50.034 - Introduction to Probability and Statistics

Week 5 – Lecture 9

January–May Term, 2019

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Lecture 9 Assigned Readings: 3.9, 4.6, 4.7

Outline of Lecture

- Functions of multiple R.V.'s
- Expectation of functions of multiple R.V.'s
- Covariance and correlation
- Conditional expectation

Functions of several R.V.'s

Recall: Any real-valued function of a R.V. is a R.V. (Lecture 5).

More generally, any real-valued function of any number of R.V.'s on the **same** sample space is always a R.V.

Examples:

- An investment portfolio consists of 10 stocks, whose prices are modeled by R.V.'s $X_1, \ldots, X_{10}$. Then the portfolio value can be modeled by

$$Y = a_1 X_1 + \cdots + a_{10} X_{10},$$

  where each a_i is the quantity of the stock X_i in the portfolio.
- In a shop, there are 5 customer service assistants. They have different rates for serving customers, modeled by $X_1, \ldots, X_5$. Then the customer waiting time can be modeled by

$$Y = \frac{1}{\frac{1}{X_1} + \cdots + \frac{1}{X_5}}.$$

Example 1

Consider the roll outcome of a fair die. Let

$$X = \begin{cases} 1, & \text{if the outcome is even;} \\ 0, & \text{otherwise;} \end{cases}$$

$$Y = \begin{cases} 1, & \text{if the outcome is prime;} \\ 0, & \text{otherwise.} \end{cases}$$

Question: What is the distribution of the new R.V. XY?

Answer: We have seen X and Y before!
In Example 1 of Lecture 7, we saw that $\Pr(X = 1, Y = 1) = \frac{1}{6}$.

- Only one outcome (the outcome 2), among the six possible outcomes, satisfies both $X = 1$ and $Y = 1$.

Example 1

$\Pr(X = 0, Y = 0) = \Pr(\text{odd and not prime}) = \frac{1}{6}$ (outcome 1).
$\Pr(X = 0, Y = 1) = \Pr(\text{odd and prime}) = \frac{2}{6}$ (outcomes $3, 5$).
$\Pr(X = 1, Y = 0) = \Pr(\text{even and not prime}) = \frac{2}{6}$ (outcomes $4, 6$).
$\Pr(X = 1, Y = 1) = \Pr(\text{even and prime}) = \frac{1}{6}$ (outcome 2).

The joint pmf of X and Y is given by:

	$Y = 0$	$Y = 1$
$X = 0$	$\frac{1}{6}$	$\frac{2}{6}$
$X = 1$	$\frac{2}{6}$	$\frac{1}{6}$

Note: XY is a discrete R.V. with possible values 0 and 1.

- $XY = 1$ if and only if both $X = 1$ and $Y = 1$.
- $\Pr(XY = 1) = \frac{1}{6}$, $\Pr(XY = 0) = \frac{5}{6}$.
- Thus, XY has a Bernoulli distribution with parameter $p = \frac{1}{6}$.

A closer look at Example 1

Questions:

(1): What is the mean μ_X and variance σ_X^2 of X?
(2): What is the mean μ_Y and variance σ_Y^2 of Y?
(3): Can we determine the mean and variance of XY just from $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$?

Solution:

(1): $\mu_X = \mathbf{E}[X] = \frac{1}{2}$. ($1, 3, 5$ are not even; $2, 4, 6$ are even.)

$$\mathbf{E}[X^2] = 0^2 \cdot \Pr(X = 0) + 1^2 \cdot \Pr(X = 1) = 0^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{2} = \frac{1}{2}.$$

Thus, $\sigma_X^2 = \mathrm{var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2 = \frac{1}{2} - (\frac{1}{2})^2 = \frac{1}{4}$.
(2): $\mu_Y = \mathbf{E}[Y] = \frac{1}{2}$. ($1, 4, 6$ are not prime; $2, 3, 5$ are prime.)

$$\mathbf{E}[Y^2] = 0^2 \cdot \Pr(Y = 0) + 1^2 \cdot \Pr(Y = 1) = 0^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{2} = \frac{1}{2}.$$

Thus, $\sigma_Y^2 = \mathrm{var}(Y) = \mathbf{E}[Y^2] - (\mathbf{E}[Y])^2 = \frac{1}{2} - (\frac{1}{2})^2 = \frac{1}{4}$.

A closer look at Example 1

(3): So far, we have $\mu_X = \mu_Y = \frac{1}{2}$ and $\sigma_X^2 = \sigma_Y^2 = \frac{1}{4}$.

Let $Z = X_1 X_2$, where X_1, X_2 are discrete R.V.'s with means $\mu_{X_1} = \mu_{X_2} = \frac{1}{2}$ and variances $\sigma_{X_1}^2 = \sigma_{X_2}^2 = \frac{1}{4}$. Is there enough information to determine the mean and variance of Z?

Case 1: $X_1 = X$, $X_2 = Y$.

- We know that XY is a Bernoulli R.V. with parameter $p = \frac{1}{6}$.
- This implies $\mu_{XY} = p = \frac{1}{6}$ and $\sigma_{XY}^2 = p(1-p) = \frac{5}{36}$.

Case 2: $X_1 = X$, $X_2 = X$.

- $\mu_{X^2} = \mathbf{E}[X^2] = \frac{1}{2}$ (from previous slide).
- $\mathbf{E}[X^4] = 0^4 \cdot \frac{1}{2} + 1^4 \cdot \frac{1}{2} = \frac{1}{2}$.
- $\sigma_{X^2}^2 = \mathrm{var}(X^2) = \mathbf{E}[X^4] - (\mathbf{E}[X^2])^2 = \frac{1}{2} - (\frac{1}{2})^2 = \frac{1}{4}$.

Conclusion: The examples $Z = XY$ and $Z = X^2$ show that we cannot determine the mean and variance of $Z = X_1 X_2$ simply from knowing the means and variances of both X_1 and X_2.

Mean and variance: Why are they insufficient?

Given two R.V.'s X and Y, the mean and variance of each R.V. give useful information about each individual R.V.

- μ_X and σ_X^2 give info about the marginal distribution of X.
- μ_Y and σ_Y^2 give info about the marginal distribution of Y.

However, even if $\mu_X = \mu_Y$, $\sigma_X^2 = \sigma_Y^2$, there are different scenarios:

- X and Y could be iid.
- X and Y could tend to vary together (not indepedently).
- X and Y could have completely different distributions.

As our previous example shows, X and Y could have the same mean and variance, even if $X \neq Y$!

- The means and variances of individual R.V.'s do not give information about how the R.V.'s are **jointly** distributed.

Covariance

Let X and Y be R.V.'s with finite means μ_X and μ_Y respectively.

The covariance of X and Y, denoted by $\text{cov}(X, Y)$, is defined to be

$$\text{cov}(X, Y) = \mathbf{E}[(X - \mu_X)(Y - \mu_Y)],$$

provided that this expectation $\mathbf{E}[(X - \mu_X)(Y - \mu_Y)]$ exists.

- Both μ_X and μ_Y must exist and be finite for $\text{cov}(X, Y)$ to make sense.
- **Fact:** If both $\text{var}(X)$ and $\text{var}(Y)$ exist and are finite, then $\text{cov}(X, Y)$ exists and is finite.

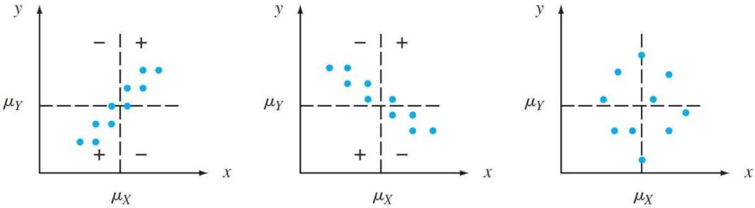
Recall: The variance of X is $\text{var}(X) = \mathbf{E}[(X - \mu_X)^2]$.

- In other words, $\text{var}(X) = \text{cov}(X, X)$.
- Covariance is an extension of variance!

Visualization of Covariance

Fact: Covariance can be positive, negative, or zero.

Let X and Y be R.V.'s such that $\mathrm{cov}(X, Y)$ exists.
Then the sign of $\mathrm{cov}(X, Y)$ can be perceived from the data.



positive | negative | near zero

Interpretation of Covariance

The covariance of X and Y gives a measure of how strongly X are Y are **linearly** related.

- If $\mathrm{cov}(X, Y) > 0$, then there is a **positive linear relationship**.
  - As the X-value increases, the Y-value tends to increase.
- If $\mathrm{cov}(X, Y) < 0$, then there is a **negative linear relationship**.
  - As the X-value increases, the Y-value tends to decrease.
- If $\mathrm{cov}(X, Y) = 0$, then there is **no linear relationship**.
  - **Note:** There could still be some non-linear relationship.

Examples of Covariance

Positive Covariance: X-value $\nearrow$ implies tendency for Y-value $\nearrow$

- X = height of person
- Y = weight of person

The taller the person, the heavier the person is likely to be.

Negative Covariance: X-value $\nearrow$ implies tendency for Y-value $\searrow$

- X = Number of days a student absent from school
- Y = GPA of student

The higher the number of days a student is absent from school, the lower his/her GPA is likely to be.

Zero Covariance: No linear relationship between X and Y

- X = height of student
- Y = GPA of student

Whether a student is taller or shorter (vs mean height), we cannot infer whether his/her GPA is more likely to be higher or lower.

A closer look at the definition of covariance

By definition, $\mathrm{cov}(X, Y) = \mathbf{E}[(X - \mu_X)(Y - \mu_Y)]$.

- It is the expected value of $(X - \mu_X)(Y - \mu_Y)$.
- For different observed values $X = x$ and $Y = y$, the value of $(x - \mu_X)(y - \mu_Y)$ could be positive, negative, or zero.
- As we collect more and more observed values, we can get a better sense of the average value of $(x - \mu_X)(y - \mu_Y)$.

Case 1: More outcomes with $\left\{\begin{matrix}\text{large } X\text{-value}\\ \text{large } Y\text{-value}\end{matrix}\right\}$ and/or $\left\{\begin{matrix}\text{small } X\text{-value}\\ \text{small } Y\text{-value}\end{matrix}\right\}$.

- $\left\{\begin{matrix}\text{large } X\text{-value}\\ \text{large } Y\text{-value}\end{matrix}\right\}$ implies $\left\{\begin{matrix}(X-\mu_X)>0\\ (Y-\mu_Y)>0\end{matrix}\right\}$, so $(X - \mu_X)(Y - \mu_Y)$ tends to be positive, i.e. $\mathrm{cov}(X, Y) > 0$.
- $\left\{\begin{matrix}\text{small } X\text{-value}\\ \text{small } Y\text{-value}\end{matrix}\right\}$ implies $\left\{\begin{matrix}(X-\mu_X)<0\\ (Y-\mu_Y)<0\end{matrix}\right\}$, so $(X - \mu_X)(Y - \mu_Y)$ tends to be positive, i.e. $\mathrm{cov}(X, Y) > 0$.

A closer look at the definition of covariance

Case 2: More outcomes with $\left\{\begin{matrix}\text{large } X\text{-value} \\ \text{small } Y\text{-value}\end{matrix}\right\}$ and/or $\left\{\begin{matrix}\text{small } X\text{-value} \\ \text{large } Y\text{-value}\end{matrix}\right\}$.

- $\left\{\begin{matrix}\text{large } X\text{-value} \\ \text{small } Y\text{-value}\end{matrix}\right\}$ implies $\left\{\begin{matrix}(X-\mu_X)>0 \\ (Y-\mu_Y)<0\end{matrix}\right\}$, so $(X-\mu_X)(Y-\mu_Y)$ tends to be negative, i.e. $\operatorname{cov}(X,Y)<0$.
- $\left\{\begin{matrix}\text{small } X\text{-value} \\ \text{large } Y\text{-value}\end{matrix}\right\}$ implies $\left\{\begin{matrix}(X-\mu_X)<0 \\ (Y-\mu_Y)>0\end{matrix}\right\}$, so $(X-\mu_X)(Y-\mu_Y)$ tends to be negative, i.e. $\operatorname{cov}(X,Y)<0$.

Case 3: All four possibilities are equally likely:

$$\left\{\begin{matrix}\text{large } X\text{-value} \\ \text{large } Y\text{-value}\end{matrix}\right\}, \left\{\begin{matrix}\text{small } X\text{-value} \\ \text{small } Y\text{-value}\end{matrix}\right\}, \left\{\begin{matrix}\text{large } X\text{-value} \\ \text{small } Y\text{-value}\end{matrix}\right\}, \left\{\begin{matrix}\text{small } X\text{-value} \\ \text{large } Y\text{-value}\end{matrix}\right\}$$

- First two subcases: increase in $(X-\mu_X)(Y-\mu_Y)$.
- Last two subcases: decrease in $(X-\mu_X)(Y-\mu_Y)$.

Overall effect: The contributions cancel out, i.e. $\operatorname{cov}(X,Y)=0$.

Calculation of covariance

Let X and Y be R.V.'s with finite means μ_X and μ_Y respectively. Suppose that the covariance $\mathrm{cov}(X, Y)$ exists.

How to calculate covariance?

- If X and Y are **discrete** R.V.'s with joint pmf $p(x, y)$, then

$$\mathrm{cov}(X, Y) = \sum_{x \in D_X} \sum_{y \in D_Y} (x - \mu_X)(y - \mu_Y) p(x, y),$$

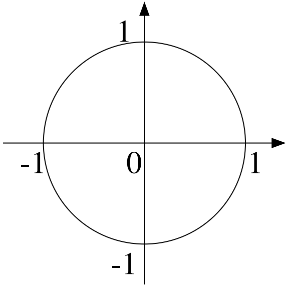
  where D_X, D_Y are the possible values of X, Y respectively.

- If X and Y are **continuous** R.V.'s with joint pdf $f(x, y)$, then

$$\mathrm{cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x, y) \, dx \, dy.$$

Example 2

A point is chosen randomly from a disk of radius 1. Let X and Y denote the x-coordinate and y-coordinate respectively of the point.
The joint pdf of X and Y is



$$f(x,y) = \begin{cases} \frac{1}{\pi}, & \text{if } x^2 + y^2 \leq 1; \\ 0, & \text{otherwise.} \end{cases}$$

(1): What is the covariance of X and Y?

(2): Are X and Y independent?

Example 2

Solution:

(1): First, we note that $\mu_X = \mu_Y = 0$ by symmetry, since the pdf $f(x, y)$ does not depend on x and y.

By the definition of covariance, we get

$$
\begin{aligned}
\mathrm{cov}(X, Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x, y)\, dx\, dy \\
&= \int_{-1}^{1} \int_{-1}^{1} (x - 0)(y - 0) \tfrac{1}{\pi}\, dx\, dy \\
&= \tfrac{1}{\pi} \int_{-1}^{1} \left[\tfrac{1}{2} x^2 y \right]_{x=-1}^{x=1} dy \\
&= \tfrac{1}{\pi} \int_{-1}^{1} [\tfrac{1}{2} y - \tfrac{1}{2} y]\, dy \\
&= 0.
\end{aligned}
$$

Example 2

(2): In Example 3 of Lecture 7, we have already calculated the marginal pdf of Y:

$$f_Y(y) = \begin{cases} \frac{2}{\pi}\sqrt{1-y^2}, & \text{if } -1 \le y \le 1; \\ 0, & \text{otherwise.} \end{cases}$$

By symmetry, we also have the marginal pdf of X:

$$f_X(x) = \begin{cases} \frac{2}{\pi}\sqrt{1-x^2}, & \text{if } -1 \le x \le 1; \\ 0, & \text{otherwise.} \end{cases}$$

Consequently,

$$f_X(x)f_Y(y) = \begin{cases} \frac{4}{\pi^2}\sqrt{1-x^2}\sqrt{1-y^2}, & \text{if } -1 \le y \le 1; \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, $f_X(x)f_Y(y) \neq f(x,y)$, and we conclude that X and Y are not independent.

Zero covariance does NOT mean independence!

Warning! As the previous example shows,

$\mathrm{cov}(X, Y) = 0$ does **NOT** mean X and Y are independent!

- $\mathrm{cov}(X, Y) = 0$ only means that there is a complete absence of any **linear** relationship between X and Y.
- Two R.V.'s could have zero covariance but still be strongly dependent, because there is a **non-linear** relationship.

In the previous example, there is a relationship between X and Y.

- Given a randomly chosen point in this disk, if its x-coordinate is k, then the square of its y-coordinate must be $\leq 1 - k^2$.
- This is not a linear relationship, but it is still a relationship!

However, if two R.V.'s are independent, then we can conclude that they have zero covariance.

Very useful formula for covariance

Theorem: Let X and Y be R.V.'s with finite means μ_X and μ_Y respectively.
If the covariance $\mathrm{cov}(X, Y)$ exists, then

$$\mathrm{cov}(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y].$$

Proof: Since "mean of sum" = "sum of means", we have

$$\begin{aligned}\mathrm{cov}(X, Y) &= \mathbf{E}[(X - \mu_X)(Y - \mu_Y)] \\ &= \mathbf{E}[XY - X\mu_Y - \mu_X Y + \mu_X\mu_Y] \\ &= \mathbf{E}[XY] - \mu_Y\mathbf{E}[X] - \mu_X\mathbf{E}[X] + \mu_X\mu_Y\end{aligned}$$

Since $\mathbf{E}[X] = \mu_X$ and $\mathbf{E}[Y] = \mu_Y$, we can simplify the last line to:

$$\begin{aligned}&\mathbf{E}[XY] - \mu_Y\mu_X - \mu_X\mu_Y + \mu_X\mu_Y \\ &= \mathbf{E}[XY] - \mu_X\mu_Y \\ &= \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y].\end{aligned}$$

How to calculate $\mathbf{E}[XY]$?

Case: Let X and Y be **discrete** R.V.'s with joint pmf $p(x, y)$. Suppose the sets of possible values for X and Y are D_X and D_Y respectively. Then,

$$\mathbf{E}[XY] = \sum_{x \in D_X} \sum_{y \in D_Y} xy \cdot p(x, y).$$

Case: Let X and Y be **continuous** R.V.'s with joint pdf $f(x, y)$. Then,

$$\mathbf{E}[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot f(x, y)\, dx\, dy.$$

Example 3

Let X and Y be continuous R.V.'s with joint pdf

$$f(x,y) = \begin{cases} \dfrac{x+3y}{4}, & \text{if } 0 \le x \le 2, -1 \le y \le 1; \\ 0, & \text{otherwise.} \end{cases}$$

What is $\mathbf{E}[XY]$? What is the covariance of X and Y?

Solution:

$$\begin{aligned} \mathbf{E}[XY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot f(x,y)\, dx\, dy = \int_{-1}^{1} \int_{0}^{2} \tfrac{1}{4}(x^2 y + 3xy^2)\, dx\, dy \\ &= \int_{-1}^{1} \left[\tfrac{x^3 y}{12} + \tfrac{3x^2 y^2}{8} \right]_{x=0}^{x=2} dy = \int_{-1}^{1} \left(\tfrac{3}{2} y^2 + \tfrac{2}{3} y^3 \right) dy \\ &= \left[\tfrac{y^3}{2} + \tfrac{y^2}{2} \right]_{y=-1}^{y=1} = 1. \end{aligned}$$

Example 3

Next, let $f_X(x)$ and $f_Y(y)$ be the marginal pdf's of X and Y.

$$\begin{aligned}
\mathbf{E}[X] &= \int_{-\infty}^{\infty} x \cdot f_X(x)\,dx = \int_{-\infty}^{\infty} x \cdot \left(\int_{-\infty}^{\infty} f(x,y)\,dy\right)\,dx \\
&= \int_0^2 x \cdot \left(\int_{-1}^{1} \tfrac{x+3y}{4}\,dy\right)\,dx = \int_0^2 x \cdot \left[\tfrac{xy}{4} + \tfrac{3y^2}{8}\right]_{y=-1}^{y=1} dx \\
&= \int_0^2 x \cdot \tfrac{x}{2}\,dx = \left[\tfrac{x^3}{6}\right]_{x=0}^{x=2} = \frac{4}{3}.
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}[Y] &= \int_{-\infty}^{\infty} y \cdot f_Y(y)\,dy = \int_{-\infty}^{\infty} y \cdot \left(\int_{-\infty}^{\infty} f(x,y)\,dx\right)\,dy \\
&= \int_{-1}^{1} y \cdot \left(\int_0^2 \tfrac{x+3y}{4}\,dx\right)\,dy = \int_{-1}^{1} y \cdot \left[\tfrac{x^2}{8} + \tfrac{3xy}{4}\right]_{x=0}^{x=2} dx \\
&= \int_{-1}^{1} y\left(\tfrac{1}{2} + \tfrac{3y}{2}\right)\,dx = \left[\tfrac{y^2}{4} + \tfrac{y^3}{2}\right]_{y=-1}^{y=1} = 1.
\end{aligned}$$

Therefore, $\operatorname{cov}(X,Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y] = 1 - \frac{4}{3} \cdot 1 = -\frac{1}{3}$.

Cauchy–Schwarz Inequality and definition of correlation

Pronunciation: "co-shee shh-wart-zz" inequality.

Theorem: Let X and Y be R.V.'s with finite variances. Then

$$[\mathrm{cov}(X, Y)]^2 \leq \mathrm{var}(X)\mathrm{var}(Y).$$

This inequality is famously called the Cauchy–Schwarz inequality.

Definition: Let X and Y be R.V.'s with finite variances σ_X^2, σ_Y^2. Then the correlation of X and Y, denoted by $\rho(X, Y)$, is

$$\rho(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}} = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Important Consequence of Cauchy–Schwarz inequality:
The correlation $\rho(X, Y)$ of R.V.'s X and Y, if it exists, satisfies:

$$-1 \leq \rho(X, Y) \leq 1.$$

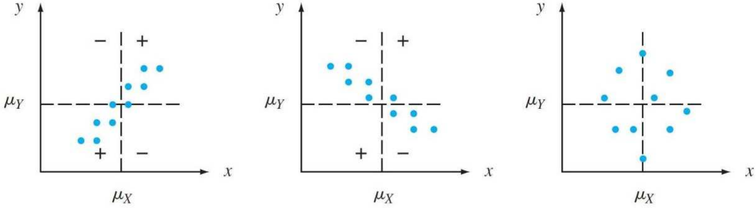
Interpretation of Correlation

Both covariance and correlation have similar interpretations.

Both $\operatorname{cov}(X, Y)$ and $\rho(X, Y)$ measure how strongly two R.V.'s X and Y are **linearly** related.

- If $\rho(X, Y) > 0$, then there is a **positive linear relationship**.
  - As the X-value increases, the Y-value tends to increase.
- If $\rho(X, Y) < 0$, then there is a **negative linear relationship**.
  - As the X-value increases, the Y-value tends to decrease.
- If $\rho(X, Y) = 0$, then there is **no linear relationship**.

The sign/rough value of $\rho(X, Y)$ can be perceived from the data.



close to 1 | close to -1 | close to 0

Covariance versus Correlation

Difference: $\mathrm{cov}(X, Y)$ depends on **scale**, but $\rho(X, Y)$ does not.
Example: Consider the roll outcome of a fair die. Let

$$X = \begin{cases} 1, & \text{if the outcome is even;} \\ 0, & \text{otherwise;} \end{cases}$$

$$Y = \begin{cases} 100, & \text{if the outcome is even;} \\ 0, & \text{otherwise.} \end{cases}$$

- Clearly, $Y = 100X$.
- As expected, X and Y have correlation $\rho(X, Y) = 1$.
- The covariance is $\mathrm{cov}(X, Y) = 2500 = 100^2 \cdot \rho(X, Y)$.

More generally, if we replace 100 by some K, then the correlation remains as 1, while the covariance changes to $\mathrm{cov}(X, Y) = \frac{1}{4}K^2$.

Intuition: Correlation does NOT depend on the scale (e.g. 1 point vs 100 points) or choice of units (e.g. centimeters vs inches).

Conditional distributions have conditional expectations

Let X, Y be R.V.'s, and let $C' \subseteq \mathbb{R}$ such that $\Pr(Y \in C') > 0$.

Recall: The conditional distribution of X given $Y \in C'$ is defined to be the collection of all **conditional probabilities** of the form $\Pr(X \in C | Y \in C')$ for all sets $C \subseteq \mathbb{R}$.

- For any $y \in \mathbb{R}$ such that $p_Y(y) > 0$, the conditional pmf of X given $Y = y$ is $p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)}$. (discrete R.V.'s)
- For any $y \in \mathbb{R}$ such that $f_Y(y) > 0$, the conditional pdf of X given $Y = y$ is $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$. (continuous R.V.'s)

Definition: The conditional expectation of X given $Y \in C'$ is the expectation of the conditional distribution of X given $Y \in C'$.

- **Note:** A conditional distribution is a legitimate distribution.
- Expectation can be defined on any distribution.
- Hence, it makes sense to consider $\mathbf{E}$[conditional distribution].
- "conditional expectation" = "conditional mean".

Conditional expectation

If $C' \subseteq \mathbb{R}$ such that $\Pr(Y \in C') > 0$, then the conditional expection of X given $Y \in C'$ is denoted by $\mathbf{E}[X|Y \in C']$.

- If C' is fixed, then $\mathbf{E}[X|Y \in C']$ is a fixed value.

If we are given a specific value $Y = y$, then the conditional expectation of X given $Y = y$ is denoted by $\mathbf{E}[X|Y = y]$.

- Similarly, if y is fixed, then $\mathbf{E}[X|Y = y]$ is a fixed value.

Conditional expectation as a variable:

- We can think of $\mathbf{E}[X|Y \in C']$ as a function in terms of C'.
  - Different values of C' give different values for $\mathbf{E}[X|Y \in C']$.
- Similarly, we can think of $\mathbf{E}[X|Y = y]$ as a function of y.
  - Different values of y give different values for $\mathbf{E}[X|Y = y]$.

More generally, we can think of $\mathbf{E}[X|Y]$ as a **function of** Y.

- **Recall:** Any real-valued function of a R.V. is a R.V.
- In other words, $\mathbf{E}[X|Y]$ is a random variable!

Law of total probability for expectations

The **law of total probability for expectations** states that if X and Y are **arbitrary** R.V.'s such that X has finite mean, then

$$\mathbf{E}[\mathbf{E}[X|Y]] = \mathbf{E}[X].$$

- In other words, the mean of the R.V. $\mathbf{E}[X|Y]$ equals $\mathbf{E}[X]$.
- Similarly, $\mathbf{E}[\mathbf{E}[Y|X]] = \mathbf{E}[Y]$ (provided Y has finite mean).

Intuition:

- The conditional expectation $\mathbf{E}[X|Y = y]$ would be different for different values of y.
- Different values of y have different probabilities of occurring. These probabilities are given by the distribution of Y.
- As we vary y over all possible values with their corresponding probabilities, the overall expected value of $\mathbf{E}[X|Y = y]$ would no longer be conditioned on Y being fixed at certain values, and hence should equal $\mathbf{E}[X]$.

Example 4

A blood test has two outcomes: positive, and negative. Let X and Y be R.V.'s defined on a group of patients, as follows:

- $X(\text{patient}) = \begin{cases} 0, & \text{if patient's blood test is negative;} \\ 1, & \text{if patient's blood test is positive.} \end{cases}$
- $Y(\text{patient}) = \begin{cases} 0, & \text{if patient is healthy;} \\ 1, & \text{if patient is sick.} \end{cases}$

The joint pmf $p(x, y)$ of X and Y is given as follows:

	Y=0	Y=1
X=0	0.72	0.005
X=1	0.18	0.095

What does $\mathbf{E}[\mathbf{E}[X|Y]]$ mean, and what is its value?

Solution: $\mathbf{E}[\mathbf{E}[X|Y]]$ is the overall mean of the patient's expected blood test result (conditioned on health status), as we vary over all patients, some healthy ($Y = 0$) and some sick ($Y = 1$). By the law of total probability for expectations, $\mathbf{E}[\mathbf{E}[X|Y]] = \mathbf{E}[X] = 0.275$.

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \qquad f^{n} \text{ of } x \qquad E_Y\left[E_X[X|Y]\right]$$

$$E[X|Y] = \int x\, f_{X|Y}(x|y)\, dx \qquad f^{n} \text{ of } y \qquad = E_X[X]$$

$$E\left[E[X|Y]\right] = \int\left[\int x\, f_{X|Y}(x|y)\, dx\right] f_Y(y)\, dy$$

$$= \int x\left[\int f_{X,Y}(x,y)\, dy\right] dx = \int x\, f_X(x)\, dx = E[X]$$

$$E\left[E[X|Y]\right] = E[X]$$

Find $E[X|Y]$?

$$E[X|Y=0] = \sum x\, P_{X|Y}(x|y=0) = \overset{x=0}{0}\,\frac{P_{X,Y}(0,0)}{P_Y(0)} + \overset{x=1}{1}\,\frac{P_{X,Y}(1,0)}{P_Y(0)}$$

$$E[X|Y=1] = \sum x\, P_{X|Y}(x|y=1) = 0\,\frac{P_{X,Y}(0,1)}{P_Y(1)} + 1\,\frac{P_{X,Y}(1,1)}{P_Y(1)}$$

$$E\left[E[X|Y]\right] = E[X|Y=0]\,P_Y(y=0) + E[X|Y=1]\,P_Y(y=1)$$

$$= E[X]$$