# Review of Probability: Part 1

Osman DOĞAN

# Outline: Review of Probability - Part 1

- Review of Probability - Part 1:
  1. Random variables (RVs) and probability distributions
  2. Expected values and moments
  3. Two random variables: Joint, marginal and conditional distributions
  4. Covariance, correlation and independence
- Readings:
  1. Stock and Watson (2020, Chapter 2).
  2. Hanck et al. (2021, Chapter 2).

## The probabilistic framework

- Why do we need a probabilistic framework?
- Because the analysis require dealing with uncertainty inherent in data due to sampling, measurement error, and so on.
- We need to understand why we often deal with random variables in econometrics.
- We will also see how to characterize them.
- The methods we will develop will be functions of these random variables.
- Therefore, the characterization will extend to these methods as well.

## Random variables

- Consider a random process (experiment): for example, your letter grade from this course, Turkiye's GDP in 2023, the unemployment rate in 2023Q3, and so on.

**Definition 1**

An experiment is the process by which an outcome (or observation) is obtained.

- The mutually exclusive potential results of a random process are called the outcomes.

**Definition 2**

The **sample space** associated with an experiment is the set consisting of all possible outcomes. A sample space will be denoted by $S$ (or $\Omega$).

- The sample space may or may not be countable.

**Definition 3**

A subset of sample space, that is, a collection of possible outcomes, is called an **event**.

## Random variables

- For example, consider the following random process (experiment): the sex of the next new person you meet.
- The sample space consists of two outcomes: $S = \{F, M\}$.
- What are the subsets (events) of the sample space?

$$\{\phi, \{F\}, \{M\}, S\}.$$

- What is the event then when the subset is the empty set? You do not meet anyone new?

**Definition 4**

The probability of an outcome is the proportion of the time that the outcome occurs in the long run.

- If $S = \{F, M\}$, then $P(M) = P(F) = 1/2$.

## Random variables

- For example, consider the following random process (experiment): the number of times your wireless network connection fails while you are writing a term paper.
- Then, the sample space consists of infinite (but countable) outcomes, $S = \{0, 1, 2, \ldots\}$.
- What are the subsets of the sample space? We cannot list all the subsets.
- What is the event then when the subset is $\{0, 1, 2\}$? Connection fails at most twice.

**Definition 5**

A random variable (RV) is a numerical summary of the sample space of a random process.

## Example 1

Define an experiment as tossing two coins and observing the results. Let $Y$ be a random variable that equals the number of heads obtained. Identify $S$ and the values $Y$ can take. Compute the probabilities for each value of $Y$.

## Solution 1 (Solution of Example 1)

*The sample space is*

$$S = \{HH, HT, TH, TT\}$$

$Y$ *can take three values,* $Y = 0,\ 1,$ *and* $2$*, which are events defined by specific collections of sample points:*

$$\{Y = 0\} = \{TT\}, \quad \{Y = 1\} = \{HT, TH\}, \quad \{Y = 2\} = \{HH\}.$$

*Then,*

$$P(Y = 0) = P(\{TT\}) = 1/4,\ P(Y = 1) = P(\{HT, TH\}) = 1/2,$$
$$P(Y = 2) = P(\{HH\}) = 1/4.$$

## Random variables

- There are 3 types of RVs:
  - Discrete RVs: takes a countable set of values,
  - Continuous RVs: takes an uncountable set of values,
  - Mixed RVs: a combination of discrete and continuous.
- An example for the discrete type would be: the number of times your wireless network connection fails while you are writing a term paper.
- An example for the continues type would be: your salary in your new job after graduation.

## Probability distribution of the RV

- One way to characterize the RV is through its probability distribution.

**Definition 6**

The probability distribution of a discrete random variable is the list of all possible values of the variable and the probability that each value will occur.

**Definition 7**

The cumulative probability distribution (CDF) is the probability that the random variable is less than or equal to a particular value.

- Let $Y$ be the number of times your wireless network connection fails while you are writing a term paper.

- Then, we can characterize the probability distribution and CDF of $Y$ as

$$\begin{array}{lcl} P(Y=0)=0.80 & & P(Y\leq 0)=0.80 \\ P(Y=1)=0.10 & & P(Y\leq 1)=0.90 \\ P(Y=2)=0.06 & \text{and} & P(Y\leq 2)=0.96 \\ P(Y=3)=0.03 & & P(Y\leq 3)=0.99 \\ P(Y=4)=0.01 & & P(Y\leq 4)=1 \end{array}$$

## Probability distribution of the RV

- We can use a table to characterize the probability distribution and the CDF of $Y$:

Table 1: The pdf and cdf of $Y$

| | number of failures | | | | |
|---|---|---|---|---|---|
| | $0$ | $1$ | $2$ | $3$ | $4$ |
| probability distribution | 0.80 | 0.10 | 0.06 | 0.03 | 0.01 |
| cumulative probability distribution | 0.80 | 0.90 | 0.96 | 0.99 | 1.00 |

- The probability of an event can be computed from the probability distribution.

- For example, the probability of the event of at least one failure is the sum of the probabilities of the constituent outcomes.

$$\begin{aligned} P(Y \geq 1) &= P(Y = 1 \text{ or } Y = 2 \text{ or } Y = 3 \text{ or } Y = 4) \\ &= 1 - P(Y = 0) = 1 - 0.80 = 0.20. \end{aligned}$$

## Probability distribution of the RV

- An important special case of a discrete random variable is when the random variable is binary; that is, the outcome is $0$ or $1$.
- A binary random variable is called a Bernoulli random variable.
- For example, let $G$ be the sex of the next new person you meet: $1$ for female and $0$ for male.
- Then, the probability distribution of $G$ is

$$G = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases} \tag{1}$$

## Probability distribution of the RV

- For a continuous RV, it makes no sense to list all outcomes and associated probabilities, because the range of values it takes is not countable.
- But, the CDF for a continuous variable is defined just as it is for a discrete random variable: the probability that the RV is less than or equal to a particular value.
- From the CDF, under some suitable scenarios, we can derive what is called, the probability density function (PDF).
- The pdf summarizes the probability of the RV taking on an interval of values.
- For example, consider the commuting time of a student from home to school. Figure 1 shows the cdf and the pdf of commuting times.

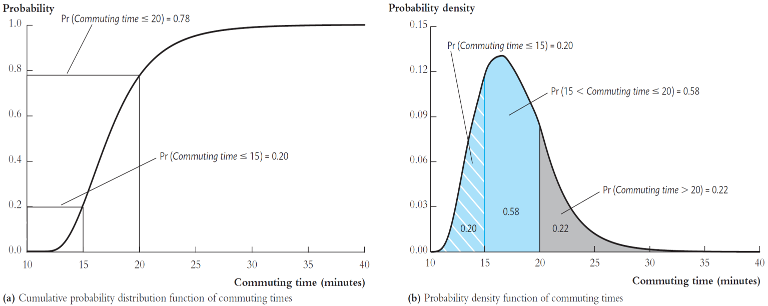



Figure 1: The cdf and pdf of commuting time

- From the cdf and the pdf, we can compute $P(\text{Commuting time} \leq 15) = 0.20$ and $P(15 < \text{Commuting time} \leq 20) = 0.58$.

## Moments of a RV

- Another way to characterize a RV is through its moments.

**Definition 8**

The expected value of a random variable $Y$, denoted by $\mathrm{E}(Y)$, is the average value of the random variable over many repeated trials or occurrences.

- Let $Y$ be a discrete random variable that can take on $k$ values, $y_1, y_2, \ldots, y_k$. Then,

$$\mathrm{E}(Y) = \mu_Y = \sum_{j=1}^{k} y_j \times P(Y = y_j). \tag{2}$$

- For $r > 1$, the $r$'th central moment of $Y$ is defined by

$$\mathrm{E}\left(Y - \mu_Y\right)^r = \sum_{j=1}^{k} \left(y_j - \mu_Y\right)^r \times P(Y = y_j). \tag{3}$$

- Setting $r = 2$, we will get the variance of $Y$: $\sigma_Y^2 = E\left(Y - \mu_Y\right)^2$.
- The positive square root of the variance is the standard deviation, denoted often as $\sigma_Y$.

## Moments of a RV

### Example 2

Let $Y$ be the number of times your wireless network connection fails while you are writing a term paper. Use Table 1 to compute $\mu_Y$ and $\sigma_Y^2$.

- Using definitions, we have

$$\begin{aligned}
\mathrm{E}(Y) = \mu_Y = \sum_{j=1}^{5} y_j \times P(Y = y_j) &= (0 \times 0.80) + (1 \times 0.10) + (2 \times 0.06) \\
&+ (3 \times 0.03) + (4 \times 0.01) = 0.35
\end{aligned}$$

$$\begin{aligned}
\mathrm{var}(Y) = \sigma_Y^2 = \sum_{j=1}^{5} (y_j - \mu)^2 \times P(Y = y_j) &= (0 - 0.35)^2 \times 0.80 \\
&+ (1 - 0.35)^2 \times 0.10 + (2 - 0.35)^2 \times 0.06 + (3 - 0.35)^2 \times 0.03 \\
&+ (4 - 0.35)^2 \times 0.01 = 0.6475
\end{aligned}$$

### Exercise 1

Consider the Bernoulli random variable $G$ in (1). Determine its mean and variance.

## Moments of a RV

- Using central moments, we can also characterize the probability distribution of random variables.
- The skewness of the distribution of $Y$ is

$$\text{Skewness} = \frac{\mathrm{E}((Y - \mu_Y)^3)}{\sigma_Y^3}. \tag{4}$$

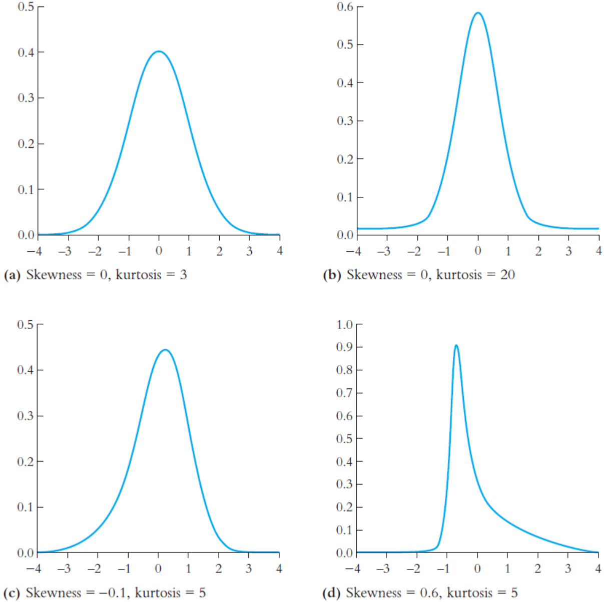
  1. Skewness$= 0$ means that the distribution is symmetric around $\mu_Y$,
  2. Skewness$> 0$ means the distribution has long right tail,
  3. Skewness$< 0$ means the distribution has long left tail,
  4. Skewness is unit free, i.e., changing the units of $Y$ does not change its skewness.

- The kurtosis of a distribution is a measure of how much mass is in its tails:

$$\text{Kurtosis} = \frac{\mathrm{E}((Y - \mu_Y)^4)}{\sigma_Y^4}. \tag{5}$$

  1. Kurtosis $= 3$: normal distribution,
  2. Kurtosis $> 3$ means the distribution has heavy (fat) tails (leptokurtotic).

- Figure 2 shows four distributions that have a mean of $0$ and a variance of $1$. The distributions in (a) and (b) are symmetric while those in (c) and (d) are not.




**(a)** Skewness = 0, kurtosis = 3

**(b)** Skewness = 0, kurtosis = 20

**(c)** Skewness = −0.1, kurtosis = 5

**(d)** Skewness = 0.6, kurtosis = 5

Figure 2: Distributions with different skewness and kurtosis

## Moments of a RV

- Let $Y$ be a **continuous random variable** that take values on a subset of $D \subset \mathbb{R}$. Let $f_Y(y)$ be the density of $Y$. Then, the expectation of $Y$ is defined by

$$\mu_Y = \mathrm{E}(Y) = \int_D y f_Y(y) \mathsf{d}y.$$

- For $r > 1$, the $r$'th central moment $Y$ is

$$\mathrm{E}\left(Y - \mu_Y\right)^r = \int_D \left(y - \mu_Y\right)^r f_Y(y) \mathsf{d}y.$$

**Example 3**

Consider $Y$ with PDF $f_Y(y) = 3/y^4$ for $y > 1$. We will show that this is a proper density.

$$\int_1^\infty \frac{3}{z^4} \mathsf{d}z = -\frac{3}{3} z^{-3}\Big|_{z=1}^{\infty} = -z^{-3}\big|_{z=1}^{\infty} = -\left(\lim_{t\to\infty} \frac{1}{t^3} - 1\right) = 1.$$

## Moments of a RV

- R and Python provides functions for numerical integration methods. The integral in Example 3 can be computed in the following way.

Listing 1: Numerical integration in R

```
fun = function(z) {
  return(3/z^4)
  }
integrate(fun,lower=1,upper=Inf)
```

Listing 2: Numerical integration in Python

```
from scipy.integrate import quad
import numpy as np

# Define the function
def fun(z):
    return 3 / z**4

# Perform the integration
result, error = quad(fun, 1, np.inf)

print("Integration result:", result)
print("Estimated error:", error)
```

## Moments of a RV

### Example 4

Consider the random variable in Example 3. Find $\mu_Y = \mathrm{E}(Y)$ and $\sigma_Y^2$.

$$\mathrm{E}(Y) = \int_1^\infty y \frac{3}{y^4} \mathrm{d}y = \int_1^\infty \frac{3}{y^3} \mathrm{d}y = -\frac{3}{2} y^{-2} \Big|_{y=1}^{\infty} = -\frac{3}{2} \left( \lim_{y\to\infty} \frac{1}{y^2} - 1 \right) = \frac{3}{2}.$$

$$\mathrm{E}(Y^2) = \int_1^\infty y^2 \frac{3}{y^4} \mathrm{d}y = \int_1^\infty \frac{3}{y^2} \mathrm{d}y = -3y^{-1} \big|_{y=1}^{\infty} = -3 \left( \lim_{y\to\infty} \frac{1}{y} - 1 \right) = 3.$$

Note that

$$\begin{aligned} \sigma^2 &= \mathrm{E}\left( (Y - \mu_Y)^2 \right) = \mathrm{E}\left( Y^2 - 2Y\mu_Y + \mu_Y^2 \right) \\ &= \mathrm{E}\left(Y^2\right) - 2\,\mathrm{E}\left(Y\right)\mu_Y + \mu_Y^2 = \mathrm{E}\left(Y^2\right) - 2\mu_Y\mu_Y + \mu_Y^2 \\ &= \mathrm{E}\left(Y^2\right) - \mu_Y^2. \end{aligned}$$

Then, $\sigma^2 = \mathrm{E}\left(Y^2\right) - \mu_Y^2 = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4}$.

- Example 4 shows that $\sigma_Y^2 = \mathrm{E}((Y - \mu_Y)^2) = \mathrm{E}(Y^2) - \mu_Y^2$.

## Two RVs

- In this course, except for Chapters 2 and 3, you will deal with scalar RVs on occasion. Often, you will encounter vectors containing scalar RVs.
- Suppose we are dealing with two RVs. One way to characterize these two RVs is using what we call their joint probability distribution.
- From the joint distribution, one can calculate the marginal probability distribution for each RV.
- From the joint and the marginal distributions, we can calculate the conditional probability distribution for one RV conditional on the other RV.

**Definition 9**

The joint probability distribution of two discrete random variables $X$ and $Y$ is the probability that the random variables simultaneously take on certain values, say $x$ and $y$. The joint probability distribution is written as $P(X = x, Y = y)$.

## Two RVs

- Let $Y$ be a binary random variable that equals $1$ if the commute to school is short (less than 20 minutes) and that equals $0$ otherwise, and let $X$ be a binary random variable that equals $0$ if it is raining and $1$ if not.
- Assume that the joint probability distribution of $Y$ and $X$ is given in Table 2.

Table 2: Joint probability distribution of $Y$ and $X$

| | Rain $(X = 0)$ | No Rain $(X = 1)$ | Total |
|---|---|---|---|
| Long commute $(Y = 0)$ | 0.15 | 0.07 | 0.22 |
| Short commute $(Y = 1)$ | 0.15 | 0.63 | 0.78 |
| Total | 0.30 | 0.70 | 1.00 |

- According to the table, the likelihood of short commute on a rainy day is $15\%$, i.e., $P(Y = 1, X = 0) = 0.15$.
- These four possible outcomes are mutually exclusive and constitute the sample space, so four probabilities sum to $1$:

$$\sum_{i=1}^{k}\sum_{j=1}^{l} P(Y = y_i, X = x_j) = 1.$$

## Two RVs

- The marginal probability distribution of $Y$ can be computed from the joint distribution of $X$ and $Y$ by adding up the probabilities of all possible outcomes for which $Y$ takes on a specified value.
- If $X$ can take on $l$ different values $x_1, \ldots, x_l$, then the marginal probability that $Y$ takes on the value $y$ is

$$P(Y = y) = \sum_{i=1}^{l} P(Y = y, X = x_i). \tag{6}$$

- In Table 2, the row sums give the marginal probability distribution of $Y$, while the column sums give the marginal probability distribution of $X$.

**Example 5**

Consider Table 2. What is the probability of a long commute, i.e., $P(Y = 0)$?

$$P(Y = 0) = P(Y = 0, X = 0) + P(Y = 0, X = 1) = 0.15 + 0.07 = 0.22.$$

## Two RVs

- Use the following formula to calculate the conditional probabilities when you are dealing with two discrete RVs.

$$P(Y = y|X = x) = \frac{P(Y = y, X = x)}{P(X = x)}. \tag{7}$$

- In other words, $P(Y = y|X = x)$ refers to the likelihood of the event that $Y$ takes the value $y$ given that $X = x$ has happened.

**Example 6**

Consider the RVs in Table 2. What is the likelihood of a short commute given that it is raining outside?

$$P(Y = 1|X = 0) = \frac{P(Y = 1, X = 0)}{P(X = 0)} = 0.15/0.30 = 0.50.$$

What is the likelihood of no rain outside given that it has been a short commute?

$$P(X = 1|Y = 1) = \frac{P(X = 1, Y = 1)}{P(Y = 1)} = 0.63/0.70 = 0.90.$$

## Two RVs

- Using the conditional distribution (the list of conditional probabilities), one can calculate the conditional moments.
- For example, conditional expectation of $Y$ given $X$, also called the conditional mean of $Y$ given $X$, is the mean of the conditional distribution of $Y$ given $X$.
- We use the following notation $\mathrm{E}(Y|X)$ to denote the conditional mean of $Y$ given $X$.
- If you do not specify the value $X$ takes on then $\mathrm{E}(Y|X)$ becomes random.
- If you specify the value $X$ takes on, say $X = x$ then $\mathrm{E}(Y|X = x)$ is not random.
- Suppose both $Y$ and $X$ are discrete and $Y$ takes on $k$ values, $y_1, y_2, \ldots, y_k$. Then,

$$\mathrm{E}(Y|X = x) = \sum_{j=1}^{k} y_j P(Y = y_j|X = x). \tag{8}$$

## Two RVs

- A similar argument applies to higher moments of $Y$ given $X$. For example, we can construct the conditional variance of $Y$ given $X$.

$$\mathrm{var}(Y|X=x) = \sum_{j=1}^{k} \big(y_j - \mathrm{E}(Y|X=x)\big)^2 P(Y=y_j|X=x). \tag{9}$$

**Example 7**

Consider the RVs in Table 2. What is the expected value of commute given that it is not a rainy day? What is the variance of commute given that it is not a rainy day?

$$\mathrm{E}(Y|X=1) = 1 \times P(Y=1|X=1) + 0 \times P(Y=0|X=1) = 0.63/0.70 = 0.9.$$

$$\begin{aligned}\mathrm{var}(Y|X=1) &= (1-0.9)^2 \times P(Y=1|X=1) + (0-0.9)^2 \times P(Y=0|X=1)\\ &= (0.1)^2 \times 0.9 + (-0.9)^2 \times 0.1 = 0.09.\end{aligned}$$

## Two RVs

- Indeed, the unconditional expectation of $Y$, i.e., $E(Y)$, is the weighted average of the conditional expectations of $Y$ given $X$, where the weights come from the marginal distribution of $X$.

$$\mathrm{E}(Y) = \sum_{i=1}^{l} \mathrm{E}(Y|X = x_i) P(X = x_i) = \mathrm{E}\left(\mathrm{E}(Y|X)\right). \tag{10}$$

where the outer expectation is with respect to the marginal distribution of $X$.

- This result in (10) is known as the law of iterated expectations (LIE).

1. The LIE implies that if $\mathrm{E}(Y|X) = 0$, then $\mathrm{E}(Y) = \mathrm{E}\left(\mathrm{E}(Y|X)\right) = \mathrm{E}(0) = 0$.
2. The LIE also holds for multiple RVs. Let $X$, $Y$ and $Z$ be RVs that are jointly distributed. Then

$$\mathrm{E}(Y) = \mathrm{E}\left(\mathrm{E}(Y|X, Z)\right). \tag{11}$$

where the outer expectation is with respect to the joint distribution of $X$ and $Z$.

## Two RVs: Independence

- Two random variables $X$ and $Y$ are independently distributed, or independent, if knowing the value of one of the variables provides no information about the other.

- In other words, $X$ and $Y$ are independently distributed if, for all values of $x$ and $y$, we have

$$P(Y=y|X=x)=P(Y=y) \text{ or } P(Y=y,X=x)=P(Y=y)P(X=x).$$

- In econometric analysis, we will often be interested in the extent to which two random variables move together (i.e., a measure of linear association).

- The covariance between $X$ and $Y$ is simply given by

$$\begin{aligned} \operatorname{cov}(X,Y)=\sigma_{XY} &= \mathrm{E}\left((X-\mu_X)(Y-\mu_Y)\right) \\ &= \mathrm{E}(XY)-\mu_X\mu_Y. \end{aligned} \tag{12}$$

1. Note that the covariance of $X$ with its self is its variance, i.e., $\operatorname{cov}(X,X)=\sigma_X^2$.
2. If $X$ and $Y$ are independent, then $\operatorname{cov}(X,Y)=0$, because $\operatorname{cov}(X,Y)=\mathrm{E}(XY)-\mu_X\mu_Y=\mathrm{E}(X)\,\mathrm{E}(Y)-\mu_X\mu_Y=0$.

## Two RVs: Covariance

- Suppose again both $Y$ and $X$ are discrete and $Y$ takes on $k$ values and $X$ takes on $l$ values. Then,

$$\text{cov}(X,Y) = \sum_{i=1}^{k}\sum_{j=1}^{l}(x_j - \mu_X)(y_i - \mu_Y)P(X = x_j, Y = y_i).$$

- Since the covariance is the product of $X$ and $Y$, deviated from their respective means, its units are the units of $X$ multiplied by the units of $Y$.
- This can make the interpretation of covariance difficult. To get around this problem, we introduce the correlation measure:

$$\text{corr}(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\,\text{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}. \qquad (13)$$

- The correlation measure is unit free and always $-1 \leq \text{corr}(X,Y) \leq 1$.

## Two RVs: Covariance

### Example 8

If the conditional mean of $Y$ does not depend on $X$, then $\text{cov}(Y, X) = \text{corr}(Y, X) = 0$, i.e., if $\mathrm{E}(Y|X) = \mu_Y$, then $\text{cov}(Y, X) = \text{corr}(Y, X) = 0$.

### Proof of Example 8.

First, suppose that $Y$ and $X$ have mean $0$ so that $\text{cov}(X, Y) = \mathrm{E}(YX)$. Then, by LIE,

$$\mathrm{E}(YX) = \mathrm{E}\left(\mathrm{E}(YX|X)\right) = \mathrm{E}\left(\mathrm{E}(Y|X)X\right) = \mu_Y \times \mathrm{E}\left(X\right) = 0,$$

because $\mathrm{E}(X) = 0$. Thus, $\text{cov}(X, Y) = \mathrm{E}(YX) = 0$. Also, from (13), we have $\text{corr}(Y, X) = 0$. If $X$ and $Y$ do not have mean $0$, subtract off their means, and then the preceding proof applies. □

## Sums of RVs

- Let $X$, $Y$, and $V$ be scalar random variables and let $a$, $b$ and $c$ be constants. Then, the following results hold:

1. $\mathrm{E}(a + bX + cY) = a + b\mu_X + c\mu_Y$,
2. $\mathrm{var}(a + bY) = b^2\sigma_Y^2$,
3. $\mathrm{var}(aX + bY) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_X\sigma_Y$,
4. $\mathrm{E}(Y^2) = \sigma_Y^2 + \mu_Y^2$,
5. $\mathrm{cov}(a + bX + cY, V) = b\sigma_{XV} + c\sigma_{YV}$,
6. $\mathrm{E}(XY) = \sigma_{XY} + \mu_X\mu_Y$,
7. $|\,\mathrm{corr}(X, Y)| \leq 1$ and $|\sigma_{XY}| \leq \sigma_X\sigma_Y$.

## Example 9

Let $X$ and $Y$ be discrete random variables. Table 3 presents the joint probabilities. For example, $P(X = 1, Y = 14) = 0.02$, and so forth.

Table 3: Joint distribution of $X$ and $Y$

| | | Value of Y | | | | |
|---|---|---|---|---|---|---|
| | | 14 | 22 | 30 | 40 | 65 |
| Value of X | 1 | 0.02 | 0.05 | 0.10 | 0.03 | 0.01 |
| | 5 | 0.17 | 0.15 | 0.05 | 0.02 | 0.01 |
| | 8 | 0.02 | 0.03 | 0.15 | 0.10 | 0.09 |

1. Calculate the probability distribution, mean and variance of $Y$.
2. Calculate the probability distribution, mean and variance of $Y$ given $X = 8$.
3. Calculate the covariance and correlation between $X$ and $Y$.

## Solution of Example 9

- Using $P(X = x_i) = \sum_{j=1}^{l} P(X = x_i, Y = y_j)$ and $P(Y = y_j) = \sum_{i=1}^{k} P(Y = y_j, X = x_i)$, we obtain the marginal distributions given in Table 4.

Table 4: Marginal distributions of $Y$ and $X$

| | **14** | **22** | **30** | **40** | **65** | | **1** | **5** | **8** |
|---|---|---|---|---|---|---|---|---|---|
| Marginal distribution of Y | 0.21 | 0.23 | 0.30 | 0.15 | 0.11 | Marginal distribution of X | 0.21 | 0.40 | 0.39 |

- Part 1: Using the marginal distribution of $Y$:

$$\mathrm{E}(Y) = 14 \times 0.21 + 22 \times 0.23 + 30 \times 0.30 + 40 \times 0.15 + 65 \times 0.11 = 30.15$$
$$\mathrm{E}(Y^2) = 14^2 \times 0.21 + 22^2 \times 0.23 + 30^2 \times 0.30 + 40^2 \times 0.15 + 65^2 \times 0.11 = 1127.23$$
$$\implies \text{Var}(Y) = \mathrm{E}(Y^2) - (\mathrm{E}(Y))^2 = 218.21.$$

- Part 2: Using $P(Y = y_j | X = 8) = \frac{P(Y=y_j, X=8)}{P(X=8)}$, we obtain

Table 5: Distribution of $Y|X = 8$

| Distribution of $Y\|X = 8$ | | | | |
|---|---|---|---|---|
| 14 | 22 | 30 | 40 | 65 |
| 0.02/0.39 | 0.03/0.39 | 0.15/0.39 | 0.10/0.39 | 0.09/0.39 |

## Solution of Example 9 continue...

- Part 2:

$$\mathrm{E}(Y|X=8) = 14 \times (0.02/0.39) + 22 \times (0.03/0.39) + 30 \times (0.15/0.39)$$
$$+ 40 \times (0.10/0.39) + 65 \times (0.09/0.39) = 39.21,$$
$$\mathrm{E}(Y^2|X=8) = 14^2 \times (0.02/0.39) + 22^2 \times (0.03/0.39) + 30^2 \times (0.15/0.39),$$
$$+ 40^2 \times (0.10/0.39) + 65^2 \times (0.09/0.39) = 1778.7$$
$$\implies \mathsf{Var}(Y|X=8) = \mathrm{E}(Y^2|X=8) - (\mathrm{E}(Y|X=8))^2 - 1778.7 - (39.21)^2 = 241.65.$$

- Part 3: We need to first compute $\mathrm{E}(XY) = \sum_j^k \sum_i^l x_i y_j P(X = x_i, Y = y_j)$:

$$\mathrm{E}(XY) = (1 \times 14 \times 0.02) + (1 \times 22 \times 0.05) + \ldots + (8 \times 65 \times 0.09) = 171.7,$$
$$\mathsf{cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y) = 171.7 - 5.33 \times 30.15 = 11.0,$$
$$\mathsf{corr}(X,Y) = \mathsf{cov}(X,Y)/(\sigma_X \sigma_Y) = 11.0/(2.60 \times 14.77) = 0.286.$$

## Bibliography I

Hanck, Christoph et al. (2021). *Introduction to Econometrics with R.* URL: https://www.econometrics-with-r.org/index.html.

Stock, James H. and Mark W. Watson (2020). *Introduction to Econometrics.* Fourth. Pearson.