Random Variables

Colby Community College

Example 1

Two books are assigned for a course:

- A textbook, which costs \$137.
- The accompanying study guide, which costs \$33.

The university bookstore determined:

- 20% of enrolled students do not buy either book.
- 55% of enrolled students only buy the textbook.
- 25% of the enrolled students buy both books.

How many books should be expected to sell if 100 students enrolled?

We expect about 25 students will buy none, 55 will buy just the textbook, and 25 will buy both. A total of $1 \cdot 55 + 2 \cdot 25 = 105$ books.

How much revenue should be expected to be made?

$\$0 \cdot 25 + \$137 \cdot 55 + (\$137 + \$33) \cdot 25 = \$7,535 + \$4,250 = \$11,785$

What is the average revenue per student?

$\$11,785 \div 100$ students $= \$117.85$ per student.

Definition

A **random variable** is a random process or variable that has a numerical value, determined by chance, for each outcome.

Note

Often capital letters such as X, Y, or Z are used for random variables.

Corresponding lower case letters are used for the possible outcomes.

Example 2

Consider tossing a coin: We could get either a heads or a tails.

If we let X be the number of tails we get in a single flip, then the two possible outcomes are: $x_1 = 0$ and $x_2 = 1$.

The probability distribution is:

i	1	2	Total
x_i	0	1	–
$P(X = x_i)$	0.50	0.50	1.00

Example 3

If we let X be the amount a student spends in Example 1, then the probability distribution is:

i	1	2	3	Total
x_i	\$0	\$137	\$170	–
$P(X = x_i)$	0.20	0.55	0.25	1.00

Definition

A **discrete random variable** has a collection of values that is finite or countable.

Definition

A **continuous random variable** has infinitely many values, and the collection of values is not countable.

Note

The random variable in Example 3 is a discrete random variable.

Example 4

Let's consider tossing two coins, with the following random variable:

$$X = \text{number of heads when two coins are tossed}$$

Is X is a discrete or continuous random variable?

X is discrete since the number of heads can be 0, 1, or 2.

The probability distribution is:

i	1	2	3	Total
x_i	0	1	2	–
$P(X = x_i)$	0.25	0.50	0.25	1.00

Example 5

Hiring managers were asked to identify the biggest mistakes that job applicants make during an interview.

Consider the following table:

i	x_i	$P(X = x_i)$
1	Inappropriate attire	0.50
2	Being late	0.44
3	Lack of Eye Contact	0.33
4	Checking phone or texting	0.30
	Total	1.57

Is X and random variable?

1. The outcomes are not numerical, they are categorical.
2. $\sum P(X = x_i) = 1.57 \neq 1$

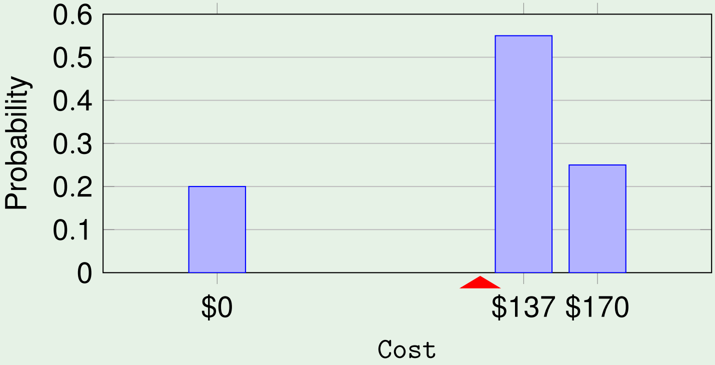
So, we see that X is not a random variable.

Definition

The average outcome of X is called the **expected value** of X.

Example 6

In Example 1 the average revenue, \$117.85 per student, is the expected value for the bookstore's revenue.



Expected Value Of A Discrete Random Variable

If X takes outcomes $x_1, \ldots, x_k$ with probabilities $P(X = x_1), \ldots, P(X = x_k)$, the expected value of X is:

$$\begin{aligned} E(X) &= x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) + \cdots + x_k \cdot P(X = x_k) \\ &= \sum_{i=1}^{k} x_i \cdot P(X = x_i) \end{aligned}$$

Note

The Greek letter μ is sometimes used in place of $E(X)$.

Example 7

In the casino game Roulette, a wheel with 38 numbered spaces is spun. There are 18 red spaces, 18 black spaces, and 2 green spaces.

In one possible bet, the players bet \$1 on a single number. If that number is spun on the wheel, then they receive \$36. Otherwise, they lose their \$1.

If X is the net winnings, then the probability distribution is:

i	1	2	Total
x_i	\$35	-\$1	–
$P(X = x_i)$	$\frac{1}{38}$	$\frac{37}{38}$	1

The expected value of X is:

$$E(X) = x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) + x_3 \cdot P(X = x_3)$$

$$= \$35 \cdot \frac{1}{38} + -\$1 \cdot \frac{37}{38} = \$0.9211 - \$0.9737 \approx -\$0.053$$

On *average*, the player will lose 5.3 cents per bet.

Example 8

Suppose an individual has a 0.242% risk of dying during the next year.

An insurance company charges $275 for a life-insurance policy that pays $100,000 death benefit.

What is the expected value for the person buying the insurance?

The probabilities distribution is:

i	1	2	Total
x_i	$99,725	-$275	–
$P(X = x_i)$	0.00242	0.99758	1.0

The expected value is:

$$\begin{aligned} E(X) &= x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) \\ &= \$99,725 \cdot 0.00242 - \$275 \cdot 0.99758 = -\$33 \end{aligned}$$

Note

It makes sense that an insurance policy would have a negative expected value, otherwise the insurance company couldn't stay in business.

Example 9

A company estimates that 0.7% of their products will fail, with a replacement cost of \$350, after the original warranty period but within 2 years of the purchase.

If they offer a 2 year extended warranty for \$48, what is the company's expected value?

The probabilities and values for the two outcomes are:

i	1	2	Total
x_i	-\$302	\$48	–
$P(X = x_i)$	0.007	0.993	1.0

The expected value is:

$$\begin{aligned} E(X) &= x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) \\ &= -\$302 \cdot 0.007 + \$48 \cdot 0.993 = \$45.55 = -\$33 \end{aligned}$$

The company makes, on average, \$45.55 for each extended warranty.

Note

If you ran the university bookstore in Example 1, then not only would you want to know your expected revenue, but also how much variability there is in your revenue.

General Variance Formula

If X takes outcomes $x_1, \ldots, x_k$ with probabilities $P(X = x_1), \ldots,$ $P(X = x_k)$ and expected value $\mu = E(x)$, then the variance of X, denoted by $Var(X)$ or the symbol σ^2, is:

$$\begin{aligned}\sigma^2 &= (x_1 - \mu)^2 \cdot P(X = x_1) + \cdots + (x_2 - \mu)^2 \cdot P(X = x_2) \\ &= \sum_{j=1}^{k} (x_j - \mu)^2 \cdot P\left(X = x_j\right)\end{aligned}$$

The standard deviation of X, denoted σ, is the square root of the variance. i.e. $\sigma = \sqrt{\sigma^2}$

Example 10

Let us find the expected value of the bookstore in Example 1.

It is useful to construct a table to hold the computations:

i	1	2	3	Total	
x_i	\$0.00	\$137.00	\$170.00	–	
$P(x = x_i)$	0.20	0.55	0.25	–	
$x_i \cdot P(X = x_i)$	0.00	75.35	42.50	117.85	$= \mu$
$x_i - \mu$	-117.85	19.15	52.15	–	
$(x_i - \mu)^2$	13888.62	366.72	2719.62	–	
$(x_i - \mu)^2 \cdot P(X = x_i)$	2777.72	201.70	679.91	3659.33	$= \sigma^2$

The variance of X is $\sigma^2 = 3659.33$ and so the standard deviation is $\sigma = \sqrt{3659.33} = \$60.49$.

So, on average, we can expect a variability of revenue of around \$60.49 per student.

Example 11

The bookstore also offers a chemistry textbook for \$159 with a supplement for \$41. From past expereince, they know about 25% of chemistry students just buy the textbook while 60% buy both.

Let us construct the same type of table:

i	1	2	3	Total	
x_i	\$0.00	\$159.00	\$200.00	–	
$P(x = x_i)$	0.15	0.25	0.60	–	
$x_i \cdot P(X = x_i)$	0.00	39.75	120.00	159.75	$= \mu$
$x_i - \mu$	-159.75	-0.75	40.25	–	
$(x_i - \mu)^2$	25520.06	0.56	1620.06	–	
$(x_i - \mu)^2 \cdot P(X = x_i)$	3828.01	0.14	972.04	4800.19	$= \sigma^2$

The variance of X is $\sigma^2 = 4800.19$ and so the standard deviation is $\sigma = \sqrt{4800.19} = \$69.28$.

So, on average, we can expect a variability of revenue of around \$69.28 per student.

Example 12

John travels to work five days a week.

We will use:

- X_1 to represent his travel time on Monday
- X_2 to represent his travel time on Tuesday
- X_3 to represent his travel time on Wednesday
- X_4 to represent his travel time on Thursday
- X_5 to represent his travel time on Friday

His total travel time for the week is the sum of the daily five values:

$$W = X_1 + X_2 + X_3 + X_4 + X_5$$

Note

By breaking the week into the individual days we can better understand the source of each randomness and is useful for modeling W.

Example 13

It takes John an average of $E(X_i) = 18$ minutes each day to commute to work.

To find the expected value of his average commute times for the week we can sum the expected time for each day:

$$\begin{aligned} E(W) &= E(X_1 + X_2 + X_3 + X_4 + X_5) \\ &= E(X_1) + E(X_2) + E(X_3) + E(X_4) + E(X_5) \\ &= 18 + 18 + 18 + 18 + 18 \\ &= 90 \text{ minutes} \end{aligned}$$

Would you be surprised if John's weekly commute wasn't exactly 90 minutes long?

There is always some variability with probabilities, so we can reasonably expect his commute to be a bit different from 90 minutes.

Example 14

Elena is selling a TV on eBay and plans to use the money to buy a toaster oven.

Let X represent the profit for selling the TV and Y represent the cost of the toaster oven.

What equation represents the net change in Elena's eBay account?

The net change is money earned minus money spent: $X - Y$.

Based on past auctions, Elena figures she should expect to get about \$175 on the TV and pay about \$23 fo the toaster oven.

In total, how much should she expect to make or spend?

$$E(X - Y) = E(X) - E(Y) = 175 - 23 = 152$$

So, she should expect to make about \$152.

Definition

A **linear combination** of two random variables X and Y is

$$aX + bY$$

where a and b are some fixed and known numbers.

Example 15

In Example 14, Elena's net change is the linear combination

$$1 \cdot X + (-1) \cdot Y$$

Example 16

In Example 12, John's weekly commute time is the linear combination

$$1 \cdot X_1 + 1 \cdot X_2 + 1 \cdot X_3 + 1 \cdot X_4 + 1 \cdot X_5$$

Expected Value of Linear Combinations of Random Variables

If X and Y are random variables, then

$$E(aX + bY) = a \cdot E(X) + b \cdot E(Y)$$

Example 17

Leonard has invested \$6000 in Caterpiller Inc. (CAT) and \$2000 in Exxon Mobile Corp. (XOM).

- X represents the change in Caterpiller's stock next month, and
- Y represents the change in Exxon Mobile's stock next month

What is the equation describing how much Leonard will make or lose next month?

$$\$6000 \cdot X + \$2000 \cdot Y$$

Caterpillar piller stock has recently been rising at 2.0% and Exxon Mobil's at 0.2% per month.

What is the expected change in Leonard's stock portfolio?

$$\begin{aligned} E(\$6000 \cdot X + \$2000 \cdot Y) &= \$6000 \cdot E(X) + \$2000 \cdot Y \\ &= \$6000 \cdot 0.02 + \$2000 \cdot 0.002 = \$124 \end{aligned}$$

Would it be surprising to learn Leonard actually had a loss this month?

While stocks tend to rise over time, they are often volatile in the short term.

Variability of Linear Combinations of Random Variables

The variance of a linear combination is

$$Var(aX + bY) = a^2 \cdot Var(X) + b^2 \cdot Var(y)$$

Note

This formula only works if X and Y are independent random variables.

Note

Why the formula works and what happens if X and Y are dependent will be left to a dedicated probability course.

Note

As usual, you can get the standard deviation by taking the square root of the variance.

Example 18

Suppose John's daily commute has a standard deviation of 4 minutes.

What is the uncertainty in his total weekly commute?

We start by noting that $Var(X_i) = \sigma^2 = 4^2 = 16$.

Recall that that:

$$W = 1 \cdot X_1 + 1 \cdot X_2 + 1 \cdot X_3 + 1 \cdot X_4 + 1 \cdot X_5$$

So,

$$\begin{aligned} Var(W) &= Var(1 \cdot X_1 + 1 \cdot X_2 + 1 \cdot X_3 + 1 \cdot X_4 + 1 \cdot X_5) \\ &= 1^2 \cdot 16 + 1^2 \cdot 16 + 1^2 \cdot 16 + 1^2 \cdot 16 + 1^2 \cdot 16 \\ &= 80 \end{aligned}$$

$$\sigma = \sqrt{Var(W)} = \sqrt{80} = 8.94$$

The standard deviation for John's commute is about 9 minutes.

Is is reasonable to assume each daily commute is independent?

It depends on traffic patterns and what mode of transport John uses.

Example 19

It is reasonable to assume that an eBay auction for a TV is more or less independent from one for a toaster oven.

Suppose that the standard deviation of TV auctions is \$25 and toaster oven auctions is \$8.

What is the standard deviation of Elena's net gain?

Recall:

- X is the profit from the TV, so $Var(X) = 25^2 = 625$, and
- Y is the cost of the toaster oven, so $Var(Y) = 8^2 = 64$.
- The linear combination is $X - Y = 1 \cdot X + (-1) \cdot Y$.

Then,

$$\begin{aligned} Var(1 \cdot X + (-1) \cdot Y) &= 1^2 \cdot Var(X) + (-1)^2 \cdot Var(Y) \\ &= 1 \cdot 625 + 1 \cdot 64 = 689 \\ \sigma &= \sqrt{689} = 26.25 \end{aligned}$$

The standard deviation for Elena's net gain is about \$26.25.