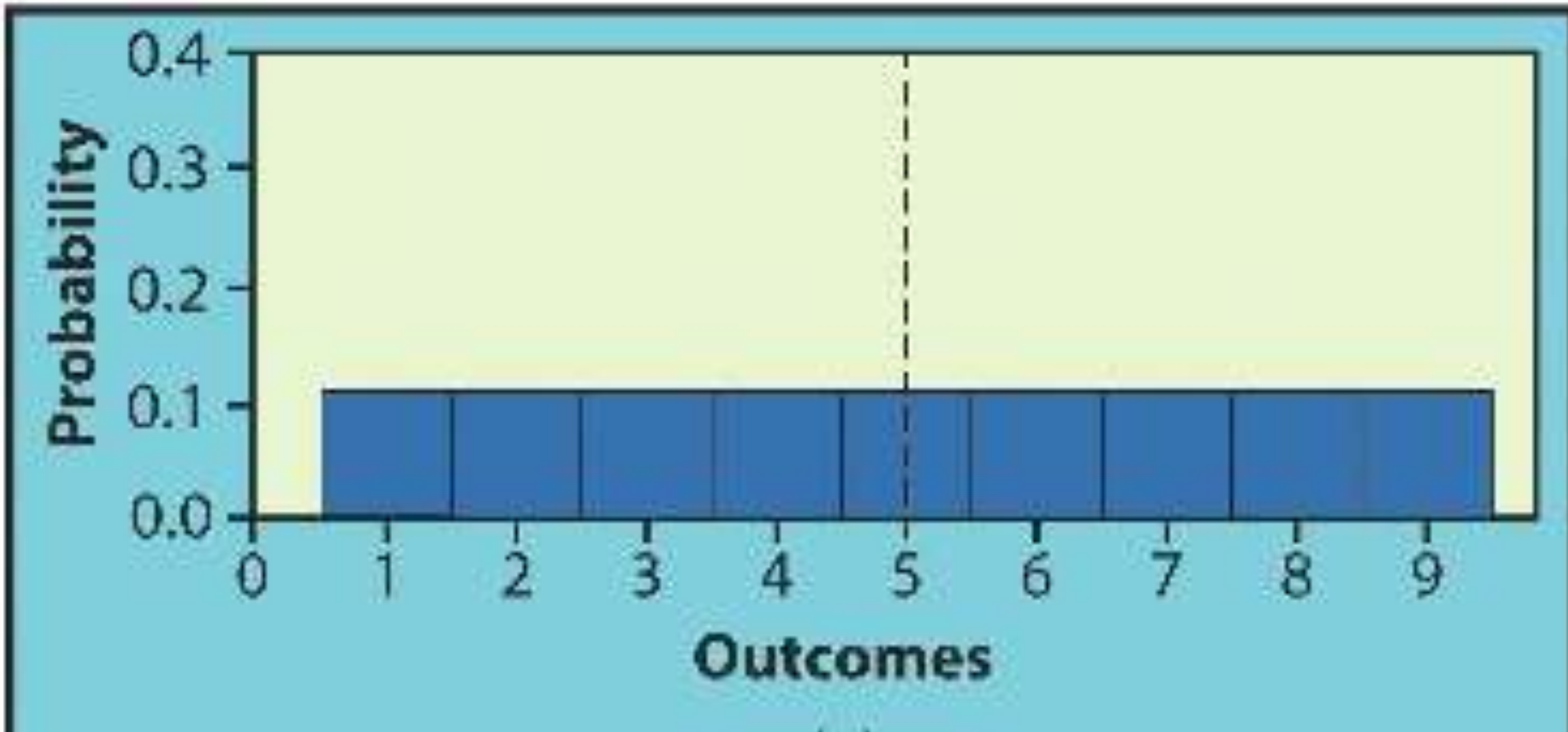
It is useful for calculating the total when you have a large number of observations.

Say you own a company with a large number (N) of employees.

The expected salary is E[X]

The total salary that you have to pay out is N E[X]

Find the mean of this probability distribution,



(a)

X						
Prob						

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

Example: Roll two dice

$Y = X_1 + X_2$ total number of points

y	2	3	4	5	6	7	8	9	10	11	12
P(Y=y)											

E(X) =

Example: Roll one die

Let X be outcome of rolling one die. The frequency function is

$$p(x) = \frac{1}{6}, \qquad x = 1, \ldots, 6,$$

and hence

$$\mathbb{E}(X) = \sum_{x=1}^{6} \frac{x}{6} = \frac{7}{2} = 3.5$$

Example

The following probability describes the number of persons living in households in America. Given the following probability distribution, find its mean, variance and standard deviation.

Houschold size X	1	2	3	4	5	6	7
Probability	0.25	0.32	0.17	0.15	0.07	0.03	0.01

Linearity of the expected value

RULES FOR MEANS

Rule 1. If X is a random variable and a and b are fixed numbers, then

$$\mu_{a+bX} = a + b\mu_X$$

Rule 2. If X and Y are random variables, then

$$\mu_{X+Y} = \mu_X + \mu_Y$$

This is the addition rule for means.

PROOF:

$$\begin{aligned}
\mathbb{E}(a\,X + b\,Y) &= \sum_{x,y} (a\,x + b\,y) p(x,y) \\
&= a \sum_{x,y} x\, p(x,y) + b \sum_{x,y} y\, p(x,y) \\
\sum_x p(x,y) = p(y) \quad &= a \sum_{x} x\, p(x) + b \sum_{y} y\, p(y) \\
&= a\mathbb{E}(X) + b\mathbb{E}(Y)
\end{aligned}$$

VARIANCE OF A DISCRETE RANDOM VARIABLE

Suppose that X is a discrete random variable whose distribution is

Value of X	x_1	x_2	x_3	$\cdots$	x_k
Probability	p_1	p_2	p_3	$\cdots$	p_k

and that μ is the mean of X. The variance of X

$$\sigma_X^2 = (x_1 - \mu_X)^2 p_1 + (x_2 - \mu_X)^2 p_2 + \cdots + (x_k - \mu_X)^2 p_k$$
$$= \sum (x_i - \mu_X)^2 p_i$$

The standard deviation σ_X of X is the square root of the variance.

Example: Roll one die

X takes values in $\{1, 2, 3, 4, 5, 6\}$ with frequency function $p(x) = \frac{1}{6}$.

$$\mathbb{E}(X) = \sum_{x=1}^{6} x\,\frac{1}{6} = \frac{7}{2}$$

$$\text{var}(X) = \sum_{x=1}^{6} \left(x - \frac{7}{2}\right)^2 \frac{1}{6} = \frac{1}{6}\left(\frac{25}{4} + \frac{9}{4} + \frac{1}{4} + \frac{1}{4} + \frac{9}{4} + \frac{25}{4}\right) = \frac{35}{12}$$

Answer:

$$\mu = \sum x\, p(x) = 1(.25) + 2\,(.32) + 3\,(.17) + 4\,(.15) + 5\,(.07) + 6\,(.03) + 7\,(.01) = 2.6$$

$$\sigma^2 = \sum (x - \mu)^2\, p(x) = (1 - 2.6)^2\,(.25) + \ldots = 2.02$$

$$\sigma = \text{sqrt}(\sigma^2) = \text{sqrt}(2.02) = 1.42$$

Household size X	1	2	3	4	5	6	7
Probability	0.25	0.32	0.17	0.15	0.07	0.03	0.01

Properties of the Variance

RULES FOR VARIANCES

Rule 1. If X is a random variable and a and b are fixed numbers, then

$$\sigma^2_{a+bX} = b^2\sigma^2_X$$

Rule 2. If X and Y are independent random variables, then

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$$
$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$$

This is the **addition rule for variances of independent random variables.**

Rule 3. If X and Y have correlation ρ, then

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\rho\sigma_X\sigma_Y$$
$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y - 2\rho\sigma_X\sigma_Y$$

This is the **general addition rule for variances of random variables.**

Covariance

For *independent* random variables X and Y we have

$$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y).$$

Question: What about *dependent* random variables?

It can be shown that

$$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y) + 2\,\text{cov}(X,Y)$$

where

$$\text{cov}(X,Y) = \mathbb{E}\big[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y)\big]$$

is the **covariance** of X and Y.

Properties of the covariance

- $\operatorname{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y)$
- $\operatorname{cov}(X, X) = \operatorname{var}(X)$
- $\operatorname{cov}(X, 1) = 0$
- $\operatorname{cov}(X, Y) = \operatorname{cov}(Y, X)$
- $\operatorname{cov}(a\,X_1 + b\,X_2, Y) = a \operatorname{cov}(X_1, Y) + b \operatorname{cov}(X_2, Y)$

Covariance

Important:

$\text{cov}(X, Y) = 0$ does NOT imply that X and Y are independent.

Example:

Suppose $X \in \{-1, 0, 1\}$ with probabilities $\mathbb{P}(X = x) = \frac{1}{3}$ for $x = -1, 0, 1$. Then $\mathbb{E}(X) = 0$ and

$$\text{cov}(X, X^2) = \mathbb{E}(X^3) = \mathbb{E}(X) = 0$$

Answer:

$\mu = \sum x\, p(x) = 1(.25) + 2\,(.32) + 3\,(.17) + 4\,(.15) + 5\,(.07) +$

$6\,(.03) + 7\,(.01) = 2.6$

$\sigma^2 = \sum (x - \mu)^2\, p(x) = (1 - 2.6)^2\,(.25) + \ldots = 2.02$

$\sigma = \text{sqrt}(\sigma^2) = \text{sqrt}(2.02) = 1.42$

Household size X	1	2	3	4	5	6	7
Probability	0.25	0.32	0.17	0.15	0.07	0.03	0.01

Multivariate Distributions: Discrete Case

Discrete Case

Let X and Y be discrete random variables.

Joint frequency function of X and Y

$$p_{XY}(x, y) = \mathbb{P}(X = x, Y = y) = \mathbb{P}(\{X = x\} \cap \{Y = y\})$$

Marginal frequency function of X

$$p_X(x) = \sum_i p_{XY}(x, y_i)$$

Marginal frequency function of Y

$$p_Y(y) = \sum_i p_{XY}(x_i, y)$$

Conditional probability of $X = x$ given $Y = y$

$$\mathbb{P}(X = x | Y = y) = p_{X|Y}(x|y) = \frac{p_{XY}(x, y)}{p_Y(y)} = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}$$

Example: Three Toss of a coin

Suppose we have two random variables X and Y:

- X - number of heads on the first toss (values in $\{0, 1\}$)
- Y - total number of heads (values in $\{0, 1, 2, 3\}$)

The joint frequency function $p_{XY}(x, y)$ is given by the following table

Let's fill in this table:

X/Y	0	1	2	3	
0					
1					

$x\backslash y$	0	1	2	3	
0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0	$\frac{1}{2}$
1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{1}{2}$
	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

Marginal frequency function of Y:

P(Y=0)= P(Y=0, X=0) + P(Y=0, X=1)

=

P(Y=1)= P(Y=1, X=0) + P(Y=1, X=1)

…….

<u>Statistical inference</u> is concerned with **patterns that occur in a group of individuals (population)** rather than the individuals themselves. It is very important to note that a characteristic of the population as a whole is not necessarily a characteristic of each individual.

In what case is the previous statement false ?

Ex.

- The average height of a basketball team is more than 6 feet.
  - What can we say about each player's height?
- The trend of the market is downward
  - What we say about the behavior of the market tomorrow?

Similarly, when we say that the trend of the market is downward, we are really saying that the **probability** is higher that tomorrow the market will be down than the **probability** that it will be up."

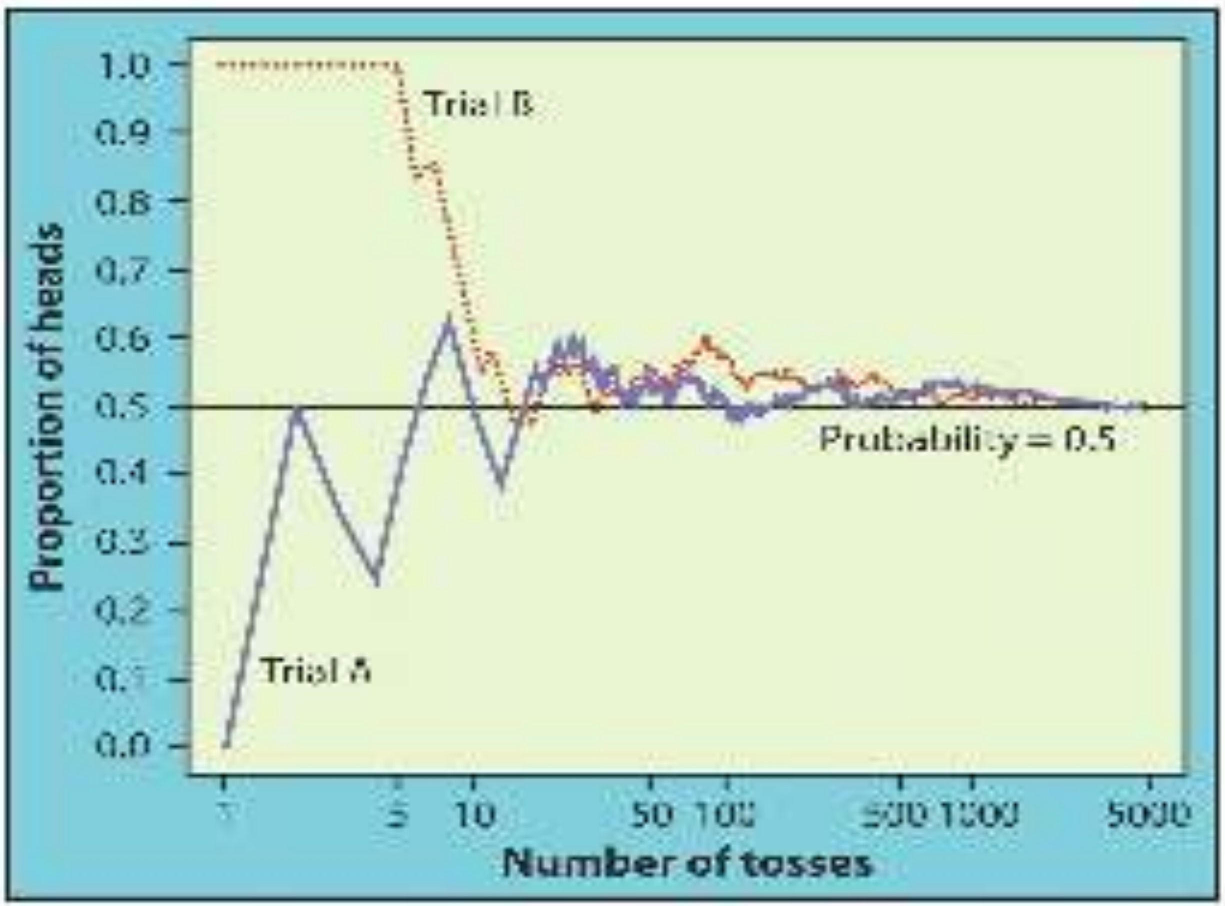
- we are saying that, over time, the expected value is decreasing.

Q: If you toss a coin, what are the chances (probability) that it will be heads?

A:

What the answer means is that if the coin is tossed a large number of times, ***about*** 50% of the time the result will be heads.

A computer simulation showing that the proportion of heads approaches .5 as the number of tosses increases
(Law of large numbers)



The **Law of Large Numbers** for a Fair Coin Toss Implies…

- The proportion of heads will near ½ as the number of tosses increases
- Do the following data contradict this law?

No. Tosses	**100**	**10,000**	**1,000,000**
No. Heads	55	4,900	500,500
Expected No. of Heads	50	5,000	500,000
Difference	+5	-50	+500

- Often times, we are interested in the mean (average) of a population. Such a quantity is called a **parameter** since it pertains to the whole population.

- It is often impractical to calculate a mean (or any thing else) based on the whole population. Thus, we calculate a mean from a sample. Quantities computed from a sample are called **statistics.** The hope that the value of the statistic is close to that of the parameter (but we will never know because the value of the parameter is unknown).

Law of the Large number

The law of large numbers connects the sample size with the closeness of the sample mean to the population mean.

Theorem

LAW OF LARGE NUMBERS

Draw independent observations at random from any population with finite mean μ. As the number of observations drawn increases, the mean $\bar{x}$ of the observed values gets closer and closer to the mean μ of the population.

Suppose that from a population of size **N** we take **k** SRS's of size **n** (in practice we can afford only one sample). The mean is then computed for each sample, yielding **k** means.

The distribution (histogram) of these **k** means is called the **sampling distribution of the sample mean.**

The following results give information about the sampling distribution of the sample mean.

Important Theorem

MEAN AND STANDARD DEVIATION OF A SAMPLE MEAN

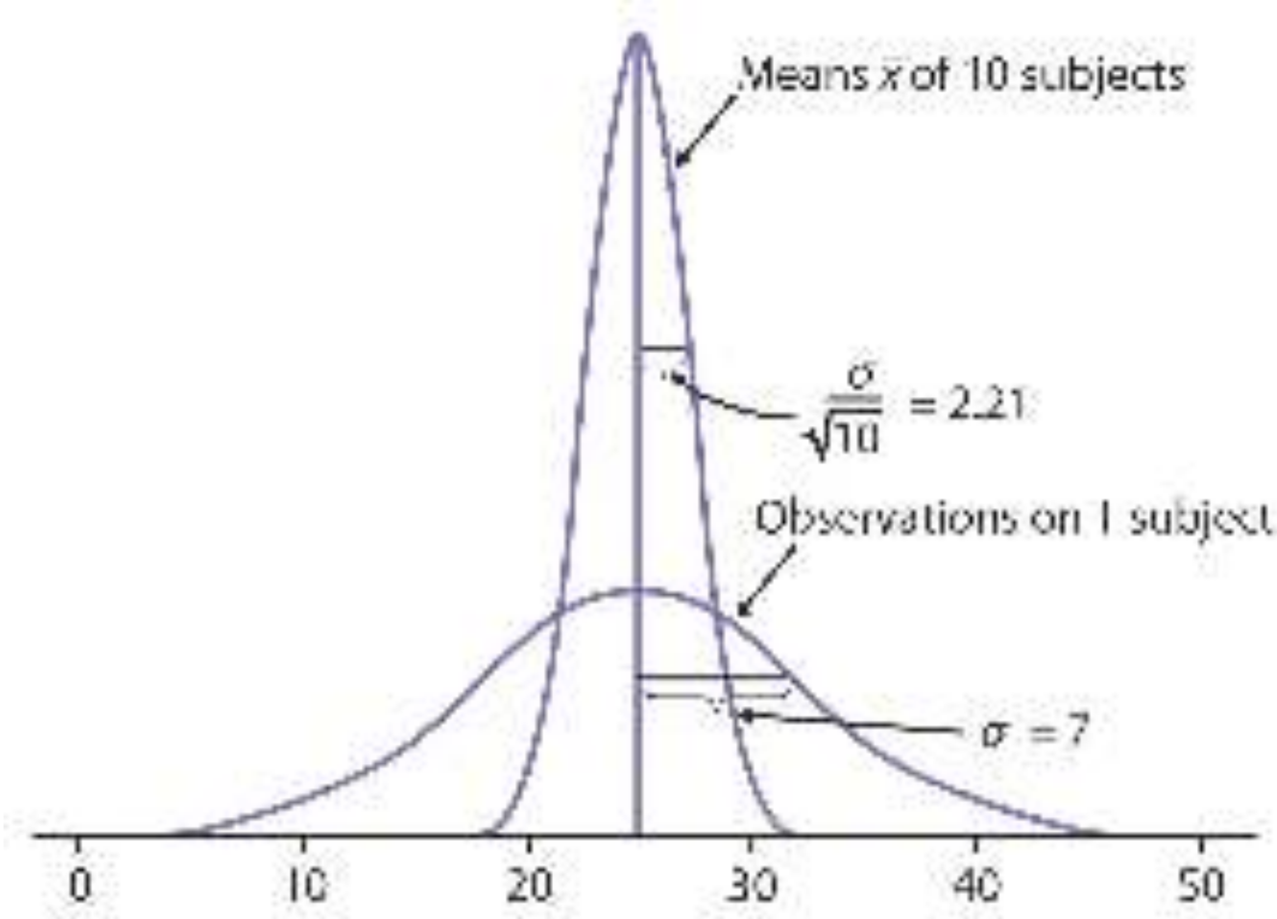
Suppose that $\bar{x}$ is the mean of an SRS of size n drawn from a large population with mean μ and standard deviation σ. Then the mean of the sampling distribution of $\bar{x}$ is μ and its standard deviation is $\sigma/\sqrt{n}$.

If the population distribution is **normal**, then the **sampling distribution of the mean is also normal** regardless of the sample size

SAMPLING DISTRIBUTION OF A SAMPLE MEAN

If a population has the $N(\mu, \sigma)$ distribution, then the sample mean $\bar{x}$ of n independent observations has the $N(\mu, \sigma / \sqrt{n})$ distribution.

The variation of the mean is reduced as the sample size increases



- The previous result says that if the population has a normal distribution, then the sampling distribution of the mean is also normal. This is really not that surprising.

- The following amazing result, called the **Central Limit Theorem (CLT),** states that, <u>regardless</u> of the distribution of the population, the sampling distribution of the mean is normal when **n is large**.

Theorem

CENTRAL LIMIT THEOREM

Draw an SRS of size n from any population with mean μ and finite standard deviation σ. When n is large, the sampling distribution of the sample mean $\bar{x}$ is approximately Normal:

$$\bar{x} \text{ is approximately } N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Sample distribution of the mean
With sample sizes of 1 (a), 2 (b), 10 (c) and 25(4)