Math 447 - Probability

Dikran Karagueuzian

SUNY-Binghamton

BINGHAMTON UNIVERSITY

STATE UNIVERSITY OF NEW YORK

Chapter 1

Introduction

Introduction

Statistical techniques are employed in almost every phase of life.

- Surveys are designed to collect early returns on election day and forecast the outcome of an election.
- Consumers are sampled to provide information for predicting product preferences.
- Research physicians conduct experiments to determine the effect of various drugs and controlled environmental conditions on humans in order to infer the appropriate treatment for various illnesses.
- Engineers sample a product quality characteristic and various controllable process variables to identify key variables related to product quality.
- Newly manufactured electronic devices are sampled before shipping to decide whetherto ship or hold individual lots.
- Economists observe various indices of economic health over a period of time and use the information to forecast the condition of the economy in the future.

Statistical techniques play an important role in achieving the objective of each of these practical situations.

The objective of statistics is to make an *inference* about a population based on information contained in a sample from that population and to provide an associated measure of goodness for the inference.

In the broadest sense, making an inference implies partially or completely describing a phenomenon or physical object.
Little difficulty is encountered when appropriate and meaningful descriptive measures are available, but this is not always the case.

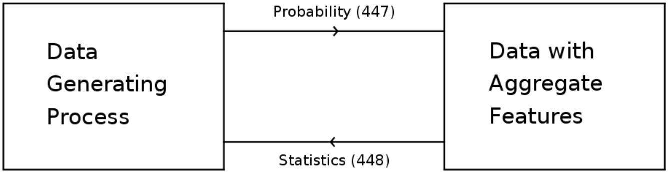
We can characterize the available data

- Graphically, e.g. using a histogram to plot relative frequencies of, say, GPAs of students in the class,
  or
- Numerically, e.g. finding the average annual rainfall in California over the past 50 years and the deviation from this average quantity in a particular year.

We may also be interested in the *likelihood* of a certain event, e.g. drawing the Royalty (King and Queen) of *different* suits from a standard deck of cards.

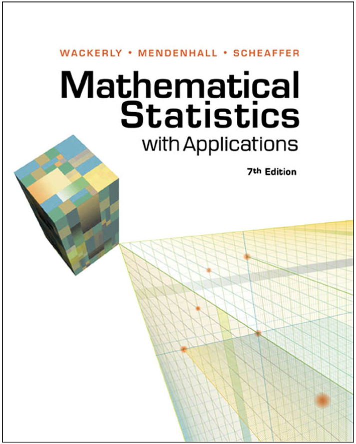
Basic to inference making is the problem of calculating the *probability* of an observed sample.
As a result, probability is the mechanism used in making statistical inferences.

Intuitive assessments of probabilities can often turn out to be unsatisfactory, and we need a rigorous theory of probability in order to develop methods of inference.



We will begin with a study of the mechanism employed in making inferences, the theory of probability.
This theory provides theoretical models for generating experimental data and thereby provides the basis for our study of statistical inference.

Reference



I have prepared these notes from the book "Mathematical Statistics with Applications, 7th Edition" by Wackerley, Mendenhall, and Scheaffer. (Thomson Brooks/Cole)

For the course, this shall be the reference book.

Throughout the notes, the words "Text" and "Book" will refer to the book mentioned above.

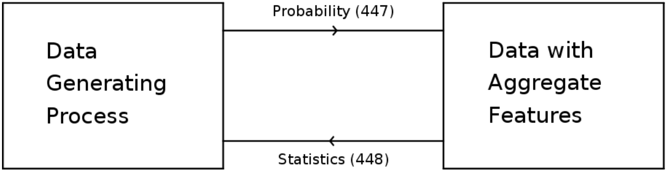
Two problems which will be in the Final Exam (in some form or other)

1. The "Monty Hall Problem " (Text 2.20, Wikipedia)
2. Bayes' Theorem Problem (e.g. Text 2.125)

Prerequisite: C or better in Math 323.

A C- is acceptable if your first semester at Binghamton was before Fall 2014.

Please try to log in to the homework system. For a link, look at the syllabus page: www.dikran.com



Remarks on how to study:

- We want to get to a point where we can do almost any problem from the book.
- Use the fact that no problem in the book is terribly difficult. Usually they don't take more than 10 minutes each.
- Know the tricks and techniques from each section.
- Try to do the problems. Don't spend huge amounts of time hung up on a single problem.
  If you can't do it in 20 minutes, time to look things up:
  is there a technique you forgot? A formula for expectation or variance?
- If you have all the techniques, maybe it's time to look at the solution. (Solutions to every problem are available.)
- After reading the solution, close the solution book and try to solve the problem again.
  Can you explain the steps?
  Could you solve the problem 2 days later without referring to the solutions?
- Over time you will build a library of problems you can solve.
  You will notice patterns – the problems will fit into into categories.

Why should you do the applet exercises from the book?

To check whether you know the various distributions, I can show you graphs and ask "Which of the distributions we studied is this?" or "What are the parameter values?".
You can prepare for this using the Applet exercises.
(Look at the book; there are dozens.)
You could also prepare by using R.

Please attempt every problem I assign, and try a few more: Many have answers in the back of the text.
If you can't do the problem:

1. Look at examples from the relevant section.
2. Answers for odd-numbered problems in the back of the text.
3. Student Solution Manual (All odd-numbered problems solved).
4. Look for analogous problems.
5. Yahoo! answers.
6. Instructor's Solution Manual.

You have to do the work.

An example to show that what is easy is not obvious:
The University of California at Berkeley was sued.
Discrimination against women in graduate admissions was alleged.
Women were admitted to graduate school at a much lower rate.

The university attempted to find the culprit(s).

Every department was required to report admission rates for men and women.
The reasoning: if women applicants are admitted at a lower rate overall, there must be some department(s) which are discriminating against women.
Find those departments and institute appropriate remedies.

Surprise: *every* department reports that women are admitted at *higher* rates than men.

Detailed records prove it.
How is this possible?

- Toy example with two departments: Engineering and Humanities
- Engineering admits 40% of women and 30% of men
- Humanities admits 20% of women and 10% of men
- Engineering applicants: 90 men, 10 women
- Humanities applicants: 90 women, 10 men
- Overall: women 24% men 28%

"Simpson's Paradox"

Probability: simple, but not obvious. You have to do the work!

End of Chapter 1

Chapter 2

Probability

An "Interview Problem"

We're going to play a game. The player is allowed to flip a fair coin repeatedly, and decide after each flip whether to stop. When the player stops, if they have so far flipped k heads in n flips, they are paid $\frac{k}{n}$ dollars.

Example (1)

If the player flips H on the first try and stops, their payout is $\frac{1}{1} = \$1$.

Example (2)

If the player flips T, T and then stops, their payout is $\frac{0}{2} = \$0$.

Example (3)

If the player flips T, H, T and then stops, their payout is $\frac{1}{3} = \$0.33\ldots$.

Questions:

- How much would you pay to play this game?
- How much would you charge someone else?
- What is the best strategy?

A strategy:

- If we flip H on the first try, stop.
- If we flip T, flip 1000 times; the result will be very close to 50% heads, very likely.
  (If not, flip a few thousand times more.)

Approximate Payout:		50% chance of \$ 1
	$+$	50% chance of about 50 ¢
Total:		<u>75 ¢</u>.

This is called the "Chow-Robbins Game". The exact value is unknown.

- It may be correct to keep playing even though you have more than 50% heads; for example, with $1T$ and $2H$, it is correct to keep playing.
- The correct strategy is unknown; to see the complexity, consider questions like: Would you continue with $5H$ and $3T$?
  How about $66H$ and $59T$?
- The idea that the payout will "eventually be 50% or close to it" is a limit theorem – called "The Law of Large Numbers".
- The fair price for the game is called an "Expected Value" or "Mean".

"Subjective Probability"

- What is the probability that Hillary Clinton will be the next President of The United States?
- What is the probability that this patient survives the operation?

We study "Axiomatic Probability" (Kolmogorov, circa 1931).

Definition (Probability)

A probability is an assignment of numbers (probabilities) to sets of possible outcomes satisfying certain axioms.

Example: Coin Flip

On flipping two fair coins, the possible outcomes are HH, HT, TH, and TT, all equally likely. So the probability of each outcome is $\frac{1}{4} = 0.25$.

Axiomatic Probability

Definition ((Axiomatic) Probability)

Suppose S is a sample space associated with an experiment.
To every event A in S (A is a subset of S), we assign a number, $P(A)$, called the probability of A, so that the following axioms hold:

Axiom 1: $P(A) \geq 0$.

Axiom 2: $P(S) = 1$.

Axiom 3: If $A_1, A_2, \dots$ form a sequence of pairwise mutually exclusive events in S (that is, $A_i \cap A_j = \emptyset$ if $i \neq j$), then

$$P(A_1 \cup A_2 \cup \dots) = \sum_{n=1}^{\infty} P(A_n).$$

Example (Rolling a die)

The "sample space" S of possible outcomes is

$$S = \{1, 2, 3, 4, 5, 6\}$$

An event is a subset of S.

Example (continued)

Let A be the event "The result of the die roll is an even number". Then

$$A = \{2, 4, 6\}$$

We already know how to assign a probability $P(A)$:

$$P(A) = \frac{1}{2}.$$

Define three events E_2, E_4, E_6 by

- E_2 = Result is a 2,
- E_4 = Result is a 4,
- E_6 = Result is a 6.

Then

$$A = E_2 \cup E_4 \cup E_6, \qquad P(E_2) = P(E_4) = P(E_6) = \frac{1}{6},$$

and

$$P(A) = P(E_2) + P(E_4) + P(E_6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

You already know the axioms of probability.
The fact above is a special case of the most complex axiom.

Problems in Discrete Probability

Here is a type of problem where there is a sample space S (which is a finite set) and we know, or can assume, that every individual outcome in S is equally likely.
We have an event $A \subset S$, and we want to find $P(A)$.

Solution:

Count the elements of S and those of A. Then

$$P(A) = \frac{|A|}{|S|}$$

Problems in Discrete Probability

Here is a type of problem where there is a sample space S (which is a finite set) and we know, or can assume, that every individual outcome in S is equally likely.
We have an event $A \subset S$, and we want to find $P(A)$.

Solution:

Count the elements of S and those of A. Then

$$P(A) = \frac{|A|}{|S|}$$

Problems in Discrete Probability

Here is a type of problem where there is a sample space S (which is a finite set) and we know, or can assume, that every individual outcome in S is equally likely.
We have an event $A \subset S$, and we want to find $P(A)$.

Solution:

Count the elements of S and those of A. Then

$$P(A) = \frac{|A|}{|S|}$$

Problems in Discrete Probability

Here is a type of problem where there is a sample space S (which is a finite set) and we know, or can assume, that every individual outcome in S is equally likely.
We have an event $A \subset S$, and we want to find $P(A)$.

Solution:

Count the elements of S and those of A. Then

$$P(A) = \frac{|A|}{|S|}$$

Problems in Discrete Probability

Here is a type of problem where there is a sample space S (which is a finite set) and we know, or can assume, that every individual outcome in S is equally likely.

We have an event $A \subset S$, and we want to find $P(A)$.

Solution:

Count the elements of S and those of A. Then

$$P(A) = \frac{|A|}{|S|}$$

Problems in Discrete Probability

Here is a type of problem where there is a sample space S (which is a finite set) and we know, or can assume, that every individual outcome in S is equally likely.

We have an event $A \subset S$, and we want to find $P(A)$.

Solution:

Count the elements of S and those of A. Then

$$P(A) = \frac{|A|}{|S|} \quad \longleftarrow \quad \text{number of elements of } A$$

Here is a type of problem where there is a sample space S (which is a finite set) and we know, or can assume, that every individual outcome in S is equally likely.
We have an event $A \subset S$, and we want to find $P(A)$.

Solution:

Count the elements of S and those of A. Then

$$P(A) = \frac{|A|}{|S|} \quad \begin{array}{ll} \longleftarrow & \text{number of elements of } A \\ \longleftarrow & \text{number of elements of } S \end{array}$$

Problems in Discrete Probability

Here is a type of problem where there is a sample space S (which is a finite set) and we know, or can assume, that every individual outcome in S is equally likely.
We have an event $A \subset S$, and we want to find $P(A)$.

Solution:

Count the elements of S and those of A. Then

$$P(A) = \frac{|A|}{|S|} \quad \begin{array}{ll} \longleftarrow & \text{number of elements of } A \\ \longleftarrow & \text{number of elements of } S \end{array}$$

This applies to our example of rolling the die.

Problems in Discrete Probability

Here is a type of problem where there is a sample space S (which is a finite set) and we know, or can assume, that every individual outcome in S is equally likely.

We have an event $A \subset S$, and we want to find $P(A)$.

Solution:

Count the elements of S and those of A. Then

$$P(A) = \frac{|A|}{|S|} \quad \begin{array}{ll} \longleftarrow & \text{number of elements of } A \\ \longleftarrow & \text{number of elements of } S \end{array}$$

This applies to our example of rolling the die. There

$$S = \{1, 2, 3, 4, 5, 6\} \qquad |S| = 6$$

Problems in Discrete Probability

Here is a type of problem where there is a sample space S (which is a finite set) and we know, or can assume, that every individual outcome in S is equally likely.
We have an event $A \subset S$, and we want to find $P(A)$.

Solution:

Count the elements of S and those of A. Then

$$P(A) = \frac{|A|}{|S|} \quad \begin{array}{ll} \longleftarrow & \text{number of elements of } A \\ \longleftarrow & \text{number of elements of } S \end{array}$$

This applies to our example of rolling the die. There

$$\begin{array}{rlrl} S &= \{1, 2, 3, 4, 5, 6\} & |S| &= 6 \\ A &= \{2, 4, 6\} & |A| &= 3 \end{array}$$

Problems in Discrete Probability

Here is a type of problem where there is a sample space S (which is a finite set) and we know, or can assume, that every individual outcome in S is equally likely.
We have an event $A \subset S$, and we want to find $P(A)$.

Solution:

Count the elements of S and those of A. Then

$$P(A) = \frac{|A|}{|S|} \quad \begin{array}{ll} \longleftarrow & \text{number of elements of } A \\ \longleftarrow & \text{number of elements of } S \end{array}$$

This applies to our example of rolling the die. There

$$\begin{array}{ll} S = \{1,2,3,4,5,6\} & |S| = 6 \\ A = \{2,4,6\} & |A| = 3 \end{array} \Longrightarrow P(A) = \frac{3}{6} = \frac{1}{2}.$$

Problems in Discrete Probability

Here is a type of problem where there is a sample space S (which is a finite set) and we know, or can assume, that every individual outcome in S is equally likely.
We have an event $A \subset S$, and we want to find $P(A)$.

Solution:

Count the elements of S and those of A. Then

$$P(A) = \frac{|A|}{|S|} \quad \begin{array}{l} \longleftarrow \quad \text{number of elements of } A \\ \longleftarrow \quad \text{number of elements of } S \end{array}$$

This applies to our example of rolling the die. There

$$\begin{array}{cc} S = \{1,2,3,4,5,6\} & |S| = 6 \\ A = \{2,4,6\} & |A| = 3 \end{array} \implies P(A) = \frac{3}{6} = \frac{1}{2}.$$

Note that this process only works when we know that all members of S are equally likely outcomes.

The "Sample-Point Method"

The following steps are used to find the probability of an event:

1. Define the experiment and clearly determine how to describe one simple event.
2. List the simple events associated with the experiment and test each to make certain that it cannot be decomposed. This defines the sample space S.
3. Assign reasonable probabilities to the sample points in S, making certain that $P(E_i) \geq 0$ and $\sum_i P(E_i) = 1$.
4. Define the event of interest, A, as a specific collection of sample points
   (A sample point is in A if A occurs when the sample point occurs. Test *all* sample points in S to identify those in A.)
5. Find $P(A)$ by summing the probabilities of the sample points in A.

It is possible to define probabilities in a different way so that not all members of S are equally likely.
This might correspond to, say, loading the die so that 6 is more likely to come up.

Exercise 2.12

A vehicle arriving at an intersection can turn right, turn left, or continue straight ahead. The experiment consists of observing the movement of a single vehicle through the intersection.

(a) List the sample space for this experiment.

(b) Assuming that all sample points are equally likely, find the probability that the vehicle turns.

Solution:

$S = \{\text{turns right}\,,\ \text{turns left}\,,\ \text{straight ahead}\}$.
Assuming all sample points are equally likely, find the probability that the vehicle turns.
Here $T = \{\text{turns}\} = \{\text{turns right}\,,\ \text{turns left}\}$. So $P(T) = \frac{|T|}{|S|} = \frac{2}{3}$.

We can still define probabilities even if not all points in the sample space are equally likely.

Exercise 2.10

The proportions of blood phenotypes, A, B, AB, and O, in the population of all Caucasians in the Unites States are approximately .41, .10, .04, and .45, respectively. A single Caucasian is chosen at random from the population.

(a) List the sample space for this experiment.

(b) Make use of the information given above to assign probabilities to each of the simple events.

(c) What is the probability that the person chosen at random has either type A or type AB blood?

Solution:

$S = \{A, B, AB, O\}$;
$P(A) = 0.41,\ P(B) = 0.10,\ P(AB) = 0.04,\ P(O) = 0.45$;
$E = \{\text{person has type } A \text{ or } AB \text{ blood}\}$.
Then

$$P(E) = P(A) + P(AB) = 0.41 + 0.04 = 0.45.$$

Remark:

In a situation like this (not all simple events are equally likely), we need extra information to find the probabilities.

So far, we have thought of S as a finite set of points ("simple events"). We can also think of S as a continuous space.

The probability then becomes something like a measurement of area. Note that all axioms of probability are satisfied if $S =$ unit square and the event A is a subset of the unit square; then $P(A) =$ area of A.

In Section 2.8, we will see various probability formulas to get an idea of what's going on.
Pretend we are in the situation of subsets of the unit square and that probability = area.

Theorem 2.6: "Additive Law"

Theorem 2.6: "Additive Law"

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Theorem 2.6: "Additive Law"

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof:

Theorem 2.6: "Additive Law"

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof:

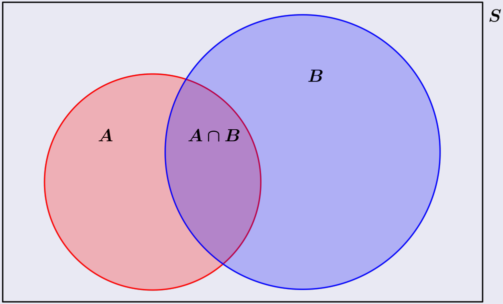
Let's draw a picture:

Theorem 2.6: "Additive Law"

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof:

Let's draw a picture:

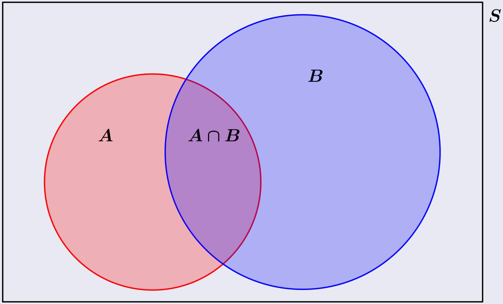


Theorem 2.6: “Additive Law”

$P(A \cup B) = P(A) + P(B) - P(A \cap B).$

Proof:

Let’s draw a picture:



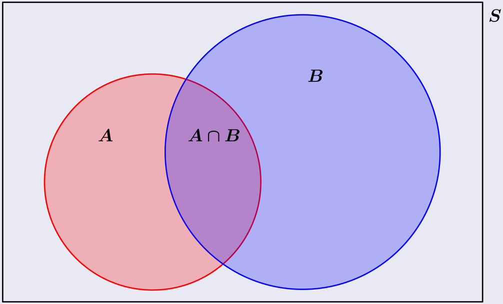
$\text{area}(A \cup B) = \text{area}(A) + \text{area}(B) - \text{double counted part.}$

Theorem 2.6: "Additive Law"

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

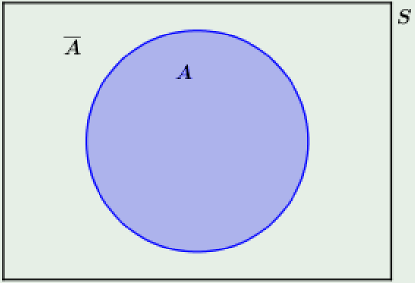
Proof:

Let's draw a picture:



$$\text{area}(A \cup B) = \text{area}(A) + \text{area}(B) - \underbrace{\text{double counted part}}_{\text{area}(A \cap B)}.$$

Example (Theorem 2.7: $P(A) = 1 - P\left(\overline{A}\right)$)



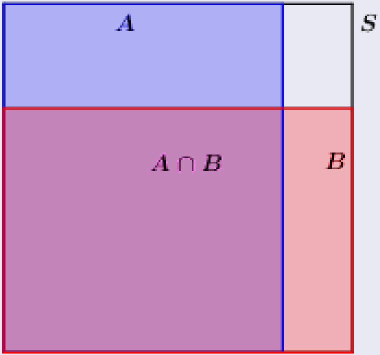
$$A \cup \overline{A} = S, A \cap \overline{A} = \emptyset \implies P(A) + P(\overline{A}) = P(S) = 1.$$

Exercise:

Suppose A and B are two events with $P(A) = 0.8, P(B) = 0.7$.
Is it possible that $P(A \cap B) = 0.3$?

Solution:

Answer: NO!



Combine the two statements
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
and $P(A \cup B) \leq 1$:

More complex example: Counting (Example 2.10)

A labor dispute has arisen concerning the distribution of 20 laborers to four different construction jobs.
The first job (considered to be very undesirable) required 6 laborers; the second, third, and fourth utilized 4, 5, and 5 laborers, respectively.
The dispute arose over an alleged random distribution of the laborers to the jobs that placed all 4 members of a particular ethnic group on job 1.
In considering whether the assignment represented injustice, a mediation panel desired the probability of the observed event.

(a) Determine the number of sample points in the sample space S for this experiment, that is, determine the number of ways the 20 laborers can be divided into groups of the appropriate sizes to fill all of the jobs.

(b) Find the probability of the observed event if it is assumed that the laborers are randomly assigned to jobs.

Analysis:

How many ways can you assign 20 laborers to 4 construction jobs requiring 6, 4, 5, and 5 laborers, respectively?

Notice that $6 + 4 + 5 + 5 = 20$.

Question: How many ways can we divide a 20-element set into 4 subsets of size 6, 4, 5, and 5, respectively?

Answer: Theorem 2.3: There are

$$\binom{20}{6 \quad 4 \quad 5 \quad 5} = \frac{20!}{6! \cdot 4! \cdot 5! \cdot 5!} \qquad (= 4.89 \times 10^{10}) \text{ ways.}$$

How many ways can we do this so that all 4 members of the minority group are assigned to the most "unpleasant" job?

Analysis: (continued)

Note that the "unpleasant" job requires 6 people.

Analysis: (continued)

Note that the “unpleasant” job requires 6 people. We now have 16 people to assign to 4 jobs; the first job still needs 2 more people.

Analysis: (continued)

Note that the "unpleasant" job requires 6 people. We now have 16 people to assign to 4 jobs; the first job still needs 2 more people.

Answer:

$$\binom{16}{2 \quad 4 \quad 5 \quad 5}$$

Analysis: (continued)

Note that the "unpleasant" job requires 6 people. We now have 16 people to assign to 4 jobs; the first job still needs 2 more people.

Answer:

$$\binom{16}{2 \quad 4 \quad 5 \quad 5} = \frac{16!}{2! \cdot 4! \cdot 5! \cdot 5!}$$

Analysis: (continued)

Note that the "unpleasant" job requires 6 people. We now have 16 people to assign to 4 jobs; the first job still needs 2 more people.

Answer:

$$\binom{16}{2 \quad 4 \quad 5 \quad 5} = \frac{16!}{2! \cdot 4! \cdot 5! \cdot 5!} \qquad (= 1.51 \times 10^8).$$

Analysis: (continued)

Note that the "unpleasant" job requires 6 people. We now have 16 people to assign to 4 jobs; the first job still needs 2 more people.

Answer:

$$\binom{16}{2 \quad 4 \quad 5 \quad 5} = \frac{16!}{2! \cdot 4! \cdot 5! \cdot 5!} \qquad (= 1.51 \times 10^8).$$

What is the probability that this would happen if every worker were assigned randomly to a job?

Analysis: (continued)

Note that the "unpleasant" job requires 6 people. We now have 16 people to assign to 4 jobs; the first job still needs 2 more people.

Answer:

$$\binom{16}{2 \quad 4 \quad 5 \quad 5} = \frac{16!}{2! \cdot 4! \cdot 5! \cdot 5!} \qquad (= 1.51 \times 10^8).$$

What is the probability that this would happen if every worker were assigned randomly to a job?

$$P(A) = \frac{|A|}{|S|}$$

Analysis: (continued)

Note that the "unpleasant" job requires 6 people. We now have 16 people to assign to 4 jobs; the first job still needs 2 more people.

Answer:

$$\binom{16}{2 \quad 4 \quad 5 \quad 5} = \frac{16!}{2! \cdot 4! \cdot 5! \cdot 5!} \qquad (= 1.51 \times 10^8).$$

What is the probability that this would happen if every worker were assigned randomly to a job?

$$P(A) = \frac{|A|}{|S|} = \frac{\frac{16!}{2! \cdot 4! \cdot 5! \cdot 5!}}{\frac{20!}{6! \cdot 4! \cdot 5! \cdot 5!}}$$

Analysis: (continued)

Note that the "unpleasant" job requires 6 people. We now have 16 people to assign to 4 jobs; the first job still needs 2 more people.

Answer:

$$\binom{16}{2 \quad 4 \quad 5 \quad 5} = \frac{16!}{2! \cdot 4! \cdot 5! \cdot 5!} \qquad (= 1.51 \times 10^8).$$

What is the probability that this would happen if every worker were assigned randomly to a job?

$$P(A) = \frac{|A|}{|S|} = \frac{\frac{16!}{2! \cdot \cancel{4!} \cdot \cancel{5!} \cdot \cancel{5!}}}{\frac{20!}{6! \cdot \cancel{4!} \cdot \cancel{5!} \cdot \cancel{5!}}} = \frac{16! \cdot 6!}{20! \cdot 2!}$$

Analysis: (continued)

Note that the "unpleasant" job requires 6 people. We now have 16 people to assign to 4 jobs; the first job still needs 2 more people.

Answer:

$$\binom{16}{2 \quad 4 \quad 5 \quad 5} = \frac{16!}{2! \cdot 4! \cdot 5! \cdot 5!} \qquad (= 1.51 \times 10^8).$$

What is the probability that this would happen if every worker were assigned randomly to a job?

$$P(A) = \frac{|A|}{|S|} = \frac{\frac{16!}{2! \cdot \cancel{4!} \cdot \cancel{5!} \cdot \cancel{5!}}}{\frac{20!}{6! \cdot \cancel{4!} \cdot \cancel{5!} \cdot \cancel{5!}}} = \frac{16! \cdot 6!}{20! \cdot 2!}$$

$$= \frac{16! \cdot \ 6 \ \cdot \ 5 \cdot 4 \ \cdot \ 3 \ \cdot 2!}{20 \ \cdot 19 \cdot \ 18 \ \cdot 17 \cdot 16! \cdot 2!}$$

Analysis: (continued)

Note that the "unpleasant" job requires 6 people. We now have 16 people to assign to 4 jobs; the first job still needs 2 more people.

Answer:

$$\binom{16}{2 \quad 4 \quad 5 \quad 5} = \frac{16!}{2! \cdot 4! \cdot 5! \cdot 5!} \qquad (= 1.51 \times 10^8).$$

What is the probability that this would happen if every worker were assigned randomly to a job?

$$P(A) = \frac{|A|}{|S|} = \frac{\frac{16!}{2! \cdot \cancel{4!} \cdot \cancel{5!} \cdot \cancel{5!}}}{\frac{20!}{6! \cdot \cancel{4!} \cdot \cancel{5!} \cdot \cancel{5!}}} = \frac{16! \cdot 6!}{20! \cdot 2!}$$

$$= \frac{\cancel{16!} \cdot (\cancel{6}) \cdot \{\cancel{5 \cdot 4}\} \cdot (\cancel{3}) \cdot \cancel{2!}}{\{\cancel{20}\} \cdot 19 \cdot (\cancel{18}) \cdot 17 \cdot \cancel{16! \cdot 2!}} = \frac{1}{19 \cdot 17}$$

Analysis: (continued)

Note that the "unpleasant" job requires 6 people. We now have 16 people to assign to 4 jobs; the first job still needs 2 more people.

Answer:

$$\binom{16}{2 \quad 4 \quad 5 \quad 5} = \frac{16!}{2! \cdot 4! \cdot 5! \cdot 5!} \qquad (= 1.51 \times 10^8).$$

What is the probability that this would happen if every worker were assigned randomly to a job?

$$P(A) = \frac{|A|}{|S|} = \frac{\frac{16!}{2! \cdot \cancel{4!} \cdot \cancel{5!} \cdot \cancel{5!}}}{\frac{20!}{6! \cdot \cancel{4!} \cdot \cancel{5!} \cdot \cancel{5!}}} = \frac{16! \cdot 6!}{20! \cdot 2!}$$

$$= \frac{\cancel{16!} \cdot (\cancel{6}) \cdot \{\cancel{5 \cdot 4}\} \cdot (\cancel{3}) \cdot \cancel{2!}}{\{\cancel{20}\} \cdot 19 \cdot (\cancel{18}) \cdot 17 \cdot \cancel{16! \cdot 2!}} = \frac{1}{19 \cdot 17} = \frac{1}{323}.$$

Note:

Observe that it is easier to expand the factorial and cancel the common factors out, than to compute the numerator and the denominator separately and then do the division.

Remarks:

- In Example 2.10, we worked out the probability of assigning 20 laborers to jobs requiring 6, 4, 5, and 5 laborers, respectively, such that 4 particular laborers are assigned to the first job, with respect to the random assignment.
- If the question is, "Are the 4 laborers of the particular ethnic group being treated uniformly?", this evidence is not conclusive by itself.
- There are two possible mitigating factors:
  - Maybe not all assignments of laborers to the jobs are equally likely.
  - Maybe there were many chances to observe this event.

Recall:

Theorem (2.3)

If $n = n_1 + \cdots + n_k$, the number of ways of partitioning n objects into subsets of size $n_1, \ldots, n_k$ is the "Multinomial Coefficient"

$$\binom{n}{n_1\ n_2\ \ldots\ n_k} = \frac{n!}{n_1! \cdot n_2! \cdot \cdots \cdot n_k!}.$$

Remark: Binomial Coefficients

Binomial coefficients are a special case of Multinomial coefficients (k=2). Recall the binomial coefficient is written $\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$. This is the same as $\binom{n}{k\quad n-k} = \frac{n!}{k! \cdot (n-k)!}$.

Binomial Theorem:

$$(x+y)^n = x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \cdots + \binom{n}{n-1}xy^{n-1} + y^n$$
$$= \sum_{k=0}^{n} \binom{n}{k} x^{n-k}y^k.$$

There is an analogous "Multinomial Theorem":

$$(x_1 + \cdots + x_k)^n = \sum_{\substack{n_1,\ldots,n_k \\ \sum_i n_i = n}} \binom{n}{n_1 \ \ldots \ n_k} x_1^{n_1} \ldots x_k^{n_k}.$$

Exercise 2.43

A fleet of nine taxis is to be dispatched to three airports in such a way that three go to airport A, five go to airport B, and one goes to airport C. In how many ways can this be accomplished?

Exercise 2.43

A fleet of nine taxis is to be dispatched to three airports in such a way that three go to airport A, five go to airport B, and one goes to airport C. In how many ways can this be accomplished?

Answer:

Exercise 2.43

A fleet of nine taxis is to be dispatched to three airports in such a way that three go to airport A, five go to airport B, and one goes to airport C. In how many ways can this be accomplished?

Answer:

$$\binom{9}{3 \quad 5 \quad 1} = \frac{9!}{3! \cdot 5! \cdot 1!}$$

Exercise 2.43

A fleet of nine taxis is to be dispatched to three airports in such a way that three go to airport A, five go to airport B, and one goes to airport C. In how many ways can this be accomplished?

Answer:

$$\binom{9}{3 \quad 5 \quad 1} = \frac{9!}{3! \cdot 5! \cdot 1!} = \frac{9 \cdot 8 \cdot 7 \cdot \quad 6 \quad \cdot 5!}{3 \cdot 2 \quad \cdot 5!}$$

Exercise 2.43

A fleet of nine taxis is to be dispatched to three airports in such a way that three go to airport A, five go to airport B, and one goes to airport C. In how many ways can this be accomplished?

Answer:

$$\binom{9}{3 \quad 5 \quad 1} = \frac{9!}{3! \cdot 5! \cdot 1!} = \frac{9 \cdot 8 \cdot 7 \cdot (\cancel{6}) \cdot \cancel{5!}}{(\cancel{3 \cdot 2}) \cdot \cancel{5!}} = 9 \cdot 8 \cdot 7$$

Exercise 2.43

A fleet of nine taxis is to be dispatched to three airports in such a way that three go to airport A, five go to airport B, and one goes to airport C. In how many ways can this be accomplished?

Answer:

$$\binom{9}{3 \quad 5 \quad 1} = \frac{9!}{3! \cdot 5! \cdot 1!} = \frac{9 \cdot 8 \cdot 7 \cdot (\cancel{6}) \cdot \cancel{5!}}{(\cancel{3 \cdot 2}) \cdot \cancel{5!}} = 9 \cdot 8 \cdot 7 = 504.$$

Exercise 2.43

A fleet of nine taxis is to be dispatched to three airports in such a way that three go to airport A, five go to airport B, and one goes to airport C. In how many ways can this be accomplished?

Answer:

$$\binom{9}{3 \quad 5 \quad 1} = \frac{9!}{3! \cdot 5! \cdot 1!} = \frac{9 \cdot 8 \cdot 7 \cdot (\cancel{6}) \cdot \cancel{5!}}{(\cancel{3 \cdot 2}) \cdot \cancel{5!}} = 9 \cdot 8 \cdot 7 = 504.$$

Exercise 2.45

What is the coefficient of $x^2 y^5 z^{10}$ in the expansion of $(x + y + z)^{17}$?

Exercise 2.43

A fleet of nine taxis is to be dispatched to three airports in such a way that three go to airport A, five go to airport B, and one goes to airport C. In how many ways can this be accomplished?

Answer:

$$\binom{9}{3 \quad 5 \quad 1} = \frac{9!}{3! \cdot 5! \cdot 1!} = \frac{9 \cdot 8 \cdot 7 \cdot (\cancel{6}) \cdot \cancel{5!}}{(\cancel{3 \cdot 2}) \cdot \cancel{5!}} = 9 \cdot 8 \cdot 7 = 504.$$

Exercise 2.45

What is the coefficient of $x^2y^5z^{10}$ in the expansion of $(x + y + z)^{17}$?

Answer:

Exercise 2.43

A fleet of nine taxis is to be dispatched to three airports in such a way that three go to airport A, five go to airport B, and one goes to airport C. In how many ways can this be accomplished?

Answer:

$$\binom{9}{3 \quad 5 \quad 1} = \frac{9!}{3! \cdot 5! \cdot 1!} = \frac{9 \cdot 8 \cdot 7 \cdot (\cancel{6}) \cdot \cancel{5!}}{(\cancel{3 \cdot 2}) \cdot \cancel{5!}} = 9 \cdot 8 \cdot 7 = 504.$$

Exercise 2.45

What is the coefficient of $x^2 y^5 z^{10}$ in the expansion of $(x+y+z)^{17}$?

Answer:

"17 choose 2, 5, 10"

Exercise 2.43

A fleet of nine taxis is to be dispatched to three airports in such a way that three go to airport A, five go to airport B, and one goes to airport C. In how many ways can this be accomplished?

Answer:

$$\binom{9}{3\quad 5\quad 1} = \frac{9!}{3!\cdot 5!\cdot 1!} = \frac{9\cdot 8\cdot 7\cdot (\cancel{6})\cdot \cancel{5!}}{(\cancel{3\cdot 2})\cdot \cancel{5!}} = 9\cdot 8\cdot 7 = 504.$$

Exercise 2.45

What is the coefficient of $x^2y^5z^{10}$ in the expansion of $(x+y+z)^{17}$?

Answer:

$$\text{“17 choose 2, 5, 10”} = \binom{17}{2\quad 5\quad 10}$$

Exercise 2.43

A fleet of nine taxis is to be dispatched to three airports in such a way that three go to airport A, five go to airport B, and one goes to airport C. In how many ways can this be accomplished?

Answer:

$$\binom{9}{3 \quad 5 \quad 1} = \frac{9!}{3! \cdot 5! \cdot 1!} = \frac{9 \cdot 8 \cdot 7 \cdot (\cancel{6}) \cdot \cancel{5!}}{(\cancel{3 \cdot 2}) \cdot \cancel{5!}} = 9 \cdot 8 \cdot 7 = 504.$$

Exercise 2.45

What is the coefficient of $x^2y^5z^{10}$ in the expansion of $(x + y + z)^{17}$?

Answer:

$$\text{"17 choose 2, 5, 10"} = \binom{17}{2 \quad 5 \quad 10} = \frac{17!}{2! \cdot 5! \cdot 10!}$$

Exercise 2.43

A fleet of nine taxis is to be dispatched to three airports in such a way that three go to airport A, five go to airport B, and one goes to airport C. In how many ways can this be accomplished?

Answer:

$$\binom{9}{3\quad 5\quad 1} = \frac{9!}{3!\cdot 5!\cdot 1!} = \frac{9\cdot 8\cdot 7\cdot (\cancel{6})\cdot \cancel{5!}}{(\cancel{3\cdot 2})\cdot \cancel{5!}} = 9\cdot 8\cdot 7 = 504.$$

Exercise 2.45

What is the coefficient of $x^2y^5z^{10}$ in the expansion of $(x+y+z)^{17}$?

Answer:

$$\text{“17 choose 2, 5, 10”} = \binom{17}{2\quad 5\quad 10} = \frac{17!}{2!\cdot 5!\cdot 10!} \overset{\text{easy!}}{=} 408408.$$

Conditional Probability and Independence

Definition

The "Conditional probability of A given B" is

$$P(A \mid B) \stackrel{\text{def}}{=} \frac{P(A \cap B)}{P(B)}.$$

Remark:

This is defined only if $P(B) > 0$.

Example

Two dice are thrown: let A be the event that the total showing is ≥ 10. Let B be the event that one of the dice shows a 1.
Then

$$P(A) = \frac{1}{6}, \qquad P(A \mid B) = 0.$$

Independent Events

Definition

Two events A and B are independent if $P(A \cap B) = P(A) \cdot P(B)$. Otherwise the events are dependent.

Remarks:

- If $P(B) > 0$, this is equivalent to $P(A \mid B) = P(A)$. Also, if $P(A) > 0$, this is equivalent to $P(B \mid A) = P(B)$.
- Independence is very special — similar to orthogonality.
- Sometimes independence is implicit: "We throw two dice"; it is assumed here that the two dice are independent.

The Multiplicative Law of Probability:

$$P(A \cap B) = P(A \mid B) \cdot P(B) = P(B \mid A) \cdot P(A).$$

Remark: $P(A \mid B) \neq P(B \mid A)$ in general.

The "Law of Total Probability"

Definition (Partition of a Set)

A "partition" of S is a division of S into disjoint pieces: sets $B_1, \ldots, B_k$ contained in S so that

1. $S = B_1 \cup \cdots \cup B_k$,
2. $B_i \cap B_j = \emptyset$ for $i \neq j$.

Law of Total Probability:

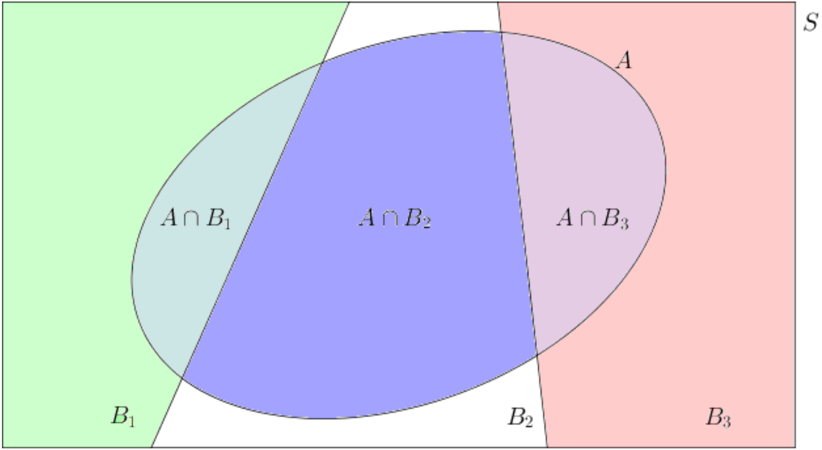
If $B_1, \ldots, B_k$ is a partition of S, then

$$P(A) = \sum_{i=1}^{k} P(A \mid B_i) \cdot P(B_i).$$

Remark:

This is really saying $P(A) = \sum_{i=1}^{k} P(A \cap B_i)$.

Example:



Theorem (Bayes' Rule)

If $B_1, \ldots, B_k$ is a partition of S and $P(B_i) > 0$ for all i, then

$$P(B_j \mid A) = \frac{P(A \mid B_j) \cdot P(B_j)}{\sum_{i=1}^{k} P(A \mid B_i) \cdot P(B_i)}.$$

Proof.

By definition,

$$P(B_j \mid A) = \frac{P(A \cap B_j)}{P(A)}.$$

Now apply the law of total probability in the denominator. $\square$

Example:

You are a doctor, you have a 90% accurate test for a disease. The prevalence of this disease in the population is 1%. A patient tests positive. What is the probability that the patient actually has the disease?

We write B_1 = patient has the disease,
B_2 = patient does not have the disease, A = patient tests positive.
The problem asks: what is $P(B_1 \mid A)$?

$$P(B_1 \mid A) = \frac{P(A \mid B_1) \cdot P(B_1)}{P(A \mid B_1) \cdot P(B_1) + P(A \mid B_2) \cdot P(B_2)}.$$

Interpretation: There are two ways of testing positive:

1. have the disease $(P(A \mid B_1) \cdot P(B_1))$,
   or
2. false positive $(P(A \mid B_2) \cdot P(B_2))$.

Example (continued):

We are given $P(A \mid B_1) = 90\%$, $P(A \mid B_2) = 10\%$, $P(B_1) = 1\%$.
We can deduce $P(B_2) = 99\%$.
We plug this information in Bayes' formula (previous slide):

$$\begin{aligned} P(B_1 \mid A) &= \frac{P(A \mid B_1) \cdot P(B_1)}{P(A \mid B_1) \cdot P(B_1) + P(A \mid B_2) \cdot P(B_2)} \\ &= \frac{(0.9)(0.01)}{(0.9)(0.01) + (0.1)(0.99)} \\ &= \frac{0.009}{0.009 + 0.099} = \frac{0.009}{0.108} \\ &= \frac{1}{12} \approx 0.0833. \end{aligned}$$

Conclusion:

The probability that the patient actually has the disease is only about <u>8%</u>.

Another way to think about this:

Suppose in the same setup that we have 1000 patients, of which 10 actually have the disease.
In this group, 9 will test positive. In the remaining 990 patients, we will get 99 positive tests.

If we know that a patient tests positive, we know that they are one of the $108 = 9 + 99$ patients identified above.
The number of those who actually have the disease is 9.
So the probability that a patient who tests positive actually has the disease is $\dfrac{9}{108} \approx 0.0833$.

Remark:

The key to the analysis is: there are two ways to test positive: have disease, or false positive.

Analysis: What is the relative likelihood of these two events?

Problem:

Vlad is to play a 2-game chess match with Gary and wishes to maximize his chances of winning, and minimize Gary's chances of winning. To do this, he may select a strategy right before he plays each game: timid or bold.

Unfortunately, Gary is the superior player.
If Vlad plays timidly, Gary will still win 10% of those games, and the rest will be draws. If Vlad plays boldly, Gary will win $\frac{5}{9}$ of those games, and lose the rest.

Describe Vlad's optimal strategy in this 2-game match.

Analysis of the problem:

- Scoring of a chess match: win $= 1$, loss $= 0$, draw $= \frac{1}{2}$.
  After 2 games, the player with more points wins the match.
  If the players have the same number of points, the match is tied.
- Gary is the better player, but Vlad can vary his strategy:
  - (T) Timid: Gary wins 10%, draw 90%.
  - (B) Bold: Gary wins $\frac{5}{9}$, Vlad wins $\frac{4}{9}$.
- Conclusion of the problem: With the correct strategy, Vlad has better chances of winning the match.

Correct Strategy:

Play boldly in the first game. If win, play timidly in the second game. Otherwise, play boldly again.

Exercise 2.133

A student answers a multiple-choice examination question that offers four possible answers.

Suppose the probability that the student knows the answer to the question is .8 and the probability that the student will guess is .2.

Assume that if the student guesses, the probability of selecting the correct answer is .25.

If the student correctly answers a question, what is the probability that the student really knew the correct answer?

Solution:

- Setup: name the events: N = student knows answer, $\overline{\text{N}}$ = student does not know answer, C = student answers correctly.
- Translate info from problem into notation:

$$P(N) = 0.8, P(\overline{N}) = 0.2, P(C \mid N) = 1, P(C \mid \overline{N}) = 0.25.$$

- What do we want?
  $P(N \mid C)$.
- Bayes' Formula:

$$P(N \mid C) = \frac{P(C \mid N)P(N)}{P(C \mid N)P(N) + P(C \mid \overline{N})P(\overline{N})}.$$

- Now plug in the numbers:

$$P(N \mid C) = \frac{(1)(0.8)}{(1)(0.8) + (0.25)(0.2)} = \frac{0.80}{0.85} = \frac{16}{17} \approx 94.12\%.$$

Where does Bayes' Formula come from?

- Setup: $B_1, \ldots, B_n$ = partition of S, A = separate event. Bayes' Formula:

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{P(A \mid B_1)P(B_1) + \cdots + P(A \mid B_n)P(B_n)}.$$

- To derive this, use the definition of conditional probability:

$$P(B_i \mid A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A \mid B_i)P(B_i)}{P(A)} \quad \text{(Multiplicative Law).}$$

- Now apply "Law of Total Probability" in the denominator:

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{P(A \mid B_1)P(B_1) + \cdots + P(A \mid B_n)P(B_n)}.$$

Monty Hall Problem (Exercise 2.20)

Monty Hall Problem (Exercise 2.20)

You **MUST** learn this problem.

Monty Hall Problem (Exercise 2.20)

You **MUST** learn this problem.

The following game was played on a popular television show.

Monty Hall Problem (Exercise 2.20)

You **MUST** learn this problem.

The following game was played on a popular television show. The host showed a contestant three large curtains. Behind one of the curtains was a nice prize (maybe a new car)

You **MUST** learn this problem.

The following game was played on a popular television show. The host showed a contestant three large curtains. Behind one of the curtains was a nice prize (maybe a new car)

Good Prize

Monty Hall Problem (Exercise 2.20)

You **MUST** learn this problem.

The following game was played on a popular television show. The host showed a contestant three large curtains. Behind one of the curtains was a nice prize (maybe a new car) and behind the other two curtains were worthless prizes (duds).

Good Prize

Monty Hall Problem (Exercise 2.20)

You **MUST** learn this problem.

The following game was played on a popular television show. The host showed a contestant three large curtains. Behind one of the curtains was a nice prize (maybe a new car) and behind the other two curtains were worthless prizes (duds).

Good Prize

Dud 1

Dud 2

You **MUST** learn this problem.

The following game was played on a popular television show. The host showed a contestant three large curtains. Behind one of the curtains was a nice prize (maybe a new car) and behind the other two curtains were worthless prizes (duds). The contestant was asked to choose one curtain.

Good Prize

Dud 1

Dud 2

You **MUST** learn this problem.

The following game was played on a popular television show. The host showed a contestant three large curtains. Behind one of the curtains was a nice prize (maybe a new car) and behind the other two curtains were worthless prizes (duds). The contestant was asked to choose one curtain. If the curtains are identified by their prizes, they could be labeled G, D_1, and D_2 (Good Prize, Dud 1, and Dud 2).

Good Prize

Dud 1

Dud 2

Monty Hall Problem (Exercise 2.20)

You **MUST** learn this problem.

The following game was played on a popular television show. The host showed a contestant three large curtains. Behind one of the curtains was a nice prize (maybe a new car) and behind the other two curtains were worthless prizes (duds). The contestant was asked to choose one curtain. If the curtains are identified by their prizes, they could be labeled G, D_1, and D_2 (Good Prize, Dud 1, and Dud 2). Thus, the sample space for the contestant's choice is $S = G, D_1, D_2$.

Good Prize

Dud 1

Dud 2

Monty Hall Problem (Exercise 2.20)

(a) If the contestant has no idea which curtains hide the various prizes and selects a curtain at random, assign reasonable probabilities to the simple events and calculate the probability that the contestant selects the curtain hiding the nice prize.

(b) Before showing the contestant what was behind the curtain initially chosen, the game show host would open one of the curtains and show the contestant one of the duds (he could always do this because he knew the curtain hiding the good prize). He then offered the contestant the option of changing from the curtain initially selected to the other remaining unopened curtain.

Which strategy maximizes the c ontestant's probability of winning the good prize: stay with the initial choice or switch to the other curtain? In answering the following sequence of questions, you will discover that, perhaps surprisingly, this question can be answered by considering only the sample space above and using the probabilities that you assigned to answer part (a).

Monty Hall Problem (Exercise 2.20)

(i) If the contestant choses to stay with her initial choice, she wins the good prize if and only if she initially chose curtain G. If she stays with her initial choice, what is the probability that she wins the good prize?

(ii) If the host shows her one of the duds and she switches to the other unopened curtain, what will be the result if she had initially selected G?

(iii) Answer the question in part (ii) if she had initially selected one of the duds.

(iv) If the contestant switches from her initial choice (as the result of being shown one of the duds), what is the probability that the contestant wins the good prize?

(v) Which strategy maximizes the contestant's probability of winning the good prize: stay with the initial choice or switch to the other curtain?

Analysis:

- Let G be the event that the initially selected curtain hides the good prize: $P(G) = \dfrac{1}{3}$.
- Let W be the event that we win (assuming we choose to switch curtains).
  What is $P(W)$?
- Law of Total Probability:

$$P(W) = P(W \mid G)P(G) + P(W \mid \overline{G})P(\overline{G}) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}.$$

Conclusion:

It is correct to switch curtains: the probability of winning by switching is $\dfrac{2}{3}$, while that of winning by not switching is $\dfrac{1}{3}$.

Remark:

You can experiment with cards: You'll need a friend to act as the game show host.

For example,

- Queen of Hearts = good prize,
- 3 of Spades, 5 of Clubs = bad prizes.

Question:

What if there were 4 curtains instead of 3?

Analysis:

Similar to the previous case!

Analysis (4 curtains):

Again, let G be the event that the initially selected curtain holds the good prize.

Assume we use the switching strategy. Let W be the event that we win.

By Law of Total Probability,

$$P(W) = P(W \mid G)P(G) + P(W \mid \overline{G})P(\overline{G}) = 0 \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{3}{4} = \frac{3}{8}.$$

$P(G) = \dfrac{1}{4}$, $P(W \mid G) = 0$ as before.

What is $P(W \mid \overline{G})$?

Two curtains are eliminated: one because the event $\overline{G}$ is "our selection does NOT hide the good prize", and one because we see a dud.

We should switch because $\dfrac{3}{8} > \dfrac{1}{4}$.

5 curtains, but at a cost:

Now suppose that there are 5 curtains. The good prize is \$1000, but it costs \$100 to switch.
Should you still switch? Why or why not?

Same analysis: G and W as before.

$$P(W) = P(W \mid G)P(G) + P(W \mid \overline{G})P(\overline{G}) = 0 \cdot \frac{1}{5} + \frac{1}{3} \cdot \frac{4}{5} = \frac{4}{15}.$$

The probability of winning by NOT switching is $P(G) = \frac{1}{5}$.

Since $\frac{4}{15} > \frac{1}{5}$, we would switch if it were free.

We have, by NOT switching, a $\frac{1}{5}$ chance of \$1000; this is worth about \$200.

By switching, we get a $\frac{4}{15}$ chance of \$1000, minus the switching cost of \$100;
that is, $\frac{4}{15} \cdot \$1000 - \$100 \approx \$166.67$.

As $\$166.67 < \$200$, we should NOT switch for the cost.

End of Chapter 2

Chapter 3

Discrete Random Variables and Their Probability Distributions

Random Variables and Expected Values

Definition

A Random Variable is a real-valued function on a sample space.

In practice, we may write a random variable like this:

$$X = \begin{cases} 2 & \text{with probability } 2/3 \\ -1 & \text{with probability } 1/3 \end{cases}$$

Definition

The average value of a random variable (over a large number of trials, say) is called the Expected Value of the random variable.

This is written $E[X]$ and we speak of "the expectation of X" or "the mean of X".

The formal definition captures some properties and subtleties not seen in our format. But this format is very convenient for computing the expected value $E[X]$ or "mean of X": in this case,

$$E[X] = 2 \cdot \frac{2}{3} + (-1) \cdot \frac{1}{3} = \frac{4}{3} - \frac{1}{3} = 1.$$

Here X can take on values 2 and -1, and $P(X = 2) = {}^2\!/_3$, $P(X = -1) = {}^1\!/_3$.

So the "probability function" $p(x) := P(X = x)$ is

$$p(x) = \begin{cases} {}^2\!/_3 & (x = 2) \\ {}^1\!/_3 & (x = -1) \\ 0 & \text{otherwise.} \end{cases}$$

The probability function must have the properties

1. $0 \le p(x) \le 1$ for all x,
2. $\sum_x p(x) = 1$.

Further, $E[X] = \sum_x xp(x)$ (Definition 3.4).

Example (Monty Hall: 5 curtains at a cost)

In this example, where we win \$1000 or \$0 (and maybe pay \$100), our winnings are a random variable.

Let X be our winnings if we don't switch:

$$X = \begin{cases} \$1000 & \text{with probability } 1/5 \\ \$0 & \text{with probability } 4/5. \end{cases}$$

Let Y be our winnings in the switch at a cost:

$$Y = \begin{cases} \$1000 - \$100 & \text{with probability } 4/15 \\ -\$100 & \text{with probability } 11/15. \end{cases}$$

Example (Monty Hall: 5 curtains at a cost)

We find

$$E[X] = \$1000 \cdot \frac{1}{5} + \$0 \cdot \frac{4}{5} = \$200,$$

$$\begin{aligned} E[Y] &= (\$1000 - \$100) \cdot \frac{4}{15} + (-\$100) \cdot \frac{11}{15} \\ &= \$1000 \cdot \frac{4}{15} - \$100 \cdot \left(\frac{4}{15} + \frac{11}{15}\right) \\ &= \$1000 \cdot \frac{4}{15} - \$100 \approx \$166.67. \end{aligned}$$

The expected value in switching is <u>less</u> than that without switching. So we should NOT switch.

Some simple types of Exercises

- Write down the probability function for a random variable. Find the mean.

Exercise 3.1:

When the health department tested private wells in a county for two impurities commonly found in drinking water, it found that 20% of the wells had neither impurity, 40% had impurity A, and 50% had impurity B. (Obviously, some had both impurities.)
If a well is randomly chosen from those in the county, find the probability distribution for Y , the number of impurities found in the well.

Solution:

Y can take the values 0, 1, or 2. We must find $P(Y = 0)$, $P(Y = 1)$, and $P(Y = 2)$.
From the problem statement, $P(Y = 0) = 20\%$.

Solution: (continued)

To find $P(Y=1)$ and $P(Y=2)$, use the "Event-Composition Method" (recall Monty Hall: 5 curtains).

We define events

- A = well has impurity A,
- B = well has impurity B.

Now translate the problem statement into probability statements about these events:

$$P(\overline{A} \cap \overline{B}) = 20\%, P(Y=2) = P(A \cap B).$$

How can we write down $P(Y=1)$?

$$P(Y=1) = P(A \cup B) - P(A \cap B).$$

Solution: (continued)

We know that $P(A) = 40\%$, and $P(B) = 50\%$. Also

$$P(Y = 1) + P(Y = 2) = P(A \cup B) = 80\%.$$

Plug this in $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ to deduce $P(A \cap B) = 10\%$.

We find the probability distribution of Y as

y	$p(y) = P(Y = y)$
0	20%
1	70%
2	10%

Note that if some y is not listed in the table, then $p(y) = 0$.

We can now find $E[Y]$:

$$E[Y] = 0 \cdot 20\% + 1 \cdot 70\% + 2 \cdot 10\% = 0.7 + 0.2 = \boxed{0.9}.$$

Exercise 3.3

A group of four components is known to contain two defectives. An inspector tests the components one at a time until the two defectives are located. Once she locates the two defectives, she stops testing, but the second defective is tested to ensure accuracy.
Let Y denote the number of the test on which the second defective is found. Find the probability distribution for Y.

Solution:

Observe that Y must be 2, 3, or 4.
Where can the defective components be? We can distribute 2 defectives among 4 components in $\binom{4}{2} = 6$ ways. Use $\times$ to represent a defective component and $\circ$ to represent a good component. There are 6 possibilities:

Solution: (continued)

$$\begin{array}{lllll} \times & \times & \circ & \circ & (Y=2) \\ \times & \circ & \circ & \times & (Y=4) \\ \circ & \times & \circ & \times & (Y=4) \end{array} \qquad \begin{array}{lllll} \times & \circ & \times & \circ & (Y=2) \\ \circ & \times & \times & \circ & (Y=3) \\ \circ & \circ & \times & \times & (Y=4) \end{array}$$

In each of these cases, we can write down the number of the test on which the second defective is found. (Proceed left to right).

Find the probability distribution of Y using the "Sample Point Method" from Chapter 2:

$$P(Y=2) = \frac{1}{6}, P(Y=3) = \frac{2}{6}, P(Y=4) = \frac{3}{6}.$$

Probability function $p(y)$ is then given by

y	2	3	4
$p(y)$	$1/6$	$1/3$	$1/2$

The expected value $E[Y]$ is

$$E[Y] = 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{3} + 4 \cdot \frac{1}{2} = \boxed{3\frac{1}{3}}.$$

Properties of Expected Value:

Ways to think about the expected or mean value:

- Long-run average value.
- Like an integral or sum.

Properties of integrals and sums usually hold for expected value.
In particular,

- Expected value is linear,
- Expected value of a nonrandom variable (like a constant) is just that variable / value.

We can use these properties to give a standard formula for "variance".

Definition (Variance, Standard Deviation)

If Y is a random variable with mean $E[Y] = \mu$, the variance of the random variable Y is defined to be the expected value of $(Y - \mu)^2$. That is,

$$V[Y] = E[(Y - \mu)^2].$$

The standard deviation of Y, denoted by σ, is the positive square root of $V[Y]$.

By the properties of expectation, we can prove the formula

$$V[Y] = E[Y^2] - \mu^2.$$

Proof of $V[Y] = E[Y^2] - \mu^2$:

By definition,

$$\begin{aligned} V[Y] &= E[(Y - \mu)^2] = E[Y^2 - 2\mu Y + \mu^2] \\ &= E[Y^2] - 2\mu E[Y] + E[\mu^2] \qquad \text{(linearity of } E\text{)} \\ &= E[Y^2] - 2\mu \cdot \mu + \mu^2 \qquad \text{(properties of expected value)} \\ &= E[Y^2] - \mu^2. \end{aligned}$$

We noted that expectation was linear:
if X, Y are random variables and a, b are real numbers (constants), then

$$E[aX + bY] = aE[X] + bE[Y].$$

This is NOT true for variance:
if a is a constant, then $V[aY] = a^2V[Y]$.
This is because $E[aY] = aE[Y] = a\mu$,
so

$$\begin{aligned} V[aY] &= E[(aY - a\mu)^2] \\ &= E[a^2(Y - \mu)^2] \\ &= a^2E[(Y - \mu)^2] \\ &= a^2V[Y]. \end{aligned}$$

There is a concept of independence for the random variables:
If X and Y are independent, then $V[X+Y] = V[X] + V[Y]$.

Remark:

Think of variance as being like "norm-squared" and independence as being the orthogonality. The above equation is the Pythagorean Theorem.

In the rest of this Chapter, several random variables are introduced, and we compute means and variances.

For the tests, know the variables, their means and variances, and how to derive them. The derivation of these results is often an exercise in Calculus.

Note: We use one of the equivalent terms "distribution", "probability distribution", "probability function", and "probability mass function".

Definition (Bernoulli Random Variable)

A <u>Bernoulli random variable</u> with parameter p is one which has probability function $p(1) = p$ and $p(0) = 1 - p$.

This means if X is such a variable, then

$$X = \begin{cases} 1 & \text{(with probability } p\text{)} \\ 0 & \text{(with probability } 1-p\text{)} \end{cases}$$

We can find $E[X]$ and $V[X]$:
By definition,

$$E[X] = \sum_x xp(x) = 0p(0) + 1p(1) = 0 + p = \boxed{p},$$

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

To compute $E[X^2]$, we apply Theorem 3.2:

Theorem (3.2)

Let Y be a discrete random variable with probability function $p(y)$ and $g(Y)$ be a real-valued function of Y.
Then the expected value of $g(Y)$ is given by

$$E[g(Y)] = \sum_y g(y)p(y).$$

We get

$$E[X^2] = 0^2 \cdot p(0) + 1^2 \cdot p(1) = p, V[X] = p - p^2 = \boxed{p(1-p)}.$$

Notation: Sometimes we write q for $1 - p$; we need context to know if q is being used this way.

Many of the variables of Chapter 3 are built from repeated independent Bernoulli trials. Example: Binomial random variable with parameters n and p.

Definition (Binomial Random Variable)

We do n Bernoulli trials with parameter p which are all independent, and count the number of 1s.
The total Y is a Binomial random variable with parameters n and p.

Equivalently, Y is Binomial with parameters n and p if Y has the probability function

$$p(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & (k = 0, 1, \ldots, n), \\ 0 & \text{otherwise.} \end{cases}$$

Notation: $Y \sim \mathrm{Bin}(n, p)$.

Why must Y have the above probability function?

If we have k successes in n trials, there are $\binom{n}{k}$ ways to distribute these.

Once we consider a particular pattern of successes, each success has probability p and each failure has it $1-p$.

Since the trials are independent, we can multiply to get $p^k(1-p)^{n-k}$.

In the text, the definition is that $Y \sim \text{Bin}(n,p)$ if Y has the probability function given above.

Terminology: Sometimes we say "Y has the binomial distribution with parameters n and p".

We can now compute $E[Y]$ and $V[Y]$.

If $X_1, \ldots, X_n$ are the random variables for Bernoulli trials, then $Y = X_1 + \cdots + X_n$.

$$\begin{aligned} E[Y] &= E[X_1 + \cdots + X_n] \\ &= E[X_1] + \cdots + E[X_n] \\ &= \underbrace{p + \cdots + p}_{n \text{ times}} \\ &= \boxed{np}, \end{aligned}$$

$$\begin{aligned} V[Y] &= V[X_1 + \cdots + X_n] \\ &= V[X_1] + \cdots + V[X_n] \qquad \text{(because } X_1, \ldots, X_n \text{ are independent)} \\ &= \underbrace{p(1-p) + \cdots + p(1-p)}_{n \text{ times}} \\ &= \boxed{np(1-p)}. \end{aligned}$$

Alternate derivation is given in Theorem 3.7.

Theorem (3.7)

If $Y \sim Bin(n, p)$, then $E[Y] = np$ and $V[Y] = npq$.

Proof:

We have

$$
\begin{aligned}
E[Y] &= \sum_y y p(y) \qquad \text{(by definition of expectation)} \\
&= \sum_{y=0}^{n} y \binom{n}{y} p^y q^{n-y} \qquad \text{(by definition of ``binomial'')} \\
&= \sum_{y=1}^{n} \frac{yn!}{y!(n-y)!} p^y q^{n-y} \qquad \text{(note that the first term is zero)} \\
&= \sum_{y=1}^{n} \frac{n(n-1)!}{(y-1)!(n-y)!} p p^{y-1} q^{n-y} \\
&= np \sum_{y=1}^{n} \frac{(n-1)!}{(y-1)!((n-1)-(y-1))!} p^{y-1} q^{(n-1)-(y-1)}.
\end{aligned}
$$

Proof (continued):

Now write $z = y - 1$, and change the variables in the sum:

$$E[Y] = np \sum_{y=1}^{n} \frac{(n-1)!}{(y-1)!((n-1)-(y-1))!} p^{y-1} q^{(n-1)-(y-1)}$$

$$= np \sum_{z=0}^{n-1} \frac{(n-1)!}{z!((n-1)-z)!} p^{z} q^{(n-1)-z}$$

$$= np \sum_{z=0}^{n-1}$$

Proof (continued):

Now write $z = y - 1$, and change the variables in the sum:

$$E[Y] = np \sum_{y=1}^{n} \frac{(n-1)!}{(y-1)!((n-1)-(y-1))!} p^{y-1} q^{(n-1)-(y-1)}$$

$$= np \sum_{z=0}^{n-1} \frac{(n-1)!}{z!((n-1)-z)!} p^{z} q^{(n-1)-z}$$

$$= np \sum_{z=0}^{n-1}$$

Proof (continued):

Now write $z = y - 1$, and change the variables in the sum:

$$
\begin{aligned}
E[Y] &= np \sum_{y=1}^{n} \frac{(n-1)!}{(y-1)!((n-1)-(y-1))!} p^{y-1} q^{(n-1)-(y-1)} \\
&= np \sum_{z=0}^{n-1} \frac{(n-1)!}{z!((n-1)-z)!} p^{z} q^{(n-1)-z} \\
&= np \sum_{z=0}^{n-1}
\end{aligned}
$$

Proof (continued):

Now write $z = y - 1$, and change the variables in the sum:

$$\begin{aligned} E[Y] &= np \sum_{y=1}^{n} \frac{(n-1)!}{(y-1)!((n-1)-(y-1))!} p^{y-1} q^{(n-1)-(y-1)} \\ &= np \sum_{z=0}^{n-1} \frac{(n-1)!}{z!((n-1)-z)!} p^{z} q^{(n-1)-z} \\ &= np \sum_{z=0}^{n-1} \binom{n-1}{z} p^{z} q^{(n-1)-z}. \end{aligned}$$

This is the probability function for $\text{Bin}(n-1, p)$, so the whole sum above is

$$(p+q)^{n-1} = 1^{n-1} = 1.$$

So $E[Y] = np$.

Proof (continued):

Now write $z = y - 1$, and change the variables in the sum:

$$E[Y] = np \sum_{y=1}^{n} \frac{(n-1)!}{(y-1)!((n-1)-(y-1))!} p^{y-1} q^{(n-1)-(y-1)}$$

$$= np \sum_{z=0}^{n-1} \frac{(n-1)!}{z!((n-1)-z)!} p^{z} q^{(n-1)-z}$$

$$= np \sum_{z=0}^{n-1} \underbrace{\binom{n-1}{z} p^{z} q^{(n-1)-z}}.$$

This is the probability function for $\mathrm{Bin}(n-1, p)$, so the whole sum above is

$$(p+q)^{n-1} = 1^{n-1} = 1.$$

So $E[Y] = np$.

Proof (continued):

For the computation of $V[Y]$, we could write similar sums of binomial coefficients. But the book introduces another method.

Work with the expectation $E[Y(Y-1)] = E[Y^2] - E[Y]$.

Since we already know $E[Y]$, knowing $E[Y(Y-1)]$ is equivalent to knowing $E[Y^2]$.

We find

$$E[Y(Y-1)] = \sum_{y=0}^{n} y(y-1)\binom{n}{y} p^y q^{n-y}.$$

When we expand the binomial coefficient, we can cancel the $y(y-1)$ with the $y!$.

Do the computations to obtain $V[Y] = npq$.

□

Definition (Geometric Random Variable)

We do repeated independent Bernoulli trials until we get a success. Let Y be the number of the trial on which the first success occurs. Then Y is a Geometric random variable with parameter p, written $Y \sim \text{Geom(p)}$. (p is the parameter in all the independent Bernoulli trials.)

Example:

Think of flipping a coin again and again until we get "heads".

Remark:

We use the convention that the number of the first trial is 1 (not zero). So $Y \geq 1$.

From the description above, we can find the probability function, mean, and variance of Y.

$$p(1) = P(Y = 1) = p,$$
$$p(2) = P(Y = 2) = P(\text{Failure in 1 and success in 2}).$$

Since trials are independent,

$P(\text{Failure in 1 and success in 2}) = P(\text{Failure in 1}) \cdot P(\text{success in 2}) = qp.$

Thus $p(2) = qp$.

We use similar reasoning to deduce $p(3) = q^2 p, \ldots, p(y) = q^{y-1} p$.

From the above probability function, we can write down an infinite series for $E[Y]$:

$$E[Y] = \sum_{y=1}^{\infty} y q^{y-1} p.$$

(This is an exercise in Calculus 2.)

Warm-up Exercise:

What is $\sum_{y=1}^{\infty} p(y) = \sum_{y=1}^{\infty} q^{y-1}p$?

This sum is

$$p\sum_{y=1}^{\infty} q^{y-1} = p\sum_{z=0}^{\infty} q^{z} = p\frac{1}{1-q} = \frac{p}{1-(1-p)} = \frac{p}{p} = \boxed{1}.$$

Review of Calculus: for $|x| < 1$,

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

Warm-up Exercise:

What is $\sum_{y=1}^{\infty} p(y) = \sum_{y=1}^{\infty} q^{y-1} p$?

This sum is

$$p \sum_{y=1}^{\infty} q^{y-1} = p \sum_{z=0}^{\infty} q^{z} = p \frac{1}{1-q} = \frac{p}{1-(1-p)} = \frac{p}{p} = \boxed{1}.$$

Review of Calculus: for $|x| < 1$,

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^{n}$$

Warm-up Exercise:

What is $\sum_{y=1}^{\infty} p(y) = \sum_{y=1}^{\infty} q^{y-1}p$?

This sum is

$$p\sum_{y=1}^{\infty} q^{y-1} = p\sum_{z=0}^{\infty} q^{z} = p\frac{1}{1-q} = \frac{p}{1-(1-p)} = \frac{p}{p} = \boxed{1}.$$

Review of Calculus: for $|x| < 1$,

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

Warm-up Exercise:

What is $\sum_{y=1}^{\infty} p(y) = \sum_{y=1}^{\infty} q^{y-1} p$?

This sum is

$$p \sum_{y=1}^{\infty} q^{y-1} = p \sum_{z=0}^{\infty} q^z = p \frac{1}{1-q} = \frac{p}{1-(1-p)} = \frac{p}{p} = \boxed{1}.$$

Review of Calculus: for $|x| < 1$,

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

Warm-up Exercise:

What is $\sum_{y=1}^{\infty} p(y) = \sum_{y=1}^{\infty} q^{y-1}p$?

This sum is

$$p\sum_{y=1}^{\infty} q^{y-1} = p\sum_{z=0}^{\infty} q^{z} = p\frac{1}{1-q} = \frac{p}{1-(1-p)} = \frac{p}{p} = \boxed{1}.$$

Review of Calculus: for $|x| < 1$,

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

Warm-up Exercise:

What is $\sum_{y=1}^{\infty} p(y) = \sum_{y=1}^{\infty} q^{y-1}p$?

This sum is

$$p\sum_{y=1}^{\infty} q^{y-1} = p\sum_{z=0}^{\infty} q^z = p\frac{1}{1-q} = \frac{p}{1-(1-p)} = \frac{p}{p} = \boxed{1}.$$

Review of Calculus: for $|x| < 1$,

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

Warm-up Exercise:

What is $\sum_{y=1}^{\infty} p(y) = \sum_{y=1}^{\infty} q^{y-1}p$?

This sum is

$$p\sum_{y=1}^{\infty} q^{y-1} = p\sum_{z=0}^{\infty} q^{z} = p\frac{1}{1-q} = \frac{p}{1-(1-p)} = \frac{p}{p} = \boxed{1}.$$

Review of Calculus: for $|x| < 1$,

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^{n}$$

Warm-up Exercise:

What is $\sum_{y=1}^{\infty} p(y) = \sum_{y=1}^{\infty} q^{y-1}p$?

This sum is

$$p\sum_{y=1}^{\infty} q^{y-1} = p\sum_{z=0}^{\infty} q^z = p\frac{1}{1-q} = \frac{p}{1-(1-p)} = \frac{p}{p} = \boxed{1}.$$

Review of Calculus: for $|x| < 1$,

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

Warm-up Exercise:

What is $\sum_{y=1}^{\infty} p(y) = \sum_{y=1}^{\infty} q^{y-1}p$?

This sum is

$$p\sum_{y=1}^{\infty} q^{y-1} = p\sum_{z=0}^{\infty} q^{z} = p\frac{1}{1-q} = \frac{p}{1-(1-p)} = \frac{p}{p} = \boxed{1}.$$

Review of Calculus: for $|x| < 1$,

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

Warm-up Exercise:

What is $\sum_{y=1}^{\infty} p(y) = \sum_{y=1}^{\infty} q^{y-1}p$?

This sum is

$$p\sum_{y=1}^{\infty} q^{y-1} = p\sum_{z=0}^{\infty} q^{z} = p\frac{1}{1-q} = \frac{p}{1-(1-p)} = \frac{p}{p} = \boxed{1}.$$

Review of Calculus: for $|x| < 1$,

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

$$\therefore \qquad \frac{d}{dx}\left(\frac{1}{1-x}\right)$$

Warm-up Exercise:

What is $\sum_{y=1}^{\infty} p(y) = \sum_{y=1}^{\infty} q^{y-1}p$?

This sum is

$$p\sum_{y=1}^{\infty} q^{y-1} = p\sum_{z=0}^{\infty} q^{z} = p\frac{1}{1-q} = \frac{p}{1-(1-p)} = \frac{p}{p} = \boxed{1}.$$

Review of Calculus: for $|x| < 1$,

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

$$\therefore \quad -1(1-x)^{-2}(-1) = \frac{d}{dx}\left(\frac{1}{1-x}\right) = \sum_{n=0}^{\infty} nx^{n-1}$$

Warm-up Exercise:

What is $\sum_{y=1}^{\infty} p(y) = \sum_{y=1}^{\infty} q^{y-1} p$?

This sum is

$$p \sum_{y=1}^{\infty} q^{y-1} = p \sum_{z=0}^{\infty} q^{z} = p \frac{1}{1-q} = \frac{p}{1-(1-p)} = \frac{p}{p} = \boxed{1}.$$

Review of Calculus: for $|x| < 1$,

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^{n}$$

$$\therefore \quad \cancel{-}1(1-x)^{-2}(\cancel{-}1) = \frac{d}{dx}\left(\frac{1}{1-x}\right) = \sum_{n=0}^{\infty} n x^{n-1} = \sum_{n=1}^{\infty} n x^{n-1}.$$

Now change x to q and n to y to get

$$\frac{1}{(1-q)^2} = \sum_{y=1}^{\infty} y q^{y-1}$$

$$\therefore \sum_{y=1}^{\infty} y q^{y-1} p = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}.$$

Conclusion: $E[Y] = \dfrac{1}{p}$.

Remark:

This is very intuitive: if the chance of success is $\frac{1}{3}$, the expected waiting time until success is 3 trials.

What about $V[Y]$?

$$V[Y] = E[Y^2] - E[Y]^2.$$

We need to compute $E[Y^2]$; we already know $E[Y]^2 = \frac{1}{p^2}$.

This is $\sum_{y=1}^{\infty} y^2 q^{y-1} p$.
This doesn't look like the derivative of something we know.
But we could use the technique of the last calculation to sum

$$\sum_{y=1}^{\infty} y(y-1)q^{y-2} \quad \text{or even} \quad \sum_{y=1}^{\infty} y(y-1)q^{y-1}p \quad (\star).$$

This last guy $(\star)$ is $E[Y(Y-1)] = E[Y^2] - E[Y]$.
So, by computing this sum $(\star)$, we can find $E[Y^2]$, because we know $E[Y]$ already.

Exercise:

Do the work outlined above to get

$$\boxed{V[Y] = \frac{q}{p^2}}.$$

$V[Y] = \frac{q}{p^2}$

Step 1: Compute

$$\frac{d^2}{dx^2}\left(\frac{1}{1-x}\right) = \frac{2}{(1-x)^3} = \frac{d^2}{dx^2}\sum_{n=0}^{\infty} x^n = \sum_{n=0}^{\infty} n(n-1)x^{n-2}$$

Step 2: Write in terms of q and y and multiply by qp.

$$\frac{2qp}{(1-q)^3} = \sum_{y=1}^{\infty} y(y-1)q^{y-1}p = E[Y(Y-1)]$$

Step 3: Use $E[Y(Y-1)] = E[Y^2] - E[Y]$, $V[Y] = E[Y^2] - E[Y]^2$, and $E[Y] = 1/p$ to find $V[Y]$.

$$E[Y^2] = E[Y(Y-1)] + E[Y] = \frac{2q}{p^2} + \frac{1}{p}$$

$$V[Y] = E[Y^2] - E[Y]^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2}$$

Negative Binomial Random Variable

We consider repeated independent Bernoulli trials, all with parameter p. Let Y be the number of the trial on which the r^{th} success occurs. From this description, we can find the probability distribution of Y, $E[Y]$, and $V[Y]$.

If the r^{th} success is on trial y, i.e. $Y = y$, then

(1) The y^{th} trial must be a success,

(2) There must be exactly $r-1$ successes in the first $y-1$ trials.

What is the probability of (2)?

There are $\binom{y-1}{r-1}$ ways to distribute $r-1$ successes in $y-1$ trials.

The probability of any particular way occurring is $p^{r-1}q^{(y-1)-(r-1)} = p^{r-1}q^{y-r}$.

The probability of (1) is just p, and (1) and (2) are independent.

Put this all together to get

$$p(y) = \binom{y-1}{r-1} p^{r-1} q^{y-r} p = \binom{y-1}{r-1} p^r q^{y-r}.$$

Remarks:

1. In the text, the definition of a negative binomial random variable is something which has the probability distribution as above.
2. If $r = 1$, this is just a geometric random variable.
3. The waiting time for r successes is the waiting time for the first, plus the waiting time for the second, ..., plus the waiting time for the r^{th} success.
4. The waiting time for each success is a geometric random variable with parameter p.
   This means that the negative binomial random variable with parameters r and p is the sum of r independent geometric random variables, each with parameter p.

Now we can write down $E[Y]$ and $V[Y]$ where Y is a negative binomial random variable with parameters r and p.

Let $X_1, \ldots, X_r$ be independent geometric RVs with parameter p. Then

$$E[Y] = E[X_1 + \cdots + X_r] = E[X_1] + \cdots + E[X_r] = \underbrace{\frac{1}{p} + \cdots + \frac{1}{p}}_{r \text{ times}} = \frac{r}{p},$$

$$V[Y] = V[X_1 + \cdots + X_r] \stackrel{*}{=} V[X_1] + \cdots + V[X_r] = \underbrace{\frac{q}{p^2} + \cdots + \frac{q}{p^2}}_{r \text{ times}} = \frac{rq}{p^2}.$$

*: because $X_1, \ldots, X_r$ are independent.

So if we remember that for a geometric RV the mean is $\frac{1}{p}$ and the variance is $\frac{q}{p^2}$, then the mean and variance of a negative binomial RV are easy to remember.

Exercise 3.72:

Given that we have already tossed a balanced coin ten times and obtained zero heads, what is the probability that we must toss it at least two more times to obtain the first head?

One solution:

- Note that the information about the first 10 flips is not relevant.
- The probability that we need at least 2 flips to get the first head is the same as the probability that the first flip is tails.
- The word "balanced" in the problem statement means that this probability is $\frac{1}{2}$.

Formal solution: The problem is about a geometric RV with $p = 1/2$. The question asks: What is $P(Y \geq 12 \mid Y \geq 11)$?

$$P(Y \geq 12 \mid Y \geq 11) = \frac{P(Y \geq 12 \cap Y \geq 11)}{P(Y \geq 11)} = \frac{P(Y \geq 12)}{P(Y \geq 11)}$$

For a geometric RV, $P(Y \geq k) =$ the probability of at least $k - 1$ successive failures $= q^{k-1}$.

$$P(Y \geq 12 \mid Y \geq 11) = \frac{P(Y \geq 12)}{P(Y \geq 11)} \frac{q^{12-1}}{q^{11-1}} = q = \frac{1}{2}$$

What if we allowed for some possibility that it was a trick coin? Then the first ten flips do make a difference, because they give us evidence that the coin is biased.

New problem, 3.72 revised

We have a bag of 100 coins. One is "double-tails" and 99 are normal. We pick one coin from the bag and flip it 10 times. It comes up tails 10 times in a row. What are the chances that it is actually the trick coin?

Name the events:
T = event that the coin selected is the trick coin.
R = event that we get a run of 10 tails.
Compute with Bayes' Rule.

Calculations with Bayes' Rule

$P(T) = 0.01$, $P(\bar{T}) = 0.99$, $P(R \mid T) = 1$, $P(R \mid \bar{T}) = 1/2^{10}$
Problem asks for $P(T \mid R)$

$$P(T \mid R) = \frac{P(T \cap R)}{P(R)} = \frac{P(R \mid T)P(T)}{P(R \mid T)P(T) + P(R \mid \bar{T})P(\bar{T})}$$

$$P(T \mid R) = \frac{1 \cdot 0.01}{1 \cdot 0.01 + 2^{-10} \cdot 0.99} \approx 0.912$$

So far we have studied the following distributions:

(1) Bernoulli random variable (Bernoulli trial).

(2) Binomial RV, $\text{Bin}(n, p)$.

"Number of successes in n independent Bernoulli trials".

(3) Geometric RV.

"Number of Bernoulli trials required to get first success".

(4) Negative Binomial RV.

"Number of Bernoulli trials required to get exactly r successes".

For each of these RVs, you should be able to produce the "probability function" (sometimes called distribution or PDF) $p(y) = P(Y = y)$.

You should also be able to produce the mean and variance. Note that a few tricks may be required to complete the computation.

Let's try now to produce a table with the probability function, mean, and variance for all of these RVs.

Distribution	**p(y)**	**E[Y]**	**V[Y]**	
Bernoulli	$p^y q^{1-y}\ (y=0,1)$ or $\begin{cases} p & y=1 \\ q & y=0 \end{cases}$	p	pq	$[q=1-p]$
Geometric	$q^{y-1}p$	$\frac{1}{p}$	$\frac{q}{p^2}$	
Binomial	$\binom{n}{y} p^y q^{n-y}$	np	npq	
Negative Binomial	$\binom{y-1}{r-1} q^{y-r} p^r$	$\frac{r}{p}$	$\frac{rq}{p^2}$	

Hypergeometric Random Variable

Suppose we have an urn with r red balls and $N - r$ black balls.
We select (without replacement) n balls from the urn and count the number Y of red balls.
Then Y is said to have the hypergeometric distribution with parameters N, r, n.

The hypergeometric RV has probability function

$$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} = \frac{\begin{array}{c}(\#\text{ of ways to get } y \text{ red from } r)\times \\ (\#\text{ of ways to get } n-y \text{ black from } N-r)\end{array}}{\#\text{ of ways to take } n \text{ balls from } N}.$$

There is a close analogy between the binomial and hypergeometric RVs.
In the limit as N becomes large, they are almost the same.

Suppose for the moment we only take one ball from the urn, so $n = 1$. Then the probability that we get a red ball is $\frac{r}{N}$ and the probability that we get a black ball is $\frac{N-r}{N}$.

So the expected number of red balls is $\frac{r}{N}$.

If we increase n, and do the selection n times, then the expected number of reds is

$$\begin{array}{r}(\text{Expected } \# \text{ on try } 1) + (\text{Expected } \# \text{ on try } 2) \\ + \cdots + (\text{Expected } \# \text{ on try } n)\end{array} = n \cdot \frac{r}{N}.$$

Notice the analogy with the Binomial RV.

The probability of success, p, is $\frac{r}{N}$, and the mean is $n \cdot \frac{r}{N}$.

We would therefore expect to get for the variance $V[Y]$ the result $n \cdot \frac{r}{N} \cdot \frac{N-r}{N}$.

BE CAREFUL!

There is a correction term:

The variance is less, because when we remove reds, we make blacks more likely, and vice versa.

Computation (Chapter 5) shows

$$V[Y] = n \cdot \frac{r}{N} \cdot \frac{N-r}{N} \cdot \frac{N-n}{N-1}.$$

$\leftarrow$ This we just have to remember for now.

Notice $\lim_{N\to\infty} \frac{N-n}{N-1} = 1$, i.e. the results for hypergeometric and binomial RVs are the same in the limit $N \to \infty$.

A convention for binomial coefficients:

$$\binom{n}{k} = 0 \quad \text{if} \quad k > n.$$

This convention is relevant in evaluating the probability function for the hypergeometric distribution.

Exercise 3.103:

A warehouse contains ten printing machines, four of which are defective. A company selects five of the machines at random, thinking all are in working condition. What is the probability that all five of the machines are nondefective?

Solution:

- $N = 10 \leftarrow$ total # of machines.
- "Red ball" corresponds to a nondefective machine: $r = 6$.
- $n = 5 \leftarrow$ # of machines in the sample.
- $Y =$ # of nondefective machines in the sample.

We have to find $P(Y = 5) = p(5)$.
Use the hypergeometric probability function:

$$p(5) = \frac{\binom{6}{5}\binom{10-6}{5-5}}{\binom{10}{5}} = \frac{6\binom{4}{0}}{252} = \frac{6}{252} \quad \left(= \frac{1}{42}\right).$$

Distribution	p(y)	E[Y]	V[Y]
Bernoulli	$p^y q^{1-y}$ $(y = 0, 1)$ or $\begin{cases} p & y = 1 \\ q & y = 0 \end{cases}$	p	pq
Geometric	$q^{y-1} p$	$\frac{1}{p}$	$\frac{q}{p^2}$
Binomial	$\binom{n}{y} p^y q^{n-y}$	np	npq
Negative Binomial	$\binom{y-1}{r-1} q^{y-r} p^r$	$\frac{r}{p}$	$\frac{rq}{p^2}$
Hypergeometric	$\frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$	$\frac{nr}{N}$	$\frac{nr}{N} \cdot \frac{N-r}{N} \cdot \frac{N-n}{N-1}$

$[q = 1 - p]$

Poisson Distribution

Suppose we have a type of event (say, tornadoes in Kansas) which occurs randomly at a certain rate, say 5 per year.

In any given year, there might not be exactly 5. We mean only that 5 per year is somehow an average rate; in a particular year, the actual number might be 0, 10, 3, or 17.

The Poisson random variable is a model for this situation. There is a parameter λ corresponding to the average rate. The variable Y is the number of events in a given period.

Definition (Poisson Distribution)

Y has the Poisson distribution with parameter λ if Y has the probability function

$$p(y) = \frac{\lambda^y e^{-\lambda}}{y!} \qquad (y = 0, 1, 2, \dots).$$

Notice that Y can take on any positive integer value.

For this to make sense, the probabilities need to add up to 1.
Let's check:

$$\sum_{y=0}^{\infty} \frac{\lambda^y e^{-\lambda}}{y!} = e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = e^{-\lambda} e^{\lambda} = 1.$$

So the probabilities do make sense.

Next step: What are the mean and variance of Y?
This is the easiest part of the table: if $Y \sim \text{Pois}(\lambda)$, then $E[Y] = \lambda$ and $V[Y] = \lambda$.

Easy to remember, but the derivation requires some work with power series:

$$
\begin{aligned}
E[Y] &= \sum_{y=0}^{\infty} y p(y) = \sum_{y=0}^{\infty} y \frac{\lambda^y e^{-\lambda}}{y!} \\
&= \sum_{y=1}^{\infty}
\end{aligned}
$$

$$E[Y] = \sum_{y=0}^{\infty} yp(y) = \sum_{y=0}^{\infty} y \frac{\lambda^y e^{-\lambda}}{y!}$$

$$= \sum_{y=1}^{\infty}$$

$$
\begin{aligned}
E[Y] &= \sum_{y=0}^{\infty} y p(y) = \sum_{y=0}^{\infty} y \frac{\lambda^y e^{-\lambda}}{y!} \\
&= \sum_{y=1}^{\infty} y \frac{\lambda^{y-1}\lambda}{y \cdot (y-1)!} e^{-\lambda} = e^{-\lambda} \sum_{y=1}^{\infty} \frac{\lambda^{y-1}\lambda}{(y-1)!} \\
&= e^{-\lambda} \cdot \lambda \cdot \sum_{z=0}^{\infty} \frac{\lambda^z}{z!} \qquad (\text{here } z = y-1) \\
&=
\end{aligned}
$$

$$
\begin{aligned}
E[Y] &= \sum_{y=0}^{\infty} y p(y) = \sum_{y=0}^{\infty} y \frac{\lambda^y e^{-\lambda}}{y!} \\
&= \sum_{y=1}^{\infty} \cancel{y} \frac{\lambda^{y-1}\lambda}{\cancel{y} \cdot (y-1)!} e^{-\lambda} = e^{-\lambda} \sum_{y=1}^{\infty} \frac{\lambda^{y-1}\lambda}{(y-1)!} \\
&= e^{-\lambda} \cdot \lambda \cdot \sum_{z=0}^{\infty} \frac{\lambda^z}{z!} \qquad (\text{here } z = y-1) \\
&=
\end{aligned}
$$

$$E[Y] = \sum_{y=0}^{\infty} yp(y) = \sum_{y=0}^{\infty} y \frac{\lambda^y e^{-\lambda}}{y!}$$

$$= \sum_{y=1}^{\infty} \cancel{y} \frac{\lambda^{y-1}\lambda}{\cancel{y} \cdot (y-1)!} e^{-\lambda} = e^{-\lambda} \sum_{y=1}^{\infty} \frac{\lambda^{y-1}\lambda}{(y-1)!}$$

$$= e^{-\lambda} \cdot \lambda \cdot \sum_{z=0}^{\infty} \frac{\lambda^z}{z!} \qquad (\text{here } z = y-1)$$

$$=$$

$$E[Y] = \sum_{y=0}^{\infty} yp(y) = \sum_{y=0}^{\infty} y \frac{\lambda^y e^{-\lambda}}{y!}$$

$$= \sum_{y=1}^{\infty} \cancel{y} \frac{\lambda^{y-1}\lambda}{\cancel{y} \cdot (y-1)!} e^{-\lambda} = e^{-\lambda} \sum_{y=1}^{\infty} \frac{\lambda^{y-1}\lambda}{(y-1)!}$$

$$= e^{-\lambda} \cdot \lambda \cdot \sum_{z=0}^{\infty} \frac{\lambda^z}{z!} \qquad (\text{here } z = y - 1)$$

$$=$$

$$
\begin{aligned}
E[Y] &= \sum_{y=0}^{\infty} y p(y) = \sum_{y=0}^{\infty} y \frac{\lambda^y e^{-\lambda}}{y!} \\
&= \sum_{y=1}^{\infty} \cancel{y} \frac{\lambda^{y-1}\lambda}{\cancel{y} \cdot (y-1)!} e^{-\lambda} = e^{-\lambda} \sum_{y=1}^{\infty} \frac{\lambda^{y-1}\lambda}{(y-1)!} \\
&= e^{-\lambda} \cdot \lambda \cdot \sum_{z=0}^{\infty} \frac{\lambda^z}{z!} \qquad \text{(here } z = y-1\text{)} \\
&= e^{-\lambda} \cdot \lambda \cdot e^{\lambda}
\end{aligned}
$$

$$
\begin{aligned}
E[Y] &= \sum_{y=0}^{\infty} y p(y) = \sum_{y=0}^{\infty} y \frac{\lambda^y e^{-\lambda}}{y!} \\
&= \sum_{y=1}^{\infty} \cancel{y} \frac{\lambda^{y-1}\lambda}{\cancel{y} \cdot (y-1)!} e^{-\lambda} = e^{-\lambda} \sum_{y=1}^{\infty} \frac{\lambda^{y-1}\lambda}{(y-1)!} \\
&= e^{-\lambda} \cdot \lambda \cdot \sum_{z=0}^{\infty} \frac{\lambda^z}{z!} \qquad (\text{here } z = y-1) \\
&= \cancel{e^{-\lambda}} \cdot \lambda \cdot \cancel{e^{\lambda}} = \lambda.
\end{aligned}
$$

The derivation of $V[Y]$ involves similar tricks.

Note:

This result is important in understanding problems which say "Y is a Poisson distributed with average rate ____".

Moment Generating Functions

Definitions:

- The k^{th} moment (also moment about the origin) of a random variable X is $\mu'_k := E[X^k]$.
- The k^{th} central moment (or moment about the mean) of X is $\mu_k := E[(X-\mu)^k]$, where $\mu = E[X]$.

Notation: moment $= \mu'_k$, central moment $= \mu_k$.

Definition:

The moment generating function of a RV X is $m_X(t) = E[e^{tX}]$. This is a function of t.

Remark:

We do this because the distribution is determined by the moment generating function (MGF).

How is this of any use?
Sometimes we can determine the MGF of an unknown distribution.
We will use this technique in the proof of the Central Limit Theorem.

Remark:

Knowing the moments μ'_k and knowing the MGF are equivalent, using the power series for e^{tX}:

$$e^{tX} = 1 + tX + \frac{(tX)^2}{2!} + \cdots = \sum_{k=0}^{\infty} \frac{(tX)^k}{k!}$$

$$\begin{aligned}
\therefore \quad E[e^{tX}] &= \sum_x p(x)e^{tx} = \sum_x p(x) \left(\sum_{k=0}^{\infty} \frac{(tx)^k}{k!} \right) \\
&= \sum_{k=0}^{\infty} \left(\sum_x p(x)x^k \right) \frac{(t)^k}{k!} = \sum_{k=0}^{\infty} E[X^k] \frac{(t)^k}{k!} \\
&= \sum_{k=0}^{\infty} \mu'_k \frac{(t)^k}{k!}.
\end{aligned}$$

What about other random variables?

Let Y be a Bernoulli RV with parameter p.
Then

$$E[e^{tY}] = \sum_y p(y)e^{ty} = pe^{t(1)} + qe^{t(0)} = q + pe^t.$$

Now suppose $Y \sim \text{Bin}(n, p)$.
Then Y is the sum of n independent Bernoulli RVs, each with parameter p.

Remark:

If X and Y are independent, then $E[XY] = E[X]E[Y]$.
Thus if X and Y are independent, we compute

$$\begin{aligned} m_{X+Y}(t) &= E[e^{t(X+Y)}] = E[e^{tX}e^{tY}] \\ &= E[e^{tX}]E[e^{tY}] \qquad \text{(by independence)} \\ &= m_X(t)m_Y(t). \end{aligned}$$

So if $Y \sim \text{Bin}(n, p)$, then $\boxed{m_Y(t) = (q + pe^t)^n}$.

We could, of course, do exercises from Calculus 2 to get this result without the remark.

Remark on rigor:

In the text, we may exchange the order of limits without justification. This does not generally work, but works in the context.

Example:

$$\lim_{n\to\infty}\lim_{m\to\infty}\frac{n}{n+m}=\lim_{n\to\infty}0=0$$

We could, of course, do exercises from Calculus 2 to get this result without the remark.

Remark on rigor:

In the text, we may exchange the order of limits without justification. This does not generally work, but works in the context.

Example:

$$\lim_{n\to\infty}\lim_{m\to\infty}\frac{n}{n+m}=\lim_{n\to\infty}0=0$$

We could, of course, do exercises from Calculus 2 to get this result without the remark.

Remark on rigor:

In the text, we may exchange the order of limits without justification. This does not generally work, but works in the context.

Example:

$$\lim_{n\to\infty} \lim_{m\to\infty} \frac{n}{n+m} = \lim_{n\to\infty} 0 = 0$$

We could, of course, do exercises from Calculus 2 to get this result without the remark.

Remark on rigor:

In the text, we may exchange the order of limits without justification. This does not generally work, but works in the context.

Example:

$$\lim_{n\to\infty}\lim_{m\to\infty}\frac{n}{n+m}=\lim_{n\to\infty}0=0$$

$$\lim_{m\to\infty}\lim_{n\to\infty}\frac{n}{n+m}=\lim_{m\to\infty}1=1.$$

Where did we do this?

Note that an infinite sum is a limit:

$$\sum_{n=1}^{\infty}a_n\stackrel{\text{def}}{=}\lim_{N\to\infty}\left(\sum_{n=1}^{N}a_n\right).$$

We could, of course, do exercises from Calculus 2 to get this result without the remark.

Remark on rigor:

In the text, we may exchange the order of limits without justification. This does not generally work, but works in the context.

Example:

$$\lim_{n\to\infty} \lim_{m\to\infty} \frac{n}{n+m} = \lim_{n\to\infty} 0 = 0$$

$$\not\parallel$$

$$\lim_{m\to\infty} \lim_{n\to\infty} \frac{n}{n+m} = \lim_{m\to\infty} 1 = 1.$$

Where did we do this?

Note that an infinite sum is a limit:

$$\sum_{n=1}^{\infty} a_n \stackrel{\text{def}}{=} \lim_{N\to\infty} \left(\sum_{n=1}^{N} a_n \right).$$

A derivative is a limit:

$$\frac{d}{dx}f(x)) = \lim_{h\to 0}\left[\frac{f(x+h)-f(x)}{h}\right].$$

In computing the mean of a geometric RV, we exchanged derivative and infinite sum.

Remark:

We will show for the geometric RV that

$$m(t) = \frac{pe^t}{1-qe^t}.$$

Using the connection between the geometric and the negative binomial RVs, namely that the negative binomial RV is the sum of r independent geometric RVs, we get, for the negative binomial RV,

$$m(t) = \left(\frac{pe^t}{1-qe^t}\right)^r.$$

Problem:

Players A and B compete in a game in which they alternate throws of a die. The objective is to be the first player to throw a 1.

A goes first. What is the probability that B wins?

One way:

Apply what we know about the geometric RV.

Let Y be the number of the turn on which the game ends.
Then Y is geometric, with $p = 1/6$.
Let A be the event that the player A wins. Then

$$A = \{Y = 1, 3, 5, \ldots \text{(an odd number)}\}.$$

Let B be the event that B wins, that is, Y is an even number.

For a geometric RV, $P(Y = k) = q^{k-1}p$.
Then

$$P(B) = \sum_{k \text{ even}} q^{k-1}p = \sum_{l=1}^{\infty} q^{2l-1}p = \sum_{m=0}^{\infty} q^{2m+1}p$$

$$= pq \sum_{m=0}^{\infty} (q^2)^m = pq \cdot \frac{1}{1-q^2}.$$

Solution: (continued)

Now plug in $p = 1/6, q = 5/6$ and find

$$P(B) = \frac{1}{6}\frac{5}{6} \cdot \frac{1}{1 - 25/36} = \frac{5}{36}\frac{1}{(11/36)} = \frac{5}{11}.$$

Other ways to think about this:

$P(B) = 1 - P(A)$, so we can compute $P(A)$ instead.

Also, instead of using what we know about geometric RVs, we could write down directly the probability that B wins on turn $2 + B$ wins on turn $4 + \dots$:

$$P(B \text{ wins on turn } 2) = \begin{array}{l} P(A \text{ doesn't win on turn } 1) \cdot \\ \quad P(B \text{ wins at } 2 \mid A \text{ doesn't win at } 1) \end{array} = \frac{5}{6}\frac{1}{6},$$

and similarly for the rest of the terms.

This is basically re-deriving the probability function for the geometric RV.

Recall:

We studied the "moment generating function" $m_Y(t) = E[e^{tY}]$.
We saw that the "central moments" $\mu'_k = E[Y^k]$ are related to this function:
Expand $E[e^{tY}]$ as a power series and the moments appear:

$$
\begin{aligned}
E[e^{tY}] &= \sum_y e^{ty} p(y) = \sum_y \sum_{k=0}^{\infty} \frac{(ty)^k}{k!} p(y) \\
&= \sum_{k=0}^{\infty} \sum_y \frac{t^k y^k}{k!} p(y) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \left(\sum_y y^k p(y) \right) \\
&= \sum_{k=0}^{\infty} \frac{t^k}{k!} E[Y^k] = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mu'_k.
\end{aligned}
$$

Consequence: $m_Y^{(k)}(0) = \mu'_k = E[Y^k]$.

In particular, $m_Y^{(1)}(0) = \mu_1' = E[Y]$.

The series is $\frac{t^0}{0!}\mu_0' + \frac{t^1}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + \dots$.

Take one derivative: $0 + \mu_1' + \frac{2t}{2!}\mu_2' + \frac{3t^2}{3!}\mu_3' + \dots$.

Evaluate at 0: $\mu_1' + 0 + 0 + \cdots = \mu_1'$.

Continue in this way to show $m_Y^{(k)}(0) = \mu_k' = E[Y^k]$.

An example of this phenomenon: the geometric RV.
What is the MGF of $Y \sim \text{Geom}(p)$?
Let's derive this now (Exercise 3.147):

$$E[e^{tY}] = \sum_y e^{ty} p(y) = \sum_{y=1}^{\infty} e^{ty} q^{y-1} p = \sum_{z=0}^{\infty} e^{t(z+1)} q^z p$$

$$= \sum_{z=0}^{\infty} p e^t e^{tz} q^z = \sum_{z=0}^{\infty} p e^t (e^t q)^z = p e^t \frac{1}{1 - q e^t}$$

$$=$$

In particular, $m_Y^{(1)}(0) = \mu_1' = E[Y]$.

The series is $\frac{t^0}{0!}\mu_0' + \frac{t^1}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + \dots$.

Take one derivative: $0 + \mu_1' + \frac{2t}{2!}\mu_2' + \frac{3t^2}{3!}\mu_3' + \dots$.

Evaluate at 0: $\mu_1' + 0 + 0 + \cdots = \mu_1'$.

Continue in this way to show $m_Y^{(k)}(0) = \mu_k' = E[Y^k]$.

An example of this phenomenon: the geometric RV.
What is the MGF of $Y \sim \text{Geom}(p)$?
Let's derive this now (Exercise 3.147):

$$E[e^{tY}] = \sum_y e^{ty} p(y) = \sum_{y=1}^{\infty} e^{ty} q^{y-1} p = \sum_{z=0}^{\infty} e^{t(z+1)} q^z p$$

$$= \sum_{z=0}^{\infty} pe^t e^{tz} q^z = \sum_{z=0}^{\infty} pe^t (e^t q)^z = pe^t \frac{1}{1 - qe^t}$$

$$=$$

In particular, $m_Y^{(1)}(0) = \mu_1' = E[Y]$.

The series is $\frac{t^0}{0!}\mu_0' + \frac{t^1}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + \dots$.

Take one derivative: $0 + \mu_1' + \frac{2t}{2!}\mu_2' + \frac{3t^2}{3!}\mu_3' + \dots$.

Evaluate at 0: $\mu_1' + 0 + 0 + \cdots = \mu_1'$.

Continue in this way to show $m_Y^{(k)}(0) = \mu_k' = E[Y^k]$.

An example of this phenomenon: the geometric RV.
What is the MGF of $Y \sim \mathrm{Geom}(p)$?
Let's derive this now (Exercise 3.147):

$$E[e^{tY}] = \sum_y e^{ty} p(y) = \sum_{y=1}^{\infty} e^{ty} q^{y-1} p = \sum_{z=0}^{\infty} e^{t(z+1)} q^z p$$

$$= \sum_{z=0}^{\infty} pe^t e^{tz} q^z = \sum_{z=0}^{\infty} pe^t (e^t q)^z = pe^t \frac{1}{1-qe^t}$$

$$=$$

In particular, $m_Y^{(1)}(0) = \mu_1' = E[Y]$.

The series is $\frac{t^0}{0!}\mu_0' + \frac{t^1}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + \dots$.

Take one derivative: $0 + \mu_1' + \frac{2t}{2!}\mu_2' + \frac{3t^2}{3!}\mu_3' + \dots$.

Evaluate at 0: $\mu_1' + 0 + 0 + \cdots = \mu_1'$.

Continue in this way to show $m_Y^{(k)}(0) = \mu_k' = E[Y^k]$.

An example of this phenomenon: the geometric RV.
What is the MGF of $Y \sim \text{Geom}(p)$?
Let's derive this now (Exercise 3.147):

$$E[e^{tY}] = \sum_y e^{ty} p(y) = \sum_{y=1}^{\infty} e^{ty} q^{y-1} p = \sum_{z=0}^{\infty} e^{t(z+1)} q^z p$$

$$= \sum_{z=0}^{\infty} pe^t e^{tz} q^z = \sum_{z=0}^{\infty} pe^t (e^t q)^z = pe^t \frac{1}{1 - qe^t}$$

$$=$$

In particular, $m_Y^{(1)}(0) = \mu_1' = E[Y]$.

The series is $\frac{t^0}{0!}\mu_0' + \frac{t^1}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + \dots$.

Take one derivative: $0 + \mu_1' + \frac{2t}{2!}\mu_2' + \frac{3t^2}{3!}\mu_3' + \dots$.

Evaluate at 0: $\mu_1' + 0 + 0 + \cdots = \mu_1'$.

Continue in this way to show $m_Y^{(k)}(0) = \mu_k' = E[Y^k]$.

An example of this phenomenon: the geometric RV.
What is the MGF of $Y \sim \text{Geom}(p)$?
Let's derive this now (Exercise 3.147):

$$
\begin{aligned}
E[e^{tY}] &= \sum_y e^{ty} p(y) = \sum_{y=1}^{\infty} e^{ty} q^{y-1} p = \sum_{z=0}^{\infty} e^{t(z+1)} q^z p \\
&= \sum_{z=0}^{\infty} pe^t e^{tz} q^z = \sum_{z=0}^{\infty} pe^t (e^t q)^z = pe^t \frac{1}{1 - qe^t} \\
&=
\end{aligned}
$$

In particular, $m_Y^{(1)}(0) = \mu_1' = E[Y]$.

The series is $\frac{t^0}{0!}\mu_0' + \frac{t^1}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + \dots$.

Take one derivative: $0 + \mu_1' + \frac{2t}{2!}\mu_2' + \frac{3t^2}{3!}\mu_3' + \dots$.

Evaluate at 0: $\mu_1' + 0 + 0 + \cdots = \mu_1'$.

Continue in this way to show $m_Y^{(k)}(0) = \mu_k' = E[Y^k]$.

An example of this phenomenon: the geometric RV.
What is the MGF of $Y \sim \text{Geom}(p)$?
Let's derive this now (Exercise 3.147):

$$E[e^{tY}] = \sum_y e^{ty} p(y) = \sum_{y=1}^{\infty} e^{ty} q^{y-1} p = \sum_{z=0}^{\infty} e^{t(z+1)} q^z p$$

$$= \sum_{z=0}^{\infty} pe^t e^{tz} q^z = \sum_{z=0}^{\infty} pe^t (e^t q)^z = pe^t \frac{1}{1 - qe^t}$$

$$=$$

In particular, $m_Y^{(1)}(0) = \mu_1' = E[Y]$.

The series is $\frac{t^0}{0!}\mu_0' + \frac{t^1}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + \dots$.

Take one derivative: $0 + \mu_1' + \frac{2t}{2!}\mu_2' + \frac{3t^2}{3!}\mu_3' + \dots$.

Evaluate at 0: $\mu_1' + 0 + 0 + \cdots = \mu_1'$.

Continue in this way to show $m_Y^{(k)}(0) = \mu_k' = E[Y^k]$.

An example of this phenomenon: the geometric RV.
What is the MGF of $Y \sim \text{Geom}(p)$?
Let's derive this now (Exercise 3.147):

$$E[e^{tY}] = \sum_y e^{ty}p(y) = \sum_{y=1}^{\infty} e^{ty}q^{y-1}p = \sum_{z=0}^{\infty} e^{t(z+1)}q^z p$$

$$= \sum_{z=0}^{\infty} pe^t e^{tz}q^z = \sum_{z=0}^{\infty} pe^t(e^t q)^z = pe^t \frac{1}{1-qe^t}$$

$$=$$

In particular, $m_Y^{(1)}(0) = \mu_1' = E[Y]$.

The series is $\frac{t^0}{0!}\mu_0' + \frac{t^1}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + \dots$.

Take one derivative: $0 + \mu_1' + \frac{2t}{2!}\mu_2' + \frac{3t^2}{3!}\mu_3' + \dots$.

Evaluate at 0: $\mu_1' + 0 + 0 + \dots = \mu_1'$.

Continue in this way to show $m_Y^{(k)}(0) = \mu_k' = E[Y^k]$.

An example of this phenomenon: the geometric RV.
What is the MGF of $Y \sim \text{Geom}(p)$?
Let's derive this now (Exercise 3.147):

$$E[e^{tY}] = \sum_y e^{ty} p(y) = \sum_{y=1}^{\infty} e^{ty} q^{y-1} p = \sum_{z=0}^{\infty} e^{t(z+1)} q^z p$$

$$= \sum_{z=0}^{\infty} pe^t e^{tz} q^z = \sum_{z=0}^{\infty} pe^t (e^t q)^z = pe^t \frac{1}{1 - qe^t}$$

$$=$$

In particular, $m_Y^{(1)}(0) = \mu_1' = E[Y]$.

The series is $\frac{t^0}{0!}\mu_0' + \frac{t^1}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + \dots$.

Take one derivative: $0 + \mu_1' + \frac{2t}{2!}\mu_2' + \frac{3t^2}{3!}\mu_3' + \dots$.

Evaluate at 0: $\mu_1' + 0 + 0 + \cdots = \mu_1'$.

Continue in this way to show $m_Y^{(k)}(0) = \mu_k' = E[Y^k]$.

An example of this phenomenon: the geometric RV.
What is the MGF of $Y \sim \text{Geom}(p)$?
Let's derive this now (Exercise 3.147):

$$E[e^{tY}] = \sum_y e^{ty} p(y) = \sum_{y=1}^{\infty} e^{ty} q^{y-1} p = \sum_{z=0}^{\infty} e^{t(z+1)} q^z p$$

$$= \sum_{z=0}^{\infty} pe^t e^{tz} q^z = \sum_{z=0}^{\infty} pe^t (e^t q)^z = pe^t \frac{1}{1 - qe^t}$$

$$=$$

In particular, $m_Y^{(1)}(0) = \mu_1' = E[Y]$.

The series is $\frac{t^0}{0!}\mu_0' + \frac{t^1}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + \dots$.

Take one derivative: $0 + \mu_1' + \frac{2t}{2!}\mu_2' + \frac{3t^2}{3!}\mu_3' + \dots$.

Evaluate at 0: $\mu_1' + 0 + 0 + \cdots = \mu_1'$.

Continue in this way to show $m_Y^{(k)}(0) = \mu_k' = E[Y^k]$.

An example of this phenomenon: the geometric RV.
What is the MGF of $Y \sim \text{Geom}(p)$?
Let's derive this now (Exercise 3.147):

$$E[e^{tY}] = \sum_y e^{ty} p(y) = \sum_{y=1}^{\infty} e^{ty} q^{y-1} p = \sum_{z=0}^{\infty} e^{t(z+1)} q^z p$$
$$= \sum_{z=0}^{\infty} pe^t e^{tz} q^z = \sum_{z=0}^{\infty} pe^t (e^t q)^z = pe^t \frac{1}{1 - qe^t}$$
$$=$$

In particular, $m_Y^{(1)}(0) = \mu_1' = E[Y]$.

The series is $\frac{t^0}{0!}\mu_0' + \frac{t^1}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + \dots$.

Take one derivative: $0 + \mu_1' + \frac{2t}{2!}\mu_2' + \frac{3t^2}{3!}\mu_3' + \dots$.

Evaluate at 0: $\mu_1' + 0 + 0 + \cdots = \mu_1'$.

Continue in this way to show $m_Y^{(k)}(0) = \mu_k' = E[Y^k]$.

An example of this phenomenon: the geometric RV.
What is the MGF of $Y \sim \text{Geom}(p)$?
Let's derive this now (Exercise 3.147):

$$E[e^{tY}] = \sum_y e^{ty} p(y) = \sum_{y=1}^{\infty} e^{ty} q^{y-1} p = \sum_{z=0}^{\infty} e^{t(z+1)} q^z p$$

$$= \sum_{z=0}^{\infty} pe^t e^{tz} q^z = \sum_{z=0}^{\infty} pe^t (e^t q)^z = pe^t \frac{1}{1 - qe^t}$$

$$=$$

In particular, $m_Y^{(1)}(0) = \mu_1' = E[Y]$.

The series is $\frac{t^0}{0!}\mu_0' + \frac{t^1}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + \dots$.

Take one derivative: $0 + \mu_1' + \frac{2t}{2!}\mu_2' + \frac{3t^2}{3!}\mu_3' + \dots$.

Evaluate at 0: $\mu_1' + 0 + 0 + \cdots = \mu_1'$.

Continue in this way to show $m_Y^{(k)}(0) = \mu_k' = E[Y^k]$.

An example of this phenomenon: the geometric RV.
What is the MGF of $Y \sim \text{Geom}(p)$?
Let's derive this now (Exercise 3.147):

$$E[e^{tY}] = \sum_y e^{ty} p(y) = \sum_{y=1}^{\infty} e^{ty} q^{y-1} p = \sum_{z=0}^{\infty} e^{t(z+1)} q^z p$$

$$= \sum_{z=0}^{\infty} pe^t e^{tz} q^z = \sum_{z=0}^{\infty} pe^t (e^t q)^z = pe^t \frac{1}{1 - qe^t}$$

$$=$$

In particular, $m_Y^{(1)}(0) = \mu_1' = E[Y]$.

The series is $\frac{t^0}{0!}\mu_0' + \frac{t^1}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + \dots$.

Take one derivative: $0 + \mu_1' + \frac{2t}{2!}\mu_2' + \frac{3t^2}{3!}\mu_3' + \dots$.

Evaluate at 0: $\mu_1' + 0 + 0 + \dots = \mu_1'$.

Continue in this way to show $m_Y^{(k)}(0) = \mu_k' = E[Y^k]$.

An example of this phenomenon: the geometric RV.
What is the MGF of $Y \sim \text{Geom}(p)$?
Let's derive this now (Exercise 3.147):

$$E[e^{tY}] = \sum_y e^{ty}p(y) = \sum_{y=1}^{\infty} e^{ty}q^{y-1}p = \sum_{z=0}^{\infty} e^{t(z+1)}q^z p$$

$$= \sum_{z=0}^{\infty} pe^t e^{tz} q^z = \sum_{z=0}^{\infty} pe^t (e^t q)^z = pe^t \frac{1}{1-qe^t}$$

$$=$$

In particular, $m_Y^{(1)}(0) = \mu_1' = E[Y]$.

The series is $\frac{t^0}{0!}\mu_0' + \frac{t^1}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + \dots$.

Take one derivative: $0 + \mu_1' + \frac{2t}{2!}\mu_2' + \frac{3t^2}{3!}\mu_3' + \dots$.

Evaluate at 0: $\mu_1' + 0 + 0 + \cdots = \mu_1'$.

Continue in this way to show $m_Y^{(k)}(0) = \mu_k' = E[Y^k]$.

An example of this phenomenon: the geometric RV.
What is the MGF of $Y \sim \text{Geom}(p)$?
Let's derive this now (Exercise 3.147):

$$E[e^{tY}] = \sum_y e^{ty} p(y) = \sum_{y=1}^{\infty} e^{ty} q^{y-1} p = \sum_{z=0}^{\infty} e^{t(z+1)} q^z p$$

$$= \sum_{z=0}^{\infty} pe^t e^{tz} q^z = \sum_{z=0}^{\infty} pe^t (e^t q)^z = pe^t \frac{1}{1 - qe^t}$$

$$=$$

In particular, $m_Y^{(1)}(0) = \mu_1' = E[Y]$.

The series is $\frac{t^0}{0!}\mu_0' + \frac{t^1}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + \dots$.

Take one derivative: $0 + \mu_1' + \frac{2t}{2!}\mu_2' + \frac{3t^2}{3!}\mu_3' + \dots$.

Evaluate at 0: $\mu_1' + 0 + 0 + \cdots = \mu_1'$.

Continue in this way to show $m_Y^{(k)}(0) = \mu_k' = E[Y^k]$.

An example of this phenomenon: the geometric RV.
What is the MGF of $Y \sim \text{Geom}(p)$?
Let's derive this now (Exercise 3.147):

$$
\begin{aligned}
E[e^{tY}] &= \sum_y e^{ty} p(y) = \sum_{y=1}^{\infty} e^{ty} q^{y-1} p = \sum_{z=0}^{\infty} e^{t(z+1)} q^z p \\
&= \sum_{z=0}^{\infty} pe^t e^{tz} q^z = \sum_{z=0}^{\infty} pe^t (e^t q)^z = pe^t \frac{1}{1-qe^t} \\
&= \frac{pe^t}{1-qe^t} = m_Y(t)
\end{aligned}
$$

In particular, $m_Y^{(1)}(0) = \mu_1' = E[Y]$.

The series is $\frac{t^0}{0!}\mu_0' + \frac{t^1}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + \ldots$.

Take one derivative: $0 + \mu_1' + \frac{2t}{2!}\mu_2' + \frac{3t^2}{3!}\mu_3' + \ldots$.

Evaluate at 0: $\mu_1' + 0 + 0 + \cdots = \mu_1'$.

Continue in this way to show $m_Y^{(k)}(0) = \mu_k' = E[Y^k]$.

An example of this phenomenon: the geometric RV.
What is the MGF of $Y \sim \text{Geom}(p)$?
Let's derive this now (Exercise 3.147):

$$E[e^{tY}] = \sum_y e^{ty}p(y) = \sum_{y=1}^{\infty} e^{ty}q^{y-1}p = \sum_{z=0}^{\infty} e^{t(z+1)}q^z p$$

$$= \sum_{z=0}^{\infty} pe^t e^{tz} q^z = \sum_{z=0}^{\infty} pe^t (e^t q)^z = pe^t \frac{1}{1-qe^t}$$

$$= \boxed{\frac{pe^t}{1-qe^t} = m_Y(t)}.$$

What about $m'_Y(0)$?
We should get $E[Y]$.
We find

$$m'_Y(t) = \frac{pe^t \cdot (1 - qe^t) - pe^t \cdot (-qe^t)}{(1 - qe^t)^2}.$$

Evaluate this at 0:
use $e^0 = 1$ and get

$$m'_Y(0) = \frac{p(1-q) - p(-q)}{(1-q)^2} = \frac{p^2 + pq}{p^2} = \frac{p(p+q)}{p^2} = \frac{1}{p};$$

indeed this is $E[Y]$.

Notes:

- This may or may not be a good way to compute $E[Y^k]$.
- Using the identity $V[Y] = E[Y^2] - E[Y]^2$, we can compute $V[Y] = m_Y^{(2)}(0) - [m'_Y(0)]^2$.

Caution: $V[Y]$ is not $m_Y^{(2)}(0)$, unless $E[Y] = 0$.

Tchebysheff's Theorem

Theorem (Tchebysheff)

Let Y be a RV with mean μ and variance σ^2. Then for any $k > 0$,

$$P(|Y-\mu| \geq k\sigma) \leq \frac{1}{k^2}, \quad P(|Y-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

Note that the two events are complementary: if $|Y-\mu|$ is not $\geq k\sigma$, then $|Y-\mu| < k\sigma$.

What does this really say?

Let's set $k = 3$.
Then the first statement says

$$P(|Y-\mu| \geq 3\sigma) \leq \frac{1}{3^2} = \frac{1}{9}.$$

The result says that $P(Y$ is far from mean$)$ is small.
In particular, the probability that Y is 3 or more standard deviations away from its mean is less than or equal to $\frac{1}{9}$.

Remarks:

1. The theorem has a hypothesis that Y has a mean and variance.
   Not every RV has a mean and variance.
   Virtually all of the RVs we study will have means and variances.

2. We said nothing else about Y, so this applies to all the RVs we have studied, plus anything you can imagine (with a μ and σ^2).

3. This bound is sharp, in the sense that it cannot be improved for general RVs.

4. The bound is weak, in the sense that it can be greatly improved with the knowlegde of the distribution.

Example

Let

$$Y = \begin{cases} 0 & \text{with probability } 16/18 \\ 1 & \text{with probability } 1/18 \\ -1 & \text{with probability } 1/18 \end{cases}$$

Then $E[Y] = 0, V[Y] = E[Y^2] - E[Y]^2 = \frac{1}{9}$.

So in applying Tchebysheff, $\mu = 0, \sigma^2 = \frac{1}{9}$.

The theorem says $P(|Y - 0| \geq 3\sigma) \leq \frac{1}{9}$.

Here $\sigma = \sqrt{\frac{1}{9}} = \frac{1}{3}$, so $3\sigma = 3 \cdot \frac{1}{3} = 1$.

So $P(|Y| \geq 1) \leq \frac{1}{9}$.

But for our RV Y, $P(|Y| \geq 1) = \frac{1}{18} + \frac{1}{18} = \frac{1}{9}$. So the bound cannot be possibly improved.

Remark:

We specialized $k = 3$ for clarity. But a similar example can be

Example (Exercise 3.123)

The random variable Y has a Poisson distribution and is such that $p(0) = p(1)$. What is $p(2)$?

Solution:

$Y \sim$ Poisson(λ), so $p(y) = \dfrac{\lambda^y e^{-\lambda}}{y!}$.

We are given $p(0) = p(1)$. So

$$\frac{\lambda^0 e^{-\lambda}}{0!} = \frac{\lambda^1 e^{-\lambda}}{1!} \implies \lambda^0 = \lambda^1 \implies \lambda = 1.$$

Thus

$$p(2) = \frac{1^2 e^{-1}}{2!} = \frac{e^{-1}}{2} = \boxed{\frac{1}{2e}}.$$

Notice that we had to know the probability function for the Poisson RV.

Another thing to try:
Derive!

Examples: MGF of geometric RV, the mean and variance of the Poisson RV (and lots of ther examples).

MGF of the Geometric RV

What is the probability function of the geometric RV?

$$p(y) = q^{y-1}p \qquad y = 1, 2, 3, \ldots.$$

What is the definition of MGF?

$$m_Y(t) = E[e^{tY}] \overset{\text{Thm 3.2}}{=} \sum_y p(y)e^{ty} = \sum_{y=1}^{\infty} q^{y-1}pe^{ty}$$

$$= \sum_{z=0}^{\infty} q^z pe^{t(z+1)} = \sum_{z=0}^{\infty} e^{tz}q^z pe^t = pe^t \sum_{z=0}^{\infty} (e^t q)^z$$

$$= \frac{pe^t}{1 - qe^t}.$$

Famous Problem: "St. Petersburg Paradox"

You are to play a game where a fair coin is flipped repeatedly. If the first flip is heads, you get pais \$ 1 and the game ends. If the first flip is tails and the second is heads, you get \$2. If the first 2 flips are tails and the third is heads, you get \$4; and so on.

1. How much should you be willing to pay to play this game?
2. How much should you charge if someone else wants to play and you are responsible for the payouts?

Interpretation:

We are looking at the expectation of a function of a geometric RV.

Let $Y \sim \text{Geom}\left(\frac{1}{2}\right)$.

The naïve "fair value" of this game is $E[2^{Y-1}]$.

Notice that

$$E[2^{Y-1}] = \sum_{y=1}^{\infty} q^{y-1} p 2^{y-1} = \frac{1}{2} \sum_{z=0}^{\infty} \left(\frac{1}{2} \cdot 2\right)^z$$

$$= \frac{1}{2} \sum_{z=0}^{\infty} 1 = \frac{1}{2}(1 + 1 + \dots) \to \infty.$$

Before we move on to Chapter 4:

Challenging problems in Probability: “Interview Puzzles”.

Examples

1. The Chow-Robbins Game (From the 1^{st} slide).
2. 100 passengers get on a plane with 100 seats. The first passenger has lost his boarding pass and chooses a seat at random. Subsequent passengers sit in their assigned seats (if empty) or choose a seat at random (if the assigned seat is occupied). What is the probability that the 100^{th} passenger is able to sit in their assigned seat?
3. 4 points are chosen at random on the unit sphere in $\mathbb{R}^3$. They form a tetrahedron. What is the probability that the origin (the center of the sphere) lies in the tetrahedron?

So far we have only treated "discrete probability". In Chapter 4, we will discuss continuous probability.

Remark:

What does a random point on the surface of a sphere mean?

"Random point on the surface of the sphere" means that the probability that the point lies in a set S is proportional to the area of S.

Second thing which is not part of this course but very important in practice:
Statistical software.
In practice "R" is most common, and there are many online courses on how to use R.

Knowing R is an "employable skill"!.

Almost every question we discuss here is empirical: it can be addressed by experiment and simulation.
This exposition is to give you enough background to understand what R is doing.

End of Chapter 3

Chapter 4

Continuous Variables and Their Probability Distributions

Continuous Probability

If we have a random variable Y, we can define a real-valued function $F_Y : \mathbb{R} \to \mathbb{R}$:

$$F_Y(y) = P(Y \leq y).$$

Example

Suppose $Y \sim \text{Bin}\left(2, \frac{1}{2}\right)$.
The probability function of Y is

$$p(0) = \frac{1}{4}, \quad p(1) = \frac{1}{2}, \quad p(2) = \frac{1}{4}.$$

What is F_Y?
Try to draw a graph!

First note that if $y < 0$, then $P(Y \leq y) = 0$.
So $F_Y(y) = 0$ for $y < 0$.

What is $F_Y(0)$?

$$F_Y(0) \overset{\text{def}}{=} P(Y \leq 0) = \frac{1}{4}.$$

Example (continued)

For $0 < y < 1$, what is $F_Y(y)$?

$$F_Y(y) \stackrel{\text{def}}{=} P(Y \leq y) = \frac{1}{4},$$

because $\text{Bin}\left(2, \frac{1}{2}\right)$ can only take values 0, 1, 2.

What is $F_Y(1)$?

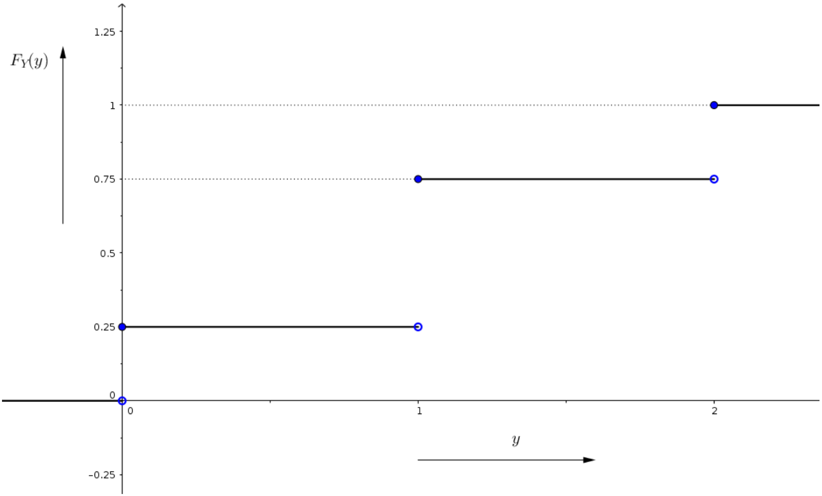
$$\begin{aligned} F_Y(1) &= P(Y \leq 1) = P(Y = 0) + P(Y = 1) \\ &= p(0) + p(1) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}. \end{aligned}$$

And if $1 < y < 2$, then $F_Y(y) = \frac{3}{4}$.

What is $F_Y(2)$?

$$F_Y(2) = P(Y \leq 2) = P(Y = 0) + P(Y = 1) + P(Y = 2) = 1.$$

Graph of $F_Y(y)$:



This F_Y is called the "Distribution function" of Y (sometimes called "Cumulative Distribution Function").

For all the random variables we have studied thus far, it has this step-function structure.

Remarks:

1. We could just as well define a RV by giving F_Y rather than the probability function.
2. If we define an RV by a CDF F_Y, then a <u>continuous random variable</u> is one for which the function F_Y is continuous.
3. If we define a RV this way, then for any interval $[a, b]$, we can write $P(a \leq y \leq b) = F_Y(b) - F_Y(a)$.

For a continuous RV Y, what is the probability that $Y = 1$?

$$P(Y = 1) = P(1 \leq Y \leq 1) = F_Y(1) - F_Y(1) = \boxed{0}.$$

Remark:

By definition, $P(Y \le a) = F_Y(a)$.
Thus

$$\begin{aligned} F_Y(b) - F_Y(a) &= P(Y \le b) - P(Y \le a) \\ &= P(\{Y \le b\} \setminus \{Y \le a\}) \\ &= P(a < Y \le b). \end{aligned}$$

But the difference does not matter, because, as we saw before, $P(Y = a) = 0$, and $\{a\} \cap \{a < Y \le b\} = \emptyset$.

Intuition: For a continuous RV Y we really don't want to talk about $P(Y = a)$. Remember the analogy between probability and length (or area).
$P(Y = a)$ is like the length of a single point (zero). But a line segment, which is made up of points, has a nonzero length.
So just as we only want to talk about the length of sets that are "non-discrete" (discrete sets have zero length), for a continuous RV Y we only want to talk about $P(Y \in S)$ if S is a set that makes sense in the context.

Example (Choosing a point at random in the unit interval)

Define

$$F_Y(y) = \begin{cases} 0 & y \le 0 \\ y & 0 < y < 1 \\ 1 & y \ge 1 \end{cases}.$$

Note that for this F_Y,

$$P(0 \le y \le 1) = F_Y(1) - F_Y(0) = 1 - 0 = 1.$$

If $[a, b] \subset [0, 1]$, then

$$P(a \le y \le b) = F_Y(b) - F_Y(a) = b - a = \text{length}([a, b]).$$

So the probability that a point chosen according to this distribution lies in a subinterval $[a, b]$ is proportional to the length of that subinterval.

Such distributions are called Uniform.

Here is another way we could define the uniform distribution on $[0,1]$:

Let's consider the function

$$f_Y(y) = \begin{cases} 1 & y \in [0,1] \\ 0 & y \notin [0,1] \end{cases}.$$

Then

$$F_Y(y) = \int_{-\infty}^{y} f_Y(x)\,dx,$$

and we have

$$F_Y(b) - F_Y(a) = \int_a^b f_Y(x)\,dx.$$

This f_Y is called the "Probability Distribution Function" (PDF) of Y, and we could just as well define Y by giving the PDF.

And this is what we will do for most of Chapter 4. Note by the Fundamental Theorem of Calculus,

$$f_Y(y) = \frac{d}{dy} F_Y(y).$$

We defined a "Cumulative Distribution Function" (CDF) or sometimes just "Distribution Function" of a RV Y to be $F_Y(y) = P(Y \leq y)$.

Notice that $F_Y : \mathbb{R} \to \mathbb{R}$. This is different from the RV Y in the sense that the domain of Y is some "sample space", i.e. $Y : S \to \mathbb{R}$.

Properties of a distribution function:

(1)

$$\lim_{y \to -\infty} F_Y(y) = 0. \qquad \text{[Book writes } F_Y(-\infty) = 0\text{]}$$

(2)

$$\lim_{y \to \infty} F_Y(y) = 1. \qquad \text{[Book writes } F_Y(\infty) = 1\text{]}$$

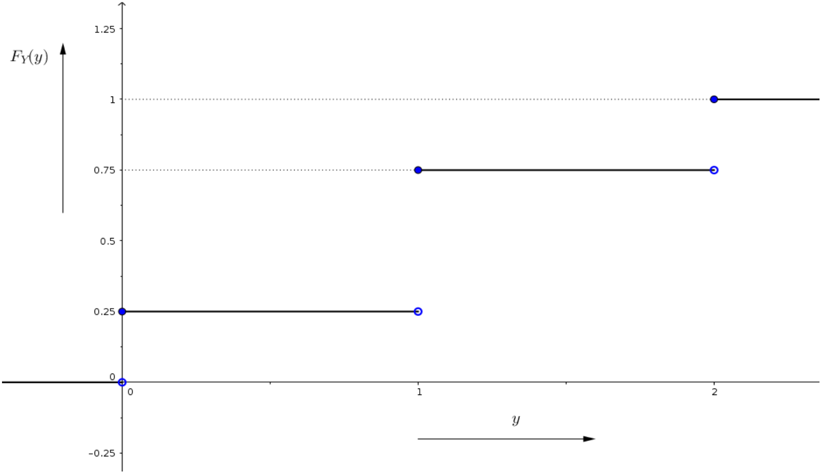
(3) F_Y is nondecreasing: if $y_1 < y_2$, then $F_Y(y_1) \leq F_Y(y_2)$.

(4) F_Y is "right continuous".

Recall that we studied $F(Y)$ where $Y \sim \text{Bin}(2, \frac{1}{2})$.

We saw that F_Y is a step function.

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ 1/4 & 0 \le y < 1 \\ 3/4 & 1 \le y < 2 \\ 1 & y \ge 2. \end{cases}$$



$F(Y)$ is defined for general RVs Y.

Y is said to be "continuous" if $F_Y$is continuous. This is the definition of a "continuous RV".

Counter-intuitive point: For a countinuous RV Y, $P(Y = a) = 0$ for any $a \in \mathbb{R}$.

If $P(Y = a)$ were some non-zero number, say $\frac{1}{10}$, then we would have, for any $y < a$,

$$P(Y \leq a) - P(Y \leq y) \geq \frac{1}{10}.$$

So

$$\lim_{y \to a_-} P(Y \leq y) \leq P(Y \leq a) - \frac{1}{10} < P(Y \leq a).$$

But if F_Y is continuous,

$$\lim_{y \to a_-} P(Y \leq y) = P(Y \leq a).$$

So this is impossible, because if F_Y is continuous,

$$\lim_{y \to a_-} F_Y(y) = F_Y\left(\lim_{y \to a_-} y\right) = F_Y(a).$$

Recall the intuition: Probability is like length or area.
The length of any single point is zero. But the length of $[0, 1]$ (which is made up of points) is 1.

Similarly, the probability (for a continuous RV Y) that $Y = a$ is zero, but $P(Y \in [0, 1])$ can be positive.

We defined the "Probability Density Function" (PDF) (sometimes "Density Function") to be

$$f_Y = \frac{d}{dy} F_Y, \qquad \text{that is, } f_Y(y) = F'_Y(y).$$

In this book we can just assume that F_Y is differentiable.

For a continuous RV Y with CDF F and PDF f, we have

$$P(a \leq y \leq b) = P(a < y < b) = \int_a^b f(y)dy.$$

This follows directly from the definitions we have given:

$$\begin{aligned} \int_{-\infty}^{y} f(x)dx &= F(y) - F(-\infty) \qquad \text{(By FToC)} \\ &= F(y) - 0 \qquad \text{(By properties of CDF)} \\ &= P(Y \leq y). \end{aligned}$$

So

$$\begin{aligned} \int_a^b f(x)dx &= P(Y \leq b) - P(Y \leq a) \\ &= P(a < y \leq b) = P(a \leq y \leq b), \end{aligned}$$

because $P(Y = a)$ is zero.

Mostly we will define our random variables by giving the PDF.

Properties of PDF:

(1)

$$f_Y(y) \geq 0. \qquad \text{[Recall } F_Y \text{ is nondecreasing]}$$

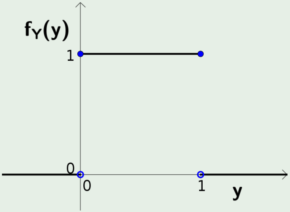
(2)

$$\int_{-\infty}^{\infty} f_Y(y)dy = 1. \quad [= P(-\infty \leq y \leq \infty]$$

Example

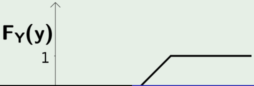
Let's define a RV by giving a PDF.
Suppose

$$f_Y(y) = \begin{cases} 0 & y < 0 \\ 1 & 0 \le y \le 1 \\ 0 & y > 1 \end{cases}$$



Notice that (1) and (2) are satisfied.
Define

$$F_Y(y) = \int_{-\infty}^{Y} f_Y(x)dx = \begin{cases} 0 & y \le 0 \\ y & 0 \le y \le 1 \\ 0 & y \ge 1 \end{cases}$$



Much of the rest of this chapter will go according to the following scheme:

- We define a RV Y by giving some PDF.
- Then we compute the mean and variance of this RV.
- Then you have a bunch of problems in which you use the properties of this RV to figure something out.
- This is much the same as Chapter 3, but all sums will be replaced by integrals, because

Definition 4.5

For a continuous RV Y,

$$E(Y) = \int_{-\infty}^{\infty} y f(y)dy,$$

where f is the PDF of Y.

As before, not every RV has an expectation: we need this integral to be convergent.

- Also, analogously to Chapter 3,

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)f(y)dy.$$

Now we know enough to begin computing means and variances.

We already defined a RV $Y \sim \text{Unif}([0,1])$ by the PDF

$$f_Y = \begin{cases} 1 & y \in [0,1] \\ 0 & y \notin [0,1] \end{cases}$$

What is $E[Y]$?

$$E[Y] = \int_{-\infty}^{\infty} yf(y)dy = \int_{-\infty}^{0} y \cdot 0dy + \int_{0}^{1} y \cdot 1dy + \int_{1}^{\infty} y \cdot 0dy$$
$$= \int_0^1 ydy = \frac{y^2}{2}\bigg|_0^1 = \boxed{\frac{1}{2}}.$$

What is $V[Y]$?

We know that $V[Y] = E[(Y-\mu)^2]$, and the expectation E has the same linearity properties as it did for discrete RVs.

So

$$\begin{aligned} E[(Y-\mu)^2] &= E[Y^2 - 2\mu Y + \mu^2] \\ &= E[Y^2] - 2\mu E[Y] + \mu^2 \qquad \text{(Linearity of } E\text{)} \\ &= E[Y^2] - 2\mu \cdot \mu + \mu^2 = E[Y^2] - \mu^2. \end{aligned}$$

Thus

$$\begin{aligned} V[Y] = E[Y^2] - \mu^2 &= \int_{-\infty}^{\infty} y^2 f(y)dy - \left(\frac{1}{2}\right)^2 \\ &= \int_{-\infty}^{0} y^2 \cdot 0dy + \int_{0}^{1} y^2 \cdot 1dy + \int_{1}^{\infty} y^2 \cdot 0dy - \frac{1}{4} \\ &= \int_0^1 y^2 dy - \frac{1}{4} = \left.\frac{y^3}{3}\right|_0^1 - \frac{1}{4} \\ &= \frac{1}{3} - \frac{1}{4} = \boxed{\frac{1}{12}}. \end{aligned}$$

Exercise 4.11.a

Suppose that Y possesses the density function

$$f(y) = \begin{cases} cy & 0 \leq y \leq 2 \\ 0 & \text{elsewhere.} \end{cases}$$

Find the value of c that makes $f(y)$ a probability density function.

Solution:

Recall that, from property (2) of PDFs,

$$\int_{-\infty}^{\infty} f(y)dy = 1.$$

So we have

$$\int_{0}^{2} cy\, dy = 1,$$

that is,

$$1 = \left.\frac{cy^2}{2}\right|_0^2 = \frac{c(2)^2}{2} - \frac{c(0)^2}{2} = 2c \implies \boxed{c = \frac{1}{2}}.$$

Note that we could also have an undefined c in the definition of a CDF F_Y, which is determined by some property of CDFs, e.g. $\lim_{y \to \infty} F_Y(y) = 1$.

This might give us an equation for c.

Definitions: "Quantile", "Median"

Let Y denote any random variable. If $0 < p < 1$, the p^{th} quantile of Y, denoted by ϕ_p, is the smallest value such that $P(Y \leq \phi_p) = F_Y(\phi_p) \geq p$. If Y is continuous, ϕ_p is the smallest value such that

$$F_Y(\phi_p) = P(Y \leq \phi_p) = p.$$

The quantity corresponding to 0.5, $\phi_{0.5}$, is called the "Median" of Y.

Many problems use this notation.

Remark:

We said that E is linear. But E does not commute with arbitrary functions:

it is <u>not</u> true that $E[g(Y)] = g(E[Y])$.

This only works for linear functions.

Plan:

- We will study a few continuous probability distributions.
- We will find means, variances, and MGFs.
- We will study the continuous version of Tchebysheff's Theorem.

Remark:

Usually the distribution will be defined by the density function (PDF). Recall that a PDF has 2 properties:

(1) $f_Y(y) \geq 0$ for all $y \in \mathbb{R}$.

(2)

$$\int_{-\infty}^{\infty} f_Y(y)dy = 1.$$

The relationship of the density function to the values of the random variable is:

$$P(a \leq Y \leq b) = \int_a^b f_Y(y)dy.$$

Remark:

Any function satisfying properties (1) and (2) is a valid density function, and defines a random variable.
We will study a few special ones (useful in applications).
Thus a random variable is about as general as a function.
Analogy to calculus: In principle, you can talk about the integral of any function, but we study certain special functions that are useful in applications.

The main random variables we will study:

- The Uniform distribution, Unif(S).
- The Normal distribution, $\mathcal{N}(\mu, \sigma^2)$.
- The Beta distribution, Beta(α, β).
- The Gamma distribution, $\Gamma(\alpha, \beta)$.
- The Exponential distribution, Exp(λ).
- The Chi-Squared distribution, $\chi^2(k)$.

In various exercises
you may also see
Weibull distribution,

We defined the uniform distribution on $[0, 1]$ by the PDF

$$f(y) = \begin{cases} 1 & y \in [0, 1] \\ 0 & y \notin [0, 1]. \end{cases}$$

and we computed

$$E[Y] = \boxed{\frac{1}{2}}, V[Y] = \boxed{\frac{1}{12}}.$$

The uniform distribution on $[\theta_1, \theta_2]$ is defined by the PDF

$$f(y) = \begin{cases} \dfrac{1}{\theta_2 - \theta_1} & y \in [\theta_1, \theta_2] \\ 0 & y \notin [\theta_1, \theta_2]. \end{cases}$$

The same computation as before will give us

$$E[Y] = \boxed{\frac{\theta_1 + \theta_2}{2}}, V[Y] = \boxed{\frac{(\theta_2 - \theta_1)^2}{12}}.$$

Question:

If Y is in units v, is $V[Y]$ in units v?

The uniform distribution may seem trivial, but there are many nontrivial problems based on it.

Examples:

1. If we choose 2 points at random in the unit interval, we get a random subinterval. Suppose we choose n random subintervals. What is the probability that there is one which has nonempty intersection with all of the others?
2. (Classic Problem:) The uptown and downtown trains come equally often to our station, but we wind up taking the uptown train 90% of the time. How is this possible?
(You can look up many versions of this.)

The Normal Distribution: “Bell Curve”

- This distribution is very important: in practice, many distributions are “approximately normal”.
- Also, the main theorem of this class, the “Central Limit Theorem”, says that if we add up a bunch of independent and identically distributed (IID) RVs, the result is approximately normal.

Definition

The normal distribution is defined by its PDF

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}},$$

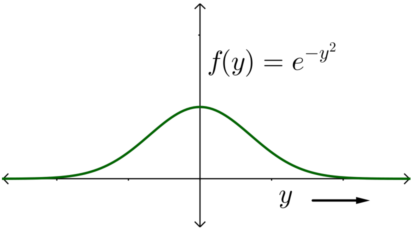
where $\sigma > 0$ and $-\infty < \mu < \infty$ are two parameters.

If Y is normal with parameters μ and σ (denoted $Y \sim \mathcal{N}(\mu, \sigma^2)$), then $E[Y] = \mu$ and $V[Y] = \sigma^2$.

Remarks on the function:

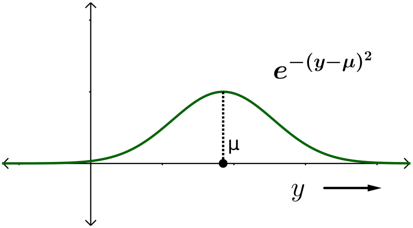
Start with e^{-y^2}.

This is a "Bell Curve".



Suppose we wanted to center it on μ:

We could change this to $e^{-(y-\mu)^2}$.



What is

$$\int_{-\infty}^{\infty} e^{-(y-\mu)^2} dy = \int_{-\infty}^{\infty} e^{-y^2} dy \ ?$$

This integral is usually done in Calculus 2 or 3.

Method:

Write

$$I = \int_{-\infty}^{\infty} e^{-y^2}\, dy.$$

Then

$$I^2 = \left(\int_{-\infty}^{\infty} e^{-x^2} dx\right)\left(\int_{-\infty}^{\infty} e^{-y^2} dy\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\, dx\, dy.$$

Convert to polar coordinates, and deduce that

$$I^2 = \pi \implies \int_{-\infty}^{\infty} e^{-y^2} dy = \sqrt{\pi}.$$

Our PDF needs to be

$$f(y) = \frac{1}{\sqrt{\pi}} e^{-y^2},$$

so that

$$\int_{-\infty}^{\infty} f(y) dy = 1.$$

Next, note that if

$$f(y) = \frac{1}{\sqrt{\pi}} e^{-y^2},$$

then the mean is 0:

$$E[Y] = \int_{-\infty}^{\infty} y \cdot \frac{1}{\sqrt{\pi}} e^{-y^2} dy = 0,$$

by symmetry – “odd function”.

Now normalize to have variance 1:

What is the variance of the RV Y defined by the PDF $f(y) = \frac{1}{\sqrt{\pi}} e^{-y^2}$?

$$V[Y] = E[Y^2] -$$

Next, note that if

$$f(y) = \frac{1}{\sqrt{\pi}} e^{-y^2},$$

then the mean is 0:

$$E[Y] = \int_{-\infty}^{\infty} y \cdot \frac{1}{\sqrt{\pi}} e^{-y^2} dy = 0,$$

by symmetry – "odd function".

Now normalize to have variance 1:

What is the variance of the RV Y defined by the PDF $f(y) = \frac{1}{\sqrt{\pi}} e^{-y^2}$?

$$V[Y] = E[Y^2] -$$

Next, note that if

$$f(y) = \frac{1}{\sqrt{\pi}} e^{-y^2},$$

then the mean is 0:

$$E[Y] = \int_{-\infty}^{\infty} y \cdot \frac{1}{\sqrt{\pi}} e^{-y^2} dy = 0,$$

by symmetry – "odd function".

Now normalize to have variance 1:

What is the variance of the RV Y defined by the PDF $f(y) = \frac{1}{\sqrt{\pi}} e^{-y^2}$?

$$V[Y] = E[Y^2] -$$

Next, note that if

$$f(y) = \frac{1}{\sqrt{\pi}} e^{-y^2},$$

then the mean is 0:

$$E[Y] = \int_{-\infty}^{\infty} y \cdot \frac{1}{\sqrt{\pi}} e^{-y^2} dy = 0,$$

by symmetry – "odd function".

Now normalize to have variance 1:

What is the variance of the RV Y defined by the PDF $f(y) = \dfrac{1}{\sqrt{\pi}} e^{-y^2}$?

$$V[Y] = E[Y^2] -$$

Next, note that if

$$f(y) = \frac{1}{\sqrt{\pi}} e^{-y^2},$$

then the mean is 0:

$$E[Y] = \int_{-\infty}^{\infty} y \cdot \frac{1}{\sqrt{\pi}} e^{-y^2} dy = 0,$$

by symmetry – "odd function".

Now normalize to have variance 1:

What is the variance of the RV Y defined by the PDF $f(y) = \dfrac{1}{\sqrt{\pi}} e^{-y^2}$?

$$V[Y] = E[Y^2] - E[Y]^2$$

Next, note that if

$$f(y) = \frac{1}{\sqrt{\pi}} e^{-y^2},$$

then the mean is 0:

$$E[Y] = \int_{-\infty}^{\infty} y \cdot \frac{1}{\sqrt{\pi}} e^{-y^2} dy = 0,$$

by symmetry – "odd function".

Now normalize to have variance 1:

What is the variance of the RV Y defined by the PDF $f(y) = \frac{1}{\sqrt{\pi}} e^{-y^2}$?

$$V[Y] = E[Y^2] - \cancelto{0}{E[Y]^2} = E[Y^2]$$

Next, note that if

$$f(y) = \frac{1}{\sqrt{\pi}} e^{-y^2},$$

then the mean is 0:

$$E[Y] = \int_{-\infty}^{\infty} y \cdot \frac{1}{\sqrt{\pi}} e^{-y^2} dy = 0,$$

by symmetry – "odd function".

Now normalize to have variance 1:

What is the variance of the RV Y defined by the PDF $f(y) = \frac{1}{\sqrt{\pi}} e^{-y^2}$?

$$V[Y] = E[Y^2] - \cancelto{0}{E[Y]^2} = E[Y^2] = \int_{-\infty}^{\infty} y^2 \cdot \frac{1}{\sqrt{\pi}} e^{-y^2} dy.$$

Next, note that if

$$f(y) = \frac{1}{\sqrt{\pi}} e^{-y^2},$$

then the mean is 0:

$$E[Y] = \int_{-\infty}^{\infty} y \cdot \frac{1}{\sqrt{\pi}} e^{-y^2} dy = 0,$$

by symmetry – "odd function".

Now normalize to have variance 1:

What is the variance of the RV Y defined by the PDF $f(y) = \frac{1}{\sqrt{\pi}} e^{-y^2}$?

$$V[Y] = E[Y^2] - \cancelto{0}{E[Y]^2} = E[Y^2] = \int_{-\infty}^{\infty} y^2 \cdot \frac{1}{\sqrt{\pi}} e^{-y^2} dy.$$

This is a simple exercise in integration by parts.

Next, note that if

$$f(y) = \frac{1}{\sqrt{\pi}} e^{-y^2},$$

then the mean is 0:

$$E[Y] = \int_{-\infty}^{\infty} y \cdot \frac{1}{\sqrt{\pi}} e^{-y^2} dy = 0,$$

by symmetry – "odd function".

Now normalize to have variance 1:

What is the variance of the RV Y defined by the PDF $f(y) = \dfrac{1}{\sqrt{\pi}} e^{-y^2}$?

$$V[Y] = E[Y^2] - \cancelto{0}{E[Y]^2} = E[Y^2] = \int_{-\infty}^{\infty} y^2 \cdot \frac{1}{\sqrt{\pi}} e^{-y^2} dy.$$

This is a simple exercise in integration by parts.

When we are done with all this "adjusting" and "normalizing", we get

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}.$$

If Y is a normal RV with mean μ and variance σ^2, then

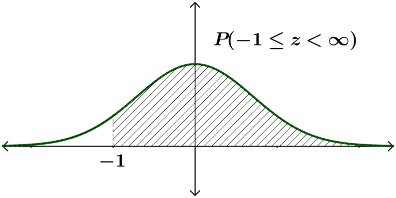
$$P(a \leq Y \leq b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}\, dy.$$

Unfortunately it is not possible to express this integral in closed form. So we use tables or library functions. (e.g. Table 4, Appendix 3, Text). The table gives, for a "standard normal RV" ($\mu = 0, \sigma = 1$),

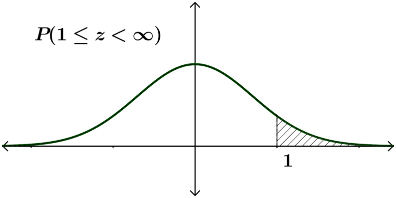
$$P(z \leq Y < \infty) = \int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy \qquad \text{for } Y \sim \mathcal{N}(0,1).$$

Two tricks:

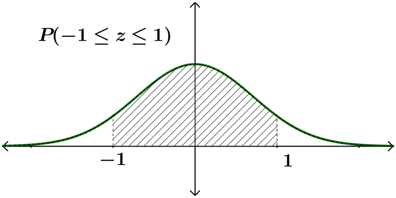
1. If Y is normal with mean μ and standard deviation σ (variance σ^2), then $\dfrac{Y-\mu}{\sigma}$ is <u>also normal</u> with mean 0 and variance 1.
2. Suppose, for a standard normal RV Z, we want $P(-1 \leq z \leq 1)$. This is the same as $P(-1 \leq z < \infty) - P(1 \leq z < \infty)$: (pictures follow:)


$P(-1 \le z < \infty)$
-1

$-$


$P(1 \le z < \infty)$
1

$=$


$P(-1 \le z \le 1)$
-1
1

Look up in the table: $P(1 \leq z < \infty) = 0.1587$.
What about $P(-1 \leq z < \infty)$?

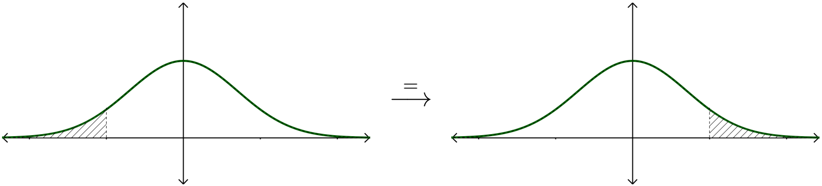
There are no negative numbers in Table 4, Appendix 3. But

$$P(-1 \leq z < \infty) + P(-\infty < z \leq -1) = P(-\infty < z < \infty) = 1.$$

So

$$P(-1 \leq z < \infty) = 1 - P(-\infty < z \leq -1) = 1 - P(1 < z < \infty),$$

by symmetry:



And we can look this up in the table:

$$P(-1 \leq z < \infty) = 1 - 0.1587 = 0.8413.$$

Thus

$$P(-1 \leq z \leq 1) = 0.8413 - 0.1587 \approx 68\%.$$

One more trick / method:

Suppose we have a normal RV Y with mean 3 and variance 4, and we are asked to find $P(1 \leq Y \leq 5)$.

Transform Y so that it is a standard normal RV.
Don't forget to transform the 1 and the 5!

$$1 \leq Y \leq 5 \iff \frac{1-3}{2} \leq \frac{Y-3}{2} \leq \frac{5-3}{2} \iff -1 \leq \frac{Y-3}{2} \leq 1.$$

Now $Z := \dfrac{Y-3}{2}$ is a standard normal RV, so that the probabilities can be found by table lookup and symmetry as before.
So

$$P(1 \leq Y \leq 5) = P(-1 \leq Z \leq 1).$$

(We already saw how to compute this.)

Remark:

That Z has $E[Z] = 0, V[Z] = 1$ is easy (linearity of E).
That Z is also normal is not trivial or obvious, but we will see it later using MGFs.

Example 4.9:

The achievement scores for a college entrance examination are normally distributed with mean 75 and standard deviation 10. What fraction of the scores lies between 80 and 90?

Interpretation:

Y is normal with $\mu = 75, \sigma = 10$. Find $P(80 \leq Y \leq 90)$.

Solution

We write

$$P(80 \leq Y \leq 90) = P\left(\frac{80-75}{10} \leq \frac{Y-75}{10} \leq \frac{90-75}{10}\right).$$

Noting that $Z = \dfrac{Y-75}{10}$ is standard normal, we need to find $P(0.5 \leq Z \leq 1.5)$.

This is

$$\int_{0.5}^{1.5} f_Z(y)dy = \int_{0.5}^{\infty} f_Z(y)dy - \int_{1.5}^{\infty} f_Z(y)dy.$$

where f_Z is the standard normal PDF.

Solution: (continued)

Both the integrals can be looked up in table.

Answer: 0.3085-0.0668 $= \boxed{0.2417}$.

Remarks:

- This is the most common application of probability theory. Because this is so basic, there are many ways to do and describe this computation.
- The numbers $0.5 = \dfrac{80-75}{10}$ and $1.5 = \dfrac{90-75}{10}$ are called z-scores: a "raw score" is converted to a "z-score", which is measures in standard deviations away from the mean.
- Z is a common notation for a standard normal RV.
- The table in the book is "complementary error function":

$$\mathrm{erfc}(z) = \int_z^\infty f_Z(y)dy.$$

Remarks: (continued)

- Sometimes you'll get a table of
$$\mathcal{N}(z) = \int_{-\infty}^{z} f_Z(y)dy \quad (= \text{erf}(z)).$$
Since $\mathcal{N}(z) + \text{erfc}(z) = 1$, it is equivalent to use either $\mathcal{N}(z)$ or $\text{erfc}(z)$ for calculations.
- Warning: Make sure you know what kind of table you have.

Exercise 4.73(a):

The width of bolts of fabric is normally distributed with mean 950 mm (millimeters) and standard deviation 10 mm. What is the probability that a randomly chosen bolt has a width of between 947 and 958 mm?

Interpretation:

Find $P(947 \leq Y \leq 958)$, where $Y \sim \mathcal{N}(950, 100)$.

Solution:

We have

$$
\begin{aligned}
P(947 \le Y \le 958) &= P\left(\frac{947-950}{10} \le \frac{Y-950}{10} \le \frac{958-950}{10}\right) \\
&= P(-0.3 \le Z \le 0.8) \quad \left(\text{where } Z = \frac{Y-950}{10}\right) \\
&= \operatorname{erfc}(-0.3) - \operatorname{erfc}(0.8) \\
&\text{(we cannot look up } \operatorname{erfc}(-0.3)!) \\
&= (1 - \operatorname{erfc}(0.3)) - \operatorname{erfc}(0.8) \\
&= (1 - 0.3821) - (0.2119) = \boxed{0.406}.
\end{aligned}
$$

Remark:

The book does not use the erfc notation.

Remark: The "95% Rule" aka the "$68 - 95 - 99.7\%$ Rule"

Common probabilities for a standard normal RV are often memorized. In particular:

$$\begin{aligned} P(-1 \leq Z \leq 1) &\approx 0.68, \\ P(-2 \leq Z \leq 2) &\approx 0.954, \\ P(-1.96 \leq Z \leq 1.96) &\approx 0.95, \\ P(-3 \leq Z \leq 3) &\approx 0.997. \end{aligned}$$

These probabilities $P(-n \leq Z \leq n)$ converge very rapidly to 1 as n grows.

Sometimes people say that the normal distribution has "thin tails". By this they mean that $P(Z \geq n) + P(Z \leq -n)$ is very small for values of n greater than, say 4.

In practical problems, it may be appropriate to use a different distribution, if we are interested in, say, $P(Z \geq 5)$.

Gamma Distribution

Observe that a normally distributed RV can take on any real value. We might be interested in a situation where we know that the RV is positive (or at least non-negative).

Example: Time-to-failure, or "first repair".

The Gamma distribution is a model for this.
There is a hump "near 0", and a "tail" going out to $+\infty$.

Definition (The Gamma Distribution)

Y has the Gamma distribution with parameters α and β if the PDF is

$$f_Y(y) = \begin{cases} \dfrac{y^{\alpha-1} e^{-\frac{y}{\beta}}}{\beta^\alpha \Gamma(\alpha)} & y \geq 0, \\ 0 & y < 0, \end{cases}$$

where

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$$

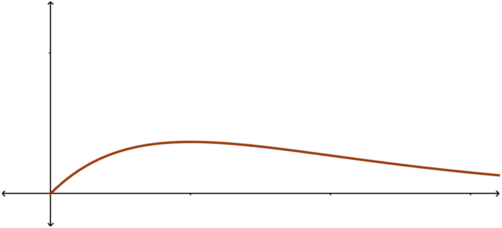
is the "generalized factorial function".

How to make sense of this?
Simplify to a special case: $\alpha = 2, \beta = 1$.
Then

$$f_Y(y) = \begin{cases} ye^{-y} & y \geq 0, \\ 0 & y < 0, \end{cases}$$

and we get this shape:

$\longrightarrow$



Note that the number $\beta^\alpha \Gamma(\alpha)$ is put in the PDF so that

$$\int_{-\infty}^{\infty} f_Y(y) dy = \int_0^{\infty} \frac{y^{\alpha-1} e^{-\frac{y}{\beta}}}{\beta^\alpha \Gamma(\alpha)} dy = 1.$$

This is true by the change-of-variable $v = \dfrac{y}{\beta}$ and the definition of Γ:

First

$$dv = \frac{dy}{\beta}, dy = \beta dv, y = \beta v.$$

Using the change-of-variable, we get

$$\int_0^\infty \frac{y^{\alpha-1}e^{-\frac{y}{\beta}}}{\beta^\alpha\Gamma(\alpha)}dy = \int_{v=0}^{v=\infty} \frac{(\beta v)^{\alpha-1}e^{-v}}{\beta^\alpha\Gamma(\alpha)}\beta dv$$

Using the change-of-variable, we get

$$\int_0^\infty \frac{y^{\alpha-1}e^{-\frac{y}{\beta}}}{\beta^\alpha \Gamma(\alpha)}dy = \int_{v=0}^{v=\infty} \frac{(\beta v)^{\alpha-1}e^{-v}}{\beta^\alpha \Gamma(\alpha)}\beta dv$$

$$= \int_0^\infty \frac{\beta^{\alpha-1}v^{\alpha-1}e^{-v}\beta}{\beta^\alpha \Gamma(\alpha)}dv$$

Using the change-of-variable, we get

$$\int_0^\infty \frac{y^{\alpha-1}e^{-\frac{y}{\beta}}}{\beta^\alpha\Gamma(\alpha)}dy = \int_{v=0}^{v=\infty} \frac{(\beta v)^{\alpha-1}e^{-v}}{\beta^\alpha\Gamma(\alpha)}\beta dv$$

$$= \int_0^\infty \frac{\cancel{\beta^{\alpha-1}}v^{\alpha-1}e^{-v}\cancel{\beta}}{\cancel{\beta^\alpha}\Gamma(\alpha)}dv$$

$$= \int_0^\infty \frac{v^{\alpha-1}e^{-v}}{\Gamma(\alpha)}dv$$

Using the change-of-variable, we get

$$\begin{aligned}
\int_0^\infty \frac{y^{\alpha-1}e^{-\frac{y}{\beta}}}{\beta^\alpha\Gamma(\alpha)}dy &= \int_{v=0}^{v=\infty} \frac{(\beta v)^{\alpha-1}e^{-v}}{\beta^\alpha\Gamma(\alpha)}\beta dv \\
&= \int_0^\infty \frac{\cancel{\beta^{\alpha-1}}v^{\alpha-1}e^{-v}\cancel{\beta}}{\cancel{\beta^\alpha}\Gamma(\alpha)}dv \\
&= \int_0^\infty \frac{v^{\alpha-1}e^{-v}}{\Gamma(\alpha)}dv \\
&= \frac{1}{\Gamma(\alpha)}\int_0^\infty v^{\alpha-1}e^{-v}dv \quad (\Gamma(\alpha) \text{ is a number})
\end{aligned}$$

Using the change-of-variable, we get

$$\begin{aligned}
\int_0^\infty \frac{y^{\alpha-1}e^{-\frac{y}{\beta}}}{\beta^\alpha\Gamma(\alpha)}dy &= \int_{v=0}^{v=\infty} \frac{(\beta v)^{\alpha-1}e^{-v}}{\beta^\alpha\Gamma(\alpha)}\beta dv \\
&= \int_0^\infty \frac{\cancel{\beta^{\alpha-1}}v^{\alpha-1}e^{-v}\cancel{\beta}}{\cancel{\beta^\alpha}\Gamma(\alpha)}dv \\
&= \int_0^\infty \frac{v^{\alpha-1}e^{-v}}{\Gamma(\alpha)}dv \\
&= \frac{1}{\Gamma(\alpha)}\underbrace{\int_0^\infty v^{\alpha-1}e^{-v}dv} \quad (\Gamma(\alpha) \text{ is a number}) \\
&= \frac{1}{\Gamma(\alpha)}\cdot\overbrace{\Gamma(\alpha)}
\end{aligned}$$

Using the change-of-variable, we get

$$\begin{aligned}
\int_0^\infty \frac{y^{\alpha-1}e^{-\frac{y}{\beta}}}{\beta^\alpha\Gamma(\alpha)}dy &= \int_{v=0}^{v=\infty} \frac{(\beta v)^{\alpha-1}e^{-v}}{\beta^\alpha\Gamma(\alpha)}\beta dv \\
&= \int_0^\infty \frac{\cancel{\beta^{\alpha-1}}v^{\alpha-1}e^{-v}\cancel{\beta}}{\cancel{\beta^\alpha}\Gamma(\alpha)}dv \\
&= \int_0^\infty \frac{v^{\alpha-1}e^{-v}}{\Gamma(\alpha)}dv \\
&= \frac{1}{\Gamma(\alpha)}\underbrace{\int_0^\infty v^{\alpha-1}e^{-v}dv} \quad (\Gamma(\alpha) \text{ is a number}) \\
&= \frac{1}{\Gamma(\alpha)} \cdot \overbrace{\Gamma(\alpha)} = 1.
\end{aligned}$$

You can expect a lot of computations like this:
e.g.

Theorem

If $Y \sim \Gamma(\alpha, \beta)$, then $E[Y] = \alpha\beta$ and $V[Y] = \alpha\beta^2$.

Proof is an exercise in integration similar to above, but more complicated.

Remarks:

- α and β are called the "shape" and the "scale" parameters, respectively.
- $\alpha > 0, \beta > 0$.
- If $Y \sim \Gamma(\alpha, \beta)$, then $Y \geq 0$.
- Other sources may work in terms of $\lambda = \dfrac{1}{\beta}$.

The definition of the Gamma PDF involves the Gamma function, a "generalized factorial":

$$\Gamma(\alpha) \stackrel{\text{def}}{=} \int_0^\infty y^{\alpha-1} e^{-y} dy.$$

Integration by parts proves the formula $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.

$$\Gamma(1) = \int_0^\infty y^{1-1} e^{-y} dy = \int_0^\infty e^{-y} dy = 1.$$

Using the two facts above, we find

$$\begin{aligned} \Gamma(2) = (2-1)\Gamma(2-1) &= 1 \cdot \Gamma(1) = 1, \\ \Gamma(3) = (3-1)\Gamma(3-1) &= 2 \cdot \Gamma(2) = 2 \quad = 2!, \\ \Gamma(4) = (4-1)\Gamma(4-1) &= 3 \cdot \Gamma(3) = 6 \quad = 3!. \end{aligned}$$

Induction shows that $\Gamma(n) = (n-1)!$ for all $n \in \mathbb{N}$.

Notice that $\Gamma(\alpha)$ is defines for real values of α, whereas the factorial is defined only for nonnegative integers.

This is the sense in which the Γ function is a generalized factorial.

Remark:

The values of $\Gamma(\alpha)$ when α is not an integer can be difficult to determine, e.g.

Exercises 4.194 and 4.196

4.194 If $u > 0$, then

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{uy^2}{2}} dy = \frac{1}{\sqrt{u}}.$$

4.196 Show that

$$\Gamma\left(\frac{1}{2}\right) = \int_0^{\infty} y^{-\frac{1}{2}} e^{-y} dy = \sqrt{\pi}$$

by making the transformation $y = \frac{x^2}{2}$ and employing Exercise 4.194.

The way we handle the Γ function in this course is to treat it as a black box:
The Gamma PDF uses the Γ function as a normalization.
We generally will not evaluate $\Gamma(\alpha)$ unless α is a nonnegative integer.

Theorem

If $Y \sim \Gamma(\alpha, \beta)$*, then* $E[Y] = \alpha\beta$ *and* $V[Y] = \alpha\beta^2$*.*

Sketch of Proof:

Use integration by parts and the standard formula $V[Y] = E[Y^2] - E[Y]^2$.
We find

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^{\infty} y \cdot \frac{y^{\alpha-1} e^{-\frac{y}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)} dy.$$

In evaluating such an integral, remember what is a function of y and what is a number:

$$E[Y] = \frac{1}{\Gamma(\alpha)} \int_{-\infty}^{\infty} \frac{y^{\alpha} e^{-\frac{y}{\beta}}}{\beta^{\alpha}} dy.$$

Substitute $v = \dfrac{y}{\beta}$, so that $y = \beta v$ and $dy = \beta dv$.

$$\begin{aligned}
\therefore \quad E[Y] &= \frac{1}{\Gamma(\alpha)} \int_0^\infty \frac{(\beta v)^\alpha e^{-v}}{\beta^\alpha} \beta dv \\
&= \frac{1}{\Gamma(\alpha)} \int_0^\infty v^\alpha e^{-v} \cdot \beta dv \\
&= \frac{1}{\Gamma(\alpha)} \cdot \beta \int_0^\infty v^{(\alpha+1)-1} e^{-v} dv \\
&= \frac{1}{\Gamma(\alpha)} \cdot \beta \cdot \Gamma(\alpha+1) \qquad \text{by definition of } \Gamma \text{ function.} \\
&= \frac{1}{\Gamma(\alpha)} \cdot \beta \cdot \alpha\Gamma(\alpha) \qquad \text{by the recursive formula for } \Gamma. \\
&= \boxed{\alpha\beta}.
\end{aligned}$$

Finding $V[Y]$ is similar, but much more complicated.

Here is a trick which is useful:

Recall that any PDF $f_Y(y)$ has

$$\int_{-\infty}^{\infty} f_Y(y)dy = 1 \iff P(-\infty < Y < \infty) = 1.$$

We checked that this was true for the Γ PDF:

$$\int_0^{\infty} \frac{y^{\alpha-1}e^{-\frac{y}{\beta}}}{\beta^\alpha \Gamma(\alpha)} dy = 1.$$

But β^α and $\Gamma(\alpha)$ are numbers. So

$$\int_0^{\infty} y^{\alpha-1}e^{-\frac{y}{\beta}} dy = \beta^\alpha \Gamma(\alpha).$$

So we have a scheme for evaluating any integral of this form.
For example,

$$\int_0^{\infty} y^3 e^{-\frac{y}{2}} dy = \int_0^{\infty} y^{\alpha-1}e^{-\frac{y}{\beta}} dy \qquad \begin{matrix} \alpha = 4 \\ \beta = 2 \end{matrix}$$

$$= \beta^\alpha \Gamma(\alpha) = 2^4 \Gamma(4) = 2^4 \cdot 3! = 16 \cdot 6 = \boxed{96}.$$

Special cases of the Gamma distribution:

The Exponential Distribution:

The exponential distribution is the case $\alpha = 1$, and is denoted $\text{Exp}(\beta)$.
If $Y \sim \text{Exp}(\beta)$, then $E[Y] = \beta$ and $V[Y] = \beta^2$.

The Chi-Squared (χ^2) Distribution with k degrees of freedom:

This is the Γ distribution with $\alpha = \frac{k}{2}, \beta = 2$.
If $Y \sim \chi^2[k]$, then $E[Y] = k$ and $V[Y] = 2k$.

Remark:

The reason we study this separately is in Theorem 7.2:
If $Z_1, \ldots, Z_k$ are independent standard normal RVs and
$Y = Z_1^2 + \cdots + Z_k^2$ (think of Y as a sum of squared errors), then
$Y \sim \chi^2[k]$.

Exercise 4.91 (a):

If Y has an exponential distribution and $P(Y > 2) = .0821$, what is $\beta = E[Y]$?

Solution:

$Y \sim \text{Exp}(\beta), P(Y > 2) = .0821.$
Now find β.

Notice that $P(Y > 2) = \int_2^\infty f_Y(y)dy$.

Plug in the PDF, evaluate the integral, and solve (expression in β) $= 0.0821$ for β.
By definition, the PDF is

$$f_Y(y) = \begin{cases} \dfrac{y^{1-1}e^{-\frac{y}{\beta}}}{\beta^1\Gamma(1)} & y \geq 0 \\ 0 & y < 0 \end{cases} = \begin{cases} \dfrac{1}{\beta}e^{-\frac{y}{\beta}} & y \geq 0 \\ 0 & y < 0 \end{cases}.$$

$$\therefore\ 0.0821 = \int_2^\infty \frac{1}{\beta}e^{-\frac{y}{\beta}}dy = -e^{-\frac{y}{\beta}}\Big|_2^\infty = -0 - (-e^{-\frac{2}{\beta}}) = e^{-\frac{2}{\beta}}.$$

So $-\frac{2}{\beta} = \ln(0.0821) \implies \beta = -\dfrac{2}{\ln(0.0821)} \approx \boxed{0.8}$.

Exercise 4.89 (a):

The operator of a pumping station has observed that demand for water during early afternoon hours has an approximately exponential distribution with mean 100 cfs (cubic feet per second). Find the probability that the demand will exceed 200 cfs during the early afternoon on a randomly selected day.

Solution:

$Y \sim \text{Exp}(\beta), E[Y] = 100.$
Since $E[Y] = \beta$, we have $\beta = 100$.
Now find $P(Y > 200)$.

$$\begin{aligned} P(Y > 200) &= \int_{200}^{\infty} f_Y(y)dy = \int_{200}^{\infty} \frac{1}{\beta} e^{-\frac{y}{\beta}} dy \\ &= -e^{-\frac{y}{\beta}} \Big|_{200}^{\infty} = -0 - (-e^{-\frac{200}{\beta}}) \\ &= e^{-\frac{200}{\beta}} = e^{-\frac{200}{100}} = \boxed{e^{-2}}. \end{aligned}$$

Recall:

The Gamma PDF uses the Γ function, a generalized factorial:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y}\, dy.$$

Integrate by parts:

$$u = y^{\alpha-1}, \quad dv = e^{-y} dy$$

$$du = (\alpha-1) y^{\alpha-2} dy, \qquad v = -e^{-y}.$$

$$\begin{aligned}
\therefore \int_0^\infty u\, dv = uv\Big|_0^\infty - \int_0^\infty v\, du &= y^{\alpha-1}(-e^{-y})\Big|_0^\infty - \int_0^\infty v du \\
&= 0 - \int_0^\infty (-e^{-y})(\alpha-1) y^{\alpha-2} dy \\
&= (\alpha-1) \int_0^\infty y^{(\alpha-1)-1} e^{-y} dy \\
&= (\alpha-1)\Gamma(\alpha-1).
\end{aligned}$$

Thus the recursion formula $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.
Also $\Gamma(n) = (n-1)!$ for positive integers n.

The key formula for many problems and exercises is

$$\int_0^\infty y^{\alpha-1} e^{-\frac{y}{\beta}}\,dy = \beta^\alpha \Gamma(\alpha). \qquad (\star)$$

$\Gamma(\alpha)$ can be evaluated by the recursion formula we just proved, if $\alpha \in \mathbb{N}$.

The formula $(\star)$ is equivalent to the statement that the Gamma PDF integrates to 1.

The above remarks will save you from having to integrate by parts several times.

All of this is based on integration by parts, it's just made easier with clever packaging.

Exercise 4.111 (a):

Suppose that Y has a Gamma distribution with parameters α and β. If a is any positive or negative value such that $\alpha + a > 0$, show that

$$E[Y^a] = \frac{\beta^a \Gamma(\alpha + a)}{\Gamma(\alpha)}.$$

Solution:

$$\begin{aligned} E[Y^a] &= \int_{-\infty}^{\infty} y^a f_Y(y) dy \qquad \text{where } f_Y \text{ is the Gamma PDF} \\ &= \int_0^{\infty} y^a \frac{y^{\alpha-1} e^{-\frac{y}{\beta}}}{\beta^\alpha \Gamma(\alpha)} dy = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^{\infty} y^{\alpha+a-1} e^{-\frac{y}{\beta}} \\ &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \beta^a \Gamma(\alpha + a) \qquad \text{by formula } (\star) \\ &= \boxed{\frac{\beta^a \Gamma(\alpha + a)}{\Gamma(\alpha)}} \qquad \text{by cancelling powers of } \beta. \end{aligned}$$

Definition (Beta Distribution)

A random variable Y is said to have the Beta distribution (denoted $Y \sim \text{Beta}(\alpha, \beta)$ if the PDF is

$$f_Y(y) = \begin{cases} \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)} & 0 \leq y \leq 1 \\ 0 & \text{elsewhere}, \end{cases}$$

where

$$B(\alpha, \beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1} dy = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

is the "Beta function".

Remark:

If $Y \sim \text{Beta}(\alpha, \beta)$, then $0 \leq Y \leq 1$.

Theorem

If $Y \sim \textit{Beta}(\alpha, \beta)$, *then*

$$E[Y] = \frac{\alpha}{\alpha + \beta}$$

Theorem

If $Y \sim Beta(\alpha, \beta)$*, then*

$$E[Y] = \frac{\alpha}{\alpha+\beta} \quad \textit{and} \quad V[Y] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Theorem

If $Y \sim Beta(\alpha, \beta)$*, then*

$$E[Y] = \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad V[Y] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Proof:

Use the formulas for the Beta-integral and the Γ function. We'll look at $E[Y]$; $V[Y]$ is left as an exercise.

Theorem

If $Y \sim Beta(\alpha, \beta)$, then

$$E[Y] = \frac{\alpha}{\alpha+\beta} \quad \textit{and} \quad V[Y] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Proof:

Use the formulas for the Beta-integral and the Γ function. We'll look at $E[Y]$; $V[Y]$ is left as an exercise.

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy$$

Theorem

If $Y \sim Beta(\alpha, \beta)$, then

$$E[Y] = \frac{\alpha}{\alpha+\beta} \quad \textit{and} \quad V[Y] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Proof:

Use the formulas for the Beta-integral and the Γ function. We'll look at $E[Y]$; $V[Y]$ is left as an exercise.

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^1 y \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)} dy$$

Theorem

If $Y \sim Beta(\alpha, \beta)$, then

$$E[Y] = \frac{\alpha}{\alpha+\beta} \quad \textit{and} \quad V[Y] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Proof:

Use the formulas for the Beta-integral and the Γ function. We'll look at $E[Y]$; $V[Y]$ is left as an exercise.

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^1 y \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)} dy$$

$$= \frac{1}{B(\alpha,\beta)} \int_0^1 y^{(\alpha+1)-1}(1-y)^{\beta-1}$$

Theorem

If $Y \sim Beta(\alpha, \beta)$, then

$$E[Y] = \frac{\alpha}{\alpha+\beta} \quad \textit{and} \quad V[Y] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Proof:

Use the formulas for the Beta-integral and the Γ function. We'll look at $E[Y]$; $V[Y]$ is left as an exercise.

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^1 y \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)} dy$$

$$= \frac{1}{B(\alpha,\beta)} \int_0^1 y^{(\alpha+1)-1}(1-y)^{\beta-1} \quad \text{(Now use B-integral formula.)}$$

Theorem

If $Y \sim Beta(\alpha, \beta)$, then

$$E[Y] = \frac{\alpha}{\alpha+\beta} \quad \textit{and} \quad V[Y] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Proof:

Use the formulas for the Beta-integral and the Γ function. We'll look at $E[Y]$; $V[Y]$ is left as an exercise.

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^1 y \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)} dy$$

$$= \frac{1}{B(\alpha,\beta)} \int_0^1 y^{(\alpha+1)-1}(1-y)^{\beta-1} \quad \text{(Now use B-integral formula.)}$$

$$= \frac{1}{B(\alpha,\beta)} \cdot B(\alpha+1,\beta)$$

Theorem

If $Y \sim Beta(\alpha, \beta)$, then

$$E[Y] = \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad V[Y] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Proof:

Use the formulas for the Beta-integral and the Γ function. We'll look at $E[Y]$; $V[Y]$ is left as an exercise.

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y)dy = \int_0^1 y \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)} dy$$

$$= \frac{1}{B(\alpha,\beta)} \int_0^1 y^{(\alpha+1)-1}(1-y)^{\beta-1} \quad \text{(Now use B-integral formula.)}$$

$$= \frac{1}{B(\alpha,\beta)} \cdot B(\alpha+1,\beta) = \frac{1}{B(\alpha,\beta)} \cdot \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}$$

Theorem

If $Y \sim Beta(\alpha, \beta)$, then

$$E[Y] = \frac{\alpha}{\alpha+\beta} \quad \textit{and} \quad V[Y] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Proof:

Use the formulas for the Beta-integral and the Γ function. We'll look at $E[Y]$; $V[Y]$ is left as an exercise.

$$\begin{aligned}
E[Y] &= \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^1 y \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)} dy \\
&= \frac{1}{B(\alpha,\beta)} \int_0^1 y^{(\alpha+1)-1}(1-y)^{\beta-1} \quad \text{(Now use B-integral formula.)} \\
&= \frac{1}{B(\alpha,\beta)} \cdot B(\alpha+1,\beta) = \frac{1}{B(\alpha,\beta)} \cdot \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} \cdot \frac{\alpha\Gamma(\alpha)\cdot\Gamma(\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)}
\end{aligned}$$

Theorem

If $Y \sim Beta(\alpha, \beta)$, then

$$E[Y] = \frac{\alpha}{\alpha+\beta} \quad \textit{and} \quad V[Y] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Proof:

Use the formulas for the Beta-integral and the Γ function. We'll look at $E[Y]$; $V[Y]$ is left as an exercise.

$$\begin{aligned}
E[Y] &= \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^1 y \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)} dy \\
&= \frac{1}{B(\alpha,\beta)} \int_0^1 y^{(\alpha+1)-1}(1-y)^{\beta-1} \quad \text{(Now use B-integral formula.)} \\
&= \frac{1}{B(\alpha,\beta)} \cdot B(\alpha+1,\beta) = \frac{1}{B(\alpha,\beta)} \cdot \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \\
&= \frac{\cancel{\Gamma(\alpha+\beta)}}{\cancel{\Gamma(\alpha)} \cdot \Gamma(\beta)} \cdot \frac{\alpha\cancel{\Gamma(\alpha)} \cdot \cancel{\Gamma(\beta)}}{(\alpha+\beta)\cancel{\Gamma(\alpha+\beta)}} = \boxed{\frac{\alpha}{\alpha+\beta}}.
\end{aligned}$$

□

Remark and Terminology:

- The CDF for the Beta distribution is called the "incomplete Beta function".
- If $0 \leq y \leq 1$, we can write this as

$$F(y) = \int_0^y \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)} dy.$$

- If α and β are both integers, and we write $n = \alpha + \beta - 1$, then

$$F(y) = \sum_{i=0}^{n} \binom{n}{i} y^i (1-y)^{n-i}.$$

  This, like so much else, is proved by integration by parts.

Exercise 4.125:

The percentage of impurities per batch in a chemical product is a random variable Y with density function

$$f_Y(y) = \begin{cases} 12y^2(1-y) & 0 \leq y \leq 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Find the mean and variance of the percentage of impurities in a randomly selected batch of the chemical.

Solution:

Note that the given PDF is the Beta distribution with parameters $\alpha = 3, \beta = 2$.
Therefore the mean is

$$\frac{\alpha}{\alpha+\beta} = \frac{3}{3+2} = \boxed{\frac{3}{5}},$$

and the variance is

$$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{3 \cdot 2}{(3+2)^2(3+2+1)} = \frac{6}{5^2 \cdot 6} = \boxed{\frac{1}{25}}.$$

So far we have studied the following continuous distribtions:

- Uniform,
- Normal,
- Gamma,
  and
- Beta.

Note that the Gamma distribution includes the special cases: The χ^2 distribution and the Exponential distribution.

Now we move on to cover some things that are relevant to all distributions:

- Moment Generating Functions,
  and
- Tchebysheff's Theorem.

Definition (Moment Generating Functions)

The moment generating function (MGF) of a RV Y is $E[e^{tY}] = m_Y(t)$ or $m(t)$.

The theoretical importance of this is in part that the MGF determines the RV;
that is, if we want to check that a RV is (say) normal with mean a and variance b, it is enough to check that this RV has the right MGF.

Why is the MGF called that?

It "generates" the moments $\mu'_k = E[Y^k]$, in the sense that

$$\left(\frac{d}{dt}\right)^k [m(t)]\bigg|_{t=0} = \mu'_k.$$

Remark:

Sometimes people think in terms of "central moments"
$\mu_k = E[(Y-\mu)^k]$, but we can derive these from the μ'_k (and vice versa).

Remark:

It is not necessary that the integral defined by the expectation $E[e^{tY}]$ converges for all values of t; it is enough that it converges for some t.

What will we use this for?

(1) A linear function of a normal RV is normal. $(\star)$

In particular, if $Y \sim \mathcal{N}(\mu, \sigma^2)$, then $Z = \dfrac{Y - \mu}{\sigma}$ is a standard normal RV.

We know this because Z has the right MGF.

Note that $(\star)$ would NOT be true if we replaced normal by Beta, Gamma, Exponential, etc.

(2) Later we will see using similar ideas that a linear combination of independent normal RVs is normal.

Again, this is special to the normal distribution.

How do we compute an MGF?

Of course we can look it up in a table.
But, if you are asked to derive the result, the answer is that you do an exercise in integration.

Example

$Y \sim \Gamma(\alpha, \beta)$. Find the MGF of Y.

Solution:

$$m_Y(t) = E[e^{tY}] = \int_{-\infty}^{\infty} e^{ty} \underbrace{f_Y(y)}_{\text{The } \Gamma \text{ PDF}} dy$$

$$= \int_{0}^{\infty} e^{ty} \frac{y^{\alpha-1} e^{-\frac{y}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)} dy \quad (\text{as } \Gamma \text{ PDF is zero for } y < 0).$$

Now rearrange this to get an integral we can evaluate using the standard formula for Γ integrals.

Solution: (continued)

Now

$$\int_0^\infty y^{\alpha-1} e^{-\frac{y}{\beta}}\, dy = \beta^\alpha \Gamma(\alpha). \quad \text{(Integral of } \Gamma \text{ PDF must be 1.)}$$

So

$$m(t) = \int_0^\infty \frac{1}{\beta^\alpha \Gamma(\alpha)} \cdot y^{\alpha-1} e^{-\frac{y}{X}}\, dy.$$

What should go in place of X?

$$ty - \frac{y}{\beta} = -\frac{y}{X} \Rightarrow t - \frac{1}{\beta} = -\frac{1}{X} \Rightarrow X = -\frac{1}{t - \frac{1}{\beta}} = \frac{\beta}{1 - \beta t}.$$

$$\begin{aligned}
\therefore\ m(t) &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-\frac{y}{\beta/(1-\beta t)}}\, dy \\
&= \frac{1}{\beta^\alpha \Gamma(\alpha)} \left(\frac{\beta}{1-\beta t}\right)^\alpha \Gamma(\alpha) = \boxed{\frac{1}{(1-\beta t)^\alpha}}.
\end{aligned}$$

Further trickery in integration will show that for a normal RV with mean 0 and variance σ^2 that $m(t) = e^{\left(\frac{\sigma^2 t^2}{2}\right)}$.
From this, we can derive the MGF of any normal RV.

Suppose Y is a RV with MGF $m_Y(t)$ and $X = aY + b$.
What is $m_X(t)$?

$$\begin{aligned} m_X(t) &= E[e^{tX}] &&= E[e^{t(aY+b)}] \\ &= E[e^{atY}e^{bt}] &&= e^{bt}E[e^{(at)Y}] \\ &= e^{bt}m_Y(t). && (\star) \end{aligned}$$

So if Y is normal with mean 0 and variance σ^2, and $X = Y + \mu$, then

$$m_X(t) = e^{\mu t}m_Y(t) = e^{\mu t}e^{\left(\frac{\sigma^2 t^2}{2}\right)} = e^{\left(\mu t + \frac{\sigma^2 t^2}{2}\right)}.$$

Exercise in "trickery": (Example 4.16)

Find the MGF for $g(Y) = Y - \mu$, where $Y \sim \mathcal{N}(\mu, \sigma^2)$.

Exercise 4.139

If $Y \sim \mathcal{N}(\mu, \sigma^2)$ and $X = -3Y + 4$, find $m_X(t)$.

Solution:

Using $(\star)$ from the previous slide, we find

$$\begin{aligned} m_X(t) = e^{4t} m_Y(-3t) &= e^{4t} e^{\left(\mu(-3t) + \frac{\sigma^2(-3t)^2}{2}\right)} \\ &= e^{\left(4t + (-3\mu)t + \frac{\sigma^2(9t^2)}{2}\right)} \\ &= e^{\left((-3\mu+4)t + \frac{(3\sigma)^2 t^2}{2}\right)}. \end{aligned}$$

What is the distribution of X?

X is normal with mean $-3\mu + 4$ and variance $(3\sigma)^2$ (or standard deviation 3σ),
because X has the same MGF as a normal RV mean $-3\mu + 4$ and standard deviation 3σ.

Theorem (Tchebysheff's Theorem)

If Y is any RV with mean μ and standard deviation σ, then

$$P(|Y-\mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Equivalently,

$$P(|Y-\mu| < k\sigma) \geq 1-\frac{1}{k^2}.$$

Note that "with mean μ and standard deviation σ" is part of the hypothesis; not every RV has a mean and a standard deviation. (Recall the St. Petersburg Paradox.)

Proof:

Start with the equation for σ^2, and estimate to get an inequality:

$$\sigma^2 = E[(Y-\mu)^2] = \int_{-\infty}^{\infty} (y-\mu)^2 f_Y(y)dy$$
$$= \int_{y \le \mu - k\sigma} (y-\mu)^2 f_Y(y)dy + \int_{|y-\mu|<k\sigma} (y-\mu)^2 f_Y(y)dy$$
$$+ \int_{y \ge \mu - k\sigma} (y-\mu)^2 f_Y(y)dy.$$

Note that all 3 parts are nonnegative.
In particular, the middle one is ≥ 0.
Also, in the first and the third part, $(y-\mu)^2 \ge k^2\sigma^2$ (Check this!).
So

$$\sigma^2 \ge \int_{y \le \mu - k\sigma} k^2\sigma^2 f_Y(y)dy + 0 + \int_{y \ge \mu - k\sigma} k^2\sigma^2 f_Y(y)dy$$
$$\ge k^2\sigma^2 \left(\int_{y \le \mu - k\sigma} f_Y(y)dy + \int_{y \ge \mu - k\sigma} f_Y(y)dy \right)$$
$$\ge k^2\sigma^2 \cdot P(|Y-\mu| \ge k\sigma) \implies P(|Y-\mu| \ge k\sigma) \le \frac{1}{k^2}.$$

Remark:

Similar proof works for discrete RVs.

When would you use Tchebysheff's theorem?
Mainly when you don't know the distribution of the RV being studied.

Exercise 4.147:

A machine used to fill cereal boxes dispenses, on the average, μ ounces per box. The manufacturer wants the actual ounces dispensed Y to be within 1 ounce of μ at least 75% of the time. What is the largest value of σ, the standard deviation of Y, that can be tolerated if the manufacturer's objectives are to be met?

Solution:

We want $P(|Y-\mu| \le 1) \ge 0.75$.
Tchebysheff tells us that

$$P(|Y-\mu| < k\sigma) \ge 1 - \frac{1}{k^2}.$$

From the context, the RV Y is continuous, so we won't worry about the difference between $\le$ and $<$. (Why?)

To use Tchebysheff's Theorem, we need $1 - \frac{1}{k^2} = 0.75$.
So we can solve for k and then solve for σ:

we find $k = 2$ and $\sigma = \frac{1}{2}$.

Remark:

- There is a lot of stuff buried in the exercises. For example, there are relationships between the RVs we have studied. (Poisson-Gamma relationship, and others.)
- Hazard rate functions are frequently used in practice.
- Moments of the normal distribution:

  The standard normal RV Z has mean 0 and variance 1.
  But then what about $E[Z^3]$ or $E[Z^4]$ or $E[Z^5]$?

  In fact we can show using integration by parts, Gamma integrals, or thinking about $\chi^2(1)$, that $E[Z^3] = 0, E[Z^4] = 3,$ and $E[Z^5] = 0$. (Exercise 4.199 in the book explores further on this.)
- If you want to go on, consider doing some of these exercises.

Problem: A model for the height of adult American men.

Our model will be, that the distribution is normal with mean 5'9" and standard deviation 3".

(a) According to the model, how many American men have height between 5'9" and 6'0"?

Solution:

Notice that we need to know the number of adult American men.
(This is typical of real-world problems: you need some extra information which is not given to you.)
We look this up, or we estimate from the population of the US.
Say the US population is 320 million, of which 100 million are children, and half of the rest are adult men. This gives 110 million adult men.

According to the model, heights between 5'9" and 6'0' have z-scores between 0 and 1.
According to the $68\% - 95\% - 99.7\%$ rule, and symmetry, this gives $34\% \times 110$ million ≈ 37.4 million.

(This could be improved with better data from the Census Bureau.)

Problem: A model for the height of adult American men. (continued)

(b) According to the model, how many adult American men have heights greater than 7'3"?

Solution:

First note that

$$7'3'' = 87'', 5'9'' = 69'' \implies \frac{87'' - 69''}{3''} = \frac{18''}{3''} = 6.$$

So this is 6 standard deviations away from the mean.

What fraction of the adult male population is 6 standard deviations away from the mean of the model?

Our table does not go this far.

We cannot do the integral $\displaystyle\int_6^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2} dx$.

We could use numerical approximation or look up using R or a more extensive table.

Remark:

Here again, the problem requires that we find more information.
We have to know which sources of information are good.
Let's say, after looking up, we find that the answer is roughly 10^{-9}.
Then our model estimates that the number of such men is
$10^{-9} \times 110$ million ≈ 0.11.
The number of men must be a nonnegative integer, so the model prediction is practically zero.

Problem: (continued)

(c) According to your own method of estimation, how many adult American men actually are taller than 7'3"?

Solution:

We have to look for data.
Important point we found very quickly is that the answer was NOT zero.
We think something in the order of dozens is right, but need to do more research.
Question: Who counts as American? (Green Card? Only Citizens? etc.)

Remark: "The map is not the territory."

The model is not reality. For some questions, the model will yield a reasonable answer, and for others, it won't.

As applied mathematicians, we are responsible for building and understanding models. But we are also responsible for understanding the difference between the model and reality.

Sometimes the model isn't appropriate for answering the question being asked – question (b) is one of those times.

End of Chapter 4

Chapter 5

Multivariate Probability Distributions

“Multivariate Probability Distributions”

If we have 2 (or n) random variables Y_1, Y_2 $(, \ldots, Y_n)$, they may have some relationship to one another, other than “independence”.
This relationship is specified by the “Joint Distribution” of Y_1 and Y_2.

In the discrete case, there is a joint proability function

$$p(y_1, y_2) = P(Y_1 = y_1 \text{ and } Y_2 = y_2).$$

Y_1 and Y_2 are independent if and only if

$$p(y_1, y_2) = P(Y_1 = y_1) \cdot P(Y_2 = y_2).$$

This joint probability function satisfies many of the same (or similar) axioms as an ordinary probability function.
In particular,

1. $p(y_1, y_2) \geq 0$.
2. $$\sum_{y_1} \sum_{y_2} p(y_1, y_2) = 1.$$

Example 5.1:

A local supermarket has three checkout counters. Two customers arrive at the counters at different times when the counters are serving no other customers. Each customer chooses a counter at random, independently of the other. Let Y_1 denote the number of customers who choose counter 1 and Y_2, the number who select counter 2. Find the joint probability function of Y_1 and Y_2.

Observe that $Y_1 = 0, 1,$ or 2.
Also $Y_2 = 0, 1,$ or 2.
So the "joint probability function" $p(y_1, y_2)$ can be written down in a 3×3 table:

y_2 \ y_1	0	1	2
0	$1/9$	$2/9$	$1/9$
1	$2/9$	$2/9$	0
2	$1/9$	0	0

Note that $y_1 = 0$ and $y_2 = 0$ iff both customers select counter 3. So $p(0,0) = {}^1/_3 \times {}^1/_3 = {}^1/_9$.

Notice that $y_1 = 1$ and $y_2 = 0$ can happen in 2 ways:

Cust #1	selects counter #1	and cust #2	selects counter #3.
or: #1	selects counter #1	and cust #1	selects counter #3.

Question: Are Y_1 and Y_2 independent RVs?

Answer: No. If Y_1 and Y_2 were independent, $p(y_1, y_2)$ would be $P(Y_1 = y_1) \cdot P(Y_2 = y_2)$.
So if we had a zero in the table, there would be a whole row or column of zeroes.

What is the probability $P(Y_1 = 0)$?

Look at this in 2 ways:

(1) $Y_1 = 0$ means 1st customer went to counter 2 or 3, and 2nd went to counter 2 or 3.

The decisions of the customers are independent, so the probability is ${}^2/_3 \times {}^2/_3 = {}^4/_9$.

(2) Look at the table:

$$P(Y_1 = 0) = p(0,0) + p(0,1) + p(0,2) = \sum_{y_2} p(0, y_2),$$

that is, consider $Y_1 = 0$ and all possibilities for Y_2, giving $1/9 + 2/9 + 1/9 = 4/9$.

From the joint probability function, we can derive the individual probability functions ("Marginal Probability Functions"):

$$p_1(y_1) = \sum_{\text{all } y_2} p(y_1, y_2).$$

In our example, $p_1(y_1)$ is given by the table:

y_1	0	1	2
$p_1(y_1)$	$4/9$	$4/9$	$1/9$

So the terminology "marginal distribution" comes from the fact that this is the distribution on the margin of the original table.

y_2 \ y_1	0	1	2	$p_2(y_2)\ \downarrow$
0	$1/9$	$2/9$	$1/9$	$4/9$
1	$2/9$	$2/9$	0	$4/9$
2	$1/9$	0	0	$1/9$
$p_1(y_1)\ \rightarrow$	$4/9$	$4/9$	$1/9$	

The "marginal distribution of Y_1" is written $p_1(y_1)$ in the text.
The condition for independence is $p(y_1, y_2) = p_1(y_1) \cdot p_2(y_2)$.

All of this is also treated for continuous RVs.
Recall that when working with continuous RVs, we introduced the "cumulative distribution function" F_Y: $F_Y(y) = P(Y \leq y)$.
When studying more than 1 RV, we have a "joint distribution function" $F(y_1, y_2) = P(Y_1 \leq y_1 \text{ and} Y_2 \leq y_2)$.

There is a "joint density function" $f(y_1, y_2)$, and the relationship is similar to the one-variable case:

$$F(y_1, y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(t_1, t_2)\, dt_2\, dt_1.$$

Everything that we did before will be repeated in the context of 2 random variables.

Marginal Probability Functions: Discrete Case

In the discrete case, the joint probability function is

$$p(y_1, y_2) = P(Y_1 = y_1 \text{ and } Y_2 = y_2),$$

and the marginal probability functions are

$$p_1(y_1) = \sum_{y_2} p(y_1, y_2), \quad p_2(y_2) = \sum_{y_1} p(y_1, y_2).$$

Remark:

The marginal probability functions are the single-RV probability functions obtained by "ignoring" the other variable.

The conditional probability functions are $p(y_1 \mid y_2)$ and $p(y_2 \mid y_1)$:

$$p(y_1 \mid y_2) = \frac{p(y_1, y_2)}{p_2(y_2)} = \frac{P(Y_1 = y_1 \text{ and } Y_2 = y_2)}{P(Y_2 = y_2)}; \quad (\star)$$

and similarly for $p(y_2 \mid y_1)$.

Remarks:

- This corresponds to
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)},$$
because the comma in $(\star)$ means “and”, and “and” corresponds to intersection.
- This notation is not very good:
Note that $p(y_1 \mid y_2)$ and $p(y_2 \mid y_1)$ are different functions.
In other words, $P(A \mid B) \neq P(B \mid A)$.

  So, in this notation of the text, if we write $P(5 \mid 3)$, what does this mean?
Does it mean $P(Y_1 = 5 \text{ and } Y_2 = 3)$ or $P(Y_2 = 5 \text{ and } Y_1 = 3)$?
We have to trust that no confusion will arise.
- Conditional probability $p(y_1 \mid y_2)$ is only defined if $p_2(y_2) > 0$.

Marginal Probability Functions: Continuous Case

All of these concepts exist for continuous RVs.
In addition, continuous RVs are often defined by a "Joint Distribution Function"

$$F(y_1, y_2) = P(Y_1 \le y_1, Y_2 \le y_2).$$

The (joint) probability function for the discrete RVs corresponds to the joint density function for two continuous RVs:

$$F(y_1, y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(t_1, t_2)\, dt_2\, dt_1.$$

The marginal density functions are defined by replacing the sums in the discrete case with integrals, e.g.

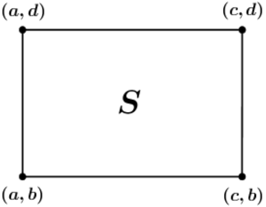
$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_2.$$

Also the conditional density function $f(y_1 \mid y_2)$ is

$$\frac{f(y_1, y_2)}{f_2(y_2)} \qquad \text{analogous to} \qquad \frac{p(y_1, y_2)}{p_2(y_2)}.$$

Recall that $F(y_1, y_2) = P(Y_1 \le y_1, Y_2 \le y_2)$.

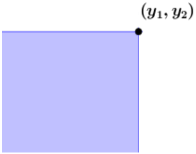
Suppose we have a rectangle in the (y_1, y_2)-plane.



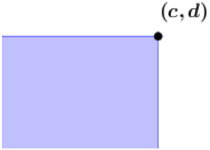
What is $P((Y_1, Y_2) \in S)$

$= P(a \le Y_1 \le c, b \le Y_2 \le d)$ in terms of F?

Notice that $F(y_1, y_2)$ gives the probability of a set like this:

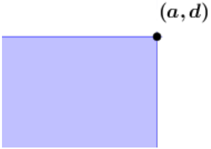


$X =$

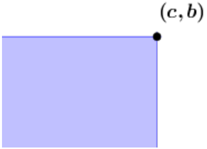


$$\implies \quad P((Y_1, Y_2) \in X) = F(c, d).$$

$A =$



$B =$



$$S = X \setminus (A \cup B).$$

$$\therefore P(S) = P(X) - P(A \cup B) = P(X) - P(A) - P(B) + P(A \cap B)$$
$$= \boxed{F(c,d) - F(a,d) - F(c,b) + F(a,b)}. \ (\star)$$

Remark:

In all of this, we take $P(Y_1 = a)$ (for example) to be 0.
Technically $(\star)$ above is $P(a < Y_1 \leq c, b < Y_2 \leq d)$.

Consequence:

Any joint distribution function (JDF) F must satisfy

$$F(c,d) - F(a,d) - F(c,b) + F(a,b) \geq 0$$

whenever $d \geq b$ and $c \geq a$, because $P((Y_1, Y_2) \in S) \geq 0$.

Note that this JDF has other properties which are analogous to the properties of a distribution function for a single RV, e.g.

$$\lim_{y_1 \to \infty} \lim_{y_2 \to \infty} F(y_1, y_2) = 1. \qquad \text{(see p.228 in the text.)}$$

Game plan for problems in this chapter:

1. Translate paragraph into $P(Y_1, Y_2) = X$.
2. Set up a multiple integral.
3. Do multiple integral.

Example

Define the joint distribution of two RVs Y_1, Y_2 by taking them to be the coordinates of a point chosen at random from the unit square $[0,1] \times [0,1]$. Find $P(0.1 \leq Y_1 \leq 0.3, 0 \leq Y_2 \leq 0.5)$.

Solution:

We must find the joint density function $f(y_1, y_2)$.
"Chosen at random" means "uniform distribution", which in turn implies that "The density function is constant in some region and is 0 elsewhere".

Here the region is the unit square $[0,1] \times [0,1]$.

Solution: (continued)

The density function is required to satisfy

$$\text{Total Probability } = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1\, dy_2 = 1.$$

If this f is zero outside $[0,1] \times [0,1]$ and $f = c$ inside it, then

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1\, dy_2 &= \int_0^1\int_0^1 c\, dy_1\, dy_2 = \int_0^1 cy_1\Big|_0^1 dy_2 \\
&= \int_0^1 (c \cdot 1 - c \cdot 0) dy_2 = \int_0^1 c\, dy_2 \\
&= cy_2\Big|_0^1 = c \cdot 1 - c \cdot 0 = c.
\end{aligned}$$

So $c = 1$, because the density function must integrate to 1.

Or, more simply,

$$\int_0^1\int_0^1 c\, dy_1\, dy_2 = \text{area of unit square} \cdot c = c.$$

Remark:

Later you will just be able to write down the density function almost immediately, in cases like this.

Solution: (continued)

We're supposed to find $P(0.1 \leq Y_1 \leq 0.3, 0 \leq Y_2 \leq 0.5)$.
This is

$$\int_{0.1}^{0.3} \int_{0}^{0.5} f(y_1, y_2)\, dy_1\, dy_2.$$

But $f \equiv 1$ here.
So the probability is

$$\int_{0.1}^{0.3} \int_{0}^{0.5} f(y_1, y_2)\, dy_1\, dy_2 = (0.3 - 0.1) \times (0.5 - 0) = 0.2 \times 0.5 = \boxed{0.1}.$$

Remark:

The double integrals and setup can get more complicated; not every region is a rectangle.

Also, the questions can take a different form:
In the example we saw just now, what is $F(0.2, 0.4)$?

$$F(0.2, 0.4) = P(Y_1 \leq 0.2, Y_2 \leq 0.4) = \int_{-\infty}^{0.2} \int_{-\infty}^{0.4} f(y_1, y_2)\, dy_2\, dy_1.$$

Since $f \equiv 1$ in the unit square and 0 elsewhere, this is

$$F(0.2, 0.4) = \int_0^{0.2} \int_0^{0.4} 1\, dy_2\, dy_1 = \boxed{0.08}.$$

Remark:

In the definition of density functions, do not forget about the "$f = 0$ elsewhere" clause:

$$\int_{-\infty}^{0.2} 1\, dy_1$$

does not converge.

Example 5.4:

Gasoline is to be stocked in a bulk tank once at the beginning of each week and then sold to individual customers. Let Y_1 denote the proportion of the capacity of the bulk tank that is available after the tank is stocked at the beginning of the week. Because of the limited supplies, Y_1 varies from week to week. Let Y_2 denote the proportion of the capacity of the bulk tank that is sold during the week. Because Y_1 and Y_2 are both proportions, both variables take on values between 0 and 1. Further, the amount sold, y_2, cannot exceed the amount available, y_1. Suppose that the joint density function for Y_1 and Y_2 is given by

$$f(y_1, y_2) = \begin{cases} 3y_1 & 0 \leq y_2 \leq y_1 \leq 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Find the probability that less than one-half of the tank will be stocked and more than one-quarter of the tank will be sold.

We convert the above paragraph into some simpler-looking math:

Interpretation:

Suppose Y_1 and Y_2 have the joint density function

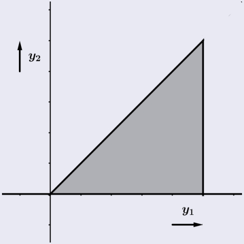
$$f(y_1, y_2) = \begin{cases} 3y_1 & 0 \le y_2 \le y_1 \le 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Find $P(0 \le Y_1 \le 0.5, Y_2 > 0.25)$.

Solution:

Start by graphing the region of integration.

Where is $f(y_1, y_2)$ nonzero?

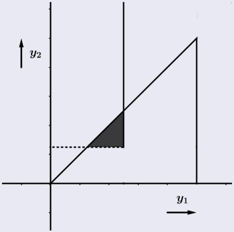


Where is the region of integration?

Solution: (continued)

But f is mostly zero in this region.

So really the region for integration is this:

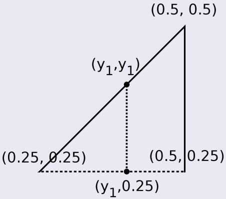


$$P(0 \leq Y_1 \leq 0.5, Y_2 > 0.25) = \int_0^{0.5} \int_{0.25}^{\infty} f(y_1, y_2)\, dy_2\, dy_1.$$

But, since $f = 0$ outside the small triangle, this is

$$\int_0^{0.5} \int_{0.25}^{y_1} (3y_1)\, dy_2\, dy_1.$$

Close-up of the small triangle:



Solution: (continued)

Suppressing the work, the integral comes out as $\frac{5}{128}$,
So the answer is

$$P(0 \leq Y_1 \leq 0.5, Y_2 > 0.25) = \boxed{\frac{5}{128}}.$$

For an example where you must find the density function, see problems 5.8-5.11 in Section 5.2.

Recall:

We have discussed marginal and conditional distributions.
It is much easier to think about definitions in the discrete case:

$p(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2)$. (Recall that comma means "and".)

If we are interested in $P(Y_1 = y_1)$, this is the same as

$$P(Y_1 = y_1, Y_2 = \text{ anything}) = \sum_{\text{all } y_2} p(y_1, y_2).$$

So the "marginal distribution" $p_1(y_1)$ is given by

$$p_1(y_1) = \sum_{\text{all } y_2} p(y_1, y_2),$$

and in continuous case, by analogy,

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_2.$$

Here the integration with respect to y_2 replaces the sum over y_2.

Recall:

Next we looked at the conditional distribution $p(y_1 \mid y_2)$ and conditional density $f(y_1 \mid y_2)$.

Recall the analogy with conditional probability

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} :$$

$$p(y_1 \mid y_2) = \frac{p(y_1, y_2)}{p_2(y_2)}, \qquad f(y_1 \mid y_2) = \frac{f(y_1, y_2)}{f_2(y_2)}.$$

Remark:

Remembering these definitions is essential to being able to do problems without the aid of the text.

Exercise 5.33 (a):

Suppose that Y_1 is the total time between a customer's arrival in the store and departure from the service window, Y_2 is the time spent in line before reaching the window, and the joint density of these variables is

$$f(y_1, y_2) = \begin{cases} e^{-y_1} & 0 \leq y_2 \leq y_1 < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

Find the marginal density functions for Y_1 and Y_2.

Solution:

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_2 = \int_0^{y_2} e^{-y_1} dy_2$$
$$= e^{-y_1} \int_0^{y_2} 1\, dy_2 = y_1 e^{-y_1}.$$

For what values of y_1 does this calculation work?

Notice that if $y_1 \leq 0$, then $f(y_1, y_2) = 0$ for all y_2.
This means that $f_1(y_1) = 0$ if $y_1 < 0$.

Solution: (continued)

If $y_1 \geq 0$, then $f(y_1, y_2)$ is nonzero for $0 \leq y_2 \leq y_1$.
So our calculation works for $y_1 \geq 0$, and the marginal density function is

$$f_1(y_1) = \begin{cases} y_1 e^{-y_1} & y_1 \geq 0 \\ 0 & y_1 < 0. \end{cases}$$

Now we find $f_2(y_2)$:

$$\begin{aligned} f_2(y_2) &= \int_{-\infty}^{\infty} f(y_1, y_2) dy_1 = \int_{y_2}^{\infty} e^{-y_1} dy_1 \\ &= -e^{-y_1}\Big|_{y_2}^{\infty} = e^{-y_2}. \end{aligned}$$

This computation works if $y_2 \geq 0$.
If $y_2 < 0$, then $f(y_1, y_2) = 0$ for all y_1.
So

$$f_2(y_2) = \begin{cases} e^{-y_2} & y_2 \geq 0 \\ 0 & y_2 < 0. \end{cases}$$

Exercise 5.33 (b):

What is the conditional density function of Y_1 given that $Y_2 = y_2$? Be sure to specify the values of y_2 for which this conditional density is defined.

Solution:

$$f(y_1 \mid y_2) = \frac{f(y_1, y_2)}{f_2(y_2)}$$

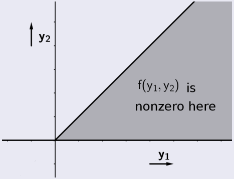
is defined for $f_2(y_2) > 0$.

$$\therefore f(y_1 \mid y_2) = \begin{cases} {}^{e^{-y_1}}\!/_{e^{-y_2}} & 0 \le y_2 \le y_1 < \infty \\ \text{undefined} & y_2 < 0 \\ 0 & 0 \le y_1 \le y_2 < \infty. \end{cases}$$

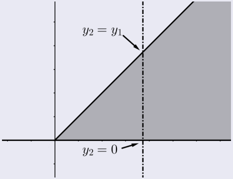
Note that ${}^{e^{-y_1}}\!/_{e^{-y_2}} = e^{-(y_1 - y_2)}$ (in order to reconcile with back of text). It may help to keep a picture of the plane in which the functions are defined.

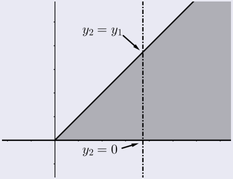
Solution: (continued)

Let's consider $f(y_1, y_2)$:



Recall when we found $f_1(y_1)$ we integrated over y_2.
This corresponds to the picture alongside if $y_1 \geq 0$:





If $y_2 < 0$, we integrate over a line, and $f(y_1, y_2)$ is identically zero on this line.

Remark on multiplication / division of case-defined functions:

If f has 2 cases and g has 2 cases, then fg has 4 cases.

Example:

$$f(x) = |x| = \begin{cases} x & x \geq 0 \\ -x & x < 0 \end{cases}, \quad g(x) = \begin{cases} x & -1 \leq x \leq 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Then

$$f(x)g(x) = \begin{cases} x \cdot x & x \geq 0, -1 \leq x \leq 1 \\ x \cdot 0 & x \geq 0, x \notin [-1, 1] \\ -x \cdot x & x < 0, -1 \leq x \leq 1 \\ -x \cdot 0 & x < 0, x \notin [-1, 1]. \end{cases}$$

Thus

$$f(x)g(x) = \begin{cases} x^2 & x \in [0, 1] \\ -x^2 & x \in [-1, 0) \\ 0 & x \notin [-1, 1]. \end{cases}$$

What we did for 5.33 (b) was much like this.

Independence of Random Variables

Recall:

RVs are independent iff the probability function (likewise, density function, and distribution function) is a product.

Remark:

There is a long series of problems which have not been assigned (but which you should look at anyway).
They ask "If Y_1, Y_2 have joint density function BLAH, are Y_1 and Y_2 independent?"

An important property of independent RVs:

If Y_1, Y_2 are independent, then $E[Y_1 Y_2] = E[Y_1]E[Y_2]$.

Remark:

Independence is a somewhat subtle property.
It is possible, for example, to construct RVs Y_1, Y_2, and Y_3, such that Y_1 and Y_2 are independent, Y_2 and Y_3 are independent, and Y_1 and Y_3 are independent; BUT Y_1, Y_2, Y_3 are not independent.

Definition (Expectation of Functions of RVs)

If Y_1, Y_2 have joint probability function $p(y_1, y_2)$, and g is a function of Y_1 and Y_2, then

$$E[g(Y_1, Y_2)] = \sum_{y_1} \sum_{y_2} g(y_1, y_2) p(y_1, y_2).$$

Analogously, for continuous RVs,

$$E[g(Y_1, Y_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y_1, y_2) f(y_1, y_2)\, dy_1\, dy_2.$$

Expectation has the same linearity properties we studied before. These can be used to simplify many problems.

Exercise 5.74:

Suppose that a radioactive particle is randomly located in a square with sides of unit length. A reasonable model for the joint density function for Y_1 and Y_2 is

$$f(y_1, y_2) = \begin{cases} 1 & 0 \le y_1 \le 1, 0 \le y_2 \le 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Find (a) $E[Y_1 - Y_2]$. (b) $E[Y_1 Y_2]$. (c) $E[Y_1^2 + Y_2^2]$. (d) $V[Y_1 Y_2]$.

Solution: (a)

Observe that the RVs are independent, and the marginal distributions are uniform on $[0, 1]$.
So

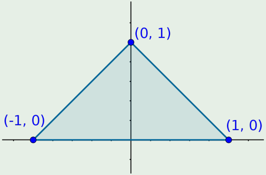
$$E[Y_1 - Y_2] = E[Y_1] - E[Y_2] = \frac{1}{2} - \frac{1}{2} = 0,$$

without any integration at all.

Tricks for Double-Integral Problems

Example

The joint distribution of Y_1, Y_2 is uniform over the triangle shown alongside. Find $E[Y_1]$.



Solution by formal procedure and integration:

We must integrate

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 f(y_1, y_2)\, dy_1\, dy_2,$$

where f is the joint density function.
What is this joint density function?
"Uniform" means f is a constant in the triangle (let's call it T) and 0 outside.

Solution by formal procedure and integration: (continued)

What is this constant?
We can find this constant by using the fact that the "total probability" is 1,
i.e.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1\, dy_2 = 1.$$

$$\therefore \quad \iint_T c\, dy_1\, dy_2 = 1 \qquad \text{because } f = \begin{cases} c & \text{inside } T \\ 0 & \text{elsewhere.} \end{cases}$$

$$\therefore \quad c \iint_T 1\, dy_1\, dy_2 = 1 \implies c \cdot \text{area}(T) = 1.$$

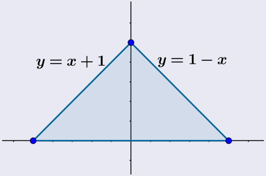
But $\text{area}(T) = \frac{1}{2} \cdot b \cdot h = \frac{1}{2} \cdot 2 \cdot 1 = 1.$
So $c = 1$.
So now we know the joint density f.

To find $E[Y_1]$, we must integrate

$$E[Y_1] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 f(y_1, y_2)\, dy_1\, dy_2 = \iint_T y_1\, dy_1\, dy_2.$$

Solution by formal procedure and integration: (continued)

Now we must find the limits of integration for the triangle:



$$\iint_T y_1 \, dy_1 \, dy_2 = \int_{y_1=-1}^{0} \int_{y_2=0}^{y_1+1} y_1 \, dy_2 \, dy_1 + \int_0^1 \int_0^{1-y_1} y_1 \, dy_2 \, dy_1$$

$$= \int_{-1}^{0} y_1 \int_0^{y_1+1} dy_2 \, dy_1 + \int_0^1 y_1 \int_0^{1-y_1} dy_2 \, dy_1$$

$$= \int_{-1}^{0} y_1(y_1+1) dy_1 + \int_0^1 y_1(1-y_1) dy_1$$

$$= \int_{-1}^{0} (y_1^2 + y_1) dy_1 + \int_0^1 (y_1 - y_1^2) dy_1$$

Solution by formal procedure and integration: (continued)

$$= \left(\frac{y^3}{3} + \frac{y^2}{2} \right) \Bigg|_{-1}^{0} + \left(\frac{y^2}{2} - \frac{y^3}{3} \right) \Bigg|_{0}^{1}$$

Solution by formal procedure and integration: (continued)

$$= \left(\frac{y^3}{3} + \frac{y^2}{2}\right)\Big|_{-1}^{0} + \left(\frac{y^2}{2} - \frac{y^3}{3}\right)\Big|_{0}^{1}$$

$$= -\left[\frac{(-1)^3}{3} + \frac{(-1)^2}{2}\right] + \left[\frac{1^2}{2} - \frac{1^3}{3}\right]$$

Solution by formal procedure and integration: (continued)

$$= \left(\frac{y^3}{3} + \frac{y^2}{2}\right)\Big|_{-1}^{0} + \left(\frac{y^2}{2} - \frac{y^3}{3}\right)\Big|_{0}^{1}$$

$$= -\left[\frac{(-1)^3}{3} + \frac{(-1)^2}{2}\right] + \left[\frac{1^2}{2} - \frac{1^3}{3}\right]$$

$$= -\left[\frac{1}{2} - \frac{1}{3}\right] + \left[\frac{1}{2} - \frac{1}{3}\right]$$

Solution by formal procedure and integration: (continued)

$$
\begin{aligned}
&= \left(\frac{y^3}{3} + \frac{y^2}{2}\right)\Big|_{-1}^{0} + \left(\frac{y^2}{2} - \frac{y^3}{3}\right)\Big|_{0}^{1} \\
&= -\left[\frac{(-1)^3}{3} + \frac{(-1)^2}{2}\right] + \left[\frac{1^2}{2} - \frac{1^3}{3}\right] \\
&= -\left[\frac{1}{2} - \frac{1}{3}\right] + \left[\frac{1}{2} - \frac{1}{3}\right] = \boxed{0}.
\end{aligned}
$$

Solution by formal procedure and integration: (continued)

$$= \left(\frac{y^3}{3} + \frac{y^2}{2}\right)\Big|_{-1}^{0} + \left(\frac{y^2}{2} - \frac{y^3}{3}\right)\Big|_{0}^{1}$$

$$= -\left[\frac{(-1)^3}{3} + \frac{(-1)^2}{2}\right] + \left[\frac{1^2}{2} - \frac{1^3}{3}\right]$$

$$= -\left[\frac{1}{2} - \frac{1}{3}\right] + \left[\frac{1}{2} - \frac{1}{3}\right] = \boxed{0}.$$

Solution by symmetry:

Solution by formal procedure and integration: (continued)

$$= \left(\frac{y^3}{3} + \frac{y^2}{2} \right) \Bigg|_{-1}^{0} + \left(\frac{y^2}{2} - \frac{y^3}{3} \right) \Bigg|_{0}^{1}$$

$$= - \left[\frac{(-1)^3}{3} + \frac{(-1)^2}{2} \right] + \left[\frac{1^2}{2} - \frac{1^3}{3} \right]$$

$$= - \left[\frac{1}{2} - \frac{1}{3} \right] + \left[\frac{1}{2} - \frac{1}{3} \right] = \boxed{0}.$$

Solution by symmetry:

What is the geometric interpretation of $E[Y_1]$?

Solution by formal procedure and integration: (continued)

$$= \left(\frac{y^3}{3} + \frac{y^2}{2}\right)\Bigg|_{-1}^{0} + \left(\frac{y^2}{2} - \frac{y^3}{3}\right)\Bigg|_{0}^{1}$$

$$= -\left[\frac{(-1)^3}{3} + \frac{(-1)^2}{2}\right] + \left[\frac{1^2}{2} - \frac{1^3}{3}\right]$$

$$= -\left[\frac{1}{2} - \frac{1}{3}\right] + \left[\frac{1}{2} - \frac{1}{3}\right] = \boxed{0}.$$

Solution by symmetry:

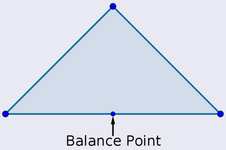
What is the geometric interpretation of $E[Y_1]$?
This is the y_1-coordinate such that we can "balance" the triangle at this point,

Solution by formal procedure and integration: (continued)

$$
\begin{aligned}
&= \left(\frac{y^3}{3} + \frac{y^2}{2}\right)\Big|_{-1}^{0} + \left(\frac{y^2}{2} - \frac{y^3}{3}\right)\Big|_{0}^{1} \\
&= -\left[\frac{(-1)^3}{3} + \frac{(-1)^2}{2}\right] + \left[\frac{1^2}{2} - \frac{1^3}{3}\right] \\
&= -\left[\frac{1}{2} - \frac{1}{3}\right] + \left[\frac{1}{2} - \frac{1}{3}\right] = \boxed{0}.
\end{aligned}
$$

Solution by symmetry:

What is the geometric interpretation of $E[Y_1]$? This is the y_1-coordinate such that we can "balance" the triangle at this point, i.e.,

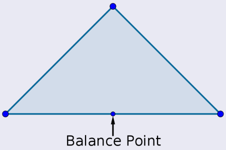


Solution by formal procedure and integration: (continued)

$$= \left(\frac{y^3}{3} + \frac{y^2}{2}\right)\Big|_{-1}^{0} + \left(\frac{y^2}{2} - \frac{y^3}{3}\right)\Big|_{0}^{1}$$
$$= -\left[\frac{(-1)^3}{3} + \frac{(-1)^2}{2}\right] + \left[\frac{1^2}{2} - \frac{1^3}{3}\right]$$
$$= -\left[\frac{1}{2} - \frac{1}{3}\right] + \left[\frac{1}{2} - \frac{1}{3}\right] = \boxed{0}.$$

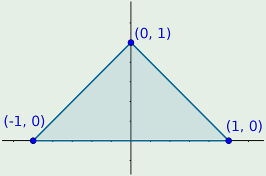
Solution by symmetry:

What is the geometric interpretation of $E[Y_1]$?
This is the y_1-coordinate such that we can "balance" the triangle at this point, i.e.,
By symmetry, $E(Y_1) = 0$.



Example

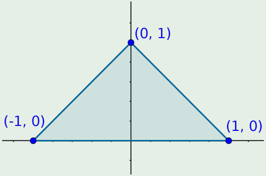
With (Y_1, Y_2) uniformly distributed over the triangle T alongside, find $E[Y_2]$.



Example

With (Y_1, Y_2) uniformly distributed over the triangle T alongside, find $E[Y_2]$.
You can use the results obtained earlier, e.g.

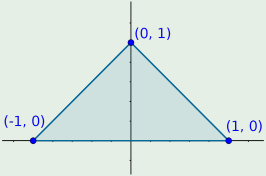
$$f(y_1, y_2) = \begin{cases} 1 & \text{in } T \\ 0 & \text{elsewhere.} \end{cases}$$



Example

With (Y_1, Y_2) uniformly distributed over the triangle T alongside, find $E[Y_2]$.
You can use the results obtained earlier, e.g.

$$f(y_1, y_2) = \begin{cases} 1 & \text{in } T \\ 0 & \text{elsewhere.} \end{cases}$$

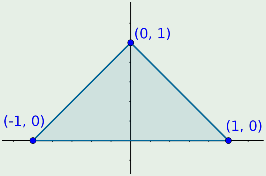


Solution:

Example

With (Y_1, Y_2) uniformly distributed over the triangle T alongside, find $E[Y_2]$.
You can use the results obtained earlier, e.g.

$$f(y_1, y_2) = \begin{cases} 1 & \text{in } T \\ 0 & \text{elsewhere.} \end{cases}$$



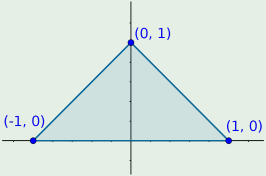
Solution:

$$E[Y_2] = \iint_T y_2 \, dy_1 \, dy_2$$

Example

With (Y_1, Y_2) uniformly distributed over the triangle T alongside, find $E[Y_2]$.
You can use the results obtained earlier, e.g.

$$f(y_1, y_2) = \begin{cases} 1 & \text{in } T \\ 0 & \text{elsewhere.} \end{cases}$$



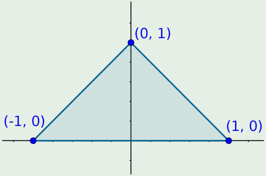
Solution:

$$E[Y_2] = \iint_T y_2 \, dy_1 \, dy_2 \qquad \text{(just like before)}$$

Example

With (Y_1, Y_2) uniformly distributed over the triangle T alongside, find $E[Y_2]$.
You can use the results obtained earlier, e.g.

$$f(y_1, y_2) = \begin{cases} 1 & \text{in } T \\ 0 & \text{elsewhere.} \end{cases}$$



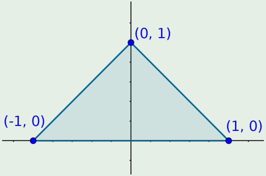
Solution:

$$E[Y_2] = \iint_T y_2 \, dy_1 \, dy_2 \qquad \text{(just like before)}$$
$$= \int_{y_1=-1}^{0} \int_{y_2=0}^{y_1+1} y_2 \, dy_2 \, dy_1 + \int_0^1 \int_0^{1-y_1} y_2 \, dy_2 \, dy_1$$

Example

With (Y_1, Y_2) uniformly distributed over the triangle T alongside, find $E[Y_2]$.
You can use the results obtained earlier, e.g.

$$f(y_1, y_2) = \begin{cases} 1 & \text{in } T \\ 0 & \text{elsewhere.} \end{cases}$$



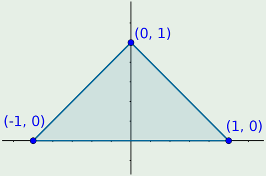
Solution:

$$\begin{aligned} E[Y_2] &= \iint_T y_2 \, dy_1 \, dy_2 \qquad \text{(just like before)} \\ &= \int_{y_1=-1}^{0} \int_{y_2=0}^{y_1+1} y_2 \, dy_2 \, dy_1 + \int_0^1 \int_0^{1-y_1} y_2 \, dy_2 \, dy_1 \\ &= \int_{-1}^{0} \left. \frac{y_2^2}{2} \right|_0^{y_1+1} dy_1 + \int_0^1 \left. \frac{y_2^2}{2} \right|_0^{1-y_1} dy_1 \end{aligned}$$

Example

With (Y_1, Y_2) uniformly distributed over the triangle T alongside, find $E[Y_2]$.
You can use the results obtained earlier, e.g.

$$f(y_1, y_2) = \begin{cases} 1 & \text{in } T \\ 0 & \text{elsewhere.} \end{cases}$$



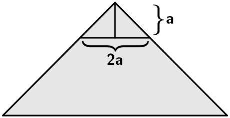
Solution:

$$\begin{aligned} E[Y_2] &= \iint_T y_2 \, dy_1 \, dy_2 \qquad \text{(just like before)} \\ &= \int_{y_1=-1}^{0} \int_{y_2=0}^{y_1+1} y_2 \, dy_2 \, dy_1 + \int_0^1 \int_0^{1-y_1} y_2 \, dy_2 \, dy_1 \\ &= \int_{-1}^{0} \left. \frac{y_2^2}{2} \right|_0^{y_1+1} dy_1 + \int_0^1 \left. \frac{y_2^2}{2} \right|_0^{1-y_1} dy_1 \\ &= \int_{-1}^{0} \frac{(y_1+1)^2}{2} \, dy_1 + \int_0^1 \frac{(1-y_1)^2}{2} \, dy_1 \end{aligned}$$

Solution: (continued)

$$\begin{aligned}
&= \int_{-1}^{0} \left(\frac{(y_1^2}{2} + y_1 + \frac{1}{2} \right) \, dy_1 + \int_0^1 \left(\frac{1}{2} - y_1 + \frac{y_1^2}{2} \right) \, dy_1 \\
&= \left(\frac{y_1^3}{6} + \frac{y_1^2}{2} + \frac{y_1}{2} \right) \Bigg|_{-1}^{0} + \left(\frac{y_1}{2} - \frac{y_1^2}{2} + \frac{y_1^3}{6} \right) \Bigg|_{0}^{1} \\
&= -\left(\frac{(-1)^3}{6} + \frac{(-1)^2}{2} + \frac{(-1)}{2} \right) + \left(\frac{1}{2} - \frac{1^2}{2} + \frac{1^3}{6} \right) \\
&= -\left(-\frac{1}{6} + \frac{1}{2} - \frac{1}{2} \right) + \left(\frac{1}{2} - \frac{1}{2} + \frac{1}{6} \right) = \frac{1}{6} + \frac{1}{6} = \boxed{\frac{1}{3}}.
\end{aligned}$$

Suppose we do this problem by symmetry like this:



Find a so that the area of the top triangle is the same as the area of the bottom trapezoid.

The a that works has

$$\frac{1}{2}\cdot 2a\cdot a = 1-\frac{1}{2}\cdot 2a\cdot a \implies a^2 = 1-a^2 \implies a = \frac{\sqrt{2}}{2} \neq \frac{2}{3}.$$

Why is this wrong?

We are not asking for "area of top Δ" $=$ "area of the trapezoid".
We are asking geometrically for the balance point.
Weight far from the balance point disturbs the balance more than the weight near the balance point.

This problem shows the weakness of symmetry methods:
It's possible to find an attractive argument which is just wrong.

Covariance and Correlation

Definition (Covariance)

The covariance of two RVs Y_1 and Y_2 is

$$\mathrm{Cov}(Y_1, Y_2) = E[(Y_1 - \mu)(Y_2 - \nu)],$$

wher $\mu = E[Y_1]$ and $\nu = E[Y_2]$.

Definition (Correlation)

The correlation of the RVs Y_1 and Y_2 is measured by

$$\rho_{Y_1, Y_2} = \frac{\mathrm{Cov}(Y_1, Y_2)}{\sigma_1 \sigma_2},$$

where σ_1, σ_2 are the standard deviations of Y_1, Y_2, respectively.

Remark:

In order to compute these, we need to know the joint distribution of Y_1 and Y_2.

Remarks:

- Using the linearity properties of E, we have
$$E[(Y_1 - \mu)(Y_2 - \nu)] = E[Y_1 Y_2] - \mu\nu.$$
This is basically the same calculation that gave us
$$V[Y] = E[(Y - \mu)^2] = E[Y^2] - \mu^2.$$
- If Y_1, Y_2 are independent, then $E[Y_1 Y_2] = E[Y_1]E[Y_2]$, so that
$$\mathrm{Cov}(Y_1, Y_2) = E[Y_1 Y_2] - \mu\nu = E[Y_1]E[Y_2] - \mu\nu = \mu\nu - \mu\nu = 0.$$
Thus
$$Y_1, Y_2 \text{ independent } \implies \mathrm{Cov}(Y_1, Y_2) = 0.$$
But the converse is not true!

Example

Suppose that Y_1, Y_2 are discrete RVs whose joint probability function is given by the following table:

$\downarrow Y_2 \quad Y_1 \rightarrow$	-1	0	$+1$	$p_2(Y_2) \downarrow$
-1	$^1/_{16}$	$^3/_{16}$	$^1/_{16}$	$^5/_{16}$
0	$^3/_{16}$	0	$^3/_{16}$	$^6/_{16}$
$+1$	$^1/_{16}$	$^3/_{16}$	$^1/_{16}$	$^5/_{16}$
$p_1(Y_1) \rightarrow$	$^5/_{16}$	$^6/_{16}$	$^5/_{16}$	

$\mu = E[Y_1] = 0, \nu = E[Y_2] = 0.$
What is $E[Y_1 Y_2]$?

$$+1 \cdot \frac{1}{16} + 1 \cdot \frac{1}{16} - 1 \cdot \frac{1}{16} - 1 \cdot \frac{1}{16} = 0.$$

$+1$	0	-1
0	0	0
-1	0	$+1$

Example (continued)

$$\mathrm{Cov}(Y_1, Y_2) = E[Y_1 Y_2] - \mu\nu = 0 - 0 = \boxed{0}.$$

But Y_1, Y_2 are NOT independent.

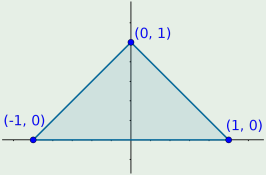
Recall that Y_1 and Y_2 are independent if the joint probability function $p(y_1, y_2)$ is the product of the marginal distributions:

$$p(y_1, y_2) = p_1(y_1) \cdot p_2(y_2).$$

But $0 \neq {}^6\!/\!_{16} \cdot {}^6\!/\!_{16}$.
So Y_1 and Y_2 are NOT independent.

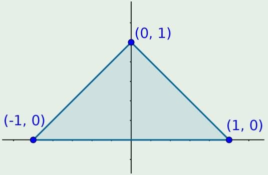
Example (continuous RVs)

The joint distribution of Y_1, Y_2 is uniform over the triangle shown alongside. Are Y_1 and Y_2 independent?



Example (continuous RVs)

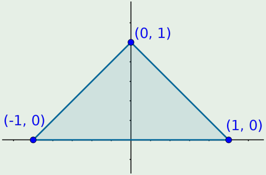
The joint distribution of Y_1, Y_2 is uniform over the triangle shown alongside. Are Y_1 and Y_2 independent?



No, because if they were, the joint density function would be nonzero in a rectangle.

Example (continuous RVs)

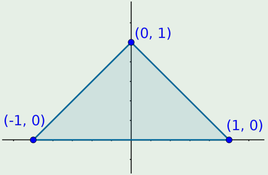
The joint distribution of Y_1, Y_2 is uniform over the triangle shown alongside. Are Y_1 and Y_2 independent?



No, because if they were, the joint density function would be nonzero in a rectangle. What is $\text{Cov}(Y_1, Y_2)$?

Example (continuous RVs)

The joint distribution of Y_1, Y_2 is uniform over the triangle shown alongside. Are Y_1 and Y_2 independent?

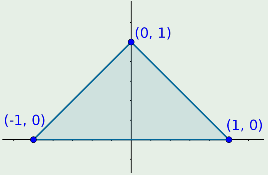


No, because if they were, the joint density function would be nonzero in a rectangle. What is $\text{Cov}(Y_1, Y_2)$?

$$\text{Cov}(Y_1, Y_2) = E[Y_1 Y_2] - E[Y_1] \cdot E[Y_2]$$

Example (continuous RVs)

The joint distribution of Y_1, Y_2 is uniform over the triangle shown alongside. Are Y_1 and Y_2 independent?

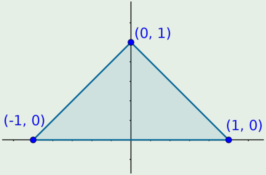


No, because if they were, the joint density function would be nonzero in a rectangle. What is $\mathrm{Cov}(Y_1, Y_2)$?

$$\mathrm{Cov}(Y_1, Y_2) = E[Y_1 Y_2] - \overset{0}{\cancel{E[Y_1]}} \cdot E[Y_2] = E[Y_1 Y_2].$$

Example (continuous RVs)

The joint distribution of Y_1, Y_2 is uniform over the triangle shown alongside. Are Y_1 and Y_2 independent?



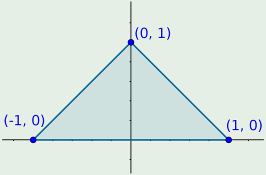
No, because if they were, the joint density function would be nonzero in a rectangle. What is $\mathrm{Cov}(Y_1, Y_2)$?

$$\mathrm{Cov}(Y_1, Y_2) = E[Y_1 Y_2] - \cancelto{0}{E[Y_1]} \cdot E[Y_2] = E[Y_1 Y_2].$$

After some computing, we find that $E[Y_1 Y_2] = 0$.

Example (continuous RVs)

The joint distribution of Y_1, Y_2 is uniform over the triangle shown alongside. Are Y_1 and Y_2 independent?



No, because if they were, the joint density function would be nonzero in a rectangle. What is $\text{Cov}(Y_1, Y_2)$?

$$\text{Cov}(Y_1, Y_2) = E[Y_1 Y_2] - \cancelto{0}{E[Y_1]} \cdot E[Y_2] = E[Y_1 Y_2].$$

After some computing, we find that $E[Y_1 Y_2] = 0$.

So $\text{Cov}(Y_1, Y_2) = 0$, but Y_1 and Y_2 are NOT independent.

Recall:

$$\boxed{\mathrm{Cov}(X, Y) = E[(X-\mu)(Y-\nu)]},$$

where $\mu = E[X]$ and $\nu = E[Y]$.
We noted the convenience that we can express $\mathrm{Cov}(X, Y)$ as $E[XY] - \mu\nu$:

$$\begin{aligned} E[(X-\mu)(Y-\nu)] &= E[XY - \mu Y - \nu X + \mu\nu] \\ &= E[XY] - \mu E[Y] - \nu E[X] + \mu\nu \\ &= E[XY] - \mu\nu - \mu\nu + \mu\nu = E[XY] - \mu\nu. \end{aligned}$$

Remark:

The covariance is a measure of the extent to which X and Y "vary together".
Note that the covariance can be negative.

Theorem (5.12)

Let $Y_1, \ldots, Y_n$ and $X_1, \ldots, X_m$ be random variables with $E(Y_i) = \mu_i$ and $E(X_j) = \nu_j$. Define

$$U_1 = \sum_{i=1}^{n} a_i Y_i \qquad \text{and} \qquad U_2 = \sum_{j=1}^{m} b_j X_j$$

for constants $a_1 \ldots a_n$ and $b_1 \ldots b_m$.
Then

(a) $E[U_1] = \sum_{i=1}^{n} a_i \mu_i$.

(b) $V[U_1] = \sum_{i=1}^{n} a_i^2 V[Y_i] + 2 \sum\sum_{1 \leq i < j \leq n} a_i a_j \; Cov(Y_i, Y_j)$.

(c) $Cov(U_1, U_2) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \; Cov(Y_i, X_j)$.

The above theorem presents an important property of covariance:
Covariance is bilinear,
that is, it is a function of two variables which is separately linear in each variable.

Thus

$$\mathrm{Cov}(2X+1, Y) = \mathrm{Cov}(2X, Y) + \mathrm{Cov}(1, Y).$$

Also

$$\mathrm{Cov}(X, 3Y+4) = \mathrm{Cov}(X, 3Y) + \mathrm{Cov}(X, 4).$$

Also

$$\mathrm{Cov}(2X, Y) = 2\,\mathrm{Cov}(X, Y), \mathrm{Cov}(X, 3Y) = 3\,\mathrm{Cov}(X, Y).$$

Remark:

Note that it is "separately" linear in each variable:
it is NOT true that

$$\mathrm{Cov}(2X+1, 2Y+1) = 2\,\mathrm{Cov}(X, Y) + 1.$$

How does this give us the complicated statement of Theorem 5.12?

Observe that $V[X] = \text{Cov}(X, X)$.

If $U = b_1 X_1 + \cdots + b_m X_m$, then we can apply bilinearity to compute

$$
\begin{aligned}
& V[U] = \text{Cov}(U, U) \\
&= \text{Cov}(b_1 X_1 + \cdots + b_m X_m, b_1 X_1 + \cdots + b_m X_m) \\
&= b_1 \text{Cov}(X_1, b_1 X_1 + \cdots + b_m X_m) + \dots \\
&\quad + b_m \text{Cov}(X_m, b_1 X_1 + \cdots + b_m X_m) \\
&= b_1 b_2 \text{Cov}(X_1, X_1) + b_1 b_2 \text{Cov}(X_1, X_2) + \cdots + b_1 b_m \text{Cov}(X_1, X_m) \\
&\quad + b_2 b_1 \text{Cov}(X_2, X_1) + b_2 b_2 \text{Cov}(X_2, X_2) + \cdots + b_2 b_m \text{Cov}(X_2, X_m) \\
&\quad + \qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
&\quad + b_m b_1 \text{Cov}(X_m, X_1) + b_m b_2 \text{Cov}(X_m, X_2) + \cdots + b_m b_m \text{Cov}(X_m, X_m) \\
&= \sum_{j=1}^{m} b_j^2 V[X_j] + 2 \sum\sum_{1 \le i < j \le n} b_i b_j \text{Cov}(X_i, X_j).
\end{aligned}
$$

Remark:

The proof of Theorem 5.12 consists of calculations like this one.
The key point is the verification (directly from the definition) that the covariance is bilinear.

How will this come up?
You may be asked to compute some variance (or covariance) and the easiest way to do it will be to use this result.

Exercise 5.112

Let Y_1 and Y_2 denote the lengths of life, in hundreds of hours, for components of types I and II, respectively, in an electronic system. The joint density of Y_1 and Y_2 is

$$f(y_1, y_2) = \begin{cases} \dfrac{y_1}{8} e^{-(y_1 + y_2)/2} & y_1 > 0, y_2 > 0, \\ 0 & \text{elsewhere.} \end{cases}$$

The cost C of replacing the two components depends upon their length of life at failure and is given by $C = 50 + 2Y_1 + 4Y_2$. Find $E[C]$ and $V[C]$.

Solution:

Solution:

Additive constants don't matter for variance and covariance, i.e. $V[X+a] = V[X]$. Also $\mathrm{Cov}(X+a, Y) = \mathrm{Cov}(X, Y)$.

Solution:

Additive constants don't matter for variance and covariance, i.e. $V[X + a] = V[X]$. Also $\mathrm{Cov}(X + a, Y) = \mathrm{Cov}(X, Y)$. So $V[C]$ above is

$$\mathrm{Cov}(2Y_1 + 4Y_2, 2Y_1 + 4Y_2)$$

Solution:

Additive constants don't matter for variance and covariance, i.e. $V[X+a]=V[X]$. Also $\mathrm{Cov}(X+a,Y)=\mathrm{Cov}(X,Y)$. So $V[C]$ above is

$$\begin{aligned}&\mathrm{Cov}(2Y_1+4Y_2,2Y_1+4Y_2)\\ =\ &2\mathrm{Cov}(Y_1,2Y_1+4Y_2)+4\mathrm{Cov}(Y_2,2Y_1+4Y_2)\end{aligned}$$

Solution:

Additive constants don't matter for variance and covariance, i.e. $V[X+a] = V[X]$. Also $\text{Cov}(X+a, Y) = \text{Cov}(X, Y)$. So $V[C]$ above is

$$\begin{aligned}
&\text{Cov}(2Y_1 + 4Y_2, 2Y_1 + 4Y_2) \\
&= 2\text{Cov}(Y_1, 2Y_1 + 4Y_2) + 4\text{Cov}(Y_2, 2Y_1 + 4Y_2) \\
&= 2 \cdot 2\text{Cov}(Y_1, Y_1) + 4 \cdot 2\text{Cov}(Y_1, Y_2) + 2 \cdot 4\text{Cov}(Y_2, Y_1) \\
&\quad + 4 \cdot 4\text{Cov}(Y_2, Y_2)
\end{aligned}$$

Solution:

Additive constants don't matter for variance and covariance, i.e. $V[X+a] = V[X]$. Also $\mathrm{Cov}(X+a, Y) = \mathrm{Cov}(X, Y)$. So $V[C]$ above is

$$\begin{aligned}
&\mathrm{Cov}(2Y_1 + 4Y_2, 2Y_1 + 4Y_2) \\
&= 2\mathrm{Cov}(Y_1, 2Y_1 + 4Y_2) + 4\mathrm{Cov}(Y_2, 2Y_1 + 4Y_2) \\
&= 2 \cdot 2\mathrm{Cov}(Y_1, Y_1) + 4 \cdot 2\mathrm{Cov}(Y_1, Y_2) + 2 \cdot 4\mathrm{Cov}(Y_2, Y_1) \\
&\quad + 4 \cdot 4\mathrm{Cov}(Y_2, Y_2) \qquad [\text{Remark: } \mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)]
\end{aligned}$$

Solution:

Additive constants don't matter for variance and covariance, i.e. $V[X+a]=V[X]$. Also $\text{Cov}(X+a,Y)=\text{Cov}(X,Y)$. So $V[C]$ above is

$$\begin{aligned}
&\text{Cov}(2Y_1+4Y_2,2Y_1+4Y_2)\\
&=2\text{Cov}(Y_1,2Y_1+4Y_2)+4\text{Cov}(Y_2,2Y_1+4Y_2)\\
&=2\cdot 2\text{Cov}(Y_1,Y_1)+4\cdot 2\text{Cov}(Y_1,Y_2)+2\cdot 4\text{Cov}(Y_2,Y_1)\\
&\quad+4\cdot 4\text{Cov}(Y_2,Y_2) \qquad [\text{Remark: Cov}(X,Y)=\text{Cov}(Y,X)]\\
&=2\cdot 2V[Y_1]+(4\cdot 2+2\cdot 4)\text{Cov}(Y_1,Y_2)+4\cdot 4V[Y_2]
\end{aligned}$$

Solution:

Additive constants don't matter for variance and covariance, i.e. $V[X+a] = V[X]$. Also $\mathrm{Cov}(X+a, Y) = \mathrm{Cov}(X, Y)$. So $V[C]$ above is

$$\begin{aligned}
&\mathrm{Cov}(2Y_1 + 4Y_2, 2Y_1 + 4Y_2) \\
&= 2\mathrm{Cov}(Y_1, 2Y_1 + 4Y_2) + 4\mathrm{Cov}(Y_2, 2Y_1 + 4Y_2) \\
&= 2 \cdot 2\mathrm{Cov}(Y_1, Y_1) + 4 \cdot 2\mathrm{Cov}(Y_1, Y_2) + 2 \cdot 4\mathrm{Cov}(Y_2, Y_1) \\
&\quad + 4 \cdot 4\mathrm{Cov}(Y_2, Y_2) \qquad [\text{Remark: } \mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)] \\
&= 2 \cdot 2V[Y_1] + (4 \cdot 2 + 2 \cdot 4)\mathrm{Cov}(Y_1, Y_2) + 4 \cdot 4V[Y_2] \\
&= 4V[Y_1] + 16\mathrm{Cov}(Y_1, Y_2) + 16V[Y_2].
\end{aligned}$$

Solution:

Additive constants don't matter for variance and covariance, i.e. $V[X+a] = V[X]$. Also $\text{Cov}(X+a, Y) = \text{Cov}(X, Y)$. So $V[C]$ above is

$$
\begin{aligned}
&\text{Cov}(2Y_1 + 4Y_2, 2Y_1 + 4Y_2) \\
&= 2\text{Cov}(Y_1, 2Y_1 + 4Y_2) + 4\text{Cov}(Y_2, 2Y_1 + 4Y_2) \\
&= 2 \cdot 2\text{Cov}(Y_1, Y_1) + 4 \cdot 2\text{Cov}(Y_1, Y_2) + 2 \cdot 4\text{Cov}(Y_2, Y_1) \\
&\quad + 4 \cdot 4\text{Cov}(Y_2, Y_2) \qquad [\text{Remark: } \text{Cov}(X, Y) = \text{Cov}(Y, X)] \\
&= 2 \cdot 2V[Y_1] + (4 \cdot 2 + 2 \cdot 4)\text{Cov}(Y_1, Y_2) + 4 \cdot 4V[Y_2] \\
&= 4V[Y_1] + 16\text{Cov}(Y_1, Y_2) + 16V[Y_2].
\end{aligned}
$$

Remarks:

Solution:

Additive constants don't matter for variance and covariance, i.e. $V[X+a] = V[X]$. Also $\text{Cov}(X+a, Y) = \text{Cov}(X, Y)$. So $V[C]$ above is

$$\begin{aligned}
&\text{Cov}(2Y_1 + 4Y_2, 2Y_1 + 4Y_2) \\
&= 2\text{Cov}(Y_1, 2Y_1 + 4Y_2) + 4\text{Cov}(Y_2, 2Y_1 + 4Y_2) \\
&= 2 \cdot 2\text{Cov}(Y_1, Y_1) + 4 \cdot 2\text{Cov}(Y_1, Y_2) + 2 \cdot 4\text{Cov}(Y_2, Y_1) \\
&\quad + 4 \cdot 4\text{Cov}(Y_2, Y_2) \qquad [\text{Remark: Cov}(X, Y) = \text{Cov}(Y, X)] \\
&= 2 \cdot 2V[Y_1] + (4 \cdot 2 + 2 \cdot 4)\text{Cov}(Y_1, Y_2) + 4 \cdot 4V[Y_2] \\
&= 4V[Y_1] + 16\text{Cov}(Y_1, Y_2) + 16V[Y_2].
\end{aligned}$$

Remarks:

- This dupliates the calculation for Theorem 5.12, but with 4 terms and not m^2 terms.

Solution:

Additive constants don't matter for variance and covariance, i.e. $V[X+a]=V[X]$. Also $\text{Cov}(X+a,Y)=\text{Cov}(X,Y)$. So $V[C]$ above is

$$\begin{aligned}
&\text{Cov}(2Y_1+4Y_2, 2Y_1+4Y_2)\\
&=2\text{Cov}(Y_1, 2Y_1+4Y_2)+4\text{Cov}(Y_2, 2Y_1+4Y_2)\\
&=2\cdot 2\text{Cov}(Y_1,Y_1)+4\cdot 2\text{Cov}(Y_1,Y_2)+2\cdot 4\text{Cov}(Y_2,Y_1)\\
&\quad +4\cdot 4\text{Cov}(Y_2,Y_2) \qquad [\text{Remark: } \text{Cov}(X,Y)=\text{Cov}(Y,X)]\\
&=2\cdot 2V[Y_1]+(4\cdot 2+2\cdot 4)\text{Cov}(Y_1,Y_2)+4\cdot 4V[Y_2]\\
&=4V[Y_1]+16\text{Cov}(Y_1,Y_2)+16V[Y_2].
\end{aligned}$$

Remarks:

- This dupliates the calculation for Theorem 5.12, but with 4 terms and not m^2 terms.
- The biggest "difficulty" in some of the problems is setting up and doing double integrals.

Practice Problem 1: Exercise 5.27 (b)

Given that the joint density function of Y_1 and Y_2 is

$$f(y_1, y_2) = \begin{cases} 6(1 - y_2) & 0 \le y_1 \le y_2 \le 1, \\ 0 & \text{elsewhere,} \end{cases}$$

find $P(Y_2 \le 1/2 \mid Y_1 \le 3/4)$.

Solution:

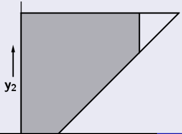
We need to know $\dfrac{P(Y_2 \le 1/2, Y_1 \le 3/4)}{P(Y_1 \le 3/4)}$.

To get the two parts of this fraction, we must compute

$$P(Y_1 \le 3/4) = \iint_{Y_1 \le \frac{3}{4}} f(y_1, y_2)\, dy_1\, dy_2.$$

We need to graph the region of integration.

First draw the region where $f \neq 0$:



Solution: (continued)

$$
\begin{aligned}
P\left(Y_1 \le \frac{3}{4}\right) &= \iint_{\text{shaded region}} 6(1-y_2)\,dy_1\,dy_2 \\
&= \int_{y_1=0}^{3/4} \int_{y_2=y_1}^{1} 6(1-y_2)\,dy_2\,dy_1 \\
&= \int_0^{3/4} \left(6y_2 - 3y_2^2 \Big|_{y_1}^{1} \right) dy_1 \\
&= \int_0^{3/4} \left[(6-3) - (6y_1 - 3y_1^2) \right] dy_1 \\
&= \int_0^{3/4} \left[3 - 6y_1 + 3y_1^2 \right] dy_1 \\
&= \int_0^{3/4} 3(1-y_1)^2\,dy_1 = 3 \cdot \left(-\frac{1}{3}\right)(1-y_1)^3 \Big|_0^{3/4}
\end{aligned}
$$

Solution: (continued)

$$= -(1-y_1)^3\Big|_0^{3/4}$$

Solution: (continued)

$$= -(1-y_1)^3\Big|_0^{3/4} = \left[-\left(1-\frac{3}{4}\right)^3\right] - \left[-(1-0)^3\right]$$

Solution: (continued)

$$= -(1-y_1)^3\Big|_0^{3/4} = \left[-\left(1-\frac{3}{4}\right)^3\right] - \left[-(1-0)^3\right]$$

$$= -\left(\frac{1}{4}\right)^3 + 1$$

Solution: (continued)

$$= -(1-y_1)^3\Big|_0^{3/4} = \left[-\left(1-\frac{3}{4}\right)^3\right] - \left[-(1-0)^3\right]$$

$$= -\left(\frac{1}{4}\right)^3 + 1 = -\frac{1}{64} + 1$$

Solution: (continued)

$$= -(1-y_1)^3\Big|_0^{3/4} = \left[-\left(1-\frac{3}{4}\right)^3\right] - \left[-(1-0)^3\right]$$

$$= -\left(\frac{1}{4}\right)^3 + 1 = -\frac{1}{64} + 1 = \frac{63}{64}.$$

Solution: (continued)

$$= -(1-y_1)^3\Big|_0^{3/4} = \left[-\left(1-\frac{3}{4}\right)^3\right] - \left[-(1-0)^3\right]$$

$$= -\left(\frac{1}{4}\right)^3 + 1 = -\frac{1}{64} + 1 = \frac{63}{64}.$$

Next: $P\left(Y_2 \le {}^1\!/_2, Y_1 \le {}^3\!/_4\right)$.

Solution: (continued)

$$= -(1-y_1)^3\Big|_0^{3/4} = \left[-\left(1-\frac{3}{4}\right)^3\right] - \left[-(1-0)^3\right]$$

$$= -\left(\frac{1}{4}\right)^3 + 1 = -\frac{1}{64} + 1 = \frac{63}{64}.$$

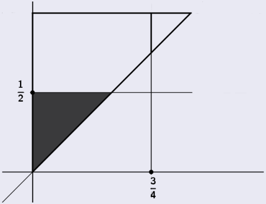
Next: $P\left(Y_2 \le {}^1\!/_2, Y_1 \le {}^3\!/_4\right)$.
Again, draw the region:

Solution: (continued)

$$= -(1-y_1)^3\Big|_0^{3/4} = \left[-\left(1-\frac{3}{4}\right)^3\right] - \left[-(1-0)^3\right]$$

$$= -\left(\frac{1}{4}\right)^3 + 1 = -\frac{1}{64} + 1 = \frac{63}{64}.$$

Next: $P\left(Y_2 \le 1/2, Y_1 \le 3/4\right)$.
Again, draw the region:

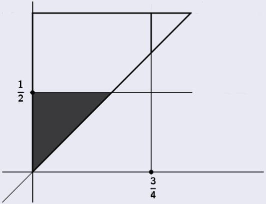


Solution: (continued)

$$= -(1-y_1)^3\Big|_0^{3/4} = \left[-\left(1-\frac{3}{4}\right)^3\right] - \left[-(1-0)^3\right]$$

$$= -\left(\frac{1}{4}\right)^3 + 1 = -\frac{1}{64} + 1 = \frac{63}{64}.$$

Next: $P\left(Y_2 \le 1/2, Y_1 \le 3/4\right)$.
Again, draw the region:
Notice for this region, the cutoff $Y_1 \le 3/4$ doesn't matter!

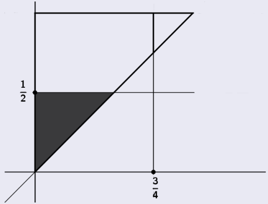


Solution: (continued)

$$= -(1-y_1)^3\Big|_0^{3/4} = \left[-\left(1-\frac{3}{4}\right)^3\right] - \left[-(1-0)^3\right]$$

$$= -\left(\frac{1}{4}\right)^3 + 1 = -\frac{1}{64} + 1 = \frac{63}{64}.$$

Next: $P\left(Y_2 \leq {}^1\!/_2, Y_1 \leq {}^3\!/_4\right)$.
Again, draw the region:
Notice for this region, the cutoff $Y_1 \leq {}^3\!/_4$ doesn't matter!



Remark: We might have missed this if we hadn't drawn the region: this is a common mistake.

Solution: (continued)

$$P\left(Y_2 \le \frac{1}{2}, Y_1 \le \frac{3}{4}\right) = \iint_{\text{shaded region}} 6(1-y_2)\, dy_1\, dy_2$$

$$= \int_{y_1=0}^{1/2} \int_{y_2=y_1}^{1/2} 6(1-y_2)\, dy_2\, dy_1$$

$$= \int_0^{1/2} \left(6y_2 - 3y_2^2 \Big|_{y_1}^{1/2} \right) dy_1$$

$$= \int_0^{1/2} \left[\left(6 \cdot \frac{1}{2} - 3 \cdot \left(\frac{1}{2}\right)^2 \right) - (6y_1 - 3y_1^2) \right] dy_1$$

$$= \int_0^{1/2} \left[\left(3 - \frac{3}{4} \right) - 6y_1 + 3y_1^2 \right] dy_1$$

$$= \int_0^{1/2} \left[\left(-\frac{3}{4} \right) + (3 - 6y_1 + 3y_1^2) \right] dy_1$$

Solution: (continued)

$$
\begin{aligned}
&= \int_0^{1/2} \left(-\frac{3}{4}\right) dy_1 + \int_0^{1/2} 3(1-y_1^2)dy_1 \\
&= 3 \cdot \left(-\frac{1}{3}\right)(1-y_1)^3\Big|_0^{1/2} + \frac{1}{2} \cdot \left(-\frac{3}{4}\right) \\
&= -\left(1-\frac{1}{2}\right)^3 - \left(-(1-0)^3\right) - \frac{3}{8} \\
&= -\frac{1}{8} - (-1) - \frac{3}{8} = \boxed{\frac{1}{2}}.
\end{aligned}
$$

To finish, divide this by the previous fraction.

Final answer: $\frac{32}{63}$.

Practice Problem 2:

Given that the joint density function of Y_1 and Y_2 is

$$f(y_1, y_2) = \begin{cases} \frac{y_1}{8} e^{-\frac{(y_1+y_2)}{2}} & y_1 > 0, y_2 > 0, \\ 0 & \text{elsewhere,} \end{cases}$$

find $E\left[\frac{Y_2}{Y_1}\right]$.

[Hint: Y_1, Y_2 are independent.]

Solution:

Since Y_1, Y_2 are independent, so are Y_2, $\frac{1}{Y_1}$.

So $E\left[\frac{Y_2}{Y_1}\right] = E[Y_2] \cdot E\left[\frac{1}{Y_1}\right]$.

To find $E[Y_2]$, we first find the marginal density function

$$f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1.$$

Then

$$E[Y_2] = \int_{-\infty}^{\infty} y_2 f_2(y_2)\, dy_2.$$

Solution: (continued)

We compute

$$f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1$$

Solution: (continued)

We compute

$$f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1$$
$$= \int_{0}^{\infty} \frac{y_1}{8} e^{-\frac{(y_1+y_2)}{2}}\, dy_1$$

Solution: (continued)

We compute

$$
\begin{aligned}
f_2(y_2) &= \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1 \\
&= \int_0^{\infty} \frac{y_1}{8} e^{-\frac{(y_1+y_2)}{2}}\, dy_1 = \frac{1}{8} e^{-\frac{y_2}{2}} \int_0^{\infty} y_1 e^{-\frac{y_1}{2}}\, dy_1
\end{aligned}
$$

Solution: (continued)

We compute

$$f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1$$

$$= \int_0^{\infty} \frac{y_1}{8} e^{-\frac{(y_1+y_2)}{2}}\, dy_1 = \frac{1}{8} e^{-\frac{y_2}{2}} \int_0^{\infty} y_1 e^{-\frac{y_1}{2}}\, dy_1$$

$$v = \frac{y_1}{2} \qquad 2v = y$$

Solution: (continued)

We compute

$$
\begin{aligned}
f_2(y_2) &= \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1 \\
&= \int_0^{\infty} \frac{y_1}{8} e^{-\frac{(y_1+y_2)}{2}}\, dy_1 = \frac{1}{8} e^{-\frac{y_2}{2}} \int_0^{\infty} y_1 e^{-\frac{y_1}{2}}\, dy_1
\end{aligned}
$$

$$
\begin{array}{rr}
v = \frac{y_1}{2} & 2v = y \\
dv = \frac{1}{2} dy_1 & 2dv = dy_1
\end{array}
$$

Solution: (continued)

We compute

$$
\begin{aligned}
f_2(y_2) &= \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1 \\
&= \int_0^{\infty} \frac{y_1}{8} e^{-\frac{(y_1+y_2)}{2}}\, dy_1 = \frac{1}{8} e^{-\frac{y_2}{2}} \int_0^{\infty} y_1 e^{-\frac{y_1}{2}}\, dy_1 \\
&= \frac{1}{8} e^{-\frac{y_2}{2}} \int_{v=0}^{\infty} 2v e^{-v}\, 2dv \qquad \begin{aligned} v &= \tfrac{y_1}{2} \\ dv &= \tfrac{1}{2} dy_1 \end{aligned} \quad \begin{aligned} 2v &= y \\ 2dv &= dy_1 \end{aligned}
\end{aligned}
$$

Solution: (continued)

We compute

$$
\begin{aligned}
f_2(y_2) &= \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1 \\
&= \int_0^{\infty} \frac{y_1}{8} e^{-\frac{(y_1+y_2)}{2}}\, dy_1 = \frac{1}{8} e^{-\frac{y_2}{2}} \int_0^{\infty} y_1 e^{-\frac{y_1}{2}}\, dy_1 \\
&= \frac{1}{8} e^{-\frac{y_2}{2}} \int_{v=0}^{\infty} 2v e^{-v}\, 2dv \qquad \begin{aligned} v &= \tfrac{y_1}{2} \\ dv &= \tfrac{1}{2} dy_1 \end{aligned} \quad \begin{aligned} 2v &= y \\ 2dv &= dy_1 \end{aligned} \\
&= \frac{1}{2} e^{-\frac{y_2}{2}} \int_{v=0}^{\infty} v e^{-v}\, dv
\end{aligned}
$$

Solution: (continued)

We compute

$$
\begin{aligned}
f_2(y_2) &= \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1 \\
&= \int_0^{\infty} \frac{y_1}{8} e^{-\frac{(y_1+y_2)}{2}}\, dy_1 = \frac{1}{8} e^{-\frac{y_2}{2}} \int_0^{\infty} y_1 e^{-\frac{y_1}{2}}\, dy_1 \\
&= \frac{1}{8} e^{-\frac{y_2}{2}} \int_{v=0}^{\infty} 2ve^{-v}\, 2dv \qquad \begin{array}{rl} v &= \frac{y_1}{2} \\ dv &= \frac{1}{2} dy_1 \end{array} \quad \begin{array}{rl} 2v &= y \\ 2dv &= dy_1 \end{array} \\
&= \frac{1}{2} e^{-\frac{y_2}{2}} \int_{v=0}^{\infty} ve^{-v}\, dv = \frac{1}{2} e^{-y_2/2} \left[-ve^{-v} \Big|_0^{\infty} - \int_0^{\infty} e^{-v}\, dv \right]
\end{aligned}
$$

Solution: (continued)

We compute

$$
\begin{aligned}
f_2(y_2) &= \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1 \\
&= \int_0^{\infty} \frac{y_1}{8} e^{-\frac{(y_1+y_2)}{2}}\, dy_1 = \frac{1}{8} e^{-\frac{y_2}{2}} \int_0^{\infty} y_1 e^{-\frac{y_1}{2}}\, dy_1 \\
&= \frac{1}{8} e^{-\frac{y_2}{2}} \int_{v=0}^{\infty} 2v e^{-v}\, 2dv \qquad \begin{array}{rl} v = \frac{y_1}{2} & \quad 2v = y \\ dv = \frac{1}{2} dy_1 & \quad 2dv = dy_1 \end{array} \\
&= \frac{1}{2} e^{-\frac{y_2}{2}} \int_{v=0}^{\infty} v e^{-v}\, dv = \frac{1}{2} e^{-y_2/2} \left[-v e^{-v}\Big|_0^{\infty} - \int_0^{\infty} e^{-v}\, dv \right] \\
&= \frac{1}{2} e^{-\frac{y_2}{2}}.
\end{aligned}
$$

Solution: (continued)

We compute

$$
\begin{aligned}
f_2(y_2) &= \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1 \\
&= \int_0^{\infty} \frac{y_1}{8} e^{-\frac{(y_1+y_2)}{2}}\, dy_1 = \frac{1}{8} e^{-\frac{y_2}{2}} \int_0^{\infty} y_1 e^{-\frac{y_1}{2}}\, dy_1 \\
&= \frac{1}{8} e^{-\frac{y_2}{2}} \int_{v=0}^{\infty} 2v e^{-v}\, 2dv \qquad \begin{array}{rl} v = & \frac{y_1}{2} \\ dv = & \frac{1}{2} dy_1 \end{array} \quad \begin{array}{rl} 2v = & y \\ 2dv = & dy_1 \end{array} \\
&= \frac{1}{2} e^{-\frac{y_2}{2}} \int_{v=0}^{\infty} v e^{-v}\, dv = \frac{1}{2} e^{-y_2/2} \left[-v e^{-v} \Big|_0^{\infty} - \int_0^{\infty} e^{-v}\, dv \right] \\
&= \frac{1}{2} e^{-\frac{y_2}{2}}. \qquad (\text{Note } f_2(y_2) = 0 \text{ for } y_2 < 0.)
\end{aligned}
$$

Solution: (continued)

So

$$E[Y_2] = \int_{-\infty}^{\infty} y_2 f_2(y_2)\, dy_2$$

Solution: (continued)

So

$$E[Y_2] = \int_{-\infty}^{\infty} y_2 f_2(y_2)\, dy_2 = \int_0^{\infty} \frac{1}{2} y_2 e^{-\frac{y_2}{2}}\, dy_2$$

Solution: (continued)

So

$$E[Y_2] = \int_{-\infty}^{\infty} y_2 f_2(y_2)\, dy_2 = \int_0^{\infty} \frac{1}{2} y_2 e^{-\frac{y_2}{2}}\, dy_2$$

$$= \frac{1}{2} \int_0^{\infty} y_2 e^{-\frac{y_2}{2}}\, dy_2$$

Solution: (continued)

So

$$E[Y_2] = \int_{-\infty}^{\infty} y_2 f_2(y_2)\, dy_2 = \int_0^{\infty} \frac{1}{2} y_2 e^{-\frac{y_2}{2}}\, dy_2$$

$$= \frac{1}{2} \underbrace{\int_0^{\infty} y_2 e^{-\frac{y_2}{2}}\, dy_2} \longleftarrow$$

This is exactly the integral we just did with y_1 replaced by y_2.

Solution: (continued)

So

$$E[Y_2] = \int_{-\infty}^{\infty} y_2 f_2(y_2)\, dy_2 = \int_0^{\infty} \frac{1}{2} y_2 e^{-\frac{y_2}{2}}\, dy_2$$

$$= \frac{1}{2} \underbrace{\int_0^{\infty} y_2 e^{-\frac{y_2}{2}}\, dy_2} \longleftarrow$$

This is exactly the integral we just did with y_1 replaced by y_2.

$$= \frac{1}{2} \cdot 4$$

Solution: (continued)

So

$$E[Y_2] = \int_{-\infty}^{\infty} y_2 f_2(y_2)\, dy_2 = \int_0^{\infty} \frac{1}{2} y_2 e^{-\frac{y_2}{2}}\, dy_2$$

$$= \frac{1}{2} \underbrace{\int_0^{\infty} y_2 e^{-\frac{y_2}{2}}\, dy_2} \quad \longleftarrow \quad \text{This is exactly the integral we just did with } y_1 \text{ replaced by } y_2.$$

$$= \frac{1}{2} \cdot 4 = \boxed{2}.$$

Solution: (continued)

So

$$E[Y_2] = \int_{-\infty}^{\infty} y_2 f_2(y_2)\, dy_2 = \int_0^{\infty} \frac{1}{2} y_2 e^{-\frac{y_2}{2}}\, dy_2$$

$$= \frac{1}{2} \underbrace{\int_0^{\infty} y_2 e^{-\frac{y_2}{2}}\, dy_2} \quad \longleftarrow$$

This is exactly the integral we just did with y_1 replaced by y_2.

$$= \frac{1}{2} \cdot 4 = \boxed{2}.$$

Similarly, compute $E\left[\frac{1}{Y_1}\right]$.

Solution: (continued)

So

$$E[Y_2] = \int_{-\infty}^{\infty} y_2 f_2(y_2)\, dy_2 = \int_0^{\infty} \frac{1}{2} y_2 e^{-\frac{y_2}{2}}\, dy_2$$

$$= \frac{1}{2} \underbrace{\int_0^{\infty} y_2 e^{-\frac{y_2}{2}}\, dy_2} \quad \longleftarrow$$

This is exactly the integral we just did with y_1 replaced by y_2.

$$= \frac{1}{2} \cdot 4 = \boxed{2}.$$

Similarly, compute $E\left[\frac{1}{Y_1}\right]$.

Now multiply these to get the final answer:

Solution: (continued)

So

$$E[Y_2] = \int_{-\infty}^{\infty} y_2 f_2(y_2)\, dy_2 = \int_0^{\infty} \frac{1}{2} y_2 e^{-\frac{y_2}{2}}\, dy_2$$

$$= \frac{1}{2} \underbrace{\int_0^{\infty} y_2 e^{-\frac{y_2}{2}}\, dy_2} \qquad \longleftarrow \quad \text{This is exactly the integral we just did with } y_1 \text{ replaced by } y_2.$$

$$= \frac{1}{2} \cdot 4 = \boxed{2}.$$

Similarly, compute $E\left[\frac{1}{Y_1}\right]$. $\left(= \frac{1}{2}\right)$.

Now multiply these to get the final answer:

Solution: (continued)

So

$$E[Y_2] = \int_{-\infty}^{\infty} y_2 f_2(y_2)\,dy_2 = \int_0^{\infty} \frac{1}{2} y_2 e^{-\frac{y_2}{2}}\,dy_2$$

$$= \frac{1}{2}\underbrace{\int_0^{\infty} y_2 e^{-\frac{y_2}{2}}\,dy_2} \quad \longleftarrow$$

This is exactly the integral we just did with y_1 replaced by y_2.

$$= \frac{1}{2}\cdot 4 = \boxed{2}.$$

Similarly, compute $E\left[\frac{1}{Y_1}\right]$. $\left(= \frac{1}{2}\right)$.

Now multiply these to get the final answer: $E\left[\frac{Y_2}{Y_1}\right] = 1$.

Exercise 5.27: (Relevant for MATH 448)

Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with $E[Y_i] = \mu$ and $V[Y_i] = \sigma^2$. Consider the new RV $\overline{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$. What are the mean and the variance of $\overline{Y}$?

Solution:

$$E\left[\overline{Y}\right] = E\left[\frac{1}{n}\sum_{i=1}^n Y_i\right] = \frac{1}{n}\sum_{i=1}^n E[Y_i] = \frac{1}{n}\sum_{i=1}^n \mu = \mu,$$

$$\begin{aligned}
V\left[\overline{Y}\right] &= V\left[\frac{1}{n}\sum_{i=1}^n Y_i\right] = \mathrm{Cov}\left(\frac{1}{n}\sum_{i=1}^n Y_i, \frac{1}{n}\sum_{i=1}^n Y_i\right) \\
&= \frac{1}{n}\mathrm{Cov}\left(\sum_{i=1}^n Y_i, \frac{1}{n}\sum_{i=1}^n Y_i\right) = \frac{1}{n^2}\mathrm{Cov}\left(\sum_{i=1}^n Y_i, \sum_{i=1}^n Y_i\right) \\
&= \frac{1}{n^2}\left[\sum_{i=1}^n \mathrm{Cov}\left(Y_i, \sum_{j=1}^n Y_j\right)\right] = \frac{1}{n^2}\left[\sum_{i=1}^n\sum_{j=1}^n \mathrm{Cov}\left(Y_i, Y_j\right)\right].
\end{aligned}$$

Solution: (continued)

But Y_i and Y_j are independent if $i \neq j$.

Solution: (continued)

But Y_i and Y_j are independent if $i \neq j$. So $\mathrm{Cov}(Y_i, Y_j) = 0$ if $i \neq j$.

Solution: (continued)

But Y_i and Y_j are independent if $i \neq j$. So $\mathrm{Cov}(Y_i, Y_j) = 0$ if $i \neq j$. Thus

$$V[\overline{Y}] = \frac{1}{n^2}\left[\sum_{i=1}^{n} \mathrm{Cov}\left(Y_i, Y_i\right) + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \mathrm{Cov}\left(Y_i, Y_j\right)\right]$$

Solution: (continued)

But Y_i and Y_j are independent if $i \neq j$. So $\mathrm{Cov}(Y_i, Y_j) = 0$ if $i \neq j$. Thus

$$V[\overline{Y}] = \frac{1}{n^2}\left[\sum_{i=1}^{n} \mathrm{Cov}\left(Y_i, Y_i\right) + \cancelto{0}{\sum_{\substack{i,j=1 \\ i \neq j}}^{n} \mathrm{Cov}\left(Y_i, Y_j\right)}\right]$$

$$= \frac{1}{n^2} \cdot n \cdot \sigma^2$$

Solution: (continued)

But Y_i and Y_j are independent if $i \neq j$. So $\mathrm{Cov}(Y_i, Y_j) = 0$ if $i \neq j$. Thus

$$V[\overline{Y}] = \frac{1}{n^2}\left[\sum_{i=1}^{n} \mathrm{Cov}\left(Y_i, Y_i\right) + \underbrace{\sum_{\substack{i,j=1 \\ i\neq j}}^{n} \cancel{\mathrm{Cov}\left(Y_i, Y_j\right)}}_{0}\right]$$

$$= \frac{1}{n^2} \cdot n \cdot \sigma^2 = \boxed{\frac{\sigma^2}{n}}.$$

Solution: (continued)

But Y_i and Y_j are independent if $i \neq j$. So $\operatorname{Cov}(Y_i, Y_j) = 0$ if $i \neq j$. Thus

$$V[\overline{Y}] = \frac{1}{n^2}\left[\sum_{i=1}^{n} \operatorname{Cov}(Y_i, Y_i) + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \cancelto{0}{\operatorname{Cov}(Y_i, Y_j)}\right]$$

$$= \frac{1}{n^2} \cdot n \cdot \sigma^2 = \boxed{\frac{\sigma^2}{n}}.$$

Shortcut: If random variables X, Y are independent, then $V[X + Y] = V[X] + V[Y]$.

Solution: (continued)

But Y_i and Y_j are independent if $i \neq j$. So $\text{Cov}(Y_i, Y_j) = 0$ if $i \neq j$. Thus

$$V[\overline{Y}] = \frac{1}{n^2}\left[\sum_{i=1}^{n} \text{Cov}(Y_i, Y_i) + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \underbrace{\text{Cov}(Y_i, Y_j)}_{0}\right]$$

$$= \frac{1}{n^2} \cdot n \cdot \sigma^2 = \boxed{\frac{\sigma^2}{n}}.$$

Shortcut: If random variables X, Y are independent, then $V[X+Y] = V[X] + V[Y]$.

Warning: It does NOT follow that $V[X-Y] = V[X] - V[Y]$.

Solution: (continued)

But Y_i and Y_j are independent if $i \neq j$. So $\text{Cov}(Y_i, Y_j) = 0$ if $i \neq j$. Thus

$$V[\overline{Y}] = \frac{1}{n^2}\left[\sum_{i=1}^{n} \text{Cov}\,(Y_i, Y_i) + \cancelto{0}{\sum_{\substack{i,j=1 \\ i \neq j}}^{n} \text{Cov}\,(Y_i, Y_j)}\right]$$

$$= \frac{1}{n^2} \cdot n \cdot \sigma^2 = \boxed{\frac{\sigma^2}{n}}.$$

<u>Shortcut:</u> If random variables X, Y are independent, then $V[X+Y] = V[X] + V[Y]$.

<u>Warning:</u> It does NOT follow that $V[X-Y] = V[X] - V[Y]$. In fact

$$V[X-Y] = V[X + (-1)Y]$$

Solution: (continued)

But Y_i and Y_j are independent if $i \neq j$. So $\mathrm{Cov}(Y_i, Y_j) = 0$ if $i \neq j$. Thus

$$V[\overline{Y}] = \frac{1}{n^2}\left[\sum_{i=1}^{n} \mathrm{Cov}\,(Y_i, Y_i) + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \cancelto{0}{\mathrm{Cov}\,(Y_i, Y_j)}\right]$$

$$= \frac{1}{n^2} \cdot n \cdot \sigma^2 = \boxed{\frac{\sigma^2}{n}}.$$

<u>Shortcut:</u> If random variables X, Y are independent, then $V[X + Y] = V[X] + V[Y]$.

<u>Warning:</u> It does NOT follow that $V[X - Y] = V[X] - V[Y]$. In fact

$$V[X - Y] = V[X + (-1)Y] = V[X] + V[(-1)Y]$$

Solution: (continued)

But Y_i and Y_j are independent if $i \neq j$. So $\mathrm{Cov}(Y_i, Y_j) = 0$ if $i \neq j$. Thus

$$V[\overline{Y}] = \frac{1}{n^2}\left[\sum_{i=1}^{n} \mathrm{Cov}(Y_i, Y_i) + \underbrace{\sum_{\substack{i,j=1\\ i\neq j}}^{n} \mathrm{Cov}(Y_i, Y_j)}_{0}\right]$$

$$= \frac{1}{n^2} \cdot n \cdot \sigma^2 = \boxed{\frac{\sigma^2}{n}}.$$

<u>Shortcut:</u> If random variables X, Y are independent, then $V[X+Y] = V[X] + V[Y]$.

<u>Warning:</u> It does NOT follow that $V[X-Y] = V[X] - V[Y]$. In fact

$$V[X-Y] = V[X+(-1)Y] = V[X] + V[(-1)Y]$$
$$= V[X] + (-1)^2 V[Y]$$

Solution: (continued)

But Y_i and Y_j are independent if $i \neq j$. So $\mathrm{Cov}(Y_i, Y_j) = 0$ if $i \neq j$. Thus

$$V[\overline{Y}] = \frac{1}{n^2}\left[\sum_{i=1}^{n} \mathrm{Cov}\,(Y_i, Y_i) + \underbrace{\sum_{\substack{i,j=1 \\ i \neq j}}^{n} \mathrm{Cov}\,(Y_i, Y_j)}_{0}\right]$$

$$= \frac{1}{n^2} \cdot n \cdot \sigma^2 = \boxed{\frac{\sigma^2}{n}}.$$

Shortcut: If random variables X, Y are independent, then $V[X+Y] = V[X] + V[Y]$.

Warning: It does NOT follow that $V[X-Y] = V[X] - V[Y]$. In fact

$$V[X-Y] = V[X + (-1)Y] = V[X] + V[(-1)Y]$$

$$= V[X] + (-1)^2 V[Y] = V[X] + V[Y].$$

Solution: (continued)

But Y_i and Y_j are independent if $i \neq j$. So $\operatorname{Cov}(Y_i, Y_j) = 0$ if $i \neq j$. Thus

$$V[\overline{Y}] = \frac{1}{n^2}\left[\sum_{i=1}^{n} \operatorname{Cov}(Y_i, Y_i) + \cancelto{0}{\sum_{\substack{i,j=1 \\ i\neq j}}^{n} \operatorname{Cov}(Y_i, Y_j)}\right]$$

$$= \frac{1}{n^2} \cdot n \cdot \sigma^2 = \boxed{\frac{\sigma^2}{n}}.$$

Shortcut: If random variables X, Y are independent, then $V[X + Y] = V[X] + V[Y]$.

Warning: It does NOT follow that $V[X - Y] = V[X] - V[Y]$. In fact

$$V[X - Y] = V[X + (-1)Y] = V[X] + V[(-1)Y]$$
$$= V[X] + (-1)^2 V[Y] = V[X] + V[Y].$$

Note that in the correct version of this computation, we used $V[aY] = a^2 V[Y]$, and that if X, Y are independent, then X and $-Y$ are independent.

Example 5.29:

Suppose that an urn contains r red balls and $N - r$ black balls. A random sample of n balls is drawn without replacement and Y, the number of red balls in the sample, is observed. Find the mean and variance of Y.

[Hint: From Chapter 3 we know that Y has a hypergeometric probability distribution.]

Solution:

We have learnt that for a hypergeometric distribution Y,

$$E[Y] = \frac{nr}{N}, \qquad V[Y] = \frac{nr}{N} \cdot \frac{N-r}{N} \cdot \frac{N-n}{N-1}.$$

Now we will prove this:

Let

$$X_i = \begin{cases} 1 & \text{if } i^{\text{th}} \text{ ball in the sample is red,} \\ 0 & \qquad\qquad \text{if it is black.} \end{cases}$$

Let $Y = X_1 + \cdots + X_n$.

Consider each X_i separately.

$P(X_i = 1) = \frac{r}{N}$, so $E[X_i] = \frac{r}{N}$.

By linearity of Expectation, we find

Solution: (continued)

$$E[Y] = E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n] = \frac{nr}{N}.$$

Note that the X_i are dependent on one another and we can use linearity of E anyway.

Let's consider the dependence more carefully:

$$P(X_2 = 1, X_1 = 1) = P(X_2 = 1 \mid X_1 = 1) \cdot P(X_1 = 1) = \frac{r-1}{N-1} \cdot \frac{r}{N}.$$

More generally, $P(X_j = 1, X_i = 1) = \dfrac{(r-1)r}{(N-1)N}$.

Since $X_i = 0$ or 1, $E[X_i X_j] = \dfrac{(r-1)r}{(N-1)N}$.

We can now start thinking about

$$V[Y] = \mathrm{Cov}(Y, Y) = \mathrm{Cov}(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j).$$

Solution: (continued)

If $i \neq j$, then

$$\mathrm{Cov}(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j] = \frac{(r-1)r}{(N-1)N} - \frac{r}{N} \cdot \frac{r}{N}.$$

If $i = j$, then

$$\mathrm{Cov}(X_i, X_i) = E[X_i^2] - E[X_i]^2.$$

But $Y_i = 0$ or 1.
So $X_i^2 = X_i$.
Thus $E[X_i^2] = \dfrac{r}{N}$,
and $\mathrm{Cov}(X_i, X_i) = \dfrac{r}{N} - \left(\dfrac{r}{N}\right)^2$.
Therefore

$$\begin{aligned} V[Y] &= n \cdot \left[\frac{r}{N} - \left(\frac{r}{N}\right)^2\right] + n(n-1) \cdot \left[\frac{(r-1)r}{(N-1)N} - \left(\frac{r}{N}\right)^2\right] \\ &= \frac{nr}{N} \cdot \left(\left[1 - \frac{r}{N}\right] + (n-1)\left[\frac{r-1}{N-1} - \frac{r}{N}\right]\right) \\ &= \frac{nr}{N} \cdot \left(\frac{N-r}{N} + \frac{(n-1)(r-1)}{N-1} - \frac{(n-1)r}{N}\right) \end{aligned}$$

Solution: (continued)

$$= \frac{nr}{N} \cdot \left(\frac{N - r - (n-1)r}{N} + \frac{(n-1)(r-1)}{N-1} \right)$$

Solution: (continued)

$$= \frac{nr}{N} \cdot \left(\frac{N - r - (n-1)r}{N} + \frac{(n-1)(r-1)}{N-1} \right)$$

$$= \frac{nr}{N} \cdot \left(\frac{N - r - nr + r}{N} + \frac{N}{N} \frac{(n-1)(r-1)}{N-1} \right)$$

Solution: (continued)

$$= \frac{nr}{N} \cdot \left(\frac{N - r - (n-1)r}{N} + \frac{(n-1)(r-1)}{N-1} \right)$$

$$= \frac{nr}{N} \cdot \left(\frac{N - \cancel{r} - nr + \cancel{r}}{N} + \frac{N}{N} \frac{(n-1)(r-1)}{N-1} \right)$$

$$= \frac{nr}{N} \cdot \left(\frac{N - nr}{N} \frac{N-1}{N-1} + \frac{N(n-1)(r-1)}{N(N-1)} \right)$$

Solution: (continued)

$$= \frac{nr}{N} \cdot \left(\frac{N - r - (n-1)r}{N} + \frac{(n-1)(r-1)}{N-1} \right)$$

$$= \frac{nr}{N} \cdot \left(\frac{N - \cancel{r} - nr + \cancel{r}}{N} + \frac{N}{N}\frac{(n-1)(r-1)}{N-1} \right)$$

$$= \frac{nr}{N} \cdot \left(\frac{N - nr}{N}\frac{N-1}{N-1} + \frac{N(n-1)(r-1)}{N(N-1)} \right)$$

$$= \frac{nr}{N} \cdot \left(\frac{(N-nr)(N-1) + N(n-1)(r-1)}{N(N-1)} \right)$$

Solution: (continued)

$$= \frac{nr}{N} \cdot \left(\frac{N - r - (n-1)r}{N} + \frac{(n-1)(r-1)}{N-1} \right)$$

$$= \frac{nr}{N} \cdot \left(\frac{N - \cancel{r} - nr + \cancel{r}}{N} + \frac{N}{N}\frac{(n-1)(r-1)}{N-1} \right)$$

$$= \frac{nr}{N} \cdot \left(\frac{N - nr}{N}\frac{N-1}{N-1} + \frac{N(n-1)(r-1)}{N(N-1)} \right)$$

$$= \frac{nr}{N} \cdot \left(\frac{(N - nr)(N-1) + N(n-1)(r-1)}{N(N-1)} \right)$$

$$= \frac{nr}{N} \cdot \left(\frac{(N^2 - Nnr - N + nr) + (Nnr - Nn - Nr + N)}{N(N-1)} \right)$$

Solution: (continued)

$$
\begin{aligned}
&= \frac{nr}{N} \cdot \left(\frac{N - r - (n-1)r}{N} + \frac{(n-1)(r-1)}{N-1} \right) \\
&= \frac{nr}{N} \cdot \left(\frac{N - \cancel{r} - nr + \cancel{r}}{N} + \frac{N}{N} \frac{(n-1)(r-1)}{N-1} \right) \\
&= \frac{nr}{N} \cdot \left(\frac{N - nr}{N} \frac{N-1}{N-1} + \frac{N(n-1)(r-1)}{N(N-1)} \right) \\
&= \frac{nr}{N} \cdot \left(\frac{(N - nr)(N-1) + N(n-1)(r-1)}{N(N-1)} \right) \\
&= \frac{nr}{N} \cdot \left(\frac{(N^2 - \cancel{Nnr} - \cancel{N} + nr) + (\cancel{Nnr} - Nn - Nr + \cancel{N})}{N(N-1)} \right) \\
&= \frac{nr}{N} \left[\frac{(N-r)}{N} \frac{(N-n)}{(N-1)} \right], \qquad \text{as claimed.}
\end{aligned}
$$

Definition (Multinomial Experiment)

A multinomial experiment possesses the following properties:

(1) The experiment consists of n identical trials.

(2) The outcome of each trial falls into one of k classes or cells.

(3) The probability that the outcome of a single trial falls into cell i, is p_i, $i = 1, \ldots, k$ and remains the same from trial to trial. Notice that $p_1 + \cdots + p_k = 1$.

(4) The trials are independent.

(5) The random variables of interest are $Y_1, \ldots, Y_k$, where Y_i equals the number of trials for which the outcome falls into cell i. Notice that $Y_1 + \cdots + Y_k = n$.

Multinomial experiment is like a binomial experiment, but there are k possible outcomes, not just 2.

These RVs $Y_1, \ldots, Y_k$ are said to have a multinomial distribution. Formally,

Definition (Multinomial Distribution)

Assume that $p_1, \ldots, p_k$ are such that $\sum_{i=1}^{n} p_i = 1$, and $p_i > 0$ for $i = 1, \ldots, k$. The random variables $Y_1, \ldots, Y_k$ are said to have a multinomial distribution with parameters n and $p_1, \ldots, p_k$ if the joint probability function of $Y_1, \ldots, Y_k$ is given by

$$p(y_1, \ldots, y_k) = \frac{n!}{y_1! \ldots y_k!} p_1^{y_1} \ldots p_k^{y_k},$$

where, for each i, $y_i = 0, 1, \ldots, n$ and $\sum_{i=1}^{k} y_i = n$.

By thinking of outcome type i as success, and anything else as failure, we see that the marginal distribution of each Y_i is binomial with parameters n (the number of trials) and p_i (the probability of outcome type i).

Theorem (5.13)

If $Y_1, \dots, Y_k$ have a multinomial distribution with parameters n and $p_1, \dots, p_k$, then

(1) $E[Y_i] = np_i, V[Y_i] = np_i q_i$, *where* $q_i = 1 - p_i$.

(2) $Cov(Y_s, Y_t) = -np_s p_t$, *if* $s \neq t$.

Remark:

The hard part of Theorem 5.13 is the statement (2).
Notice that this covariance is negative; this is intuitive from
$Y_1 + \cdots + Y_k = n$.
(Since the sum is constant, if one Y_i is large the others are more likely to be small.)

Proof: (Part (2) of Theorem 5.13)

Define

$$U_i = \begin{cases} 1 & \text{if trial } i \text{ results in outcome } s, \\ 0 & \text{otherwise} \end{cases}, \quad V_j = \begin{cases} 1 & \text{if trial } j \text{ results in outcome } t, \\ 0 & \text{otherwise.} \end{cases}$$

Proof: (continued)

Then

$$Y_s = \sum_{i=1}^{n} U_i,\ Y_t = \sum_{j=1}^{n} V_j, \quad \mathrm{Cov}(Y_s, Y_t) = \mathrm{Cov}\left(\sum_{i=1}^{n} U_i, \sum_{j=1}^{n} V_j\right).$$

Now we can use the bilinearity of covariance.
Notice also that if $i \neq j$, then trial i is independent of trial j, by definition of multinomial experiment.
So U_i is independent of V_j if $i \neq j$.
So $\mathrm{Cov}(U_i, V_j) = 0$,
and

$$\mathrm{Cov}(Y_s, Y_t) = \sum_{i=1}^{n} \mathrm{Cov}(U_i, V_i) + \sum_{i \neq j} \cancelto{0}{\mathrm{Cov}(U_i, V_j)}.$$

Now $\mathrm{Cov}(U_i, V_i) = E[U_i V_i] - E[U_i]E[V_i]$.
Since U_i and V_i cannot both be 1, $E[U_i V_i] = 0$.
Also $E[U_i] = p_s, E[V_i] = p_t$
implies $\mathrm{Cov}(U_i, V_j) = -p_s p_t$.
Thus

$$\mathrm{Cov}(Y_s, Y_t) = \sum^{n} \mathrm{Cov}(U_i, V_j) = -np_s p_t,$$

Exercise 5.119

A learning experiment requires a rat to run a maze (a network of pathways) until it locates one of three possible exits. Exit 1 presents a reward of food, but exits 2 and 3 do not. (If the rat eventually selects exit 1 almost every time, learning may have taken place.) Let Y_i denote the number of times exit i is chosen in successive runnings. For the following, assume that the rat chooses an exit at random on each run.

(a) Find the probability that $n = 6$ runs result in $Y_1 = 3$, $Y_2 = 1$, and $Y_3 = 2$.

(b) For general n, find $E[Y_1]$ and $V[Y_1]$.

(c) Find $\mathrm{Cov}(Y_2, Y_3)$ for general n.

(d) To check for the rat's preference between exits 2 and 3, we may look at $Y_2 - Y_3$. Find $E[Y_2 - Y_3]$ and $V[Y_2 - Y_3]$ for general n.

Hints and Answers:

Note that $k = 3$ and $p_1 = p_2 = p_3 = \frac{1}{3}$.

- For part (a), apply the joint probability function of the multinomial distribution:

  Answer: $\binom{6}{3\,1\,2} \cdot p_1^3 p_2^1 p_3^3 = \frac{6!}{3!1!2!}\left(\frac{1}{3}\right)^3\left(\frac{1}{3}\right)\left(\frac{1}{3}\right)^2 = \frac{20}{243}$.

- For part (b), apply our knowledge of binomial distributions:

$$E[Y_1] = \frac{n}{3}, \qquad V[Y_1] = \frac{2n}{9}.$$

- For part (c), apply Theorem 5.13:
  $\mathrm{Cov}(Y_2, Y_3) = -\frac{n}{9}$.

- For part (d),

$$E[Y_2 - Y_3] = E[Y_2] - E[Y_3] = \frac{n}{3} - \frac{n}{3} = 0.$$

Correlation

There is one issue with covariance as a measure of how much X and Y 'vary together":
It is larger when X and Y are larger, even if the connection between X and Y is not very strong.

For example, if $\mathrm{Cov}(X, Y) = 1$, then $\mathrm{Cov}(2X, 2Y) = 4$, by bilinearity of covariance.
But the connection between $2X$ and $2Y$ is no better than the connection between X and Y.

One way to measure this (undefined) "connection" is <u>correlation</u>.
For two RVs X, Y with some joint distribution,

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}, \qquad \text{where } \sigma_X = \sqrt{V[X]}, \sigma_Y = \sqrt{V[Y]}.$$

For the example above with $2X$ and $2Y$, we have

$$\rho_{2X,2Y} = \frac{\mathrm{Cov}(2X, 2Y)}{\sigma_2 X \sigma_2 Y} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} = \rho_{X,Y}.$$

Remark:

Think of this as a covariance "normalized" for the size of X and Y.

$$\text{FACT: } -1 \le \rho_{X,Y} \le 1.$$

To see this, we use the Cauchy-Schwarz inequality adapted to random variables.

Cauchy-Schwarz Inequality:

$$|\langle v, w\rangle| \le \|v\| \cdot \|w\|.$$

You may recall that $\langle v, w\rangle = \|v\| \cdot \|w\| \cdot \cos\theta$, where θ is the angle between v and w.
If you are willing to assume this fact, then the Cauchy-Schwarz inequality follows from $|\cos\theta| \le 1$.

If we translate this to the language of random variables by saying that the inner product of two RVs X, Y is $\mathrm{Cov}(X, Y)$
– notice that $\langle\cdot,\cdot\rangle$ and $\mathrm{Cov}(\cdot,\cdot)$ are both bilinear.

The length or norm $\|v\|$ translates to σ_X, i.e. $\sqrt{V[X]}$.
Then the Cauchy-Schwarz inequality translates to $|\mathrm{Cov}(X, Y)| \leq \sigma_X \sigma_Y$.
So

$$|\rho_{X,Y}| = \frac{|\mathrm{Cov}(X,Y)|}{\sigma_X \sigma_Y} \leq 1.$$

Remark:

If the correlation is exactly 1 or -1, this implies a perfect linear relationship between X and Y, i.e. $Y = aX + b$ with probability 1.

Remark: "Correlation does not imply causation"

There is no value of the correlation that implies a causal connection between X and Y.
There might be, for example, some common cause of X and Y that explains the correlation.

Remark:

If you think of X and Y as being like vectors, you can think of $\rho_{X,Y}$ as being like the cosine of the angle between them.

Example

Suppose $\rho_{X,Y} = 0.9$ and $\rho_{Y,Z} = 0.8$. What is the minimum possible value of $\rho_{X,Z}$?

Solution:

Suppose that we write $\theta_{X,Y}$ and $\theta_{Y,Z}$ for the angles between the RVs X, Y and Y, Z.
Then $\theta_{X,Y} = \cos^{-1}(0.9)$ and $\theta_{Y,Z} = \cos^{-1}(0.8)$.
(Note that the principal range of $\cos^{-1}(t)$ is $[0, \pi]$.)

The maximum possible angle between X and Z is less than or equal to $\theta_{X,Y} + \theta_{Y,Z}$, that is, $\theta_{X,Z} \leq \theta_{X,Y} + \theta_{Y,Z}$.
So

$$\begin{aligned}\rho_{X,Z} = \cos\theta_{X,Z} &\leq \cos(\theta_{X,Y} + \theta_{Y,Z}) \\ &\leq \cos(\cos^{-1}(0.9) + \cos^{-1}(0.8)) \leq 0.458.\end{aligned}$$

Example (Properties of Correlation)

Suppose $\rho_{X,Y} = 0.2$ and $Z = 2Y + 3$. What is $\rho_{X,Z}$?

Solution:

From the definitions, we calculate

$$\begin{aligned}\rho_{X,Z} &= \frac{\text{Cov}(X,Z)}{\sigma_X \sigma_Z} = \frac{\text{Cov}(X, 2Y+3)}{\sigma_X \sigma_Z} \\ &= \frac{2\text{Cov}(X,Y)}{\sigma_X \sigma_Z} \qquad \text{but } \sigma_Z = \sqrt{V[2Y+3]} = 2\sqrt{V[Y]} \\ &= \frac{2\text{Cov}(X,Y)}{\sigma_X \cdot 2\sigma_Y} = \rho_{X,Y}.\end{aligned}$$

What if $Z = -3Y + 4$ instead?

Same calculation shows that $\rho_{X,Z} = -\rho_{X,Y}$.

This means that a linear change of variable can only change the sign of the correlation and not the magnitude.

The Bivariate Normal Distribution

No discussion of multivariate probability distributions would be complete without reference to the multivariate normal distribution, which is a keystone of much modern statistical theory.
We look at the simplest case, the Bivariate Normal Distribution:

Definition (Bivariate Normal Distribution)

Two continuous RVs Y_1, Y_2 are said to have the bivariate normal distribution if the density function is given by

$$f(y_1, y_2) = \frac{e^{-Q/2}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}, \quad -\infty < y_1 < \infty, -\infty < y_2 < \infty,$$

where

$$Q = \frac{1}{1-\rho^2}\left[\frac{(y_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2}\right].$$

Thus the bivariate normal distribution is a function of five parameters: $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$, and ρ.
The choice of notation employed for these parameters is not coincidental:

Exercise 5.128

The marginal distributions of Y_1 and Y_2 are normal distributions with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively.

Remarks:

- With a bit of somewhat tedious integration, we can also show that
$$\rho = \frac{\mathrm{Cov}(Y_1, Y_2)}{\sigma_1 \sigma_2} = \rho_{Y_1, Y_2},$$
the correlation coefficient between Y_1 and Y_2.
- This distribution is special, in the sense that, if Y_1 and Y_2 have a bivariate normal distribution, they are independent if and only if their covariance (equivalently, $\rho = \rho_{Y_1, Y_2}$) is zero.
Zero covariance does not imply independence in general.

The expression for the joint density function of the k-variate normal distribution ($k > 2$) is most easily expressed by using the matrix algebra.

Definition (The k-variate Normal Distribution ($k \geq 2$))

Let $\mathbf{Y} = (Y_1, \ldots, Y_k)$ denote a k-dimensional *random vector* (i.e. $Y_1, \ldots, Y_k$ are k random variables).
Also let $\boldsymbol{\mu} = E[\mathbf{Y}] = (E[Y_1], \ldots, E[Y_k])$ denote the k-dimensional *mean vector*,
and
$\mathbf{\Sigma} := E[(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})^T] = [[\mathrm{Cov}(Y_i, Y_j)]]_{1 \leq i,j \leq k}$ the $k \times k$ *Covariance Matrix*.
Then $Y_1, \ldots, Y_k$ have the k-variate normal distribution if their joint density function is

$$f_{\mathbf{Y}}(y_1, \ldots, y_k) = \frac{1}{(2\pi)^{k/2}\sqrt{\det \mathbf{\Sigma}}} e^{\left(-\frac{1}{2}(\mathbf{Y}-\boldsymbol{\mu})^T \mathbf{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})\right)}$$

$$= \frac{1}{\sqrt{\det(2\pi\mathbf{\Sigma})}} e^{\left(-\frac{1}{2}(\mathbf{Y}-\boldsymbol{\mu})^T \mathbf{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})\right)}.$$

Check that $k = 2$ gives the bivariate normal distribution we have just seen.

Conditional Expectation

$$E[Y_1 \mid Y_2 = y_2] \qquad \text{or} \qquad E[g(Y_1) \mid Y_2 = y_2].$$

This is defined by integrating with respect to the conditional density defined earlier:

$$E[g(Y_1) \mid Y_2 = y_2] = \int_{-\infty}^{\infty} g(y_1) f(y_1 \mid y_2)\, dy_1.$$

Notice that this is a function of y_2.

Recall that $f(y_1 \mid y_2) = \dfrac{f(y_1, y_2)}{f_2(y_2)}$

Assuming that everything is defined and we haven't divided by zero, we could compute the expectation of $E[Y_1 \mid Y_2 = y_2]$, because it is a function of Y_2.

Theorem (5.14)

Let Y_1 *and* Y_2 *denote random variables. Then*

$$E[Y_1] = E[E[Y_1 \mid Y_2]],$$

where on the right-hand side the inside expectation is with respect to the conditional distribution of Y_1 *given* Y_2 *and the outside expectation is with respect to the distribution of* Y_2*.*

Theorem (5.14)

Let Y_1 and Y_2 denote random variables. Then

$$E[Y_1] = E[E[Y_1 \mid Y_2]],$$

where on the right-hand side the inside expectation is with respect to the conditional distribution of Y_1 given Y_2 and the outside expectation is with respect to the distribution of Y_2.

Proof: (continuous case; the discrete case is analogous.)

Theorem (5.14)

Let Y_1 and Y_2 denote random variables. Then

$$E[Y_1] = E[E[Y_1 \mid Y_2]],$$

where on the right-hand side the inside expectation is with respect to the conditional distribution of Y_1 given Y_2 and the outside expectation is with respect to the distribution of Y_2.

Proof: (continuous case; the discrete case is analogous.)

$$E[E[Y_1 \mid Y_2]] = \int_{-\infty}^{\infty} E[Y_1 \mid Y_2] f_2(y_2)\, dy_2$$

Theorem (5.14)

Let Y_1 and Y_2 denote random variables. Then

$$E[Y_1] = E[E[Y_1 \mid Y_2]],$$

where on the right-hand side the inside expectation is with respect to the conditional distribution of Y_1 given Y_2 and the outside expectation is with respect to the distribution of Y_2.

Proof: (continuous case; the discrete case is analogous.)

$$\begin{aligned} E[E[Y_1 \mid Y_2]] &= \int_{-\infty}^{\infty} E[Y_1 \mid Y_2] f_2(y_2)\, dy_2 \\ &= \int_{y_2=-\infty}^{\infty} \left[\int_{y_1=-\infty}^{\infty} y_1 f(y_1 \mid y_2)\, dy_1 \right] f_2(y_2)\, dy_2 \end{aligned}$$

Theorem (5.14)

Let Y_1 and Y_2 denote random variables. Then
$$E[Y_1] = E[E[Y_1 \mid Y_2]],$$
where on the right-hand side the inside expectation is with respect to the conditional distribution of Y_1 given Y_2 and the outside expectation is with respect to the distribution of Y_2.

Proof: (continuous case; the discrete case is analogous.)

$$\begin{aligned}
E[E[Y_1 \mid Y_2]] &= \int_{-\infty}^{\infty} E[Y_1 \mid Y_2] f_2(y_2)\, dy_2 \\
&= \int_{y_2=-\infty}^{\infty} \left[\int_{y_1=-\infty}^{\infty} y_1 f(y_1 \mid y_2)\, dy_1 \right] f_2(y_2)\, dy_2 \\
&= \int_{y_2=-\infty}^{\infty} \int_{y_1=-\infty}^{\infty} y_1 \frac{f(y_1, y_2)}{f_2(y_2)} f_2(y_2)\, dy_1\, dy_2
\end{aligned}$$

Theorem (5.14)

Let Y_1 and Y_2 denote random variables. Then

$$E[Y_1] = E[E[Y_1 \mid Y_2]],$$

where on the right-hand side the inside expectation is with respect to the conditional distribution of Y_1 given Y_2 and the outside expectation is with respect to the distribution of Y_2.

Proof: (continuous case; the discrete case is analogous.)

$$\begin{aligned}
E[E[Y_1 \mid Y_2]] &= \int_{-\infty}^{\infty} E[Y_1 \mid Y_2] f_2(y_2)\, dy_2 \\
&= \int_{y_2=-\infty}^{\infty} \left[\int_{y_1=-\infty}^{\infty} y_1 f(y_1 \mid y_2)\, dy_1 \right] f_2(y_2)\, dy_2 \\
&= \int_{y_2=-\infty}^{\infty} \int_{y_1=-\infty}^{\infty} y_1 \frac{f(y_1, y_2)}{\cancel{f_2(y_2)}} \cancel{f_2(y_2)}\, dy_1\, dy_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 f(y_1, y_2)\, dy_1\, dy_2
\end{aligned}$$

Theorem (5.14)

Let Y_1 and Y_2 denote random variables. Then

$$E[Y_1] = E[E[Y_1 \mid Y_2]],$$

where on the right-hand side the inside expectation is with respect to the conditional distribution of Y_1 given Y_2 and the outside expectation is with respect to the distribution of Y_2.

Proof: (continuous case; the discrete case is analogous.)

$$\begin{aligned}
E[E[Y_1 \mid Y_2]] &= \int_{-\infty}^{\infty} E[Y_1 \mid Y_2] f_2(y_2)\, dy_2 \\
&= \int_{y_2=-\infty}^{\infty} \left[\int_{y_1=-\infty}^{\infty} y_1 f(y_1 \mid y_2)\, dy_1 \right] f_2(y_2)\, dy_2 \\
&= \int_{y_2=-\infty}^{\infty} \int_{y_1=-\infty}^{\infty} y_1 \frac{f(y_1, y_2)}{\cancel{f_2(y_2)}} \cancel{f_2(y_2)}\, dy_1\, dy_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 f(y_1, y_2)\, dy_1\, dy_2 = E[Y_1].
\end{aligned}$$

Remark:

Our computation may be easier with the information given in the problem.

Note:

$E[Y_1 \mid Y_2]$ can be regarded as a RV.
Then

$$V[Y_1 \mid Y_2] = E[Y_1^2 \mid Y_2] - E[Y_1 \mid Y_2]^2.$$

The formula in Theorem 5.14 was a relationship between the unconditional expectation $E[Y_1]$ and the conditional expectation $E[Y_1 \mid Y_2]$.
There is a more complicated relation between the unconditional variance $V[Y]$ and the conditional variance $V[Y_1 \mid Y_2]$:

Theorem (5.15)

$$V[Y_1] = E[V[Y_1 \mid Y_2]] + V[E[Y_1 \mid Y_2]].$$

Remark:

There are similar, but more complicated relationships between conditional and unconditional higher moments $E[Y_1^3]$, etc.

Proof of Theorem 5.15:

Recall

$$V[Y_1 \mid Y_2] = E[Y_1^2 \mid Y_2] - E[Y_1 \mid Y_2]^2.$$

Then

$$E[V[Y_1 \mid Y_2]] = E[E[Y_1^2 \mid Y_2]] - E[E[Y_1 \mid Y_2]^2].$$

By definition,

$$V[E[Y_1 \mid Y_2]] = E[E[Y_1^2 \mid Y_2]^2] - E[E[Y_1 \mid Y_2]]^2.$$

Proof: (continued)

The variance of Y_1 is

$$V[Y_1] = E[Y_1^2] - E[Y_1]^2$$

Proof: (continued)

The variance of Y_1 is

$$\begin{aligned} V[Y_1] &= E[Y_1^2] - E[Y_1]^2 \\ &= E[E[Y_1^2 \mid Y_2]] - E[E[Y_1 \mid Y_2]]^2 \end{aligned}$$

Proof: (continued)

The variance of Y_1 is

$$\begin{aligned} V[Y_1] &= E[Y_1^2] - E[Y_1]^2 \\ &= E[E[Y_1^2 \mid Y_2]] - E[E[Y_1 \mid Y_2]]^2 \qquad \text{(By Theorem 5.14)} \end{aligned}$$

Proof: (continued)

The variance of Y_1 is

$$\begin{aligned} V[Y_1] &= E[Y_1^2] - E[Y_1]^2 \\ &= E[E[Y_1^2 \mid Y_2]] - E[E[Y_1 \mid Y_2]]^2 \qquad \text{(By Theorem 5.14)} \\ &= E[E[Y_1^2 \mid Y_2]] - E[E[Y_1^2 \mid Y_2]^2] + E[E[Y_1^2 \mid Y_2]^2] \\ &\qquad - E[E[Y_1 \mid Y_2]]^2 \end{aligned}$$

Proof: (continued)

The variance of Y_1 is

$$\begin{aligned} V[Y_1] &= E[Y_1^2] - E[Y_1]^2 \\ &= E[E[Y_1^2 \mid Y_2]] - E[E[Y_1 \mid Y_2]]^2 \qquad \text{(By Theorem 5.14)} \\ &= \underbrace{E[E[Y_1^2 \mid Y_2]] - E[E[Y_1^2 \mid Y_2]^2]} + E[E[Y_1^2 \mid Y_2]^2] \\ &\qquad - E[E[Y_1 \mid Y_2]]^2 \\ &= \overbrace{E[V[Y_1 \mid Y_2]]} \end{aligned}$$

- By the definition of "conditional variance".

Proof: (continued)

The variance of Y_1 is

$$
\begin{aligned}
V[Y_1] &= E[Y_1^2] - E[Y_1]^2 \\
&= E[E[Y_1^2 \mid Y_2]] - E[E[Y_1 \mid Y_2]]^2 \qquad \text{(By Theorem 5.14)} \\
&= \underbrace{E[E[Y_1^2 \mid Y_2]] - E[E[Y_1^2 \mid Y_2]^2]} + \underbracket{E[E[Y_1^2 \mid Y_2]^2] - E[E[Y_1 \mid Y_2]]^2} \\
&= \overbrace{E[V[Y_1 \mid Y_2]]} + \overbracket{V[E[Y_1 \mid Y_2]]},
\end{aligned}
$$

- By the definition of "conditional variance".
- Because $E[Y_1 \mid Y_2]$ is a RV and $V[X] = E[X^2] - E[X]^2$.

Proof: (continued)

The variance of Y_1 is

$$
\begin{aligned}
V[Y_1] &= E[Y_1^2] - E[Y_1]^2 \\
&= E[E[Y_1^2 \mid Y_2]] - E[E[Y_1 \mid Y_2]]^2 \qquad \text{(By Theorem 5.14)} \\
&= \underbrace{E[E[Y_1^2 \mid Y_2]] - E[E[Y_1^2 \mid Y_2]^2]} + \underbracket{E[E[Y_1^2 \mid Y_2]^2]} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \underbracket{- E[E[Y_1 \mid Y_2]]^2} \\
&= \overbrace{E[V[Y_1 \mid Y_2]]} + \overbracket{V[E[Y_1 \mid Y_2]]}, \qquad \text{as claimed.}
\end{aligned}
$$

- By the definition of "conditional variance".
- Because $E[Y_1 \mid Y_2]$ is a RV and $V[X] = E[X^2] - E[X]^2$.

Proof: (continued)

The variance of Y_1 is

$$\begin{aligned}
V[Y_1] &= E[Y_1^2] - E[Y_1]^2 \\
&= E[E[Y_1^2 \mid Y_2]] - E[E[Y_1 \mid Y_2]]^2 \qquad \text{(By Theorem 5.14)} \\
&= \underbrace{E[E[Y_1^2 \mid Y_2]] - E[E[Y_1^2 \mid Y_2]^2]} + \underbracket{E[E[Y_1^2 \mid Y_2]^2]} \\
&\qquad \underbracket{- E[E[Y_1 \mid Y_2]]^2} \\
&= \overbrace{E[V[Y_1 \mid Y_2]]} + \overbracket{V[E[Y_1 \mid Y_2]]}, \qquad \text{as claimed.}
\end{aligned}$$

- By the definition of “conditional variance”.
- Because $E[Y_1 \mid Y_2]$ is a RV and $V[X] = E[X^2] - E[X]^2$.

Note: Make sure to remember this result: it will help with Exercises 5.136 and 5.138.

Exercise 5.136

The number of defects per yard in a certain fabric, Y, has a Poisson distribution with parameter λ, which is assumed to be a random variable with a density function given by

$$f(\lambda) = \begin{cases} e^{-\lambda} & \lambda \geq 0, \\ 0 & \text{elsewhere.} \end{cases}$$

Find (a) the expectation, and (b) the variance of Y. (c) Is it likely that $Y \geq 9$?

Exercise 5.138

Assume that Y denotes the number of bacteria per cubic centimeter in a particular liquid and that Y has a Poisson distribution with parameter λ. Further assume that λ varies from location to location and has a Gamma distribution with parameters α and β, where α is a positive integer. If we randomly select a location, what is the

(a) expected number of bacteria per cubic centimeter?

(b) standard deviation of the number of bacteria per cubic centimeter?

Hints for Exercises 5.136 and 5.138:

- Use Theorem 5.14 for parts (a) and Theorem 5.15 for parts (b).
- Now 5.136(c) is easy.

Exercise 5.167

Let Y_1 and Y_2 be jointly distributed random variables with finite variances.

(a) Show that

$$E[Y_1 Y_2]^2 \leq E[Y_1^2]E[Y_2^2]$$

by observing that

$$E[(tY_1 - Y_2)^2] \geq 0$$

for any real number t.

(b) Hence prove that

$$-1 \leq \rho_{Y_1,Y_2} \leq 1$$

.

Solution: (Exercise 5.167(a))

If Y_1, Y_2 are RVs, then note that $E[(tY_1 - Y_2)^2] \geq 0$.
So

$$E[t^2 Y_1^2 - 2tY_1 Y_2 + Y_2^2] \geq 0.$$
$$\therefore \quad t^2 E[Y_1^2] - 2tE[Y_1 Y_2] + E[Y_2^2] \geq 0.$$

This is a quadratic $at^2 + bt + c \geq 0$.
Since this is true for all real t, $b^2 - 4ac \leq 0$.
Now

$$b = -2E[Y_1 Y_2], \qquad a = E[Y_1^2], \qquad c = E[Y_2^2].$$

So

$$(-2E[Y_1 Y_2])^2 - 4E[Y_1^2]E[Y_2^2] \leq 0.$$
$$\therefore \quad 4(E[Y_1 Y_2])^2 - E[Y_1^2]E[Y_2^2]) \leq 0.$$
$$\therefore \quad E[Y_1 Y_2]^2 \leq E[Y_1^2]E[Y_2^2].$$

Solution: (Exercise 5.167(b))

Now recall, for RVs X_1, X_2,

$$\rho_{X_1,X_2} = \frac{\mathrm{Cov}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}} = \frac{E[(X_1-\mu_1)(X_2-\mu_2)]}{\sqrt{E[(X_1-\mu_1)^2]E[(X_2-\mu_2)^2]}},$$

where $\mu_1 = E[X_1]$ and $\mu_2 = E[X_2]$.

Now let $Y_1 = X_1 - \mu_1$ and $Y_2 = X_2 - \mu_2$.

By Exercise 5.167(a), we know that $E[Y_1Y_2]^2 \leq E[Y_1^2]E[Y_2^2]$, that is,

$$\frac{E[Y_1Y_2]^2}{E[Y_1^2]E[Y_2^2]} \leq 1 \implies \rho^2_{Y_1,Y_2} \leq 1,$$

where we have used the linearity of Expectations.
Thus

$$-1 \leq \rho_{Y_1,Y_2} \leq 1,$$

as desired.

Exercise 5.31:

The joint density function of Y_1 and Y_2 is given by

$$f(y_1, y_2) = \begin{cases} 30y_1y_2^2 & y_1 - 1 \le y_2 \le 1 - y_1, 0 \le y_1 \le 1, \\ 0 & \text{elsewhere.} \end{cases}$$

(a) Show that the marginal density of Y_1 is a beta density with $\alpha = 2$ and $\beta = 4$.

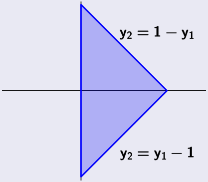
(b) Derive the marginal density of Y_2.

(c) Derive the conditional density of Y_2 given $Y_1 = y_1$.

(d) Find $P(Y_2 > 0 \mid Y_1 = .75)$.

Solution:

First, graph the region in which the density function is nonzero:



Solution: (a)

Recall the definition of marginal density function:

$$
\begin{aligned}
f_1(y_1) &= \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_2 \\
&= \begin{cases} \displaystyle\int_{y_1-1}^{1-y_1} 30 y_1 y_2^2 \, dy_2 & 0 \le y_1 \le 1, \\ 0 & \text{otherwise.} \end{cases} \\
&= \begin{cases} 30 y_1 \left. \dfrac{y_2^3}{3} \right|_{y_1-1}^{1-y_1} & 0 \le y_1 \le 1, \\ 0 & \text{otherwise.} \end{cases} \\
&= \begin{cases} 20 y_1 (1-y_1)^3 & 0 \le y_1 \le 1, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

All of this is part of the definition of $f_1(y_1)$!

Solution: (b)

$f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1.$

Now there are 3 cases!

$$\text{If } -1 \leq y_2 \leq 0, \quad \int_0^{1+y_2} 30 y_1 y_2^2 \, dy_1 = 30 y_2^2 \left. \frac{y_1^2}{2} \right|_0^{1+y_2} = 15 y_2^2 (1 + y_2)^2,$$

$$\text{if } 0 \leq y_2 \leq -1, \quad \int_0^{1-y_2} 30 y_1 y_2^2 \, dy_1 = 30 y_2^2 \left. \frac{y_1^2}{2} \right|_0^{1-y_2} = 15 y_2^2 (1 - y_2)^2,$$

and 0 otherwise.

That is,

$$f_2(y_2) = \begin{cases} 0 & y_2 \notin [-1, 1], \\ 15 y_2^2 (1 + y_2)^2 & y_2 \in [-1, 0], \\ 15 y_2^2 (1 - y_2)^2 & y_2 \in [0, 1]. \end{cases}$$

Solution: (c)

Solution: (c)

$$f(y_2 \mid y_1) = \frac{f(y_1, y_2)}{f_1(y_1)}.$$

Solution: (c)

$f(y_2 \mid y_1) = \dfrac{f(y_1, y_2)}{f_1(y_1)}$. This is defined only if $y_1 \in (0, 1)$.

Solution: (c)

$f(y_2 \mid y_1) = \dfrac{f(y_1, y_2)}{f_1(y_1)}$. This is defined only if $y_1 \in (0,1)$. In the triangle $\{y_1 - 1 < y_2 < 1 - y_1, 0 < y_1 < 1\}$, it is $\dfrac{30y_1 y_2^2}{20y_1(1-y_1)^3}$

Solution: (c)

$f(y_2 \mid y_1) = \dfrac{f(y_1, y_2)}{f_1(y_1)}$. This is defined only if $y_1 \in (0, 1)$. In the triangle $\{y_1 - 1 < y_2 < 1 - y_1, 0 < y_1 < 1\}$, it is $\dfrac{30 y_1 y_2^2}{20 y_1 (1 - y_1)^3} = \dfrac{3}{2} y_2^2 (1 - y_1)^{-3}$.

Solution: (c)

$f(y_2 \mid y_1) = \dfrac{f(y_1, y_2)}{f_1(y_1)}$. This is defined only if $y_1 \in (0,1)$. In the triangle $\{y_1 - 1 < y_2 < 1 - y_1, 0 < y_1 < 1\}$, it is $\dfrac{30 y_1 y_2^2}{20 y_1 (1-y_1)^3} = \dfrac{3}{2} y_2^2 (1-y_1)^{-3}$.

Thus

$$f(y_2 \mid y_1) = \begin{cases} \text{undefined} & y_1 \notin (0,1), \\ \dfrac{3}{2} y_2^2 (1-y_1)^{-3} & 0 < y_1 < 1 \text{ and } y_1 - 1 < y_2 < 1 - y_1, \\ 0 & \text{otherwise.} \end{cases}$$

Solution: (c)

$f(y_2 \mid y_1) = \dfrac{f(y_1, y_2)}{f_1(y_1)}$. This is defined only if $y_1 \in (0,1)$. In the triangle $\{y_1 - 1 < y_2 < 1 - y_1, 0 < y_1 < 1\}$, it is $\dfrac{30 y_1 y_2^2}{20 y_1 (1-y_1)^3} = \dfrac{3}{2} y_2^2 (1-y_1)^{-3}$.

Thus

$$f(y_2 \mid y_1) = \begin{cases} \text{undefined} & y_1 \notin (0,1), \\ \dfrac{3}{2} y_2^2 (1-y_1)^{-3} & 0 < y_1 < 1 \text{ and } y_1 - 1 < y_2 < 1 - y_1, \\ 0 & \text{otherwise.} \end{cases}$$

Solution: (d)

Solution: (c)

$f(y_2 \mid y_1) = \dfrac{f(y_1, y_2)}{f_1(y_1)}$. This is defined only if $y_1 \in (0,1)$. In the triangle $\{y_1 - 1 < y_2 < 1 - y_1, 0 < y_1 < 1\}$, it is $\dfrac{30 y_1 y_2^2}{20 y_1 (1-y_1)^3} = \dfrac{3}{2} y_2^2 (1-y_1)^{-3}$.

Thus

$$f(y_2 \mid y_1) = \begin{cases} \text{undefined} & y_1 \notin (0,1), \\ \dfrac{3}{2} y_2^2 (1-y_1)^{-3} & 0 < y_1 < 1 \text{ and } y_1 - 1 < y_2 < 1 - y_1, \\ 0 & \text{otherwise.} \end{cases}$$

Solution: (d)

$$P(Y_2 > 0 \mid Y_1 = .75) = \int_0^{\infty} f(y_2 \mid y_1)\, dy_2$$

Solution: (c)

$f(y_2 \mid y_1) = \dfrac{f(y_1, y_2)}{f_1(y_1)}$. This is defined only if $y_1 \in (0,1)$. In the triangle $\{y_1 - 1 < y_2 < 1 - y_1, 0 < y_1 < 1\}$, it is $\dfrac{30 y_1 y_2^2}{20 y_1 (1-y_1)^3} = \dfrac{3}{2} y_2^2 (1-y_1)^{-3}$.
Thus

$$f(y_2 \mid y_1) = \begin{cases} \text{undefined} & y_1 \notin (0,1), \\ \dfrac{3}{2} y_2^2 (1-y_1)^{-3} & 0 < y_1 < 1 \text{ and } y_1 - 1 < y_2 < 1 - y_1, \\ 0 & \text{otherwise.} \end{cases}$$

Solution: (d)

$$P(Y_2 > 0 \mid Y_1 = .75) = \int_0^\infty \underbrace{f(y_2 \mid y_1)}\, dy_2$$

This is 0 unless
$y_2 < 1 - 0.75 = 0.25$

Solution: (c)

$f(y_2 \mid y_1) = \dfrac{f(y_1, y_2)}{f_1(y_1)}$. This is defined only if $y_1 \in (0,1)$. In the triangle $\{y_1 - 1 < y_2 < 1 - y_1, 0 < y_1 < 1\}$, it is $\dfrac{30y_1y_2^2}{20y_1(1-y_1)^3} = \dfrac{3}{2}y_2^2(1-y_1)^{-3}$.

Thus

$$f(y_2 \mid y_1) = \begin{cases} \text{undefined} & y_1 \notin (0,1), \\ \frac{3}{2}y_2^2(1-y_1)^{-3} & 0 < y_1 < 1 \text{ and } y_1 - 1 < y_2 < 1 - y_1, \\ 0 & \text{otherwise.} \end{cases}$$

Solution: (d)

$$P(Y_2 > 0 \mid Y_1 = .75) = \int_0^\infty \underbrace{f(y_2 \mid y_1)} \, dy_2 \quad \begin{matrix}\text{This is 0 unless} \\ y_2 < 1 - 0.75 = 0.25\end{matrix}$$

$$= \int_0^{0.25} \frac{3}{2}y_2^2(1-0.75)^{-3} \, dy_2$$

Solution: (c)

$f(y_2 \mid y_1) = \dfrac{f(y_1, y_2)}{f_1(y_1)}$. This is defined only if $y_1 \in (0,1)$. In the triangle $\{y_1 - 1 < y_2 < 1 - y_1, 0 < y_1 < 1\}$, it is $\dfrac{30 y_1 y_2^2}{20 y_1 (1-y_1)^3} = \dfrac{3}{2} y_2^2 (1-y_1)^{-3}$.

Thus

$$f(y_2 \mid y_1) = \begin{cases} \text{undefined} & y_1 \notin (0,1), \\ \frac{3}{2} y_2^2 (1-y_1)^{-3} & 0 < y_1 < 1 \text{ and } y_1 - 1 < y_2 < 1 - y_1, \\ 0 & \text{otherwise.} \end{cases}$$

Solution: (d)

$$P(Y_2 > 0 \mid Y_1 = .75) = \int_0^\infty \underbrace{f(y_2 \mid y_1)}\, dy_2 \quad \begin{array}{c}\text{This is 0 unless} \\ y_2 < 1 - 0.75 = 0.25\end{array}$$

$$= \int_0^{0.25} \frac{3}{2} y_2^2 (1 - 0.75)^{-3}\, dy_2 = \frac{1}{2}.$$

Example 5.32:

A quality control plan for an assembly line involves sampling $n = 10$ finished items per day and counting Y, the number of defectives. If p denotes the probability of observing a defective, then Y has a binomial distribution, assuming that a large number of items are produced by the line. But p varies from day to day and is assumed to have a uniform distribution on the interval from 0 to $\frac{1}{4}$. Find the expected value of Y.

Solution:

We employ Theorem 5.14:
$E[Y] = E[E[Y \mid p]]$.

$$E[Y \mid p] = np \quad \text{because we know the expectation of a binomial RV.}$$

$$E[Y] = E[np] \quad \text{where } p \sim \text{Unif}\left(0, \frac{1}{4}\right).$$

We know $E[p] = \frac{1}{8}$, because we know the expectation of a uniform RV.

So $E[Y] = \boxed{\frac{n}{8}}$.

Example 5.33:

In Example 5.32, find the variance of Y.

Solution:

Here we apply Theorem 5.15:
$V[Y] = E[V[Y \mid p]] + V[E[Y \mid p]]$.
We know that for any particular value of p, Y is a binomial RV, whose mean and variance are known:

$$E[Y \mid p] - np, V[Y \mid p] = npq, \text{ (where } q = 1 - p).$$

So $V[Y] = E[npq] - V[np]$.
Remember that $p \sim \text{Unif}\left(0, \frac{1}{4}\right)$.
So

$$E[npq] = nE[pq] = n\int_0^{1/4} y(1-y)\frac{1}{1/4}\,dy$$
$$= n\int_0^{1/4} y \cdot 4dy - n\int_0^{1/4} y^2\, 4dy = nE[p] - nE[p^2].$$

So this can be done using the known mean and variance of a uniform RV.
Instead, we directly do the above integrals:

Solution: (continued)

$$E[npq] = n \cdot 4 \cdot \left.\frac{y^2}{2}\right|_0^{1/4} - n \cdot 4 \cdot \left.\frac{y^3}{3}\right|_0^{1/4}$$
$$= \frac{n}{8} - \frac{n}{48} = \frac{5n}{48}.$$

The other term is $V[E[Y \mid p]]$,
which is

$$V[np] = n^2 V[p], \qquad \text{where } p \sim \text{Unif}\left(0, \frac{1}{4}\right)$$
$$= n^2 \cdot \frac{1}{12} \cdot \left(\frac{1}{4} - 0\right)^2 = \frac{n^2}{192}.$$

So $V[Y] = \boxed{\frac{5n}{48} + \frac{n^2}{192}}$.

Solution: (Exercise 5.136 (c))

The simplest solution is to apply Tchebysheff's theorem.

You previously found μ and σ^2 for Y in parts (a) and (b).
The value 9 is far from μ, measured in units of σ.
So applying Tchebysheff makes sense.

Here $\mu = 1$, so

$$P(Y > 9) = P((Y - \mu) > (9 - \mu)) = P(Y - \mu > 8) \leq P(|Y - \mu| > 8).$$

Now use the theorem.

Slogan for Tchebysheff's Theorem:

The probability that Y is far from its mean, where "far" is measured in units of σ, is small.

Example:

Suppose Y is a normal RV with mean μ and variance σ^2.
What kind of RV is $Z = \dfrac{Y-\mu}{\sigma}$?
Standard Normal.

Why?
Notice that we can compute

$$E[Z] = E\left[\frac{Y-\mu}{\sigma}\right] = \frac{1}{\sigma}E[Y-\mu] = \frac{1}{\sigma}(\mu-\mu) = 0,$$

$$V[Z] = V\left[\frac{Y-\mu}{\sigma}\right] = \frac{1}{\sigma^2}V[Y-\mu] = \frac{1}{\sigma^2}V[Y] = \frac{1}{\sigma^2}\sigma^2 = 1.$$

Does this show that Z is standard normal?
No. This shows the "standard" part, but not the "normal" part.
That is, we showed that Z has mean 0 and variance 1, but not that Z is normally distributed.

Can you think of another RV which has mean 0 and variance 1?

Two examples:

1. Uniform RV on $[-a, a]$.

   This has mean $\dfrac{a+(-a)}{2} = 0$ and variance $\dfrac{(a-(-a))^2}{12} = \dfrac{4a^2}{12} = \dfrac{a^2}{3}$.
   So if we take $a = \sqrt{3}$, this has variance 1.

2. Let

$$X = \begin{cases} +1 & \text{with probability } 1/2, \\ -1 & \text{with probability } 1/2. \end{cases}$$

   Then $E[X] = 0$, $V[X] = E[X^2] - 0^2 = 1$.

This shows that it is not possible to "recognize" a RV using only its mean and variance.

One way to "recognize" a RV is to use the moment generating function. That is, if we know (for whatever reason) that $m_X(t) = m_Y(t)$ for all t near $t = 0$, then we have that X and Y have the same distribution.

The way we prove that $Z = \dfrac{Y-\mu}{\sigma}$ is normal is we show that Z has the right MGF.
What is the MGF of a normal RV Y with mean μ and variance σ^2?
$e^{\mu t + \frac{\sigma^2 t^2}{2}}$.

This means that the MGF of Z is

$$\begin{aligned} E[e^{tZ}] &= E[e^{t\left(\frac{Y-\mu}{\sigma}\right)}] = E[e^{\frac{t}{\sigma}(Y-\mu)}] = E[e^{\left(\frac{t}{\sigma}Y\right)}]E[e^{\left(\frac{t}{\sigma}(-\mu)\right)}] \\ &= e^{-\frac{\mu t}{\sigma}} E[e^{\left(\frac{t}{\sigma}\right)Y}] = e^{-\frac{\mu t}{\sigma}} m_Y\left(\frac{t}{\sigma}\right) = e^{-\frac{\mu t}{\sigma}} e^{\left(\mu\left(\frac{t}{\sigma}\right) + \frac{\sigma^2\left(\frac{t}{\sigma}\right)^2}{2}\right)} \\ &= e^{\left(-\frac{\mu t}{\sigma} + \frac{\mu t}{\sigma} + \frac{t^2}{2}\right)} = e^{\frac{t^2}{2}}. \end{aligned}$$

And thus $m_2(t) = e^{(0)t + \frac{(1)^2 t^2}{2}}$.
This is the MGF of a normal RV with mean 0 and variance 1.
Therefore by "uniqueness of MGF", Z is normal, with mean 0 and variance 1.

End of Chapter 5

Chapter 6

Functions of Random Variables

Functions of Random Variables

Suppose we have random variables X, Y with some joint distribution. We can construct a new RV U from X and Y by combining them somehow.
For example, $U = \frac{X}{Y}$, $U = 4X + 3$, $U = X + Y$, $U = \arctan X$, $U = X^2 + Y^2$, etc.

Suppose we "know" X and Y in the sense that we know the joint density or joint distribution function.
How do we figure out the distribution or density function of U?
This problem is addressed by the methods of Chapter 6.

Simple Example:

Suppose Y has the density function

$$f(y) = \begin{cases} 2y & y \in [0,1], \\ 0 & y \notin [0,1]. \end{cases}$$

Let $U = 3Y - 1$.
Find the PDF of U.

Finding the Probability Distribution of a Function of RVs

There are three key methods for finding the probability distribution for a function of random variables:

(1) The method of Distribution Functions,

(2) The method of Transformations,
and

(3) The method of Moment-Generating Functions.

There is also a fourth method for finding the *joint* distribution of several functions of random variables.

The method that works "best" varies from one application to another. Hence, acquaintance with the first three methods is desirable.

Consider random variables $Y_1, \ldots, Y_n$ and a function $U(Y_1, \ldots, Y_n)$, denoted simply as U.
Then three of the methods for finding the probability distribution of U are as follows:

"Method of Distribution Functions"

Outline:

- Find the distribution function (CDF) $F_Y(y) = P(Y \leq y)$.
- Use this CDF to find the CDF $F_U(u)$, by transforming the inequality $Y \leq y$ so that $U(= 3Y - 1)$ is on the LHS.
- Now differentiate $F_U(u)$ to get the density function for U, $f_U(u)$.

Solution: (Simple Example)

<u>Step 1</u>: Find $F_Y = P(Y \leq y) = \displaystyle\int_{-\infty}^{y} f(t)\,dt$.

$F_Y \equiv 0$ if $y < 0$.
Also, since the PDF f_Y integrates to 1, $F_Y \equiv 1$ if $y > 1$.

If $y \in [0,1]$, then $\displaystyle\int_{-\infty}^{y} f(t)\,dt = \int_0^y 2t\,dt = t^2\Big|_0^y$.

So $F_Y(y) = y^2$.
Thus

$$F_Y(y) = \begin{cases} y^2 & y \in [0,1], \\ 0 & y < 0 \end{cases}$$

Solution:(continued)

Step 2:
We know $P(Y \le y)$; we will use this to find $P(U \le u)$.

We write

$$F_U(u) = P(U \le u) = P(3Y - 1 \le u)$$

Now we start transforming: $$= P(3Y \le u+1) = P\left(Y \le \frac{u+1}{3}\right)$$

Now we use Step 1: $$= F_Y\left(\frac{u+1}{3}\right)$$

$$= \begin{cases} \left(\frac{u+1}{3}\right)^2 & \frac{u+1}{3} \in [0,1], \\ 0 & \frac{u+1}{3} < 0, \\ 1 & \frac{u+1}{3} > 1. \end{cases}$$

Notice $\frac{u+1}{3} < 0$ corresponds to $u < -1$, while $\frac{u+1}{3} > 1$ corresponds to $u > 2$.

Thus

$$F_U(u) = \begin{cases} \frac{(u+1)^2}{9} & u \in [-1,2], \\ 0 & u < -1, \\ 1 & u > 2. \end{cases}$$

Solution:(continued)

Step 3:

Find the PDF $f_U(u) = \dfrac{d}{du} F_U$.

Now that we know F_U, this is straightforward:

$$f_U(u) = \begin{cases} \frac{2(u+1)}{9} & u \in [-1, 2], \\ 0 & u \notin [-1, 2]. \end{cases}$$

A more complicated example:

Suppose that X and Y are independent and have the unniform distribution on the unit interval $[0, 1]$. Let $U = X + Y$. Find the density function $f_U(u)$.

Remark:

(X, Y) is a random point in $[0, 1] \times [0, 1]$.

Solution:

Write $F_U(u) = P(U \leq u) = P(X + Y \leq u)$.

Solution: (continued)

Since (X, Y) is a random point in $[0, 1] \times [0, 1]$, we have $0 \leq X + Y \leq 2$.

Solution: (continued)

Since (X, Y) is a random point in $[0,1] \times [0,1]$, we have $0 \leq X + Y \leq 2$. Thus $F_U(u) = 0$ if $u < 0$ and $F_U(u) = 1$ if $u > 2$.

Solution: (continued)

Since (X, Y) is a random point in $[0,1] \times [0,1]$, we have $0 \leq X + Y \leq 2$. Thus $F_U(u) = 0$ if $u < 0$ and $F_U(u) = 1$ if $u > 2$. In between we can draw a picture and solve geometrically.

Solution: (continued)

Since (X, Y) is a random point in $[0,1] \times [0,1]$, we have $0 \leq X + Y \leq 2$. Thus $F_U(u) = 0$ if $u < 0$ and $F_U(u) = 1$ if $u > 2$. In between we can draw a picture and solve geometrically.

We want to draw a square and the region $x + y \leq u$:

Solution: (continued)

Since (X, Y) is a random point in $[0,1] \times [0,1]$, we have $0 \leq X + Y \leq 2$. Thus $F_U(u) = 0$ if $u < 0$ and $F_U(u) = 1$ if $u > 2$. In between we can draw a picture and solve geometrically.

We want to draw a square and the region $x + y \leq u$:

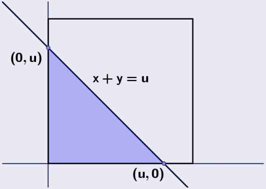
Shaded region: $(x, y) \in [0,1] \times [0,1]$

and $x + y \leq u$.

Solution: (continued)

Since (X, Y) is a random point in $[0,1] \times [0,1]$, we have $0 \leq X + Y \leq 2$. Thus $F_U(u) = 0$ if $u < 0$ and $F_U(u) = 1$ if $u > 2$. In between we can draw a picture and solve geometrically.

We want to draw a square and the region $x + y \leq u$:



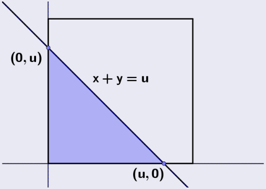
Shaded region: $(x, y) \in [0,1] \times [0,1]$

and $x + y \leq u$.

Solution: (continued)

Since (X, Y) is a random point in $[0,1] \times [0,1]$, we have $0 \leq X + Y \leq 2$. Thus $F_U(u) = 0$ if $u < 0$ and $F_U(u) = 1$ if $u > 2$. In between we can draw a picture and solve geometrically.

We want to draw a square and the region $x + y \leq u$:



Shaded region: $(x, y) \in [0,1] \times [0,1]$

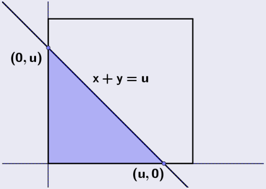
and $x + y \leq u$.

What is the area of the shaded region?

Solution: (continued)

Since (X, Y) is a random point in $[0,1] \times [0,1]$, we have $0 \leq X + Y \leq 2$. Thus $F_U(u) = 0$ if $u < 0$ and $F_U(u) = 1$ if $u > 2$. In between we can draw a picture and solve geometrically.

We want to draw a square and the region $x + y \leq u$:



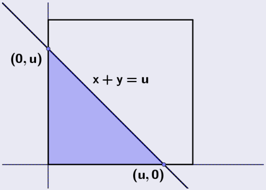
Shaded region: $(x, y) \in [0,1] \times [0,1]$ and $x + y \leq u$.

What is the area of the shaded region? $\dfrac{u^2}{2}$.

Solution: (continued)

Since (X, Y) is a random point in $[0, 1] \times [0, 1]$, we have $0 \leq X + Y \leq 2$. Thus $F_U(u) = 0$ if $u < 0$ and $F_U(u) = 1$ if $u > 2$. In between we can draw a picture and solve geometrically.

We want to draw a square and the region $x + y \leq u$:



Shaded region: $(x, y) \in [0, 1] \times [0, 1]$ and $x + y \leq u$.

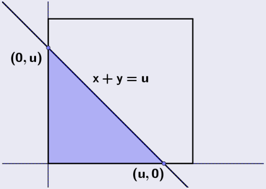
What is the area of the shaded region? $\dfrac{u^2}{2}$.

This works for $0 \leq u \leq 1$.

Solution: (continued)

Since (X, Y) is a random point in $[0,1] \times [0,1]$, we have $0 \leq X + Y \leq 2$. Thus $F_U(u) = 0$ if $u < 0$ and $F_U(u) = 1$ if $u > 2$. In between we can draw a picture and solve geometrically.

We want to draw a square and the region $x + y \leq u$:



Shaded region: $(x, y) \in [0,1] \times [0,1]$ and $x + y \leq u$.

What is the area of the shaded region? $\dfrac{u^2}{2}$.

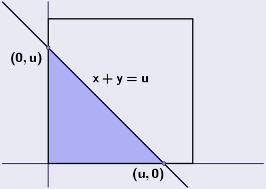
This works for $0 \leq u \leq 1$.

If $1 \leq u \leq 2$, the picture is different:

Solution: (continued)

Since (X, Y) is a random point in $[0,1] \times [0,1]$, we have $0 \leq X + Y \leq 2$. Thus $F_U(u) = 0$ if $u < 0$ and $F_U(u) = 1$ if $u > 2$. In between we can draw a picture and solve geometrically.

We want to draw a square and the region $x + y \leq u$:



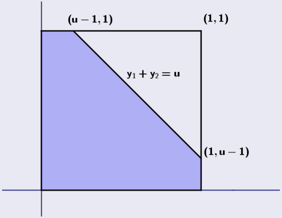
Shaded region: $(x, y) \in [0,1] \times [0,1]$

and $x + y \leq u$.

What is the area of the shaded region? $\dfrac{u^2}{2}$.

This works for $0 \leq u \leq 1$.

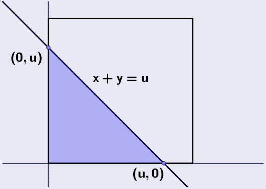
If $1 \leq u \leq 2$, the picture is different:



Solution: (continued)

Since (X, Y) is a random point in $[0,1] \times [0,1]$, we have $0 \leq X + Y \leq 2$. Thus $F_U(u) = 0$ if $u < 0$ and $F_U(u) = 1$ if $u > 2$. In between we can draw a picture and solve geometrically.

We want to draw a square and the region $x + y \leq u$:



Shaded region: $(x, y) \in [0,1] \times [0,1]$

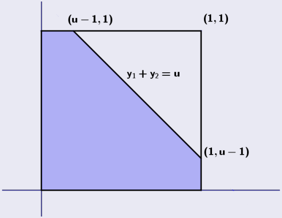
and $x + y \leq u$.

What is the area of the shaded region? $\dfrac{u^2}{2}$.

This works for $0 \leq u \leq 1$.

If $1 \leq u \leq 2$, the picture is different:

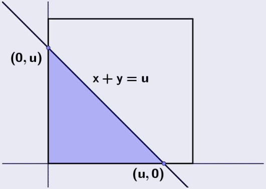
Notice that this is no longer a triangle; it is a square with a triangle removed.



Solution: (continued)

Since (X, Y) is a random point in $[0,1] \times [0,1]$, we have $0 \leq X + Y \leq 2$. Thus $F_U(u) = 0$ if $u < 0$ and $F_U(u) = 1$ if $u > 2$. In between we can draw a picture and solve geometrically.

We want to draw a square and the region $x + y \leq u$:



Shaded region: $(x, y) \in [0,1] \times [0,1]$

and $x + y \leq u$.

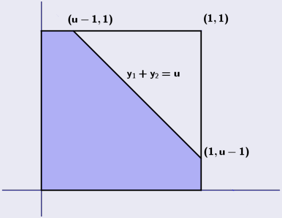
What is the area of the shaded region? $\dfrac{u^2}{2}$.

This works for $0 \leq u \leq 1$.

If $1 \leq u \leq 2$, the picture is different:

Notice that this is no longer a triangle; it is a square with a triangle removed.

The removed triangle has area $\dfrac{1}{2}(2-u)(2-u)$.



Solution: (continued)

This tells us that

$$F_U(u) = \begin{cases} 0 & u < 0, \\ 1 & u > 2, \\ \frac{u^2}{2} & u \in [0,1], \\ 1 - \frac{1}{2}(2-u)^2 & u \in [1,2]. \end{cases}$$

Notice that values match at endpoints: F_U is continuous.

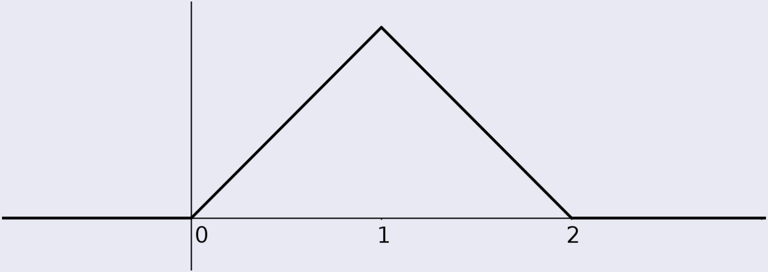
The PDF $f_U(u)$ is $\dfrac{d}{du}F_U$.

So

$$\begin{aligned} &\text{if } u \notin [0,2], \quad f_U(u) = 0, \\ &\text{if } u \in [0,1], \quad f_U(u) = u, \qquad \text{and} \\ &\text{if } u \in [1,2], \quad f_U(u) = \frac{d}{du}\left(1 - \frac{1}{2}(2-u)^2\right) \\ &\qquad\qquad\qquad\qquad = 2 \cdot \left(-\frac{1}{2}\right) \cdot (2-u) \cdot (-1) = 2 - u. \end{aligned}$$

Solution: (continued)

So the graph of $f_U(u)$ is



Remark:

This is a special case of a general fact:
if $U = X + Y$, then the density of U is the convolution of the densities of X and Y.

Remark:

Why are we interested in this problem of understanding a RV U which is a function of other RVs?
From the point of view of this book, the answer is "statistics".

If $Y_1, \ldots, Y_n$ are IID samples from some distribution, we take our samples and compute, for example, $U = \frac{1}{n}(Y_1 + \cdots + Y_n)$.
What is the distribution of U?
Note that it is NOT the the same as the distribution of one of the Y_i!

We can ask the same question for, e.g., $U = Y_1^2, \ldots, Y_n^2$, $U =$ standard deviation or variance of the Y_i.

Since the Y_i are RVs, U is a RV.
What is the distribution of U?

Example

Suppose Y_1, Y_2 have joint density

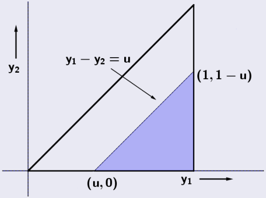
$$f(y_1, y_2) = \begin{cases} 3y_1 & 0 \le y_2 \le y_1 \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find the PDF of $U = Y_1 - Y_2$.

Solution:

$$F_U(u) = P(U \le u) = P(Y_1 - Y_2 \le u).$$

To find this probability, we will need to do a double integral, so we draw a picture:



Shaded region is where $y_1 - y_2 \ge u$ and the PDF is nonzero.

Solution: (continued)

$P(Y_1 - Y_2 \le u)$ is one minus the integral of the PDF over the shaded region.

Assuming that $u \in [0,1]$, we see that

$$P(Y_1 - Y_2 \le u) = 1 - \int_u^1 \int_0^{y_1 - u} 3y_1 \, dy_2 \, dy_1.$$

This is an exercise in integration.

<u>Answer:</u> $\frac{1}{2}(3u - u^3)$.

Now we can write down the CDF $F_U(u)$ by handling the "stupid" cases.

Recall that (Y_1, Y_2) is a point in the triangle, so $0 \le Y_1 - Y_2 \le 1$. Thus

$$F_U(u) = \begin{cases} 0 & u \le 0, \\ 1 & u \ge 1, \\ \frac{1}{2}(3u - u^3) & u \in [0,1]. \end{cases}$$

To find the density function $f_U(u)$, use $f_U(u) = \frac{d}{du} F_U(u)$.

"Method of Transformations"

Recall our first example:
Y was a RV with the PDF

$$f_Y(y) = \begin{cases} 2y & 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \qquad U = 3Y - 1.$$

We found $F_U(u)$ by writing $F_U(u) = P(U \le u) = P(3Y - 1 \le u)$ and then rearranging to get $P\left(Y \le \dfrac{u+1}{3}\right) = F_Y\left(\dfrac{u+1}{3}\right)$.

Abstractly, we know f_Y and F_Y.
Also $u = h(Y)$, where h is an increasing function, so it preserves inequalities.
So we can write

$$F_U(u) = P(U \le u) = P(h(Y) \le u) = P(Y \le h^{-1}(u)) = F_Y(h^{-1}(u)).$$

Thus

$$f_U(u) = \frac{d}{du} F_U(u) = \frac{d}{du} F_Y(h^{-1}(u)) = f_Y(h^{-1}(u)) \cdot \frac{d}{du}(h^{-1})(u).$$

This also works if h is decreasing:
if h is decreasing, it <u>reverses</u> inequalities.

So for such an h,

$$P(U \leq u) = P(h(Y) \leq u) = P(h^{-1}(h(y)) \geq h^{-1}(u))$$
$$= P(Y \geq h^{-1}u) = 1 - F_Y(h^{-1}(u)).$$

So

$$\frac{d}{du}F_U(u) = \frac{d}{du}\left[1 - F_Y(h^{-1}(u))\right] = -f_Y(h^{-1}(u)) \cdot \frac{d}{du}(h^{-1})(u).$$

Thus, we can combine these two observations into the main formula:

$$f_U(u) = f_Y(h^{-1}(u)) \cdot \left|\frac{d}{du}(h^{-1})(u)\right|.$$

Note that the absolute value covers both the cases.

Going back to our example, how do we apply this?
What is h^{-1}?

$$U = h(Y) = 3Y - 1 \implies Y = \frac{U+1}{3} \implies h^{-1}(u) = \frac{u+1}{3}.$$

So $\left|\frac{d}{du}(h^{-1})\right| \equiv \frac{1}{3}$.

Thus $f_U(u) = \frac{1}{3}f_Y\left(\frac{u+1}{3}\right)$, which is

$$f_U(u) = \begin{cases} \dfrac{2}{3}\left(\dfrac{u+1}{3}\right) & 0 \leq \dfrac{u+1}{3} \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

$$\therefore \quad f_U(u) = \begin{cases} \dfrac{2(u+1)}{9} & -1 \leq u \leq 2, \\ 0 & \text{otherwise.} \end{cases}$$

Notice that we have seen cases where $U = h(X, Y)$
– this will not be invertible.
Even cases with one variable might not be invertible.

The method can be adapted for non-invertible cases; we consider such an example:

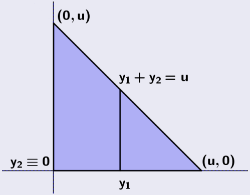
Example

Let Y_1, Y_2 be independent exponential RVs with parameter $\; = 1$. Let $U = Y_1 + Y_2$. Find the PDF $f_U(u)$.

Solution:

We already know the method:
Write $F_U(u) = P(U \le u) = P(Y_1 + Y_2 \le u)$, and then do a double integral.
Note that we need to know the joint density function of Y_1, Y_2.
They are independent, so the joint density is the product of the marginal densities:

$$f(y_1, y_2) = \begin{cases} e^{-(y_1+y_2)} & y_1, y_2 \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$



$$P(Y_1 + Y_2 \le u) = \iint_{\text{shaded region}} f(y_1, y_2)\, dy_2\, dy_1$$

$$= \int_0^u \int_0^{u-y_1} e^{-(y_1+y_2)} dy_2\, dy_1.$$

Rather than do this integral, we will apply the method of transformations.

Where is our function invertible and increasing?
For fixed y_1, $U = y_1 + Y_2 = h(Y_2)$.

Regard this as a function of Y_2.

Solution: (continued)

Now we will use the method to obtain the joint density of U and Y_1:

$$g(y_1,u)=\begin{cases} f(y_1,h^{-1}(u))\cdot\left|\dfrac{d}{du}(h^{-1})(u)\right| & y_1\ge 0, y_2\ge 0,\\ 0 & \text{otherwise.}\end{cases}$$

Note that $h(Y_2)=Y_2+y_1$, so $h^{-1}(U)=U-y_2$.

Thus $\dfrac{d}{du}(h^{-1})\equiv 1$.

The condition $y_2\ge 0$ gives us $u-y_1\ge 0$, or $u\ge y_1$.

So "$y_1\ge 0, y_2\ge 0$" translates to $0\le y_1\le u$.

Thus

$$g(y_1,u)=\begin{cases} e^{-(y_1+(u-y_1))}\cdot 1 & 0\le y_1\le u,\\ 0 & \text{otherwise.}\end{cases}$$

$$\therefore\quad g(y_1,u)=\begin{cases} e^{-u} & 0\le y_1\le u,\\ 0 & \text{otherwise.}\end{cases}$$

Solution: (continued)

How do we obtain the density $f_U(u)$?

This is the marginal density:
Take the joint PDF and integrate out the y_1:

$$
\begin{aligned}
f_U(u) &= \int_{-\infty}^{\infty} g(y_1, u)\, dy_1 \\
&= \begin{cases} \displaystyle\int_0^u e^{-u}\, dy_1 & u \geq 0, \\ 0 & \text{otherwise.} \end{cases} \\
&= \begin{cases} ue^{-u} & u \geq 0, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

Exercise 6.26(a)

Let $\alpha, m > 0$ be constants. Suppose that Y has the Weibull distribution, whose density function is

$$f(y) = \begin{cases} \frac{1}{\alpha} m y^{m-1} e^{-\frac{y^m}{\alpha}} & y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Find the density function of $U = Y^m$.

Solution:

We use the method of transformations:

Note that $h(Y) = Y^m$ is an increasing function for $Y > 0$, and $h^{-1}(U) = U^{\frac{1}{m}}$.

What is the density $f_U(u)$?

$$f_U(u) = \begin{cases} f_Y(h^{-1}(u)) \cdot \left| \frac{d}{du}(h^{-1})(u) \right| & y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Now translate this to obtain the desired $f_U(u)$:

Solution: (continued)

$f_U(u)$

Solution: (continued)

$$f_U(u) = \begin{cases} \dfrac{1}{\alpha} m \left(u^{\frac{1}{m}}\right)^{m-1} e^{-\frac{(u^{1/m})^m}{\alpha}} \left|\dfrac{d}{du}(h^{-1})(u)\right| & y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Solution: (continued)

$$f_U(u) = \begin{cases} \frac{1}{\alpha} m \left(u^{\frac{1}{m}}\right)^{m-1} e^{-\frac{(u^{1/\cancel{m}})^{\cancel{m}}}{\alpha}} \left|\frac{d}{du}(h^{-1})(u)\right| & y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{1}{\alpha} m \left(u^{\frac{1}{m}}\right)^{m-1} e^{-\frac{u}{\alpha}} \cdot \frac{1}{m} u^{\left(\frac{1}{m}-1\right)} & u > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Solution: (continued)

$$f_U(u) = \begin{cases} \frac{1}{\alpha} m \left(u^{\frac{1}{m}}\right)^{m-1} e^{-\frac{(u^{1/\cancel{m}})^{\cancel{m}}}{\alpha}} \left| \frac{d}{du}(h^{-1})(u) \right| & y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{1}{\alpha} \cancel{m} \left(u^{\frac{1}{m}}\right)^{m-1} e^{-\frac{u}{\alpha}} \cdot \frac{1}{\cancel{m}} u^{\left(\frac{1}{m}-1\right)} & u > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{1}{\alpha} e^{-\frac{u}{\alpha}} u^{\left(1-\frac{1}{m}\right)} \cdot u^{\left(\frac{1}{m}-1\right)} & u > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Solution: (continued)

$$f_U(u) = \begin{cases} \frac{1}{\alpha} m \left(u^{\frac{1}{m}}\right)^{m-1} e^{-\frac{(u^{1/\cancel{m}})^{\cancel{m}}}{\alpha}} \left|\frac{d}{du}(h^{-1})(u)\right| & y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{1}{\alpha} \cancel{m} \left(u^{\frac{1}{m}}\right)^{m-1} e^{-\frac{u}{\alpha}} \cdot \frac{1}{\cancel{m}} u^{\left(\frac{1}{m}-1\right)} & u > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{1}{\alpha} e^{-\frac{u}{\alpha}} \cancel{u^{\left(1-\frac{1}{m}\right)}} \cdot \cancel{u^{\left(\frac{1}{m}-1\right)}} & u > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{1}{\alpha} e^{-\frac{u}{\alpha}} & u > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Solution: (continued)

$$f_U(u) = \begin{cases} \frac{1}{\alpha} m \left(u^{\frac{1}{m}}\right)^{m-1} e^{-\frac{(u^{1/\cancel{m}})^{\cancel{m}}}{\alpha}} \left|\frac{d}{du}(h^{-1})(u)\right| & y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{1}{\alpha} \cancel{m} \left(u^{\frac{1}{m}}\right)^{m-1} e^{-\frac{u}{\alpha}} \cdot \frac{1}{\cancel{m}} u^{\left(\frac{1}{m}-1\right)} & u > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{1}{\alpha} e^{-\frac{u}{\alpha}} \cancel{u^{\left(1-\frac{1}{m}\right)}} \cdot \cancel{u^{\left(\frac{1}{m}-1\right)}} & u > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{1}{\alpha} e^{-\frac{u}{\alpha}} & u > 0, \\ 0 & \text{otherwise.} \end{cases}$$

So U is exponential with parameter α.

Solution: (continued)

$$f_U(u) = \begin{cases} \frac{1}{\alpha} m \left(u^{\frac{1}{m}}\right)^{m-1} e^{-\frac{(u^{1/\cancel{m}})^{\cancel{m}}}{\alpha}} \left| \frac{d}{du}(h^{-1})(u) \right| & y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{1}{\alpha} \cancel{m} \left(u^{\frac{1}{m}}\right)^{m-1} e^{-\frac{u}{\alpha}} \cdot \frac{1}{\cancel{m}} u^{\left(\frac{1}{m}-1\right)} & u > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{1}{\alpha} e^{-\frac{u}{\alpha}} \cancel{u^{\left(1-\frac{1}{m}\right)}} \cdot \cancel{u^{\left(\frac{1}{m}-1\right)}} & u > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{1}{\alpha} e^{-\frac{u}{\alpha}} & u > 0, \\ 0 & \text{otherwise.} \end{cases}$$

So U is exponential with parameter α.

The text covers only the 4 distributions and 2 special cases mentioned in Chapter 4, because many others can be obtained from these by simple transformations.

Solution: (continued)

$$f_U(u) = \begin{cases} \dfrac{1}{\alpha} m \left(u^{\frac{1}{m}}\right)^{m-1} e^{-\frac{(u^{1/\cancel{m}})^{\cancel{m}}}{\alpha}} \left| \dfrac{d}{du}(h^{-1})(u) \right| & y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \dfrac{1}{\alpha} \cancel{m} \left(u^{\frac{1}{m}}\right)^{m-1} e^{-\frac{u}{\alpha}} \cdot \dfrac{1}{\cancel{m}} u^{\left(\frac{1}{m}-1\right)} & u > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \dfrac{1}{\alpha} e^{-\frac{u}{\alpha}} \cancel{u^{\left(1-\frac{1}{m}\right)}} \cdot \cancel{u^{\left(\frac{1}{m}-1\right)}} & u > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \dfrac{1}{\alpha} e^{-\frac{u}{\alpha}} & u > 0, \\ 0 & \text{otherwise.} \end{cases}$$

So U is exponential with parameter α.

The text covers only the 4 distributions and 2 special cases mentioned in Chapter 4, because many others can be obtained from these by simple transformations. In the book, you will find numerous exercises with good material: Poisson-Gamma relationship, Hazard Rates, etc.

Method of Moment Generating Functions

Method of MGF:

- Try to determine the MGF of $U = f(X, Y)$.
- Then try to "recognize" this MGF as one we already know (see tables in the text).
- Then, by uniqueness of MGF, we know the distribution of U.

This might go wrong: the MGF for U might not be one in our tables. But for simple functions $f(X, Y)$ e.g. $f(X, Y) = X + Y$ and the right RVs X, Y, we will be "lucky".

Example

Suppose Z is a standard normal RV.
Suppose Y is a normal RV with mean μ and variance σ^2.
Notice that the MGF of Y is $e^{\mu t+\frac{\sigma^2 t^2}{2}}$, and that of Z is $e^{(0)t+\frac{(1)^2 t^2}{2}} = e^{\frac{t^2}{2}}$.
What is the distribution of $X = \frac{Y-\mu}{\sigma}$?

Solution:

We use the method of MGFs.

Solution:

We use the method of MGFs.

$$m_X(t) = E\left[e^{tX}\right]$$

Solution:

We use the method of MGFs.

$$m_X(t) = E\left[e^{tX}\right] = E\left[e^{t\left(\frac{Y-\mu}{\sigma}\right)}\right]$$

Solution:

We use the method of MGFs.

$$m_X(t) = E\left[e^{tX}\right] = E\left[e^{t\left(\frac{Y-\mu}{\sigma}\right)}\right]$$
$$= E\left[e^{\left(\frac{t}{\sigma}\right)Y} e^{-\frac{\mu t}{\sigma}}\right]$$

Solution:

We use the method of MGFs.

$$
\begin{aligned}
m_X(t) = E\left[e^{tX}\right] &= E\left[e^{t\left(\frac{Y-\mu}{\sigma}\right)}\right] \\
&= E\left[e^{\left(\frac{t}{\sigma}\right)Y} e^{-\frac{\mu t}{\sigma}}\right] = e^{-\frac{\mu t}{\sigma}} \cdot E\left[e^{\left(\frac{t}{\sigma}\right)Y}\right]
\end{aligned}
$$

Solution:

We use the method of MGFs.

$$m_X(t) = E\left[e^{tX}\right] = E\left[e^{t\left(\frac{Y-\mu}{\sigma}\right)}\right]$$

$$= E\left[e^{\left(\frac{t}{\sigma}\right)Y}e^{-\frac{\mu t}{\sigma}}\right] = e^{-\frac{\mu t}{\sigma}} \cdot E\left[e^{\left(\frac{t}{\sigma}\right)Y}\right]$$

$$= e^{-\frac{\mu t}{\sigma}} m_Y\left(\frac{t}{\sigma}\right)$$

Solution:

We use the method of MGFs.

$$
\begin{aligned}
m_X(t) &= E\left[e^{tX}\right] = E\left[e^{t\left(\frac{Y-\mu}{\sigma}\right)}\right] \\
&= E\left[e^{\left(\frac{t}{\sigma}\right)Y} e^{-\frac{\mu t}{\sigma}}\right] = e^{-\frac{\mu t}{\sigma}} \cdot E\left[e^{\left(\frac{t}{\sigma}\right)Y}\right] \\
&= e^{-\frac{\mu t}{\sigma}} m_Y\left(\frac{t}{\sigma}\right) = e^{-\frac{\mu t}{\sigma}} e^{\left(\mu\frac{t}{\sigma} + \frac{\sigma^2\left(\frac{t}{\sigma}\right)^2}{2}\right)}
\end{aligned}
$$

Solution:

We use the method of MGFs.

$$
\begin{aligned}
m_X(t) &= E\left[e^{tX}\right] = E\left[e^{t\left(\frac{Y-\mu}{\sigma}\right)}\right] \\
&= E\left[e^{\left(\frac{t}{\sigma}\right)Y} e^{-\frac{\mu t}{\sigma}}\right] = e^{-\frac{\mu t}{\sigma}} \cdot E\left[e^{\left(\frac{t}{\sigma}\right)Y}\right] \\
&= e^{-\frac{\mu t}{\sigma}} m_Y\left(\frac{t}{\sigma}\right) = e^{-\frac{\mu t}{\sigma}} e^{\left(\mu\frac{t}{\sigma} + \frac{\cancel{\sigma^2}\left(\frac{t}{\cancel{\sigma}}\right)^2}{2}\right)} \\
&= e^{\left(-\frac{\mu t}{\sigma} + \frac{\mu t}{\sigma} + \frac{t^2}{2}\right)}
\end{aligned}
$$

Solution:

We use the method of MGFs.

$$
\begin{aligned}
m_X(t) &= E\left[e^{tX}\right] = E\left[e^{t\left(\frac{Y-\mu}{\sigma}\right)}\right] \\
&= E\left[e^{\left(\frac{t}{\sigma}\right)Y} e^{-\frac{\mu t}{\sigma}}\right] = e^{-\frac{\mu t}{\sigma}} \cdot E\left[e^{\left(\frac{t}{\sigma}\right)Y}\right] \\
&= e^{-\frac{\mu t}{\sigma}} m_Y\left(\frac{t}{\sigma}\right) = e^{-\frac{\mu t}{\sigma}} e^{\left(\mu \frac{t}{\sigma} + \frac{\cancel{\sigma}^2 \left(\frac{t}{\cancel{\sigma}}\right)^2}{2}\right)} \\
&= e^{\left(-\cancel{\frac{\mu t}{\sigma}} + \cancel{\frac{\mu t}{\sigma}} + \frac{t^2}{2}\right)} \\
&= e^{\frac{t^2}{2}}, \qquad \text{exactly the same as the MGF of } Z.
\end{aligned}
$$

Solution:

We use the method of MGFs.

$$\begin{aligned} m_X(t) &= E\left[e^{tX}\right] = E\left[e^{t\left(\frac{Y-\mu}{\sigma}\right)}\right] \\ &= E\left[e^{\left(\frac{t}{\sigma}\right)Y} e^{-\frac{\mu t}{\sigma}}\right] = e^{-\frac{\mu t}{\sigma}} \cdot E\left[e^{\left(\frac{t}{\sigma}\right)Y}\right] \\ &= e^{-\frac{\mu t}{\sigma}} m_Y\left(\frac{t}{\sigma}\right) = e^{-\frac{\mu t}{\sigma}} e^{\left(\mu \frac{t}{\sigma} + \frac{\cancel{\sigma^2}\left(\frac{t}{\cancel{\sigma}}\right)^2}{2}\right)} \\ &= e^{\left(-\cancel{\frac{\mu t}{\sigma}} + \cancel{\frac{\mu t}{\sigma}} + \frac{t^2}{2}\right)} \\ &= e^{\frac{t^2}{2}}, \qquad \text{exactly the same as the MGF of } Z. \end{aligned}$$

<u>Conclusion:</u> $X = \dfrac{Y-\mu}{\sigma}$ has the standard normal distribution.

Example

Let Z be a standard normal RV, and let $Y = Z^2$. What is the distribution of Y?

Solution:

We compute

$$
\begin{aligned}
m_Y(t) &= E\left[e^{tY}\right] = E\left[e^{tZ^2}\right] \\
&= \int_{-\infty}^{\infty} e^{tz^2} f(z)\,dz \qquad \text{where } f \text{ is the standard normal PDF} \\
&= \int_{-\infty}^{\infty} e^{tz^2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}}\,dz \qquad \text{where } \mu = 0, \sigma = 1 \\
&= \int_{-\infty}^{\infty} e^{tz^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz \qquad \begin{array}{l}\text{To integrate, use the trick:} \\ \text{The integral of a PDF is 1.}\end{array}
\end{aligned}
$$

If we can rearrange this into something that looks like (factor) $\cdot$ (normal PDF), then $\int$(factor)(PDF) $=$ (factor).

Solution: (continued)

$$\therefore \quad m_Y(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{z^2}{2}+tz^2\right)}\, dz.$$

Solution: (continued)

$$\therefore \quad m_Y(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{z^2}{2}+tz^2\right)}\, dz.$$

We'd like $-\dfrac{z^2}{2} + tz^2 = \dfrac{z^2}{2\sigma^2}$ for some σ.

Solution: (continued)

$$\therefore \quad m_Y(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{z^2}{2}+tz^2\right)}\, dz.$$

We'd like $-\dfrac{z^2}{2} + tz^2 = \dfrac{z^2}{2\sigma^2}$ for some σ. What does σ have to be?

Solution: (continued)

$$\therefore \quad m_Y(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{z^2}{2}+tz^2\right)}\, dz.$$

We'd like $-\dfrac{z^2}{2} + tz^2 = \dfrac{z^2}{2\sigma^2}$ for some σ. What does σ have to be?

$$-\frac{1}{2} + t = -\frac{1}{2\sigma^2}$$

Solution: (continued)

$$\therefore \quad m_Y(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{z^2}{2}+tz^2\right)}\, dz.$$

We'd like $-\dfrac{z^2}{2} + tz^2 = \dfrac{z^2}{2\sigma^2}$ for some σ. What does σ have to be?

$$-\frac{1}{2} + t = -\frac{1}{2\sigma^2} \implies \frac{1}{2} - t = \frac{1}{2\sigma^2}$$

Solution: (continued)

$$\therefore \quad m_Y(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{z^2}{2}+tz^2\right)}\, dz.$$

We'd like $-\dfrac{z^2}{2} + tz^2 = \dfrac{z^2}{2\sigma^2}$ for some σ. What does σ have to be?

$$-\frac{1}{2} + t = -\frac{1}{2\sigma^2} \implies \frac{1}{2} - t = \frac{1}{2\sigma^2}$$

$$\implies 1 - 2t = \frac{1}{\sigma^2}$$

Solution: (continued)

$$\therefore \quad m_Y(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{z^2}{2}+tz^2\right)}\, dz.$$

We'd like $-\dfrac{z^2}{2} + tz^2 = \dfrac{z^2}{2\sigma^2}$ for some σ. What does σ have to be?

$$-\frac{1}{2} + t = -\frac{1}{2\sigma^2} \implies \frac{1}{2} - t = \frac{1}{2\sigma^2}$$

$$\implies 1 - 2t = \frac{1}{\sigma^2} \implies \sigma^2 = \frac{1}{1-2t}.$$

Solution: (continued)

$$\therefore \quad m_Y(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{z^2}{2}+tz^2\right)}\, dz.$$

We'd like $-\dfrac{z^2}{2} + tz^2 = \dfrac{z^2}{2\sigma^2}$ for some σ. What does σ have to be?

$$-\frac{1}{2} + t = -\frac{1}{2\sigma^2} \implies \frac{1}{2} - t = \frac{1}{2\sigma^2}$$

$$\implies 1 - 2t = \frac{1}{\sigma^2} \implies \sigma^2 = \frac{1}{1-2t}.$$

$$\therefore \quad m_Y(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}}\, dz \qquad \text{where } \sigma^2 = \frac{1}{1-2t}.$$

Solution: (continued)

$$\therefore \quad m_Y(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{z^2}{2}+tz^2\right)}\, dz.$$

We'd like $-\frac{z^2}{2} + tz^2 = \frac{z^2}{2\sigma^2}$ for some σ. What does σ have to be?

$$-\frac{1}{2} + t = -\frac{1}{2\sigma^2} \implies \frac{1}{2} - t = \frac{1}{2\sigma^2}$$

$$\implies 1 - 2t = \frac{1}{\sigma^2} \implies \sigma^2 = \frac{1}{1-2t}.$$

$$\therefore \quad m_Y(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}}\, dz \qquad \text{where } \sigma^2 = \frac{1}{1-2t}.$$

$$\therefore \quad \frac{1}{\sigma} m_Y(t) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}}\, dz$$

Solution: (continued)

$$\therefore \quad m_Y(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{z^2}{2}+tz^2\right)}\, dz.$$

We'd like $-\dfrac{z^2}{2} + tz^2 = \dfrac{z^2}{2\sigma^2}$ for some σ. What does σ have to be?

$$-\frac{1}{2} + t = -\frac{1}{2\sigma^2} \implies \frac{1}{2} - t = \frac{1}{2\sigma^2}$$

$$\implies 1 - 2t = \frac{1}{\sigma^2} \implies \sigma^2 = \frac{1}{1-2t}.$$

$$\therefore \quad m_Y(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}}\, dz \qquad \text{where } \sigma^2 = \frac{1}{1-2t}.$$

$$\therefore \quad \frac{1}{\sigma} m_Y(t) = \int_{-\infty}^{\infty} \underbrace{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}}}_{\text{normal PDF}}\, dz = 1.$$

Solution: (continued)

$$\therefore \quad m_Y(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{z^2}{2}+tz^2\right)}\, dz.$$

We'd like $-\dfrac{z^2}{2} + tz^2 = \dfrac{z^2}{2\sigma^2}$ for some σ. What does σ have to be?

$$-\frac{1}{2} + t = -\frac{1}{2\sigma^2} \implies \frac{1}{2} - t = \frac{1}{2\sigma^2}$$

$$\implies 1 - 2t = \frac{1}{\sigma^2} \implies \sigma^2 = \frac{1}{1-2t}.$$

$$\therefore \quad m_Y(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}}\, dz \qquad \text{where } \sigma^2 = \frac{1}{1-2t}.$$

$$\therefore \quad \frac{1}{\sigma} m_Y(t) = \int_{-\infty}^{\infty} \underbrace{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}}}_{\text{normal PDF}}\, dz = 1.$$

Thus $m_Y(t) = \sigma = \dfrac{1}{(1-2t)^{1/2}}$.

Solution: (continued)

$$\therefore \quad m_Y(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{z^2}{2}+tz^2\right)}\, dz.$$

We'd like $-\dfrac{z^2}{2} + tz^2 = \dfrac{z^2}{2\sigma^2}$ for some σ. What does σ have to be?

$$-\frac{1}{2} + t = -\frac{1}{2\sigma^2} \implies \frac{1}{2} - t = \frac{1}{2\sigma^2}$$

$$\implies 1 - 2t = \frac{1}{\sigma^2} \implies \sigma^2 = \frac{1}{1-2t}.$$

$$\therefore \quad m_Y(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}}\, dz \qquad \text{where } \sigma^2 = \frac{1}{1-2t}.$$

$$\therefore \quad \frac{1}{\sigma} m_Y(t) = \int_{-\infty}^{\infty} \underbrace{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}}}_{\text{normal PDF}}\, dz = 1.$$

Thus $m_Y(t) = \sigma = \dfrac{1}{(1-2t)^{1/2}}$. This is the MGF for a Gamma RV with $\alpha = \frac{1}{2}$ and $\beta = 2$. The same is a MGF of a $\chi^2[1]$ RV.

Solution: (continued)

Conclusion: The distribution of $Y = Z^2$ is $\Gamma\left(\frac{1}{2}, 2\right)$ which is the same as $\chi^2[1]$.

Problem:

In the above setting, find $E[Z^4]$.

Solution:

We have just shown that if Z is standard normal, then Z^2 is $\chi^2[1]$.
$\therefore \quad E[Z^4] = E[(Z^2)^2] = V[Z^2] + E[Z^2]^2$ using $V[X] = E[X^2] - E[X]^2$.
Now let $X = Z^2$.
Then

$$V[Z^2] = \alpha\beta^2, \quad E[Z^2] = \alpha\beta, \qquad \text{where } \alpha = \frac{1}{2}, \beta = 2.$$

$$\therefore \quad E[Z^4] = E[(Z^2)^2] = \alpha\beta^2 + (\alpha\beta)^2 = \frac{1}{2}2^2 + \left(\frac{1}{2}2\right)^2 = \boxed{3}.$$

Example

Suppose X, Y are RVs with $X \sim \Gamma(\alpha_1, \beta)$ and $Y \sim \Gamma(\alpha_2, \beta)$, and X, Y are independent.
Let $U = X + Y$.
What is the distribution of U?

Solution:

Note that

$$\begin{aligned} m_U(t) &= E[e^{tU}] = E[e^{t(X+Y)}] = E[e^{tX}e^{tY}] \\ &= E[e^{tX}]E[e^{tY}] \quad \text{(by independence)} \quad = m_X(t)M_Y(t) \\ &= \frac{1}{(1-\beta t)^{\alpha_1}} \cdot \frac{1}{(1-\beta t)^{\alpha_2}} = \frac{1}{(1-\beta t)^{\alpha_1+\alpha_2}}. \end{aligned}$$

This is the MGF of a $\Gamma(\alpha_1 + \alpha_2, \beta)$ RV.
So if $X \sim \Gamma(\alpha_1, \beta)$ and $Y \sim \Gamma(\alpha_2, \beta)$ are independent, then $U = X + Y \sim \Gamma(\alpha_1 + \alpha_2, \beta)$.

Theorem (6.3)

Let $Y_1, \ldots, Y_n$ be independent normally distributed random variables with $E[Y_i] = \mu_i$ and $V[Y_i] = \sigma_i^2$, for $i = 1, \ldots, n$, and let $a_1, \ldots, a_n$ be constants. If

$$U = \sum_{i=1}^{n} a_i Y_i,$$

then U is a normally distributed random variable with

$$E[U] = \sum_{i=1}^{n} a_i \mu_i \qquad \text{and} \qquad V[U] = \sum_{i=1}^{n} a_i^2 \sigma_i^2.$$

Sketch of Proof:

The normally distributed RV Y_i has the MGF

$$m_{Y_i}(t) = e^{\left(\mu_i t + \frac{\sigma_i^2 t^2}{2}\right)}$$

for each $i = 1, \ldots, n$.

So the RV $a_i Y_i$ has the MGF ____. (Find it!)

Now use the independence of Y_i (thus that of $a_i Y_i$) to find

$$m_U(t) = \prod_{i=1}^{n} m_{a_i Y_i}(t) = e^{\left(t \sum_{i=1}^{n} a_i \mu_i + \frac{t^2}{2} \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)}. \quad \text{(Verify!)}$$

By uniqueness of MGF, U is a normally distributed random variable with

$$E[U] = \sum_{i=1}^{n} a_i \mu_i \qquad \text{and} \qquad V[U] = \sum_{i=1}^{n} a_i^2 \sigma_i^2.$$

$\square$

Theorem (6.4)

If $Z_1, \dots, Z_n$ are independent standard normal RVs, then $U = Z_1^2 + \cdots + Z_n^2$ has the distribution $\chi^2[n]$. (Same as $\Gamma\left(\frac{n}{2}, 2\right)$.)

Proof:

We use the method of MGFs.

$$\text{Claim:} \quad m_U(t) = \frac{1}{(1-2t)^{n/2}}, \quad \text{the MGF of } \Gamma\left(\frac{n}{2}, 2\right).$$

$$\begin{aligned} m_U(t) &= E[e^{tU}] = E[e^{t(Z_1^2+\cdots+Z_n^2)}] \\ &= E[e^{tZ_1^2} \dots e^{tZ_n^2}] \\ &= E[e^{tZ_1^2}] \dots E[e^{tZ_n^2}] \qquad \text{(by independence)} \\ &= \underbrace{\frac{1}{(1-2t)^{1/2}} \cdot \dots \cdot \frac{1}{(1-2t)^{1/2}}}_{n \text{ times}} \qquad \text{(as each } Z_i^2 \text{ has the } \chi^2[1] \text{ distribution)} \\ &= \frac{1}{(1-2t)^{n/2}}. \end{aligned}$$

Multivariate Transformations using Jacobians

Let's consider the case of two random variables first.

The Bivariate Transform Method

Suppose that Y_1 and Y_2 are continuous random variables with joint density function $f_{Y_1,Y_2}(y_1, y_2)$ and that for all (y_1, y_2), such that $f_{Y_1,Y_2}(y_1, y_2) > 0$,

$$u_1 = h_1(y_1, y_2) \quad \text{and} \quad u_2 = h_2(y_1, y_2)$$

is a one-to-one transformation from (y_1, y_2) to (u_1, u_2) with inverse

$$y_1 = h_1^{-1}(u_1, u_2) \quad \text{and} \quad y_2 = h_2^{-1}(u_1, u_2).$$

If $h_1^{-1}(u_1, u_2)$ and $h_2^{-1}(u_1, u_2)$ have continuous partial derivatives with respect to u_1 and u_2 and the *Jacobian*

$$J = \det \begin{bmatrix} \partial h_1^{-1}/\partial u_1 & \partial h_1^{-1}/\partial u_2 \\ \partial h_2^{-1}/\partial u_1 & \partial h_2^{-1}/\partial u_2 \end{bmatrix} = \frac{\partial h_1^{-1}}{\partial u_1}\frac{\partial h_2^{-1}}{\partial u_2} - \frac{\partial h_2^{-1}}{\partial u_1}\frac{\partial h_1^{-1}}{\partial u_2} \neq 0,$$

then the joint density of U_1 and U_2 is

$$f_{U_1,U_2}(u_1, u_2) = f_{Y_1,Y_2}\left(h_1^{-1}(u_1, u_2), h_2^{-1}(u_1, u_2)\right)|J|,$$

where $|J|$ is the absolute value of J.

The transformation follows from calculus results used for change of variables in multiple integration.

The absolute value of the Jacobian, $|J|$, in the multivariate transformation is analogous to the quantity $\left|\frac{dh^{-1}(u)}{du}\right|$ that is used when making the one-variable transformation $U = h(Y)$.

Caution:

Be sure that the bivariate transformation $u_1 = h_1(y_1, y_2)$, $u_2 = h_2(y_1, y_2)$ is a one-to-one transformation for all (y_1, y_2) such that $f_{Y_1,Y_2}(y_1, y_2) > 0$. If not, then the resulting "density" function will not have the necessary properties of a valid density function.

Let's use this method for the following example:

Example 6.13

Let Y_1 and Y_2 be independent standard normal random variables. If $U_1 = Y_1 + Y_2$ and $U_2 = Y_1 - Y_2$, then what is the joint density of U_1 and U_2?

Solution:

The density functions for Y_1 and Y_2 are

$$f_1(y_1) = \frac{e^{-\frac{1}{2}y_1^2}}{\sqrt{2\pi}}, \quad f_2(y_2) = \frac{e^{-\frac{1}{2}y_2^2}}{\sqrt{2\pi}}, \quad \begin{array}{l} -\infty < y_1 < \infty, \\ -\infty < y_2 < \infty, \end{array}$$

and the independence of Y_1 and Y_2 implies that their joint density is

$$f_{Y_1,Y_2}(y_1, y_2) = \frac{1}{2\pi} e^{-\frac{1}{2}(y_1^2+y_2^2)}, \qquad \begin{array}{l} -\infty < y_1 < \infty, \\ -\infty < y_2 < \infty. \end{array}$$

In this case $f_{Y_1,Y_2}(y_1, y_2) > 0$ for all $-\infty < y_1 < \infty$ and $-\infty < y2 < \infty$. We are interested in the transformation

$$u_1 = y_1 + y_2 = h_1(y_1, y_2) \quad \text{and} \quad u_2 = y_1 - y_2 = h_2(y_1, y_2),$$

with the inverse transformation

$$y_1 = \frac{u_1 + u_2}{2} = h_1^{-1}(u_1, u_2) \quad \text{and} \quad y_2 = \frac{u_1 - u_2}{2} h_2^{-1}(u_1, u_2).$$

Because $\frac{\partial h_1^{-1}}{\partial u_1} = \frac{1}{2}$, $\frac{\partial h_1^{-1}}{\partial u_2} = \frac{1}{2}$, $\frac{\partial h_2^{-1}}{\partial u_1} = \frac{1}{2}$, and $\frac{\partial h_2^{-1}}{\partial u_2} = -\frac{1}{2}$, the Jacobian of this transformation is

Solution: (continued)

$$J = \det \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix} = \left(\frac{1}{2}\right)\left(-\frac{1}{2}\right) - \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = -\frac{1}{2} \neq 0,$$

and the joint density of U_1 and U_2 is

$$f_{U_1,U_2}(u_1,u_2) = \frac{1}{2\pi} e^{-\frac{1}{2}\left[\left(\frac{u_1+u_2}{2}\right)^2 + \left(\frac{u_1-u_2}{2}\right)^2\right]} \left|-\frac{1}{2}\right|, \quad \begin{array}{l} -\infty < \frac{u_1+u_2}{2} < \infty, \\ -\infty < \frac{u_1-u_2}{2} < \infty. \end{array}$$

A little algebra manipulation yields

$$f_{U_1,U_2}(u_1,u_2) = \frac{e^{-\frac{1}{2}\left(\frac{u_1^2}{2}\right)}}{\sqrt{2}\sqrt{2\pi}} \frac{e^{-\frac{1}{2}\left(\frac{u_2^2}{2}\right)}}{\sqrt{2}\sqrt{2\pi}}, \quad \begin{array}{l} -\infty < u_1 < \infty, \\ -\infty < u_2 < \infty. \end{array}$$

Notice that U_1 and U_2 are *independent* and normally distributed, both with mean 0 and variance 2.

The extra information provided by the joint distribution of U_1 and U_2 is that the two variables are independent!

The k-variate Transformation

If $Y_1, \ldots, Y_k$ are jointly continuous random variables and

$$U_1 = h_1(Y_1, \ldots, Y_k), \quad \ldots \quad , U_k = h_k(Y_1, \ldots, Y_k),$$

where the transformation

$$u_1 = h_1(y_1, \ldots, y_k), \quad \ldots \quad , u_k = h_k(y_1, \ldots, y_k)$$

is a one-to-one transformation from $(y_1, \ldots, y_k)$ to $(u_1, \ldots, u_k)$, which has the inverse transformations

$$y_1 = h_1^{-1}(u_1, \ldots, u_k), \quad \ldots \quad , y_k = h_k^{-1}(u_1, \ldots, u_k),$$

such that $h_1^{-1}(u_1, \ldots, u_k), \ldots, h_k^{-1}(u_1, \ldots, u_k)$ have continuous partial derivatives with respect to $u_1, \ldots, u_k$,
and the Jacobian

$$J = \det \begin{bmatrix} \partial h_1^{-1}/\partial u_1 & \cdots & \partial h_1^{-1}/\partial u_k \\ \vdots & \ddots & \vdots \\ \partial h_k^{-1}/\partial u_1 & \cdots & \partial h_k^{-1}/\partial u_k \end{bmatrix} \neq 0,$$

then there is a result analogous to the bivariate case that can be used to find the joint density of $U_1, \ldots, U_k$.

Why does this matter for statistics?

If $Y_1, \ldots, Y_n$ are independent normal RVs with means μ_i and variances σ_i^2, we can write $Z_i = \dfrac{Y_i - \mu_i}{\sigma_i}$
and compute

$$Z_1^2 + \cdots + Z_n^2 \qquad \text{(a sum of squared normalized "errors").}$$

We know the distribution of this quantity.

So we can do hypothesis testing by computing this quantity, and seeing how the results compare to the predicted distribution.

"Order Statistics"

Suppose we have $Y_1, \ldots, Y_n$ independent and identically distributed (IID) RVs.
We could write $Y_{(1)}$ for the smallest, $Y_{(n)}$ for the largest; so

$$Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)};$$

where $\quad Y_{(1)} = \min\{Y_1, \ldots, Y_n\},\ Y_{(n)} = \max\{Y_1, \ldots, Y_n\}.$
What is the distribution of $Y_{(1)}, \ldots, Y_{(n)}$?
The answer is given by Theorem 6.5:

Definition

A *statistic* is a function of the observable random variables in a sample and known constants.

Theorem (6.5)

Let $Y_1, \ldots, Y_n$ be independent identically distributed continuous random variables with common distribution function $F(y)$ and common density function $f(y)$.

If $Y_{(k)}$ denotes the k^{th}-order statistic, then the density function of $Y_{(k)}$ is given by

$$g_{(k)}(y_k) = \frac{n!}{(k-1)!(n-k)!}[F(y_k)]^{k-1}[1-F(y_k)]^{n-k} f(y).$$

If j and k are two integers such that $1 \leq j < k \leq n$, the joint density of $Y_{(j)}$ and $Y_{(k)}$, for $y_j < y_k$, is given by

$$\begin{aligned} g_{(j)(k)}(y_j, y_k) = &\frac{n!}{(j-1)!(k-1-j)!(n-k)!}[F(y_j)]^{j-1} \\ &\times [F(y_k) - F(y_j)]^{k-1-j} \times [1-F(y_k)]^{n-k} f(y_j) f(y_k). \end{aligned}$$

We will look at the simplest cases $Y_{(n)}$ and $Y_{(1)}$.

Proof:

Assume that the distribution of each Y_i is known, with CDF $F(y)$ and PDF $f(y)$.
What is the distribution of $Y_{(n)}$?

$$\text{distribution of } Y_{(n)} = G_{(n)}(y) := P(Y_{(n)} \leq y).$$

Write $g_{(n)}(y)$ for the PDF of $Y_{(n)}$.
Then

$$\begin{aligned} G_{(n)}(y) &= P(Y_{(n)} \leq y) = P(\max\{Y_1, \ldots, Y_n\} \leq y) \\ &= P(Y_1 \leq y \text{ and } \ldots \text{ and } Y_n \leq y) \\ &= P(Y_1 \leq y) \cdot \cdots \cdot P(Y_n \leq y) \qquad \text{(by independence)} \\ &= \underbrace{F(y) \cdot \cdots \cdot F(y)}_{n \text{ times}} = (F(y))^n . \end{aligned}$$

So $G_{(n)}(y) = (F(y))^n$.
It follows that

$$g_{(n)}(y) = G'_{(n)}(y) = n\,(F(y))^{n-1}\,f(y).$$

What about the CDF and PDF for $Y_{(1)} = \min\{Y_1, \ldots, Y_n\}$?
Work it out in a similar manner as above to discover

$$G_{(1)}(y) = 1 - (1 - F(y))^n , \quad \text{and} \quad g_{(1)}(y) = n\,(1 - F(y))^{n-1}\,f(y).$$

Exercise: (steps to find distributions of the minimum)

Let $Y_1, \ldots, Y_n$ be independent identically distributed continuous random variables with common distribution function $F(y)$ and common density function $f(y)$. Let $Y_m = \min\{Y_1, \ldots, Y_n\}$.

(1) In terms of the distribution function F, what is $P(Y_1 > y)$?

(2) In terms of the distribution function F, what is $P(Y_m > y)$?

(3) In terms of the distribution function F, what is $P(Y_m \leq y)$?

(4) Find the probability density function f_m of Y_m in terms of F and f.

Solutions:

(1) $P(Y_1 > y) = 1 - P(Y_1 \le y) = \boxed{1 - F(y)}$.

(2)

$$Y_m > y \iff Y_1 > y \text{ and } \dots \text{ and } Y_n > y.$$

Therefore

$$\begin{aligned} P(Y_m > y) &= P(Y_1 > y \text{ and } \dots \text{ and } Y_n > y) \\ &= P(Y_1 > y) \cdot \dots \cdot P(Y_n > y) \qquad \text{(by independence)} \\ &= \underbrace{(1 - F(y)) \cdot \dots \cdot (1 - F(y))}_{n \text{ times}} = \boxed{(1 - F(y))^n}. \end{aligned}$$

(3) $P(Y_m \le y) = 1 - P(Y_m > y) = \boxed{1 - (1 - F(y))^n}$.

(4)

$$\begin{aligned} f_m(y) &= \frac{d}{dy} F_m(y) = \frac{d}{dy}\left[1 - (1 - F(y))^n\right] \\ &= -n[1 - F(y)]^{n-1} \frac{d}{dy}(-F(y)) = \boxed{n\,(1 - F(y))^{n-1} f(y)}. \end{aligned}$$

Exercise 6.6:

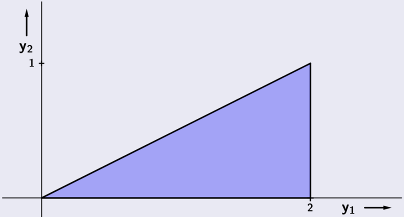
The joint distribution of amount of pollutant emitted from a smokestack without a cleaning device (Y_1) and a similar smokestack with a cleaning device (Y_2) is

$$f(y_1, y_2) = \begin{cases} 1 & 0 \le y_1 \le 2, 0 \le y \le 1 \text{ and } 2y_2 \le y_1, \\ 0 & \text{elsewhere.} \end{cases}$$

The reduction in amount of pollutant due to the cleaning device is given by $U = Y_1 - Y_2$.
Find the probability density function for U.

Solution:

Note that the region where the PDF $f(y_1, y_2) \neq 0$ is as shown along:



Solution: (continued)

Now we find the PDF for $U = Y_1 - Y_2$.

Solution: (continued)

Now we find the PDF for $U = Y_1 - Y_2$.

$$F_U(u) = P(U \leq u)$$

Solution: (continued)

Now we find the PDF for $U = Y_1 - Y_2$.

$$F_U(u) = P(U \leq u) = P(Y_1 - Y_2 \leq u)$$

Solution: (continued)

Now we find the PDF for $U = Y_1 - Y_2$.

$$F_U(u) = P(U \leq u) = P(Y_1 - Y_2 \leq u) = P(Y_2 \geq Y_1 - u).$$

Solution: (continued)

Now we find the PDF for $U = Y_1 - Y_2$.

$$F_U(u) = P(U \leq u) = P(Y_1 - Y_2 \leq u) = P(Y_2 \geq Y_1 - u).$$

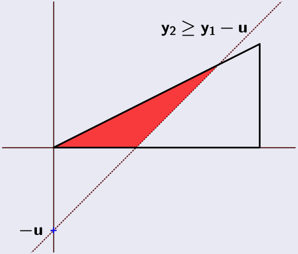
If $0 \leq u \leq 1$, the region looks like this:

Solution: (continued)

Now we find the PDF for $U = Y_1 - Y_2$.

$$F_U(u) = P(U \leq u) = P(Y_1 - Y_2 \leq u) = P(Y_2 \geq Y_1 - u).$$

If $0 \leq u \leq 1$, the region looks like this:

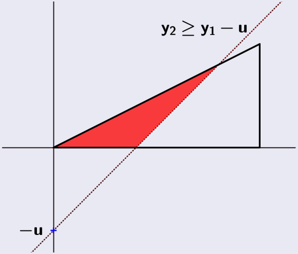


Solution: (continued)

Now we find the PDF for $U = Y_1 - Y_2$.

$$F_U(u) = P(U \leq u) = P(Y_1 - Y_2 \leq u) = P(Y_2 \geq Y_1 - u).$$

If $0 \leq u \leq 1$, the region looks like this:
How does the picture change depending on the value of u?

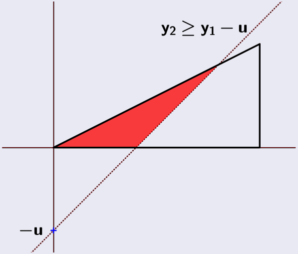


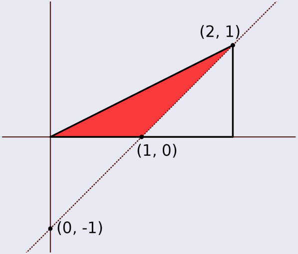
Solution: (continued)

Now we find the PDF for $U = Y_1 - Y_2$.

$$F_U(u) = P(U \le u) = P(Y_1 - Y_2 \le u) = P(Y_2 \ge Y_1 - u).$$

If $0 \le u \le 1$, the region looks like this:
How does the picture change depending on the value of u?





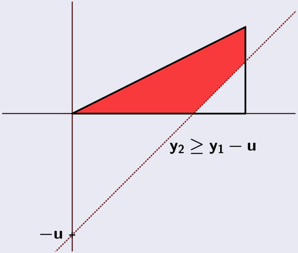
This picture shows that $u = 1$ is the "transition".

Solution: (continued)

If $1 \leq u \leq 2$, this is how the region looks:

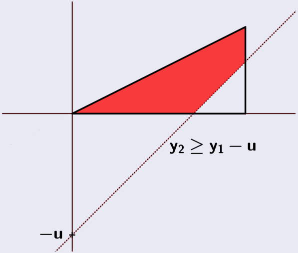
Solution: (continued)

If $1 \leq u \leq 2$, this is how the region looks:



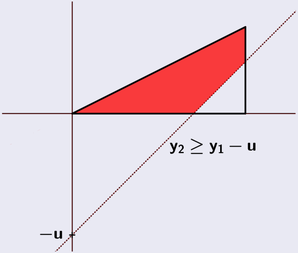
Solution: (continued)

If $1 \leq u \leq 2$, this is how the region looks:
The area of the region is 1 minus the area of the small triangle.



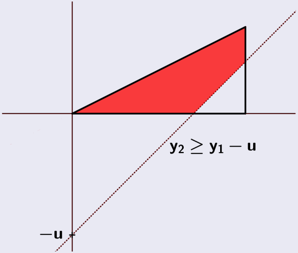
Solution: (continued)

If $1 \leq u \leq 2$, this is how the region looks:
The area of the region is 1 minus the area of the small triangle.
The area of the shaded region, as a function of u, is the PDF of U.



Solution: (continued)

If $1 \leq u \leq 2$, this is how the region looks:
The area of the region is 1 minus the area of the small triangle.
The area of the shaded region, as a function of u, is the PDF of U.



Example 6.4

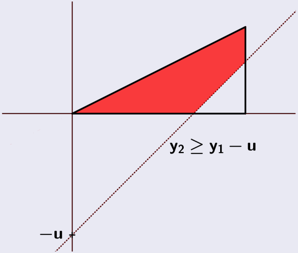
Let Y have probability density function given by

$$f_Y(y) = \begin{cases} \frac{y+1}{2} & -1 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find the density function for $U = Y^2$.

Solution: (continued)

If $1 \leq u \leq 2$, this is how the region looks:
The area of the region is 1 minus the area of the small triangle.
The area of the shaded region, as a function of u, is the PDF of U.



Example 6.4

Let Y have probability density function given by

$$f_Y(y) = \begin{cases} \frac{y+1}{2} & -1 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find the density function for $U = Y^2$.

Answer: $f_U(u) = \begin{cases} \frac{1}{2\sqrt{u}} & 0 < u \leq 1, \\ 0 & \text{elsewhere.} \end{cases}$

End of Chapter 6

Chapter 7

Sampling Distributions and the Central Limit Theorem

Sampling Distributions related to the Normal Distribution

Theorem (7.1)

Let $Y_1, \ldots, Y_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2.
Then

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

is normally distributed with mean $\mu_{\overline{Y}} = \mu$ and variance $\sigma^2_{\overline{Y}} = \sigma^2/n$.

Proof:

Because $Y_1, \ldots, Y_n$ is a random sample from a normal distribution with mean μ and variance σ^2, $Y_i, i = 1, \ldots, n$, are independent, normally distributed variables, with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2$.
Further,

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{Y_1}{n} + \cdots + \frac{Y_n}{n} = a_1 Y_1 + \cdots + a_n Y_n,$$

where $a_i = 1/n$, $i = 1, \ldots, n$.

Proof: (continued)

By Theorem 6.3, we conclude that $\overline{Y}$ is normally distributed with

$$E[\overline{Y}] = E\left[\frac{Y_1}{n} + \cdots + \frac{Y_n}{n}\right] = \underbrace{\frac{\mu}{n} + \cdots + \frac{\mu}{n}}_{n \text{ times}} = \mu,$$

and

$$V[\overline{Y}] = V\left[\frac{Y_1}{n} + \cdots + \frac{Y_n}{n}\right] = \underbrace{\frac{\sigma^2}{n^2} + \cdots + \frac{\sigma^2}{n^2}}_{n \text{ times}} = \frac{\sigma^2}{n}.$$

Remark:

Under the conditions of Theorem 7.1, $\overline{Y}$ is normally distributed with mean $\mu_{\overline{Y}} = \mu$ and variance $\sigma^2_{\overline{Y}} = \sigma^2/n$.
It follows that

$$Z = \frac{\overline{Y} - \mu_{\overline{Y}}}{\sigma_{\overline{Y}}} = \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}}$$

has the standard normal distribution.

Remarks:

- Notice that the variance of each of the random variables $Y_1, \ldots, Y_n$ is σ^2 and that of the sampling distribution of the random variable $\overline{Y}$ is σ^2/n.
- With $\overline{Y}$ as in Theorem 7.1, it follows that

$$Z = \frac{\overline{Y} - \mu_{\overline{Y}}}{\sigma_{\overline{Y}}} = \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} = \sqrt{n}\left(\frac{\overline{Y} - \mu}{\sigma}\right)$$

has a standard normal distribution.

Example 7.2:

A bottling machine can be regulated so that it discharges an average of μ ounces per bottle. It has been observed that the amount of fill dispensed by the machine is normally distributed with $\sigma = 1.0$ ounce. A sample of $n = 9$ filled bottles is randomly selected from the output of the machine on a given day (all bottled with the same machine setting), and the ounces of fill are measured for each. Find the probability that the sample mean will be within .3 ounce of the true mean μ for the chosen machine setting.

Solution:

If $Y_1, \ldots, Y_9$ denote the ounces of fill to be observed, then we know that the Y_is are normally distributed with mean μ and variance $\sigma^2 = 1$ for $i = 1, \ldots, 9$.

Therefore, by Theorem 7.1, Y possesses a normal sampling distribution with mean $\mu_Y = \mu$ and variance $\sigma_Y^2 = {}^{\sigma}/_{\sqrt{n}} = {}^{1}/_{9}$.

We want to find

$$\begin{aligned} P(|\overline{Y} - \mu| \leq 0.3) &= P(-0.3 \leq \overline{Y} - \mu \leq 0.3) \\ &= P\left(-\frac{0.3}{\sigma/\sqrt{n}} \leq \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} \leq \frac{0.3}{\sigma/\sqrt{n}}\right). \end{aligned}$$

Because $\dfrac{\overline{Y} - \mu_{\overline{Y}}}{\sigma_{\overline{Y}}} = \dfrac{\overline{Y} - \mu}{\sigma/\sqrt{n}}$ has a standard normal distribution, it follows that

$$P(|\overline{Y} - \mu| \leq 0.3) = P\left(-\frac{0.3}{1/\sqrt{9}} \leq Z \leq \frac{0.3}{1/\sqrt{9}}\right) = P(-0.9 \leq Z \leq 0.9).$$

Using Table 4, Appendix 3, we find

$$P(-0.9 \leq Z \leq 0.9) = 1 - 2P(Z > 0.9) = 1 - 2(0.1841) = 0.6318.$$

Thus, the probability is only .6318 that the sample mean will be within .3 ounce of the true population mean.

Example 7.3:

Refer to Example 7.2. How many observations should be included in the sample if we wish $\overline{Y}$ to be within .3 ounce of μ with probability .95?

Solution:

Now we want

$$P(|\overline{Y} - \mu| \le 0.3) = P(-0.3 \le \overline{Y} - \mu \le 0.3) = 0.95.$$

Divide each term of the inequality by $\sigma_{\overline{Y}} = {}^{\sigma}/_{\sqrt{n}}$ to get

$$P\left(-\frac{0.3}{\sigma/\sqrt{n}} \le \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} \le \frac{0.3}{\sigma/\sqrt{n}}\right) = P(-0.3\sqrt{n} \le Z \le 0.3\sqrt{n}) = 0.95.$$

(Recall that $\sigma = 1$).

But using Table 4, Appendix 3, we obtain $P(-1.96 \le Z \le 1.96) = 0.95$.

It must follow that

$$0.3\sqrt{n} = 1.96 \implies n = \left(\frac{1.96}{0.3}\right)^2 \approx 42.68.$$

Practically, it is impossible to take a sample of size 42.68.

Our solution indicates that a sample of size 42 is not quite large enough to reach our objective.

If $n = 43$, $P(|\overline{Y} - \mu| \le 0.3)$ slightly exceeds 0.95

Theorem (7.2)

Let $Y_1, \ldots, Y_n$ be as in Theorem 7.1. Then $Z_i = \dfrac{Y_i - \mu}{\sigma}$ are independent standard normal random variables, $i = 1, \ldots, n$, and

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{Y_i - \mu}{\sigma} \right)^2$$

has a χ^2 distribution with n degrees of freedom.

Proof.

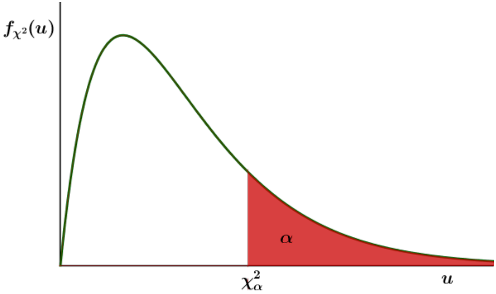
Because $Y_1, \ldots, Y_n$ is a random sample from a normal distribution with mean μ and variance σ^2, $Z_i = \dfrac{Y_i - \mu}{\sigma}$ has a standard normal distribution for $i = 1, \ldots, n$.

Further, the random variables Z_i are independent as the random variables Y_i are independent, $i = 1, \ldots, n$.

It follows directly from Theorem 6.4 that $\sum_{i=1}^{n} Z_i^2$ has the distribution $\chi^2[n]$. $\square$

Remark:

From Table 6, Appendix 3, we can find values χ^2_α so that $P(\chi^2 > \chi^2_\alpha) = \alpha$, that is, $P(\chi^2 \leq \chi^2_\alpha) = 1 - \alpha$.
Thus χ^2_α is the $(1-\alpha)$ quantile of the χ^2 RV.



The following example illustrates the combined use of Theorem 7.2 and the χ^2 tables.

Example 7.4:

If $Z_1, \dots, Z_6$ denotes a random sample from the standard normal distribution, find a number b such that

$$P\left(\sum_{i=1}^{6} Z_i^2 \le b\right) = 0.95.$$

Solution:

By Theorem 7.2, $\sum_{i=1}^{6} Z_i^2$ has the distribution $\chi^2[6]$.

Looking at Table 6, Appendix 3, in the row headed 6 df and the column headed $\chi^2_{.05}$, we see the number 12.5916.

Thus

$$P\left(\sum_{i=1}^{6} Z_i^2 > 12.5916\right) = 0.05 \iff P\left(\sum_{i=1}^{6} Z_i^2 \le 12.5916\right) = 0.95,$$

and $b = 12.5916$ is the .95 quantile (95^{th} percentile) of the sum of the squares of six independent standard normal random variables.

The χ^2 distribution plays an important role in many inferential procedures.
For example, suppose that we wish to make an inference about the population variance σ^2 based on a random sample $Y_1, \ldots, Y_n$ from a normal population.
A good estimator of σ^2 is the sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2.$$

The following theorem gives the probability distribution for a function of the statistic S^2.

Theorem (7.3)

Let $Y_1, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and variance σ^2. Then

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

has the distribution $\chi^2[n-1]$.
Also $\overline{Y}$ and S^2 are independent random variables.

Proof:

For simplicity, we only consider the case $n = 2$, and show that $\dfrac{(n-1)S^2}{\sigma^2}$ has the distribution $\chi^2[1]$.

In case $n = 2$, $\overline{Y} = \dfrac{Y_1 + Y_2}{2}$,

and, therefore,

$$S^2 = \frac{1}{2-1}\sum_{i=1}^{2}(Y_i - \overline{Y})^2 = \left[Y_1 - \frac{Y_1 + Y_2}{2}\right]^2 + \left[Y_2 - \frac{Y_1 + Y_2}{2}\right]^2$$

$$= \left[\frac{Y_1 - Y_2}{2}\right]^2 + \left[\frac{Y_2 - Y_1}{2}\right]^2 = 2\left[\frac{Y_1 - Y_2}{2}\right]^2 = \frac{(Y_1 - Y_2)^2}{2}.$$

It follows that, when $n = 2$,

$$\frac{(n-1)S^2}{\sigma^2} = \frac{(Y_1 - Y_2)^2}{2\sigma^2} = \left(\frac{Y_1 - Y_2}{\sqrt{2\sigma^2}}\right)^2.$$

We will show that this quantity is equal to the square of a standard normal random variable; that is, it is a Z^2, which possesses the distribution $\chi^2[1]$.

Proof: (continued)

Because $Y_1 - Y_2$ is a linear combination of independent, normally distributed random variables ($Y_1 - Y_2 = a_1 Y_1 + a_2 Y_2$ with $a_1 = 1$ and $a_2 = -1$),
Theorem 6.3 tells us that $Y_1 - Y_2$ has a normal distribution with mean $1\mu - 1\mu = 0$ and variance $(1)^2\sigma^2 + (-1)^2\sigma^2 = 2\sigma^2$.
Therefore, $Z = \dfrac{Y_1 - Y_2}{\sqrt{2\sigma^2}}$ has a standard normal distribution.
Because for $n = 2$,

$$\frac{(n-1)S^2}{\sigma^2} = \left(\frac{Y_1 - Y_2}{\sqrt{2\sigma^2}}\right)^2 = Z^2,$$

it follows that $\dfrac{(n-1)S^2}{\sigma^2}$ has the distribution $\chi^2[1]$.

In Example 6.13, we proved that $U_1 = \dfrac{Y_1 + Y_2}{\sigma}$ and $U_2 = \dfrac{Y_1 - Y_2}{\sigma}$ are independent.
Notice that, because $n = 2$,

$$\overline{Y} = \frac{Y_1 + Y_2}{2} = \frac{\sigma U_1}{2}, \quad S^2 = \frac{(Y_1 - Y_2)^2}{2} = \frac{(\sigma U_2)^2}{2}.$$

Because $\overline{Y}$ is a function of only U_1 and S^2 is a function of only U_2, the

The t-Distribution

Definition (7.2)

Let Z be a standard normal random variable and let W be a $\chi^2[\nu]$-distributed variable. Then, if Z and W are independent, $T = \dfrac{Z}{\sqrt{W/\nu}}$ is said to have the t-distribution with ν degrees of freedom (or parameter ν).

If $Y_1, \ldots, Y_n$ constitute a random sample from a normal population with mean μ and variance σ^2, Theorem 7.1 may be applied to show that $Z = \dfrac{\sqrt{n}(\overline{Y} - \mu)}{\sigma}$ has a standard normal distribution.

Theorem 7.3 tells us that $W = \dfrac{(n-1)S^2}{\sigma^2}$ has a χ^2 distribution with $\nu = n - 1$ df and that Z and W are independent (because $\overline{Y}$ and S^2 are independent).

Therefore, by Definition 7.2,

$$T = \frac{Z}{\sqrt{W/\nu}} = \frac{\sqrt{n}\left((\overline{Y} - \mu)/\sigma\right)}{\sqrt{\left((n-1)S^2/\sigma^2\right)/(n-1)}} = \sqrt{n}\left(\frac{\overline{Y} - \mu}{S}\right)$$

Exercise 7.98 outlines a method to find the density function of a t-distribution.

Exercise 7.98:

Suppose that T is defined as in Definition 7.2.

(a) If W is fixed at w, then T is given by ${}^{Z}/_{c}$, where $c = {}^{w}/_{\nu}$. Use this idea to find the conditional density of T for a fixed $W = w$.

(b) Find the joint density of T and W, $f(t, w)$, by using $f(t, w) = f(t \mid w)f(w)$.

(c) Integrate over w to show that

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \qquad -\infty < t < \infty.$$

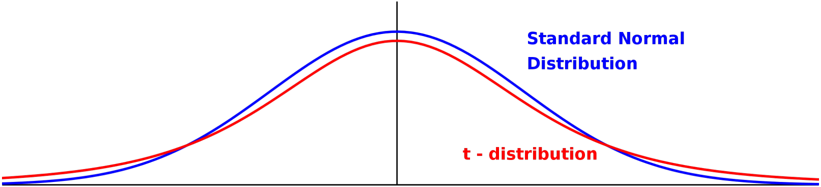


Table 5, Appendix 3 lists the values of t_α such that $P(T > t_\alpha)$.

In general, $t_\alpha = \phi_{1-\alpha}$, the $(1-\alpha)$ quantile (the $100(1-\alpha)^{\text{th}}$ percentile) of a t-distributed RV.

Example 7.6:

The tensile strength for a type of wire is normally distributed with unknown mean μ and unknown variance σ^2. Six pieces of wire were randomly selected from a large roll; Y_i, the tensile strength for portion i, is measured for $i = 1, \ldots, 6$. The population mean μ and variance σ^2 can be estimated by $\overline{Y}$ and S^2, respectively. Because $\sigma^2_{\overline{Y}} = {}^{\sigma^2}/_{n}$, it follows that $\sigma^2_{\overline{Y}}$ can be estimated by ${}^{\sigma^2}/_{n}$. Find the approximate probability that $\overline{Y}$ will be within ${}^{2S}/_{\sqrt{n}}$ of the true population mean μ.

Solution:

We want to find

$$P\left(-\frac{2S}{\sqrt{n}} \leq \overline{Y} - \mu \leq \frac{2S}{\sqrt{n}}\right) = P\left(-2 \leq \sqrt{n}\left(\frac{\overline{Y} - \mu}{S}\right) \leq 2\right)$$
$$= P(-2 \leq T \leq 2),$$

where T has a t-distribution with, in this case, $n - 1 = 5$ df.
Table 5, Appendix 3 suggests that the upper-tail area to the right of 2.015 is 0.05.
Hence $P(-2.015 \leq T \leq 2.015) = 0.9$, and the probability that $\overline{Y}$ will be within 2 estimated standard deviations of μ is slightly less than 0.9.

Remark:

If σ^2 were known, the probability that $\overline{Y}$ will fll within $2\sigma_{\overline{Y}}$ of μ would be

$$P\left(-\frac{2\sigma}{\sqrt{n}} \leq \overline{Y} - \mu \leq \frac{2\sigma}{\sqrt{n}}\right) = P\left(-2 \leq \sqrt{n}\left(\frac{\overline{Y} - \mu}{\sigma}\right) \leq 2\right)$$
$$= P(-2 \leq Z \leq 2) = 0.9544.$$

The F-Distribution

Suppose that we want to compare the variances of two normal populations based on information contained in independent random samples from the two populations.
Samples of sizes n_1 and n_2 are taken from the two populations with variances σ_1^2 and σ_2^2, respectively.
From the observations in the samples, we can estimate σ_1^2 and σ_2^2 from S_1^2 and S_2^2, respectively.
Thus it seems intuitive that the ratio ${S_1^2}/{S_2^2}$ could be used to make inferences about the relative magnitudes of σ_1^2 and σ_2^2.
The ratio $\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \dfrac{\sigma_2^2}{\sigma_1^2}\left(\dfrac{S_1^2}{S_2^2}\right)$ has the F-distribution with $n_1 - 1$ numerator degrees of freedom and $n_2 - 1$ denominator degrees of freedom.

Definition (7.3)

Let W_1 and W_2 be independent χ^2-distributed random variables with ν_1 and ν_2 df, respectively. Then $F = \dfrac{W_1/\nu_1}{W_2/\nu_2}$ is said to have an F-distribution with ν_1 numerator degrees of freedom and ν_2 denominator degrees of freedom.

Exercise 7.99 outlines a method to find the probability distribution function of an F-distribution.

Exercise 7.99:

Suppose F is defined as in Definition 7.3.

(a) If W_2 is fixed at w_2, then $F = {}^{W_1}/_{c}$, where $c = {}^{w_2 \nu_1}/_{\nu_2}$. Find the conditional density of F for fixed $W_2 = w_2$.

(b) Find the joint density of F and W_2.

(c) Integrate over w_2 to show that the probability density function of F – say, $g(y)$ – is given by

$$g(y) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)\left({}^{\nu_1}/_{\nu_2}\right)^{\nu_1/2}}{\Gamma\left({}^{\nu_1}/2\right)\Gamma\left({}^{\nu_2}/2\right)} y^{\frac{\nu_1}{2}-1}\left(1+\frac{\nu_1 y}{\nu_2}\right)^{-\frac{\nu_1+\nu_2}{2}}, \quad 0 < y < \infty.$$

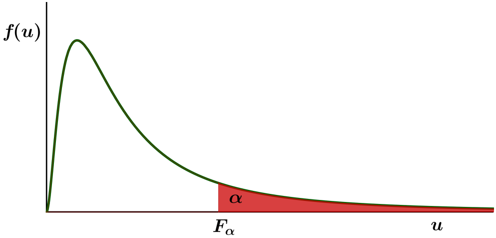


Table 7, Appendix 3 lists values of F_α such that $P(F > F_\alpha)$.

In general, $F_\alpha = \phi_{1-\alpha}$, the $(1-\alpha)$ quantile (the $100(1-\alpha)^{\text{th}}$ percentile) of an F-distributed RV.

Example 7.7:

If we take independent samples of size $n_1 = 6$ and $n_2 = 10$ from two normal populations with equal population variances, find the number b such that

$$P\left(\frac{S_1^2}{S_2^2} \le b\right) = 0.95.$$

Solution:

Because $n_1 = 6$ and $n_2 = 10$, and the population variances are equal, $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2}{S_2^2}$ has an F-distribution with $\nu_1 = n_1 - 1 = 5$ numerator degrees of freedom and $\nu_2 = n_2 - 1 = 9$ denominator degrees of freedom.

Also,

$$P\left(\frac{S_1^2}{S_2^2} \leq b\right) = 1 - P\left(\frac{S_1^2}{S_2^2} > b\right).$$

Therefore, we want to find the number b cutting off an upper-tail area of 0.05 under the F density function with 5 numerator degrees of freedom and 9 denominator degrees of freedom.

Looking in column 5 and row 9 in Table 7, Appendix 3, we see that the appropriate value of b is 3.48.

Remark:

Even when the population variances are equal, the probability that the ratio of the sample variances exceeds 3.48 is still 0.05 (assuming sample sizes of $n_1 = 6$ and $n_2 = 10$).

The Central Limit Theorem

Heuristically, if you add up a lot of IID RVs and normalize appropriately, the result is a standard normal RV.
More specifically,

Theorem (7.4)

Let $Y_1, \dots, Y_n$ be independent and identically distributed random variables with $E[Y_i] = \mu$ and $V[Y_i] = \sigma^2 < \infty$.
Define

$$U_n = \frac{\sum_{i=1}^n Y_i - n\mu}{\sigma\sqrt{n}} = \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} \quad \text{where } \overline{Y} = \frac{1}{n}\sum_{i=1}^n Y_i.$$

Then the distribution function of U_n converges to the standard normal distribution function as $n \to \infty$.
That is,

$$\lim_{n\to\infty} P(U_n \le u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{t^2/2}\, dt \qquad \text{for all } u.$$

Note: The formula $U_n = \dfrac{\overline{Y} - \mu}{\sigma/\sqrt{n}}$, where $\overline{Y} = \dfrac{1}{n}\sum_{i=1}^n Y_i$, guarantees that

$E[U_n] = 0$

Exercises:

(1) Show that $E[U_n] = 0$ using linearity of expectation and $E[Y_i] = \mu$.

(2) Also show that $V[U_n] = 1$ using independence, properties of variance, and $V[Y_i] = \sigma^2$.

So U_n is "correctly normalized" to approach a standard normal RV, in the sense of convergence in distribution.

Conclusion:

The distribution function of U_n converges to the distribution function of a standard normal RV.
That is,

$$\lim_{n\to\infty} P(U_n \leq u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{t^2/2}\, dt.$$

Note: This result is true for any distribution of the Ys that satisfies the hypotheses $V[Y] = \sigma^2 < \infty$, etc.

Remark: There are other senses of convergence, such as "weak convergence", "almost sure convergence", etc.

How do we apply this theorem in the context of Chapter 7?
The problems will not specifically say "Apply the CLT".

Exercise 7.43:

An anthropologist wishes to estimate the average height of men for a certain race of people. If the population standard deviation is assumed to be 2.5 inches and if she randomly samples 100 men, find the probability that the difference between the sample mean and the true population mean will not exceed .5 inch.

Interpretation:

- Rule of thumb: For most distributions, we get good convergence of U_ns to normal after about $n = 30$.
  The exercise has 100 samples.
  Since $100 > 30$, it is OK to apply the CLT.
- No specific distribution is mentioned.
  So if we don't apply the CLT, how can we do the problem?
- We are given the population standard deviation, which we need to apply the CLT for this exercise.

Solution:

Let $Y_1, \ldots, Y_{100}$ be the heights, and let $\overline{Y} = \dfrac{1}{100} \sum_{i=1}^{100} Y_i$ be the sample mean.
We are interested in $P\left(\left|\overline{Y} - \mu\right| < 0.5\right)$.

To solve, translate this into $P\left(|U_{100}| < \text{ something}\right)$, and then use that U_{100} is approximately standard normal (by CLT).

What is U_{100}?
$U_{100} \approx \dfrac{\overline{Y} - \mu}{\sigma/\sqrt{n}} = \dfrac{\overline{Y} - \mu}{(2.5)/\sqrt{100}} = \dfrac{\overline{Y} - \mu}{0.25}$.
So $\left|\overline{Y} - \mu\right| < 0.5$ iff $|U_{100}| < 2$.

Thus our question "What is $P\left(\left|\overline{Y} - \mu\right| < 0.5\right)$?" is the same as "What is $P\left(|U_{100}| < 2\right)$?".

But U_{100} is approximately standard normal.
Using the "95% rule" approximation, the answer is
$P\left(|U_{100}| < 2\right) = 1 - 2(0.0228) \approx 95.4\%$.

Remark:

We said $V[Y_i] = \sigma^2 < \infty$ was a hypothesis of the CLT.
The CLT is not true for distributions without a variance.
What kind of distribution doesn't have $V[Y] < \infty$?

There is one we will study in this chapter: The t-distribution with "1 degree of freedom" (or with parameter 1).
This is also called the Cauchy distribution, and it comes up in physics.

Definition (The t-distribution)

Let Z be a standard normal random variable and let W be a $\chi^2[\nu]$-distributed variable. Then, if Z and W are independent, $T = \dfrac{Z}{\sqrt{W/\nu}}$ is said to have the t-distribution with ν degrees of freedom (or parameter ν).

Remark:

The t-distribution with larger ν does have a variance.
$\nu = 1$ is something of an exceptional case.

The t-distribution with $\nu = 1$ has the PDF $f(y) = \dfrac{1}{\pi}\dfrac{1}{1+y^2}$.

If we take n IID samples $Y_1, \ldots, Y_n$ from a t-distribution with $\nu = 1$, then $\overline{Y} = \dfrac{1}{n}\sum_{i=1}^n Y_i$ also has the t-distribution with $\nu = 1$ for <u>any</u> n.
So this sum will never converge to a standard normal RV.

Recall that if $Z_1, \ldots, Z_\nu$ are independent standard normal RVs, then $Z_1^2 + \cdots + Z_\nu^2$ has the $\chi^2[\nu]$ distribution.
So think of T as the observation Z divided by "observed normalized errors".
This comes up as a "regression coefficient".

The reason for defining T is that it is a "Sampling Distribution derived from the normal distribution".

Remark: (for those interested in finance)

The standard model for "log-returns" in risk management is the t-distribution with $\nu = 5$ (maybe 4 or 6).

Reason: "fatter tails".

Recall: The Central Limit Theorem

Theorem (7.4)

Let $Y_1, \ldots, Y_n$ be independent and identically distributed random variables with $E[Y_i] = \mu$ and $V[Y_i] = \sigma^2 < \infty$.
Define

$$U_n = \frac{\sum_{i=1}^n Y_i - n\mu}{\sigma\sqrt{n}} = \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} \quad \textit{where } \overline{Y} = \frac{1}{n}\sum_{i=1}^n Y_i.$$

Then the distribution function of U_n converges to the standard normal distribution function as $n \to \infty$.
That is,

$$\lim_{n\to\infty} P(U_n \le u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{t^2/2}\, dt \qquad \textit{for all } u.$$

Proof:

Start with some key ingredients:

(a) Theorem 7.5, which we use as a black box.

(b) Taylor's theorem with remainder, which is another black box.

(c) Limit definition of the exponential function.

Theorem (7.5)

Let Y and $Y_1, Y_2, \ldots$ be random variables with moment-generating functions $m(t)$ and $m_1(t), m_2(t), \ldots,$ respectively. If $\lim_{n\to\infty} m_n(t) = m(t)$ for all real t, then the distribution function of Y_n converges to the distribution function of Y as $n \to \infty$.

Theorem (Taylor's theorem with remainder)

$$f(t) = \underbrace{f(0) + f'(0)\cdot t}_{\text{linear approximation to } f} + \underbrace{\frac{f''(\xi)}{2}\cdot t^2}_{\text{error term}}, \quad \text{where } 0 < \xi < t.$$

We will bound the error term by knowing something about f''.

The Exponential Function:

Recall that $\lim\limits_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e$ and $\lim\limits_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$,

and if $\lim\limits_{n\to\infty} b_n = b$, then $\lim\limits_{n\to\infty}\left(1+\frac{b_n}{n}\right)^n = e^b$.

Proof: (continued)

How does this apply to the CLT?

(a) shows that it is enough to show that $m_{U_n}(t) \to m_Z(t) = e^{t^2/2}$.

We know that $U_n = \dfrac{1}{\sqrt{n}} \sum_{i=1}^n Z_i$, where $Z_i = \dfrac{Y_i - \mu}{\sigma}$.

The Z_i are independent since Y_i are.

So

$$m_{U_n}(t) = m_{Z_1}\left(\frac{t}{\sqrt{n}}\right) \cdot \cdots \cdot m_{Z_n}\left(\frac{t}{\sqrt{n}}\right).$$

Also the Z_i are identically distributed.

So this means $m_{U_n}(t) = \left[m_{Z_1}\left(\dfrac{t}{\sqrt{n}}\right)\right]^n$.

Now apply

(b) to m_{Z_1}:

$m_{Z_1}(t) = m_{Z_1}(0) + m'_{Z_1}(0)t + m''_{Z_1}(\xi)\dfrac{t^2}{2}, \quad 0 < \xi < t.$

But we know that the derivatives of MGF are the moments:

$m_{Z_1}(0) = 1,$

$m'_{Z_1}(0) = E[Z_1] = 0,$

$m''_{Z_1}(0) = E[Z_1^2] = V[Z_1] + E[Z_1]^2 = 1.$

Proof: (continued)

So

$$m_{Z_1}(t) = m_{Z_1}(0) + m'_{Z_1}(0)t + m''_{Z_1}(\xi)\frac{t^2}{2} = 1 + 0 + m''_{Z_1}(\xi)\frac{t^2}{2}.$$

Thus

$$m_{U_n}(t) = \left[m_{Z_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n = \left[1 + m''_{Z_1}(\xi_n)\frac{\left(t/\sqrt{n}\right)^2}{2}\right]^n$$

$$= \left[1 + \frac{m''_{Z_1}(\xi_n)\frac{t^2}{2}}{n}\right]^n. \qquad \text{Here } 0 < \xi_n < \frac{t}{\sqrt{n}}.$$

Now $\lim_{n\to\infty} m_{U_n}(t) = \lim_{n\to\infty}\left[1 + \frac{m''_{Z_1}(\xi_n)\frac{t^2}{2}}{n}\right]^n.$

Since the MGF is continuous and $0 < \xi_n < \frac{t}{\sqrt{n}}$, we have $\xi_n \to 0$ and $m''_{Z_1}(\xi_n) \to m''_{Z_1}(0) = 1$.
So

$$m''_{Z_1}(\xi_n)\frac{t^2}{2} \to \frac{t^2}{2}.$$

Proof: (continued)

Now use

(c) :

This tells us that $\lim_{n\to\infty} m_{U_n}(t) = \lim_{n\to\infty}\left(1+\frac{b_n}{n}\right)^n$, where

$b_n = m''_{Z_1}(\xi_n)\frac{t^2}{2}$, and $\lim_{n\to\infty} b_n = \frac{t^2}{2}$.

So $\lim_{n\to\infty} m_{U_n}(t) = e^b = e^{t^2/2}$. This proves the Central Limit Theorem.

What do we need to know for solving CLT problems?

- We need to know facts about MGF used in the proof.
  You should be able to prove, for example, that $m_{aX}(t) = m_X(at)$.
- Also using the facts about mean and variance used, you should be able to show $E[Z_i] = 0$ and $V[Z_i] = 1$.
- You should be able to produce a correct statement of the CLT.
  Most importantly, using this, you should be able to do the problems of the form "Apply the CLT", even if the problem statements do not explicitly mention the CLT.

Application of the Central Limit Theorem

Exercise 7.45:

Workers employed in a large service industry have an average wage of \$7.00 per hour with a standard deviation of \$0.50. The industry has 64 workers of a certain ethnic group. These workers have an average wage of \$6.90 per hour. Is it reasonable to assume that the wage rate of the ethnic group is equivalent to that of a random sample of workers from those employed in the service industry?

[Hint: Calculate the probability of obtaining a sample mean less than or equal to \$6.90 per hour.]

Solution:

Let $Y_1, \ldots, Y_{64}$ be the pay rates of the workers in the ethnic group.
We are interested in the probability $P\left(\overline{Y} \leq \$6.90\right)$, where
$\overline{Y} = \frac{1}{64}\sum_{i=1}^{64} Y_i$ is the average.
We can apply the CLT because of the rule of thumb - "30 samples is good enough", and $64 > 30$.

Solution: (continued)

$$\overline{Y} \le 6.9 \iff \overline{Y} - 7 \le 6.9 - 7 = -0.1.$$

But $U_n = \dfrac{\overline{Y} - \mu}{\sigma/\sqrt{n}}$.

Here $\sigma = 0.5$ and $\sqrt{n} = \sqrt{64} = 8$.

So

$$\overline{Y} - 7 \le -0.1 \iff \frac{\overline{Y} - 7}{0.5/8} \le \frac{-0.1}{0.5/8} \iff U_{64} \le -\frac{8}{5} \iff U_n \le -1.6.$$

The approximation of the CLT is $P(U_n \le -1.6) \approx P(Z \le -1.6)$.

From the table, $P(Z \le -1.6) \overset{\text{symmetry}}{=} P(Z \ge 1.6) \approx \boxed{0.0548}$.

That is, the probability that we would observe such a wage variation by chance is about 5.5% in this model.

Remark:

Is this sufficient to conclude that there is differential pay for this ethnic group?

Answer: The calculation above is not sufficient – You have to believe that the model is valid.

Remark:

The most important part of the CLT to memorize is the formula

$$U_n = \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}}, \qquad \text{where } \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

The problems are all of a form where you start from "U_n is approximately standard normal" (for large n) and derive some conclusion.
(In the example we just did, it was about $P\left(\overline{Y} \leq 6.9\right)$.)

You cannot do this without the formula for U_n.

The formula U_n is NOT arbitrary; it is the simplest thing it could possibly be:
we start with $\overline{Y}$ and normalize it to have mean 0 and variance 1.

Exercise 7.37(a):

Let $Y_1, \ldots, Y_5$ be a random sample of size 5 from a normal population with mean 0 and variance 1 and let $Y = \frac{1}{5}\sum_{i=1}^{5} Y_i$. What is the distribution of $W = \sum_{i=1}^{5} Y_i^2$? Why?

Answer: Using MGFs, we can show that $W \sim \chi^2[5]$.

Exercise 7.51:

Exercise 7.51:

Shear strength measurements for spot welds have been found to have standard deviation 10 pounds per square inch (psi). How many test welds should be sampled if we want the sample mean to be within 1 psi of the true mean with probability approximately .99?

Exercise 7.51:

Shear strength measurements for spot welds have been found to have standard deviation 10 pounds per square inch (psi). How many test welds should be sampled if we want the sample mean to be within 1 psi of the true mean with probability approximately .99?

Solution:

Exercise 7.51:

Shear strength measurements for spot welds have been found to have standard deviation 10 pounds per square inch (psi). How many test welds should be sampled if we want the sample mean to be within 1 psi of the true mean with probability approximately .99?

Solution:

The above exercise tells us that $\sigma = 10$, and that we want $P\left(\left|\overline{Y} - \mu\right| \leq 1\right) \approx 0.99$.

Exercise 7.51:

Shear strength measurements for spot welds have been found to have standard deviation 10 pounds per square inch (psi). How many test welds should be sampled if we want the sample mean to be within 1 psi of the true mean with probability approximately .99?

Solution:

The above exercise tells us that $\sigma = 10$, and that we want $P\left(\left|\overline{Y} - \mu\right| \leq 1\right) \approx 0.99$. We are asked to find a suitable n for this.

Exercise 7.51:

Shear strength measurements for spot welds have been found to have standard deviation 10 pounds per square inch (psi). How many test welds should be sampled if we want the sample mean to be within 1 psi of the true mean with probability approximately .99?

Solution:

The above exercise tells us that $\sigma = 10$, and that we want $P\left(\left|\overline{Y} - \mu\right| \le 1\right) \approx 0.99$. We are asked to find a suitable n for this.

We can guess, since $\sigma = 10$, that the required n is large, so the CLT applies.

Exercise 7.51:

Shear strength measurements for spot welds have been found to have standard deviation 10 pounds per square inch (psi). How many test welds should be sampled if we want the sample mean to be within 1 psi of the true mean with probability approximately .99?

Solution:

The above exercise tells us that $\sigma = 10$, and that we want $P\left(\left|\overline{Y} - \mu\right| \leq 1\right) \approx 0.99$. We are asked to find a suitable n for this.

We can guess, since $\sigma = 10$, that the required n is large, so the CLT applies. To use the CLT and normal tables, we need to translate the probability requirement of the problem to something involving U_n, and then figure out what n we need.

Exercise 7.51:

Shear strength measurements for spot welds have been found to have standard deviation 10 pounds per square inch (psi). How many test welds should be sampled if we want the sample mean to be within 1 psi of the true mean with probability approximately .99?

Solution:

The above exercise tells us that $\sigma = 10$, and that we want $P\left(\left|\overline{Y} - \mu\right| \leq 1\right) \approx 0.99$. We are asked to find a suitable n for this.

We can guess, since $\sigma = 10$, that the required n is large, so the CLT applies. To use the CLT and normal tables, we need to translate the probability requirement of the problem to something involving U_n, and then figure out what n we need.

Start with

$$\left|\overline{Y} - \mu\right| \leq 1$$

Exercise 7.51:

Shear strength measurements for spot welds have been found to have standard deviation 10 pounds per square inch (psi). How many test welds should be sampled if we want the sample mean to be within 1 psi of the true mean with probability approximately .99?

Solution:

The above exercise tells us that $\sigma = 10$, and that we want $P\left(\left|\overline{Y} - \mu\right| \leq 1\right) \approx 0.99$. We are asked to find a suitable n for this.

We can guess, since $\sigma = 10$, that the required n is large, so the CLT applies. To use the CLT and normal tables, we need to translate the probability requirement of the problem to something involving U_n, and then figure out what n we need.

Start with

$$\left|\overline{Y} - \mu\right| \leq 1 \iff \left|\frac{\overline{Y} - \mu}{\sigma/\sqrt{n}}\right| \leq \frac{1}{\sigma/\sqrt{n}}$$

Exercise 7.51:

Shear strength measurements for spot welds have been found to have standard deviation 10 pounds per square inch (psi). How many test welds should be sampled if we want the sample mean to be within 1 psi of the true mean with probability approximately .99?

Solution:

The above exercise tells us that $\sigma = 10$, and that we want $P\left(\left|\overline{Y} - \mu\right| \leq 1\right) \approx 0.99$. We are asked to find a suitable n for this.

We can guess, since $\sigma = 10$, that the required n is large, so the CLT applies. To use the CLT and normal tables, we need to translate the probability requirement of the problem to something involving U_n, and then figure out what n we need.

Start with

$$\left|\overline{Y} - \mu\right| \leq 1 \iff \left|\frac{\overline{Y} - \mu}{\sigma/\sqrt{n}}\right| \leq \frac{1}{\sigma/\sqrt{n}} \implies \left|\left(\frac{\overline{Y} - \mu}{10/\sqrt{n}}\right)\right| \leq \frac{1}{10/\sqrt{n}}.$$

Exercise 7.51:

Shear strength measurements for spot welds have been found to have standard deviation 10 pounds per square inch (psi). How many test welds should be sampled if we want the sample mean to be within 1 psi of the true mean with probability approximately .99?

Solution:

The above exercise tells us that $\sigma = 10$, and that we want $P\left(\left|\overline{Y} - \mu\right| \leq 1\right) \approx 0.99$. We are asked to find a suitable n for this.

We can guess, since $\sigma = 10$, that the required n is large, so the CLT applies. To use the CLT and normal tables, we need to translate the probability requirement of the problem to something involving U_n, and then figure out what n we need.

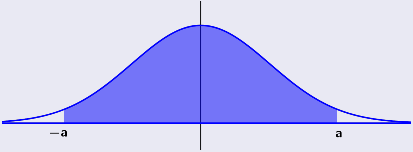
Start with

$$\left|\overline{Y} - \mu\right| \leq 1 \iff \left|\frac{\overline{Y} - \mu}{\sigma/\sqrt{n}}\right| \leq \frac{1}{\sigma/\sqrt{n}} \implies \left|\underbrace{\left(\frac{\overline{Y} - \mu}{10/\sqrt{n}}\right)}\right| \leq \frac{1}{10/\sqrt{n}}.$$

By CLT, the braced expression above is standard normal for large n.

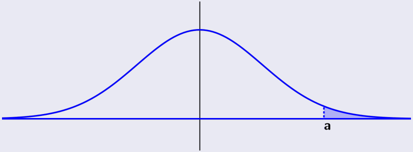
Solution: (continued)

Now we need to find n such that $P\left(|Z| \leq \frac{1}{10/\sqrt{n}}\right) = 0.99$.



For which a is the shaded area $= 0.99$?

By symmetry, this is the same as:



For which a is the shaded area $= 0.005$?

This we can look up!
$a \approx 2.575$ from the table.
So we solve for n:

$$\frac{1}{10/\sqrt{n}} \approx 2.575 \implies \frac{\sqrt{n}}{10} \approx 2.575 \implies \sqrt{n} \approx 25.75$$

$$\implies n \approx (25.75)^2 \approx 663.$$

Thus $n = 663$ is good enough.

Remark:

The CLT only requires that the samples $Y_1, \ldots, Y_n$ are IID and have $\sigma^2 < \infty$.

Exercise 7.53 (b):

One-hour carbon monoxide concentrations in air samples from a large city average 12 ppm (parts per million) with standard deviation 9 ppm. Find the probability that the average concentration in 100 randomly selected samples will exceed 14 ppm.

In such problems, CLT applies even though $Y_1, \ldots, Y_n$ are NOT normally distributed.

What is normally distributed is $U_n = \dfrac{\overline{Y} - \mu}{\sigma / \sqrt{n}}$.

The Normal Approximation to the Binomial Distribution

The central limit theorem also can be used to approximate probabilities for some discrete random variables when the exact probabilities are tedious to calculate.
One useful example involves the binomial distribution for large values of the number of trials n.

Suppose that Y has a binomial distribution with n trials and probability of success on any one trial denoted by p.
If we want to find $P(Y \leq b)$, we can use the binomial probability function to compute $P(Y = y)$ for each nonnegative integer $y \leq b$ and then sum these probabilities.
Tables are available for some values of the sample size n, but direct calculation is cumbersome for large values of n for which tables may be unavailable.

Alternatively, we can view Y, the number of successes in n trials, as a sum of a sample consisting of 0s and 1s;
that is,

$$Y = \sum_{i=1}^{n} X_i, \qquad \text{where } X_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ trial is success,} \\ 0 & \text{otherwise.} \end{cases}$$

The random variables X_i for $i = 1, \ldots, n$ are independent (because the trials are independent), and it is easy to show that $E[X_i] = p$ and $V[X_i] = p(1-p)$ for$i = 1, \ldots, n$.
Consequently, when n is large, the sample fraction of successes,

$$\frac{Y}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X},$$

possesses an approximately normal sampling distribution with mean $\mu_{\overline{X}} = E[X_i] = p$ and variance $V_{\overline{X}} = \dfrac{V[X_i]}{n} = \dfrac{p(1-p)}{n}$.

Thus, Theorem 7.4 (the central limit theorem) helps us establish that if $Y \sim \mathrm{Bin}(n, p)$ and if n is large, then $\dfrac{Y}{n}$ has approximately the same distribution as $U \sim \mathcal{N}\left(p, \dfrac{p(1-p)}{n}\right)$.
Equivalently, for large n, we can think of Y as having approximately the same distribution as $W \sim \mathcal{N}(np, np(1-p))$.

The normal approximation to binomial probabilities works well even for moderately large n as long as p is not close to zero or one.
A useful rule of thumb is that the normal approximation to the binomial distribution is appropriate when

$$0 < p - 3\sqrt{\frac{p(1-p)}{n}} \quad \text{and} \quad p + 3\sqrt{\frac{p(1-p)}{n}} < 1.$$

Equivalently, the normal approximation is adequate if

$$n > 9\left(\frac{\max\{p, 1-p\}}{\min\{p, 1-p\}}\right).$$

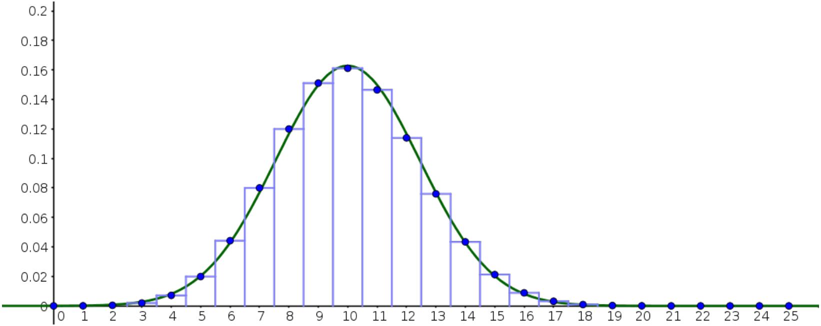
For example, suppose that Y has a binomial distribution with $n = 25$ and $p = 0.4$ (we will see this in the example that follows).
We have

$$\max\{0.4, 1-0.4\} = 0.6, \quad \min\{0.4, 1-0.4\} = 0.4$$

$$\Longrightarrow \; 9\left(\frac{\max\{p, 1-p\}}{\min\{p, 1-p\}}\right) = 9\left(\frac{0.6}{0.4}\right) = 13.5.$$

Since $n = 25 > 13.5$, the normal approximation is indeed adequate.

Here is a comparison of the distributions $Y \sim \text{Bin}(25, 0.4)$ (histogram in blue) and the normal approximation $W \sim \mathcal{N}(10, 6)$ (green):



Note that

$$\mu_W = np = 25(0.4) = 10,$$

and

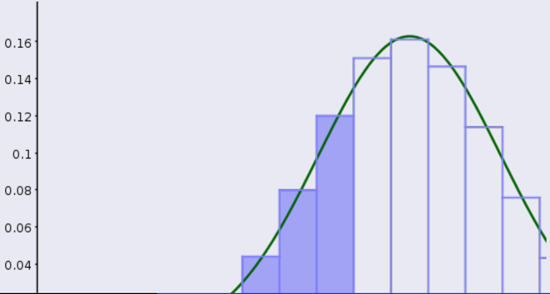
$$\sigma_W^2 = np(1-p) = 25(0.4)(0.6) = 6.$$

Example 7.11

Suppose that $Y \sim \text{Bin}(25, 0.4)$. Find the exact probabilities that $Y \leq 8$ and $Y = 8$ and compare these to the corresponding values found by using the normal approximation.

Solution:

The exact probability that $Y \leq 8$ is the blue (filled) area of the histogram shown along:



Solution: (continued)

We look up Table 1, Appendix 3, to find $P(Y \leq 8) = 0.274$.

Solution: (continued)

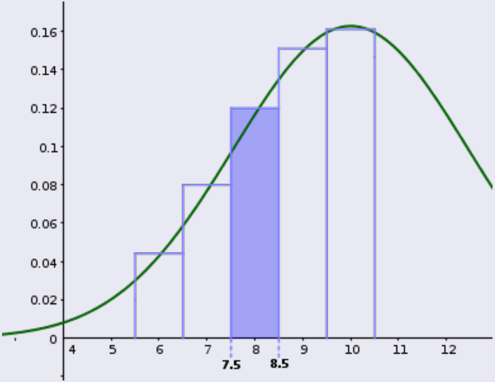
We look up Table 1, Appendix 3, to find $P(Y \leq 8) = 0.274$.

The exact probability that $Y = 8$ is the difference between $P(Y \leq 8)$ and $P(Y \leq 7)$.

Solution: (continued)

We look up Table 1, Appendix 3, to find $P(Y \leq 8) = 0.274$.

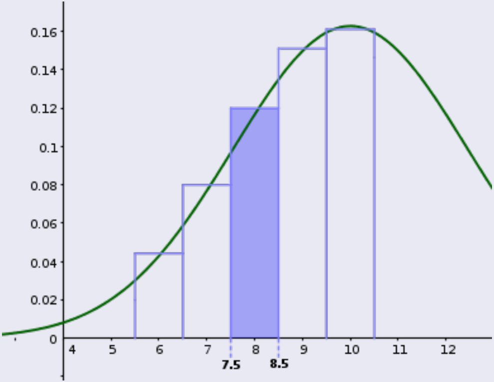
The exact probability that $Y = 8$ is the difference between $P(Y \leq 8)$ and $P(Y \leq 7)$. This is the blue (filled) strip in the picture:



Solution: (continued)

We look up Table 1, Appendix 3, to find $P(Y \leq 8) = 0.274$.

The exact probability that $Y = 8$ is the difference between $P(Y \leq 8)$ and $P(Y \leq 7)$. This is the blue (filled) strip in the picture:



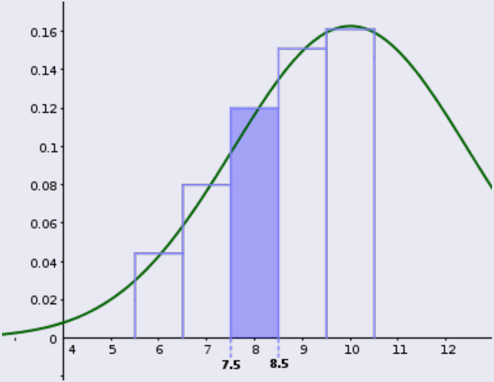
From the table, we find

$$P(Y = 8)$$
$$= P(Y \leq 8) - P(Y \leq 7)$$

Solution: (continued)

We look up Table 1, Appendix 3, to find $P(Y \leq 8) = 0.274$.

The exact probability that $Y = 8$ is the difference between $P(Y \leq 8)$ and $P(Y \leq 7)$. This is the blue (filled) strip in the picture:



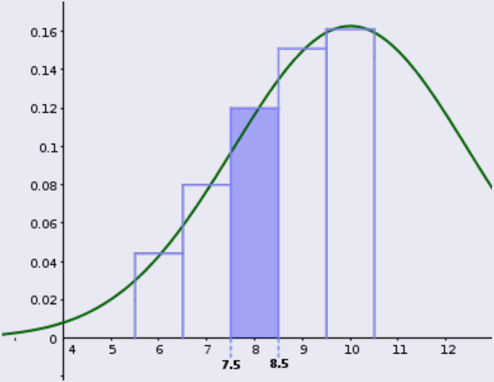
From the table, we find

$$
\begin{aligned}
&P(Y = 8) \\
=&P(Y \leq 8) - P(Y \leq 7) \\
=&0.274 - 0.154
\end{aligned}
$$

Solution: (continued)

We look up Table 1, Appendix 3, to find $P(Y \leq 8) = 0.274$.

The exact probability that $Y = 8$ is the difference between $P(Y \leq 8)$ and $P(Y \leq 7)$. This is the blue (filled) strip in the picture:



From the table, we find

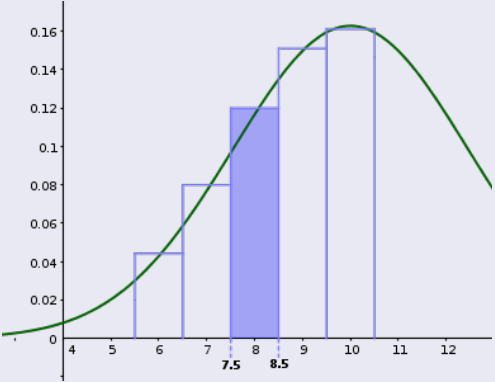
$$P(Y = 8)$$
$$= P(Y \leq 8) - P(Y \leq 7)$$
$$= 0.274 - 0.154 = 0.120.$$

Now our normal approximation is $W \sim \mathcal{N}(10, 6)$.

Solution: (continued)

We look up Table 1, Appendix 3, to find $P(Y \leq 8) = 0.274$.

The exact probability that $Y = 8$ is the difference between $P(Y \leq 8)$ and $P(Y \leq 7)$. This is the blue (filled) strip in the picture:

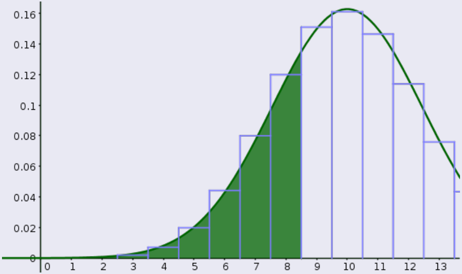


From the table, we find

$$P(Y = 8)$$
$$= P(Y \leq 8) - P(Y \leq 7)$$
$$= 0.274 - 0.154 = 0.120.$$

Now our normal approximation is $W \sim \mathcal{N}(10, 6)$. Looking at the picture, we need to find $P(Y \leq 8) \approx P(W \leq 8.5)$, and $P(Y = 8) \approx P(7.5 \leq W \leq 8.5)$; the half-integers accouting for the (obvious) correction.

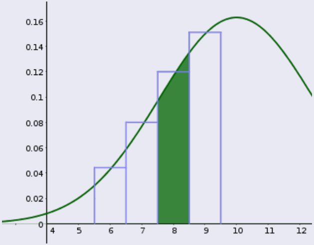
Solution: (continued)



Thus

$$P(W \leq 8.5)$$
$$= P\left(\frac{W-10}{\sqrt{6}} \leq \frac{8.5-10}{\sqrt{6}}\right)$$
$$= P(Z \leq -0.61) = 0.2709.$$

from Table 4, Appendix 3.



Likewise,

$$P(7.5 \leq W \leq 8.5)$$
$$= P\left(\frac{7.5-10}{\sqrt{6}} \leq \frac{W-10}{\sqrt{6}} \leq \frac{8.5-10}{\sqrt{6}}\right)$$
$$= P(-1.02 \leq Z \leq -0.61)$$
$$= 0.2709 - 0.1539 = 0.1170.$$

Note that the approximate values (0.2709 and 0.1170) are very close to the actual values (0.274 and 0.120) calculated earlier.

End of Chapter 7