Ch1.5 - Probability Theory (Short Notes)

(*ECO 104 Recap)*

ECO 204
Statistics For Business and Economics - II

Shaikh Tanvir Hossain

East **W**est **U**niversity, Dhaka
Last updated: February 24, 2025

Outline

Outline

1. **Random Variable**
   - Discrete and Continuous Random Variables and Probability Distributions
   - Expectation, Variance and Some Rules
2. **Theoretical Distributions**
   - Bernoulli and Normal Distribution
3. **Appendix: Notation List**

Recap Slides

- This is chapter 1.5, we already started talking about random sample, but I realized we need some topics from ECO104 in this course, so that's why this quick and dirty recap.
- In particular we will need
  - 1) Some understanding of basic descriptive statistics, e.g., sample mean, median, variance, standard deviation, etc, also what is bar chart, histogram, etc. Since we have already discussed this in the class, I won't go over this here
  - 2) Some understanding of the contents related to Probability Theory, in particular,
    - random variable, discrete and continuous random variable,
    - probability distribution of a random variable
    - probability mass function (pmf) and probability density function (pdf)
    - CDF and quantiles
    - expectation and variance of a distribution
    - some expectation and variance Rules
    - joint, marginal and conditional distribution.
    - conditional expectation
- Please don't get scared, this seem a lot, but believe me, not hard, so let's tart

"Probability theory is nothing but common sense reduced to calculation."

— *Pierre-Simon Laplace*

Random Variable

Discrete and Continuous Random Variables and Probability Distributions

Random Variables and Probability Distributions

- A **random variable** is something which can randomly take different values, the randomness is coming from an *experiment*, as concrete example you can think *taking a random EWU student from all EWU students* and writing his / her monthly income.
- Here the experiment is taking a random student and the random variable is the monthly income of the student, and we will write this with X

sl.	Monthly Family Income (tk)	R.V.
1.	70,150	$X = ?$

- *or taking a random EWU student from all EWU students* and writing his / her gender, and here the random variable is gender (We will code Male with 0 and Female with 1)

sl.	Gender (Male: 0 and Female: 1)	R.V.
1.	0	$X = ?$

- Note, the idea of the random variable comes before performing the experiment, after we have observed the monthly income / gender of the student, we will know the value of the random variables, that is we know $X = 70,150$ or $X = 1$. Then X is not random anymore, it has a fixed value.

Random Variables and Probability Distributions

- *Ques: If we think the random variable X is monthly income, then what are all possible values of X*
- *Ans: Whatever possible income values are available in the population.*
- In reality X can take only finite number of values, because our population is finite. However when the population is large in Statistics we will often assume our

  population is infinite

- You might *protest* by saying - *"What nonsense? How can we assume population is infinite, when it is not"*, well, this is a simplifying assumption to makes our life easier. Reality is hard to model, so often in Statistics and also in Economics we make many simplifying assumptions and then we have nice theoretical results.

Random Variables and Probability Distributions

- When our random variable X has infinite possible values we say it is a **continuous random variable**, for example it could happen that the monthly income of a student X can be any value in the interval

  $$X \text{ can take any value in the interval } [0, \infty) \text{ or } [0, 1000) \text{ or } (-\infty, 1000) \text{ or even } (-\infty, \infty)$$

- Again X is a random variable, before performing the experiment, you know all *possible values*, but you don't know exactly which value will be *observed or realized*.

Random Variables and Probability Distributions

- A random variable can be either **discrete** or **continuous**, here the random variable X, which is the monthly income, is a continuous random variable, because it can take any value in an interval on the real line.
- We have already given an example of a *discrete random variable* at the beginning of the slides, that is *gender of a student*, which can take only two values 0 (for Male) and 1 (for female), and nothing in between. Note here we coded Female and Male to 1 and 0 respectively (as a side note, notice this kind of encoding categories in numbers a common practice in Statistics)
- You know what is an *interval* in the real line, right? For example, $[0,2]$ is an interval, $[0,2)$ is an interval, $(-\infty,2]$ is an interval, $(-\infty,\infty)$ is also an interval.

Random Variables and Probability Distributions

- Now let's talk about *probability distribution* of a random variable.
- Roughly the probability distribution of a random variable gives *how the probabilities are distributed over all possible values of the random variable*.
- If X is a discrete random variable, the idea of probability distribution of X is an easy concept to understand, for example if X takes value $0, 1$, only two values, and we know

$$\mathbb{P}(X = 0) = 0.7$$
$$\mathbb{P}(X = 1) = 0.3$$

  then we know the probability distribution of X (notice the probability has to be summed to 1).
- You might question - *"How are these probabilities calculated?"*
- You can this is coming from the population. So in the entire population 70% of the students are Make and 30% are Female, so we get the probabilities from there.
- Here the distribution of probabilities also tell us, if we randomly pick a student, then there is 70% chance that the student will be male and 30% chance that the student will be female.
- Keep in Mind this these probabilities are calculated in population, there is no sample here.

Random Variables and Probability Distributions

- Also important only with this information, we can also calculate probabilities of certain events, where the random variable will take value within certain interval, e.g.,

  $$\mathbb{P}(-5 < X < 0.5) \text{ or } \mathbb{P}(X > 0.5)$$

  In this case

  $$\begin{aligned} \mathbb{P}(-0.5 < X < 0.5) &= 0.7, \\ \mathbb{P}(X > 0.5) &= 0.3. \end{aligned}$$

- Why? or how did we calculated this?

Random Variables and Probability Distributions

- When X is a continuous random variable, the idea of probability distribution is a bit tricky, because in this case although X can take any value in an interval but the *probability of X taking a specific value is 0*, i.e.,

$$\mathbb{P}(X = a) = 0 \text{ no matter whatever } a \text{ is}$$

- Why? Intuitively there are so many possibilities that we cannot specifically calculate $\mathbb{P}(X = a)$ for any fixed a, so just set it to 0
- In this case the probability distribution of X can be represented using a function called *probability density function* (pdf), which is a function of all values that the random variable can take, i.e., it's a function of x, for example it could be that the *density function of a continuous random variable X* looks like this (notice it's just a function, you know what is a function right?)

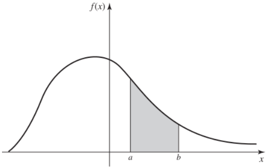


Figure 1: density of function of a continuous random variable

- Notice it's a function of x (where x is a value of the random variable X)
- Now how do we calculate probability of X takes value in the interval $[a, b]$, in particular how do we calculate $\mathbb{P}(a < X < b) =?$
- The idea is for a continuous random variable we *integrate the density function in that region*, so

$$\mathbb{P}(a < X < b) = \int_a^b f(x)dx$$

- So probability in any range is just an integration (we will see why in a minute?)
- Also since *definite integration means finding area under the curve*, so the probability of X taking value in the interval $[a, b]$ will be equal to the area under the density function $f(x)$ in the interval $[a, b]$, this means

$$\mathbb{P}(a < X < b) = \int_a^b f(x)dx = \text{Area under the function } f(x) \text{ in the interval } [a, b]$$

Random Variables and Probability Distributions

- This is what you have been using in ECO 104, recall, when you are asked to calculate certain probability then you were looking at some table, the table has some specific probabilities in certain interval, so this means someone did some kind of integration for you and you are just using the result.
- *Relax*, I won't ask you to do integration in this course, you just have to understand that
  - probability in any interval for a continuous random variable can be calculated using *definite integration*,
  - and integration is area under the curve, so probability is area under the curve or the function.
- We will see some examples of probabilities for a normally distributed random variable soon, then you will understand this better.

- Now let's explain why for a continuous random variable X, when we find probability in any interval we integrate,
- Recall when X is a discrete random variable takes values $1, 2$ and 3, then we can calculate,

$$\begin{aligned}\mathbb{P}(1 \leq X \leq 3) &= \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3) \\ &= \sum_{i=1}^{3} \mathbb{P}(X = a_i),\end{aligned}$$

- where $a_1 = 1, a_2 = 2$, and $a_3 = 3$, this means it's a sum of probabilities of X taking specific values.
- For a continuous random variable, we cannot just summation if X takes value in the interval $[1, 3]$, since there are infinite values X can take, then in this case, *the summation becomes integral*, but we integrate the density function, and we get

$$\mathbb{P}(1 < X < 3) = \int_1^3 f(x)dx$$

- As a passing note, note that when we are talking about a continuous random variable,

$$\mathbb{P}(1 < X < 3) = \mathbb{P}(1 \leq X \leq 3)$$

- However this is not the case for a discrete random variable, can you think why???

Random Variable

Expectation, Variance and Some Rules

- The idea of an expected value or in short expectation is very similar to population mean (note the idea of a distribution is how probabilities are distributed over the population)
- The difference from the average is - *we give proper weights to the values*, Here is the definition,

Definition 1-5.1: (Expected Value)

If X is a *discrete random variable* taking values $a_1, a_2, \ldots, a_N$, then the *Expectation* or the *Expected Value* of X is defined as

$$\mathbb{E}(X) = \sum_{i=1}^{N} a_i \mathbb{P}(X = a_i)$$

If X is a *continuous random variable* takes value within $(-\infty, \infty)$, where the density function is $f(x)$ then the expected value is,

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

We will usually use the notation $\mathbb{E}(\cdot)$ to denote that we are performing *Expectation* on X.

- You should always remember the expectation sign works on random variable, not fixed number

Expected Value or Mean is a constant

$$\mathbb{E}(X) = 2.5$$

Random Variable (which is NOT a constant, and it possible that it takes many values). But if we take expectation, we are asking for its Mean value. The Mean or Expected value is always a constant.

- SO never write $\mathbb{E}(2.5)$, this is wrong!

- So let's see an example... assume X takes values $0, 1, 2, 3$, and the probability distribution is given by

$$\mathbb{P}(X=0)=1/8, \quad \mathbb{P}(X=1)=3/8, \quad \mathbb{P}(X=2)=3/8, \quad \mathbb{P}(X=3)=1/8$$

- So we can calculate the expected value as,

$$\mathbb{E}(X)=(0\times 1/8)+(1\times 3/8)+(2\times 3/8)+(3\times 1/8)=1.5$$

- So calculation is very easy, now we may ask *what does expected value mean*? As we said, The Expectation (or Expected value) is like average, but it is for a population, so it's a *population average* or *population mean*,

- Now let's see how to calculate the expected value of a continuous random variable X, suppose X is a continuous random variable taking values in $(-\infty, \infty)$ and it has following density function,

$$f(x) = \begin{cases} 2x & \text{for } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

- We can calculate the expected value (just by *replacing sum with integration!*)

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 x(2x)dx = \int_0^1 2x^2 dx \quad \ldots$$
$$\ldots = 2\int_0^1 x^2 dx = 2 \times \left[\frac{x^3}{3}\right]_0^1 = \frac{2}{3} \times \left[x^3\right]_0^1 \quad \ldots$$
$$\ldots = \frac{2}{3} \times \left[1^3 - 0^3\right] = \frac{2}{3} \times 1 = \frac{2}{3}.$$

- Like Expectation, variance is also a summary measure, where the expectation gives an idea of the central value, variance gives the idea how *dispersed the values are*.

Definition 1-5.2: (Variance)

If X is a discrete random variable with taking values $a_1, a_2, \ldots, a_N$, then the *Variance* of X is defined as

$$\mathbb{V}(X) = \mathbb{E}\left(\left(X - \mathbb{E}\left(X\right)\right)^2\right) = \sum_{i=1}^{N} (a_i - \mathbb{E}(X))^2 \mathbb{P}(X = a_i)$$

If X is a continuous random variable with density function $f(x)$, then the *Variance* of X is defined as

$$\mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 f(x) dx.$$

- First note that, Variance is also an Expectation, but it is an *Expectation of* $(X - \mathbb{E}(X))^2$, NOT X, here $\mathbb{E}(X)$ is just a constant that we calculated before
- The interpretation of the variance is already same and you already know it,

- Let's calculate $\mathbb{V}(X)$ for the random variable X, which counts the number of heads (PMF in page 27).

$$\begin{aligned}\mathbb{V}(X) = &\left((0-1.5)^2 \times \mathbb{P}(X=0)\right) + \left((1-1.5)^2 \times \mathbb{P}(X=1)\right) + \\ &\left((2-1.5)^2 \times \mathbb{P}(X=2)\right) + \left((3-1.5)^2 \times \mathbb{P}(X=3)\right) \\ = &\left((-1.5)^2 \times 1/8\right) + \left((-0.5)^2 \times 3/8\right) + \left((0.5)^2 \times 3/8\right) + \left((1.5)^2 \times 1/8\right) \\ = &(2.25 \times 1/8) + (0.25 \times 3/8) + (0.25 \times 3/8) + (2.25 \times 1/8) \\ = &0.75\end{aligned}$$

- So calculating Variance is really easy
- Like discrete random variables we can also calculate Expected values and Variance for a continuous random variable.
- The expectation and the variance of a continuous random variable can be calculated the same way we did for discrete, however, we need *Integration*, how do you do this?

Theoretical Distributions

Bernoulli and Normal Distribution

Theoretical Distributions - Bernoulli and Normal

- Now let's see examples of two theoretical distributions of, these distributions will play important roles in the coming chapter.
  - For *Discrete Random Variables* - *Bernoulli Distribution*
  - For *Continuous Random Variables* - *Normal Distribution*
- You may ask *"Why we learn theoretical distributions?"*

  This is because based on the nature of the random variable coming from the population, we can model the population data (or random variable) with different theoretical distributions. and then learn about them!

- An important point: For any theoretical distributions there is a thing called *parameter* which will control the distribution, and you should remember changing the parameter will change the distribution, of course we don't the parameter, since it's a population quantity, but we can estimate this of course.
- Let's see more about *parameter* and *distributions*....

Theoretical Distributions - Bernoulli and Normal

Bernoulli Distribution with Parameter p **(Ber(p))**

When a random variable X follows Bernoulli distribution, then X takes only two values 0 and 1, and the probability of X taking value 1 is p, i.e.,

$$\mathbb{P}(X = 1) = p \quad \text{and} \quad \mathbb{P}(X = 0) = 1 - p$$

We write this as

$$X \sim \text{Bern}(p)$$

- Note that, changing the parameter will change the distribution, for example it could be that $p = 0.3$, then we know

$$\mathbb{P}(X = 1) = 0.3 \quad \text{and} \quad \mathbb{P}(X = 0) = 0.7$$

- Or, $p = 0.5$, then we have

$$\mathbb{P}(X = 1) = 0.5 \quad \text{and} \quad \mathbb{P}(X = 0) = 0.5$$

- Lastly you should read the notation $X \sim \text{Bern}(p)$ as *"X is distributed as Bernoulli with parameter p"*
- **Change of Notation:** previously I wrote Ber(p), although this is also written sometimes, but I found Bern(p) is more common, if you write with Binomial then you would write B(n, p) (what is this?)

- Now if $X \sim \text{Bern}(p)$, then we can easily calculate

$$\mathbb{E}(X) = p, \quad \mathbb{V}(X) = p(1-p)$$

- Can you do the calculation?
- First let's do the expectation,

$$\mathbb{E}(X) = 1 \times p + 0 \times (1-p) = p$$

- Let's do the variance calculation,

$$\mathbb{V}(X) = (1-p)^2 \times p + (0-p)^2 \times (1-p) = p(1-p)$$

- Now let's see the famous *Normal Distribution*,

Normal Distribution with Parameters μ and σ^2

A random variable X is said to follow a normal distribution with parameters μ and σ^2, if the density function of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \; e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

We write this as

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

- Here for any population, there is a fixed μ and σ^2, and then the density function will only be a function of x, for example if we fix $mu = 0$ and $\sigma^2 = 1$ (which means $\sigma = 1$), then we get

$$f(x) = \frac{1}{\sqrt{2\pi \times 1}} \; e^{-\frac{1}{2}\left(\frac{x-0}{1}\right)^2} = \frac{1}{\sqrt{2\pi}} \; e^{-\frac{1}{2}x^2}$$

which is the density function of the famous distribution, known as *Standard Normal Distribution*, and the density function will look like,

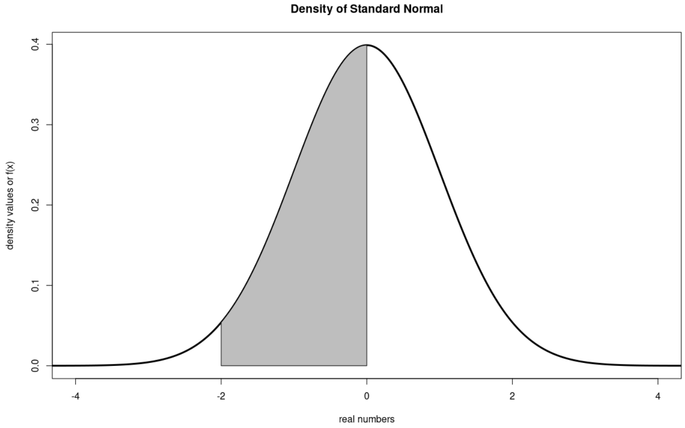
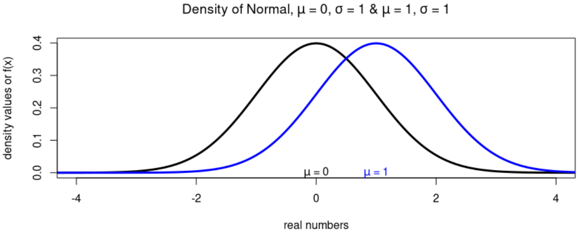
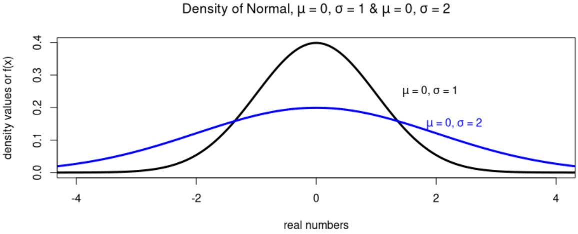


Figure 2: density function of the standard normal distribution, this is the density function for $X \sim \mathcal{N}(0, 1)$, also we showed the shaded area which is $\mathbb{P}(-2 < X < 0)$

Note that changing μ and σ, will change the shape of the density function and also change the probability and hence will change the distribution, following picture will reveal this,





If $X \sim \mathcal{N}(\mu, \sigma^2)$, then we can calculate

$$\mathbb{E}(X) = \mu, \quad \mathbb{V}(X) = \sigma^2$$

In this case we need to do integration, but you don't need to do this. Just keep in mind that μ adn σ^2 are the parameters of the distribution, and they control the distribution, and we can calculate $\mathbb{E}(X) = \mu$ and $\mathbb{V}(X) = \sigma^2$.

Now let's see how do we calculate the probabilities in the normal distribution using R, here is a problem

Example 1-5.3: Suppose we have a random variable, $X \sim \mathcal{N}(5, 1)$, calculate $\mathbb{P}(0 < X < 3)$

In words: This means we have a random variable which is normally distributed with mean 5 and variance 1, what is the probability that X takes value in the interval $(0, 3)$,

Ans: In R for this problem, first we calculate following cumulative probabilities, $\mathbb{P}(-\infty < X < 3)$ and $\mathbb{P}(-\infty < X < 0)$ using the `pnorm()` function, and then we take the difference,

R code - Calculate Probability in Normal Distribution

```
options(scipen = 999) # stop the scientific number


# First create some objects with the information given
mu <- 5
sigma <- 1

pnorm(3, mu, sigma)
# [1] 0.02275013

pnorm(0, mu, sigma)
# [1] 0.0000002866516

# Now we calculate the cumulative probabilities and take the difference directly
pnorm(3, mu, sigma) - pnorm(0, mu, sigma)
```

Now you can ask *why we calculated cumulative probabilities? Can't we get this directly?* The answer is NO, because the only option available in Excel / R / STATA, is calculating cumulative probabilities using the `pnorm` "type".

Definitely if you change the parameters μ and σ, you will get different probabilities.

Theoretical Distributions - Bernoulli and Normal

In R we will frequently use following functions,

- `dnorm()` - to calculate density function
- `pnorm()` - to calculate cumulative probabilities
- `qnorm()` - to calculate quantiles
- `rnorm()` - to generate random numbers

Important is in these kinds of functions you need to give the parameters of the distribution. You will do some examples in the first problem set.

Now *a question: how do you calculate the probability $\mathbb{P}(-\infty < X < 3)$ using the table in the back of your book?* The answer is you are continuously using following relation between *a standard normal distribution* and *a normal distribution with any mean and variance*,

Relationship between Standard Normal and Normal Distribution

If X is a random variable, such that

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

then we can transform X with another random variable Z, where

$$Z = \frac{X - \mu}{\sigma} \text{ such that } Z \sim \mathcal{N}(0, 1)$$

Here we apply the formula,

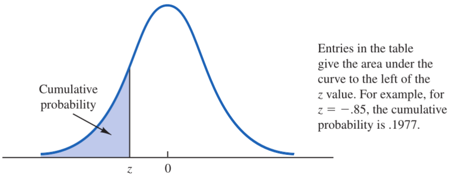
$$z = \frac{x - 5}{1} = -2$$

So if 3 is a value of the random variable $\mathcal{N}(5, 1)$, then we have an equivalent z value, which is -2 that is coming from $\mathcal{N}(0, 1)$. Now we calculate $\mathbb{P}(-\infty < Z < -2)$, for $Z \sim \mathcal{N}(0, 1)$. Can you see the two values are equal visually? (On board!)

In the table the probability $\mathbb{P}(-\infty < Z < -2) = 0.228$ is

Theoretical Distributions - Bernoulli and Normal

TABLE 1 Cumulative Probabilities for the standard Normal Distribution



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183

Figure 3: Standard Normal Distribution Table directly taken from the Anderson's book

In R if you use the function `pnorm(3, 5, 1)` you will get the same value, interestingly you can also use the function `pnorm(-2, 0, 1)`, and you will get the same value.

There are some problems in the problem set, please solve them, the next topic for is the sampling distribution of sample means, important thing to note here is

- Whether we have Normality assumption in the data
- Whether we assume independence observations
- Whether the sample size is large enough to apply to Central Limit Theorem.

We will continue with Chapter 1 now and I will discuss the details there....

Appendix: Notation List

List of Notations

Here is the list of notations that we have used in this chapter

- $\mathbb{P}(A)$ - Probability of an event A, for example,

- X Random Variable, x value of a random variable

- $X_1, X_2, \ldots, X_n$ - Sequence of Random Variables

- $\mathbb{E}(X), \mathbb{V}(X)$ - Expectation and Varince of a random variable X

- μ, σ^2 - Population Mean and Variance (Although more appropriate for Normal only, but frequently used for others)

- $f(x)$ - Density function of a random variable X (note that it's a function of x not X)

References

Abraham, B. and Ledolter, J. (2006), *Introduction to Regression Modeling*, Duxbury applied series, Thomson Brooks/Cole, Belmont, CA.

Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., Cochran, J. J., Fry, M. J. and Ohlmann, J. W. (2020), *Statistics for Business & Economics*, 14th edn, Cengage, Boston, MA.

Bertsekas, D. and Tsitsiklis, J. N. (2008), *Introduction to probability*, 2nd edn, Athena Scientific.

Blitzstein, J. K. and Hwang, J. (2015), *Introduction to Probability*.

Casella, G. and Berger, R. L. (2002), *Statistical Inference*, 2nd edn, Thomson Learning, Australia ; Pacific Grove, CA.

DeGroot, M. H. and Schervish, M. J. (2012), *Probability and Statistics*, 4th edn, Addison-Wesley, Boston.

Hansen, B. (2022), *Econometrics*, Princeton University Press, Princeton.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2023), *An introduction to statistical learning*, Vol. 112, Springer.

Newbold, P., Carlson, W. L. and Thorne, B. M. (2020), *Statistics for Business and Economics*, 9th, global edn, Pearson, Harlow, England.

Pishro-Nik, H. (2016), *Introduction to probability, statistics, and random processes*.

Ramachandran, K. M. and Tsokos, C. P. (2020), *Mathematical Statistics with Applications in R*, 3rd edn, Elsevier, Philadelphia.

References

Rice, J. A. (2007), *Mathematical Statistics and Data Analysis*, Duxbury advanced series, 3rd edn, Thomson/Brooks/Cole, Belmont, CA.

Wasserman, L. (2013), *All of statistics: a concise course in statistical inference*, Springer Science & Business Media.

Wooldridge, J. M. (2009), *Introductory Econometrics: A Modern Approach*, 4th edn, South Western, Cengage Learning, Mason, OH.