Probability

G.P.M.Sandaruwani
Consultant/Lecturer
BSc(Hons) in Management and Information Technology

What is an experiment

An experiment is a process that, when performed, results in exactly one of many observations. These observations are called the outcomes of the experiment. The set of all possible outcomes for an experiment is called a sample space, and is denoted by S. An event is a subset of the sample space.

ex. Suppose we roll a die once, and we get a 4.

Experiment: Rolling the die

Outcome: 4

Sample Space (S) = {1,2,3,4,5,6}

Example events in the above experiment:

1. Getting an even number
2. Getting a number which is greater than 3

ex. Suppose we toss a coin twice, and we get a head (H) on the first toss and a tail (T) on the second toss.

Experiment: Tossing a coin twice

Outcome: HT

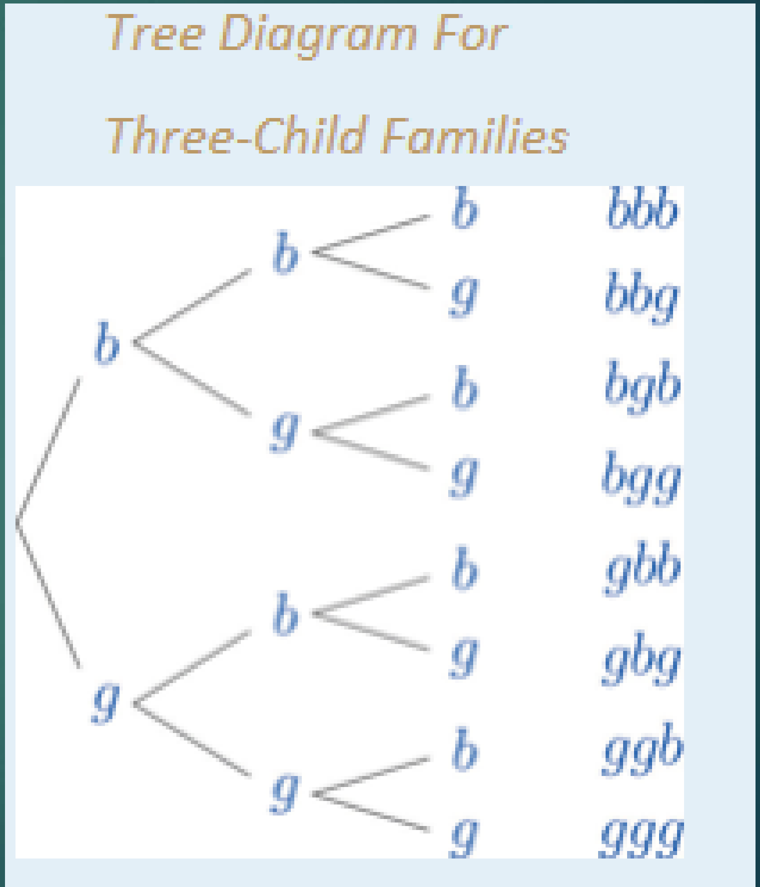
Sample Space (S) = {HH,HT,TH,TT}

The possible outcomes for the event getting the same side for the both attempts = {HH,TT}

Using tree diagrams to represent the sample space

Construct a sample space and a tree diagram that describes all three-child families according to the genders of the children with respect to birth order.

$$S = \{bbb, bbg, bgb, bgg, gbb, gbg, ggb, ggg\}$$



What is a probability

The probability of an outcome E in a sample space S is a number p between 0 and 1 that measures the likelihood that e will occur on a single trial of the corresponding random experiment. The value p = 0 corresponds to the outcome e being impossible and the value p = 1 corresponds to the outcome e being certain.

The probability of an event A is the sum of the probabilities of the individual outcomes of which it is composed. It is denoted P(A).

If an event E is $E = \{e_1, e_2, \ldots, e_k\}$, then

$$P(E) = P(e_1) + P(e_2) + \cdots + P(e_k)$$

What is a null event

A null event is an event that is impossible. Or more precisely, since an event is a subset of a sample space, the null event is the empty set. So $P(\emptyset) = 0$

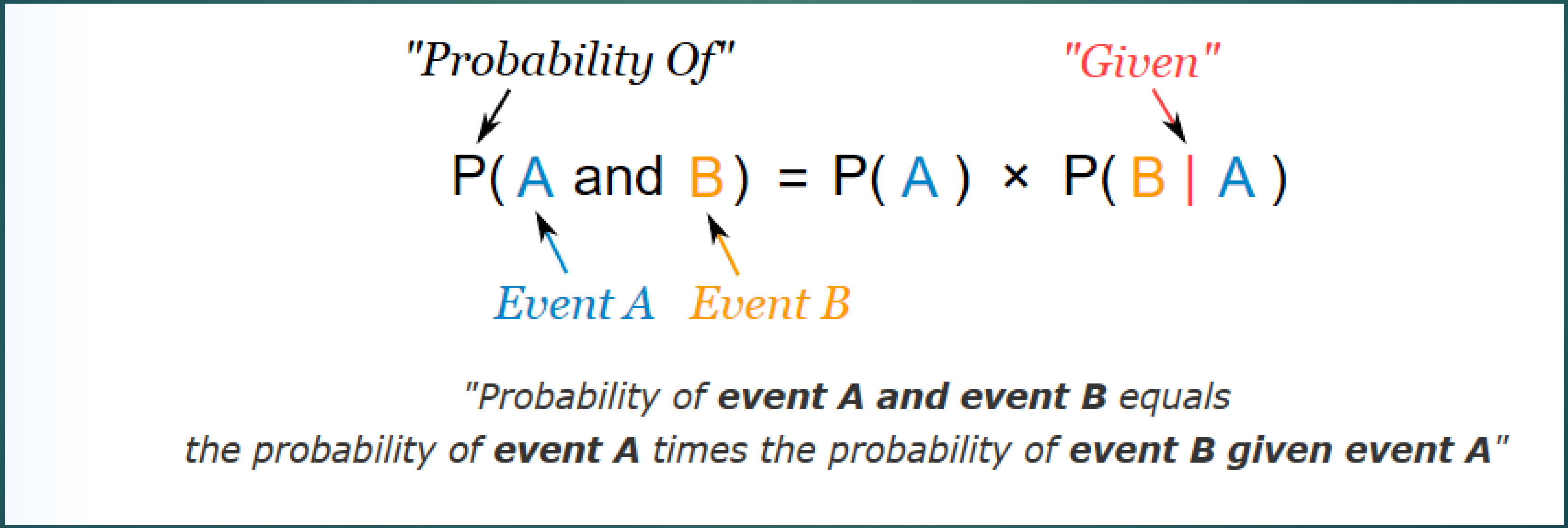
Simple & Compound Events

Simple events are those events where only a single experiment is carried out. Tossing of a coin or rolling a dice are known as simple events.

Compound events are those events which have more than one experiments occurring together. For example, rolling a dice and tossing a coin together will be known as a compound event. The sample space of compound events is obtained using lists, table and tree diagrams.

Conditional Probability

In probability theory, conditional probability is a measure of the probability of an event (some particular situation occurring) given that another event has occurred. If the event of interest is A and the event B is known or assumed to have occurred, "the conditional probability of A given B", or "the probability of A under the condition B", is usually written as P(A | B), or P(A / B).



Example: Drawing 2 Kings from a Deck

Event A is drawing a King first, and **Event B** is drawing a King second.

For the first card the chance of drawing a King is 4 out of 52 (there are 4 Kings in a deck of 52 cards):

$$P(A) = 4/52$$

But after removing a King from the deck the probability of the 2nd card drawn is **less** likely to be a King (only 3 of the 51 cards left are Kings):

$$P(B|A) = 3/51$$

And so:

$$\mathbf{P(A\ and\ B) = P(A) \times P(B|A)} = (4/52) \times (3/51) = 12/2652 = \mathbf{1/221}$$

So the chance of getting 2 Kings is 1 in 221, or about 0.5%

Big Example: Soccer Game

You are off to soccer, and want to be the Goalkeeper, but that depends who is the Coach today:

- with Coach Sam the probability of being Goalkeeper is **0.5**
- with Coach Alex the probability of being Goalkeeper is **0.3**

Sam is Coach more often ... about 6 out of every 10 games (a probability of **0.6**).

So, what is the probability you will be a Goalkeeper today?

Bayer's Theorem

Bayes' Theorem is a way of finding a probability when we know certain other probabilities.

The formula is:

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

What is the chance of rain during the day?

We will use Rain to mean rain during the day, and Cloud to mean cloudy morning.

The chance of Rain given Cloud is written P(Rain|Cloud)

So let's put that in the formula:

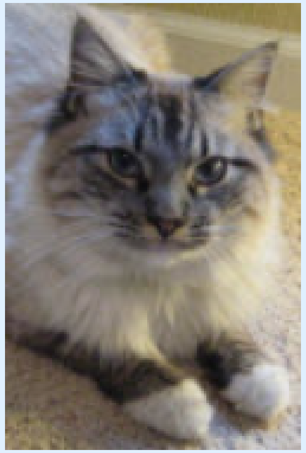
$$P(\text{Rain}|\text{Cloud}) = \frac{P(\text{Rain})\,P(\text{Cloud}|\text{Rain})}{P(\text{Cloud})}$$

- P(Rain) is Probability of Rain = 10%
- P(Cloud|Rain) is Probability of Cloud, given that Rain happens = 50%
- P(Cloud) is Probability of Cloud = 40%

$$P(\text{Rain}|\text{Cloud}) = \frac{0.1 \times 0.5}{0.4} = .125$$

Or a 12.5% chance of rain. Not too bad, let's have a picnic!

Example: Allergy or Not?



Hunter says she is itchy. There is a test for Allergy to Cats, but this test is not always right:

- For people that **really do** have the allergy, the test says "Yes" **80%** of the time
- For people that **do not** have the allergy, the test says "Yes" **10%** of the time ("false positive")

If 1% of the population have the allergy, and **Hunter's test says "Yes"**, what are the chances that Hunter really has the allergy?

In Exton School, 60% of the boys play football and 36% of the boys play ice hockey.
Given that 40% of those that play football also play ice hockey, what percent of those that play ice hockey also play football?

In Exton School, 40% of the girls like music and 24% of the girls like dance.
Given that 30% of those that like music also like dance, what percent of those that like dance also like music?

75% of the children in Exton school have a dog, and 30% have a cat.
Given that 60% of those that have cat also have a dog, what percent of those that have a dog also have a cat?

35% of the children in Exton school have a tablet, and 24% have a smart phone.
Given that 42% of those that have smart phone also have a tablet, what percent of those that have a tablet also have a smart phone?

Dr. Foster remembers to take his umbrella with him 80% of the days.
It rains on 30% of the days when he remembers to take his umbrella, and it rains on 60% of the days when he forgets to take his umbrella.

What is the probability that he remembers his umbrella when it rains?

In a factory, machine X produces 60% of the daily output and machine Y produces 40% of the daily output.

2% of machine X's output is defective, and 1.5% of machine Y's output is defective.

One day, an item was inspected at random and found to be defective. What is the probability that it was produced by machine X?

A test for a disease gives a correct positive result with a probability of 0.95 when the disease is present, but gives an incorrect positive result (false positive) with a probability of 0.15 when the disease is not present.

If 5% of the population has the disease, and Jean tests positive to the test, what is the probability Jean really has the disease?

Wire manufactured by a company is tested for strength.
The test gives a correct positive result with a probability of 0.85 when the wire is strong, but gives an incorrect positive result (false positive) with a probability of 0.04 when in fact the wire is not strong.

If 98% of the wires are strong, and a wire chosen at random fails the test, what is the probability it really is not strong enough?

A supermarket buys light globes (light bulbs) from three different manufacturers - Brightlight (35%), Glowglobe (20%) and Shinewell (45%).

In the past, the supermarket has found that 1% of Brightlight's globes are faulty, and that 1.5% of each of Glowglobe's and Shinewell's globes are faulty.

A customer buys a globe without looking at the manufacturer's name - in other words, it's a random choice. When she gets home, she finds the globe is faulty.

What is the probability she chose a Shinewell's globe?

A glazier buys his glass from four different manufacturers - Clearglass (10%), Strongpane (25%), Mirrorglass (30%) and Reflection (35%).

In the past, the glazier has found that 1% of Clearglass' product is cracked, 1.5% of Strongpane's product is cracked, and 2% of Mirrorglass' and Reflection's products are cracked.

The glazier removes the protective covering from a sheet of glass without looking at the manufacturer's name - in other words, it's a random choice. He finds the glass is cracked. What is the probability it was made by Mirrorglass?

Independent & Dependent Events

Independent events are those events where the occurrence of one event will not affect the probability of occurrence of the other event. If A and B are two independent events then the probability of both A and B will be written as P(A and B)=P(A). For example, if we toss a coin twice then the outcomes are independent of each other.

Dependent events are those where occurrence of one event will affect the other event. For example, if there are 3 red and 2 green balls in a bag and one green ball has been taken out then the probability of getting a green ball in next attempt will get affected.

Mutually Exclusive Events

Two events which cannot occur together are known as mutually exclusive events. For example, a bulb cannot be on and off at the same time. So, these events are mutually exclusive. If A and B are two mutually exclusive events then the probability of A or B happening is written as, P(A or B)=P(A)+P(B).

Exhaustive Events

Events which together exhaust the whole sample space are known as exhaustive events. For example, when we roll a die one event is to get all odd numbers and other is to get all even numbers. Both the events together will exhaust the whole sample space.

Sample space = {1,2,3,4,5,6}
Odd = {1, 3, 5}
Even = {2, 4, 6}

The union of odd and even events will add up to the sample space.

Probability Rules

There are three main rules associated with basic probability: the addition rule, the multiplication rule, and the complement rule.

1.) The Addition Rule: P(A or B) = P(A) + P(B) - P(A and B)

If A and B are mutually exclusive events, or those that cannot occur together, then the third term is 0, and the rule reduces to P(A or B) = P(A) + P(B). For example, you can't flip a coin and have it come up both heads and tails on one toss.

2.) The Multiplication Rule: P(A and B) = P(A) * P(B|A) or P(B) * P(A|B)

If A and B are independent events, we can reduce the formula to P(A and B) = P(A) * P(B). The term independent refers to any event whose outcome is not affected by the outcome of another event. For instance, consider the second of two coin flips, which still has a .50 (50%) probability of landing heads, regardless of what came up on the first flip. What is the probability that, during the two coin flips, you come up with tails on the first flip and heads on the second flip?

Let's perform the calculations: P = P(tails) * P(heads) = (0.5) * (0.5) = 0.25

3.) The Complement Rule: P(not A) = 1 - P(A)

Do you see why the complement rule can also be thought of as the subtraction rule? This rule builds upon the mutually exclusive nature of P(A) and P(not A). These two events can never occur together, but one of them always has to occur. Therefore P(A) + P(not A) = 1. For example, if the weatherman says there is a 0.3 chance of rain tomorrow, what are the chances of no rain?

Let's do the math: P(no rain) = 1 - P(rain) = 1 - 0.3 = 0.7

Ex: The breakdown of the student body in a local high school according to race and ethnicity is 51% white, 27% black, 11% Hispanic, 6% Asian, and 5% for all others. A student is randomly selected from this high school. (To select “randomly” means that every student has the same chance of being selected.) Find the probabilities of the following events:

1. B: the student is black,
2. M: the student is not white,
3. N: the student is not black.

The student body in the high school considered may be broken down into ten categories as follows: 25% white male, 26% white female, 12% black male, 15% black female, 6% Hispanic male, 5% Hispanic female, 3% Asian male, 3% Asian female, 1% male of other minorities combined, and 4% female of other minorities combined. A student is randomly selected from this high school. Find the probabilities of the following events:

1. B: the student is a black,
2. F: the student is a female,
3. A: The student is asian

1. A box contains 10 white and 10 black marbles. Construct a sample space for the experiment of randomly drawing out, with replacement, two marbles in succession and noting the color each time. (To draw "with replacement" means that the first marble is put back before the second marble is drawn.)

2. A box contains 16 white and 16 black marbles. Construct a sample space for the experiment of randomly drawing out, with replacement, three marbles in succession and noting the color each time. (To draw "with replacement" means that each marble is put back before the next marble is drawn.)

3. A box contains 8 red, 8 yellow, and 8 green marbles. Construct a sample space for the experiment of randomly drawing out, with replacement, two marbles in succession and noting the color each time.

4. A box contains 6 red, 6 yellow, and 6 green marbles. Construct a sample space for the experiment of randomly drawing out, with replacement, three marbles in succession and noting the color each time.

5. In the situation of Exercise 1, list the outcomes that comprise each of the following events.
   a. At least one marble of each color is drawn.
   b. No white marble is drawn.
6. In the situation of Exercise 2, list the outcomes that comprise each of the following events.
   a. At least one marble of each color is drawn.
   b. No white marble is drawn.
   c. More black than white marbles are drawn.
7. In the situation of Exercise 3, list the outcomes that comprise each of the following events.
   a. No yellow marble is drawn.
   b. The two marbles drawn have the same color.
   c. At least one marble of each color is drawn.

8. In the situation of Exercise 4, list the outcomes that comprise each of the following events.

    a. No yellow marble is drawn.
    b. The three marbles drawn have the same color.
    c. At least one marble of each color is drawn.

9. Assuming that each outcome is equally likely, find the probability of each event in Exercise 5.

10. Assuming that each outcome is equally likely, find the probability of each event in Exercise 6.

11. Assuming that each outcome is equally likely, find the probability of each event in Exercise 7.

12. Assuming that each outcome is equally likely, find the probability of each event in Exercise 8.

13. A sample space is $S = \{a, b, c, d, e\}$. Identify two events as $U = \{a, b, d\}$ and $V = \{b, c, d\}$. Suppose $P(a)$ and $P(b)$ are each 0.2 and $P(c)$ and $P(d)$ are each 0.1.

    a. Determine what $P(e)$ must be.
    b. Find $P(U)$.
    c. Find $P(V)$.

14. A sample space is $S = \{u, v, w, x\}$. Identify two events as $A = \{v, w\}$ and $B = \{u, w, x\}$. Suppose $P(u) = 0.22$, $P(w) = 0.36$, and $P(x) = 0.27$.

    a. Determine what $P(v)$ must be.
    b. Find $P(A)$.
    c. Find $P(B)$.

15. A sample space is $S = \{m, n, q, r, s\}$. Identify two events as $U = \{m, q, s\}$ and $V = \{n, q, r\}$. The probabilities of some of the outcomes are given by the following table:

Outcome	m	n	q	r	s
Probablity	0.18	0.16		0.24	0.21

a. Determine what $P(q)$ must be.

b. Find $P(U)$.

c. Find $P(V)$.

16. A sample space is $S = \{d, e, f, g, h\}$. Identify two events as $M = \{e, f, g, h\}$ and $N = \{d, g\}$. The probabilities of some of the outcomes are given by the following table:

Outcome	d	e	f	g	h
Probablity	0.22	0.13	0.27		0.19

17. The sample space that describes all three-child families according to the genders of the children with respect to birth order was constructed in Note 3.9 "Example 4". Identify the outcomes that comprise each of the following events in the experiment of selecting a three-child family at random.

    a. At least one child is a girl.
    b. At most one child is a girl.
    c. All of the children are girls.
    d. Exactly two of the children are girls.
    e. The first born is a girl.

18. The sample space that describes three tosses of a coin is the same as the one constructed in Note 3.9 "Example 4" with “boy” replaced by “heads” and “girl” replaced by “tails.” Identify the outcomes that comprise each of the following events in the experiment of tossing a coin three times.

    a. The coin lands heads more often than tails.
    b. The coin lands heads the same number of times as it lands tails.
    c. The coin lands heads at least twice.
    d. The coin lands heads on the last toss.

21. The following two-way contingency table gives the breakdown of the population in a particular locale according to age and tobacco usage:

Age	Tobacco Use	
	Smoker	Non-smoker
Under 30	0.05	0.20
Over 30	0.20	0.55

A person is selected at random. Find the probability of each of the following events.

a. The person is a smoker.

b. The person is under 30.

c. The person is a smoker who is under 30.

22. The following two-way contingency table gives the breakdown of the population in a particular locale according to party affiliation (*A*, *B*, *C*, or *None*) and opinion on a bond issue:

Affiliation	Opinion		
	Favors	Opposes	Undecided
A	0.12	0.09	0.07
B	0.16	0.12	0.14
C	0.04	0.03	0.06
None	0.08	0.06	0.03

A person is selected at random. Find the probability of each of the following events.

a. The person is affiliated with party *B*.

b. The person is affiliated with some party.

c. The person is in favor of the bond issue.

d. The person has no party affiliation and is undecided about the bond issue.

23. The following two-way contingency table gives the breakdown of the population of married or previously married women beyond child-bearing age in a particular locale according to age at first marriage and number of children:

Age	Number of Children		
	0	1 or 2	3 or More
Under 20	0.02	0.14	0.08
20–29	0.07	0.37	0.11
30 and above	0.10	0.10	0.01

A woman is selected at random. Find the probability of each of the following events.

a. The woman was in her twenties at her first marriage.

b. The woman was 20 or older at her first marriage.

c. The woman had no children.

d. The woman was in her twenties at her first marriage and had at least three children.

24. The following two-way contingency table gives the breakdown of the population of adults in a particular locale according to highest level of education and whether or not the individual regularly takes dietary supplements:

Education	Use of Supplements	
	Takes	Does Not Take
No High School Diploma	0.04	0.06
High School Diploma	0.06	0.44
Undergraduate Degree	0.09	0.28
Graduate Degree	0.01	0.02

An adult is selected at random. Find the probability of each of the following events.

a. The person has a high school diploma and takes dietary supplements regularly.
b. The person has an undergraduate degree and takes dietary supplements regularly.
c. The person takes dietary supplements regularly.
d. The person does not take dietary supplements regularly.

Expected Value of a Random Variable

Let X represent a discrete random variable with the probability distribution function P(X). Then the expected value of X denoted by E(X), or μ, is defined as:

$$E(X) = \mu = \Sigma\left(x_i \times P(x_i)\right)$$

To calculate this, we multiply each possible value of the variable by its probability, then add the results.

$$\Sigma\left(x_i \times P(x_i)\right) = \{\, x_1 \times P(x_1)\} + \{\, x_2 \times P(x_2)\} + \{\, x_3 \times P(x_3)\} + \ldots$$

E(X) is also called the mean of the probability distribution.

Two balls are drawn at random in succession without replacement from an urn containing 4 red balls and 6 black balls. What is the expected number of red balls?

Possible Outcome	RR	RB	BR	BB
x_i	2	1	1	0
$P(x_i)$	$\frac{2}{15}$	$\frac{4}{15}$	$\frac{4}{15}$	$\frac{1}{3}$

$$E(X) = \sum \{x_i \cdot P(x_i)\}$$

$$= 2 \times \frac{2}{15} + 1 \times \frac{4}{15} + 1 \times \frac{4}{15} + 0 \times \frac{1}{3}$$

$$= \frac{4}{5}$$

$$= 0.8$$

This means that if we performed this experiment 1000 times, we would expect to get 800 red balls.

I throw a die and get $1 if it is showing 1, and get $2 if it is showing 2\, and get $3 if it is showing 3, etc. What is the amount of money I can expect if I throw it 100100 times?

For one throw, the expected value is:

$$E(X) = \sum \{x_i \cdot P(x_i)\} = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6}$$

$$= \frac{7}{2}$$

$$= 3.5$$

So for 100 throws, I can expect to get $350.

The number of persons X, in a Singapore family chosen at random has the following probability distribution:

X	1	2	3	4	5	6	7	8
$P(X)$	0.34	0.44	0.11	0.06	0.02	0.01	0.01	0.01

Find the average family size $E(X)$.

$E(X)$

$= \sum \{x_i \cdot P(x_i)\}$

$= 1 \times 0.34 + 2 \times 0.44 + 3 \times 0.11 + 4 \times 0.06 + 5 \times 0.02 + 6 \times 0.01 + 7 \times 0.01 + 8 \times 0.01$

$= 2.1$

So the average family size is $E(X) = \mu = 2.1$ people.

In a card game with my friend, I pay a certain amount of money each time I lose. I win $4 if I draw a jack or a queen and I win $5 if I draw a king or ace from an ordinary pack of 52 playing cards. If I draw other cards, I lose. What should I pay so that we come out even? (That is, the game is "fair"?)

X	J, Q ($4)	K, A ($5)	lose ($-\$x$)
$P(X)$	$\frac{8}{52}=\frac{2}{13}$	$\frac{2}{13}$	$\frac{9}{13}$

$$
\begin{aligned}
E(X) &= \sum \{x_i \cdot P(x_i)\} \\
&= 4 \times \frac{2}{13} + 5 \times \frac{2}{13} - x \times \frac{9}{13} \\
&= \frac{18-9x}{13}
\end{aligned}
$$

Now the expected value should be $0 for the game to be fair.

So $\frac{18-9x}{13} = 0$ and this gives $x = 2$.

So I would need to pay $2 for it to be a fair game.

Variance of a Random Variable

Let X represent a discrete random variable with probability distribution function $P(X)$. The **variance** of X denoted by $V(X)$ or σ^2 is defined as:

$$V(X) = \sigma^2$$

$$= \Sigma[\{X - E(X)\}^2 \times P(X)]$$

Since $\mu = E(X)$, (or the average value), we could also write this as:

$$V(X) = \sigma^2$$

$$= \Sigma[\{X - \mu\}^2 \times P(X)]$$

Another way of calculating the variance is:

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

Standard Deviation of the Probability Distribution

$\sigma = \sqrt{V(X)}$ is called the **standard deviation** of the probability distribution. The standard deviation is a numb which describes the **spread** of the distribution. Small standard deviation means small spread, large standard deviation means large spread.

Find $V(X)$ for the following probability distribution:

X	8	12	16	20	24
$P(X)$	$\frac{1}{8}$	$\frac{1}{6}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{12}$

We have to find $E(X)$ first:

$$E(X) = 8 \times \frac{1}{8} + 12 \times \frac{1}{6} + 16 \times \frac{3}{8} + 20 \times \frac{1}{4} + 24 \times \frac{1}{12} = 16$$

Then:

$$V(X) = \sum \left[\{X - E(X)\}^2 \cdot P(X) \right]$$

$$= (8-16)^2 \times \frac{1}{8} + (12-16)^2 \times \frac{1}{6} + (16-16)^2 \times \frac{3}{8} + (20-16)^2 \times \frac{1}{4} + (24-16)^2 \times \frac{1}{12}$$

$$= 20$$

Checking this using the other formula:

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

For this, we need to work out the expected value of the **squares** of the random variable X.

X	8	12	16	20	24
X^2	64	144	256	400	576
$P(X)$	$\frac{1}{8}$	$\frac{1}{6}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{12}$

$$E(X^2) = \sum X^2 P(X)$$

$$= 64 \times \frac{1}{8} + 144 \times \frac{1}{6} + 256 \times \frac{3}{8} + 400 \times \frac{1}{4} + 576 \times \frac{1}{12}$$

$$= 276$$

We found $E(X)$ before: $E(X) = 16$

$V(X) = E(X^2) - [E(X)]^2 = 276 - 16^2 = 20$, as before.