Probability

Nathan Ponder

Louisiana Tech University

Contents

Chapter 1

Probability

Probability theory is the study of randomness. Major developments were made in this theory in the fifteenth and sixteenth centuries as individuals attempted to better understand games of chance. In more recent times, probability theory has been used to characterize random phenomena in the natural sciences, finance, health care, and other areas of study. It is worth noting too that it's necessary to have a grounding in probability in order to study statistics on a sophisticated level.

1.1. Sample Spaces

Probabilists are interested in knowing the likelihood of certain outcomes when an experiment is conducted. The set of all possible outcomes of an experiment is called the *sample space* and is labeled S. One or more outcomes taken together as a unit are referred to as an event. Events are typically labeled by upper case letters as A, B, C, etc. Some events are of such interest that we want to know their likelihood, or *probability* of occurring. For example we might want to know how likely it is that it will rain tomorrow, that a certain team will win a game, or that a certain stock price will go up during today's trading session. By convention, the probability that a certain event occurs is taken to be some number from 0 to 1. If we toss a balanced coin, for example, we say there's a probability of 1/2 we get heads. This is so because if we toss the coin over and over again, heads will come up about half of the time.

The notation for probability is as follows. If A is an event, the probability that A occurs is written $P(A)$. This is sometimes read "the probability of A" or simply "P of A". If the experiment consists of tossing a coin, and A represents the event that we get heads, then we have $P(A) = 1/2$. If the sample space S consists of n equally likely *outcomes*, and A is an event in S, then

$$P(A) = \frac{\text{the number of outcomes in A}}{n}.$$

As an example, there's a probability of 1/6 we get a 4 when we roll a die. We provide some more examples.

Example 1.1. *What's the probability you get at least two heads if you toss a coin three times?*

If you toss a coin three times, then the sample space of outcomes is

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}.$$

If A represents the event that you get at least two heads, then

$$A = \{HHT, HTH, THH, HHH\}.$$

Consequently,

$$P(A) = \frac{4}{8} = \frac{1}{2}$$

since there are just four outcomes in A and eight in S.

Example 1.2. *What's the probability that the sum of the two numbers that come up when you roll a pair of dice is 8?*

If you roll two dice to see which numbers come up, then the sample space of outcomes is $S = \{(1,1),(2,1),...,(6,6)\}$. If A represents the event we are to find the probability of, then $A = \{(6,2),(5,3),(4,4),(3,5),(2,6)\}$. Consequently, $P(A) = \frac{5}{36}$ since there are just five outcomes in A and 36 in S.

It's often desirable to study groups of events at the same time. To do this, probabilists have adopted the notation of set theory.

Groups of Events. For the events A and B in the sample space S, the following notation from set theory is used:

- $A \cap B$ is the event that both A and B occur.
- $A \cup B$ is the event that A or B occurs.
- A^C is read A *complement* and represents the event that A does not occur.
- $\emptyset$ is an impossible event.
- If $A \cap B = \emptyset$, then A and B are said to be *mutually exclusive.*

1.2. Probability Axioms and Rules

All of probability is based on three axioms that are simply formulated. Accepting them as true, we can build a whole theory of probability.

Probability Axioms

(1) $P(A) \geq 0$ for every event A in the sample space S
(2) $P(S) = 1$
(3) If the events A, B, C, etc. in the sample space S are mutually excusive one from another, then

$$P(A \cup B \cup C \cup \cdots) = P(A) + P(B) + P(C) + \cdots$$

Based on these three axioms we can derive a number of practical probability rules that will allow us to solve problems of interest. We list the most important.

Probability Rules

(1) $0 \leq P(A) \leq 1$ for every event A in the sample space S
(2) $P(A^C) = 1 - P(A)$ (*Complement Rule*)
(3) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (*Probability of a Union Rule*)
(4) If A and B are mutually exclusive, then $P(A \cup B) = P(A) + P(B)$
(5) For the three events A, B, and C,

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad -P(A \cap B) - P(B \cap C) - P(A \cap C) \\ &\quad +P(A \cap B \cap C) \end{aligned}$$

We provide some examples where these rules are put to use.

Example 1.3. *If A and B are events in a sample space for which $P(A) = 0.50$, $P(B) = 0.40$, and $P(A \cap B) = 0.25$, compute (a) $P(A \cup B)$ and (b) $P(B^C)$, and (c) $P(A^C \cap B)$.*

(a) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.50 + 0.40 - 0.25 = 0.65$

(b) $P(B^C) = 1 - P(B) = 1 - 0.40 = 0.60$

(c) Noting that B can be partitioned into the two mutually exclusive parts $A \cap B$ and $A^C \cap B$, we can write

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

Hence,

$$0.40 = 0.25 + P(A^C \cap B),$$

so that $P(A^C \cap B) = 0.15$. For a problem like this, it's worthwhile to sketch a Venn Diagram to see how B can be partitioned.

Example 1.4. *If you deal a card from a well shuffled deck of 52, what's the probability it's red or a king?*

Let A be the event that it's red and B the event that it's a king. Then

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= \frac{26}{52} + \frac{4}{52} - \frac{2}{52} \\ &= \frac{7}{13}. \end{aligned}$$

Example 1.5. *Suppose A, B, and C are events in a sample space for which $P(A) = 0.32$, $P(B) = 0.27$, $P(C) = 0.42$, $P(A \cap B) = 0.14$, $P(B \cap C) = 0.09$, $P(A \cap C) = 0.11$, and $P(A \cap B \cap C) = 0.08$. Compute (a) $P(A \cup B \cup C)$ and (b) $P(A \cap B^C \cap C)$.*

(a) $P(A \cup B \cup C) = 0.32 + 0.27 + 0.42 - 0.14 - 0.09 - 0.11 + 0.08 = 0.75$

(b) Sketching a Venn Diagram would be of great benefit here. We note that $A \cap C$ can be divided into the part that's in B and the part that's not in B to get $P(A \cap C) = P((A \cap B \cap C) \cup (A \cap B^C \cap C)) = P(A \cap B \cap C) + P(A \cap B^C \cap C)$, so that $0.11 = 0.08 + P(A \cap B^C \cap C)$ and $P(A \cap B^C \cap C) = 0.03$.

Exercises

(1) An experiment consists of tossing a coin three times. Write out the sample space of this experiment. Find the probability that you get more heads than tails.
Ans. See one of the examples above for S. $\frac{1}{2}$

(2) An experiment consists of tossing a coin four times. Write out the sample space of this experiment. Find the probability that you get (a) exactly two heads, (b) at most two heads.
Ans. $S = \{HHHH, HHHT, ..., TTTT\}$ *There should be 16 outcomes altogether. (a)* $\frac{3}{8}$, *(b)* $\frac{11}{16}$

(3) If A and B are two events in a sample space for which $P(A) = 0.30$, $P(B) = 0.60$, and $P(A \cap B) = 0.20$, compute (a) $P(A \cup B)$, (b) $P(A^C)$, and (c) $P(A^C \cap B)$.
Ans. (a) 0.70, *(b)* 0.70, *(c)* 0.40

(4) If A and B are mutually exclusive events for which $P(A) = 0.30$ and $P(B) = 0.60$, compute (a) $P(A \cup B)$ and (b) $P(A \cap B)$.
Ans. (a) 0.90, *(b)* 0

(5) Suppose 60% of college students have a VISA Card, 45% have a MasterCard, and 12% have both. What's the probability that a randomly selected college student (a) does not have a VISA or MasterCard, (b) has at least one of the two cards, (c) has a Master Card but not a VISA Card?
Ans. (a) 0.07, *(b)* 0.93, *(c)* 0.33

(6) If you roll two dice, what's the probability that the sum of the two numbers that come up is at least nine?
Ans. $\frac{5}{18}$

(7) Suppose A, B, and C are events in a sample space for which $P(A) = 0.30$, $P(B) = 0.25$, $P(C) = 0.45$, $P(A \cap B) = 0.15$, $P(B \cap C) = 0.08$, $P(A \cap C) = 0.12$, and $P(A \cup B \cup C) = 0.70$. Compute (a) $P(A \cap B \cap C)$, (b) $P(A^C \cap B \cap C)$, (c) $P((A^C \cap B) \cup C)$, and (d) $P(A^C \cap (B \cup C))$.
Ans. (a) 0.05*, (b)* 0.03*, (c)* 0.52*, (d)* 0.40

(8) If you deal a card from a well shuffled deck of 52, what's the probability (a) it's a king and (b) it's either a diamond or a two?
Ans. (a) $\frac{1}{13}$ *(b)* $\frac{4}{13}$

1.3. Odds

In probability theory, the term *odds* is defined by

$$\text{the odds of } A = \frac{P(A)}{P(A^C)}.$$

If you were to deal a card from a well shuffled deck of 52, the odds it would be an ace would be $\dfrac{4}{52} \div \dfrac{48}{52} = \dfrac{1}{12}$. Odds are typically written as a ratio of positive integers, so the odds of dealing an ace could be written as $\frac{1}{12}$, 1 : 12, or 1 *to* 12. Unfortunately, the term *odds* is not used consistently in different quarters. It is common to take "odds for the event A" to mean "odds of the event A" and "odds against the event A" to mean $P(A^C)/P(A)$. In other words, the odds against is the reciprocal of the odds for. Compounding the confusion is the fact that gamblers usually mean "odds against" when they simply say "odds".

Exercises

(1) When rolling a die, what are the odds (a) for and (b) against getting a 6.
Ans. (a) 1 *to* 5 *(b)* 5 *to* 1

(2) If you toss a fair coin six times, what are the odds (a) for and (b) against getting exactly five heads?
Ans. (a) 3/29 *(b)* 29/3

(3) If you deal a card from a well shuffled deck of 52 cards, what are the odds (a) for and (b) against getting a face card?
Ans. (a) 3/10 *(b)* 10/3

(4) If the odds against a candidate winning an election are four to one, what's the probability the candidate wins the election?
Ans. 20%

(5) You are told that the odds of the Saints winning the Super Bowl this season are 24 to 1. You recognize that you are actually being told that the odds against their winning the super bowl are 24 to 1. What is the probability they'll win the Super Bowl?
Ans. 0.04

(6) If you buy 5000 tickets, your odds for winning a lottery are 1 to 800. What is the probability to six decimal places that you win?
Ans. 0.001248

1.4. Conditional Probability and Independence

In many instances investigators want to know the likelihood of an event occurring given some other event has already occurred. In computing the probability that a football team will win the next game, for example, we would want to incorporate all information available. If the starting quarterback is injured, we should compute the probability conditioned on the event that the reserve quarterback will play. We use the notation $P(A|B)$ to represent the conditional probability that A will occur given B has occurred. It's customary to read $P(A|B)$ as simply "the probability of A given B".

By definition, the conditional probability of A given B, $P(A|B)$, is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

if $P(B) \neq 0$.

Example 1.6. *Suppose $P(A) = 0.5$ and $P(B) = 0.4$. Compute then the conditional probability $P(A|B)$ if (a) $P(A \cup B) = 0.75$ and (b) A and B are mutually exclusive.*

(a) Since, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, we have $0.75 = 0.5 + 0.4 - P(A \cap B)$ or $P(A \cap B) = 0.15$. Consequently,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.15}{0.4} = 0.375.$$

(b) Since A and B are mutually exclusive, $P(A \cap B) = 0$. As a result, $P(A|B) = \frac{0}{0.4} = 0$.

By multiplying each side of the equation $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ by $P(B)$, we obtain a useful result:

Multiplication Rule.

$$P(A \cap B) = P(A|B)P(B)$$

This rule proves useful in many probability computations.

Example 1.7. *Find the probability that the first two cards you deal from a well-shuffled deck of 52 are red.*

Let A be the event that the second card is red and B the event that the first is red. Then the probability to be computed is given by

$$P(A \cap B) = P(A|B)P(B) = \frac{25}{51} \cdot \frac{26}{52} = 0.2451.$$

If the occurrence of A is not dependent on the occurrence or nonoccurrence of B, we say that A and B are *independent*. Mathematically, we have that A and B are independent if $P(A|B) = P(A)$. Note that the multiplication rule becomes $P(A \cap B) = P(A|B)P(B) = P(A)P(B)$ if and only if A and B are independent. For this reason, the following is a common definition for independence:

Independence. The events A and B are said to be *independent* if
$$P(A \cap B) = P(A)P(B).$$

Example 1.8. *If A and B are independent events for which $P(A) = 0.6$ and $P(B) = 0.4$, compute $P(A \cup B)$.*

Since A and B are independent, $P(A \cap B) = P(A)P(B) = (0.6)(0.4) = 0.24$. As a result,
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.4 + 0.6 - 0.24 = 0.76.$$

Note that if the three events A, B, and C are independent, then $P(A \cap B \cap C) = P((A \cap B) \cap C) = P(A \cap B)P(C) = P(A)P(B)P(C)$. In more general terms, for the n independent events A_1, A_2, ...,A_n, we have
$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2)\cdots P(A_n).$$

Example 1.9. *If you toss a coin six times, what's the probability you get six heads?*

Let A_i be the event that the ith toss results in heads. Then A_1, A_2, ..., A_6, are independent. The answer is
$$P(A_1 \cap A_2 \cap \cdots \cap A_6) = P(A_1)P(A_2)\cdots P(A_6) = \left(\frac{1}{2}\right)^6 = 0.0156.$$

Example 1.10. *If you toss a coin six times what's the probability you get at least one tail?*

Note that a direct computation would be complicated. You'd have to find the probability that you get exactly one tail, that you get exactly two tails, that you get exactly three tails, etc., and then add all these probabilities together. Using the Complement Rule makes the computation much simpler. The complement of getting at least one tail is getting no tails. Another way of saying that you get no tails is to say that you get all heads. We know that the probability of getting all heads is 0.0156. Therefore, the probability of getting at least one tail is $1 - 0.0156 = 0.9844$.

Example 1.11. *Suppose one in eight soft drink bottle tops are winners. If you randomly buy eight bottles of soft drink, what's the probability you win at least once?*

According to the Complement Rule, the probability you win at least once is one minus the probability you don't win at all. Not winning at all, would mean that each of the eight bottle tops is a loser. Since each top is a loser with probability $\frac{7}{8}$, we have that the answer to the question is

$$1 - \left(\frac{7}{8}\right)^8 = 0.6564.$$

Exercises

(1) If A and B are independent, $P(A) = \frac{3}{4}$ and $P(B) = \frac{1}{3}$, compute $P(A \cap B)$.
Ans. $\frac{1}{4} = 0.25$

(2) Suppose A and B are two events for which $P(A) = 0.70$ and $P(B) = 0.25$. Compute $P(A \cap B)$ if (a) A and B are independent (b) A and B are mutually exclusive.
Ans. (a) 0.175 *(b)* 0

(3) Suppose A and B are two events for which $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{3}$, and $P(A \cap B) = \frac{1}{4}$. Compute (a) $P(A|B)$, (b) $P(B|A)$, (c) $P(A \cup B)$.
Ans. (a) $\frac{3}{4}$, *(b)* $\frac{1}{2}$, *(c)* $\frac{7}{12}$

(4) Suppose A and B are two events for which $P(A|B) = \frac{1}{2}$ and $P(A \cap B) = \frac{1}{3}$. Compute $P(B)$.
Ans. $\frac{2}{3}$

(5) When rolling two dice, what's the probability the first comes up as a 3 given the sum from the two dice is at least eight?
Ans. $\frac{2}{15}$

(6) If you toss a coin 12 times, what's the probability you get no heads? Round your answer off to five decimal places.
Ans. 0.00024

(7) Draw two cards without replacement from a well-shuffled deck of 52 playing cards. What's the probability they are both kings? Recall that there are four kings in a deck. Round your answer off to four decimal places.
Ans. $\frac{1}{221} = 0.0045$

(8) Only 15% of motorists come to a complete stop at a certain four way stop intersection. What's the probability that of the next ten motorists to go through that intersection (a) none come to a complete stop, (b) at least one comes to a complete stop, and (c) exactly two come to a complete stop.
Ans. (a) 0.1969, (b) 0.8031, (c) 0.2759

(9) Suppose one in 12 soft drink bottle tops are winners. If you randomly buy six bottles of soft drink, what's the probability you win at least once?
Ans. $1 - \left(\frac{11}{12}\right)^6 = 0.4067$

(10) Suppose one in four soft drink bottle tops are winners. If you randomly buy six bottles of soft drink, what's the probability you win at least once?
Ans. 0.8220

1.5. Bayes' Theorem

The result is named after Thomas Bayes, a Presbyterian minister from eighteenth century England. *Bayes' Theorem* is a formula that allows one to reverse the events in a conditional probability. An example where such a result is useful is when the physician tries to determine a disease based on certain symptoms. If we know the disease, we typically know the probability of various symptoms. Given a certain set of symptoms,however, we don't necessarily have a good idea of what the disease is.

To compute the conditional probability $P(B|A)$, note that A can be partitioned into the two parts $A \cap B$ and $A \cap B^C$ so that we have

$$A = (A \cap B) \cup (A \cap B^C),$$

and

$$P(A) = P(A \cap B) + P(A \cap B^C).$$

Now we can write $P(A \cap B) = P(A|B)P(B)$ and $P(A \cap B^C) = P(A|B^C)P(B^C)$ to obtain

$$\begin{aligned} P(B|A) &= \frac{P(A \cap B)}{P(A)} \\ &= \frac{P(A \cap B)}{P(A \cap B) + P(A \cap B^C)} \\ &= \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^C)P(B^C)} \end{aligned}$$

We now formulate the result.

Bayes' Theorem.

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^C)P(B^C)}$$

An application follows.

Example 1.12. *A company has developed a new drug test that tests positive on a drug user 99% of the time. It tests negative on non-drug users 99% of the time also. If only* 0.5% *of the employees in a large company are drug users, what's the probability that a tested employee is actually a drug user if the test is positive?*

Let A be the event that the employee tests positive and B the event that the employee is a drug user . Then we need to compute $P(B|A)$ to answer the question. We have

$$\begin{aligned} P(B|A) &= \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^C)P(B^C)} \\ &= \frac{(0.99)(0.005)}{(0.99)(0.005) + (0.01)(0.995)} \\ &= 0.332 \end{aligned}$$

A corollary to Bayes' Theorem is the law of total probability. It is nothing more than the additive expansion derived in the denominator of Bayes' Theorem. The formulation is as follows:

Law of Total Probability.

$$P(A) = P(A|B)P(B) + P(A|B^C)P(B^C)$$

Example 1.13. *A certain university requires all it's students to take the ACT exam before admission. Some 25% of College Algebra students at this university have an ACT math score of 26 or higher. Studies show that 90% of College Algebra students who have a math ACT score of 26 or better pass the class. For those with a math ACT score lower than 26, only 48% pass College Algebra. Assuming all the given information is accurate, compute the probability a randomly selected College Algebra student from the university in question will pass the class.*

Let A be the event that the randomly chosen College Algebra student passes the class, and let B be the event that the student made a 26 or higher on the math ACT. Then the answer is

$$P(A) = P(A|B)P(B) + P(A|B^C)P(B^C) = (0.90)(0.25) + (0.48)(0.75) = 0.585.$$

A generalized version of Bayes' Theorem can be derived if the sample space is partitioned into n events as opposed to just the two events B and B^C. The n events B_1, B_2, ..., B_n, are said to be a partition of the sample space S if

$$B_1 \cup B_2 \cup \cdots \cup B_n = S$$

and $B_i \cap B_j = \emptyset$ if $i \neq j$. To compute the conditional probability $P(B_i|A)$, we note that A can be partitioned into the n mutually exclusive parts $A \cap B_1, A \cap B_2, ..., A \cap B_n$, so that we have

$$B = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n),$$

and

$$P(B) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_n).$$

Now we can write $P(A \cap B_1) = P(A|B_1)P(B_1)$, $P(A \cap B_2) = P(A|B_2)P(B_2)$, ... , and $P(A \cap B_n) = P(A|B_n)P(B_n)$, to obtain

$$\begin{aligned}
P(B_i|A) &= \frac{P(A \cap B_i)}{P(A)} \\
&= \frac{P(A \cap B_i)}{P((A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n))} \\
&= \frac{P(A \cap B_i)}{P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_n)} \\
&= \frac{P(A|B_i)P(B_i)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n)}
\end{aligned}$$

We now formulate the general theorem.

Generalized Bayes' Theorem.

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n)}$$

Exercises

(1) If A and B are two events in a sample space for which $P(B) = 0.70$, $P(A|B) = 0.20$, $P(A|B^C) = 0.40$, compute $P(A)$.
Ans. 0.26

(2) If you deal two cards without replacement from a well shuffled deck of 52, what's the probability that the second will be red?
Ans. 0.5

(3) A blood test detects a certain disease 95% of the time when the disease is present. However, the test is also positive 1% of the time when the disease is not present. If 0.5% of the population actually has the disease, what's the probability a person has the disease given the test result is positive?
Ans. 0.3231

(4) Assume 1% of women at age forty have breast cancer. If 80% of women with breast cancer will get positive mammographies and 9.6% of women without breast cancer will also get positive mammographies, what's the probability that a 40 year old woman that has a positive mammography actually has breast cancer?
Ans. 0.0776

(5) An insurance company learns that a potential customer is smoking a cigar. The company also knows that 9.5% of males smoke cigars as do 1.7% of females. What's the probability that the potential customer is a male. Assume that half of the population to answer this question is male.

Ans. Let A be the event that the customer is a cigar smoker and B be the event that the customer is male. Then the answer is

$$\begin{aligned} P(B|A) &= \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^C)P(B^C)} \\ &= \frac{(0.095)(0.5)}{(0.095)(0.5) + (0.017)(0.5)} \\ &= 0.8482 \end{aligned}$$

The insurance company can be about 85% *sure that the cigar smoker is a male.*

(6) Suppose that just five percent of men and a quarter of one percent of women are color-blind and that there are an equal number of women and men. If a color-blind person is chosen at random, what's the probability that person is male?

Ans. 0.9524

(7) Three different factories, F_1, F_2, and F_3, are used to manufacture a large batch of cell phones. Suppose 20% of the phones are produced by F_1, 30% are produced by F_2, and 50% by F_3. Suppose also that 1% of the phones produced by F_1 are defective, as are 2% of those produced by F_2 and 3% of those produced by F_3. If one phone is selected at random from the entire batch and is found to be defective, what's the probability it was produced by (a) F_1 and (b) F_2?

Ans. (a) 0.0870

Chapter 2

Random Variables

Probabilists use *random variables* (RV's) to study random phenomena. A random variable is a rule that assigns numbers to outcomes of an experiment. Typically, X (or any other convenient upper case letter) is used to represent an RV. One can label events with RV's. For example, if an experiment consists of tossing three coins and a random variable X counts the number of heads that come up, then the event A that you get at least two heads can be written in terms of the RV by $A = \{X \geq 2\}$. Note that

$$P(\{X \geq 2\}) = \frac{3+1}{8} = \frac{1}{2},$$

since there are eight equally likely outcomes to the experiment, three of which result in two heads and one of which results in three heads. It is customary to abbreviate $P(\{X \geq 2\})$ with $P(X \geq 2)$ for ease of writing, so that we have $P(X \geq 2) = \dfrac{1}{2}$.

2.1. Discrete Random Variables

The RV we just considered is said to be *discrete* because it only takes the values 0, 1, 2, and 3. Discrete RV's only take a finite or countably infinite number of values. An example of a discrete random variable taking an infinite number of values is one that counts the number of tosses of a coin it takes for heads to come up. This random variable takes the values $1, 2, 3, 4, ...$, with positive probability. A random variable would not be discrete if it could take every value in some interval of numbers.

Associated with each discrete random variable X is a *probability mass function* $p(x)$ defined by

$$p(x) = P(X = x).$$

Probability Mass Function. The probability mass function $p(x)$ of a discrete random variable X has the following properties:

(1) $p(x) \geq 0$ for all x

(2) $\sum_x p(x) = 1$, the sum being over all values that the random variable can take

(3) $P(A) = \sum_{x \in A} p(x)$, where the sum is taken over all outcomes x in the event A

We consider a simple example of a discrete random variable.

Example 2.1. *Illustrate that the the first two properties listed above for probability mass functions are true for the random variable X that counts the number of heads you get when you toss a balanced coin three times.*

If you toss a coin three times, then the sample space of outcomes is

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}.$$

We therefore have that $\{X = 0\} = \{TTT\}$, $\{X = 1\} = \{HTT, THT, TTH\}$, $\{X = 2\} = \{HHT, THH, HTH\}$, $\{X = 3\} = \{HHH\}$, and $\{X = k\} = \emptyset$ for all $k \neq 0, 1, 2,$ or 3. Consequently,

$$p(x) = \begin{cases} 1/8 & \text{if } x = 0 \\ 3/8 & \text{if } x = 1 \\ 3/8 & \text{if } x = 2 \\ 1/8 & \text{if } x = 3 \\ 0 & \text{otherwise} \end{cases}$$

Note also that

$$\begin{aligned} \sum_x p(x) &= \sum_{k=0}^{3} p(k) \\ &= \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} \\ &= 1. \end{aligned}$$

Here's another example.

Example 2.2. *Assume the probability the home team wins the second game of the NBA finals by fewer than four points is* 0.17*. That they win by at least four points and fewer than seven points is* 0.28*, that they win by at least seven points and fewer than 10 points is* 0.25*, that they win by at least 10 points is* 0.10*, and that they lose is* 0.20*. Compute the probability that the home team wins the next NBA finals by at least four points.*

The event A that they win by at least four points is

$$\begin{aligned} P(A) &= \sum_{x \in A} p(x) \\ &= 0.28 + 0.25 + 0.10 \\ &= 0.63. \end{aligned}$$

We now consider an example for a discrete random variable that assumes an infinite number of values.

Example 2.3. *Illustrate that the the first two properties listed above for probability mass functions are true for the random variable X that counts the number tosses of a coin it takes for heads to come up.*

Clearly $P(X = 1) = 1/2$. We note that $P(X = 2)$ is the probability we get a tails on the first toss and a heads on the second. Since the two events are independent and each of probability $1/2$, we have that $P(X = 2) = (1/2)^2 = 1/4$. We note that the event $\{X = 3\}$ is the event that we get tails on the first toss, tails on the second, and heads on the third. We thus have that $P(X = 3) = (1/2)^3 = 1/8$. Continuing in this way, we obtain $P(x = k) = (1/2)^k$ for $k = 4, 5, 6, ...$ Clearly, $P(X = k) = 0$ if k is not a positive integer. As a result, $p(x) \geq 0$ for all x and

$$\begin{aligned} \sum_x p(x) &= \sum_{k=1}^{\infty} p(k) \\ &= \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k \\ &= \frac{\frac{1}{2}}{1 - \frac{1}{2}} \\ &= 1. \end{aligned}$$

For the penultimate equality we used the formula for the sum of a geometric series. The reader might wish to review this formula in the sequences and series chapter of a calculus textbook or the latter pages of a college algebra textbook.

Here is a related example.

Example 2.4. *What's the least number of times you would need to toss a coin to have at least a 99% probability of getting a heads?*

Tossing a coin once would result in a heads with probability $1/2$. If you toss a coin twice, there are four outcomes, three of which result in at least one heads. If you toss a coin three times there are eight outcomes, seven of which result in at least one heads. Continuing in this way, we obtain 2^k outcomes when the coin is tossed k time, and all but one of those outcomes has at least one heads. We therefore need to solve the inequality

$$\frac{2^k - 1}{2^k} > 0.99$$

to answer the question. We can rewrite this inequality as

$$1 - \left(\frac{1}{2}\right)^k > 0.99$$

or

$$\left(\frac{1}{2}\right)^k < 0.01.$$

Taking the natural logarithm of each side, we obtain

$$k \ln \frac{1}{2} < \ln(0.01)$$

or

$$k > \frac{\ln(0.01)}{\ln \frac{1}{2}} = 6.64.$$

We therefore need to toss the coin at least seven times.

Exercises

(1) Suppose X is a random variable that counts the number of heads you get when you toss a coin six times and that $p(x)$ is the probability mass function for X. There are methods - short of listing all outcomes in the sample space and counting the number that have exactly two or exactly four heads - to establish that $p(2) = p(4) = 15/64$. Deduce the values of $p(k)$ for $k = 0, 1, 5,$ and 6 by inspection, and then compute $p(3)$

Ans. $p(3) = \frac{5}{16}$

(2) Find the value of the constant c that makes $f(x)$ a probability mass function if $p(x)$ is given by

$$p(x) = \begin{cases} cx & \text{if } x = 2, 4, 6, \text{ or } 8 \\ 0 & \text{otherwise} \end{cases}$$

Ans. $c = 1/20$

(3) For the random variable X with the probability mass function in the immediately preceding problem, compute (a) $P(X \leq 4)$ and (b) $P(X < 4)$.

Ans. (a) 0.3 *(b)* 0.1

(4) What's the least number of times you would need to roll a die to have at least a 90% probability of getting a six?

Ans. Solve the inequality $1 - (5/6)^n \geq 9/10$ *to get* n *is at least* 13

(5) Let X be a random variable giving the sum obtained when you roll two dice. Write out a rule for the probability mass function of X. Note that $p(x)$ will be a piecewise defined function with 12 pieces. Compute the probability you get a sum of at least 10 by adding $p(10)$, $p(11)$, and $p(12)$.

Ans.

$$p(x) = \begin{cases} \frac{1}{36} & \textit{if } x = 2 \text{ or } 12 \\ \frac{2}{36} & \textit{if } x = 3 \text{ or } 11 \\ \frac{3}{36} & \textit{if } x = 4 \text{ or } 10 \\ \frac{4}{36} & \textit{if } x = 5 \text{ or } 9 \\ \frac{5}{36} & \textit{if } x = 6 \text{ or } 8 \\ \frac{6}{36} & \textit{if } x = 7 \\ 0 & \textit{otherwise} \end{cases}$$

Prob. of getting at least 10 is $\frac{1}{6}$

(6) Can a discrete random variable X have a probability mass function with rule

$$p(x) = \begin{cases} 1/x & \text{if } x \text{ is an integer greater than or equal to } 2 \\ 0 & \text{otherwise} \end{cases}$$

Explain your answer.

2.2. Continuous Random Variables

A random variable is said to be *continuous* if it can take all the values in an interval of real numbers. Continuous RV's contrast with discrete RV's in that they can take on an uncountable number of values. They are useful for measuring such phenomena as lengths, forces, time intervals, etc..

One interesting property a continuous random variables X has is that $P(X = c) = 0$ for every real number c. You might recall that this was not the case with the discrete RV X counting the number of heads that come up when you toss a coin three times. In that experiment $P(X = 3) = 1/8$. This isn't as strange as it seems. An object might appear to be 10 cm long for example. If a fine enough measuring device is used to measure its length, however, the measurer will notice that the object is actually either slightly larger or slightly shorter than 10 cm.

Probability Density Function. Each continuous RV X has associated with it a *probability density function* (pdf) $f(x)$ with the following properties:

(1) $f(x) \geq 0$ for all real x,

(2) $\displaystyle\int_{-\infty}^{\infty} f(x)dx = 1$, and

(3) $P(X \leq c) = \displaystyle\int_{-\infty}^{c} f(x)dx$ for every real number c.

We provide a few examples involving pdf's.

Example 2.5. *Show that the function*

$$f(x) = \left\{ \begin{array}{cc} 2e^{-2x} & \textit{for } x \geq 0 \\ 0 & \textit{otherwise} \end{array} \right\}$$

satisfies the first two pdf properties.

1) Since e raised to any power is positive, $2e^{-2x}$ is always positive and $f(x) \geq 0$ for all x.

2) $\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} 2e^{-2x}dx = -e^{-2x}|_0^\infty = 0 - (-e^0) = 1.$

Example 2.6. *Suppose the random variable X has the pdf*

$$f(x) = \left\{ \begin{array}{cc} \frac{1}{5} & \text{for } 1 \leq x \leq c \\ 0 & \text{otherwise} \end{array} \right\}.$$

What value must the constant c take for $f(x)$ to be a pdf? For that value, compute $P(X \leq 2.3)$

$$1 = \int_1^c \frac{1}{5}dx = \frac{x}{5}|_1^c = \frac{c}{5} - \frac{1}{5} \;\Rightarrow\; c = 6.$$

$$P(X \leq 2.3) = \int_{-\infty}^{2.3} f(x)dx = \int_{-\infty}^{1} 0dx + \int_1^{2.3} \frac{1}{5}dx = 0 + \frac{x}{5}|_1^{2.3} = \frac{1}{5}(2.3 - 1) = 0.26$$

We now derive formulas for computing probabilities on different types of intervals. Suppose $a < b$. Then

$$\int_{-\infty}^{b} f(x)dx = \int_{-\infty}^{a} f(x)dx + \int_a^b f(x)dx$$

so that

$$\begin{aligned} \int_a^b f(x)dx &= \int_{-\infty}^{b} f(x) - \int_{-\infty}^{a} f(x)dx \\ &= P(X \leq b) - P(X \leq a) \\ &= P(\{X \leq a\} \cup \{a < X \leq b\}) - P(X \leq a) \\ &= P(X \leq a) + P(a < X \leq b) - P(X \leq a) \\ &= P(a < X \leq b) \end{aligned}$$

Since $P(X = b) = 0$, we have that $P(a < X < b) = \displaystyle\int_a^b f(x)dx$. Similar derivations yield the following formulas:

Formulas Involving Probability Density Functions.

$$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b) = \int_a^b f(x)dx$$

$$P(X \leq c) = P(X < c) = \int_{-\infty}^{c} f(x)dx$$

$$P(X \geq c) = P(X > c) = \int_c^{\infty} f(x)dx$$

Example 2.7. *If the random variable X has the pdf*

$$f(x) = \left\{ \begin{array}{cc} \frac{2}{x^3} & \textit{for } x \geq 1 \\ 0 & \textit{otherwise} \end{array} \right\},$$

compute (a) $P(X < 2)$, (b) $P(2 < X < 4)$, and (c) $P(X > 3)$

(a) $P(X < 2) = \int_{-\infty}^{2} f(x)dx = \int_{1}^{2} 2x^{-3}dx = -x^{-2}|_1^2 = -\frac{1}{4} - (-1) = \frac{3}{4}$

(b) $P(2 < X < 4) = \int_{2}^{4} 2x^{-3}dx = -x^{-2}|_2^4 = -\frac{1}{16} - (-\frac{1}{4}) = \frac{3}{16}$

(c) $P(X > 3) = \int_{3}^{\infty} 2x^{-3}dx = -x^{-2}|_3^{\infty} = 0 - (-\frac{1}{9}) = \frac{1}{9}$

Exercises

(1) Suppose X is a continuous random variable with pdf

$$f(x) = \left\{ \begin{array}{cc} \frac{1}{10} & \text{for } 0 \leq x \leq 10 \\ 0 & \text{otherwise} \end{array} \right\}.$$

Compute (a) $P(X \leq 4)$, (b) $P(X > 6.5)$, (c) $P(X \geq 6.5)$, and (d) $P(5 \leq X \leq 6)$.

Ans. (a) 0.4, *(b)* 0.35, *(c)* 0.35, *(d)* 0.1

(2) What value must the constant c take for the function

$$f(x) = \left\{ \begin{array}{cc} cx^2 & \text{for } 1 \leq x \leq 2 \\ 0 & \text{otherwise} \end{array} \right\}$$

to be a pdf?

Ans. $c = \frac{3}{7}$.

(3) Suppose X is a continuous random variable with pdf

$$f(x) = \left\{ \begin{array}{cc} \frac{1}{3} & \text{for } 4 \leq x \leq 7 \\ 0 & \text{otherwise} \end{array} \right\}.$$

Compute (a) $P(5 \leq X \leq 6)$, (b) $P(X \geq 6)$, (c) $P(X \leq 1)$, (d) $P(X \geq 2)$, and (e) $P(X = 5)$.

Ans. (a) $\frac{1}{3}$, *(b)* $\frac{1}{3}$, *(c) 0, (d) 1, (e) 0*

(4) If the random variable X has the pdf

$$f(x) = \left\{ \begin{array}{cc} \frac{1}{x^2} & \text{for } x \geq 1 \\ 0 & \text{otherwise} \end{array} \right\},$$

compute (a) $P(X < 5)$, (b) $P(2 < X < 4)$, and (c) $P(X > 2)$

Ans. (a) $\frac{4}{5}$ *(b)* $\frac{1}{4}$ *(c)* $\frac{1}{2}$

(5) Suppose a random variable X has the pdf

$$f(x) = \frac{c}{1 + x^2}.$$

Compute (a) the value of c, (b) $P(X > 0)$, and (c) $P(X < 1)$.

Ans. (a) $1/\pi$ *(b)* $\frac{1}{2}$ *(c)* $\frac{3}{4}$

(6) For the random variable $X$with pdf

$$f(x) = \left\{ \begin{array}{cc} 2e^{-2x} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{array} \right\},$$

(a) compute $P(1 < X < 2)$ and (b) $P(X > 1)$ to four decimal places.
Ans. (a) 0.1170 *(b)* 0.1353

(7) For the random variable X with pdf

$$f(x) = \left\{ \begin{array}{cc} xe^{-x} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{array} \right\},$$

(a) verify $\int_{-\infty}^{\infty} f(x) = 1$ and (b) compute $P(1 < X < 2)$ to four decimal places.
Ans. (b) 0.3298

(8) Students of calculus are accustomed to evaluating definite integrals by finding an antiderivative and applying the Fundamental Theorem. Even though an antiderivative cannot be found for the function

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

the function is still the pdf of a certain random variable X. For this random variable use your calculator to approximate to four decimal places (a) $P(0 \leq X \leq 1)$ and (b) $P(X \leq 1.96)$.
Ans. (a) 0.3413 *(b)* 0.9750

2.3. Cumulative Distribution Functions

Associated with all random variables are *cumulative distribution functions* (CDF's). Some statisticians refer to them simply as *distribution functions*. The name is indeed descriptive. For the RV X, the CDF $F(x)$ is given by

$$F(x) = P(X \leq x).$$

The function gives the cumulative probability up to its argument. It therefore is a nondecreasing function.

Example 2.8. *Find a rule for the CDF of the RV X that counts the number of heads that come up when a coin is tossed three times.*

If $x < 0$, then the event $\{X \leq x\}$ cannot occur since it's not possible to get fewer than zero heads. Consequently, $F(x) = P(X \leq x) = 0$.

If $0 \leq x < 1$, then $\{X \leq x\}$ is actually the event that zero heads occur, so we have that $F(x) = P(X \leq x) = 1/8$.

If $1 \leq x < 2$, then $\{X \leq x\}$ is the event that 0 or 1 heads come up, so $F(x) = 4/8 = 1/2$.

If $2 \leq x < 3$, then $\{X \leq x\}$ is the event that 0, 1, or 2, heads come up, so $F(x) = 7/8$.

Finally, if $x > 3$, $\{X \leq x\}$ is the event that 0, 1, 2, or 3, heads come up, and we have $F(x) = 1$.

The CDF is therefore given by

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/8 & \text{if } 0 \le x < 1 \\ 1/2 & \text{if } 1 \le x < 2 \\ 7/8 & \text{if } 2 \le x < 3 \\ 1 & \text{if } x \ge 3 \end{cases}$$

Note that the function jumps upward at the values of x that X takes with positive probability. Between these x values $F(x)$ is constant. We will see later that CDF's for continuous random variables have no such jumps.

If the RV is continuous with pdf $f(x)$, we can write

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

I.e. $F(x)$ is the cumulative area under the curve up to x. One will recall that the Fundamental Theorem of Calculus says that $F'(x) = f(x)$ if the function $f(x)$ itself is continuous.

Example 2.9. *Find the CDF for the RV X that has pdf*

$$f(x) = \begin{cases} 1/10 & \text{if } 10 \le x \le 20 \\ 0 & \text{otherwise} \end{cases}.$$

If $x < 10$, then

$$F(x) = \int_{-\infty}^{x} 0dt = 0.$$

If $10 \le x < 20$, then

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{10} 0dt + \int_{10}^{x} \frac{1}{10}dt = 0 + \frac{t}{10}|_{10}^{x} = \frac{1}{10}(x-10).$$

If $x > 20$, then

$$F(x) = \int_{-\infty}^{10} 0dt + \int_{10}^{20} \frac{1}{10}dt + \int_{20}^{x} 0dt = 0 + \frac{t}{10}|_{10}^{20} + 0 = \frac{1}{10}(20-10) = 1.$$

We therefore have that

$$F(x) = \begin{cases} 0 & \text{for } x < 10 \\ \frac{1}{10}(x-10) & \text{for } 10 \le x \le 20 \\ 1 & \text{for } x > 20 \end{cases}.$$

Note that $F(x)$ is constantly zero to the left of 10, constantly one to the right of 20 and a line segment connecting $(10,0)$ to $(20,1)$ when x is between 10 and 20.

We now list some general properties of distribution functions that can be derived from the definition $F(x) = P(X \le x)$.

Cumulative Distribution Function Properties. If $F(x)$ is the cumulative distribution function of a RV X (that is either discrete or continuous), then

(1) $F(x)$ is non-decreasing,
(2) $\lim_{x\to-\infty} F(x) = 0$,
(3) $\lim_{x\to\infty} F(x) = 1$,
(4) $P(a < X \leq b) = F(b) - F(a)$, and
(5) $P(X = c) = F(c) - \lim_{x\to c^-} F(x)$.

Note furthermore that if X is a continuous RV, $P(X = c) = 0$ and

$$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b) = F(b) - F(a).$$

We provide two more examples dealing with CDF's.

Example 2.10. *If the RV X has the CDF*

$$F(x) = \begin{cases} 0 & \text{for } x < 2 \\ 1/4 & \text{for } 2 \leq x < 6 \\ 3/4 & \text{for } 6 \leq x < 7 \\ 1 & \text{for } x \geq 7 \end{cases},$$

find a rule for the probability mass function of X.

Since the CDF steps up from 0 to 1/4 at $x = 2$, we have that $f(2) = 1/4$. The CDF is then constant on $[2, 6)$ and jumps $3/4 - 1/4 = 1/2$ at $x = 6$. Hence, $f(6) = 1/2$. Using the last of the properties of CDF's, we could write this as

$$f(6) = P(X = 6) = F(6) - \lim_{x\to 6^-} F(x) = 3/4 - 1/4 = 1/2.$$

The last jump is of magnitude $1 - 3/4 = 1/4$ at $x = 7$, giving us $f(7) = 1/4$. The probability mass function is therefore given by

$$f(x) = \begin{cases} 1/4 & \text{if } x = 2 \\ 1/2 & \text{if } x = 6 \\ 1/4 & \text{if } x = 7 \\ 0 & \text{otherwise} \end{cases}$$

Example 2.11. *Find a rule for the CDF of the random variable X that has pdf*

$$f(x) = \frac{1}{\pi(1 + x^2)}.$$

Then use the CDF to compute $P(X < \sqrt{3})$ and $P(-1 \leq X \leq 1)$.

We have that

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} \frac{1}{\pi(1+t^2)} dt \\ &= \frac{1}{\pi} \tan^{-1} t|_{-\infty}^{x} \\ &= \frac{1}{\pi}(\tan^{-1} x - (-\pi/2)) \\ &= \frac{1}{2} + \frac{1}{\pi} \tan^{-1} x \end{aligned}$$

Since $F(x) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} x$, we have that

$$\begin{aligned} P(X < \sqrt{3}) &= F(\sqrt{3}) \\ &= \frac{1}{2} + \frac{1}{\pi} \cdot \frac{\pi}{3} \\ &= \frac{1}{2} + \frac{1}{3} \\ &= \frac{5}{6} \; . \end{aligned}$$

We also have that

$$\begin{aligned} P(-1 \le X \le 1) &= P(-1 < X \le 1) \\ &= F(1) - F(-1) \\ &= \frac{1}{2} + \frac{1}{\pi} \cdot \frac{\pi}{4} - (\frac{1}{2} + \frac{1}{\pi} \cdot (-\frac{\pi}{4})) \\ &= \frac{1}{2} \; . \end{aligned}$$

To conclude this section we discuss percentiles.

Percentiles of Continuous Random Variables. If X is a continuous random variable with CDF $F(x)$, we define the $100p$th percentile of X to be the value a that solves $F(a) = p$.

For example, the 75th percentile of the random variable from the last example is given by the value of a that solves $F(a) = 0.75$. We solve $\frac{1}{2} + \frac{1}{\pi} \tan^{-1} a = 0.75$ and obtain $\tan^{-1} a = \frac{\pi}{4}$ or $a = 1$.

The reader should note that the *median* of a continuous random variable is nothing more than its fiftieth percentile.

Exercises

(1) Can

$$F(x) = \begin{cases} 6x(1-x) & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

be a CDF? Explain why.

Ans. No. This function is actually decreasing on the interval $(\frac{1}{2}, 1)$.

(2) Suppose X is a random variable for which $P(X = 0) = 1/4$ and $P(X = 1) = 3/4$. Find a rule for X's CDF $F(x)$. Then graph this CDF and compute $F(-2)$, $F(0)$, $F(0.7)$, and $F(1.2)$.

Ans. $F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/4 & \text{if } 0 \le x < 1 \\ 1 & \text{if } x \ge 1 \end{cases}$, $F(-2) = 0$, $F(0) = 1/4$, $F(0.7) = 1/4$, $F(1.2) = 1$

(3) For the previous problem, compute (a) $P(X \le 2)$ and (b) $P(X > 1)$.

Ans. (a) $F(2) = 1$, *(b)* $1 - P(X \le 1) = 1 - F(1) = 1 - 1 = 0$

(4) If we let the random variable X count the number of heads you get when you toss a balanced coin 10 times, then the cumulative distribution function of X is as given in the table. When reading the table, take the function's value to be constant from one integer up to all values less than the next larger integer. For example, $F(x) = 0.172$ if $3 \le x < 4$, and $F(x) = 0.377$ if $4 \le x < 5$. Compute the probability that (a) you get at most three heads, (b) you get at least six heads, and (c) you get exactly 7 heads.

x	F(x)
0	0.001
1	0.011
2	0.055
3	0.172
4	0.377
5	0.623
6	0.828
7	0.945
8	0.989
9	0.999
10	1.000

Ans. (a) $P(X \le 3) = F(3) = 0.172$, *(b)* $P(X \ge 6) = 1 - P(X < 6) = 1 - P(X \le 5) = 1 - F(5) = 1 - 0.623 = 0.377$ *(c)* $P(X = 7) = F(7) - \lim_{x \to 7^-} F(x) = 0.945 - 0.828 = 0.117$

(5) Graph and find a rule for the CDF of the random variable with pdf

$$f(x) = \left\{ \begin{array}{cc} \frac{1}{x^2} & \text{for } x \ge 1 \\ 0 & \text{otherwise} \end{array} \right\},$$

Ans.

$$F(x) = \left\{ \begin{array}{cc} 0 & \text{if } x < 1 \\ 1 - \frac{1}{x} & \text{if } x \ge 1 \end{array} \right\},$$

(6) Suppose X is a continuous random variable with CDF

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - e^{-x} & \text{if } x \geq 0 \end{cases}.$$

Compute (a) $P(X < 2)$, (b) $P(1 < X < 2)$, (c) $P(X > 1)$.
Ans. (a) $1 - e^{-2}$, *(b)* $e^{-1} - e^{-2}$, *(c)* e^{-1}

(7) Compute the 90th percentile of X to four decimal places in the previous problem.
Ans. 2.3026

(8) For the random variable with pdf

$$f(x) = \begin{cases} 6x(1-x) & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

compute (a) the median and (b) the sixtieth percentile.
Ans. (a) 1/2 (b) Solving the equation $0.60 = \displaystyle\int_0^a 6x(1-x)dx$, *we obtain*

$$10a^3 - 15a^2 + 3 = 0.$$

A graphing calculator tells us that a is approximately 0.5671.

(9) For the continuous random variable with probability density function

$$f(x) = \begin{cases} 2e^{-2x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

compute (a) $F(1)$ and (b) the 95th percentile of X.
Ans. (a) $F(1) = P(X \leq 1) = \int_0^1 2e^{-2x}dx = 0.8647$ *(b)* $\frac{1}{2}\ln 20 = 1.4979$

(10) For the discrete random variable with probability mass function

$$f(n) = \begin{cases} \frac{3{,}600}{5{,}269n^2} & \text{if } n = 1, 2, 3, 4, \text{ or } 5 \\ 0 & \text{otherwise} \end{cases}$$

compute $F(4)$.
Ans. $F(4) = P(X \leq 4) = 1 - P(X = 5) = 1 - \frac{3600}{5269(5)^2} = 0.9727.$

2.4. Expected Values and Variance

Those who study distributions of values are often interested in central tendencies of those values. To measure such tendencies, they can take the *expected value* of a random variable modeling the distribution. For a discrete random variable X, the expected value, $E(X)$, is the sum of the values X takes weighted by their likelihood of occurring. We provide the defining formula for $E(X)$ in terms of the probability mass function.

Expected Value of a Discrete Random Variable. For the discrete random variable X with probability mass function $p(x) = P(X = x)$, the expected value of X, $E(X)$, is given by

$$E(X) = \sum_x xp(x),$$

where the sum is over all values that X takes. If the sum diverges, then the random variable is said not to have an expected value.

The expected value of X is also referred to as the *expectation* or *mean* of X, and the smaller case Greek letter μ (mu) is often used to designate it. We provide examples of expected values.

Example 2.12. *Suppose X is a random variable for which $P(X = -20) = 0.85$, $P(X = 100) = 0.14$, and $P(X = 500) = 0.01$. Compute the expected value of X.*

We have $E(X) = (-20)(0.85) + (100)(0.14) + (500)(0.01) = 2$.

Example 2.13. *Suppose X is a random variable that represents a gambler's gain on a \$100 bet on red at the roulette wheel. Compute the expected value of X.*

The ticker on a roulette wheel has equal likelihood of stopping on each of 38 slots. Eighteen of the slots are red, 18 are black, and two are green. Consequently, $P(X = 100) = \frac{18}{38}$ and $P(X = -100) = \frac{20}{38}$, so that $E(X) = (-100)(\frac{20}{38}) + (100)(\frac{18}{38}) = -5.2632$. This expected value tells us that placing many \$100 bets on red will result in an average loss of about \$5.26 for gamblers.

If $g(x)$ is a real valued function on the reals and X is a random variable, then $g(X)$ is a random variable too. We define the random variable Y by

$$Y = g(X)$$

and note that it X is discrete,

$$\begin{aligned}
E(g(X)) &= E(Y) \\
&= \sum_y yP(g(X) = y) \\
&= \sum_y y \sum_{g(x)=y} p(x) \\
&= \sum_y \sum_{g(x)=y} yp(x) \\
&= \sum_x g(x)p(x).
\end{aligned}$$

So we have a practical formula for computing the expected value of a function of a random variable.

If X is discrete with probability mass function $p(x)$, then

$$E(g(X)) = \sum_x g(x)p(x),$$

where the sum is over all values that X takes.

We provide a particular example where we illustrate the somewhat confusing notation underlying the concept.

Example 2.14. *If X is a discrete random variable taking the values -6, 6, and 12, with $P(X=-6)=\frac{1}{2}$, $P(X=6)=\frac{1}{3}$ and $P(X=12)=\frac{1}{6}$, Y is a random variable given by $Y=X^2$ - i.e. $Y=g(X)$ where $g(x)=x^2$ - we compute $E(X^2)$.*

Note that the random variable Y takes the two values 36 and 144. The probability Y takes the value 36 is the probability that $X=-6$ or 6 which is $\frac{1}{2}+\frac{1}{3}=\frac{5}{6}$. The probability that $Y=144$ is the probability that $X=12$ which is $\frac{1}{6}$. Hence,

$$E(Y) = 36 \cdot \frac{5}{6} + 144 \cdot \frac{1}{6} = 30 + 24 = 54.$$

Using the formula we derived above, we have

$$E(Y) = E(X^2) = (-6)^2 \cdot \frac{1}{2} + (6)^2 \cdot \frac{1}{3} + (12)^2 \cdot \frac{1}{6} = 18 + 12 + 24 = 54.$$

A particular formula that proves quite useful is for the linear function $g(x) = ax + b$, where a and b are constants. We have

$$\boxed{E(aX+b) = aE(X) + b.}$$

Example 2.15. *Suppose Y is a random variable that represents a gambler's gain on a \$1000 bet on red at the roulette wheel. Compute $E(Y)$.*

Since $Y = 10X$, where X is the gain on a \$100 bet on red, we can use our work in the roulette example above to obtain

$$E(Y) = E(10X) = 10E(X) = 10(-5.263) = -52.63.$$

To determine how scattered the values in a distribution are, statisticians model the distribution with a random variable and take its *variance*. The variance $V(X)$ of a random variable X is given by

$$V(X) = E((X-\mu)^2),$$

and the *standard deviation* of X, $SD(X)$, is defined to be the square root of the variance. We provide a formula for the variance in the case that X is discrete.

Variance of a Discrete Random Variable. For the discrete random variable X with probability mass function $p(x) = P(X = x)$ and expected value μ, the variance of X, $V(X)$, is given by

$$V(X) = \sum_x (x - \mu)^2 p(x),$$

where the sum is over all values that X takes. The standard deviation of X, $SD(X)$, is given by

$$SD(X) = \sqrt{V(X)}.$$

Using the fact that $E(aX + b) = aE(X) + b$, we can establish a more computationally friendly formula for $V(X)$. We have that

$$\begin{aligned} V(X) = E((X-\mu)^2) &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) + E(-2\mu X) + \mu^2 \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2\mu \cdot \mu + \mu^2 \\ &= E(X^2) - \mu^2 \\ &= E(X^2) - (E(X))^2 \ . \end{aligned}$$

We formulate this result.

Shortcut Formula for the Variance of X.

$$V(X) = E(X^2) - (E(X))^2.$$

Example 2.16. *For the random variable X above for which $P(X = -20) = 0.85$, $P(X = 100) = 0.14$, and $P(X = 500) = 0.01$, compute $V(X)$.*

We carry out the computation via the regular formula and then using the shortcut formula.

$$\begin{aligned} V(X) &= \sum_x (x-\mu)^2 f(x) \\ &= (-20-2)^2(0.85) + (100-2)^2(0.14) + (500-2)^2(0.01) \\ &= (484)(0.85) + (9604)(0.14) + (248004)(0.01) \\ &= 4236 \ . \end{aligned}$$

Now by the shortcut formula we have

$$\begin{aligned} V(X) &= \sum_x (x^2) f(x) - \mu^2 \\ &= (20)^2(0.85) + (100)^2(0.14) + (500)^2(0.01) - 2^2 \\ &= (400)(0.85) + (10000)(0.14) + (250000)(0.01) - 4 \\ &= 4236 \ . \end{aligned}$$

In the case that X is a continuous RV, the expected value is determined by an integral.

Expected Value of a Continuous Random Variable. For the continuous random variable X with probability density function $f(x)$, the expected value of X, $E(X)$, is defined to be

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

If the integral diverges, then the random variable is said not to have an expected value.

Example 2.17. *For the random variable X with pdf*

$$f(x) = \begin{cases} x/2 & \text{if } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

compute $E(X)$.

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_0^2 x \cdot \frac{x}{2} dx \\ &= \int_0^2 \frac{x^2}{2} dx \\ &= \left. \frac{x^3}{6} \right|_0^2 \\ &= \frac{4}{3}. \end{aligned}$$

If the expected value of X exists and X only takes nonnegative values, then there's an alternative formula for computing $E(X)$ that involves tail probabilities.

Note that

$$
\begin{aligned}
E(X) &= \int_0^\infty x f(x) dx \\
&= \int_0^\infty \int_0^x 1 \, dy f(x) dx \\
&= \int_0^\infty \int_y^\infty f(x) dx dy \\
&= \int_0^\infty P(X > y) dy.
\end{aligned}
$$

For a continuous random variable X that only takes nonnegative values,

$$E(X) = \int_0^\infty P(X > x) dx,$$

if the expected value exists.

Example 2.18. *Use this probability tails formula to compute the expected value of the continuous random variable X with probability density function*

$$f(x) = \begin{cases} \frac{2}{x^3} & \text{if } x \geq 1 \\ 0 & \text{if } x < 1 \end{cases}$$

We note that if $x < 1$, $P(X > x) = 1$. If $x \geq 1$,

$$P(X > x) = \int_x^\infty \frac{2}{t^3} dt = -\frac{1}{t^2}|_x^\infty = \frac{1}{x^2}.$$

So we have

$$E(X) = \int_0^\infty P(X > x) dx = \int_0^1 1 \, dx + \int_1^\infty \frac{1}{x^2} dx = 1 + (-\frac{1}{x}|_1^\infty) = 1 + 1 = 2.$$

If $g(x)$ is a real valued function on the reals and X is a continuous random variable with pdf $f(x)$, a formula analogous to the one for discrete random variables can be derived for computing the expected value of X.

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

This is the integral version of the sum formula we have for discrete random variables. As was the case with discrete random variables, we have that

$$\boxed{E(aX + b) = aE(X) + b}$$

for the constants a and b, even when X is continuous. The variance $V(X)$ of a continuous random variable X is defined by $V(X) = E((X - \mu)^2)$ just as in the discrete case.

Variance of a Continuous Random Variable. For the continuous random variable X with probability density function $f(x)$, the variance of X, $V(X)$, is given by

$$V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \left(\int_{-\infty}^{\infty} xf(x)dx\right)^2.$$

If one of the integrals diverges, then the random variable is said not to have an expected value.

Example 2.19. *For the random variable X above with pdf*

$$f(x) = \begin{cases} x/2 & \text{if } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

compute $V(X)$.

From our previous work, we have that $E(X) = \frac{4}{3}$. Now

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)dx \\ &= \int_0^2 x^2 \cdot \frac{x}{2}dx \\ &= \int_0^2 \frac{x^3}{2}dx \\ &= \frac{x^4}{8}\Big|_0^2 \\ &= 2. \end{aligned}$$

We therefore have that $V(X) = 2 - \left(\frac{4}{3}\right)^2 = \frac{2}{9}$.

Exercises

(1) Given X is a random variable for which $P(X=0) = 1/4$ and $P(X=1) = 3/4$, compute (a) $E(X)$, (b) $E(X^2)$, (c) $E(2X-3)$, (d) $V(X)$, and (e) $SD(X)$.

Ans. (a) $3/4$, *(b)* $3/4$, *(c)* $-3/2$, *(d)* $3/16$, *and (e)* $\dfrac{\sqrt{3}}{4}$

(2) If we let the random variable X count the number of heads you get when you toss a balanced coin 10 times, then the probability mass function of X is as given in the table. Compute (a) $E(X)$, (b) $V(X)$, (c) $F(2)$, and (d) $F(2.1)$, where F is the CDF.

x	f(x)
0	0.001
1	0.010
2	0.044
3	0.117
4	0.205
5	0.246
6	0.205
7	0.117
8	0.044
9	0.010
10	0.001

Ans. (a) 5 *(b)* 2.5

(3) An insurance company offers an automobile policy structured as follows: The company makes \$600 with probability 0.95, it loses \$300 with probability 0.03, and it loses \$20,000 with probability 0.02. If X represents the company's gain on one of these policies, compute $E(X)$.

Ans. \$161

(4) Suppose X is a continuous random variable with pdf

$$f(x) = \begin{cases} 1/5 & \text{for } 3 \le x \le 8 \\ 0 & \text{otherwise} \end{cases}.$$

Compute (a) $E(X)$ and (b) $V(X)$.

Ans. (a) 5.5 *(b)* $\dfrac{25}{12}$

(5) Suppose X is a continuous random variable with pdf

$$f(x) = \begin{cases} 0 & \text{for } x < 1 \\ \frac{3}{x^4} & \text{if } x \ge 1 \end{cases}.$$

Compute (a) $E(X)$ and (b) $SD(X)$.

Ans. (a) $3/2$ *(b)* $\sqrt{3}/2$

(6) Suppose X is a continuous random variable with pdf

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 3e^{-3x} & \text{if } x \ge 0 \end{cases}.$$

Compute (a) $E(X)$ and (b) $V(X)$.

Ans. (a) $1/3$ *(b)* $1/9$

(7) Compute $E(X)$ if X is a continuous random variable with pdf

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

Ans. $E(X)$ does not exist.

(8) Compute $SD(X)$ if X is a random variable representing a gambler's gain on a \$100 roulette bet on red.

Ans. 99.8614

(9) Use the probability tails formula to compute the expected value of the continuous random variable X with probability density function

$$f(x) = \begin{cases} e^{-x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

(10) Suppose X is a continuous random variable with cumulative distribution function

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{4}x^2 & \text{if } 0 \leq x < 2 \\ 1 & \text{if } x \geq 2 \end{cases}$$

Compute then (a) $E(X)$ and (b) $SD(X)$.

Ans. (a) $\frac{4}{3}$ *(b)* $\frac{\sqrt{2}}{3}$

Chapter 3

Widely Used Discrete Random Variables

3.1. Counting Techniques

In order to understand fully the discrete random variables we present in this chapter, it's necessary to know some basic counting techniques.

First we mention a multiplication rule. If there are m ways of completing one task and n ways of completing a second task, then there are $m \times n$ ways of completing both tasks. This allows us to count the number of outcomes when rolling a six sided die twice. Since there are six outcomes for each roll of the die, the total number of outcomes for two rolls would be $6 \times 6 = 36$. Applying the multiplication rule twice, we see that if you toss a coin three times, the total number of outcomes is $2 \times 2 \times 2 = 8$. The number of four digit numbers is $10 \times 10 \times 10 \times 10 = 10,000$.

A useful operation to employ for counting purposes is that of *factorials*. For natural numbers n, the notation $n!$ - read n factorial - is defined to be

$$n! = n(n-1)(n-2)\cdots 1.$$

For the special case that $n = 0$, we have the definition $0! = 1$. Factorials grow very quickly. It's the case that $6! = 120$, $10! = 3{,}628{,}800$, and $20! = 2.433(10)^{18}$, etc. A decent calculator will compute factorials. With the TI-84 for example, to get 12! input 12, select MATH-PRB-!, and press ENTER, to get $479,001,600$.

We noted the number of four digit numbers is $10,000$. This gives us the number of four digit PINs. If we want to count the number of four digit PINs with distinct numerals (i.e. where none of the numerals in the PIN repeat), we could use the multiplication rule to get

$$10 \times 9 \times 8 \times 7 = 5,040.$$

Note that this could be written

$$10 \cdot 9 \cdot 8 \cdot 7 = 10 \cdot 9 \cdot 8 \cdot 7 \cdot \frac{6!}{6!} = \frac{10!}{6!} = \frac{10!}{(10-4)!}.$$

This is referred to as the number of permutations of 10 things taken four at a time, written ${}_{10}P_4$. In general the number of permutations of n things taken r at a time is given by

$${}_nP_r = \frac{n!}{(n-r)!}.$$

Back to the four digit PIN with non-repeating numerals, the PINs 1563 and 6531 - though they contain the same four numerals - are distinct. In other words the order in which the numerals are arranged is taken into account. If we weren't concerned with order - i.e. if we simply wanted to know how many combinations of four distinct numerals there are chosen from the set $\{1,2,3,4,5,6,7,8,9,0\}$, we note that there are $4 \cdot 3 \cdot 2 \cdot 1 = 24$ orderings of the PINs consisting of the elements of the set $\{1,3,5,6\}$. We list a few of them here:

$$1356, 1365, 1536, 1563, 1635, 1653, 3156, 3165, 3516, 3561, 3615, 3651, ..., 6531$$

So the number of subsets with four elements chosen from a set with 10 elements is $\frac{10!}{4!6!}$.

One might be interested in counting more in general the number of subsets with r elements there are taken from a set with n elements. The notation for this is $\binom{n}{r}$ - read the number of combinations of n things taken r at a time. This mouthful of words is often abbreviated as "n choose r". The formula is

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

An interesting application is to count the number of five card hands in a deck of 52 cards. Noting that this is nothing more than counting the number of subsets with five elements in a set with 52 elements, we have that there are

$$\binom{52}{5} = \frac{52!}{5!47!} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2} = 52 \cdot 51 \cdot 10 \cdot 49 \cdot 2 = 2,598,960.$$

You read that right. There are roughly 2.6 million hands. The reader will note that we took advantage of a pair of cancellations to carry out the computation. We noted first that

$$\frac{52!}{47!} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48 \cdot 47!}{47!} = 52 \cdot 51 \cdot 50 \cdot 49 \cdot 48.$$

We also used the fact that $\frac{50}{5} = 10$ and $\frac{48}{4 \cdot 3 \cdot 2} = 2$.

A calculator will do the work for you. For example, with a TI-84, you input 52, select MATH-PRB-nCr, then input 5, and select ENTER. You'll note the notation nCr that Texas Instruments uses for n choose r. There are a number of ways to represent this in the literature:

$$\binom{n}{r} = nCr = C(n,r) =_n C_r = C_r^n.$$

Exercises

(1) A certain state's license plates numbers consists of three letters A-Z followed by three numerals 0-9. Compute the total number of license plates numbers this state can have.
Ans. 17,576,000

(2) How many subsets with two elements does the set $\{a, b, c, d\}$ have? List them.
Ans. $\binom{4}{2} = 6$*, The subsets are* $\{a, b\}$, $\{a, c\}$, $\{a, d\}$, $\{b, c\}$, $\{b, d\}$, $\{c, d\}$

(3) How many permutations of size two does *abcd* have? List them.
Ans. ${}_4P_2 = 12$*, The permutations are* $ab, ba, ac, ca, ad, da, bc, cb, bd, db, cd, dc$

(4) The Binomial theorem states that for natural numbers n,

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k.$$

Use this theorem to expand (a) $(x+1)^4$, (b) $(2x+3)^3$.
(a) $x^4 + 4x^3 + 6x^2 + 4x + 1$, *(b)* $8x^3 + 36x^2 + 54x + 27$

(5) How many combinations of four different letters are there? How many four letter words are there (including nonsensical ones)? How many four letter words with non-repeating letters are there.
Ans. 14,950, 456,976, 358,800

(6) How many different two person teams can you select from 10 individuals
Ans. 45

(7) Prove that $\binom{n}{n} = \binom{n}{0} = 1$.

(8) Prove that $\binom{n}{1} = n$

(9) Prove that $\binom{n}{r} = \binom{n}{n-r}$.

(10) If you toss a coin 20 times, compute the number of ways you could get exactly eight heads. [Hint: One way would be if the first eight were heads and the next 12 tails; another way would be if the first, third, fifth, seventh, ninth, eleventh, thirteenth, and fifteenth, were heads and the rest tails; etc.]
Ans. 125,970

(11) If you roll a die 10 times, count the number of ways you could get (a) exactly two 6's and (b) exactly three 6's.
Ans. (a) 45, (b) 120

3.2. Bernoulli

Perhaps the simplest of all discrete random variables used for modeling purposes is the Bernoulli. It's named after Jacob Bernoulli, a Swiss mathematician from the latter half of the seventeenth century. The random variable X is said to have a Bernoulli distribution with parameter p if $P(X = 1) = p$ and $P(X = 0) = 1-p$. The parameter p of course has to be a positive number less than 1. Simple computations will produce the mean and standard deviation of the Bernoulli. We have

$$E(X) = 0 \cdot (1-p) + 1 \cdot p = p$$

and

$$SD(X) = \sqrt{E(X^2) - (E(X))^2} = \sqrt{0^2 \cdot (1-p) + 1^2 \cdot p - p^2} = \sqrt{p(1-p)}.$$

3.3. Binomial

Suppose only 15% of airline passengers arriving at an international airport are chosen for complete baggage scrutiny and that these passengers are selected at random. What's the probability that exactly three of the next ten passengers will be selected? To answer this question we first deduce the probability that, of the next ten passengers, the first three are selected and the following seven are not. Since whether or not one passenger is selected is independent of whether or not another is selected, we have that this probability is

$$(0.15)^3(0.85)^7.$$

Now there are many different ways exactly three of the ten passengers can be selected. It could be the first three or the last three, or it could be the third, fifth, and ninth, etc. The number of such ways is equal to the number of subsets of three elements there are in a set of ten elements. In other words, there are

$$\binom{10}{3} = 120$$

ways. Each one of the ways is equally likely, so we have that the answer to our question is

$$\binom{10}{3}(0.15)^3(0.85)^7 = 0.1298.$$

The passenger baggage check example illustrates a binomial model. In order to compute probabilities associated with phenomena involving n independent trials, each of which results in success with probability p, we let the random variable X count the number of successes in the n trials. Then X is said to be a *binomial* random variable with parameters n and p. The probability mass function for this random variable is given by

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k},$$

for $k = 0, 1, 2, ..., n$. A random variable having this probability mass function is said to be binomial with parameters n and p.

We now compute the expected value of such a random variable:

$$\begin{aligned}
E(X) &= \sum_{k=0}^{n} k\tbinom{n}{k}p^k(1-p)^{n-k} \\
&= \sum_{k=1}^{n} k\tbinom{n}{k}p^k(1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{k \cdot n!}{k!(n-k)!}p^k(1-p)^{n-k} \\
&= np\sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-1-(k-1))!}p^{k-1}(1-p)^{(n-1)-(k-1)} \\
&= np\sum_{k=0}^{n-1} \frac{(n-1)!}{k!((n-1)-k)!}p^{k}(1-p)^{(n-1)-k} \\
&= np(p+1-p)^{n-1} \\
&= np.
\end{aligned}$$

We used the Binomial Theorem for the penultimate equality. A similar computation along with some clever algebra allows us to compute $E(X^2) = n(n-1)p^2+np$. We have

$$\begin{aligned}
E(X^2) &= E(X(X-1)+X) \\
&= \sum_{k=0}^{n} k(k-1)\tbinom{n}{k}p^k(1-p)^{n-k} + E(X) \\
&= \sum_{k=2}^{n} k(k-1)\tbinom{n}{k}p^k(1-p)^{n-k} + np \\
&= n(n-1)p^2\sum_{k=2}^{n} \frac{k(k-1)\cdot(n-2)!}{k!(n-k)!}p^{k-2}(1-p)^{n-k} + np \\
&= n(n-1)p^2\sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-2-(k-2))!}p^{k-2}(1-p)^{(n-2)-(k-2)} \\
&\qquad + np \\
&= n(n-1)p^2\sum_{k=0}^{n-2} \frac{(n-2)!}{k!((n-2)-k)!}p^{k}(1-p)^{(n-2)-k} + np \\
&= n(n-1)p^2(p+1-p)^{n-2} + np \\
&= n(n-1)p^2 + np.
\end{aligned}$$

This allows us to compute

$$\begin{aligned}
V(X) &= n(n-1)p^2 + np - (np)^2 \\
&= n^2p^2 - np^2 + np - n^2p^2 \\
&= np - np^2 \\
&= np(1-p).
\end{aligned}$$

Consequently, the mean and standard deviation of a binomial random variable with parameters n and p are np and $\sqrt{np(1-p)}$, respectively.

Since the binomial random variable is so widely used for modeling purposes, we repeat the distribution's basic formulas.

Binomial Distribution. If X is a binomial random variable with parameters n and p (i.e. X counts the number of successes in n independent trials, each of which results in success with probability p), then

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

for $k = 0, 1, 2, ..., n$. Moreover, $E(X) = np$ and $SD(X) = \sqrt{np(1-p)}$.

Exercises

(1) If you toss a fair coin 10 times, what's the probability you get (a) exactly five heads, (b) exactly three heads, and (c) exactly seven heads? Round your answers off to four decimal places.

Ans. (a) 0.2461 *(b)* 0.1172 *(c)* 0.1172

(2) If you roll a balanced die 20 times, what's the probability you get (a) exactly four 6's? (b) at least two 6's?

Ans. (a) 0.2022 *(b)* 0.8696

(3) If you take a 10 question true false exam by guessing on each question, what's the probability (a) you get exactly seven questions correct? (b) you pass the exam be getting at least 60% of the questions correct?

Ans. (a) $\frac{120}{1024} = 0.1172$, *(b)* 0.377

(4) Assuming it's equally likely a couple will have a boy or a girl, what's the probability that a couple having five children will have (a) exactly two boys ? (b) all boys?

Ans. (a) 0.3125 *(b)* 0.03125

(5) Only 15% of motorists come to a complete stop at a certain four way stop intersection. What's the probability that of the next ten motorists to go through that intersection (a) none come to a complete stop, (b) at least one comes to a complete stop, and (c) exactly two come to a complete stop.

Ans. (a) 0.1969, *(b)* 0.8031, *(c)* 0.2759

(6) Given X is a binomial random variable with $n = 20$ and $p = \frac{3}{4}$, compute $F(18)$ where $F(x)$ is the CDF of X.

Ans. $F(18) = P(X \le 18) = 1 - P(X = 19 \text{ or } 20) = 0.9757$

(7) Given X is a binomial random variable with $n = 10$ and $p = \frac{1}{2}$, compute $P(\mu - \sigma \le X \le \mu + \sigma)$ where μ is the expected value and σ the standard deviation of X.

Ans. $P(\mu - \sigma \le X \le \mu + \sigma) = P(5 - \sqrt{2.5} \le X \le 5 + \sqrt{2.5}) = P(X = 4, 5, \text{ or } 6) = 0.65625.$

(8) Given X is a binomial random variable with mean 6 and standard deviation 2, compute $P(X = 5)$.

Ans. 0.1812

3.4. Hypergeometric

Another discrete random variable that is closely related to the binomial is the hypergeometric. For this model, there are a number of trials where the probability of success in each trial changes depending on what happens in previous trials. Suppose we select n objects from a lot of N objects without replacement. Suppose, moreover that M of the objects in the lot of N are of a characteristic of interest. We can then compute the probability that k of the n objects we select are of the characteristic. Letting X count the number of objects of the characteristic of interest in the selection of n objects, we have

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}.$$

A random variable with this probability mass function is said to hypergeometric with parameters N, M, and n.

We provide a simple example.

Example 3.1. *Suppose five cards are dealt without replacement from a well-shuffled deck of 52 playing cards. What's the probability exactly two of the five are aces?*

We note that if X counts the number of aces dealt, then X is hypergeometric with the characteristic of interest being that the card is an ace. We have $n = 5$, $N = 52$, and $M = 4$. The answer is

$$P(X = 2) = \frac{\binom{4}{2}\binom{52-4}{5-2}}{\binom{52}{5}} = 0.0399.$$

A somewhat intricate derivation yields

$$E(X) = n(\frac{M}{N})$$

and

$$V(X) = n\Big(\frac{M}{N}\Big)\Big(1 - \frac{M}{N}\Big)\Big(\frac{N-n}{N-1}\Big).$$

It's interesting to note that if $M \to \infty$ and $N \to \infty$ in such a way that M/N stays constant, say $\frac{M}{N} = p$, then $E(X) = np$ and $\lim\limits_{M,N\to\infty} V(X) = np(1-p)$. Indeed, when M and N are really large, M/N changes very little from selection to selection in a small sample from the lot, so that the hypergeometric is approximated by the binomial with $p = \frac{M}{N}$. After, summarizing the important formulas for the hypergeometric, we illustrate with an example.

The hypergeometric random variable counts the number of objects of a characteristic of interest drawn from a lot of size N when M objects in the lot have the characteristic and the sample size is n. The hypergeometric random variable X with the parameters N, M, and n has probability mass function

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

for $k = 0, 1, 2, ..., \min\{M, n\}$. This random variable is approximately binomial with parameters n and $p = \frac{M}{N}$ if N and M are really large and n is relatively small.

Example 3.2. *Suppose a city has 322,000 registered voters, 58% of whom support a certain referendum. In a random sample of 20 voters from that city, what's the probability that exactly 12 support the referendum?*

Letting X count the number in the sample who support the referendum, we have that X is hypergeometric with $N = 322,000$, $M = (0.58)(322,000) = 186,760$, and $n = 20$. The answer is therefore

$$P(X = 12) = \frac{\binom{186760}{12}\binom{322000-186700}{20-12}}{\binom{322000}{20}} = 0.1774.$$

Since M and N are so large, the binomial with $n = 20$ and $p = 0.58$ gives us a good approximation:

$$P(X = 12) = \binom{20}{12}(0.58)^{12}(1 - 0.58)^{20-12} = 0.1768.$$

Exercises

(1) If 13 cards are to be dealt without replacement from a well-shuffled deck of 52 cards, compute the probability that (a) exactly five will be picture cards and (b) exactly six will be picture cards. Keep in mind that a deck of 52 cards has 12 picture cards.

Ans. *(a)* 0.0959 *(b)* 0.0271

(2) If five cards are to be dealt without replacement from a well-shuffled deck of 52 cards, compute the probability that (a) exactly one will be an ace, (b) exactly two will be an ace, (c) at least one will be an ace. Keep in mind that a deck of 52 cards has four aces.

Ans. *(a)* 0.2995 *(b)* 0.0399 *(c)* 0.3412

(3) A calculus professor has 100 students in her class. She randomly selects five students and tests them individually to see if they can find the antiderivative of $x \sin x$. If all five can find it, then she'll conclude the entire class knows how to find it. There are actually 20 students in the class who are not able to calculate the antiderivative. What's the probability she'll mistakenly conclude the whole class can do it?

Ans. 0.3193

(4) Suppose a unit in a certain company consists of 12 engineers, five accountants, and three administrative assistants. If a manager randomly selects four individuals from the unit to be on a committee, what's the probability the committee will have (a) exactly two engineers (b) all engineers?
Ans. (a) 0.3814 *(b)* 0.1022

(5) Suppose a vending machine is loaded with 120 soft drinks, 10 of which are past the expiration date. What's the probability that a customer who buys three drinks from the machine right after it's loaded will get at least one with an expired date? Assume the drinks were loaded randomly.
Ans. 0.2315

(6) Suppose a company manufactures 1.2 million TV sets, 7200 of which have a defective power switch. Compute the probability to seven decimal places that a restaurant that buys 15 of this company's TV sets will have at least one with a defective power switch using (a) the hypergeometric distribution and (b) the binomial distribution.
(a) .0863170 *(b)* 0.0863165

3.5. Poisson

The *Poisson* distribution - named after the great French mathematician and physicist Simeon Poisson of the early nineteenth century - arises as a discrete limiting model of the binomial. We let X be binomial with parameters n and p and consider a limiting version of this random variable where $n \to \infty$ and $p \to 0$ in such a way that np stays constant, say $np = \lambda$. We note that for $k = 0, 1, 2, ..., n$,

$$\begin{aligned}
\lim_{n\to\infty,\ p\to 0} P(X=k) &= \lim_{n\to\infty, p\to 0} \tbinom{n}{k} p^k (1-p)^{n-k} \\
&= \lim_{n\to\infty,\ p\to 0} \Big(\frac{n(n-1)\cdots(n-k+1)}{k!}\Big)\Big(\frac{\lambda}{n}\Big)^k\Big(1-\frac{\lambda}{n}\Big)^{n-k} \\
&= \frac{\lambda^k}{k!} \lim_{n\to\infty,\ p\to 0} \Big(\frac{n(n-1)\cdots(n-k+1)}{n^k}\Big)\Big(1-\frac{\lambda}{n}\Big)^{n-k} \\
&= \frac{\lambda^k}{k!} \lim_{n\to\infty,\ p\to 0} (1)\Big(1-\frac{1}{n}\Big)\Big(1-\frac{2}{n}\Big)\cdots\Big(1-\frac{k-1}{n}\Big) \\
&\qquad \cdot\Big(1-\frac{\lambda}{n}\Big)^n\Big(1-\frac{\lambda}{n}\Big)^{-k} \\
&= \frac{\lambda^k}{k!}\,(1)(e^{-\lambda})(1) \\
&= e^{-\lambda}\frac{\lambda^k}{k!}\ .
\end{aligned}$$

The factor immediately preceding $e^{-\lambda}$ in the second to last line in fact has limit 1 since it consists of the finite number k of factors, each having limit 1.

The random variable X is said to be Poisson with parameter λ if

$$P(X=k) = e^{-\lambda}\Big(\frac{\lambda^k}{k!}\Big),$$

for $k = 0, 1, 2, 3, \ldots$. As the limiting derivation above shows, the Poisson with parameter $\lambda = np$ approximates a binomial random variable with parameters n and p if n is large and p is close to zero. In practice, this approximation is good if $n \geq 100$, $p \leq 0.01$, and $np \leq 20$.

Computations show that

$$E(X) = V(X) = \lambda.$$

Note

$$\begin{aligned}
E(X) &= \sum_{k=0}^{\infty} k \cdot e^{-\lambda}\left(\frac{\lambda^k}{k!}\right) \\
&= \sum_{k=1}^{\infty} k \cdot e^{-\lambda}\left(\frac{\lambda^k}{k!}\right) \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty}\left(\frac{\lambda^{k-1}}{(k-1)!}\right) \\
&= \lambda e^{-\lambda} \sum_{k=0}^{\infty}\left(\frac{\lambda^k}{k!}\right) \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda.
\end{aligned}$$

Also,

$$\begin{aligned}
E(X(X-1)) &= \sum_{k=0}^{\infty} k(k-1) \cdot e^{-\lambda}\left(\frac{\lambda^k}{k!}\right) \\
&= \sum_{k=2}^{\infty} k(k-1) \cdot e^{-\lambda}\left(\frac{\lambda^k}{k!}\right) \\
&= \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty}\left(\frac{\lambda^{k-2}}{(k-2)!}\right) \\
&= \lambda^2 e^{-\lambda} \sum_{k=0}^{\infty}\left(\frac{\lambda^k}{k!}\right) \\
&= \lambda^2 e^{-\lambda} e^{\lambda} \\
&= \lambda^2.
\end{aligned}$$

We therefore have that the variance is

$$E(X^2) - (E(X))^2 = [E(X^2) - E(X)] + E(X) - (E(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

We provide here the basic properties of the Poisson before applying them to an example problem.

The Poisson random variable X with parameter λ has the probability mass function
$$P(X = k) = e^{-\lambda}\left(\frac{\lambda^k}{k!}\right),$$
for $k = 0, 1, 2, 3, \ldots$. The mean and standard deviation are $E(X) = \lambda$ and $SD(X) = \sqrt{\lambda}$.

Example 3.3. *If 0.8% of the population has a certain disease, compute the probability that exactly three of two hundred randomly chosen individuals will have the disease. Carry out this computation to four decimal places using (a) the binomial distribution and (b) the Poisson distribution.*

We let X count the number of individuals in the sample to have the disease. Then (a) according to the binomial model we have that
$$P(X = 3) = \tbinom{200}{3}(0.008)^3(1 - 0.008)^{200-3} = 0.1382,$$
and (b) according to the Poisson model we have that
$$P(X = 3) \cong e^{-200(0.008)} \frac{[(200)(0.008)]^3}{3!} = 0.1378.$$
You'll note that the approximation from the Poisson is good to three decimal places.

Another application of the Poisson is to count the number of occurrences of an event in a given time interval. For example one can model the number of objects in a queue over time. The application arises under three assumptions for the random variable $N(t)$ that counts the number of events occurring during the time interval $[0, t]$:

(1) $P(N(t) = 1) = \lambda t + o(t)$ where λ is a constant

(2) $P(N(t) \geq 2) = o(t)$

(3) The number of events occurring in disjoint subintervals of $[0, t]$ are independent.

The expression $o(t)$ represents a function $f(t)$ with the property that
$$\lim_{t \to 0} \frac{1}{t} f(t) = 0.$$
The reader will note, that a function meeting this requirement will have to approach zero rapidly. An example is $f(t) = t^2$. Less precisely, the assumptions are that (1) the probability that a single event occurs in a small time interval of length t is λt plus a term that is small in relation to t, (2) the probability that two or more events occur in a small interval of length t is small in relation to t, and (3) that which occurs in one interval has no probability effect on what happens in another disjoint interval.

To compute $P(N(t) = k)$ we divide the interval $[0, t]$ into n subintervals, each of length t/n. Then we can write $P(N(t) = k)$ as the sum of two probabilities which we will discuss. The first is the probability that exactly one event occurs in each of k subintervals and that no events occur in any of the other $n - k$ subintervals.

The second is the probability that $N(t) = k$ and that two or more events occur in at least one subinterval. The reader will note that the second of these probabilities is bounded above by

$$\sum_{k=1}^{n} o(t/n) = n \cdot o(t/n) = t\Big(\frac{o(t/n)}{t/n}\Big).$$

Since $t/n \to 0$ as $n \to \infty$, we have that

$$t\Big(\frac{o(t/n)}{t/n}\Big) \to 0$$

as $n \to \infty$.

The first of the two probabilities is equal to

$$\tbinom{n}{k}\Big(\frac{\lambda t}{n} + o(t/n)\Big)^k \Big(1 - \frac{\lambda t}{n} - o(t/n)\Big)^{n-k}.$$

Recalling that the binomial with large n and small p approximates the Poisson with parameter np, we consider

$$n\Big(\frac{\lambda t}{n} + o(t/n)\Big) = \lambda t + t\Big(\frac{o(t/n)}{t/n}\Big)$$

which approaches λt as $n \to \infty$. Hence, by letting $n \to \infty$, this first probability becomes

$$e^{-\lambda t}\frac{(\lambda t)^k}{k!}.$$

Consequently, we arrive at

$$P(N(t) = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}$$

for $k = 0, 1, 2, ..., .$

We provide an application.

Example 3.4. *The number of tornadoes touching down per year in the two parish Caddo/Bossier region is Poisson with mean 19. Compute (a) the probability that there will be exactly 16 tornadoes to touch down in this region next year and (b) that there will be exactly 40 to touch down over the next two years.*

(a) If X counts the number of tornadoes to touch down in Caddo/Bossier Parish region next year, then $E(X) = \lambda = 19$. We thus have that

$$P(X = 16) = e^{-19}\frac{(19)^{16}}{16!} = 0.0772.$$

If Y counts the number to touch down over the next two years, then $E(Y) = (\lambda)(2) = 38$, so that

$$P(Y = 40) = e^{-38}\frac{(38)^{40}}{40!} = 0.0598.$$

Exercises

(1) If X is Poisson with parameter $\lambda = 3$, compute (a) $P(X = 4)$, (b) $P(X \leq 2)$, (c) $E(X)$, (d) $SD(X)$, and (e) $P(X \geq 1)$.
Ans. (a) 0.1680*, (b)* 0.4232*, (c)* 3*, (d)* $\sqrt{3}$*, and (e)* 0.9502

(2) The number of cars that go through Griff's drive-through during the noon hour is Poisson with mean 21.6. What's the probability that during the noon hour tomorrow (a) exactly 22 go through the drive-through (b) at least seven pass through the drive-through.
Ans. (a) 0.0844 *(b)* 0.99992

(3) Suppose 0.3% of the population has a certain disease. In a random sample of 400 people, what's the probability that (a) at least one has the disease and (b) exactly two have the disease? Compute the answer for each part using both the binomial distribution and the Poisson.
Ans. (a) 0.69935*,* 0.69881*, (b)* 0.21723*,* 0.21686

(4) Suppose 0.6% of the TV sets a company manufactures have a defective power switch. Compute the probability that a university that buys 150 of this company's TV sets will have just one with a defective power switch using (a) the binomial distribution and (b) the Poisson distribution.
Ans. (a) 0.3671 *(b)* 0.3659

(5) Suppose there are 6.2 accidents per year on average on a certain interstate entrance ramp. Use the Poisson distribution to compute the probability there will be at most four accidents on this entrance ramp over the next two years.
Ans. 0.0057

3.6. Geometric

Suppose only eight percent of New Orleans residents attended the last Saints football game. If you randomly select residents of the city trying to find one who attended the game, what's the probability that you don't encounter an attendee until the sixth selection? For this to happen, the first five selected would have had to be non-attendees and the sixth an attendee. The probability would therefore be

$$(1 - 0.08)^5(0.08) = 0.0527.$$

The fact that residents are selected randomly makes the selections independent, allowing us to arrive at the answer by multiplying the probabilities of each of the five non-attendees times the probability of the the subsequent attendee. Those who study random phenomena use what is called the *geometric* random variable to model this situation.

We consider a succession of independent trials, each of which results in success with probability p. The geometric random variable X counts the number of trials required to encounter the first success. In the example just related, we have that $P(X = 6) = 0.0527$. In general we have:

> The geometric random variable X counts the number of independent trials it takes to obtain a success when an individual trial results in success with probability p. It's mass function is given by
> $$P(X = n) = (1-p)^{n-1}p,$$
> for $n = 1, 2, 3, ...$ The mean and standard deviation are $E(X) = 1/p$ and $SD(X) = \frac{\sqrt{1-p}}{p}$.

The reader will note that the probability mass function is in fact legitimate since the probabilities of all the values X can take sum to 1:

$$\sum_{n=1}^{\infty}(1-p)^{n-1}p = p\sum_{n=1}^{\infty}(1-p)^{n-1} = p \cdot \frac{1}{1-(1-p)} = 1.$$

To compute the mean, note that

$$\begin{aligned}
E(X) &= \sum_{n=1}^{\infty} n(1-p)^{n-1}p \\
&= p + \sum_{n=2}^{\infty} n(1-p)^{n-1}p \\
&= p + (1-p)\sum_{n=2}^{\infty} n(1-p)^{n-2}p \\
&= p + (1-p)\sum_{n=1}^{\infty}(n+1)(1-p)^{n-1}p \\
&= p + (1-p)(E(X) + \sum_{n=1}^{\infty} 1(1-p)^{n-1}p) \\
&= p + (1-p)(E(X) + p \cdot \frac{1}{1-(1-p)}) \\
&= p + (1-p)(E(X) + 1) \\
&= (1-p)E(X) + 1.
\end{aligned}$$

Solving $E(X) = (1-p)E(X) + 1$ for $E(X)$ yields $E(X) = \frac{1}{p}$. Computing the variance of X requires a bit more ingenuity. The classic derivation involves writing $X^2 = X(X-1) + X$ to get a value for $E(X^2)$.

Exercises

(1) When flipping a balanced coin, what's the probability that you get the first heads on the fourth try?

Ans. $\frac{1}{16}$

(2) If X is a random variable counting the number of times you have to flip a balanced coin to get the first heads, (a) determine the probability mass function for X, (b) compute $E(X)$, and (c) compute $V(X)$.

Ans. $p(n) = \frac{1}{2^n}$, $E(X) = V(X) = 2$

(3) If X is a random variable counting the number of tosses of a balanced die it takes to get the first six, (a) determine the probability mass function for X, (b) compute $E(X)$, and (c) compute $V(X)$.
Ans. $p(n) = \frac{5^{n-1}}{6^n}$, $E(X) = 6$, $V(X) = 30$

(4) The probability an adult male in Finland has a interpupillary distance (IPD) of less than 60 mm is 22%. If you measure the IPD of randomly selected Finnish adult males, what's the probability the tenth one selected is the first one to have an IPD less than 60?
Ans. 0.0235

3.7. Negative Binomial

The negative binomial random variable generalizes the geometric. Again we consider a number of independent trials, each resulting in success with probability p. This time our random variable X counts the number of trials it takes to obtain r successes (instead of just one success as was the case with the geometric). To compute $P(X = n)$ we'll note that in the first $n-1$ trials there need to be $r-1$ successes, and the nth trial needs to result in success. Consequently,

$$P(X = n) = \binom{n-1}{r-1} p^{r-1}(1-p)^{n-1-(r-1)} \cdot p = \binom{n-1}{r-1} p^r (1-p)^{n-r}.$$

A technique we introduce later will allow us to derive the formulas for the mean and standard deviation. We summarize in the table.

For the negative binomial random variable X counting the number of independent trials (each with success probability p) necessary to obtain r successes,

$$P(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r},$$

for $n = 1, 2, 3, \ldots.$ The mean and standard deviation are $E(X) = r/p$ and $SD(X) = \frac{\sqrt{r(1-p)}}{p}$.

Exercises

(1) What's the probability you have to roll a die eight times in order to get two 6's?
Ans. 0.0651

(2) What's the probability you have to roll a die 15 times in order to get four 1's?
Ans. 0.0378

(3) A real estate analyst knows that in a certain county 23% of the houses have a selling price higher than \$300,000. What's the probability that during the next month in that county, the fifth house sold at a price of more than \$300,000 is the twentieth house sold that month?
Ans. 0.0495

Chapter 4

Widely Used Continuous Random Variables

4.1. Uniform

In previous chapters we studied theoretical continuous random variables just to gain an understanding of probability density functions. Now we look at several continuous random variables that are used extensively to model random phenomena.

The uniform distribution on the interval [a,b], $a < b$, is given by the random variable X with probability density function

$$f(x) = \left\{ \begin{array}{cc} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{otherwise} \end{array} \right\}.$$

Note that $f(x)$ is in fact a probability density function since $f(x) \ge 0$ for all x and

$$\int_{-\infty}^{\infty} f(x)dx = \int_a^b \frac{1}{b-a}dx = \frac{1}{b-a} \cdot x|_a^b = \frac{1}{b-a}(b-a) = 1.$$

In one of the exercises, the reader is asked to show that the random variable's mean and variance are $\frac{a+b}{2}$ and $\frac{(b-a)^2}{12}$, respectively. We box in the basic formulas for the uniform distribution before presenting an example.

Uniform Distribution. The random variable X is said to be uniform on the interval [a,b], $a < b$, if its probability density function is

$$f(x) = \left\{ \begin{array}{cc} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{otherwise} \end{array} \right\}.$$

For this random variable $E(X) = \dfrac{a+b}{2}$ and $SD(X) = \dfrac{b-a}{\sqrt{12}}$.

Example 4.1. *When a clock's battery runs out, the location at which the minute hand stops is uniform with pdf*

$$f(x) = \left\{ \begin{array}{cc} \frac{1}{12} & \textit{for } 0 \leq x \leq 12 \\ 0 & \textit{otherwise} \end{array} \right\}.$$

The 0 and 12 refer to the hours on the clock face, of course. Compute the probability that the minute hand stops somewhere between 10 and 11 o'clock.

$$P(10 < X < 11) = \int_{10}^{11} \tfrac{1}{12} dx = \tfrac{x}{12}\Big|_{10}^{11} = \tfrac{1}{12}.$$

Exercises

(1) Given that X is a uniform random variable on the interval $[-15, 185]$, compute (a) $P(X < 0)$, (b) $P(10 < X < 50)$, and (c) $P(X = 85)$.
Ans. (a) 0.075 (b) 0.2 (c) 0

(2) Computed the 75th percentile of X in the previous problem.
Ans. 135

(3) If the random variable X is uniform on $[a, b]$, show that (a) $E(X) = \frac{a+b}{2}$ and (b) $V(X) = \frac{(b-a)^2}{12}$.

(4) Use the formulas in the previous problem to compute (a) the mean and (b) the standard deviation of the random variable in the first exercise in this set.
Ans. (a) 85 *(b)* 57.73503

(5) If $F(x)$ is the CDF of a uniform random variable on the interval $[2, 6]$, compute (a) $F(3)$, (b) $F(4)$, and (c) $F(100)$.
Ans. (a) 0.25 *(b)* 0.5 *(c)* 1

(6) Find a rule for the function $F(x)$ in the previous exercise.
Ans. $F(x) = \left\{ \begin{array}{cc} 0 & \textit{if } x < 2 \\ \frac{x-2}{4} & \textit{if } 2 \leq x < 6 \\ 1 & \textit{if } x \geq 6 \end{array} \right.$

(7) Find a rule for the CDF $F(x)$ of the random variable X that is uniform on the interval $[a, b]$. The rule should consist of the three pieces where $x < a$, $a \leq x < b$, and $x \geq b$.
Ans. $F(x) = \left\{ \begin{array}{cc} 0 & \textit{if } x < a \\ \frac{x-a}{b-a} & \textit{if } a \leq x < b \\ 1 & \textit{if } x \geq b \end{array} \right.$

4.2. Exponential

The random variable X with parameter $\lambda > 0$ is said to be an exponential random variable if it has probability density function

$$f(x) = \left\{ \begin{array}{cc} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{array} \right. .$$

Many phenomena, such as life times of organisms and interarrival times of customers at a store or calls made to telephone company, etc., can be modeled accurately with exponential distributions. That this is the case is understandable if we consider interarrival times of events modeled by the Poisson.

The reader will recall that we studied the random variable $N(t)$ that counts the number of events occurring during the time interval $[0, t]$. We imposed the assumptions that (1) the probability that a single event occurs in the time interval is λt plus a term that is small in relation to t, (2) the probability that two or more events occur in the interval is small in relation to t, and (3) that which occurs in one interval has no probability effect on what happens in another disjoint interval. The resulting probability mass function at which we arrived for $N(t)$ was the Poisson:

$$P(N(t) = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}$$

for $k = 0, 1, 2, ...$

We now consider the time T on an interval up until which the first event occurs. We have that

$$\begin{aligned} F_T(t) &= P(T \leq t) \\ &= 1 - P(T > t) \\ &= 1 - P(N(t) = 0) \\ &= 1 - e^{-\lambda t}\frac{(\lambda t)^0}{0!} \\ & \quad 1 - e^{-\lambda t} \ . \end{aligned}$$

Consequently,

$$f_T(t) = F'_T(t) = \frac{d}{dt}(1 - e^{-\lambda t}) = \lambda e^{-\lambda t}.$$

The reader will recognize this as the probability density function for the exponential random variable with parameter λ. Subsequent interarrival times are independent and follow the same distribution.

Routine computations show that if X is exponential with parameter λ, $E(X) = \frac{1}{\lambda}$ and $SD(X) = \frac{1}{\lambda}$. We carry out the integration by parts necessary to obtain the mean. By letting $u = x$ and $dv = \lambda e^{-\lambda x} dx$ we get $du = dx$ and $v = -e^{-\lambda x}$, giving us

$$\begin{aligned} E(X) &= \int_0^\infty x \cdot \lambda e^{-\lambda x} dx \\ &= -xe^{-\lambda x}|_0^\infty + \int_0^\infty e^{-\lambda x} dx \\ &= 0 - 0 - \frac{1}{\lambda}e^{-\lambda x}|_0^\infty \\ &= \frac{1}{\lambda}. \end{aligned}$$

Note that the computation of $\lim_{x\to\infty} xe^{-\lambda x}$ in this integral requires use of L'Hospital's Rule:

$$\lim_{x\to\infty} xe^{-\lambda x} = \lim_{x\to\infty} \frac{x}{e^{\lambda x}} = \lim_{x\to\infty} \frac{1}{\lambda e^{\lambda x}} = 0.$$

We summarize the basics for the exponential distribution and then present some examples.

Exponential Distribution. The exponential random variable X with parameter $\lambda > 0$ has probability density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

For this distribution, $E(X) = SD(X) = 1/\lambda$.

Example 4.2. *The time between calls at a phone company is exponentially distributed with mean 4 s. What's the probability that the time between the next two calls is more than 5 s?*

Let $X =$ the inter-arrival time of calls in seconds. Then

$$P(X \geq 5) = \int_5^\infty \frac{1}{4} e^{-\frac{x}{4}}\, dx = -e^{-\frac{x}{4}}\big|_5^\infty = e^{-\frac{5}{4}} \doteq 0.287.$$

Example 4.3. *If X is an exponential RV for which $P(X < 1) = 0.90$, find t so that $P(X > t) = 0.05$. (Note that this value of t is called the 95th percentile of X, and 1 is the 90th percentile.*

We first find λ:

$$0.90 = P(X < 1) = \int_0^1 \lambda e^{-\lambda x} dx = -e^{-\lambda x}|_0^1 = -e^{-\lambda} + 1,$$

so $e^{-\lambda} = 0.10$ and $\lambda = -\ln 0.10 \doteq 2.3026$.

Now we have

$$0.05 = P(X > t) \doteq \int_t^\infty 2.3026 e^{-2.3026x} dx = \cdots = e^{-2.3036t},$$

so that $t = -\frac{1}{2.3026} \ln 0.05 \doteq 1.301$.

Exercises

(1) Given X is an exponential random variable with mean 4, compute the probabilities (a) $P(X \leq 4)$ and (b) $P(X \leq 2)$ to four decimal places.
Ans. (a) 0.6321 *(b)* 0.3935

(2) Given that X is an exponential random variable with parameter $\lambda = 1$, find a rule for the distribution function F of X.
Ans. $F(x) = \begin{cases} 0 & \textit{if } x < 0 \\ 1 - e^{-x} & \textit{if } x \geq 0 \end{cases}$

(3) Find (a) the median and (b) the mean of the exponential random variable X that has parameter $\lambda = 1$. Why is one greater than the other?
Ans. *(a)* $\ln 2$ *(b)* 1

(4) Find (a) the mean, (b) the standard deviation, and the (c) median of the exponential random variable X that has parameter $\lambda = 3$.
Ans. *(a)* $1/3$, *(b)* $1/3$, *(c)* $-(1/3)\ln(1/2) = 0.2310$

(5) Suppose you model the lifetime of a certain battery with an exponential random variable that has mean 2.5 hrs. According to this model, what's the probability that a randomly selected such battery lasts for more than three hours?
Ans. 0.3012

(6) If X is an exponential random variable for which $P(X < 1) = 0.80$, find t so that $P(X > t) = 0.05$.
Ans. 1.8614

4.3. Gamma

Playing an integral role in the *gamma* distribution is the gamma function

$$\Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha-1} dt.$$

Integrating by parts, one can establish the relationship $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$. Using this formula when α is a positive integer n and noting that $\Gamma(1) = 1$, we have that

$$\Gamma(n) = (n-1)!.$$

The probability density function for the gamma distribution with positive parameters α and β is given by

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

The reader will note that the gamma distribution reduces to the exponential with parameter λ for $\alpha = 1$ and $\beta = \frac{1}{\lambda}$. The gamma distribution models the waiting time until the αth event occurs.

We summarize the basics for the gamma distribution.

Gamma Distribution. The gamma random variable X with parameters $\alpha > 0$ and $\beta > 0$ has probability density function

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

For this distribution, $E(X) = \alpha\beta$ and $SD(X) = \beta\sqrt{\alpha}$.

4.4. Normal

Many populations of values follow what is called a *normal* distribution. Examples are weights of female individuals, heights of corn stalks, lengths of "8.5×11" sheets of paper, etc., Moreover, one can apply a result called the *Central Limit Theorem* to model the averages of random values from any distribution with the normal distribution. Hence, the normal distribution plays a major role in statistics.

> **Normal Distribution**. The random variable X is said to be a normal with parameters μ and $\sigma > 0$ if it has probability density function
> $$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

This pdf has some notable characteristics:

1) $f(x) > 0$ for all x

2) $f(\mu) = \dfrac{1}{\sqrt{2\pi}\sigma}$

3) $f(\mu + a) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{a^2}{2\sigma^2}} = f(\mu - a)$, so f is symmetric about $x = \mu$.

4) $f'(x) = -\dfrac{(x-\mu)}{\sqrt{2\pi}\sigma^3} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, so $f'(x) < 0$ if $x > \mu$ and $f'(x) > 0$ if $x < \mu$. Thus, f is increasing for $x < \mu$ and decreasing for $x > \mu$.

5) $\lim_{x\to\infty} f(x) = 0 = \lim_{x\to-\infty} f(x)$

6) $f''(x) = [x - (\mu - \sigma)][x - (\mu + \sigma)]\dfrac{1}{\sqrt{2\pi}\sigma^5} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, so $f''(x) = 0$ if $x = \mu \pm \sigma$, $f''(x) < 0$ if $\mu - \sigma < x < \mu + \sigma$, and $f''(x) > 0$ if $x < \mu - \sigma$ or $x > \mu + \sigma$. Hence, f is concave downward on the interval $(\mu - \sigma, \mu + \sigma)$ and concave upward outside this interval, with inflection points at $\mu \pm \sigma$.

Putting all this together we see that the probability density function for the normal random variable is a bell-shaped curve symmetric about $x = \mu$. The larger the value of σ the flatter the curve.

Example. The number of ounces of Coke in a "12 oz" can is normally distributed with $\mu = 12$ and $\sigma = 0.2$. What's the probability that a randomly selected 12 oz can of Coke has (a) between 11.9 and 12.1 oz? (b) between 11.5 and 12.5 oz? (c) at least 11.8 oz?

(a) Letting $f(x)$ be the pdf for the normal random variable X with mean 12 and standard deviation 0.2, we see that

$$P(11.9 < X < 12.1) = \int_{11.9}^{12.1} f(x)dx \doteq 0.383.$$

The integral cannot be evaluated algebraically, so a graphing calculator or computer algebra system is used.

(b) $P(11.5 < X < 12.5) = \int_{11.5}^{12.5} f(x)dx \doteq 0.988$

(c) $P(X \geq 11.8) = \int_{11.8}^{\infty} f(x)dx = \int_{11.8}^{12} f(x)dx + 0.5 \doteq 0.841.$

Since $f(x)$ is symmetric about $x = 12$, half of X's probability is to the right of 12.

The *standard normal* random variable is a normal random variable with $\mu = 0$ and $\sigma = 1$. It is denoted by Z. The pdf of Z is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

A double integral in polar coordinates can be used to show that this function integrates to one:

If we let $I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$, we have

$$\begin{aligned}
I^2 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dxdy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dxdy \\
&= \int_{0}^{2\pi} \int_{0}^{\infty} \frac{1}{2\pi} e^{-\frac{r^2}{2}} rdrd\theta \\
&= \int_{0}^{2\pi} -\frac{1}{2\pi} e^{-\frac{r^2}{2}} |_0^{\infty} d\theta \\
&= \int_{0}^{2\pi} \frac{1}{2\pi} d\theta \\
&= 1
\end{aligned}$$

so that $I = 1$.

When one needs to compute a probability associated with a normal distribution, the integration cannot be done with pencil and paper. The usual procedure is to convert the integrand to a standard normal density function using the substitution $u = \frac{x - \mu}{\sigma}$ and then use a widely available standard normal distribution table to approximate the integral. Since

$$\begin{aligned}
P(\frac{X-\mu}{\sigma} < c) &= P(X < c\sigma + \mu) \\
&= \int_{-\infty}^{c\sigma+\mu} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx \\
&= \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\, du \\
&= P(Z < c),
\end{aligned}$$

it's the case that $\frac{X-\mu}{\sigma}$ and Z have the same distribution. Statisticians often write

$$Z \sim \frac{X-\mu}{\sigma}$$

and refer to $\frac{X-\mu}{\sigma}$ as a Z-score. As a result, in the preceding example we could have written

$$\begin{aligned}
P(11.9 < X < 12.1) &= P(\frac{11.9-12}{0.2} < Z < \frac{12.1-12}{0.2}) \\
&= P(-0.5 < Z < 0.5) \\
&= P(Z < 0.5) - P(Z \leq -0.5) \\
&= \int_{-\infty}^{0.5} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx - \int_{-\infty}^{-0.5} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx \\
&\doteq 0.6915 - 0.3085 \\
&= 0.383.
\end{aligned}$$

Since there's not a simple rule for the cumulative distribution function of the standard normal random variable Z, statisticians typically use a table. The tabular entries are for $P(Z \leq x)$ - for x between zero and 3.59 in increments of 0.01. Since Z has a probability density function symmetric about $x = 0$ and more than 99.9% of Z's probability fall between -3.5 and 3.5, one has more than enough values in the table to take care of business. A common notation related to the table is that of z_α. The value z_α is defined for all positive $\alpha < 1$ by

$$\alpha = P(Z > z_\alpha).$$

Standard Normal Probabilities of the Form $P(Z \leq x)$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998

Example 4.4. *We look at some particular cases for using the standard normal table.*

a) $P(Z < 1.92) = 0.9726$

b) $P(Z > 1.04) = 1 - P(Z \leq 1.04) \doteq 1 - 0.8508 = 0.1492$
We used the complement rule $P(A') = 1 - P(A)$.

c)

$$\begin{aligned} P(-1 < Z < 2) &= P(Z < 2) - P(Z \leq -1) \\ &\doteq 0.9772 - P(Z \geq 1) \\ &= 0.9772 - (1 - P(Z < 1)) \\ &\doteq 0.9772 - (1 - 0.8413) \\ &= 0.8185 \end{aligned}$$

d) If X is normal with mean 1 and standard deviation 2, then

$$\begin{aligned} P(1 < X < 2.04) &= P(\frac{1-1}{2} < Z < \frac{2.04-1}{2}) \\ &= P(0 < Z < 0.52) \\ &= P(Z < 0.52) - P(Z \leq 0) \\ &\doteq 0.6985 - 0.5 \\ &= 0.1985 \end{aligned}$$

e) $z_{.025} = 1.96$ since $P(Z \leq 1.96) = 0.975$ (so that $P(Z > 1.96) = 0.025$).

f) If IQ scores are normally distributed with $\mu = 100$ and $\sigma = 14.2$, and a "good" IQ score is considered to be in the top 20%, then we can calculate the lowest "good" score. Let it be x. Then, with X representing a normal random variable with mean 100 and standard deviation 14.2, we have that

$$0.20 = P(X \geq x) = P(Z \geq \frac{x-100}{14.2}), \text{ so}$$

$$\frac{x-100}{14.2} = z_{0.20} \Rightarrow x = 14.2z_{0.20} + 100 = 14.2(0.842) + 100 = 111.96 \doteq 112.$$

Exercises

(1) Use the table to compute (a) $P(Z < 1.32)$, (b) $P(Z > 0.64)$, (c) $P(Z \leq -2.06)$, (d) $P(0.32 < Z < 1.56)$, (e) $P(-1.54 < Z < 1.54)$, and (f) $P(Z = 0)$
Ans. (a) 0.9066 (b) 0.2611 (c) 0.0197 (d) 0.3151 (e) 0.8764 (f) 0

(2) Use the table to compute (a) $z_{.025}$, (b) $z_{.05}$, and (c) $z_{.01}$
Ans. (a) 1.96 *(b) about half way between 1.64 and 1.65 so approximately* 1.645 *(c) a bit closer to 2.33 than to 2.32 so about* 2.327

(3) If X is normal with $\mu = 1.4$ and $\sigma = 2.0$, compute (a) $P(X < 2.1)$, (b) $P(X > 1.0)$, (c) $P(0.6 < X < 2.2)$, and (d) the 90th percentile of X.

Ans. (a) 0.6368

(4) The mean for scores on a certain test is 300, and the standard deviation is 16. Assuming the test scores are normally distributed, compute the probability that a random score on this test is at least 310. What score would be in the 95th percentile?

Ans. 0.2660*,* 326.3

(5) If X is normal with $\mu = -5$ and $\sigma = 4.2$, compute (a) $P(X < -7)$, (b) $P(X > 0)$, and (c) $P(20 < X < 25)$.

Ans. (a) The table says the answer is between 0.6808 *and* 0.6844*. A decent graphing calculator will tell you the more precise answer is* 0.6830*. (c)* 0.0000

(6) If X is normal with mean μ and standard deviation σ, compute (a) $P(\mu - \sigma < X < \mu + \sigma)$, (b) $P(\mu - 2\sigma < X < \mu + 2\sigma)$, and (c) $P(\mu - 3\sigma < X < \mu + 3\sigma)$.

Ans. (a) 0.6827*, (b)* 0.9545*, (c)* 0.9973*. Note that statisticians talk about a an "Empirical Rule" which states that for normal distributions the probabilities that values are within one, two, and three, standard deviations of the mean are roughly* 68%*,* 95%*, and* 99%*, respectively.*

(7) Compute to two decimal places the 90th, 95th, and 99th percentiles of the standard normal.

Ans. 1.28*,* 1.64*,* 2.33

Chapter 5

Joint Probability Distributions

5.1. Discrete Case

Two discrete random variables X and Y can be paired as a single entity referred to as a *random vector* or *bivariate random variable*. This bivariate random variable has a *joint probability mass function* $p(x, y)$ for which we give a definition and properties here.

Joint Probability Mass Function. The joint probability mass function $p(x, y)$ for the pair of random variables X and Y is defined by

$$p(x, y) = P(X = x, Y = y).$$

Consequently, it's the case that $p(x, y) \geq 0$ for all x and y and that $\sum_x \sum_y p(x, y) = 1$. Moreover, for any event A in the sample space for $X \times Y$ we have that

$$P((X, Y) \in A) = \sum \sum_{(x,y) \in A} p(x, y).$$

The joint probability distribution function $F(x, y)$ for X and Y is given by

$$F(x, y) = P(X \leq x, Y \leq y),$$

so that the distribution function $F_X(x)$ for X is given by

$$F_X(x) = \lim_{y \to \infty} F(x, y),$$

and the distribution function for Y is

$$F_Y(y) = \lim_{x \to \infty} F(x, y).$$

The *marginal probability mass function of* X is given by

$$p_X(x) = \sum_y p(x, y),$$

where $p(x, y)$ is the joint probability mass function for X and Y. Similarly, we have the marginal probability mass function of Y given by

$$p_Y(y) = \sum_x p(x, y).$$

Example 5.1. *An experiment consists of randomly selecting two individuals from a group of 10 consisting of three accountants, five engineers, and two administrative assistants. Let* X *count the number of accountants selected and* Y *the number of engineers selected. Find the joint and marginal probability mass functions.*

We note that both X and Y can only take the values 0, 1, and 2. Thinking in terms of pertinent hypergeometric distributions, one can compute the probabilities in the table:

		y		
	$p(x,y)$	0	1	2
	0	$\frac{1}{45}$	$\frac{10}{45}$	$\frac{10}{45}$
x	1	$\frac{6}{45}$	$\frac{15}{45}$	0
	2	$\frac{3}{45}$	0	0

For example, selecting 0 accountants and 0 engineers would be the same as selecting two administrative assistants, so

$$p(0, 0) = \frac{\binom{2}{2}\binom{8}{0}}{\binom{10}{2}} = \frac{1}{45}.$$

The number of ways to select exactly one accountant and exactly one engineer, would be $\binom{3}{1}\binom{5}{1}\binom{2}{0} = 15$, so

$$p(1, 1) = \frac{15}{45}.$$

The marginal probability mass function for X is given by $f_X(0) = \frac{1}{45} + \frac{10}{45} + \frac{10}{45} = \frac{21}{45}$, $f_X(1) = \frac{6}{45} + \frac{15}{45} + 0 = \frac{21}{45}$, and $f_X(2) = \frac{3}{45} + 0 + 0 = \frac{3}{45}$.

The marginal probability mass function for Y is given by $f_Y(0) = \frac{1}{45} + \frac{6}{45} + \frac{3}{45} = \frac{10}{45}$, $f_Y(1) = \frac{10}{45} + \frac{15}{45} + 0 = \frac{25}{45}$, and $f_Y(2) = \frac{10}{45} + 0 + 0 = \frac{10}{45}$.

We continue the discussion of pairs of random variables with a definition for their independence. The discrete random variables X and Y are said to be *independent* if for all real numbers x and y,

$$F(x, y) = F_X(x)F_Y(y).$$

An equivalent definition is that

$$p(x, y) = p_X(x)p_Y(y).$$

For our discrete pair of random variables X and Y, we can define the *conditional probability mass function of X given $Y = y$* by

$$\begin{aligned} p_{X|Y}(x|y) &= P(X = x|Y = y) \\ &= \frac{P(X = x, Y = y)}{P(Y = y)} \\ &= \frac{p(x, y)}{p_Y(y)}. \end{aligned}$$

Of course, it is assumed $P(Y = y) > 0$ in this definition. In a similar fashion we have

$$p_{Y|X}(y|x) = \frac{p(x, y)}{p_X(x)}$$

when $P(X = x) > 0$. We calculate a conditional probability mass function for the immediately preceding example.

Example 5.2. *An experiment consists of randomly selecting two individuals from a group of 10 consisting of three accountants, five engineers, and two administrative assistants. Let X count the number of accountants selected and Y the number of engineers selected. Find the conditional mass function of X given $Y = 1$.*

We have

$$p_{X|Y}(0|1) = \frac{p(0,1)}{p_Y(1)} = \frac{10/45}{25/45} = \frac{10}{25} = \frac{2}{5},$$
$$p_{X|Y}(1|1) = \frac{p(1,1)}{p_Y(1)} = \frac{15/45}{25/45} = \frac{15}{25} = \frac{3}{5}, \text{ and}$$
$$p_{X|Y}(2|1) = \frac{p(2,1)}{p_Y(1)} = \frac{0}{25/45} = 0.$$

Note that if X and Y are independent discrete random variables then

$$\begin{aligned} p_{X|Y}(x|y) &= \frac{p(x, y)}{p_Y(y)} \\ &= \frac{p_X(x) \cdot p_Y(y)}{p_Y(y)} \\ &= p_X(x). \end{aligned}$$

It's even possible to have a conditional distribution for X given a function of the random variables X and Y equals some constant. We see in our next example that if X and Y are independent Poisson random variables, that the conditional distribution of X given the sum of X and Y is a constant natural number becomes a Binomial random variable.

Example 5.3. *Suppose X and Y are independent Poisson random variables with parameters λ_X and λ_Y, respectively. Calculate the probability mass function for X given $X + Y = n$*

We have

$$\begin{aligned}
P(X=k|X+Y=n) &= \frac{P(X=k, X+Y=n)}{P(X+Y=n)} \\
&= \frac{P(X=k, Y=n-k)}{P(X+Y=n)} \\
&= \frac{P(X=k)\cdot P(Y=n-k)}{P(X+Y=n)} \\
&= \frac{e^{-\lambda_X}\lambda_X^k}{k!}\cdot\frac{e^{-\lambda_Y}\lambda_Y^{n-k}}{(n-k)!} \\
&\qquad \div \quad \sum_{i=0}^{n}\frac{e^{-\lambda_X}\lambda_X^i}{i!}\cdot\frac{e^{-\lambda_Y}\lambda_Y^{n-i}}{(n-i)!} \\
&= e^{-(\lambda_X+\lambda_Y)}\frac{\lambda_X^k\lambda_Y^{n-k}}{k!(n-k)!} \\
&\qquad \div \quad e^{-(\lambda_X+\lambda_Y)}\sum_{i=0}^{n}\frac{\lambda_X^i\lambda_Y^{n-i}}{i!(n-i)!} \\
&= \frac{\lambda_X^k\lambda_Y^{n-k}}{k!(n-k)!} \div \frac{1}{n!}\sum_{i=0}^{n}\frac{n!}{i!(n-i)!}\lambda_X^i\lambda_Y^{n-i} \\
&= \frac{\lambda_X^k\lambda_Y^{n-k}}{k!(n-k)!} \div \frac{1}{n!}(\lambda_X+\lambda_Y)^n \\
&= \frac{n!}{k!(n-k)!}\Big(\frac{\lambda_X}{\lambda_X+\lambda_Y}\Big)^k\Big(\frac{\lambda_Y}{\lambda_X+\lambda_Y}\Big)^{n-k} \\
&= \frac{n!}{k!(n-k)!}\Big(\frac{\lambda_X}{\lambda_X+\lambda_Y}\Big)^k\Big(1-\frac{\lambda_X}{\lambda_X+\lambda_Y}\Big)^{n-k}.
\end{aligned}$$

So the conditional distribution of X given $X+Y=n$ is binomial with parameters n and $\frac{\lambda_X}{\lambda_X+\lambda_Y}$.

Thus far we have been discussing joint probability distributions for two discrete random variables. We can generalize the discussion to n discrete random variables. When dealing with the n random variables $X_1, X_2, ..., X_n$, for example, the function

$$p(x_1, x_2, ..., x_n) = P(X_1 = x_1, X-2 = x_2, ..., X_n = x_n)$$

serves as a probability mass function. We have $p(x_1, x_2, ..., x_n) \geq 0$. Also

$$\sum_{x_1}\cdots\sum_{x_n} p(x_1, x_2, ..., x_n) = 1$$

and

$$P((X_1, ..., X_n) \in A) = \sum \cdots \sum_{(x_1,...,x_n)\in A} p(x_1, x_2, ..., x_n).$$

We examine a particular discrete joint distribution called the *multinomial*. In this distribution one encounters an intricate counting procedure that we examine first. Suppose n is a positive integer and that the r nonnegative integers $n_1, ..., n_r$, are such that

$$n_1 + \cdots + n_r = n.$$

Then the *multinomial coefficient* $\binom{n}{n_1, ..., n_r}$ is defined by

$$\binom{n}{n_1, ..., n_r} = \frac{n!}{n_1!n_2! \cdots n_r!}.$$

We list some quantities that this coefficient counts:

- The number of ways you can split n distinct objects into r distinct groups of sizes n_1, n_2, . . . , and n_r, respectively.
- The number of n-letter words made up of r distinct letters used n_1, n_2, . . . , and n_r, times, respectively
- The coefficient on $x_1^{n_1} \cdots x_r^{n_r}$ in the expansion of $(x_1 + \cdots + x_r)^n$

For example, if you roll a die eight times, the number of ways you can obtain one 1, one 2, zero 3's, zero 4's, two 5's, and four 6's, is

$$\binom{8}{1, 1, 0, 0, 2, 4} = \frac{8!}{1!1!0!0!2!4!} = 840.$$

Another example is to count how many words (including nonsensical ones) you can make from two a's, one c, three t's, and one z. The answer is

$$\binom{7}{2, 1, 3, 1} = \frac{7!}{2!1!3!1!} = 420.$$

A couple of these 420 words are *aactttz* and *acatttz*.

Yet another example is that the coefficient on x^2y^2z in the expansion of $(x + y + z)^5$ is

$$\binom{5}{2, 2, 1} = \frac{5!}{2!2!1!} = 30.$$

By inspection the reader can see that the coefficient on x^5 in this expansion is 1. The multinomial coefficient gives us just this:

$$\binom{5}{5, 0, 0} = \frac{5!}{5!0!0!} = 1.$$

Now we consider the multinomial random vector.

Multinomial Distribution. For a sequence of n independent, identical experiments with each one of the experiments resulting in r outcomes with probabilities p_1, p_2, ... , and p_r, respectively, where $p_1 + \cdots + p_r = 1$, we let X_i count the number of the n experiments that result in the ith of the r outcomes. Then

$$P(X_1 = n_1, X_2 = n_2, ..., X_n = n_r) = \binom{n}{n_1, n_2, ..., n_r} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r},$$

where $n_1 + \cdots + n_r = n$.

Example 5.4. *Roll a die eight times. What's the probability (a) you obtain one* 1*, one* 2*, zero* 3*'s, zero* 4*'s, two* 5*'s, and four* 6*'s and (b) you obtain one* 1*, one* 2*, one* 3*, one* 4*, two* 5*'s, and two* 6*'s?*

(a)

$$\binom{8}{1,1,0,0,2,4}(1/6)^1(1/6)^1(1/6)^0(1/6)^0(1/6)^2(1/6)^4$$

or

$$\frac{8!}{1!1!0!0!2!4!}(1/6)^8 = \frac{840}{1,679,616} = 0.0005.$$

(b)

$$\binom{8}{1,1,1,1,2,2}(1/6)^1(1/6)^1(1/6)^1(1/6)^1(1/6)^2(1/6)^2$$

or

$$\frac{8!}{1!1!1!1!2!2!}(1/6)^8 = \frac{10,080}{1,679,616} = 0.0060.$$

Exercises

(1) Suppose X and Y are discrete random variables with probability mass function as given in the table.

		Y	
	$p(x,y)$	2	4
X	0	0.30	0.20
	6	0.40	0.10

Compute (a) $P(X + Y < 5)$, (b) $P(Y > X)$, and (c) $E(X)$
Ans.: (a) 0.50*, (b)* 0.50*, and (c)* 3

(2) Suppose X and Y are discrete random variables with probability mass function as given in the table.

		Y		
	$p(x,y)$	3	5	10
X	4	0.25	0.10	0.05
	6	0.20	0.02	0.38

Find rules for the two marginal probability mass functions and compute (a) $P(Y \geq 5)$, (b) $P(X+Y \leq 9)$, (c) $E(X)$, and (d) $E(Y)$.

Ans.:

$$p_X(x) = \begin{cases} 0.40 & \textit{if } x = 4 \\ 0.60 & \textit{if } x = 6 \\ 0 & \textit{otherwise} \end{cases}$$

$$p_Y(y) = \begin{cases} 0.45 & \textit{if } y = 3 \\ 0.12 & \textit{if } y = 5 \\ 0.43 & \textit{if } y = 10 \\ 0 & \textit{otherwise} \end{cases}$$

$P(Y \geq 5) = 0.55$, $P(X+Y \leq 9) = 0.55$, $E(X) = 5.2$, *and* $E(Y) = 6.25$

(3) Determine if the random variables in the preceding problem are independent.
Ans.: They are not.

(4) Roll a die four times. What's the probability (a) you obtain one 1, one 2, two 6's and (b) you obtain one 1, one 2, one 3, and one 4?
Ans.: (a) 0.0093 *(b)* 0.0185

(5) Deal a card from a well shuffled deck of 52. Replace it, shuffle and deal another. Repeat this process until you've dealt five cards. What's the probability you deal one spade, two diamonds, and two hearts? [Note that you're not dealing with a hypergeometric distribution since you're replacing dealt cards.]
Ans.: 0.0293

(6) Rogelio, Jesús, and Diego, are playing dominoes. If it's the case that Rogelio wins 50% of the time, Jesús 40% of the time, and Diego 10% of the time, what's the probability that in the next five games, Rogelio wins twice, Jesús twice, and Diego once?
Ans.: 0.12

(7) In the United States, 45% of the population has Type O blood, 40% Type A, 11% Type B, and four percent Type AB. If six US residents are randomly selected, what's the probability three would have Type O, two type A, and one Type B?
Ans.: 0.0962

5.2. Continuous Case

Two continuous random variables X and Y that are paired have a *joint probability density function* $f(x, y)$. We list properties.

Joint Probability Density Function. The joint probability density function $f(x, y)$ for the pair of random variables X and Y has the properties

$$f(x, y) \geq 0$$

for all x and y,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1,$$

and

$$P((X, Y) \in A) = \int \int_{A} f(x, y) dx dy.$$

The joint probability distribution function $F(x, y)$ for X and Y is given by

$$F(x, y) = P(X \leq x, Y \leq y).$$

As was the case with a pair of discrete random variables, here with continuous random variables, the distribution function $F_X(x)$ is given by $F_X(x) = \lim_{y\to\infty} F(x, y)$. Similarly, we have that the distribution function for Y is given by $F_Y(y) = \lim_{x\to\infty} F(x, y)$.

The *marginal probability density function of* X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy,$$

where $f(x, y)$ is the joint probability density function for X and Y. Similarly, the marginal probability density function of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

We provide an example.

Example 5.5. *Suppose X and Y are continuous random variables with joint probability density function*

$$f(x, y) = \begin{cases} \frac{1}{12}x + \frac{1}{24}y & \textit{if } 0 \leq x \leq 3 \text{ and } 0 \leq y \leq 2 \\ 0 & \textit{otherwise} \end{cases}$$

Compute then (a) $P(X \leq 2 \text{ and } Y \geq 1)$, (b) $P(X + Y \leq 1)$, (c) a rule for $f_X(x)$, (d) a rule for $f_Y(y)$, and (e) $E(X)$.

(a)

$$\begin{aligned}
P(X \leq 2, Y \geq 1)) &= \int_0^2 \int_1^2 (\frac{1}{12}x + \frac{1}{24}y)\ dydx \\
&= \int_0^2 (\frac{1}{12}xy + \frac{1}{48}y^2)|_1^2\ dx \\
&= \int_0^2 (\frac{1}{6}x + \frac{1}{12} - (\frac{1}{12}x + \frac{1}{48}))\ dx \\
&= \int_0^2 (\frac{1}{12}x + \frac{3}{48})dx \\
&= (\frac{1}{24}x^2 + \frac{3}{48}x)|_0^2 \\
&= \frac{7}{24}.
\end{aligned}$$

(b)

$$\begin{aligned}
P(X + Y \leq 1)) &= \int_0^1 \int_0^{1-x} (\frac{1}{12}x + \frac{1}{24}y)\ dydx \\
&= \int_0^1 (\frac{1}{12}xy + \frac{1}{48}y^2)|_0^{1-x}\ dx \\
&= \int_0^1 (\frac{1}{12}(x - x^2) + \frac{1}{48}(1 - 2x + x^2))\ dx \\
&= \int_0^1 (-\frac{3}{48}x^2 + \frac{1}{24}x + \frac{1}{48})dx \\
&= \frac{1}{48}.
\end{aligned}$$

(c)

$$\begin{aligned}
f_X(x) &= \int_0^2 (\frac{1}{12}x + \frac{1}{24}y)\ dy \\
&= (\frac{1}{12}xy + \frac{1}{48}y^2)|_0^2 \\
&= \frac{1}{6}x + \frac{1}{12},
\end{aligned}$$

so

$$f_X(x) = \begin{cases} \frac{1}{6}x + \frac{1}{12} & \text{if } 0 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

(d)

$$\begin{aligned}
f_Y(y) &= \int_0^3 (\frac{1}{12}x + \frac{1}{24}y)\ dx \\
&= (\frac{1}{24}x^2 + \frac{1}{24}xy)|_0^3 \\
&= \frac{3}{8} + \frac{1}{8}y,
\end{aligned}$$

so

$$f_Y(y) = \begin{cases} \frac{1}{8}y + \frac{3}{8} & \text{if } 0 \le y \le 2 \\ 0 & \text{otherwise} \end{cases}$$

(e)

$$\begin{aligned} E(X) &= \int_0^3 x(\frac{1}{6}x + \frac{1}{12})\ dx \\ &= \int_0^3 (\frac{1}{6}x^2 + \frac{1}{12}x)\ dx \\ &= (\frac{1}{18}x^3 + \frac{1}{24}x^2)|_0^3 \\ &= \frac{27}{18} + \frac{9}{24} \\ &= \frac{15}{8}. \end{aligned}$$

As was the case with pairs of discrete random variables, we can discuss conditional probability density functions associated with a joint probability density function for a pair of continuous random variables. The conditional probability density function of X given $Y = y$ is given by

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)},$$

where $f(x, y)$ is the joint probability density function for the continuous rnadom variables X and Y. Here, we're assuming y is a value for which $f_Y(y) > 0$.

Example 5.6. *For the joint probability density function in the previous example, take y to be a value in $[0, 2]$, and compute $f_{X|Y}(x|y)$.*

We have

$$f_{X|Y}(x|y) = \frac{\frac{1}{12}x + \frac{1}{24}y}{\frac{1}{8}y + \frac{3}{8}} = \frac{2x + y}{3y + 9},$$

if $0 \le x \le 3$. In the particular case that $y = \frac{1}{3}$, we have

$$f_{X|Y}(x|\frac{1}{3}) = \frac{2x + \frac{1}{3}}{3(\frac{1}{3}) + 9} = \frac{1}{5}x + \frac{1}{30},$$

if $0 \le x \le 3$.

As is the case when X and Y are discrete, The continuous random variables X and Y are said to be *independent* if for all real numbers x and y,

$$F(x, y) = F_X(x)F_Y(y).$$

An equivalent definition is that

$$f(x, y) = f_X(x)f_Y(y),$$

where the f's are probability density functions.

We can generalize the discussion to n continuous random variables instead of just two. When dealing with the n continuous random variables $X_1, X_2, ..., X_n$, the function $f(x_1, x_2, ..., x_n)$ might serve as the probability density function, and we would have that $f(x_1, x_2, ..., x_n) \geq 0$. Also,

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, ..., x_n) = 1$$

and

$$P((X_1, ..., X_n) \in A) = \int \cdots \int_{(x_1,...,x_n) \in A} f(x_1, x_2, ..., x_n).$$

Exercises

(1) Suppose X and Y are continuous random variables with joint probability density function

$$f(x,y) = \begin{cases} \frac{1}{6} & \text{if } 0 \leq x \leq 3 \text{ and } 0 \leq y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Compute the probabilities (a) $P(X \leq 2)$, (b) $P(X + Y \leq 1)$ and (c) $P(Y \geq \frac{1}{9}X^2)$.

Ans. (a) $\frac{2}{3}$, *(b)* $\frac{1}{12}$, *(c)* $\frac{5}{6}$

(2) Suppose X and Y are continuous random variables with joint probability density function

$$f(x,y) = \begin{cases} c(x^3 + y) & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Compute the value of c, find rules for the two marginal probability density functions, and compute the probabilities $P(Y \leq 1)$ and $P(X < Y)$.

Ans.: $c = 2/5$,

$$f_X(x) = \begin{cases} \frac{4}{5}(x^3 + 1) & \textit{if } 0 \leq x \leq 1 \\ 0 & \textit{otherwise} \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{2}{5}y + \frac{1}{10} & \textit{if } 0 \leq y \leq 2 \\ 0 & \textit{otherwise} \end{cases}$$

$P(Y \leq 1) = 0.3$, *and* $P(X < Y) = \frac{64}{75}$.

(3) Determine if the random variables in the preceding problem are independent.

Ans.: They are not.

(4) Suppose the continuous random variables X and Y have the joint probability density function

$$f(x,y) = \begin{cases} ab\, e^{-ax-by} & \text{if } x \geq 0 \text{ and } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Here a and b are positive constants. This can model the lifetimes of two components, the first - X - having mean life $\frac{1}{a}$ and the second - Y - having mean life $\frac{1}{b}$. Compute (a) the probability that both last past time t, i.e. $P(X > t, Y > t)$ and (b) that the first lasts longer than the second, i.e. $P(X > Y)$.

Ans.: (a) $e^{-(a+b)t}$ *(b)* $\frac{b}{a+b}$

(5) Determine if the random variables in the preceding problem are independent.
Ans.: They are. Note

$$F(s,t) = \int_0^s \int_0^t abe^{-ax-by}dydx = \int_0^s ae^{-ax}dx \int_0^t be^{-by}dy = F_X(s)F_Y(t)$$

(6) Suppose the continuous random variables X and Y have the joint probability density function

$$f(x,y) = \begin{cases} cye^{-2x-y} & \text{if } x \geq 0 \text{ and } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Compute (a) $P(X < 1, Y < 1)$, (b) $P(X > 2)$, (c) $P(X < Y)$, and (d) $P(X + Y < 1)$.
Ans.: $c = 2$, $P(X < 1, Y < 1) = 1 - \frac{2}{e} - \frac{1}{e^2} + \frac{2}{e^3} = 0.2285$, $P(X > 2) = e^{-4}$, $P(X < Y) = \frac{8}{9}$, *and* $P(X + Y < 1) = 1 - 2e^{-1} - e^{-2} = 0.1289$.

(7) Suppose X and Y are independent exponential random variables with means $\mu_X = 4.0$ and $\mu_Y = 5.0$. Compute the probability $P(X \geq 5$ and $Y \geq 5)$.
Ans. $P(X \geq 5, Y \geq 5) = 0.1059$

(8) The random vector (X, Y) has probability density function

$$f(x,y) = \begin{cases} c(x^3 + 2xy) & \text{if } 0 \leq y \leq x \text{ and } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Compute (a) the value of c, (b) $P(X > 2Y)$, (c) $E(X)$, and (d) $E(Y)$.
Ans. (a) $\frac{5}{52}$, *(b)* $\frac{21}{52}$, *(c)* $\frac{64}{39}$, *(d)* $\frac{12}{13}$

5.3. Functions of Joint Random Variables.

Suppose X and Y are discrete random variables on the same sample space and that the real valued, two variable function $g(x, y)$ is defined on the range of (X, Y). We consider $g(X, Y)$ as a random variable say W:

$$W = g(X, Y).$$

If $p_{X,Y}(x, y)$ is the joint probability mass function for X and Y, the the probability mass function for W is given by

$$\begin{aligned} p_W(w) &= P(g(X,Y) = w) \\ &= P((X,Y) \in g^{-1}(w)) \\ &= \sum \sum_{(x,y) \in g^{-1}(w)} p_{X,Y}(x,y) \\ &= \sum \sum_{g(x,y)=w} p_{X,Y}(x,y). \end{aligned}$$

If X and Y happen to be independent, then $p(x, y) = p_X(x) \cdot p_Y(y)$ so that

$$p_W(w) = \sum_{g(x,y)=w} p_X(x) \cdot p_Y(y).$$

Example 5.7. *Suppose that X and Y are independent binomial random variables with the same probability of success p for each trial - say X is binomial with parameters m and p and Y parameters n and p. Then for $g(x, y) = x + y$, find a rule for $W = g(X, Y)$.*

We have $X + Y$ taking the values $k = 0, 1, 2, ..., m + n$ according to the rule

$$\begin{aligned} p_{X+Y}(k) &= \sum\sum_{x+y=k} p_X(x) \cdot p_Y(y) \\ &= \sum_{r=0}^{k} p_X(r) \cdot p_Y(k-r), \end{aligned}$$

where $p_X(r) = 0$ if $r > m$ and $p_Y(k - r) = 0$ if $k - r > n$. Thus,

$$\begin{aligned} p_{X+Y}(k) &= \sum_{r=0}^{k} \tbinom{m}{r} p^r (1-p)^{m-r} \cdot \tbinom{n}{k-r} p^{k-r} (1-p)^{n-(k-r)} \\ &= \sum_{r=0}^{k} \tbinom{m}{r}\tbinom{n}{k-r} p^k (1-p)^{m+n-k}, \end{aligned}$$

again where where $\binom{m}{r} = 0$ if $r > m$ and $\binom{n}{k-r} = 0$ if $k - r > n$. Now consider a situation where you have m blue balls and n red balls in an urn. The number of ways you could select k of the balls from the urn is $\binom{m+n}{k}$. Note that you could categorize these ways according to the number of blue and red balls, making it apparent that these number of ways is the number of ways you could select 0 blue balls and k red balls plus the number of ways you could select 1 blue ball and $k-1$ red balls plus the number of ways you could select 2 blue balls and $k - 2$ red balls and so on up to the number of ways you could select k blue balls and 0 red balls. I.e.

$$\tbinom{m+n}{k} = \sum_{r=0}^{k} \tbinom{m}{r}\tbinom{n}{k-r}.$$

Consequently,

$$p_{X+Y}(k) = \tbinom{m+n}{k} p^k (1-p)^{m+n-k}$$

I.e. $X + Y$ is binomial with parameters $m + n$ and p.

Suppose now that X and Y are continuous random variables with joint probability density function $f_{X,Y}(x, y)$ and that $g(x, y)$ is a real valued two variable function that yields the continuous random variable $W = g(X, Y)$. To obtain the probability density function of W, we first calculate the cumulative distribution function of W and then take its derivative to get the probability density function. For example, if we let $g(x, y) = x + y$ so that $W = X + Y$, we have

$$\begin{aligned} F_{X+Y}(w) &= \int\int_{x+y\leq w} f_{X,Y}(x, y)dydx \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{w-x} f_{X,Y}(x, y)dydx. \end{aligned}$$

For the integral with respect to y, we make the substitution $u = y + x$, yielding $dy = du$ and the limits of integration $u = -\infty$ when $y = -\infty$ and $u = w$ when $y = w - x$. Then we have

$$\begin{aligned} F_{X+Y}(w) &= \int_{-\infty}^{\infty}\int_{-\infty}^{w} f_{X,Y}(x, u-x)du dx \\ &= \int_{-\infty}^{w}\int_{-\infty}^{\infty} f_{X,Y}(x, u-x)dx du. \end{aligned}$$

From the Fundamental Theorem of the Calculus,we therefore have

$$\begin{aligned} f_{X+Y}(w) &= \frac{d}{dw}\int_{-\infty}^{w}\int_{-\infty}^{\infty} f_{X,Y}(x, u-x)dx du \\ &= \int_{-\infty}^{\infty} f_{X,Y}(x, w-x)dx. \end{aligned}$$

Example 5.8. *Suppose X and Y are independent standard normal random variables. Calculate the probability density function of $X + Y$.*

$$\begin{aligned} F_{X+Y}(w) &= \int_{-\infty}^{\infty} f_{X,Y}(x, w-x)dx \\ &= \int_{-\infty}^{\infty} f_X(x)\cdot f_Y(w-x)dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{(w-x)^2}{2}}dx \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-(x^2-wx+\frac{w^2}{4})-\frac{w^2}{4}}dx \\ &= \frac{1}{2\pi}e^{-\frac{w^2}{4}}\int_{-\infty}^{\infty} e^{-(x-\frac{w}{2})^2}dx \\ &= (\sqrt{2\pi}\cdot\frac{1}{\sqrt{2}})\cdot\frac{1}{2\pi}e^{-\frac{w^2}{4}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\cdot\frac{1}{\sqrt{2}}}\cdot e^{-\frac{(x-\frac{w}{2})^2}{2(\frac{1}{\sqrt{2}})^2}}dx \\ &= \frac{1}{\sqrt{2\pi}\cdot\sqrt{2}}e^{-\frac{w^2}{2(\sqrt{2})^2}}\cdot 1. \end{aligned}$$

The integral does in fact have a value of 1 since the integrand is the pdf for a normal distribution. The reader will recognize the pdf for $X + Y$ as that of a normal random variable with mean 0 and standard deviation $\sqrt{2}$.

Now we consider the pair of real valued, two variable functions $g(x, y)$ and $h(x, y)$ defined on the range of (X, Y). Assume that their first partial derivatives are continuous on this range. Assume also that the transformation defined by $u = g(x, y)$ and $v = h(x, y)$ is one-to-one. Then, defining the two random variables U and V by $U = g(X, Y)$ and $V = h(X, Y)$, according to the change of variables formula from multidimensional calculus, the joint probability density function for U and V is given by

$$f_{U,V}(u,v) = \frac{1}{|J(x,y)|} f_{X,Y}(x,y),$$

where (x,y) is the point in the range of (X,Y) for which $g(x,y) = u$ and $h(x,y) = v$, and $J(x,y)$ is the Jacobian

$$J(x,y) = \begin{vmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{vmatrix}$$

There are no exercises for this section. We put the theory of functions of joint probability distributions to work in the next section to compute expected values.

5.4. Expected Value, Variance, and Covariance

For the single random variable X - whether it be discrete or continuous - we defined the expected value $\mu = E(X)$ in a previous chapter. We saw also that for the constants a and b, $E(aX + b) = aE(X) + b$. We also defined the variance $\sigma^2 = V(X) = E(X - \mu)^2$ and noted that $V(aX + b) = a^2 V(X)$.

Here, we consider a pair of random variables X and Y (discrete to start with) and a real valued, two variable function $g(x,y)$. As we did in the single variable case, we can establish that

$$E(g(X,Y)) = \sum_x \sum_y g(x,y) p_{X,Y}(x,y).$$

If X and Y are continuous random variables, then

$$E(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dy\,dx.$$

We note that if the two random variables X and Y are discrete and a and b are constants,

$$\begin{aligned} E(aX + bY) &= \sum_x \sum_y (ax + by) p_{X,Y}(x,y) \\ &= a \sum_x \sum_y x \cdot p_{X,Y}(x,y) + b \sum_x \sum_y y \cdot p_{X,Y}(x,y) \\ &= aE(X) + bE(Y). \end{aligned}$$

A similar derivation in the continuous case yields the same formula.

To compute the variance of $X + Y$ in both the discrete and continuous cases, we note

$$\begin{aligned} V(X + Y) &= E(X + Y - (\mu_X + \mu_Y))^2 \\ &= E((X - \mu_X) + (Y - \mu_Y))^2 \\ &= E((X - \mu_X)^2 + 2(X - \mu_X)(Y - \mu_Y) + (Y - \mu_Y)^2) \\ &= E(X - \mu_X)^2 + 2E(X - \mu_X)(Y - \mu_Y) + E(Y - \mu_Y)^2 \\ &= V(X) + 2E(X - \mu_X)(Y - \mu_Y) + V(Y). \end{aligned}$$

The middle term here sans the factor of 2 is referred to as the covariance of X and Y, labeled $Cov(X,Y)$. So we have

$$V(X+Y) = V(X) + V(Y) + Cov(X,Y).$$

When one considers how the covariance is defined -

$$\sum_x \sum_y (x-\mu_X)(y-\mu_Y)p_{X,Y}(x,y)$$

in the discrete case and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f_{X,Y}(x,y)dydx$$

in the continuous - it becomes apparent that it provides a measure of how X and Y are related. If large values of X go with large values of Y and small values of X with small values of Y with a preponderance of probability, then the product $(x-\mu_X)(y-\mu_Y)$ is most likely positive because both factors are positive (large X with large Y) or both are negative (small X with small Y). This leads to the covariance being positive with greater and greater magnitude as the probability concentrates more and more along the pattern large X with large Y and small X with small Y. If, on the other hand, small values of X go with large values of Y and large values of X with small values of Y for the majority of the probability, then the covariance is negative as the two factors of $(x-\mu_X)(y-\mu_Y)$ are opposite in sign over most of the probability. This negative value has greater magnitude as the probability that the pattern of small X with large Y and large X with small Y is greater.

We can establish a shortcut formula for the covariance in much the same way we did for the variance of a single random variable. We have

$$\begin{aligned} Cov(X,Y) &= E(X-\mu_X)(Y-\mu_Y) \\ &= E(XY-\mu_Y X-\mu_X Y+\mu_X\mu_Y) \\ &= E(XY)-\mu_Y E(X)-\mu_X E(Y)+\mu_X\mu_Y) \\ &= E(XY)-\mu_X\mu_Y. \end{aligned}$$

So the shortcut formula is as stated.

Covariance. The covariance of X and Y is defined by

$$Cov(X,Y) = E(X-E(X)\ )(Y-E(Y)\ ).$$

It can be computed via the shortcut formula

$$Cov(X,Y) = E(XY) - E(X)E(Y).$$

Note that if X and Y are independent, then (in the discrete case)

$$\begin{aligned} E(XY) &= \sum_x \sum_y xy \cdot p_{X,Y}(x,y) \\ &= \sum_x \sum_y xy \cdot p_X(x) \cdot p_Y(y) \\ &= \sum_x x \cdot p_X(x) \sum_y y \cdot p_Y(y) \\ &= E(X)E(Y). \end{aligned}$$

The same formula can be established in a similar way in the continuous case. Hence, if X and Y are independent,

$$Cov(X,Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0.$$

Mathematicians - especially Karl Pearson - eventually discovered that this covariance scaled so that it only takes values from -1 to 1 can provide a great deal of standardized information about the degree to which X and Y are linearly related. They developed the so called *correlation coefficient* ρ defined by

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

The reader will note that when X and Y are independent - i.e. they have no relationship - $\rho = 0$ because the covariance is zero. In the case that Y is a linear function of X, say $Y = aX + b$ with a and b being constants and $a \neq 0$, we have

$$\begin{aligned} \rho &= \frac{E(X(aX+b)) - E(X)E(aX+b)}{\sigma_X \sigma_{aX+b}} \\ &= \frac{E(aX^2 + bX) - E(X)(aE(X)+b)}{\sigma_X \cdot |a| \sigma_X} \\ &= \frac{aE(X^2) + bE(X) - a(E(X))^2 - bE(X)}{|a|\sigma_X^2} \\ &= \frac{a(E(X^2) - (E(X))^2)}{|a|\sigma_X^2} \\ &= \frac{a\sigma_X^2}{|a|\sigma_X^2} \\ &= \pm 1. \end{aligned}$$

So we see that $\rho = 1$ if $a > 0$ and $\rho = -1$ if $a < 0$.

Finally we note that ρ only takes values in the interval $[-1,1]$. This can be verified without too much difficulty. Note that

$$\begin{aligned}
0 &\leq V(\frac{X-\mu_X}{\sigma_X}+\frac{Y-\mu_Y}{\sigma_Y}) \\
&= (\frac{1}{\sigma_X})^2V(X)+(\frac{1}{\sigma_Y})^2V(Y)+2\cdot Cov(\frac{X-\mu_X}{\sigma_X},\frac{Y-\mu_Y}{\sigma_Y}) \\
&= 1+1+2E(\frac{X-\mu_X}{\sigma_X}-0)(\frac{Y-\mu_Y}{\sigma_Y}-0) \\
&= 2+2(\frac{1}{\sigma_X\sigma_Y})E(X-\mu_X)(Y-\mu_Y) \\
&= 2(1+\rho).
\end{aligned}$$

So $1+\rho\geq 0$ which makes $\rho\geq -1$. Similarly,

$$\begin{aligned}
0 &\leq V(\frac{X-\mu_X}{\sigma_X}-\frac{Y-\mu_Y}{\sigma_Y}) \\
&= 1+1-2\cdot Cov(\frac{X-\mu_X}{\sigma_X},\frac{Y-\mu_Y}{\sigma_Y}) \\
&= 2-2\rho \\
&= 2(1-\rho).
\end{aligned}$$

We therefore have that $1-\rho\geq 0$ so that $\rho\leq 1$. We summarize.

Correlation Coefficient. The correlation coefficient ρ given by

$$\rho=\frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y}$$

falls in the interval $[-1,1]$. If $Y=aX+b$, then $\rho=1$ for positive a and $\rho=-1$ for negative a. If X and Y are independent random variables, then $\rho=0$. Generally speaking, statisticians say X and Y have a linear relationship if ρ is close to 1 or -1. They're said to have no linear relationship if ρ is close to 0.

Example 5.9. *In a previous section we entertained an example in which X and Y are continuous random variables with joint probability density function*

$$f(x,y)=\begin{cases}\frac{1}{12}x+\frac{1}{24}y & \textit{if } 0\leq x\leq 3 \text{ and } 0\leq y\leq 2 \\ 0 & \textit{otherwise}\end{cases}$$

Compute ρ for this pair of random values X and Y.

In that previous example, we determined that

$$f_X(x)=\begin{cases}\frac{1}{6}x+\frac{1}{12} & \text{if } 0\leq x\leq 3 \\ 0 & \text{otherwise}\end{cases}$$

and

$$f_Y(y)=\begin{cases}\frac{1}{8}y+\frac{3}{8} & \text{if } 0\leq y\leq 2 \\ 0 & \text{otherwise}\end{cases}$$

In addition, we computed $E(X) = \frac{15}{8}$. We now calculate the remainder of the components needed to determine the value of ρ. First we note

$$\begin{aligned}
E(X^2) &= \int_0^3 x^2(\frac{1}{6}x + \frac{1}{12})\ dx \\
&= \int_0^3 (\frac{1}{6}x^3 + \frac{1}{12}x^2)\ dx \\
&= (\frac{1}{24}x^4 + \frac{1}{36}x^3)|_0^3 \\
&= \frac{81}{24} + \frac{27}{36} \\
&= \frac{27}{8} + \frac{3}{4} \\
&= \frac{33}{8}.
\end{aligned}$$

Next

$$\begin{aligned}
E(Y) &= \int_0^2 y(\frac{1}{8}y + \frac{3}{8})\ dy \\
&= \int_0^2 (\frac{1}{8}y^2 + \frac{3}{8}y)\ dy \\
&= (\frac{1}{24}y^3 + \frac{3}{16}y^2)|_0^2 \\
&= \frac{1}{3} + \frac{3}{4} \\
&= \frac{4}{12} + \frac{9}{12} \\
&= \frac{13}{12},
\end{aligned}$$

and

$$\begin{aligned}
E(Y^2) &= \int_0^2 y^2(\frac{1}{8}y + \frac{3}{8})\ dy \\
&= \int_0^2 (\frac{1}{8}y^3 + \frac{3}{8}y^2)\ dy \\
&= (\frac{1}{32}y^4 + \frac{1}{8}y^3)|_0^2 \\
&= \frac{1}{2} + 1 \\
&= \frac{3}{2}.
\end{aligned}$$

Finally

$$\begin{aligned}
E(XY) &= \int_0^3 \int_0^2 xy(\frac{1}{12}x + \frac{1}{24}y)\ dydx \\
&= \int_0^3 \int_0^2 (\frac{1}{12}x^2y + \frac{1}{24}xy^2)\ dydx \\
&= \int_0^3 (\frac{1}{24}x^2y^2 + \frac{1}{72}xy^3)|_0^2\ dx \\
&= \int_0^3 (\frac{1}{6}x^2 + \frac{1}{9}x)\ dx \\
&= (\frac{1}{18}x^3 + \frac{1}{18}x^2)|_0^3\ dx \\
&= \frac{3}{2} + \frac{1}{2} \\
&= 2.
\end{aligned}$$

This gives us

$$Cov(X,Y) = 2 - (\frac{15}{8})(\frac{13}{12}) = -\frac{1}{32},$$

$$\sigma_X^2 = \frac{33}{8} - (\frac{15}{8})^2 = \frac{39}{64},$$

and

$$\sigma_Y^2 = \frac{3}{2} - (\frac{13}{12})^2 = \frac{47}{144}.$$

Hence,

$$\rho = \frac{-\frac{1}{32}}{\sqrt{\frac{39}{64}} \cdot \sqrt{\frac{47}{144}}} = \frac{-8 \cdot 12}{32\sqrt{39 \cdot 47}} = \frac{-3}{\sqrt{1833}} = -0.07007.$$

To close the section, note that if we take the random variables $X_1, ..., X_n$, on the sample space, they have a joint pmf (if they're all discrete) or a joint pdf (if they're all continuous). In either case, we can expand the formulas for the expected value and variance of linear combinations of the X_i's and obtain

$$E(a_1X_1 + \cdots + a_nX_n) = a_1E(X_1) + \cdots a_nE(X_n)$$

and

$$V(X_1 + \cdots + X_n) = \sum_{k=1}^{n} V(X_k) + 2\sum\sum_{k<j} Cov(X_k, X_j).$$

If it's the case that the X_is are independent, then the second formula becomes

$$V(X_1 + \cdots + X_n) = \sum_{k=1}^{n} V(X_k).$$

Exercises

(1) Compute the correlation coefficient ρ for the pair of discrete random variables X and Y with probability mass function as given in the table.

	$p(x,y)$	Y: 2	4
X	0	0.30	0.20
	6	0.40	0.10

Ans.: -0.2182

(2) Compute the correlation coefficient ρ for the pair of discrete random variables X and Y with probability mass function as given in the table.

	$p(x,y)$	y: 0	1	2
	0	$\frac{1}{45}$	$\frac{10}{45}$	$\frac{10}{45}$
x	1	$\frac{6}{45}$	$\frac{15}{45}$	0
	2	$\frac{3}{45}$	0	0

Ans.: $E(X) = \frac{27}{45}$, $E(X^2) = \frac{33}{45}$, $E(Y) = 1$, $E(Y^2) = \frac{65}{45}$, *and* $E(XY) = \frac{15}{45}$, *so* $\rho = \frac{\frac{15}{45} - \frac{27}{45}}{\sqrt{(0.37333)(0.5555555)}} = -0.5855$

(3) Compute the correlation coefficient ρ for the pair of discrete random variables X and Y with probability mass function as given in the table.

	$p(x,y)$	Y: 3	5	10
X	4	0.25	0.10	0.05
	6	0.20	0.02	0.38

Ans.: $E(X) = 5.2$, $E(X^2) = 28$, $E(Y) = 6.25$, $E(Y^2) = 50.05$, *and* $E(XY) = 34$, *so* $\rho = \frac{34-(5.2)(28)}{\sqrt{(28-5.2^2)(50.05-6.25^2)}} = \frac{1.5}{\sqrt{(.96)(11.4375)}} = 0.4527$

(4) Suppose the continuous random variables X and Y have the joint probability density function

$$f(x,y) = \begin{cases} 2ye^{-2x-y} & \text{if } x \geq 0 \text{ and } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Compute ρ.

Ans.: $\rho = 0$. *Note that*

$$f_X(x) = 2e^{-2x}\int_0^\infty ye^{-y}dy = 2e^{-2x},$$

$$f_Y(y) = ye^{-y}\int_0^\infty 2e^{-2x}dx = ye^{-y},$$

and

$$\int_0^\infty \int_0^\infty xy \cdot 2ye^{-2x-y}dydx = \int_0^\infty x \cdot 2e^{-2x}dx \cdot \int_0^\infty y \cdot ye^{-y}dy,$$

so

$$E(XY) = E(X)E(Y).$$

Chapter 6

Sampling and Limit Theorems

6.1. Sample Mean and Variance

Statisticians draw conclusions about the overall nature of populations by analyzing samples. Suppose the random variable X represents a random selection of a value from a population that has mean μ and standard deviation σ. We can estimate $E(X)$ and $SD(X)$ by taking a random sample of values $x_1, x_2, x_3, ..., x_n$ from the population. The *sample statistics*

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

are used to estimate $E(X)$ and $SD(X)$, respectively. These two statistics are called the sample mean and the sample standard deviation.

Before we take a random sample of n values from the distribution, we are uncertain as to what the ith of the n values will be. We assign the random variable X_i to the ith selection in the random sample and note that $E(X_i) = \mu$ and $SD(X_i) = \sigma$. We then define the random variables $\bar{X}$ and S^2 as the sample mean and variance, respectively.

Sample Mean and Variance. The sample mean $\bar{X}$ and sample variance S^2 are given by by

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

We note that

$$\begin{aligned}
E(\bar{X}) &= E(\frac{1}{n}\sum_{i=1}^{n} X_i) \\
&= \frac{1}{n}E(\sum_{i=1}^{n} X_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} E(X_i) \\
&= \frac{1}{n}\cdot n\mu \\
&= \mu.
\end{aligned}$$

Another computation yields $E(S^2) = \sigma^2$. Note that

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \bar{X})^2 &= \sum_{i=1}^{n}(X_i^2 - 2\bar{X}X_i + (\bar{X})^2) \\
&= \sum_{i=1}^{n}(X_i^2) - 2\bar{X}\sum_{i=1}^{n} X_i + \sum_{i=1}^{n}(\bar{X})^2 \\
&= \sum_{i=1}^{n}(X_i^2) - 2\bar{X}\cdot n\bar{X} + n(\bar{X})^2) \\
&= \sum_{i=1}^{n}(X_i^2) - n(\bar{X})^2 \\
&= \sum_{i=1}^{n}(X_i^2) - n(\frac{1}{n}\sum_{i=1}^{n} X_i)^2 \\
&= \sum_{i=1}^{n}(X_i^2) - \frac{1}{n}(\sum_{i=1}^{n} X_i)^2.
\end{aligned}$$

Note also that for any random variable X, $V(X) = E(X^2) - (E(X))^2$ so that

$$E(X^2) = V(X) + (E(X))^2.$$

Consequently,

$$\begin{aligned}
E(S^2) &= E\Big(\frac{1}{n-1}\Big(\sum_{i=1}^{n}(X_i^2) - \frac{1}{n}(\sum_{i=1}^{n}X_i)^2\Big)\Big) \\
&= \frac{1}{n-1}\Big(\sum_{i=1}^{n}E(X_i^2) - \frac{1}{n}E(\sum_{i=1}^{n}X_i)^2)\Big) \\
&= \frac{1}{n-1}\Big(\sum_{i=1}^{n}(\sigma^2+\mu^2) - \frac{1}{n}[V(\sum_{i=1}^{n}X_i) + (E(\sum_{i=1}^{n}X_i))^2]\Big) \\
&= \frac{1}{n-1}\Big(n(\sigma^2+\mu^2) - \frac{1}{n}[n\sigma^2+(n\mu)^2]\Big) \\
&= \frac{1}{n-1}\Big(n\sigma^2+n\mu^2-\sigma^2-n\mu^2\Big) \\
&= \sigma^2.
\end{aligned}$$

The random variables $\bar{X}$ and S^2 are said to be unbiased estimators of μ and σ^2 since $E(\bar{X}) = \mu$ and $E(S^2) = \sigma^2$.

Now continuing to note that the X_i's we take as random readings from the population with mean μ and standard deviation σ are independent, we can make use of the result that says the variance of a sum of independent random variables is the sum of their variances to obtain

$$\begin{aligned}
V(\bar{X}) &= V(\frac{1}{n}\sum_{i=1}^{n}X_i) \\
&= \Big(\frac{1}{n}\Big)^2 V(\sum_{i=1}^{n}X_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}V(X_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 \\
&= \frac{1}{n^2}\cdot n\sigma^2 \\
&= \frac{\sigma^2}{n}.
\end{aligned}$$

Consequently, $SD(\bar{X}) = \dfrac{\sigma}{\sqrt{n}}$. As one might expect, the standard deviation of the sample mean $\bar{X}$ decreases as the sample size gets larger.

Exercises

(1) A random sample of size 100 is taken from a population with mean $\mu = 90.0$ and standard deviation $\sigma = 14.2$. Compute the mean and standard deviation of the sample mean.

Ans. $E(\bar{X}) = 90$ *and* $SD(\bar{X}) = 1.42$

(2) A random sample of size 30 is taken from a population with mean $\mu = 2.6$ and standard deviation $\sigma = 0.125$. Compute the mean and standard deviation of the sample mean.
Ans. $E(\bar{X}) = 2.6$ *and* $SD(\bar{X}) = 0.0228$

(3) A population has standard deviation $\sigma = 8.5$. What's the minimum size of a sample from this population that will produce a sample mean with a standard deviation (a) of less than 1 and (b) of less than 0.5?
Ans.: (a) 73 *(b)* 289

6.2. Law of Large Numbers

The Weak Law of Large Numbers is a fundamental result in probability . We derive it by making use of Chebyshev's Inequality - a formula of importance in and of itself.

Chebyshev's Inequality.
Suppose k is a positive constant and X is a random variable with finite mean μ and standard deviation σ. Then

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}.$$

Proof: (For the continuous case; the discrete case is similar) Suppose the continuous random variable X has pdf $f(x)$, mean μ, and standard deviation σ. Then we have that

$$\begin{aligned}
\sigma^2 &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \\
&\geq \int_{-\infty}^{\mu-k} (x-\mu)^2 f(x)dx + \int_{\mu+k}^{\infty} (x-\mu)^2 f(x)dx \\
&\geq \int_{-\infty}^{\mu-k} k^2 f(x)dx + \int_{\mu+k}^{\infty} k^2 f(x)dx \\
&= k^2\Big(\int_{-\infty}^{\mu-k} f(x)dx + \int_{\mu+k}^{\infty} f(x)dx\Big) \\
&= k^2 P(X \leq \mu - k \text{ or } X \geq \mu + k) \\
&= k^2 P\Big(|X-\mu| \geq k\Big). \quad \square
\end{aligned}$$

Example 6.1. *Scores on a certain standardized exam are 71.1 on average with a standard deviation of 12.1. Find a probability bound for a random score for this exam to be within 15 points of the mean.*

$$\begin{aligned} P(|X-71.1|<15) &= 1-P(|X-71.1|\geq 15) \\ &\geq 1-\frac{12.1^2}{15^2} \\ &= 0.349. \end{aligned}$$

The probability the score will be from 56.1 to 86.1 is at least 34%.

We now formulate the Law of Large Numbers.

> **The Weak Law of Large Numbers**.
> Suppose $X_1, X_2, ...$ are independent and identically distributed random variables with finite mean μ. Then for any $\epsilon > 0$,
> $$\lim_{n\to\infty} P\Big(|\frac{X_1+\cdots X_n}{n} - \mu| \geq \epsilon\Big) = 0.$$

Proof: (With the assumption that the variance σ^2 of the X_i's is finite.) We note that

$$E\Big(\frac{X_1+\cdots X_n}{n}\Big) = \mu$$

and

$$V\Big(\frac{X_1+\cdots X_n}{n}\Big) = \frac{\sigma^2}{n}.$$

Consequently, by Chebyshev's Inequality, we have

$$P\Big(|\frac{X_1+\cdots X_n}{n} - \mu| \geq \epsilon\Big) \leq \frac{\sigma^2}{n\epsilon^2}.$$

the result follows by the Squeeze Theorem.

Exercises

(1) For a random variable X with mean 10 and standard deviation 2, use Chebyshev's Inequality to find an upper bound on the probability that X is either less than or equal to 7 or greater than or equal to 13.
Ans. $4/9 = 0.4444$

(2) If the random variable X is normal with mean 10 and standard deviation 2, compute the probability that X is either less than or equal to 7 or greater than or equal to 13.
Ans. 0.1336

(3) If the random variable X is Binomial with mean 12 and standard deviation 3, compute the probability that X is either less than or equal to 8 or greater than or equal to 16. What is the upper bound that Chebyshev's Inequality gives for computing this probability?
Ans. 0.2422, 0.5625

(4) Scores on a certain standardized exam are 70 on average with a standard deviation of 10. Find an upper probability bound for a random score for this exam to be at least two standard deviation different from the mean.
Ans. 0.25

(5) Scores on a certain standardized exam are 65 on average with a standard deviation of 5. Find a lower probability bound for a random score for this exam to be between 57 and 73.
Ans. 0.6094

6.3. Moment Generating Functions

Moment generating functions are an important analytical tool when working with random variables. For one thing, they can be used to prove the Central Limit Theorem, arguably the most important result in statistics. As we shall see, there are other applications too.

Moment Generating Function. A moment generating function $\phi(t)$ is a real valued function on the reals defined for the random variable X by $\phi(t) = E(e^{tX})$. If X is discrete, we have

$$\phi(t) = \sum_x e^{tx} f(x).$$

If X is continuous, the defining formula becomes

$$\phi(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx.$$

We find a rule for one of the standard discrete random variables.

Example 6.2. *Given X is a Poisson random variable with parameter λ, find a rule for the moment generating function of X.*

$$\begin{aligned}
\phi(t) &= \sum_{n=1}^{\infty} e^{tn} e^{-\lambda} \Big(\frac{\lambda^n}{n!}\Big) \\
&= e^{-\lambda} \sum_{n=1}^{\infty} \Big(\frac{(\lambda e^t)^n}{n!}\Big) \\
&= e^{-\lambda} e^{\lambda e^t} \\
&= e^{\lambda(e^t - \lambda)}.
\end{aligned}$$

The last series in the computation is the Maclaurin series for $e^{\lambda e^t}$.

One property of moment generating functions that can be seen immediately is that

$$\phi(0) = E(e^{0 \cdot X}) = E(1) = 1.$$

Note that if X is continuous, we have

$$\begin{aligned}\phi'(t) &= \frac{d}{dt}\int_{-\infty}^{\infty} e^{tx} f(x)dx \\ &= \int_{-\infty}^{\infty} xe^{tx} f(x)dx \\ &= E(Xe^{tX}).\end{aligned}$$

Consequently, we have that $\phi'(0) = E(X)$ if the derivative of the generating function exists at zero. A similar derivation establishes the result for discrete random variables. The reader will note that the formula for the nth derivative is $\phi^{(n)}(t) = E(X^n e^{tX})$. We therefore have that $\phi^{(n)}(0) = E(X^n)$. This formula has many applications. We formulate the result once more and provide an example where it can be applied.

Moments of Random Variables. For the random variable X with moment generating function $\phi(t)$, we have that $\phi'(0) = E(X)$ and $\phi^{(n)}(0) = E(X^n)$, for $n = 2, 3, ...$, if the derivatives exist at zero.

Example 6.3. *Use moment generating functions to show that if X is Binomial with parameters n and p that $E(X) = np$ and $V(X) = np(1-p)$.*

We first note that

$$\begin{aligned}\phi(t) &= \sum_{k=0}^{n} e^{tk}\tbinom{n}{k}p^k(1-p)^{n-k} \\ &= \sum_{k=0}^{n}(e^t)^k\tbinom{n}{k}p^k(1-p)^{n-k} \\ &= \sum_{k=0}^{n}\tbinom{n}{k}(pe^t)^k(1-p)^{n-k} \\ &= (pe^t + 1 - p)^n.\end{aligned}$$

Hence,

$$\phi'(t) = n(pe^t + 1 - p)^{n-1}pe^t$$

and

$$\phi''(t) = n(n-1)(pe^t + 1 - p)^{n-2}pe^t \cdot pe^t + n(pe^t + 1 - p)^{n-1}pe^t,$$

so $\phi'(0) = np$ and $\phi''(0) = n(n-1)p \cdot p + np$. This gives us

$$E(X) = np,$$

and the variance of X is

$$E(X^2) - (E(X))^2 = n(n-1)p\cdot p + np - (np)^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p).$$

We note two more important facts about moment generating functions. First, they are unique so if one knows the moment generating function he/she can deduce what random variable he/she is dealing with. Second is that the moment generating function for the sum of two random variables is the product of the moment generating functions of the two random variables. This latter result has a simple derivation.

$$\begin{aligned}\phi_{X+Y}(t) &= E(e^{t(X+Y)}) \\ &= E(e^{tX}e^{tY}) \\ &= E(e^{tX})E(e^{tY}) \\ &= \phi_X(t)\phi_Y(t).\end{aligned}$$

We state these properties more generally:

Uniqueness of MGFs. Moment generating functions are unique. Moreover, if $X_1, X_2, ..., X_n$, are independent random variables, then

$$\phi_{X_1+X_2+\cdots+X_n}(t) = \phi_{X_1}(t) \cdot \phi_{X_2}(t) \cdots \phi_{X_n}(t).$$

We conclude this section with the formulation of a theorem needed for the proof of the Central Limit Theorem that comes up later in this chapter.

Continuity Theorem for Moment Generating Functions. Suppose $X_1, X_2, ...$ is a sequence of random variables with cumulative distribution functions $F_1, F_2, ...$, and moment generating functions $\phi_1, \phi_2, ...$. Suppose also that X is a random variable with cumulative distribution function F and moment generating function ϕ and that $\lim_{n\to\infty} \phi_n(t) = \phi(t)$ for all t in an open interval containing zero. Then $\lim_{n\to\infty} F_n(x) = F(x)$ for all x at which F is continuous.

Proving the Continuity Theorem is beyond the scope of this text. The same is the case for the fact that moment generating functions are unique. The reader should note that random variables can be studied in a superior rashion via characteristic functions $\phi_X(t) = E(e^{itx})$, where i is the imaginary number $\sqrt{-1}$. Understanding he mathematics behind these functions requires knowledge of complex analysis.

Exercises

(1) Calculate the moment generating function for a discrete random variable X for which $P(X = 2) = 3/4$ and $P(X = 5) = 1/4$.
 Ans. $\phi(t) = \frac{3e^{2t}}{4} + \frac{e^{5t}}{4}$

(2) Calculate the moment generating function for a continuous random variable X with probability density function

$$f(x) = \begin{cases} 2x & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Ans. $\phi(t) = \frac{2(te^t - e^t + 1)}{t^2}$

(3) Use moment generating functions to compute the standard deviation of a Poisson random variable with mean λ.

Ans. Since $\phi(t) = e^{\lambda(e^t - 1)}$, $E(X) = \lambda$, $E(X^2) = \lambda^2 + \lambda$, $SD(X) = \sqrt{\lambda}$. *Note that the Maclaurin series for* e^x *is needed to find* ϕ.

(4) Use moment generating functions to compute the standard deviation of an exponential random variable with mean μ.

Ans. $\phi(t) = \frac{\lambda}{\lambda - t}$, $E(X) = 1/\lambda$, $E(X^2) = 2/\lambda^2$, $SD(X) = 1/\lambda = E(X) = \mu$.

(5) Calculate the moment generating function of the uniform random variable on the interval $[a, b]$.

Ans. $\frac{e^{bt} - e^{at}}{(b-a)t}$

(6) Calculate the moment generating function of the standard normal random variable Z.

Ans. $e^{\frac{t^2}{2}}$

(7) Compute the mean and standard deviation of the random variable that has moment generating function $\phi(t) = \dfrac{1}{2} + \dfrac{1}{3}e^{6t} + \dfrac{1}{6}e^{12t}$.

Ans. 4, $\sqrt{20}$

(8) Compute the mean and standard deviation of the random variable that has moment generating function $\phi(t) = e^{2t^2 + 5t}$.

Ans. 5, 2

(9) If X and Y are independent Poisson random variables with means λ_X and λ_Y, respectively, use moment generating functions to show that $X + Y$ is Poisson with mean $\lambda_X + \lambda_Y$

We've already shown that $\phi_X(t) = e^{\lambda_X(e^t - 1)}$ *and that* $\phi_Y(t) = e^{\lambda_Y(e^t - 1)}$. *Hence,*

$$\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t) = e^{\lambda_X(e^t-1)} \cdot e^{\lambda_Y(e^t-1)} = e^{(\lambda_X + \lambda_Y)(e^t - 1)},$$

which is the mgf for a Poisson RV with mean $\lambda_X + \lambda_Y$.

6.4. Sums of Independent Random Variables

We now investigate the distribution of a sum of independent random variables. Assume $X_1, X_2, ..., X_n$ are independent random variables. We can obtain the moment generating function of the sum of these random variables by taking a product of the moment generating functions of the n random variables. We have

$$\phi_{X_1+X_2+\cdots+X_n}(t) = \phi_{X_1}(t)\phi_{X_2}(t)\cdots\phi_{X_n}(t).$$

Note that for the independent Poison random variables $X_1, X_2, ..., X_n$, with parameters $\lambda_1, \lambda_2, ..., \lambda_n$, respectively, we have

$$\begin{aligned} \phi_{X_1+X_2+\cdots+X_n}(t) &= e^{\lambda_1(e^t-1)} \cdot e^{\lambda_2(e^t-1)} \cdots e^{\lambda_n(e^t-1)} \\ &= e^{(\lambda_1+\lambda_2\cdots\lambda_n)(e^t-1)}. \end{aligned}$$

This is the probability mass function for a Poisson random variable with mean $\lambda_1 + \lambda_2 + \cdots + \lambda_n$. We state the result.

Sums of Poisson random variables.
If $X_1, X_2, ..., X_n$ are independent Poisson random variables with parameters $\lambda_1, \lambda_2, ..., \lambda_n$, respectively, then $X_1 + \cdots + X_n$ is a Poisson random variable with parameter $\lambda_1 + \cdots + \lambda_n$.

Example 6.4. *It is known that there are 19 tornadoes to touch down per year on average in Arkansas and two in Maine. The number of tornadoes to touch down in a region is accurately modeled to be Poisson. What's the probability that there will be exactly 20 tornadoes to touch down in Arkansas and Maine combined next year (assuming that the number touching down in Arkansas is independent of the number in Maine)?*

Since the sum of two independent Poisson random variables is Poisson with the mean being the sum of the individual means, the answer is

$$e^{-(19+2)}\frac{(19+2)^{20}}{20!} = 0.0867.$$

The reader will note that we are making the assumption that the number of tornadoes to touch down in Arkansas and in Maine are independent.

We now determine the probability density function for the sum of n independent normal random variables. We start with a single normal random variable X with mean μ and standard deviation σ. We incorporate the exponent on the factor e^{tx} into the exponent including x^2 and complete the square in x to obtain the rule:

$$\begin{aligned}
\phi(t) &= \int_{-\infty}^{\infty} e^{tx}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \\
&= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x^2-2(\mu+t\sigma^2)x+\mu^2)}{2\sigma^2}}dx \\
&= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x^2-2(\mu+t\sigma^2)x+(\mu+t\sigma^2)^2)}{2\sigma^2}+\frac{2\mu t\sigma^2+t^2\sigma^4}{2\sigma^2}}dx \\
&= e^{\mu t+\frac{t^2\sigma^2}{2}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-(\mu+t\sigma^2)^2}{2\sigma^2}}dx \\
&= e^{\mu t+\frac{t^2\sigma^2}{2}}\cdot 1 \\
&= e^{\frac{t^2\sigma^2}{2}+\mu t}.
\end{aligned}$$

The reader will note that the integral in the third to last line is in fact one since the integrand is the probability density function for a normal random variable with mean $\mu + t\sigma^2$ and standard deviation σ.

Note now that for the independent normal random variables $X_1, X_2, ..., X_n$, with means $\mu_1, \mu_2, ..., \mu_n$, respectively, and standard deviations $\sigma_1, \sigma_2, ..., \sigma_n$, respectively, we have

$$\begin{aligned}
\phi_{X_1+X_2+\cdots+X_n}(t) &= e^{\frac{t^2\sigma_1^2}{2}+\mu_1 t}\cdot e^{\frac{t^2\sigma_2^2}{2}+\mu_2 t}\cdots e^{\frac{t^2\sigma_n^2}{2}+\mu_n t} \\
&= e^{\frac{t^2(\sigma_1^2\cdots+\sigma_n^2)}{2}+(\mu_1+\cdots+\mu_n)t}.
\end{aligned}$$

This is the moment generating function for a normal random variable with mean $\mu_1 + \cdots + \mu_n$ and standard deviation $\sqrt{\sigma_1^2 + \cdots \sigma_n^2}$.

Sums of normal random variables. If $X_1, X_2, ..., X_n$ are independent normal random variables with means $\mu_1, ..., \mu_n$ and standard deviations $\sigma_1, ..., \sigma_n$, then the random variable $X_1 + \cdots X_n$ is normal with mean $\mu_1 + \mu_2 + \cdots + \mu_n$ and standard deviation $\sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2}$.

Using moment generating functions, it can be shown that for independent gamma random variables X with parameters λ and α and Y with parameters λ and β, it is the case that $X + Y$ is a gamma random variable with parameters λ and $\alpha + \beta$. Recalling that the exponential random variable with parameter λ is nothing more than a gamma random variable with parameters λ and 1, we list some interesting consequences:

Sums of exponential random variables.
If $X_1, X_2, ..., X_n$ are independent exponential random variables with parameter λ, then $X_1 + \cdots + X_n$ is a gamma random variable with parameters λ and n.

Now the sum $Z_1^2 + \cdots Z_n^2$ of the squares of independent standard normal random variables is referred to as the *chi-squared random variable with* n *degrees of freedom*. Note that the moment generating function for a chi-squared random variable with 1 degree of freedom, $\phi_{Z^2}(t)$ is given by

$$\phi_{Z^2}(t) = E(e^{Z^2 t}) = \int_{-\infty}^{\infty} e^{tx^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(1-2t)x^2}{2}} dx.$$

Letting $u = \sqrt{1-2t} \cdot x$, we obtain $dx = \frac{1}{\sqrt{1-2t}} du$ and

$$\phi_{Z^2}(t) = \frac{1}{\sqrt{1-2t}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{2}} du = (1-2t)^{-\frac{1}{2}}$$

for $t < \frac{1}{2}$. It's therefore the case that the chi-squared distribution with n degrees of freedom is

$$\left((1-2t)^{-\frac{1}{2}}\right)^n = (1-2t)^{-\frac{n}{2}}.$$

We will refer back to this moment generating function later when we ascertain the underlying distribution of a sample standard deviation.

Sums of squares of independent standard normal random variables.
If $Z_1, Z_2, ..., Z_n$ are independent standard normal random variables, then $Z_1^2 + \cdots + Z_n^2$ - referred to as the chi-squared random variable with n degrees of freedom and notated by X_n^2 - has moment generating function

$$\phi_{X_n^2}(t) = (1-2t)^{-\frac{n}{2}}$$

for $t < \frac{1}{2}$.

Exercises

(1) Suppose X and Y are independent Poisson RV's with parameters 2 and 3, respectively. Compute (a) $P(X+Y=6)$ and (b) $P(X+Y \geq 4)$.
Ans.: (a) $\frac{3125}{144e^5} = 0.1462$ *(b)* $1 - \frac{118}{3e^5} = 0.7350$

(2) Calls come in to customer service center at a rate of 4.3 per minute. Assuming calls arriving in two different minutes are independent, compute the probability that (a) at least six calls come in in a two minute period and (b) exactly 25 calls come in in a five minute period. Hint: Use the Poisson.
Ans.: (a) 0.8578 *(b)* 0.0607

(3) Suppose X and Y are independent normal random variables with the mean and standard deviation of X being 10 and 3, respectively, and the mean and standard deviation of Y being 14 and 4, respectively. Compute (a) $P(X+Y > 24)$ and (b) $P(X+Y < 25)$.
Ans.: (a) 0.5000 *(b)* 0.5793

(4) Suppose $X_1, ..., X_{10}$ are independent normal random variables, each with mean 2.0 and standard deviation 1.5. Compute (a) $P(X_1 + \cdots + X_{10} < 23.5)$ and (b) $P(1.8 \leq \frac{X_1+\cdots+X_{10}}{10} \leq 2.2)$
Ans.: (a) 0.7697 *(b)* 0.3267

(5) Suppose $X_1, ..., X_{20}$ are independent normal random variables, each with mean 0 and standard deviation 2. Compute (a) $P(X_1 + \cdots + X_{20} < 1)$ and (b) $P(-1.1 \leq \frac{X_1+\cdots+X_{20}}{20} \leq 1.1)$
Ans.: (a) 0.5445 *(b)* 0.9861

6.5. The Central Limit Theorem

Perhaps the most remarkable result in all of statistics is the Central Limit Theorem. We provide a rough formulation.

The Central Limit Theorem.
If n values are randomly sampled from a distribution with mean μ and standard deviation σ, then the sample mean of these values, $\bar{X}$, is approximately normal with mean μ and standard deviation $\dfrac{\sigma}{\sqrt{n}}$ for large n.

The theorem allows one to take as normal an average of a large number of values randomly selected from any distribution! We have chosen the vague term "large" here because the required size of n depends somewhat on the underlying distribution. If the distribution is symmetric, "large" might be taken as $n \geq 5$. If the distribution is highly asymmetric, however, "large" might mean $n \geq 200$. A good rule of thumb that applies in almost all cases is to take $n \geq 30$.

The Central Limit Theorem can be written more compactly as

$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim Z,$$

for $n \geq 30$.

We note that and equivalent way to write $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ is $\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$. A more precise statement of the theorem is as follows:

The Central Limit Theorem Precisely Stated. If c is a real constant and $X_1, X_2, ...,$ are independent random variables with mean μ and standard deviation σ, then

$$\lim_{n\to\infty} P\Big(\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \leq c\Big) = \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx.$$

Proof: It's sufficient to prove the theorem in the case that $\mu = 0$ because if we let $Y_n = X_n - \mu$ we have

$$\lim_{n\to\infty} P\Big(\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \leq c\Big) = \lim_{n\to\infty} P\Big(\frac{Y_1+\cdots+Y_n}{\sigma\sqrt{n}} \leq c\Big).$$

We therefore assume $\mu = 0$ and let $Z_n = \frac{X_1+\cdots+X_n}{\sigma\sqrt{n}}$. Then

$$\begin{aligned}
\phi_{Z_n}(t) &= E(e^{tZ_n}) \\
&= E(e^{\frac{t}{\sigma\sqrt{n}}(X_1+\cdots+X_n)}) \\
&= \phi_{X_1+\cdots+X_n}(\frac{t}{\sigma\sqrt{n}}) \\
&= \phi_{X_1}(\frac{t}{\sigma\sqrt{n}})\cdots\phi_{X_1}(\frac{t}{\sigma\sqrt{n}}) \\
&= [\phi_{X_1}(\frac{t}{\sigma\sqrt{n}})]^n.
\end{aligned}$$

We now take a limit of the the natural logarithm of $\phi_{Z_n}(t)$. We have

$$\begin{aligned}
\lim_{n\to\infty} \ln \phi_{Z_n}(t) &= \lim_{n\to\infty} \ln[\phi_{X_1}(\frac{t}{\sigma\sqrt{n}})]^n \\
&= \lim_{n\to\infty} n \ln[\phi_{X_1}(\frac{t}{\sigma\sqrt{n}})] \\
&= \lim_{n\to\infty} \frac{\ln[\phi_{X_1}(\frac{t}{\sigma\sqrt{n}})]}{1/n} \\
&= \lim_{n\to\infty} \frac{\phi'_{X_1}(\frac{t}{\sigma\sqrt{n}})(-\frac{1}{2})\frac{t}{\sigma n^{3/2}}/\phi_{X_1}(\frac{t}{\sigma\sqrt{n}})}{-1/n^2} \\
&= \lim_{n\to\infty} \frac{\phi'_{X_1}(\frac{t}{\sigma\sqrt{n}})t}{2\sigma\phi_{X_1}(\frac{t}{\sigma\sqrt{n}})n^{-1/2}} \\
&= \lim_{n\to\infty} \frac{\phi''_{X_1}(\frac{t}{\sigma\sqrt{n}})(-\frac{1}{2})\frac{t}{\sigma n^{3/2}}}{2\sigma\phi'_{X_1}(\frac{t}{\sigma\sqrt{n}})(-\frac{1}{2})\frac{t}{\sigma n^{3/2}}n^{-1/2} + 2\sigma\phi_{X_1}(\frac{t}{\sigma\sqrt{n}})(-\frac{1}{2})\frac{1}{n^{3/2}}} \\
&= \lim_{n\to\infty} \frac{\phi''_{X_1}(\frac{t}{\sigma\sqrt{n}})\frac{t^2}{\sigma}}{2\phi'_{X_1}(\frac{t}{\sigma\sqrt{n}})\frac{t}{\sqrt{n}} + 2\sigma\phi_{X_1}(\frac{t}{\sigma\sqrt{n}})} \\
&= \frac{\sigma^2 \cdot \frac{t^2}{\sigma}}{0 + 2\sigma \cdot 1} \\
&= \frac{t^2}{2}.
\end{aligned}$$

The reader will note that L'Hospital's Rule was used in going from line 3 to line 4 and then again from line 5 to line 6. The Rule is applicable since $\lim_{n\to\infty} \ln \phi_{X_1}(\frac{t}{\sigma\sqrt{n}}) = \ln \phi_{X_1}(0) = \ln 1 = 0$, and because $\lim_{n\to\infty} \phi'_{X_1}(\frac{t}{\sigma\sqrt{n}})t = \phi'_{X_1}(0)t = \mu t = 0 \cdot t = 0$.

Hence,

$$\lim_{n\to\infty} \phi_{Z_n}(t) = e^{\frac{t^2}{2}}.$$

Since $e^{\frac{t^2}{2}}$ is the moment generating function for the standard normal, we can apply the Continuity Theorem for Moment Generating Functions, and the proof is done. □

An application of the Central Limit theorem is provided.

Example 6.5. *It is known that ACT scores have mean of 21.0 and standard deviation of 5.2. What's the probability that the average of 40 randomly chosen ACT scores is (a) at most 19? (b) at least 20?*

(a) $P(\bar{X} \leq 19) = P(Z \leq \dfrac{19 - 21.0}{\frac{5.2}{\sqrt{40}}}) \cong P(Z \leq -2.43) \cong 0.0075$

(b) $P(\bar{X} \geq 20) = P(Z \geq \dfrac{20 - 21.0}{\frac{5.2}{\sqrt{40}}}) \cong P(Z \geq -1.22) \cong 0.8888$

A different formulation of the Central Limit Theorem is used to solve the next example problem. We compute the probability that a sum of random variables is within a certain interval. To apply the Central Limit Theorem, we convert this sum to an average.

Example 6.6. *A parcel carrier handles packages that weigh on average 16.4 lb with a standard deviation of 7.5 lb. What's the probability that the next 50 parcels she handles will weigh less than 750 lb altogether.*

Let X_i be the weight in pounds of the ith package. Then the answer is

$$\begin{aligned} P(X_1 + X_2 + \cdots X_{50} < 750) &= P\Big(\frac{X_1 + X_2 + \cdots X_{50}}{50} < \frac{750}{50}\Big) \\ &= P(\bar{X} < 15) \\ &= P\Big(Z < \frac{15 - 16.4}{\frac{7.5}{\sqrt{50}}}\Big) \\ &= P(Z < -1.32) \\ &= 0.093. \end{aligned}$$

6.6. Normal Approximation to the Binomial

An application of the Central Limit Theorem allows us to obtain a normal approximation to the binomial. Suppose X is binomial with parameters n and p. Then we can write

$$X = \sum_{i=1}^{n} X_i,$$

where the X_i's are independent Bernoulli random variables with parameter p. We therefore have that X/n is the sample mean $\bar{X}$ of n Bernoulli(p) random variables. We are converting a discrete random variable X to the continuous standard normal random variable Z, so for $k = 0, 1, 2, ..., n$ we write $P(X = k) = P(k - 1/2 < X < k + 1/2)$. According to the Central Limit Theorem (and recalling that $E(X_i) = p$ and $SD(X_i) = p(1 - p)$, we have that

$$\begin{aligned} P(k - 1/2 < X < k + 1/2) &= P(k - 1/2 < \sum_{i=1}^{n} X_i < k + 1/2) \\ &= P\Big(\frac{k - 1/2}{n} < \bar{X} < \frac{k + 1/2}{n}\Big) \\ &= P\Big(\frac{\frac{k-\frac{1}{2}}{n} - p}{\frac{\sqrt{p(1-p)}}{\sqrt{n}}} < Z < \frac{\frac{k+\frac{1}{2}}{n} - p}{\frac{\sqrt{p(1-p)}}{\sqrt{n}}}\Big) \\ &= P\Big(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}} < Z < \frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\Big) \end{aligned}$$

This allows us to write a pair of formulas that are of great use in approximating binomial probabilities with the standard normal.

Normal Approximation to the Binomial. If X is binomial with parameters n and p, n is large, and $k = 0, 1, 2, ..., n$, then

$$P(X \leq k) = P\Big(Z < \frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\Big)$$

and

$$P(X \geq k) = P\Big(Z > \frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\Big).$$

We demonstrate how these formulas might come in handy.

Example 6.7. *If you toss a fair coin 100 times, what's the probability you get at least 52 heads?*

Without the normal approximation, we'd have to compute individually the probabilities that we get exactly 52 heads, exactly 53 heads, etc., up to the probability of getting exactly 100 heads. We'd then have to add these 49 computations together to get our answer. Using the normal approximation, however, the computation is much simpler. Let X count the number of heads you get. Then

$$\begin{aligned} P(X \geq 52) &= P(Z > \frac{52 - \frac{1}{2} - 100(0.5)}{\sqrt{100(0.5)(1-0.5)}}) \\ &= P(Z > 0.3) \\ &= 0.3821. \end{aligned}$$

Exercises

(1) A certain cereal company packages "14 oz" boxes of cereal. These boxes have a mean weight of 14.1 oz with a standard deviation of 0.42 oz. What's the probability that 30 randomly selected boxes of this cereal have an average weight of at least 13.9 oz?

Ans. 0.995

(2) A bartender pours glasses of wine that are 10.5 oz on average with a standard deviation of 1.3 oz. (a) What's the probability that the bartender pours at least 10 oz for the next customer who requests a glass of wine? (b) What's the probability he pours at least 10 oz on average for the next 30 customers who request a glass of wine? (c) What's the probability that 430 oz of wine will suffice for the next 40 customers he pours glasses of wine?

Ans. (a) not enough information given to answer the problem, (b) 0.9824, *(c)* 0.888

(3) Hank sales propane tanks that have a mass of 17.46 kg with a standard deviation of 0.12 kg when empty. He loads 50 such tanks on the back of his truck. What's the probability that these 50 tanks have a mass of more than 875 kg combined?

Ans. 0.0092

(4) If you roll a balanced die 50 times, what's the probability you get (a) exactly eight 6's? (b) at least 10 6's?
Ans. *(a)* 0.1510 *(b)* 0.3290

(5) Suppose you take a true false exam by guessing on every question and that the lowest passing score is 60%. What's the probability you pass if the exam consists of (a) 30 questions, (b) 50 questions, (c) 100 questions?

(6) Suppose you take a multiple choice exam by guessing on every question and that the lowest passing score is 60%. What's the probability you pass if the exam consists of 30 questions, each with four answers?

(7) Twelve percent of customers at a fast food restaurant order a root beer. What's the probability that at least 30 of the next 200 customers at that restaurant will order a root beer?
Ans. 0.1157

(8) A facilities worker loads pieces of equipment onto a freight elevator that has a capacity of $1,800$ pounds. The pieces weigh 56.5 lb on average with a standard deviation of 18.5 lb. What's the probability that the next 30 pieces that need to be loaded on the elevator will be within the weight limit?
Ans. 0.851

(9) A particular laptop comes with a battery that will hold a charge for five hours and 32 minutes on average with a standard deviation of one hours and 56 minutes. If you purchase 40 such laptops, what's the probability that they hold a charge on average for five hours or more?
Ans. 0.959

Chapter 7

Random Processes

7.1. Markov Chains

We consider a sequence of random variables $X_0, X_1, ...$, each of which can take the values $1, 2, ..., N$. If the random variable X_n in the sequence takes the value j, we say that the process is in state j at time n. If the process is such that the value X_{n+1} takes is only dependent on the value that X_n takes (i.e. that future values are only dependent on the present value and not previous values), then it is said to be a Markov process. Using notation from probability, we have the following:

Markov Chains. The sequence $X_0, X_1, ...$ is called a Markov chain if

$$P(X_{n+1} = k | X_n = j, X_{n-1} = j_{n-1}, ..., X_0 = j_0) = P(X_{n+1} = k | X_n = j).$$

We write

$$p_{jk} = P(X_{n+1} = k | X_n = j),$$

and note that

$$p_{jk} \geq 0$$

and

$$\sum_{k=0}^{N} p_{jk} = 1$$

for $j = 1, 2, ..., N$. We call the p_{jk}'s the transition probabilities of the Markov chain.

It's helpful to list the transition probabilities in a square $N \times N$ matrix as follows:

$$\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{bmatrix}$$

This transition matrix allows us to compute probabilities of interest. The reader will note that we can write

$$\begin{aligned} P(X_n = j_n, X_{n-1} = j_{n-1}, ..., X_0 = j_0) &= P(X_n = j_n | X_{n-1} = j_{n-1}, ..., X_0 = j_0) \\ &\quad \times P(X_{n-1} = j_{n-1}, ..., X_0 = j_0) \\ &= p_{j_{n-1}j_n} \cdot P(X_{n-1} = j_{n-1}, ..., X_0 = j_0). \end{aligned}$$

Repeating this argument, we obtain

$$\begin{aligned} P(X_n = j_n, X_{n-1} = j_{n-1}, ..., X_0 = j_0) & \\ &= p_{j_0 j_1} p_{j_1 j_2} \cdots p_{j_{n-1} j_n} P(X_0 = j_0). \end{aligned}$$

Now p_{jk} is the probability of going from state j to state k in one step. The notation $p_{jk}^{(m)}$ is used to represent the probability of going from state j to state k in m steps. We have

$$p_{jk}^{(m)} = P(X_{n+m} = k | X_n = j).$$

We can arrive at the so called Chapman-Kolmogorov equations by noting that

$$\begin{aligned} p_{jk}^{(m)} &= P(X_m = k | X_0 = j) \\ & \sum_{i=1}^{N} P(X_m = k, X_r = i | X_0 = j) \\ & \sum_{i=1}^{N} P(X_m = k | X_r = i, X_0 = j) P(X_r = i | X_0 = j) \\ &= \sum_{i=1}^{N} p_{ik}^{(m-r)} p_{ji}^{(r)}. \end{aligned}$$

We have

Chapman-Kolmogorov Equations. If r is an integer between 0 and m, we have that

$$p_{jk}^{(m)} = \sum_{i=1}^{N} p_{ik}^{(m-r)} p_{ji}^{(r)}.$$

The transition matrix is of use to determine multistep transitions.

Powers of the Transition Matrix. Taking P to be the transition matrix

$$\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{bmatrix}$$

with p_{jk} representing to probability of moving from state j to state k in one step, we have that the entry in the jth row and kth column of P^n is the probability $p_{jk}^{(n)}$ of transitioning from state j to state k in exactly n steps.

Example 7.1. *Suppose that in a certain region the weather is such that a sunny day is followed by a stormy day with probability $\frac{1}{8}$ and a stormy day is followed by a stormy day with probability $\frac{1}{3}$. Taking a sunny day to be State 1 and a stormy day to be State 2, compute the probability that on the third day after a sunny day it will be stormy.*

To solve this problem we note that the transition matrix is

$$P = \begin{bmatrix} 7/8 & 1/8 \\ 2/3 & 1/3 \end{bmatrix}$$

The third step transition matrix is therefore

$$P^3 = \begin{bmatrix} 0.8435 & 0.1565 \\ 0.8345 & 0.1655 \end{bmatrix}$$

Consequently, the answer to the question is $p_{12}^{(3)} = 0.1565$.

If it's the case that for some positive integer m we have that

$$p_{jk}^{(m)} > 0$$

for all $j, k = 1, 2, ..., N$, then the Markov chain is said to be *ergodic*. When the chain is ergodic, the limit

$$\lim_{n\to\infty} p_{jk}^{(n)}$$

exists. We label it π_k. Using the Chapman-Kolmogorov equations, the following result can be established:

Ergodic Theorem. For an ergodic Markov chain,

$$\pi_k = \lim_{n\to\infty} p_{jk}^{(n)}$$

exists. Moreover,

$$\pi_k = \sum_{j=1}^{N} \pi_j p_{jk}$$

and

$$\sum_{k=1}^{N} \pi_k = 1.$$

Writing $\pi = \langle \pi_1, \pi_2, ..., \pi_N \rangle$, we have

$$\pi P = \pi.$$

Example 7.2. *Suppose that in a certain region the weather is such that a sunny day is followed by a stormy day with probability $\frac{1}{8}$ and a stormy day is followed by a stormy day with probability $\frac{1}{3}$. Taking a sunny day to be State 1 and a stormy day to be State 2, compute the probability vector π mentioned in the Ergodic Theorem.*

To solve this problem we note that

$$\lim_{n\to\infty} P^n = \begin{bmatrix} 0.8421 & 0.1579 \\ 0.8421 & 0.1579 \end{bmatrix}$$

Consequently, the answer to the question is $\pi_1 = 0.8421$ and $\pi_2 = 0.1579$ so that

$$\pi = \begin{bmatrix} 0.8421 & 0.1579 \end{bmatrix}$$

Note also that

$$\begin{bmatrix} 0.8421 & 0.1579 \end{bmatrix} \begin{bmatrix} 0.8421 & 0.1579 \\ 0.8421 & 0.1579 \end{bmatrix} = \begin{bmatrix} 0.8421 & 0.1579 \end{bmatrix}$$

The row vector $\mathbf{x} = \langle x_1, x_2, ..., x_N \rangle$ is said to be the *state vector* for the Markov chain at a given observation if the kth component x_k is the probability the system is in the kth state at that time. To ascertain the state vector for a specific observation we use the transition matrix.

State Transitions. A Markov chain with transition matrix P and state vector $\mathbf{x}^{(n)}$ at the nth observation is such that

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} P.$$

Accordingly, we have

$$\mathbf{x}^{(n)} = \mathbf{x}^{(0)} P^n.$$

Example 7.3. *Returning to the sunny/stormy example, if we start out on a sunny day, what's the probability that four days later it is (a) sunny and (b) stormy?*

Again we take State 1 to be sunny and State 2 to be stormy. We start on a sunny day so we have

$$\mathbf{x}^{(0)} = \langle 1, 0 \rangle.$$

We note that

$$\mathbf{x}P^4 = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 7/8 & 1/8 \\ 2/3 & 1/3 \end{bmatrix}^4 = \begin{bmatrix} 0.8424 & 0.1576 \end{bmatrix}$$

Hence, there's about an 84.2% chance it will be sunny and a 15.8% chance it will be stormy on the fourth day.

Example 7.4. *Consider a random walk on the real line in which an entity either moves one step to the right with probability p or one step to the left with probability $1-p$. We compute the transition probability for going from state i to state j in exactly n steps.*

To compute $p_{ij}^{(n)}$ we note that the transition has to consist of k steps to the right, where k is a nonnegative integer less than or equal to n, and $n-k$ steps to the left. This has to happen in such a way that

$$i + k - (n-k) = j.$$

This implies $2k = n - i + j$ or

$$k = \frac{n-i+j}{2}.$$

Hence, the transition probability for n steps is

$$p_{ij}^{(n)} = \begin{cases} \binom{n}{\frac{n-i+j}{2}} p^{\frac{n-i+j}{2}} (1-p)^{n-\frac{n-i+j}{2}} & \text{if } \frac{n-i+j}{2} = 0, 1, 2, ..., \text{ or } n \\ 0 & \text{otherwise} \end{cases}$$

Exercises

(1) Suppose a Markov chain $X_0, X_1, ...,$ has the two states 1 and 2, and transition matrix

$$P = \begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix}.$$

(a) What's the probability of going from State 2 to State 1 in one step? (b) What's the probability of going from State 2 to State 1 in two steps? (c) Is the process ergodic? (d) If it is ergodic, compute the stable state vector π with each component to four decimal places. (e) If the process starts with state vector $\mathbf{x}^{(0)} = \langle 1/2, 1/2 \rangle$, compute the state vector $\mathbf{x}^{(2)}$.

Ans. (a) 0.3 *(b)* 0.39 *(c) Yes (d)* $\langle 0.4286, 0.5714 \rangle$ *(e)* $\langle 0.435, 0.565 \rangle$

(2) Suppose a Markov chain $X_0, X_1, \ldots$, has the three states 1, 2, and 3, and transition matrix

$$P = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 1/2 & 1/6 \\ 1 & 0 & 0 \end{bmatrix}.$$

(a) What's the probability of going from State 2 to State 1 in one step? (b) What's the probability of going from State 2 to State 1 in two steps? (c) Is the process ergodic? (d) If it is ergodic, compute the stable state vector π with each component to four decimal places? (e) If the process starts with state vector $\mathbf{x}^{(0)} = \langle 0, 1, 0 \rangle$, compute the state vector $\mathbf{x}^{(2)}$.

Ans. *(a)* 1/3 *(b)* 1/2 *(c) Yes since every entry in* P^2 *is positive* *(d)* $\langle 0.5455, 0.2727, 0.1818 \rangle$ *(e)* $\langle 1/2, 1/3, 1/6 \rangle$

(3) Place \$100 bets on red in roulette so that you can be in a state of having either \$0, \$100, \$200, \$300, \$400, or \$500 in funds. Once you run our of money or accumulate \$500 you stay where you are. (a) Write out the transition matrix for the six states. (b) If you start with \$300, what's the probability you'll be broke after betting five times?

Ans. *(a)*

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 20/38 & 0 & 18/38 & 0 & 0 & 0 \\ 0 & 20/38 & 0 & 18/38 & 0 & 0 \\ 0 & 0 & 20/38 & 0 & 18/38 & 0 \\ 0 & 0 & 0 & 20/38 & 0 & 18/38 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

(b) 0.2548

(4) In a random walk, assume a particle can move one step to the right with probability 0.52 and one step to the left with probability 0.48. What then is the probability that the particle will move two steps to the right (a) in 20 steps and (b) in five steps.

Ans. *(a)* $\binom{20}{11}(0.52)^{11}(0.48)^9 = 0.1708$ *(b)* 0

Index