Probability Theory (Undergraduate)

Yiqiao YIN
Columbia University

December 11, 2018

Abstract

This document is prepared for students in Probability Theory (Undergraduate) offered at Columbia University in 2018 Fall semester. The course instructor is Professor Shaw-Hwa Lo. The grader for this course is Xiaotong. The document serves students by providing lecture notes as well as homework and exam guidance. I am grateful for Professor Shaw-Hwa Lo for providing materials and I thank Xiaotong for providing comments of this document. I am the TA for this class. Please email me yy2502columbia.edu if you have any questions.

Contents

1 Counting Method

Go back to Table of Contents. Please click TOC

1.1 Introduction

Let us start with an example about an event with multiple possible outcomes. An experiment, for example, can lead to n possible outcomes:

$$a_1, a_2, ..., a_n$$

For each a_i there are m possible outcomes from another experiment, then together there are nm possible outcomes. For example, how many outcomes one possible by tossing a coin twice? We can approach to answer this question from frequentist point of view, which is objective. One can start with a fair coin. Tossing the coin for the first time, one will observe either head or tail. By tossing the coin for the second time, one will observe, assuming using a fair coin, head or tail. One can continue this experiment and will observe, assuming tossing coin twice, the following $\{HH, HT, TH, TT\}$, using "H" for heads and "T" for tails.

It is not always the case that an experiment can be repeated. We collect previous data, called prior. We can observe new data as time moves on and we can use it as new information. We update our prior data with new information and we will arrive a more sophisticated analysis, which is called posterior. This school of thoughts, which is called Bayesian, are usually more subjective due to the fact that there is a prior before analysis starts.

Frequentist and Bayesian are two schools of thoughts in the field of statistics. Frequentist dominated the field in the 60's and 70's. Starting since the past 20 to 30 years, there are a good amount of Bayesian approach emerged.

For this course, we will mostly be dealing with repeatable experiments. We may have different outcomes, but the experiments we will be discussing can be replicated under the same condition.

1.2 Permutation and Combination

Let us look at the following example. Consider the word "statistics". How many different letter arrangements would you have? Let us separate these letters and count that there are three s's, three t's, two i's, one a, and one c. In this case we have permutation

$$10! = 10 \times 9 \times \cdots \times 2 \times 1$$

but we need to consider the cases that if you switch two's, example, the results will be the same in the word "statistics". Hence, we need the following

$$3!3!2!1!$$

and together we have

$$\frac{n!}{n_1! \dots n_k!}$$

different ways to arrange n objects, of which $n_1, n_2, ..., n_k$ are alike.

Definition 1.2.1. Suppose now that we have n objects. Reasoning similar to that we have just used for the letters example in lecture then shows that there are

$$n(n-1)(n-2)\dots(3)(2)(1) = n!$$

different permutations of the n objects.

Definition 1.2.2. In general, the same reasoning used in lecture shows that there are

$$\frac{n!}{n_1!n_2!\dots n_r!}$$

different permutations of n objects, of which n_1 are alike, n_2 are alike, ..., etc.

Example 1.2.3. Let us look at another example. How many different groups of 3 can be related from 5 items, A, B, C, D, and E? In this case, There are $\binom{5}{3} = 10$ number of different groups of 3.

Definition 1.2.4. We define $\binom{n}{r}$, for $r \leq n$, by

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

and say that $\binom{n}{r}$ represents the number of possible combinations of n objects taken r at a time.

Another interesting example is the following.

Example 1.2.5. For a class of 20, 12 boys and 8 girls. How many different groups consisting of 3 boys and 2 girls can be formed? What if 2 of the boys refuse to be in the same group together?

A useful formula is in the following

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

and the famous binomial formula is

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Proof. Consider the following

$$(x+y)^n = \underbrace{(x+y)\dots(x+y)}_{n \text{ times}}$$

if $n = 2$:

$$(x+y)^2 = x^2 + xy + yx + y^2, \text{(each term contributes once)}$$

while each x and y contributes $x^k y^{n-k}$. There are n choose k, i.e. $\binom{n}{k}$, number of these contributions and that's why it leads to $\binom{n}{k} x^k y^{n-k}$. □

Example 1.2.6. Ten balls marked 1 to 10 are put into 3 bags A, B, and C with 3 in A, 3 in B, and 4 in bag C. How many ways? Bags are all assumed to be distinct. We have 3! in the first bag, 3! in the second bag, and 4! in the third bag. In total, there are 10! possible outcome for all 10 balls. Thus, the answer is

$$\frac{10!}{3!3!4!}$$

Let us formally introduces binomial theorem.

Theorem 1.2.7. *IMPORTANT The binomial theorem states the following*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Proof. When $n = 1$, we have

$$x + y = \binom{1}{0} x^0 y^1 + \binom{1}{1} x^1 y^0 = y + x$$

Assume the above results hold for $n - 1$. Now, consider

$$\begin{aligned}
(x+y)^n &= (x+y)(x+y)^{n-1} \\
&= (x+y) \sum_{k=0}^{n-1} \binom{n-1}{k} x^k y^{n-1-k} \\
&= \sum_{k=0}^{n-1} \binom{n-1}{k} x^{k+1} y^{n-1-k} + \sum_{k=0}^{n-1} \binom{n-1}{k} x^k y^{n-k}
\end{aligned}$$

Letting $i = k + 1$ for the first term and $i = k$ for the second term, we have

$$\begin{aligned}
(x+y)^n &= \sum_{i=1}^{n} \binom{n-1}{i-1} x^i y^{n-i} + \sum_{i=1}^{n} \binom{n-1}{i} x^i y^{n-i} \\
&= x^n + \sum_{i=1}^{n-1} \left[\binom{n-1}{i-1} + \binom{n-1}{i} \right] x^i y^{n-i} + y^n \\
&= \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}
\end{aligned}$$

and we are done. □

Let us introduce the following propositions that may be helpful in this topic.

Proposition 1.2.8. *There are* $\binom{n-1}{r-1}$ *distinct positive integer-valued vectors* $(x_1, ..., x_r)$ *satisfying the equation*

$$x_1 + x_2 + \cdots + x_r = n, x_i > 0, i = 1, ..., r$$

Proposition 1.2.9. *There are* $\binom{n+r-1}{r-1}$ *distinct nonnegative integer-valued vectors* $(x_1, ..., x_r)$ *satisfying*

$$x_1 + x_2 + \cdots + x_r = n$$

Example 1.2.10. How many distinct nonnegative integer-valued solutions of $x_1 + x_2 = 3$ are possible?

Answer. There are $\binom{3+2-1}{2-1} = 4$ such solutions: (0,3), (1,2), (2,1), (3,0). □

2 Axioms of Probability

Go back to Table of Contents. Please click TOC

This section introduces the concepts of the probability of an event and then show how probabilities can be computed in certain situations.

2.1 Sample Space and Events

We define event as a specific situation. It is a subset of an outcome defined in an experiment. For example, tossing a fair coin once can result in head or tail. The event can be head or tail. Sample space is the superset of all the possible outcomes. For example, tossing two fair coins together, the sample space consists of $\{HH, HT, TH, TT\}$.

Example 2.1.1. Two draws are made from the following box with 3 balls, call them G, Y, and B.

1. Consider all possible arrangements with replacement. We have $3 \times 3 = 9$. That is, we have

$$S = \{(G, G), (B, B), (Y, Y), (G, B), (G, Y), (B, G), (B, Y), (Y, G), (Y, B)\}$$

   It can be anywhere in the following matrix.

1/2	G	Y	B
G	(G, G)	$()$	(G, B)
Y	$()$	(Y, Y)	$()$
B	$()$	$()$	(B, B)

Same question without replacement. Then we have $3 \times 2 = 6$. In matrix form, we do not count the diagonal because without replacement means once G is drawn it cannot appear for a second time.

Definition 2.1.2. Event E and event F are said to be mutually exclusive if $E \cap F = \emptyset$, that is, there is no element that simultaneously exists in E and F.

The operations forming unions, intersections, and complements of events obey certain rules similar to the rules of algebra.

Proposition 2.1.3. *We have the following rules*

1. *Commutative laws:* $E \cup F = F \cup E$*, or* $EF = FE$
2. *Associative laws:* $(E \cup F) \cup G = E \cup (F \cup G)$*, or* $(EF)G = E(FG)$
3. *Distribution laws:* $(E \cup F)G = EG \cup FG$ *or* $EF \cup G = (E \cup G)(F \cup G)$

Theorem 2.1.4. *DeMorgan's Law states that*

$$\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c$$

$$\left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i^c$$

Example 2.1.5. A famous example is to consider event E and F, by DeMorgan's law, we have

$$(E \cup F)^c = E^c F^c \text{ and } (EF)^c = E^c \cup F^c$$

Proposition 2.1.6. *Let us introduce the following proposition, called inclusion-exclusion identity.*

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1}E_{i_2}) + \dots + (-1)^{r+1} \sum_{i_1 < i_2 < \cdots < i_r} P(E_{i_1}E_{i_2}\dots E_{i_r}) + \cdots + (-1)^{n+1} P(E_1 E_2 \dots E_n)$$

Example 2.1.7. This problem is from text[1] page 34. An urn contains n balls, one of which is special. If k of these balls are withdrawn one at a time, with each selection being equally likely to be any of the balls that remain at the time, what is the probability that the special ball is chosen?

Solution. Since all of the balls are treated in an identical manner, it follows that the set of k balls selected is equally likely to be any of the $\binom{n}{k}$ sets of k balls. Therefore,

$$P(\text{special ball is selected}) = \frac{\binom{1}{1}\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{k}{n}$$

We could also have obtained this result by letting A_i denote the event that the special ball is the ith ball to be chosen, $i = 1, \dots, k$. Then, since each one of the n balls is equally likely to be the ith ball chosen, it follows that $P(A_i) = 1/n$. Hence, because these events are clearly mutually exclusive, we have

$$P(\text{special ball is selected}) = P\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} P(A_i) = \frac{k}{n}$$

We could also have argued that $P(A_i) = 1/n$, by noting that there are $n(n-1)\dots(n-k+1) = n!/(n-k)!$ equally likely outcomes of the experiment, of which $(n-1)!/(n-k)!$ result in the special ball being the ith one chosen. From this reasoning, it follows that

$$P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$$

□

Let us introduce some simple properties (notes from class).

1. If $E \subset F$, then $P(E) \le P(F)$ and $F = E \cup (E^c \cap F) \Rightarrow P(F) \ge P(E)$
2. $P(E \cup F) = P(E) + P(F) - P(EF)$ and also $P(E \cup F) = P(E) + P(E^c F)$ and $P(E^c F) = P(F) - P(EF)$
3.

$$P(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} P(E_i) - \sum_{i<j} P(E_i E_j) + \dots + (-1)^{k+1} \sum_{i_1 < i_2 < \cdots < i_k} P(E_{i1}E_{i2}\dots E_{ik}) + (-1)^{n+1} P(E_1 E_2 \dots E_n)$$

Sample space with equal chances for all outcome. If S is such a sample space $|S| = n$, the size of S for $w \in S$. This gives us $P(w) = \frac{1}{n}$. For all $E \subset S$, an event, we have $P(E) = \frac{|E|}{|S|} = \frac{m}{n}$ if $|E| = m$ and for $m \leq n$.

2.2 Axioms of Probability

Consider an experiment whose sample space is S. For each event E of the sample space S, we assume that a number $P(E)$ is defined and satisfies the following three axioms

Proposition 2.2.1. *The three axioms of probability*

1. *Axiom 1:* $0 \leq \mathrm{P}(E) \leq 1$
2. *Axiom 2:* $P(S) = 1$
3. *Axiom 3: For any sequence of mutually exclusive events $E_1, E_2, \ldots$ (that is, events of for which $E_i E_j = \emptyset$ when $i \neq j$),*

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

We refer to $P(E)$ as the probability of the event E.

Example 2.2.2. If our experiment consists of tossing a coin and if we assume that a head is as likely to appear as a tail, then we have

$$P(\{H\}) = \frac{1}{2} \text{ and } P(\{T\}) = \frac{1}{3}$$

However, if the coin were biased and we believed that a head were twice as likely to appear as a tail, we should have

$$P(\{H\}) = \frac{2}{3} \text{ and } P(\{T\}) = \frac{1}{3}$$

Let us elaborate the experiment a little in the following example.

Example 2.2.3. If a die is rolled and suppose that all six sides are equally likely to appear, then we have $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}$. From Axiom 3, we can compute the probability of rolling an even number to be

$$P(\{2, 4, 6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{2}$$

Let us introduce more properties.

Proposition 2.2.4.

$$P(E^c) = 1 - P(E)$$

Proposition 2.2.5. *If $E \subset F$, then $P(E) \leq P(F)$.*

Proposition 2.2.6.

$$P(E \cup F) = P(E) + P(F) - P(EF)$$

Proposition 2.2.7.

$$\begin{aligned} P(E_1 \cup E_2 \cup \ldots E_n) = & \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} E_{i_2}) \\ & + \cdots + (-1)^{r+1} \sum_{i_1 < \ldots i_r} P(E_{i_1} E_{i_2} \ldots E_{i_r}) \\ & + \cdots + (-1)^{n+1} P(E_1 E_2 \ldots E_n) \end{aligned}$$

Example 2.2.8. A committee of 5 is to be selected from a group of 6 men and 9 women. If the selection is made randomly, what is the probability that the committee consists of 3 men and 2 women?

Answer. Because each of the $\binom{15}{5}$ possible committees is equally likely, we know bottom of the fraction is $\binom{15}{5}$. Then we only need to find top of the fraction, which is the number of possible choices for men times the number of possible choices for women, e.g. $\binom{6}{3}\binom{9}{2}$. Hence, the final answer is

$$\frac{\binom{6}{3}\binom{9}{2}}{\binom{15}{5}} = \frac{240}{1001}$$

□

3 Conditional Probability and Independence

Go back to Table of Contents. Please click TOC

Let E and F denote, respectively, the event that the sum of the dice is 8 and the event that the first die is a 3, then the probability just obtained is called conditional probability that E occurs given that F has occurred. This term is denoted by

$$P(E|F)$$

A general formula for $P(E|F)$ that is valid for all events E and F is derived in the same manner: If the event F occurs, in order for E to occur, it is necessary that the actual occurrence be a point in both E and F; that is, in EF.

Definition 3.0.1. If $P(F) > 0$, then

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Proposition 3.0.2. *The multiplication rule.*

$$P(E_1E_2E_3 \dots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2) \dots P(E_n|E_1 \dots E_{n-1})$$

Example 3.0.3. Let us discuss an example. Toss a die twice n and toss two dice. There are 36 outcomes. Assume all outcomes are equally likely (fair dice). There is 1/36 chance for each outcome. Suppose that we observe that the first die is a 4. Given this information, what is the chance the sum of the two dice is no bigger than 7?

In this case, we have (4,1), (4,2), (4,3), (4,4), (4,5), (4,6). Then the chance 3/6 = 1/2.

Suppose we observe one of two dice is a 4, then the probability becomes what? The answer is 6/11.

Remark 3.0.4. Let E, F be two events and let $P(F) > 0$, and we have

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Given F already occurred, the chance of E will occur is $P(E|F)$

3.1 Bayes' Formula

Let E and F be events. We may express E as

$$E = EF \quad EF^c$$

for, in order for an outcome to be in E, it must either be in both E and F or be in E but not in F. As EF and EF^c are clearly mutually exclusive, we have, by Axiom 3, we have

$$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \\ &= P(E|F)P(F) + P(E|F^c)[1 - P(F)] \end{aligned}$$

Definition 3.1.1. The odds of an event A are defined by

$$\frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}$$

That is, the odds of an event A tell how much more likely it is that the event A occurs than it is does not occur. For instance, $P(A) = \frac{2}{3}$, then $P(A) = 2P(A^c)$, so the odds are 2. If the odds are equal to α, then it is common to say that the odds are "α to 1" in favor of the hypothesis.

Consider now a hypothesis H that is true with probability $P(H)$, and suppose that new evidence E is introduced. Then, the conditional probabilities, given evidence E, that H is that and that H is not true are respectively given by

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \text{ and } P(H^c|E) = \frac{P(E|H^c)P(H^c)}{P(E)}$$

Therefore, the new odds after the evidence E has been introduced are

$$\frac{P(H|E)}{P(H^c|E)} = \frac{P(H)}{P(H^c)}\frac{P(E|H)}{P(E|H^c)}$$

We can further generalize: Suppose that $F_1, ..., F_n$ are mutually exclusive events such that

$$\bigcup_{i=1}^{n} F_i = S$$

In other words, exactly one of the events $F_1, ..., F_n$ must occur. By writing,

$$E = \bigcup_{i=1}^{n} EF_i$$

and using the fact that the events EF_i, for $i = 1, ..., n$ are mutually exclusive, we obtain

$$P(E) = \sum_{i=1}^{n} P(EF_i) = \sum_{i=1}^{n} P(E|F_i)P(F_i)$$

Let $F_1, ..., F_n$ be a set of mutually exclusive and exhaustive events (meaning that exactly one of these events must occur). Suppose now that E has occurred and we are interested in determining which one of the F_j also occurred. Then we have

Proposition 3.1.2. *IMPORTANT We have the following proposition:*

$$\begin{aligned} P(F_j|E) &= \frac{P(EF_j)}{P(E)} \\ &= \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)} \end{aligned}$$

which is known as Bayes' formula.

Example 3.1.3. Discuss an example from class. There is a fair deck of cards (a fair deck has 52 cards and each is drawn equally likely). Deal the cards to 4 players (each has 1 card), say E, W, N, and S. If North has 6 spades, what is the chance that East has 3 spades?

Use reduced sample space: Given N has 6 spades and other 7 (non-spade), E, W, S will share other 7 spades among $13 \times 3 = 39$ cards. Hence,

$$\frac{\binom{7}{3}\binom{32}{10}}{\binom{39}{13}}$$

3.2 Independent Events

From the idea of Bayes' rule, we can discuss the notion of independent events.

Definition 3.2.1. Consider two events E and F. They are independent if equation

$$P(EF) = P(E)P(F)$$

holds. If they are not independent, we say they are dependent.

Proposition 3.2.2. *If E and F are independent, then so are E and F^c.*

Definition 3.2.3. Three events E, F, and G are said to be independent if

$$\begin{aligned} P(EFG) &= P(E)P(F)P(G) \\ P(EF) &= P(E)P(F) \\ P(EG) &= P(E)P(G) \\ P(FG) &= P(F)P(G) \end{aligned}$$

Example 3.2.4. A famous example is Gambler's Ruin. Please see [1] page 84.

Proposition 3.2.5. *We have the following properties*

1. $0 \leq P(E|F) \leq 1$
2. $P(S|F) = 1$
3. *If E_i, for $i = 1, 2, ...$ are mutually exclusive events, then*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i | F\right) = \sum_{i=1}^{\infty} P(E_i|F)$$

Let us discuss a birth problem as this problem can be related many problems in conditional probabilities.

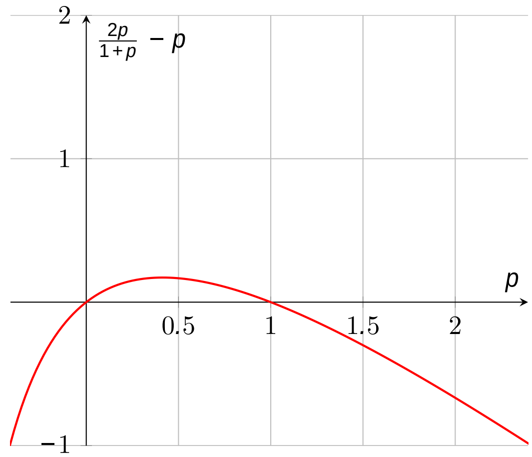
Example 3.2.6. Female chimp gave birth. It is not certain which of two male chimps is the father. Before genetic analysis, it is believed that the probability that male number 1 is the father is p and the probability that male number 2 is the father is $1 - p$. DNA obtained from the mother, male number 1, and male number 2 indicates that on one specific location of the genome, the mother has the gene pair (A,A), male number 1 has gene pair (a,a), and male number 2 has the gene pair (A,a). If a DNA test shows that the baby chimp has the gene pair (A,a), what is the probability that male number 1 is the father?

Answer. Let M_i be the event that male number i is the father. Let $B_{A,a}$ be the event that baby chimp has the gene pair (A,a). Then $P(M_1|B_{A,a})$ is obtained:

$$\begin{aligned} P(M_1|B_{A,a}) &= \frac{P(M_1B_{A,a})}{P(B_{A,a})} \\ &= \frac{P(B_{A,a}|M_1)P(M_1)}{P(B_{A,a}|M_1)P(M_1) + P(B_{A,a}|M_2)P(M_2)} \\ &= \frac{1 \cdot p}{1 \cdot p + (1/2)(1-p)} \\ &= \frac{2p}{1+p} \end{aligned}$$

Now let us compare result with p

Figure 1: The figure presents the graph of $\frac{2p}{1+p} - p$.



Hence, we arrived the inequality

$$\frac{2p}{1+p} > p$$

We conclude the information that the baby's gene pair is (A,a) increases the probability that male number 1 is the father. $\square$

4 Random Variables

Go back to Table of Contents. Please click TOC

This section we discuss random variables. We will start with definition of random variables and begin our discussion with discrete random variables. We can then discuss expectation and variance (1st moment and 2nd moment) of the random variables. Afterwards, we will move forward to discuss Bernoulli and Binomial random variables (and Poisson) as special case studies.

4.1 Random Variables

When we perform an experiment, we are often times interested in some function of the outcome. This way can generalize the situation and what will occur in future events.

Example 4.1.1. Consider an experiment of tossing 3 fair coins. Let Y denote the number of heads. Then Y is a random variable taking one of the values 0, 1, 2, and 3 with a certain probability respectively. We can write

$$\begin{aligned} P(Y = 0) &= 1/8 \\ P(Y = 1) &= 3/8 \\ P(Y = 2) &= 3/8 \\ P(Y = 3) &= 1/8 \end{aligned}$$

We notice that since Y must be one of the values from 0 through 3, then we must have

$$1 = P\left(\bigcup_{i=0}^{3} (Y = i)\right) = \sum_{i=0}^{3} P(Y = i)$$

Example 4.1.2. Consider another example. Four balls are to be randomly selected, without replacement, from an urn that contains 20 balls numbered 1 through 20. (Up to here, we know there are $\binom{20}{4}$ possible outcomes.) Let X be the largest numbered ball selected, then X is a random variable that takes on one of the values 4, 5, ..., 20. The probability that X takes on each of its possible values is

$$P(X = i) = \frac{\binom{i-1}{3}}{\binom{20}{4}}, \text{ for } i = 4, ..., 20$$

Suppose we want to determine $P(X > 10)$. One way is to compute

$$P(X > 10) = \sum_{i=11}^{20} P(X = i) = \sum_{i=11}^{20} \frac{\binom{i-1}{3}}{\binom{20}{4}}$$

We can, alternatively, compute the complement of the above event and subtract that probability from 100%. We omit the computation here. One can refer to text [1] page 113.

All of these are motivating examples that it is necessary to come up with a notion of random variable instead of paying attention to a single event.

4.2 Discrete Random Variables

A random variable that can take on at most a countable number of possible values is said to be discrete. For a discrete random variable X, we define the probability mass function $p(a)$ of X by

$$p(a) = P(X = a)$$

The probability mass function $p(a)$ is positive for at most a countable number of values of a. That is, if X must assume one of the values $x_1, ...,$ then

$$p(x_i) \geq 0 \text{ for } i = 1, 2, ...$$
$$p(x) = 0 \text{ for all other values of } x$$

Since X must take on one of the values x_i, we have

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

Example 4.2.1. The probability mass function of a random variable X is given by $p(i) = c\lambda^i/i!$, $i = 0, 1, 2, ...,$ where λ is some positive value. Find (a) $P(X = 0)$ and (b) $P(X > 2)$.

Answer. Since $\sum_{i=0}^{\infty} p(i) = 1$, we have

$$c \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = 1$$

which, using $e^x = \sum_{i=0}^{\infty} x^i/i!$, implies that

$$ce^{\lambda} = 1 \text{ or } c = e^{-\lambda}$$

Thus, we have

1. $P(X = 0) = e^{-\lambda}\lambda^0/0! = e^{-\lambda}$.
2. $P(X > 2) = 1 - P(X \leq 2) = 1 - e^{-\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2 e^{-\lambda}}{2}$

□

The cumulative distribution function F can be expressed in terms of $p(a)$ by

$$F(a) = \sum_{\text{all } x \leq a} p(x)$$

If X is a discrete random variable whose possible values are $x_1, x_2, ...,$ where $x_1 < x_2 < x_3 < ...,$ then the distribution function F of X is a step function. That is, the value of F is constant in the intervals (x_{i-1}, x_i) and then takes a step (or jump) of size $p(x_i)$ at x_i. For instance, if X has a probability mass function given by

$$p(1) = \frac{1}{4}, p(2) = \frac{1}{2}, p(3) = \frac{1}{8}, p(4) = \frac{1}{8}$$

then its cumulative distribution function is

$$F(a) = \begin{cases} 0 & a < 1 \\ \frac{1}{4} & 1 \leq a < 2 \\ \frac{3}{4} & 2 \leq a < 3 \\ \frac{7}{8} & 3 \leq a < 4 \\ 1 & 4 \leq a \end{cases}$$

4.3 Expected Value

One of the most important concepts in probability theory is that of the expectation of a random variable. If X is a discrete random variable having a probability mass function $p(x)$, or the expected value, of X, denoted by $E[X]$, is defined by

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

The expected value of X is a weighted average of the possible values that X can take on, each value being weighted by the probability that X assumes it.

Example 4.3.1. Find $E[X]$, where X is the outcome when we roll a fair die.

Answer. Since $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = 1/6$, we obtain

$$E[X] = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = 7/2$$

□

4.4 Expectation of a Function of a Random Variable

Suppose that we are given discrete random variable along with its probability mass function and we want to compute expected value of some function of X, say, $g(X)$. We can determine $E[g(X)]$ by using definition of expected value.

Example 4.4.1. Let X denote a random variable that takes on any of the values -1, 0, and 1 with respective probabilities

$$P(X = -1) = 0.2, P(X = 0) = 0.5, P(X = 1) = 0.3$$

Compute $E[X^2]$.

Answer. Let $Y = X^2$. Then the probability mass function of Y is given by

$$P(Y = 1) = P(X = -1) + P(X = 1) = 0.5$$
$$P(Y = 0) = P(X = 0) = 0.5$$

Hence,

$$E[X^2] = E[Y] = 1(0.5) + 0(0.5) = 0.5$$

□

Proposition 4.4.2. *If X is a discrete random variable that takes on one of the values x_i, $i \geq 1$, with respective probabilities $p(x_i)$, then, for any real-valued function g,*

$$E[g(X)] = \sum_i g(x_i)p(x_i)$$

Proof. Please see text [1] Page 122 for proof. □

Proposition 4.4.3. *If a and b are constants, then*

$$E[aX + b] = aE[X] + b$$

Proof. We prove

$$\begin{aligned} E[aX+b] &= \sum_{x:p(x)>0} (ax+b)p(x) \\ &= a\sum_{x:p(x)>0} xp(x) + b\sum_{x:p(x)>0} p(x) \\ &= aE[X]+b \end{aligned}$$

□

The expected value of a random variable X, $E[X]$, is also referred to as the mean or the first moment of X. The quantity $E[X^n]$, $n \geq 1$, is called the nth moment of X. By Proposition 4.1, we note that

$$E[X^n] = \sum_{x:p(x)>0} x^n p(x)$$

4.5 Variance

Besides expectation, it is also imporant to measure the variation.

Definition 4.5.1. If X is a random variable with mean μ, then the variance of X, denoted by Var(X), is defined by

$$\text{Var}(X) = E[(X-\mu)^2]$$

Alternatively, one can derive

$$\begin{aligned} \text{Var}(X) &= E[(X-\mu)^2] \\ &= \sum_x (x-\mu)^2 p(x) \\ &= \sum_x (x^2 - 2\mu x + \mu^2)p(x) \\ &= \sum_x x^2 p(x) - 2\mu \sum_x xp(x) + \mu^2 \sum_x p(x) \\ &= E[X^2] - 2\mu^2 + \mu^2 \\ &= E[X^2] - \mu^2 \end{aligned}$$

Example 4.5.2. Calculate Var(X) if X represents the outcome when a fair die is rolled.

Answer. You can easily find $E[X] = \frac{7}{2}$. Now, we find

$$\begin{aligned} E[X^2] &= 1^2(1/6) + 2^2(1/6) + 3^2(1/6) + 4^2(1/6) + 5^2(1/6) + 6^2(1/6) \\ &= (91)(1/6) \end{aligned}$$

and thus we have variance

$$\text{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

□

A useful identity is that for any constants a and B,

$$\text{Var}(aX + b) = a^2\text{Var}(X)$$

To prove this equality, let $\mu = E[X]$ and note that $E[aX + b] = a\mu + b$. Therefore, we have

$$\begin{aligned}\text{Var}(aX + b) &= E[(aX + b - a\mu - b)^2] \\ &= E[a^2(X - \mu)^2] \\ &= a^2E[(X - \mu)^2] \\ &= a^2\text{Var}(X)\end{aligned}$$

Remark 4.5.3. Please note the following.

1. Analogous to the means being the center of gravity of a distribution of mass, the variance represents, in the terminology of mechanics, the moment of inertia.
2. The square root of the $\text{Var}(X)$ is called the standard deviation of X, and we denote it by $\text{SD}(X)$. That is,

$$\text{SD}(X) = \sqrt{\text{Var}(X)}$$

Discrete random variables are often classified according to their probability mass functions. In the future, we may deal with probability distribution function (or probability density function) for continuous random variables.

4.6 The Bernoulli and Binomial Random Variables

Suppose that a trial, or an experiment, whose outcome can be classified as either a success or a failure is performed. If we let $X = 1$ when the outcome is a success and $X = 0$ when it is a failure, then the probability mass function of X is given by

$$\begin{aligned}p(0) &= P(X = 0) = 1 - p \\ p(1) &= P(X = 1) = p\end{aligned}$$

where p, $0 \le p \le 1$, is the probability that the trial is a success. A random variable X is said to be a Bernoulli random variable if its probability mass function is given by the above equation for some $p \in (0, 1)$.

Suppose now that n independent trials, each of which results in a success with probability p or in a failure with probability $1 - p$, are to be performed. If X represents the number of successes that occur in the n trials, then X is said to be a binomial random variable with parameters (n, p). Thus, a Bernoulli random variable is just a binomial random variable with parameters $(1, p)$. The probability mass function of a binomial random variable having parameters (n, p) is given by

$$p(i) = \binom{n}{i} p^i(1 - p)^{n-i} \text{ for } i = 0, 1, ..., n$$

Example 4.6.1. It is known that screws produced by a company can be defective with probability 0.01, independently of one another. The company sells the screws in packages of 10 and offers a money-back guarantee that at most 1 of the 10 screws is defective. What proportion of packages sold must the company replace?

Answer. If X is the number of defective screws in a package, then X is a binomial random variable with parameters $(10, 0.01)$. Hence, the probability that a package will have to be replaced is

$$1 - P(X = 0) - P(X = 1) = 1 - \binom{10}{0}(0.01)^0(0.99)^{10} - \binom{10}{1}(0.01)^1(0.99)^9 \equiv 0.004$$

□

Example 4.6.2. Consider a jury trial in which it takes 8 of the 12 jurors to convict the defendant; that is, in order for the defendant to be convicted, at least 8 of the jurors must vote him guilty. If we assume that jurors act independently and that whether or not the defendant is guilty, each makes the right decision with probability θ, what is the probability that the jury renders a correct decision?

Answer. The problem, as stated, is incapable of an actual solution. However, we can work out an expression to model this environment. The situation is binary and we have either guilty or not guilty. The former requires 8 votes and the latter requires 5. Hence, we have

$$\text{if he is guilty: } \sum_{i=8}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i}$$
$$\text{if he is not Guilty: } \sum_{i=5}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i}$$

and hence, by letting probability that the defendant is guilty to be α, we can write out the expression for rendering a correct decision

$$\alpha \sum_{i=8}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i} + (1-\alpha) \sum_{i=5}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i}$$

□

We can examine the properties of a binomial random variable with parameters n and p. To begin, let us compute its expected value and variance. To begin, note that

$$E[X^k] = \sum_{i=0}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i}$$
$$= \sum_{i=1}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i}$$

Using the identity

$$i \binom{n}{i} = n \binom{n-1}{i-1}$$

gives

$$\begin{aligned}
E[X^k] &= np \sum_{i=1}^{n} i^{k-1} \binom{n-1}{i-1} p^{i-1}(1-p)^{n-i} \\
&= np \sum_{j=0}^{n-1} (j+1)^{k-1} \binom{n-1}{j} p^j (1-p)^{n-1-j}, \text{ let } j = i-1 \\
&= npE[(Y+1)^{k-1}]
\end{aligned}$$

where Y is a binomial random variable with parameters $n-1$, p. Setting $k=1$, we would arrive

$$E[X] = np$$

which gives us the expected number of successes that occur in n independent trials when each is a success with probability p. Setting $k=2$, we yield

$$\begin{aligned}
E[X^2] &= npE[Y+1] \\
&= np[(n-1)p+1]
\end{aligned}$$

Since $E[X] = np$, we obtain

$$\begin{aligned}
E[X] &= np \\
\text{Var}(X) &= np(1-p)
\end{aligned}$$

4.7 Poisson random Variable

IMPORTANT A random variable X that takes on one of the values 0, 1, 2, ...is said to be Poisson random variable with parameter λ if, for some $\lambda > 0$,

$$p(i) = P(X = i) = e^{-\lambda}\frac{\lambda^i}{i!} \text{ for } i = 0, 1, 2, \dots$$

and it has property $\lambda = np$.

We can check the following:

$$\sum_{i=0}^{\infty} p(i) = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{-\lambda}e^{\lambda} = 1$$

Some general applications that obey Poisson probability are

1. The number of misprints on a page (or a group of pages) of a book.
2. The number of people in a community who survive to age 100.
3. The number of wrong telephone numbers that are dialed in a day.
4. The number of packages of a dog biscuits sold in a particular store each day
5. The number of vacancies occurring during a year in the federal judicial system.
6. The number of α-particles discharged in a fixed period of time from some radioactive material.

Example 4.7.1. Suppose number of typographical errors on a single page of this book has a Poisson distribution with parameter $\lambda = \frac{1}{2}$. Calculate the probability that there is at least one error on your page.

Answer. Letting X denote the number of error on a page, we have

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1/2} = 0.393$$

□

Example 4.7.2. Suppose that the probability that an item produced by a certain machine will be defective is 0.1. Find the probability that a sample of 10 items will contain at most 1 defective item.

Answer. The desired probability is

$$\binom{10}{0}(0.1)^0(0.9)^{10} + \binom{10}{1}(0.1)^1(0.9)^9 = 0.74$$

while Poisson approximation would yield similar results. □

5 Continuous Random Variables

Go back to Table of Contents. Please click TOC

The previous chapter discussed discrete random variables, e.g. the random variable whose set of possible values is either finite or countably infinite. There also exist random variables whose set of possible values to be uncountable. Consider X be such a random variable. We say that X is continuous random variable if there exists a nonnegative function f, defined for all real $x \in (-\infty, \infty)$, having the property that for any set B of real numbers,

$$P(X \in B) = \int_B f(x)dx$$

The function f is called the probability density function of the random variable X.

In words, the above equation states that the probability that X will be in B may be obtained by integrating the probability density function over the set B. Since X must assume some value, f must satisfy

$$1 = P(X \in (-\infty, \infty)) = \int_{-\infty}^{\infty} f(x)dx$$

All probability statements about X can be answered in terms of f.

Example 5.0.1. Letting $B = [a, b]$, we obtain

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

Example 5.0.2. IMPORTANT Suppose X is a continuous random variable whose probability density function is

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{else} \end{cases}$$

1. What is the value of C?
2. Find $P(X > 1)$.

Answer. We have the following

1. Since f is a probability density function, we must have
$$\int_{-\infty}^{\infty} f(x)dx = 1$$
and we can solve $C\int_0^2 (4x - 2x^2)dx = 1$. After integration, we have $C(2x^2 - \frac{2x^3}{3})|_{x=0}^2 = 1$ and we have result $C = \frac{3}{8}$.
2. $P(X > 1) = \int_1^{\infty} f(x)dx = \frac{3}{8}\int_1^2 (4x - 2x^2)dx = \frac{1}{2}$.

□

Example 5.0.3. The amount of time in hours that a computer functions before breaking down is a continuous random variable with probability density function given by

$$f(x) = \begin{cases} \lambda e^{-x/100} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

What is the probability that

1. a computer will function between 50 and 150 hours before breaking down?
2. it will function for fewer than 100 hours?

Answer. We solve the parts accordingly

1. Since $1 = \int_{-\infty}^{\infty} f(x)dx = \lambda \int_0^{\infty} e^{-x/100}dx$, we can take integral and obtain $1 = -\lambda(100)e^{-x/100}|_0^{\infty} = 100\lambda$. We can solve for $\lambda = \frac{1}{\lambda}$. Then we can proceed to find the probability

$$\begin{aligned} P(50 < X < 1500) &= \int_{50}^{150} \frac{1}{100}e^{-x/100}dx \\ &= -e^{-x/100}|_{50}^{150} \\ &= e^{-1/2} - e^{-3/2} \\ &= 0.383 \end{aligned}$$

2. I will leave this to you as an exercise.

□

5.1 Expectation and Variance of Continuous Random Variables

In discrete senses, we defined the expected value of a discrete random variable X by

$$E[X] = \sum_x xP(X = x)$$

If X is a continuous random variable having probability density function $f(x)$, then because

$$f(x) \approx P(x \leq X \leq x + dx) \text{ for } dx \text{ small}$$

it is easy to see that the analogous definition is to define the expected value of X by

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

Example 5.1.1. Find $\mathrm{E}(X)$ when the density function of X is

$$f(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{else} \end{cases}$$

Answer. Solve the following

$$\begin{aligned} \mathrm{E}(X) &= \int xf(x)dx \\ &= \int_0^1 2x^2 \\ &= \frac{2}{3} \end{aligned}$$

□

Proposition 5.1.2. *If X is a continuous random variable with probability density function $f(x)$, then, for any real-valued function,*

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

Example 5.1.3. The density function of X is given by

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

Find $E[e^X]$.

Answer. Let $Y = e^X$. We start by determining F_Y, the cumulative distribution function of Y. Now, for $1 \leq x \leq e$,

$$\begin{aligned} F_Y(x) &= P(Y \leq x) \\ &= P(e^X \leq x) \\ &= P(X \leq \log(x)) \\ &= \int_0^{\log(x)} f(y)dy \\ &= \log(x) \end{aligned}$$

By differentiating $F_Y(x)$, we can conclude that the probability density function of Y is given by

$$f_Y(x) = \frac{1}{x} \text{ for } 1 \leq x \leq e$$

Hence,

$$\begin{aligned} E[e^X] = E[Y] &= \int_{-\infty}^{\infty} x f_Y(x)dx \\ &= \int_1^e dx \\ &= e - 1 \end{aligned}$$

□

Lemma 5.1.4. *For a nonnegative random variable Y,*

$$E[Y] = \int_0^{\infty} P(Y > y)dy$$

Lemma 5.1.5. *If a and b are constants, then*

$$E[aX + b] = aE[X] + b$$

The variance of a continuous random variable is defined exactly as it is for a discrete random variable, namely, if X is a random variable with expected value μ, then the variance of X is defined (for any type of random variable) by

$$\text{Var}(X) = E[(X - \mu)^2]$$

The alternative formula,

$$\text{Var}(X) = E[X^2] - (E[X])^2$$

Example 5.1.6. Recall the example above,

$$f(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{else} \end{cases}$$

Find $\text{Var}(X)$ for this random variable X.

Answer. First, we compute the second moment $E[X^2]$,

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 f(x) dx \\ &= \int_0^1 2x^3 dx = \frac{1}{2} \end{aligned}$$

Hence, we obtain

$$\text{Var}(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$$

□

Note that we have the property $\text{Var}(aX + b) = a^2 \text{Var}(X)$.

5.2 Uniform Random Variable

A random variable is said to be uniformly distributed over the interval $(0, 1)$ if its probability density function is given by

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{else} \end{cases}$$

Since this is a density function, the following properties hold (1) $f(x) \geq 0$ and (2) $\int_x f(x)dx = 1$.

In general, we say that X is a uniform random variable on the interval (α, β) if the probability density function of X is given by

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \alpha < x < \beta \\ 0 & \text{else} \end{cases}$$

Since $F(a) = \int_{-\infty}^a f(x)dx$, it follows that

$$f(x) = \begin{cases} 0 & a \leq \alpha \\ \frac{1}{\beta - \alpha} & \text{if } \alpha < x < \beta \\ 1 & a \geq \beta \end{cases}$$

Example 5.2.1. Let X be uniformly distributed over (α, β). Find (a) $E[X]$ and (b) $\text{Var}(X)$.

Answer. We proceed accordingly

1. Compute

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_\alpha^\beta \frac{x}{\beta - \alpha} dx \\ &= \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} \\ &= \frac{\beta + \alpha}{2} \end{aligned}$$

2. To find Var(X), first calculate $E[X^2]$.

$$\begin{aligned} E[X^2] &= \int_\alpha^\beta \frac{1}{\beta-\alpha}x^2 dx \\ &= \frac{\beta^3-\alpha^3}{3(\beta-\alpha)} \\ &= \frac{\beta^2+\alpha\beta+\alpha^2}{3} \end{aligned}$$

Hence,

$$\text{Var}(X) = \frac{\beta^2+\alpha\beta+\alpha^2}{3} - \frac{(\alpha+\beta)^2}{4} = \frac{(\beta-\alpha)^2}{12}$$

□

Example 5.2.2. If X is uniformly distributed over $(0, 10)$, calculate the probability that $X < 3$.

Answer. Compute $P(X < 3) = \int_0^3 \frac{1}{10}dx = \frac{3}{10}$. □

Example 5.2.3. IMPORTANT Buses arrive at a specified stop at 15-minute intervals starting at 7AM. That is, they arrive at 7, 7:15, 7:30, 7:45, and so on. If a passenger arrives at the stop at a time that is uniformly distributed between 7 and 7:30, find the probability that he waits

1. less than 5 minutes for a bus;
2. more than 10 minutes for a bus.

Answer. We proceed accordingly

1. Compute

$$\begin{aligned} P(10 < X < 15) + P(25 < X < 30) &= \int_{10}^{15} \frac{1}{30}dx + \int_{25}^{30} \frac{1}{30}dx \\ &= \frac{1}{30} \end{aligned}$$

2. Compute

$$P(0 < X < 5) + P(15 < X < 20) = \frac{1}{3}$$

□

5.3 Normal Random Variables

We say that X is a normal random variable, or simply that X is normally distributed, with parameters μ and σ^2 if the density of X is given by

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

for $-\infty < x < \infty$. The density function is a bell-shaped curve that is symmetric about μ.

Example 5.3.1. Find $E[X]$ and Var(X) when X is a normal random variable with parameters μ and σ^2.

Answer. Let us start by finding the mean and variance of the standard normal random variable $Z = (X - \mu)/\sigma$. We have

$$\begin{aligned} E[Z] &= \int_{-\infty}^{\infty} x f_Z(x) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} dx \\ &= -\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \Big|_{-\infty}^{\infty} \\ &= 0 \end{aligned}$$

Thus,

$$\begin{aligned} \text{Var}(Z) &= E[Z^2] \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx, \text{ IBP: let } \mu = x \text{ and } dv = xe^{-x^2/2} \\ &= \frac{1}{\sqrt{2\pi}} \left(-xe^{-x^2/2} \Big|_{-\infty}^{\infty} + \underbrace{\int_{-\infty}^{\infty} e^{-x^2/2} dx}_{=1} \right) \\ &= 1 \end{aligned}$$

Because $X = \mu + \sigma Z$, the preceding yields the results

$$E[X] = \mu + \sigma E[Z] = \mu$$

and

$$\text{Var}(X) = \sigma^2 \text{Var}(Z) = \sigma^2$$

□

Conventionally, we denote the cumulative distribution function of a standard normal random variable by $\Phi(x)$. That is,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy$$

and we have table

Table 5.1 Area $\Phi(x)$ Under the Standard Normal Curve to the Left of X.

X	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Figure 2: Area $\Phi(x)$ from page 190 in [1]

Example 5.3.2. IMPORTANT If X is a normal random variable with parameters $\mu = 3$ and $\sigma^2 = 9$, find (a) $P(2 < X < 5)$, (b) $P(X > 0)$, and (c) $P(|X - 3| > 6)$.

Answer. We proceed accordingly

1. Compute

$$\begin{aligned} P(2 < X < 5) &= P\left(\frac{2-3}{3} < \frac{X-3}{3} < \frac{5-3}{3}\right) \\ &= \Phi(\frac{2}{3}) - \Phi(-\frac{1}{3}) \\ &= 0.3779 \end{aligned}$$

2. Compute

$$\begin{aligned} P(X > 0) &= P\left(\frac{X-3}{3} > \frac{0-3}{3}\right) \\ &= P(Z > -1) \\ &= \Phi(1) = 0.8413 \end{aligned}$$

3. Compute

$$\begin{aligned} P(|X-3| > 6) &= P(X > 9) + P(X < -3) \\ &= P\left(\frac{X-3}{3} > \frac{9-3}{3}\right) + P\left(\frac{X-3}{3} < \frac{-3-3}{3}\right) \\ &= P(X > 2) + P(Z < -2) \\ &= 0.0456 \end{aligned}$$

$\square$

Example 5.3.3. An expert witness in a paternity suit testifies that the length (in days) of human gestation is approximately normally distributed with parameters $\mu = 270$ and $\sigma^2 = 100$. The defendant in the suit is able to prove that he was out of the country during a period that begun 290 days before the birth of the child and ended 240 days before the birth. If the defendant was, in fact, the father of the child, what is the probability that the mother could have had the very long or very short gestation indicated by the testimony?

Answer. Let X denote the length of the gestation, and assume that the defendant is the father. Then the probability that the birth could occur within the indicated period is

$$\begin{aligned} P(X > 290 \text{ or } X < 240) &= P(X > 290) + P(X < 240) \\ &= P(\frac{X-270}{10} > 2) + P(\frac{X-270}{10} < -3) \\ &= 1 - \Phi(2) + 1 - \Phi(3) \\ &= 0.0241 \end{aligned}$$

$\square$

Remark 5.3.4. Please be aware that the problem can ask you “or” instead of ”and”. In that case, the properties we learned from set theory follow. You should check the intersection between the two events accordingly.

Example 5.3.5. If X, the gain from an investment, is a normal random variable with mean μ and variance σ^2, then because the loss is equal to the negative of the gain, the VAR of such an investment is that value ν such that

$$0.01 = P(-X > \nu)$$

We compute the following

$$\begin{aligned} 0.01 &= P\left(\frac{-X+\mu}{\sigma} > \frac{\nu+\mu}{\sigma}\right) \\ &= 1 - \Phi\left(\frac{\nu+\mu}{\sigma}\right) \end{aligned}$$

and from table we know $\Phi(2.33) = 0.99$ so we know $\frac{\nu+\mu}{\sigma} = 2.33$. That is, $\nu = \text{VAR} = 2.33\sigma - \mu$. Consequently, among set of investments all of whose gains are normally distributed, the investment having the smallest VAR is the one having the largest value of $\mu - 2.33\sigma$.

Theorem 5.3.6. *The DeMoivre-Laplace Theorem. If S_n denotes the number of successes that occur when n independent trials, each resulting in a success of probability p, are performed, then, for any $a < b$,*

$$P\left(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right) \to \Phi(b) - \Phi(a)$$

Example 5.3.7. Let X be the number of times that a fair coin that is flipped 40 times lands on heads. Find the probability that $X = 20$. Use the normal approximation and then compare it with the exact solution.

Answer. To employ normal approximation, note that because the binomial is a discrete integer-valued random variable, whereas the normal is a continuous random variable, it is best to write $P(X = i)$ as $P(i - 1/2 < X < i + 1/2)$ before applying the normal approximation (this is called the continuity correction). Hence, we compute

$$\begin{aligned} P(X = 20) &= P(19.5 < X < 20.5) \\ &= P\left(\frac{19.5-20}{\sqrt{10}} < \frac{X-20}{\sqrt{10}} < \frac{20.5-20}{\sqrt{10}}\right) \\ &= P\left(-1.6 < \frac{X-20}{\sqrt{10}} < 1.6\right) \\ &= \Phi(0.16) - \Phi(-0.16) \\ &= 0.1272 \end{aligned}$$

□

5.4 Exponential Random Variable

A continuous random variable whose probability density function is given, for some $\lambda > 0$, by

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

is said to be an exponential random variable with parameter λ. The cumulative distribution $F(a)$ of an exponential random variable is given by

$$\begin{aligned} F(a) &= P(X \leq a) \\ &= \int_0^a \lambda e^{-\lambda x} dx \\ &= -e^{-\lambda x}|_0^a \\ &= 1 - e^{-\lambda a} \text{ for } a \geq 0 \end{aligned}$$

Note that $F(\infty) = \int_0^\infty \lambda e^{-\lambda x} = 1$.

Example 5.4.1. IMPORTANT Let X be an exponential random variable with parameter λ. Calculate (a) $E[X]$ and (b) Var(X).

Answer. We solve the following accordingly

1. We use $E[X^n] = \int_0^\infty x^n \lambda e^{-\lambda} dx$. Integrating by parts (with $\lambda e^{-\lambda x} = dv$ and $\mu = x^n$) yields

$$\begin{aligned} E[X^n] &= -x^n e^{-\lambda x}|_0^\infty + \int_0^\infty e^{-\lambda x} n x^{n-1} dx \\ &= 0 + \frac{n}{\lambda} \int_0^\infty \lambda e^{-\lambda x} x^{n-1} dx \\ &= \frac{n}{\lambda} E[X^{n-1}] \end{aligned}$$

   Letting $n = 1$ and $n = 2$ gives us

$$E[X] = \frac{1}{2} \text{ and } E[X^2] = \frac{2}{\lambda} E[X] = \frac{2}{\lambda^2}$$

2. We have variance

$$\text{Var}(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

□

Example 5.4.2. Suppose that the number of miles that a car can run before its battery wears out is exponentially distributed with an average value of 10,000 miles. If a person desires to take a 5000-mile trip, what is the probability that he or she will be able to complete the trip without having to replace the car battery? What can be said when the distribution is not exponential? (Assume the parameter $\lambda = 1/10$).

Answer. The desired probability is

$$P(\text{remaining lifetime} > 5) = 1 - F(5) = e^{-5\lambda} = 0.607$$

However, if the lifetime distribution F is not exponential, then the relevant probability is

$$P(\text{lifetime} > t + 5 | \text{lifetime} > t) = \frac{1 - F(t+5)}{1 - F(t)}$$

where t is the number of miles that the battery had been in use prior to the start of the trip. Therefore, if the distribution is not exponential, additional information is needed (namely, the value of t) before the desired probability can be calculated. □

6 Jointly Distributed Random Variables

Go back to Table of Contents. Please click TOC

6.1 Joint Distribution Functions

We understand from above sections with probability distributions for single random variable. However, we are often interested in probability statements concerning two or more random variables. In order to deal with such probabilities, we define for any two random variables X and Y, the joint cumulative probability distribution function of X and Y by

$$F(a,b) = P(X \leq a, Y \leq b) \text{ for } -\infty < a, b < \infty$$

The distribution of X can be obtained from the joint distribution of X and Y as follows

$$\begin{aligned} F_X(a) &= P(X \leq a) \\ &= P(X \leq a, Y < \infty) \\ &= P(\lim_{b\to\infty} \{X \leq a, Y \leq b\}) \\ &= \lim_{b\to\infty} P(\{X \leq a, Y \leq b\}) \\ &= \lim_{b\to\infty} F(a,b) \\ &= F(a, \infty) \end{aligned}$$

Note that the preceding set of equalities, we have once again made use of the fact that probability is a continuous set function. Similarly, the cumulative distribution function of Y is given by

$$\begin{aligned} F_Y(b) &= P(Y \leq b) \\ &= \lim_{a\to\infty} F(a,b) \\ &= F(\infty, b) \end{aligned}$$

In the case when X and Y are both discrete random variables, it is convenient to define the joint probability mass function of X and Y by

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

the probability mass function of X can be obtained from $p(x,y)$ by

$$\begin{aligned} p_X(x) &= P(X = x) \\ &= \sum_{x:p(x,y)>0} p(x,y) \end{aligned}$$

Similarly, we have

$$p_Y(y) = \sum_{x:p(x,y)>0} p(x,y)$$

We say that X and Y are jointly continuous if there exists a function $f(x,y)$, defined for all real x and y, having the property that for every set C of pairs of real numbers (that is, C is a set in the two-dimensional plane),

$$P((X,Y) \in C) = \iint_{(x,y)\in C} f(x,y)dxdy$$

Example 6.1.1. IMPORTANT The joint density function of X and Y is given by

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

Compute (a) $P(X > 1, Y < 1)$, (b) $P(X < Y)$, and (c) $P(X < a)$.

Answer. Please refer to the following

- Compute

$$\begin{aligned} P(X > 1, Y < 1) &= \int_0^1 \int_1^\infty 2e^{-x}e^{-2y}dxdy \\ &= \int_0^1 2e^{-2y}\left(-e^{-x}|_1^\infty\right)dy \\ &= e^{-1}\int_0^1 2e^{-2y}dy \\ &= e^{-1}(1-e^{-2}) \end{aligned}$$

- Compute

$$\begin{aligned} P(X < Y) &= \iint_{(x,y):x<y} 2e^{-x}e^{-2y}dxdy \\ &= \int_0^\infty 2e^{-2y}(1-e^{-y})dy \\ &= \int_0^\infty 2e^{-2y}dy - \int_0^\infty 2e^{-3y}dy \\ &= 1 - \frac{2}{3} \\ &= \frac{1}{3} \end{aligned}$$

- Compute

$$\begin{aligned} P(X < a) &= \int_0^a \int_0^\infty 2e^{-2y}e^{-x}dydx \\ &= \int_0^a e^{-x}dx \end{aligned}$$

□

Example 6.1.2. The joint density of X and Y is given by

$$f(x,y) = \begin{cases} e^{-(x+y)} & \text{if } 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Find the density function of the random variable X/Y.

Answer. Start by computing the distribution function of X/Y. For $a > 0$,

$$\begin{aligned} F_{X/Y}(a) &= P\left(\frac{X}{Y} \le a\right) \\ &= \iint_{x/y \le a} e^{-(x+y)} dx dy \\ &= \int_0^\infty \int_0^{ay} e^{-(x+y)} dx dy \\ &= \int_0^\infty (1 - e^{-ay}) e^{-y} dy \\ &= \left\{ -e^{-y} + \frac{e^{-(a+1)y}}{a+1} \right\}\Bigg|_0^\infty \\ &= 1 - \frac{1}{a+1} \end{aligned}$$

Differentiation shows the density function of X/Y is given by $f_{X/Y}(a) = 1/(a+1)^2$ for $0 < a < \infty$. □

6.2 Independent Random Variables

The random variables X and Y are said to be independent if, for any two sets of real numbers A and B,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

In other words, X and Y are independent if, for all A and B, the events $E_A = \{X \in A\}$ and $F_B = \{Y \in B\}$ are independent.

It can be shown by using the three axioms of probability that the above equation will follow if and only if, for all a, b,

$$P(X \le a, Y \le b) = P(X \le a)P(Y \le b)$$

Hence, in terms of the joint distribution function F of X and Y, X and Y are independent if

$$F(a, b) = F_X(a)F_Y(b) \text{ for all } a, b$$

Proposition 6.2.1. *The continuous (discrete) random variables X and Y are independent if and only if their joint probability density (mass) function can be expressed as*

$$f_{X,Y}(x, y) = h(x)g(y), -\infty < x < \infty, -\infty < y < \infty$$

Answer. Let us give the proof in the continuous case. First, note that independence implies that the joint density is the product of the marginal densities of X and Y, so the preceding factorization will hold when the random variables are independent. Now, suppose that

$$f_{X,Y}(x, y) = h(x)g(y)$$

Then

$$\begin{aligned} 1 &= \int_{-\infty}^\infty \int_{-\infty}^\infty f_{X,Y}(x, y) dx dy \\ &= \int_{-\infty}^\infty h(x) dx \int_{-\infty}^\infty dy \\ &= C_1 C_2 \end{aligned}$$

where $C_1 = \int_{-\infty}^{\infty} h(x)dx$ and $C_2 = \int_{-\infty}^{\infty} g(y)dy$. Also,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = C_2 h(x)$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = C_1 g(y)$$

Since $C_1 C_2 = 1$, it follows that

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

□

Example 6.2.2. IMPORTANT Let X, Y, Z be independent and uniformly distributed over $(0,1)$. Compute $P(X \geq YZ)$.

Answer. Since

$$f_{X,Y,Z}(x,y,z) = f_X(x) f_Y(y) f_Z(z) = 1, 0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq z \leq 1$$

we have

$$\begin{aligned} P(X \geq YZ) &= \iiint_{x \geq yz} f_{X,Y,Z}(x,y,z)dxdydz \\ &= \int_0^1 \int_0^1 \int_{yz}^1 dxdydz \\ &= \int_0^1 \int_0^1 (1 - yz)dydz \\ &= \int_0^1 (1 - \frac{z}{2})dz \\ &= \frac{3}{4} \end{aligned}$$

□

6.3 Sums of Independent Random Variables

It is often important to be able to calculate the distribution of $X + Y$ from the distributions of X and Y when X and Y are independent. Suppose that X and Y are independent, continuous random variables having probability density functions f_X and

f_Y. The cumulative distribution function of $X+Y$ is obtained as follows:

$$\begin{aligned} F_{X+Y}(a) &= P(X+Y \leq a) \\ &= \iint_{x+y\leq a} f_X(x) f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x) f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x) dx f_Y(y) dy \\ &= \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy \end{aligned}$$

The cumulative distribution function F_{X+Y} is called convolution of the distributions F_X and F_Y (the cumulative distribution functions of X and Y, respectively.

By differentiating the above equation, we find that the probability density function f_{X+Y} of $X+Y$ is given by

$$\begin{aligned} f_{X+Y}(a) &= \frac{d}{da}\int_{-\infty}^{\infty} F_X(a-y) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \frac{d}{da} F_X(a-y) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy \end{aligned}$$

Let us explore the relationship of two random variables. Recall gamma random variable has a density of the form

$$f(y) = \frac{\lambda e^{-\lambda y}(\lambda y)^{t-1}}{\Gamma(t)}, 0 < y < \infty$$

An important property of this family of distributions is that for a fixed value of λ, it is closed under convolutions.

Proposition 6.3.1. *If X and Y are independent gamma random variables with respective parameters (s, λ) and (t, λ), then $X+Y$ is a gamma random variable with parameters $(s+t, \lambda)$.*

$$\begin{aligned} f_{X+Y}(a) &= \frac{1}{\Gamma(s)\Gamma(t)}\int_0^a \lambda e^{-\lambda(a-y)}[\lambda(a-y)]^{t-1}\lambda e^{-\lambda y}(\lambda y)^{t-1} dy \\ &= K e^{-\lambda a}\int_0^a (a-y)^{s-1} y^{t-1} dy \\ &= K e^{-\lambda a} a^{s+t-1}\int_0^1 (1-x)^{s-1} x^{t-1} dx, \text{ by letting } x = \frac{y}{a} \\ &= C e^{-\lambda a} a^{s+t-1} \end{aligned}$$

where C is a constant that does not depend on a. But, as the preceding is a density function and thus must integrate to 1, the value of C is determined, and we have

$$f_{X+Y}(a) = \frac{\lambda e^{-\lambda a}(\lambda a)^{s+t-1}}{\Gamma(s+t)}$$

Hence, the result is proved.

Proposition 6.3.2. *If X_i, $i = 1, ..., n$, are independent random variables that are normally distributed with respective parameters μ_i, σ_i^2, $i = 1, ..., n$, then $\sum_{i=1}^n X_i$ is normally distributed with parameters $\sum_{i=1}^n \mu_i$ and $\sum_{i=1}^n \sigma_i^2$.*

Proposition 6.3.3. *If X and Y are independent Poisson random variables with respective parameters λ_1 and λ_2, compute the distribution of $X + Y$.*

7 Properties of Expectation

Go back to Table of Contents. Please click TOC

7.1 Introduction

This section we develop and exploit additional properties of expected values. Recall expected value of the random variable X

$$E[X] = \sum_x xp(x)$$

where X is a discrete random variable with probability mass function $p(x)$, and by

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

when X is a continuous random variable with probability density function $f(x)$.

7.2 Expectation of Sums of Random Variables

Let us begin by introducing one of the most important properties in expectation of random variables.

Proposition 7.2.1. *If X and Y have a joint probability mass function $p(x, y)$, then*

$$E[g(X, Y)] = \sum_y \sum_x g(x, y)p(x, y)$$

If X and Y have a joint probability density function $f(x, y)$, then

$$E[g(X, Y)] = \int_{-\infty}^{\infty} g(x, y)f(x, y)dxdy$$

Let us prove the above property.

Proof. Suppose we have random variables X and Y that are jointly continuous with joint density function $f(x, y)$ and when $g(X, Y)$ is a nonnegative random variable. Because $g(X, Y) \geq 0$, we have

$$E[g(X, Y)] = \int_0^{\infty} P(g(X, Y) > t)dt$$

We can write

$$P(g(X, Y) > t) = \iint_{(x,y):g(x,y)>t} f(x, y)dydx$$

shows that

$$E[g(X, Y)] = \int_0^{\infty} \iint_{(x,y):g(x,y)>t} f(x, y)dydxdt$$

Interchanging the order of integration gives

$$E[g(X, Y)] = \int_x \int_y \int_{t=0}^{g(x,y)} f(x, y)dtdydx$$
$$= \int_x \int_y g(x, y)f(x, y)dydx$$

Thus, the result is proven when $g(X, Y)$ is a nonnegative random variable. □

An application of such property can be used in the following application.

Example 7.2.2. IMPORTANT An accident occurs at a point X that is uniformly distributed on a road of length L. At the time of the accident, an ambulance is at a location Y that is also uniformly distributed on the road. Assuming that X and Y are independent, find the expected distance between the ambulance and the point of the accident.

Answer. We want to compute $E[|X-y|]$. We have the joint density function of X and Y to be

$$f(x,y) = \frac{1}{L^2}, 0 < x < L, 0 < y < L$$

and it follows from the property above

$$E[|X-Y|] = \frac{1}{L^2}\int_0^L\int_0^L |x-y|dydx$$

Now we can do the math

$$\begin{aligned}\int_0^L |x-y|dy &= \int_0^x (x-y)dy + \int_x^L (y-x)dy \\ &= \frac{x^2}{2} + \frac{L^2}{2} - \frac{x^2}{2} - x(L-x) \\ &= \frac{L^2}{2} + x^2 - xL\end{aligned}$$

Therefore,

$$\begin{aligned}E[|X-Y|] &= \frac{1}{L^2}\int_0^L \left(\frac{L^2}{2} + x^2 - xL\right) dx \\ &= \frac{L}{3}\end{aligned}$$

□

An important application of the above property is the following. Suppose $E[X]$ and $E[Y]$ are both finite and let $g(X,Y) = X+Y$. Then, in the continuous case,

$$\begin{aligned}E[X+Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y)f(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dydx + \int_{-\infty}^{\infty} yf(x,y)dxdy \\ &= E[X]+E[Y]\end{aligned}$$

The same result holds in general; thus, whenever $E[X]$ and $E[Y]$ are finite,

$$E[X+Y] = E[X]+E[Y]$$

Example 7.2.3. Let $X_1, ..., X_n$ be independent and identically distributed random variables having distribution function F and expected value μ. Such as sequence of random variables is said to constitute a sample from the distribution F. The quantity

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

is called the sample mean. Compute $E[\bar{X}]$.

Answer. Compute

$$\begin{aligned}
\mathrm{E}[\bar{X}] &= \mathrm{E}\left[\sum_{i=1}^{n} \frac{X_i}{n}\right] \\
&= \frac{1}{n}\mathrm{E}\left[\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[X_i] \\
&= \mu \text{ since } \mathrm{E}X_i \equiv \mu
\end{aligned}$$

We conclude that the expected value of the sample mean is μ, the mean of the distribution. When the distribution mean μ is unknown, the sample mean if often used in statistics to estimate it. □

7.3 Moments of the Number of Events that Occur

Let us look at an example.

Example 7.3.1. Suppose that there are N distinct types of coupons and that, independently of past types collected, each new one obtained is type j with probability p_j, $\sum_{j=1}^{N} p_j = 1$. Find the expected value and variance of the number of different types of coupons that appear among the first n collected.

Answer. We will find it more convenient to work with the number of uncollected types. Let Y equal the number of types of coupons collected, and let $X = N - Y$ denote the number of uncollected types. With A_i defined as the event that there are no type i coupons in the collection, X is equal to the number of the events $A_1, ..., A_N$ that occur. Because the types of the successive coupons collected are independent, and, with probability $1 - p_i$ each new coupon is not type i, we have

$$P(A_i) = (1 - p_i)^n$$

Hence, $\mathrm{E}[X] = \sum_{i=1}^{N}(1 - p_i)^n$, from which it follows that

$$\mathrm{E}[Y] = N - \mathrm{E}[X] = N - \sum_{i=1}^{N}(1 - p_i)^n$$

Similarly, because each of the n coupons collected is neither a type i nor a type j coupon, with probability $1 - p_i - p_j$, we have

$$\mathrm{P}(A_i A_j) = (1 - p_i - p_j)^n, i = j$$

Thus,

$$\mathrm{E}[X(X-1)] = 2\sum_{i<j} \mathrm{P}(A_i A_j) = 2\sum_{i<j}(1 - p_i - p_j)^n$$

or

$$\mathrm{E}[X^2] = 2\sum_{i<j}(1 - p_i - p_j)^n + \mathrm{E}[X]$$

Hence, we obtain

$$\begin{aligned}\mathrm{var}(Y) &= \mathrm{var}(X) \\ &= E[X^2] - (EX)^2 \\ &= 2\sum_{i<j}(1-p_i-p_j)^n + \sum_{i=1}^{N}(1-p_i)^n - \left(\sum_{i=1}^{N}(1-p_i)^n\right)^2\end{aligned}$$

□

7.4 Covariance, Variance of Sums, and Correlations

The following proposition shows that the expectation of a product of independent random variables is equal to the product of their expectations.

Proposition 7.4.1. *If X and Y are independent, then, for any functions h and g,*

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

Answer. Suppose that X and Y are jointly continuous with joint density $f(x, y)$. Then we have

$$\begin{aligned}E[g(X)h(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dxdy \\ &= \int_{-\infty}^{\infty} h(y)f_Y(y)dy \int_{-\infty}^{\infty} g(x)f_X(x)dx \\ &= E[h(Y)]E[g(X)]\end{aligned}$$

□

Definition 7.4.2. IMPORTANT The covariance between X and Y, denoted by $\mathrm{Cov}(X, Y)$, is defined by

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

Upon expanding the right side of the preceding definition, we see that

$$\begin{aligned}\mathrm{Cov}(X, Y) &= E[XY - E[X]Y - XE[Y] + E[Y]E[X]] \\ &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y]\end{aligned}$$

Proposition 7.4.3. *There are the following properties:*

- $cov(X, Y) = cov(Y, X)$
- $cov(X, X) = var(X)$
- $cov(aX, Y) = aCov(X, Y)$
- $cov(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} cov(X_i, Y_j)$

7.5 Conditional Expectation

If X and Y are jointly discrete random variables, then the conditional probability mass function of X, given that $Y = y$, is defined for all y such that $P(Y = y) > 0$, by

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

It is therefore natural to define, in this case, the conditional expectation of X given that $Y = y$, for all values of y such that $p_Y(y) > 0$, by

$$\begin{aligned} E[X|Y = y] &= \sum_x xP(X = x|Y = y) \\ &= \sum_x xp_{X|Y}(x|y) \end{aligned}$$

7.6 Moment Generating Functions

IMPORTANT The moment generating function $M(t)$ of the random variable X is defined for all real values of t by

$$M(t) = E[e^{tX}] = \begin{cases} \sum_x e^{tx}p(x) & \text{if } X \text{ is discrete with mass function } p(x) \\ \int_{-\infty}^{\infty} e^{tx}f(x)dx & \text{if } X \text{ is continuous with density } f(x) \end{cases}$$

We call $M(t)$ the moment generating function because all of the moments of X can be obtained by successively differentiating $M(t)$ and then evaluating the result at $t = 0$. For example,

$$\begin{aligned} M'(t) &= \frac{d}{dt}E[e^{tX}] \\ &= E[\frac{d}{dt}(e^{tX})] \\ &= E[Xe^{tX}] \end{aligned}$$

where we have assumed that the interchange of the differentiation and expectation operators is legitimate. That is, we have assumed that

$$\frac{d}{dt}\left[\sum_x e^{tx}p(x)\right] = \sum_x \frac{d}{dt}[e^{tx}p(x)]$$

in the discrete case and

$$\frac{d}{dt}\left[\int e^{tx}f(x)dx\right] = \int \frac{d}{dt}[e^{tx}f(x)]dx$$

in the continuous case. This assumption can almost always be justified and, indeed, is valid for all of the distributions considered in this book. Hence, from above the first derivative of moment generating function, evaluated at $t = 0$, we obtain

$$M'(0) = E[X]$$

Similarly,

$$\begin{aligned} M''(t) &= \frac{d}{dt} M'(t) \\ &= \frac{d}{dt} \mathrm{E}[Xe^{tX}] \\ &= \mathrm{E}\left[\frac{d}{dt}(Xe^{tX})\right] \\ &= \mathrm{E}[X^2 e^{tX}] \end{aligned}$$

Thus, we have

$$M''(0) = \mathrm{E}[X^2]$$

In general, the nth derivative of $M(t)$ is given by

$$M^n(t) = \mathrm{E}[X^n e^{tX}], n \geq 1$$

implying that

$$M^n(0) = \mathrm{E}[X^n], n \geq 1$$

Example 7.6.1. IMPORTANT If X is a binomial random variable with parameters n and p, then

$$\begin{aligned} M(t) &= \mathrm{E}[e^{tX}] \\ &= \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k (1-p)^{n-k} \\ &= (pe^t + 1 - p)^n \end{aligned}$$

where the last equality follows from the binomial theorem. Differentiation yields

$$M'(t) = n(pe^t + 1 - p)^{n-1} pe^t$$

Thus, we have

$$\mathrm{E}[X] = M'(0) = np$$

Differentiating a second time yeilds

$$M''(t) = n(n-1)(pe^t + 1 - p)^{n-2}(pe^t)^2 + n(pe^t + 1 - p)^{n-1} pe^t$$

so

$$\mathrm{E}[X^2] = M''(0) = n(n-1)p^2 + np$$

The variance of X is given by

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{E}[X^2] - (\mathrm{E}(X))^2 \\ &= n(n-1)p^2 + np - n^2 p^2 \\ &= np(1-p) \end{aligned}$$

Example 7.6.2. IMPORTANT If X is a Poisson random variable with parameter λ, then

$$\begin{aligned} M(t) &= \mathsf{E}[e^{tX}] \\ &= \sum_{n=0}^{\infty} \frac{e^{tn} e^{-\lambda} e^{-\lambda} \lambda^n}{n!} \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} \\ &= e^{-\lambda} e^{\lambda} e^{t} \\ &= \exp(\lambda(e^t - 1)) \end{aligned}$$

Differentiating yields

$$\begin{aligned} M'(t) &= \lambda e^t \exp(\lambda(e^t - 1)) \\ M''(t) &= (\lambda e^t)^2 \exp(\lambda(e^t - 1)) + \lambda e^t \exp(\lambda(e^t - 1)) \end{aligned}$$

Thus,

$$\begin{aligned} \mathsf{E}[X] &= M'(0) = \lambda \\ \mathsf{E}[X^2] &= M''(0) = \lambda^2 + \lambda \\ \mathrm{var}(X) &= \mathsf{E}[X^2] - (\mathsf{E}(X))^2 \\ &= \lambda \end{aligned}$$

Hence, both the mean and the variance of the Poisson random variable equal λ.

Example 7.6.3. Let us find the first and second moment of exponential distribution withy parameter λ.

$$\begin{aligned} M(t) &= \mathsf{E}[e^{tX}] \\ &= \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx \\ &= \frac{\lambda}{\lambda - t} \text{ for } t < \lambda \end{aligned}$$

We note from this derivation that for the exponential distribution, $M(t)$ is defined only for values of t less than λ. Differentiation of $M(t)$ yields

$$M'(t) = \frac{\lambda}{(\lambda - t)^2}, M''(t) = \frac{2\lambda}{(\lambda - t)^3}$$

Hence,

$$\mathsf{E}[X] = M'(0) = \frac{1}{\lambda}, \mathsf{E}[X^2] = M''(0) = \frac{2}{\lambda^2}$$

The variance of X is given by

$$\begin{aligned} \mathrm{var}(X) &= \mathsf{E}[X^2] - (\mathsf{E}(X))^2 \\ &= \frac{1}{\lambda^2} \end{aligned}$$

Let us summarize the moment generating function in the following table.

Table 7.1 Discrete Probability Distribution.

	Probability mass function, $p(x)$	Moment generating function, $M(t)$	Mean	Variance
Binomial with parameters n, p; $0 \le p \le 1$	$\binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n$	$(pe^t + 1 - p)^n$	np	$np(1-p)$
Poisson with parameter $\lambda > 0$	$e^{-\lambda}\frac{\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$	$\exp\{\lambda(e^t - 1)\}$	λ	λ
Geometric with parameter $0 \le p \le 1$	$p(1-p)^{x-1}$, $x = 1, 2, \ldots$	$\frac{pe^t}{1-(1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative binomial with parameters r, p; $0 \le p \le 1$	$\binom{n-1}{r-1} p^r (1-p)^{n-r}$, $n = r, r+1, \ldots$	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

Table 7.2 Continuous Probability Distribution.

	Probability density function, $f(x)$	Moment generating function, $M(t)$	Mean	Variance
Uniform over (a, b)	$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential with parameter $\lambda > 0$	$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 \end{cases}$	$\frac{\lambda}{\lambda - t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma with parameters $(s, \lambda), \lambda > 0$	$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{s-1}}{\Gamma(s)} & x \ge 0 \\ 0 & x < 0 \end{cases}$	$\left(\frac{\lambda}{\lambda - t}\right)^s$	$\frac{s}{\lambda}$	$\frac{s}{\lambda^2}$
Normal with parameters (μ, σ^2)	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$ $-\infty < x < \infty$	$\exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$	μ	σ^2

Example 7.6.4. IMPORTANT Calculate the distribution of $X + Y$ when X and Y are independent Poisson random variables with means respectively λ_1 and λ_2.

Answer. We compute the following

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) M_Y(t) \\ &= \exp(\lambda_1(e^t - 1)) \exp(\lambda_2(e^t - 1)) \\ &= \exp((\lambda_1 + \lambda_2)(e^t - 1)) \end{aligned}$$

Hence, $X + Y$ is Poisson distributed with mean $\lambda_1 + \lambda_2$. □

It is also possible to define the joint moment generating function of two or more random variables. This is done as follows: for any n random variables $X_1, ..., X_n$, the joint moment generating function, $M(t_1, .l.., t_n)$, is defined, for all real values of $t_1, ..., t_n$, by

$$M(t_1, ..., t_n) = \mathrm{E}[e^{t_1 X_1 + \cdots + t_n X_n}]$$

The individual moment generating functions can be obtained from $M(t_1, ..., t_n)$ by letting all but one of the t_j's be 0. That is,

$$M_{X_i}(t) = \mathrm{E}[e^{tX_i}] = M(0, ..., 0, t, 0, ..., 0)$$

where the t is in the ith place.

It can be proven that the joint moment generating function $M(t_1, ..., t_n)$ uniquely determines the joint distribution of $X_1, ..., X_n$. This result can then be used to prove that the n random variables $X_1, ..., X_n$ are independent if and only if

$$M(t_1, ..., t_n) = M_{X_1}(t_1 \dots M_{X_n}(t_n)$$

For the proof in one direction, if the n random variables are independent, then

$$\begin{aligned} M(t_1, ..., t_n) &= \mathsf{E}[e^{(t_1 X_1 + \cdots + t_n X_n)}] \\ &= \mathsf{E}[e^{t_1 X_1} \dots e^{t_n X_n}] \\ &= \mathsf{E}[e^{t_1 X_1}] \dots \mathsf{E}[e^{t_n} X_n] \text{ by independence} \\ &= M_{X_1}(t_1) \dots M_{X_n}(t_n) \end{aligned}$$

8 Limit Theorems

Go back to Table of Contents. Please click TOC

8.1 Introduction

The most important theoretical results in probability theory are limit theorems. Of these, the most important are those classified either under the heading laws of large numbers of under the heading central limit theorems. Usually, theorems are considered to be laws of large numbers if they are concerned with stating conditions under which the average of a sequence of random variables converges (in some sense) to the expected average.

8.2 Chebyshev's Inequality and the Weak Law of Large Numbers

Let us start with Markov's Inequality. IMPORTANT

Proposition 8.2.1. *If X is a random variable that takes only nonnegative values, then for any value $a > 0$,*

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Proof. For $a > 0$, let

$$I = \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{otherwise} \end{cases}$$

and note that, since $X \geq 0$, then we have $I \leq \frac{X}{a}$. Taking expectations of the preceding inequality yields

$$E[I] \leq \frac{E[X]}{a}$$

which, because $E[I] = P(X \geq a)$, proves the result. □

Please see the following example.

```
# Package
library(quantmod)

# Get Data
getSymbols('FB')
data <- FB
head(data); tail(data)
plot(data[,4], main = "Chart: Stock Price ($)")

# Define Return
head(data[,4]); head(lag(data[,4]))
return <- data[,4]/lag(data[,4]) - 1
summary(return);
plot(return, main = "Chart: Return of Stock Price")
hist(return, breaks = 100, main = "Histogram of Returns")

# Markov's Inequality
a <- 0.02
p <- mean(na.omit(ifelse(return > a, 1, 0))); p
expected.value <- mean(na.omit(return)); expected.value
```

```
expected.value/a

# Summarize in table
Summary <- cbind(
  Probability.of.Event = p,
  Expectation.over.Arbitrary.Value = expected.value/a
); Summary


# Define Function
Markov.Inequality <- function(a = 0.1) {
  # Get Data
  getSymbols('FB')
  data <- FB
  head(data); tail(data)
  plot(data[,4], main = "Chart:␣Stock␣Price␣($)")

  # Define Return
  head(data[,4]); head(lag(data[,4]))
  return <- data[,4]/lag(data[,4]) - 1
  summary(return);
  plot(return, main = "Chart:␣Return␣of␣Stock␣Price")
  hist(return, breaks = 100, main = "Histogram␣of␣Returns")

  # Markov's Inequality
  a <- a
  p <- mean(na.omit(ifelse(return > a, 1, 0))); p
  expected.value <- mean(na.omit(return)); expected.value
  expected.value/a

  # Summarize in table
  Summary <- cbind(
    Probability.of.Event = p,
    Expectation.over.Arbitrary.Value = expected.value/a
  ); Summary

  # Output
  return(Summary)
}

# Run
lapply(c(0.01, 0.05, 0.1, 0.15, 0.2), Markov.Inequality)
```

However, this will not give us correct answer. Who can spot the problem? Please review the following.

Example 8.2.2.

```
# The first line does not satisfy the inequality
# Can anybody spot the mistake?
# Ans:

# Define Function
Markov.Inequality <- function(a = 0.1) {
  # Get Data
```

```
  getSymbols('AAPL')
  data <- AAPL
  head(data); tail(data)
  plot(data[,4], main = "Chart: Stock Price ($)")

  # Define Return
  head(data[,4]); head(lag(data[,4]))
  return <- data[,4]/lag(data[,4]) - 1
  summary(return); plot(return, main = "Chart: Return of Stock Price")
  hist(return, breaks = 100, main = "Histogram of Returns")

  # Markov's Inequality
  a <- a
  number.of.pos.obs <- sum(na.omit(ifelse(return > 0, 1, 0)));
  number.of.pos.obs
  number.of.event <- sum(na.omit(ifelse(return > a, 1, 0)));
  number.of.event
  p <- number.of.event/number.of.pos.obs; p
  expected.value <- mean(na.omit(return[ifelse(return > 0, 1, 0) == 1,]));
  expected.value
  expected.value/a

  # Summarize in table
  Summary <- cbind(
    Value.of.Interest = a,
    Probability.of.Event = p,
    Expectation.over.Arbitrary.Value = expected.value/a
  ); Summary

  # Output
  return(Summary)
}

# Run
lapply(c(0.001, 0.005, 0.01, 0.05, 0.1, 0.15, 0.2), Markov.Inequality)
Report <- matrix(unlist(lapply(c(0.001, 0.005, 0.01, 0.05, 0.1, 0.15, 0.2),
Markov.Inequality)), nrow = 3);
Report <- t(Report); colnames(Report) <- c("Value.of.Interest",
"Prob.of.Event", "Exp.Over.Arbi.Value"); Report
```

Proposition 8.2.3. *IMPORTANT Chebyshev's Inequality. If X is a random variable with finite mean μ and variance σ^2, then for any value $k > 0$,*

$$P((X-\mu)^2 \geq k) \leq \frac{\sigma^2}{k^2}$$

Proof. Since $(X-\mu)^2$ is a nonnegative random variable, we can apply Markov's Inequality (with $a = k^2$) to obtain

$$P((X-\mu)^2 \geq k^2) \leq \frac{E[(X-\mu)^2]}{k^2}$$

But since $(X-\mu)^2 \geq k^2$ if and only if $|X-\mu| \geq k$, than what is written above is equivalent to

$$P(|X-\mu| \geq k) \leq \frac{E[(X-\mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

and we are done. □

```
# Get Data
getSymbols('FB')

# Define Function
Chebyshev.Inequality <- function(k = 0.1) {
  # Check
  if (k < 0) {
    return(
      print(paste(
        "Error Message: Chebyshev Inequality requires k to
        be nonnegative. Please check the value of k."
        )
      )
    )
  } else {
    # Get Data
    data <- AAPL
    head(data); tail(data)
    plot(data[,4], main = "Chart: Stock Price ($)")

    # Define Return
    head(data[,4]); head(lag(data[,4]))
    return <- data[,4]/lag(data[,4]) - 1
    summary(return); plot(return, main = "Chart: Return of Stock Price")
    hist(return, breaks = 100, main = "Histogram of Returns")

    # Markov's Inequality
    mu <- mean(na.omit(return))
    p <- mean(na.omit(as.numeric(abs(return - mu) > k)))
    sigma <- var(na.omit(return))

    # Summarize in table
    Summary <- cbind(
      Value.of.Interest = k,
      Probability.of.Event = p,
      Variance.over.Arbitrary.Value.Square = sigma/k^2
    )

    # Output
    return(Summary)
  }
}

# Run
#lapply(c(0.001, 0.005, 0.01, 0.05, 0.1, 0.15, 0.2), Chebyshev.Inequality)
Report <- matrix(unlist(lapply(c(0.001, 0.005, 0.01, 0.05, 0.1, 0.15, 0.2),
Markov.Inequality)), nrow = 3);
Report <- t(Report); colnames(Report) <- c("Value.of.Interest",
"Prob.of.Event", "Var.over.Arbi.Value.Sq"); Report
```

```
# What happen if k is negative?
Chebyshev.Inequality(-0.1)
```

Example 8.2.4. If X is uniformly distributed over the interval $(0, 10)$, then what is the probability that X and value 5 has distance greater than 4.

Answer. Let us work this out a step at a time.

- First, we compute expectation: $\mathrm{E}(X) = \frac{a+b}{2} = \frac{10}{2} = 5$;
- Second, we compute variance: $\mathrm{E}(X) = \frac{(b-a)^2}{12} = \frac{100}{12} = \frac{25}{3}$
- Compute the result using Chebyshev's Inequality:

$$\mathrm{P}(|X - 5| > 4) = \frac{25/3}{16} = 0.52$$

□

Theorem 8.2.5. *Weak Law of Large Numbers. Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean $\mathrm{E}[X_i] = \mu$. Then, for any $\epsilon > 0$,*

$$\mathrm{P}\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right\} \to 0 \text{ as } n \to \infty$$

Proof. We shall prove this theorem only under the additional assumption that the random variables have a finite variance σ^2. Now, since

$$\mathrm{E}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \mu \text{ and } \mathrm{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

it follows from Chebyshev's Inequality that

$$\mathrm{P}\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right\} \to 0 \text{ as } n \to \infty$$

□

8.3 The Central Limit Theorem

The Central Limit Theorem is one of the most remarkable results in probability theory. Loosely put, it states that the sum of a large number of independent random variables has a distribution that is approximately normal. Hence, it not only provides a simple method for computing approximate probabilities for sums of independent random variables, but also helps explain the remarkable fact that the empirical frequencies of so many natural populations exhibit bell-shaped (that is, normal) curves.

Theorem 8.3.1. *The Central Limit Theorem. Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having mean μ and variance σ^2. Then the distribution of*

$$\frac{X_1 + \ldots X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal as $n \to \infty$. That is, for $-\infty < a < \infty$,

$$\mathrm{P}\left\{\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} dx \text{ as } n \to \infty$$

Lemma 8.3.2. *Let $Z_1, Z_2, \ldots$ be a sequence of random variables having distribution functions F_{Z_n} and moment generating functions M_{Z_n}, $n \geq 1$, and let Z be a random variable having distribution function F_Z and moment generating function M_Z. If $M_{Z_n}(t) \to M_Z(t)$ for all t, then $F_{Z_n}(t) \to F_Z(t)$ for all t at which $F_Z(t)$ is continuous.*

If we let Z be a standard normal random variable, then, since $M_Z(t) = e^{t^2/2}$, it follows from above lemma that if $M_{Z_n}(t) \to e^{t^2/2}$ as $n \to \infty$, then $F_{Z_n}(t) \to \Phi(t)$ as $n \to \infty$.

Now let us produce the following proof.

Proof. Suppose $\mu = 0$ and $\sigma^2 = 1$. We prove the theorem under the assumption that the moment generating function of the X_i, $M(t)$, exists and is finite. Now the moment generating function of $X_i/\sqrt{n}$ is given by

$$\mathbb{E}\left[\exp\left(\frac{tX_i}{\sqrt{n}}\right)\right] = M\left(\frac{t}{\sqrt{n}}\right)$$

Thus, the moment generating function of $\sum_{i=1}^n X_i/\sqrt{n}$ is given by $\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n$. Let

$$L(t) = \log M(t)$$

and note that

$$\begin{aligned} L(0) &= 0 \\ L'(0) &= \frac{M'(0)}{M(0)} \\ &= \mu \\ &= 0 \\ L''(0) &= \frac{M(0)M''(0) - [M'(0)]^2}{[M(0)]^2} \\ &= \mathbb{E}[X^2] \\ &= 1 \end{aligned}$$

Now, to prove the theorem, we must show that $[M(t/\sqrt{n})]^n \to e^{t^2/2}$ as $n \to \infty$, or, equivalently, that $nL(t/\sqrt{n}) \to t^2/2$ as $n \to \infty$. To show this, note that

$$\begin{aligned} \lim_{n\to\infty} \frac{L(t/\sqrt{n})}{n^{-1}} &= \lim_{n\to\infty} \frac{-L'(t/\sqrt{n})n^{-3/2}t}{-2n^{-2}}, \text{ by L'Hopital's Rule} \\ &= \lim_{n\to\infty} \left[\frac{L'(t/\sqrt{n})t}{2n^{-1/2}}\right] \\ &= \lim_{n\to\infty} \left[\frac{-L''(t/\sqrt{n})n^{-3/2}t^2}{-2n^{-3/2}}\right], \text{ again by L'Hopital's Rule} \\ &= \lim_{n\to\infty} \left[L''\left(\frac{t}{\sqrt{n}}\right)\frac{t^2}{2}\right] \\ &= \frac{t^2}{2} \end{aligned}$$

Thus, the central limit theorem is proven when $\mu = 0$ and $\sigma^2 = 1$. The result now follows in the general case by considering the standardized random variables $X_i^* = (X_i - \mu)/\sigma$ and applying the preceding result, since $\mathbb{E}[X_i^*] = 0$, $\text{var}(X_i^*) = 1$. ☐

Theorem 8.3.3. *Central limit Theorem for independent random variables. Let $X_1, X_2, \dots$ be a sequence of independent random variables having respective means and variances $\mu_i = E[X_i]$, $\sigma_i^2 = Var(X_i)$. If (a) the X_i are uniformly bounded – that is, if for some M, $P(|X_i| < M) = 1$ for all i, and (b) $\sum_{i=1}^\infty \sigma_i^2 = \infty$ then we have*

$$P\left\{\frac{\sum_{i=1}^n (X_i - \mu_i)}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \leq a\right\} \to \Phi(a) \to \text{ as } n \to \infty$$

8.4 The Strong Law of Large Numbers

The strong law of large numbers is probably the best-known result in probability theory. It states that the average of a sequence of independent random variables having a common distribution will, with probability 1, converge to the mean of that distribution.

Theorem 8.4.1. *Let $X_1, X_2, \dots$ be a sequence of independent and identically distributed random variables, each having a finite mean $\mu = E[X_i]$. Then, with probability 1,*

$$\frac{X_1 + \cdots + X_n}{n} \to \mu \text{ as } n \to \infty$$

Remark 8.4.2. Here we say converges in probability and what we mean is the following

$$P(\lim_{n\to\infty}(X_1 + \cdots + X_n)/n = \mu) = 1$$

8.5 Other Inequalities

We are sometimes confronted with situations in which we are interested in obtaining an upper bound for a probability of the form $P(X - \mu \geq a)$, where a is some positive value and when only the mean $\mu = E[X]$ and variance $\sigma^2 = \text{var}(X)$ of the distribution of X are known. Naturally, since $X - \mu \geq a > 0$ implies that $|X - \mu| \geq a$, it follows from Chebyshev's inequality that

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2} \text{ when } a > 0$$

Proposition 8.5.1. *One-sided Chebyshev Inequality. If X is a random variable with mean 0 and finite variance σ^2, then, for any $a > 0$,*

$$P(X \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

Answer. Let $b > 0$ and note that

$$X \geq a \text{ is equivalent to } X + b \geq a + b$$

Hence,

$$\begin{aligned} P(X \geq a) &= P(X + b \geq a + b) \\ &\leq P((X + b)^2 \geq (a + b)^2) \end{aligned}$$

where the inequality is obtained by noting that since $a + b > 0$, $X + b \geq a + b$ implies $(X + b)^2 \geq (a + b)^2$. Upon applying Markov's inequality, the preceding yields that

$$P(X \geq a) \leq \frac{E[(X + b)^2]}{(a + b)^2} = \frac{\sigma^2 + b^2}{(a + b)^2}$$

Letting $b = \sigma^2/a$ [which is easily seen to be the value of b that minimizes $(\sigma^2 + b^2)/(a + b)^2$] gives the desired result. □

Proposition 8.5.2. *If* $E[X] = \mu$ *and* $Var(X) = \sigma^2$*, then, for* $a > 0$*,*

$$P(X \geq \mu + a) \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

$$P(X \leq \mu - a) \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

Proposition 8.5.3. *Cherno Bounds.*

$$P(X \geq a) \leq e^{-ta} M(t) \text{ for all } t > 0$$
$$P(X \leq a) \leq e^{-ta} M(t) \text{ for all } t < 0$$

Since the Cherno bounds hold for all t *in either the positive or negative quadrant, we obtain the best bound on* $P(X \geq a)$ *by using the* t *that minimizes* $e^{-ta} M(t)$*.*

Proposition 8.5.4. *Jensen's Inequality. If* $f(x)$ *is a convex function, then*

$$E[f(X)] \geq f(E[X])$$

provided that the expectations exist and are finite.

9 Homework

Go back to Table of Contents. Please click TOC

This section attaches all the Homework solutions offered by the instructor. Please see Exam Review for more guidance.

GU4203 - Introduction to Probability

Homework 1 Solutions

1 Problems

Question 1 - a) As there are two places for letters (of which there are 26) and five places for numbers (of which there are 10), it follows that there are $26^2 \cdot 10^5 = 67,600,000$ possible license plates.

b) In the case when no letter or number can be repeated, we know that (e.g) after picking the first letter, we have $26 - 1 = 25$ possible choices of letter for the second. By applying this same reasoning for the choice of numbers, we see that there are $26 \cdot 25$ choices for the letters and $10 \cdot 9 \cdot 8 \cdot 7 \cdot 6$ choices for the numbers. Therefore the total possible number of license plates is the product of these numbers, which is $19,656,000$.

Question 7 - a) In this case, as we only care about the possible ways of ordering 6 people, the total number of ways is $6! = 720$.

b) In this case, we first realise that we can either have the boys sitting first or the girls sitting first (2 choices). Then, within each group of boys and girls, there are $3! = 6$ choices of ordering them. Therefore, the total number of ways of ordering the boys and girls in this scenario is $2 \cdot 6 \cdot 6 = 72$.

c) In this case, we note that the first boy can either be in any of the first four positions (4 choices). Then as we have 3! possible ways of ordering the boys once the position of the first boy has been fixed (and similarly so for the girls), there are $4 \cdot 3! \cdot 3! = 144$ possible choices.

d) We can use a similar reasoning as to in part b) to obtain the answer of 72; we can either have BGBGBG or GBGBGB, and then we have the 3! possible ways of ordering the boys/girls separately.

Question 10 - a) As there are no restrictions on how to seat people, there are $8! = 40,320$ possible choices of seating plans.

b) If persons A and B must sit next to each other, then there are two choices of who sits first, and seven choices of where the first person sits (giving 14 in total). Then as for the remaining six persons we have no restrictions on where they sit (so there are 6!), it follows that in total, there are $14 \cdot 6! = 10,080$ possible seating plans.

c) We can use similar reasoning to Q7d in order to deduce that the total number of seating plans is $2 \cdot 4! \cdot 4! = 1152$.

d) We can use similar reasoning to Q7c in order to deduce that the total number of seating plans is $5 \cdot 4! \cdot 4! = 2880$.

e) In this case, we have 4! possible ways of ordering the married couples, and for each married couple, there are 2 ways of ordering how they sit. Therefore, the total number of seating plans is $4! \cdot 2^4 = 384$.

Question 13 - If we have 20 people and everyone shakes hands with everyone else, then we have $\binom{20}{2} = 380$ total handshakes. This is because, for a single handshake, we need to choose 2 people from the 20 to shake hands with each other.

Question 19 - a) To simplify exposition, if a person refuses to serve with someone else, we call them "naughty". If two of the men refuse to serve together, then we need to consider the total number of possibilities when either i) 0 of them are on the committee, or ii) 1 of them is on the committee. In the first case, we are selecting 3 women from 8 and 3 men from 6, giving $\binom{8}{3}\binom{6}{3}$ combinations. In the second case, we are selecting 3 women from 8, 2 men from the 4 men who are not naughty, and 1 man from the 2 naughty men, giving $\binom{8}{3}\binom{4}{2}\binom{2}{1}$ combinations. Therefore in total there are

$$\binom{8}{3}\binom{4}{3} + \binom{8}{3}\binom{4}{2}\binom{2}{1} = 896$$

possible committees.

b) Using a similar argument to the above, we see that there are

$$\binom{6}{3}\binom{6}{3} + \binom{6}{3}\binom{6}{2}\binom{2}{1} = 1000$$

possible committees.

c) In this case, if neither of the naughty people are on the committee, there are $\binom{7}{3}\binom{5}{3}$ choices of committee. Now, if the naughty man is on the committee, there are $\binom{7}{2}\binom{5}{3}$ choices; if the naughty women is on the committee, there are $\binom{7}{3}\binom{5}{2}$ choices. Therefore there are, in total,

$$\binom{7}{3}\binom{5}{3} + \binom{7}{2}\binom{5}{3} + \binom{7}{3}\binom{5}{2} = 910$$

possible committees.

Question 21 - As the hint says, any path consists of 7 total moves, 4 of which are to the right and 3 of which are up. The choice of when to make the 4 right moves (or alternatively the 3 up moves) uniquely determines a path, meaning that there are $\binom{7}{4} = \binom{7}{3} = 35$ total paths.

Question 22 - We break the problem up into two parts; we first consider paths from A to the circled point, and then paths from the circled point to B. As for the first there are $\binom{4}{2}$ possible paths, and the second $\binom{2}{1}$, it follows that there are $\binom{4}{2}\binom{3}{1} = 18$ paths in total which go through the circled point.

Question 30 - First consider the case when we are only concerned about sitting the French and English together. In this case, we have 9 choices of where the first one sits, and 2 choices of the order in which they set. For the remaining delegates there are 8! possible choices of where they sit, giving $18 \cdot 8!$ choices in total.

Now, to get the desired number of seating arrangements, it is enough to subtract the total number of seating arrangements when both the French and English, and the Russians and US delegates, are sitting next to each other. To calculate this, suppose the chairs are labelled 1 through to 10, and the seating numbers of the countries are F, E, R and U respectively. Then in order to determine the seating positions of the pairs FE and RU, it is enough to consider only the smallest number of the pair who have the highest numbers, and the largest number of the pair with the smallest numbers. If you have trouble seeing this, draw a diagram. This corresponds to $8 \cdot 7$ possible choices (as the numbers we are picking from are 2 up to 9). We then have 2 choices each of the choice of ordering within a pair, giving $2^2 \cdot 8 \cdot 7$ possible choices for these four delegates. As we don't care about the placement of the remaining 6 delegates, we have 6! possible choices of ordering for them. In total, this means we have $2^2 \cdot 8 \cdot 7 \cdot 6! = 2^2 \cdot 8!$ possible seating arrangements.

Therefore the final answer is that there are $18 \cdot 8! - 2^2 \cdot 8! = 14 \cdot 8! = 564,440$ possible seating arrangements.

Question 31 - For the first part, we can identify that this is the same problem as asking for the total number of *non-negative* integer solutions to the equation $x_1 + x_2 + x_3 + x_4 = 8$, and so the total number of divisions is $\binom{8+4-1}{4-1} = \binom{11}{3} = 165$. For the second part, we are now after the total number of *positive* integer solutions, and so the total number of divisions is $\binom{8-1}{4-1} = \binom{7}{3} = 35$.

2 Theoretical Exercises

Question 5 - Firstly, we want to determine the number of $0-1$ vectors $(x_1, \ldots, x_n)$ such that $\sum_{i=1}^n x_i = j$. As to do so, we simply need to select j of the n entries to be equal to 1 and the rest 0, it follows that there are $\binom{n}{j}$. Therefore, as

$$\sum_{i=1}^n x_i \geq k \iff \sum_{i=1}^n x_i \in \{k, k+1, \ldots, n\},$$

it follows that the total number of vectors which satisfy the criterion are $\sum_{i=k}^n \binom{n}{i}$.

Question 8 - Using the hint provided, there are $\binom{n+m}{r}$ groups of size r in total. Furthermore, the number of groups which have i men (and therefore $r-i$ women) are $\binom{n}{i}\binom{m}{r-i}$ for $i = 0, 1, \ldots, r$. Therefore, if we do not care about the number of men in the group, we can sum over the i to get to the total number of possible groups, and so

$$\sum_{i=0}^r \binom{n}{i}\binom{m}{r-i} = \binom{n+m}{r}.$$

3

Question 9 - This is a special case of the above formula by setting $n = m$, as then we see that

$$\binom{2n}{n} = \sum_{i=0}^{n} \binom{n}{i}\binom{n}{n-i} = \sum_{i=0}^{n} \binom{n}{i}\binom{n}{i} = \sum_{i=0}^{n} \binom{n}{i}^2.$$

Question 11 - Using the hint provided, we first consider the set $[n] := \{1, \ldots, n\}$. We want to calculate the number of subsets S of size k which have i as their highest number. If i is the highest number contained in S, then we must have that $S \subseteq [i]$. Therefore as i is contained in S, and we have $k-1$ remaining choices of numbers from $[i-1]$, there are $\binom{i-1}{k-1}$ choices in total.

To conclude, it is enough to realize that if we have a subset $S \subseteq [n]$ of size k, then the highest number contained in S could be any of k through to n, and so

$$\binom{n}{k} = \sum_{i=k}^{n} \binom{i-1}{k-1}.$$

Question 13 - This is an immediate consequence of using the binomial formula to expand $0 = (1-1)^n$. Although this seems like a cute result and nothing more, it does have one useful interpretation - it tells us that the total number of subsets of even size is equal to the total number of subsets of odd size. (Why?)

GU4203 - Introduction to Probability

Homework 2 Solutions

1 Problems

Question 3 - We can describe the events as follows:

$$\begin{aligned} E \cap F &= \{(1,2),(1,4),(1,6),(2,1),(4,1),(6,1)\} \\ E \cup F &= \{(x,y) : x+y \text{ is odd or at least one of } x,y \text{ are } 1\} \\ F \cap G &= \{(1,4),(4,1)\} \\ E \cap F^c &= \{(x,y) : x+y \text{ is odd and both } x \text{ and } y \text{ are } > 1\} \\ E \cap F \cap G &= F \cap G \text{ (as } G \subset E). \end{aligned}$$

Question 6 - a) The sample space is $\Omega := \{(1,g),(1,f),(1,s),(0,g),(0,f),(0,s)\}$.

b) A is given by $\{(1,s),(0,s)\}$.

c) B is given by $\{(0,g),(0,f),(0,s)\}$.

d) $A \cup B^c$ is given by $\{(1,s),(0,s),(1,f),(1,g)\}$.

Question 11 - Let $A = \{\text{smokes cigarettes}\}$ and $B = \{\text{smokes cigars}\}$, so $\mathbb{P}(A) = 0.28$, $\mathbb{P}(B) = 0.07$ and $\mathbb{P}(A \cap B) = 0.05$. Then for each part of the question, we are interested in calculating the following probabilities:

a) $1 - \mathbb{P}(A \cup B) = 1 - (\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)) = 0.7$ after substituting the above values in;

b) $\mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B) = 0.02$ after substituting the above values in.

Question 15 - As all poker hands are assumed to be equally likely, we are only really concerned with calculating the possible number of hands with the desired property, as then we can divide by $\binom{52}{5}$ in order to get the probability. Therefore, I will explain only how to count the possible number of hands for each part of the question:

a) In this case, we have 4 possible choices of suit, and then 5 choices from 13 cards of the same suit. This gives $4\binom{13}{5}$ possible choices in total.

b) Firstly, let us focus on the pair. For the pair, we have 13 choices of number, and then $\binom{4}{2}$ choices of suit, giving us $13 \cdot \binom{4}{2}$ possible pairs in total. For the remaining three cards, we need

1

to pick 3 denominations from the remaining 12 (of which there are $\binom{12}{3}$), and we can choose any suit for these three cards (4 for each card), giving us $4^3 \cdot \binom{12}{3}$ possible choices for the remaining three cards. Therefore the total number of hands with one pair is $13 \cdot 4^3 \cdot \binom{4}{2} \cdot \binom{12}{3}$.

c) We begin by focusing on the two pairs. We need to choose 2 denominations from 13 for the two pairs, of which there are $\binom{13}{2}$; for each pair, we then have $\binom{4}{2}$ choices of the suit in each case. As we require the final denomination to be different from the first two, there are $52 - 2 \cdot 4 = 44$ cards from which we can pick the last card. Therefore the total number of hands with two pairs is $44 \cdot \binom{13}{2}\binom{4}{2}^2$.

d) We begin by focusing on the three of a kind. We need to choose 1 denomination from 13 and 3 suits from 4, giving $13 \cdot \binom{4}{3}$ possible combinations of a three of a kind. For the remaining two cards, we need to select 2 different denominations from 12, and then choose the suit of each card, giving us $4^2 \cdot \binom{12}{2}$ possible choices of the remaining two cards. Therefore, the total number of three of a kind hands is $13 \cdot 4^2 \cdot \binom{12}{2}\binom{4}{3}$.

e) We have 13 choices of the card of which we have four in our hand, and then $\binom{48}{1}$ possible choices for the last card in our hand, giving $13\binom{48}{1}$ possible hands in total.

The final numerical probabilities are then as follows: a) 0.198%, b) 42.3%, c) 4.75%, d) 2.11%, e) 0.024%.

Question 25 - As the hint suggests, we want to compute the probability of the event E_n where a 5 occurs on the n-th roll, yet neither a 5 or 7 occurs before then. We begin by noting that the only way of rolling two dice to sum to 5 or 7 are as follows:

$$5 = 4 + 1 = 3 + 2 = 2 + 3 = 1 + 4;$$
$$7 = 6 + 1 = 5 + 2 = 4 + 3 = 3 + 4 = 2 + 5 = 1 + 6.$$

Therefore the probability that, on a single roll of a pair of die, neither a 5 or a 7 occurs is 26/36, and the probability that a 5 occurs is 4/36. As consecutive rolls of the dice are independent, it follows that $\mathbb{P}(E_n) = (25/36)^{n-1} 4/36$. The desired probability is then given by

$$\begin{aligned}
\mathbb{P}(5 \text{ occurs before a } 7) &= \sum_{n=1}^{\infty} \mathbb{P}(5 \text{ occurs before a } 7, \text{ first obtain a 5 or 7 on the } n\text{-th roll}) \\
&= \sum_{n=1}^{\infty} \mathbb{P}(E_n) = \sum_{n=1}^{\infty} \left(\frac{26}{36}\right)^{n-1} \frac{4}{36} \\
&= \frac{2}{5} \text{ (after using the formula for geometric progressions).}
\end{aligned}$$

Question 27 - We can model this problem by letting A and B draw all the balls from the urn, and then compute the probability that A draws the first red ball from the urn. If we were to label the order in which balls are drawn from 1 to 10, this means that we want to compute the probability that the first red ball drawn appears in an odd numbered position.

To compute the probability, we begin by noting that there are 10! possible ways in which the balls could be drawn from the urn, and that all of these ways are equally likely. Now, in order for A to select a red ball first, the following can occur

- *The first red ball appears in the 1st position* - there are 3 choices of red ball to begin with and 9! for the remaining nine balls whose order we do not worry about, giving $3 \cdot 9!$ choices in total;
- *The first red ball appears in the 3rd position* - there are $7 \cdot 6$ choices of white ball for the first two positions, 3 choices for the first red ball, and then 7! choices for the remaining seven balls whose order we do not worry about, giving $7 \cdot 6 \cdot 3 \cdot 7!$ choices in total;
- *The first red ball appears in the 5th position* - there are $7 \cdot 6 \cdot 5 \cdot 4$ choices of white ball for the first four positions, 3 choices for the first red ball, and then 5! choices for the remaining five balls whose order we do not worry about, giving $7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 5!$ choices in total;
- *The first red appears in the 7th position* - there are $7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2$ choices of white ball for the first six positions, 3 choices for the first red ball, and 3! choices for the remaining three balls whose order we do not worry about, giving $7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 3 \cdot 3!$ choices in total.

By summing over the number of choices in each of the four scenarios here, and diving by 10!, we eventually find that the probability is equal to 7/12.

Question 33 - We begin by stating our assumptions - we assume that all of the elk are equally likely to be captured during both occasions, and that whether a elk is captured or not the first time is independent of whether the same elk is captured or not the second time. Now, we note that we have $\binom{20}{4}$ possible combinations of elk who are captured the second time around. If 2 of the captured elk are tagged, we need to select 2 elk from the 5 originally captured (giving $\binom{5}{2}$) and 2 elk from the remaining untagged 15 (giving $\binom{15}{2}$). Under the assumptions stated, the desired probability is given by $\binom{5}{2}\binom{15}{2}/\binom{20}{4} = 70/323 = 21.7\%$.

Question 42 - If two dice are thrown n times in succession, the probability that a double six never occurs is $(35/36)^n$ as successive rolls are independent and the probability that a double six is not rolled on one occasion is $1 - 1/36$. Therefore, the probability that a double six is rolled at least once is $1 - (35/36)^n$. If we want this probability to be at least 1/2, then

$$1 - \left(\frac{35}{36}\right)^n \geq \frac{1}{2} \iff \left(\frac{35}{36}\right)^n \geq \frac{1}{2} \iff n \geq \frac{\log(1/2)}{\log(35/36)},$$

meaning the smallest number of n necessary is 25.

Question 53 - We use the Inclusion-Exclusion principle. Let A_i be the event that the i-th couple sits next to each other; we are therefore interested in the probability $1 - \mathbb{P}(A_1 \cup A_2 \cup A_3 \cup A_4)$. Now, in total there are 8! possible arrangements of the 4 couples, all of which we can consider to be equally likely. As the couples are indistinguishable, the Inclusion-Exclusion formula simplifies to

$$\mathbb{P}(A_1 \cup A_2 \cup A_3 \cup A_4) = 4\mathbb{P}(A_1) - 6\mathbb{P}(A_1 \cap A_2) + 4\mathbb{P}(A_1 \cap A_2 \cap A_3) - \mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4).$$

The probabilities in the above formula are then given by:

- $\mathbb{P}(A_1) = 2 \cdot 7!/8!$ - We have 7 choices of the location of the first partner (and two choices for the order they sit in), and 6! choices of the positions of the remaining partners.
- $\mathbb{P}(A_1 \cap A_2) = 2^2 \cdot 6!/8!$ - We have 6! choices of the location for the two pairs of partners and the remaining 4 individuals. We then have 2 choices for the order a couple sits in for both couples (giving 2^2 choices). Remember that the above procedure determines the positions of the third and fourth couples.
- $\mathbb{P}(A_1 \cap A_2 \cap A_3) = 2^3 \cdot 5!/8!$ - We have 5! choices of location for the three pairs of partners and the remaining 2 individuals. We then have 2^3 choices in total for the order in which the first, second and third couples sit in.
- $\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) = 2^4 \cdot 4!/8!$ - At this point, we are only concerned with how we can order four couples next to each other (4! in total), and how we can arrange the partners in each couple (2 per married couple, giving 2^4 in total).

Substituting these in to the above formula, and then subtracting it from 1, gives a final answer of $12/35 = 34.3\%$.

Question 54 - We use the Inclusion-Exclusion principle. Let S be the event that a bridge hand is void of a spade, and similarly define events C, D and H for clubs, diamonds and hearts respectively. We want to compute the probability $\mathbb{P}(S \cup C \cup D \cup H)$. Now, as all of the suits are equally likely, we have e.g $\mathbb{P}(C \cap D) = \mathbb{P}(S \cap H)$, and so the Inclusion-Exclusion formula simplifies down to (for example)

$$\mathbb{P}(S \cup C \cup D \cup H) = 4\mathbb{P}(S) - 6\mathbb{P}(S \cap H) + 4\mathbb{P}(S \cap H \cap D).$$

Note that $\mathbb{P}(S \cup C \cup D \cup H) = 0$ as a hand cannot be devoid of all the four suits. Now, as all bridge hands are equally likely (giving $\binom{52}{13}$ in total), we see that

- $\mathbb{P}(S) = \binom{39}{13}/\binom{52}{13}$ as we need to choose 13 cards from the 39 cards which are not spades;
- $\mathbb{P}(S \cap H) = \binom{39}{13}/\binom{52}{13}$ as we need to choose 13 cards from the 26 cards which are not spades or hearts;
- $\mathbb{P}(S \cap H \cap D) = \binom{13}{13}/\binom{52}{13}$ as we need to choose 13 cards from the 13 cards which are not spades, hearts or diamonds.

Substituting these into the above formula then gives a probability of approximately 5.1%.

2 Theoretical Exercises

Question 5 - We want to find a disjoint collection of F_i such that $\cup_{i=1}^m F_i = \cup_{i=1}^m E_i$ for all $m \geq 1$, given a (potentially countable infinite) sequence of events E_i. Note that the $m = 1$ case tells us that $F_1 := E_1$. For the $m = 2$ case, note that we can write

$$F_1 \cup F_2 = E_1 \cup E_2 = E_1 \cup (E_2 \cap E_1^c)$$

so both the left and right hand side are disjoint unions; as $F_1 = E_1$, we therefore can choose $F_2 := E_2 \cap E_1^c$. If we keep repeating this, we begin to see a pattern forming from which we decide to choose

$$F_i := E_i \cap \left(\bigcap_{j=1}^{i-1} E_j^c \right).$$

To prove that this has the desired properties, first note that if $i < j$ then $F_i \cap F_j \subseteq E_i \cap E_i^c = \emptyset$, so the F_i are pairwise disjoint. Then in order to show that $\cup_{i=1}^m F_i = \cup_{i=1}^m E_i$ for all $m \geq 1$, we do so by induction. The $m = 1$ case is immediate. Then if the statement is true for $m = n$, we see that it is true for $m = n + 1$ as

$$\begin{aligned}
\bigcup_{i=1}^{n+1} F_i &= \left(\bigcup_{i=1}^{n} F_i \right) \cup F_{n+1} = \left(\bigcup_{i=1}^{n} E_i \right) \cup \left(E_n \cap \bigcap_{i=1}^{n} E_i^c \right) \\
&= \left(\bigcup_{i=1}^{n} E_i \cup E_{n+1} \right) \cap \left(\bigcup_{i=1}^{n} E_i \cup \left(\bigcup_{i=1}^{n} E_i \right)^c \right) \\
&= \bigcup_{i=1}^{n} E_i \cup E_{n+1} = \bigcup_{i=1}^{n+1} E_i,
\end{aligned}$$

where we have used de-Morgan's laws and the distributivity properties of unions/intersections.

Question 11 - Bonferroni's inequality follows as a simple consequence of the fact that probabilities are bounded above by 1, and then some rearranging:

$$1 \geq \mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F).$$

Question 12 - As the event of interest is a disjoint union of the events $E \cap F^c$ and $F \cap E^c$, we get that the desired probability is

$$\begin{aligned}
\mathbb{P}(E \cap F^c) + \mathbb{P}(F \cap E^c) &= \mathbb{P}(E) - \mathbb{P}(E \cap F) + \mathbb{P}(F) - \mathbb{P}(E \cap F) \\
&= \mathbb{P}(E) + \mathbb{P}(F) - 2\mathbb{P}(E \cap F).
\end{aligned}$$

GU4203 - Introduction to Probability

Homework 3 Solutions

1 Problems

Question 1 - Let A be the event that at least one die rolls a six and B be the event that the two die rolled are different. Then $\mathbb{P}(B) = 5/6$ (as we simply need the second die to be one of the five possible values which is different from that obtaining by the first die) and

$$\mathbb{P}(A \cap B) = \mathbb{P}(\text{1st die} = 6, \text{2nd die} \neq 6) + \mathbb{P}(\text{1st die} \neq 6, \text{2nd die} = 6) = 2 \cdot \frac{1}{6} \cdot \frac{5}{6}.$$

Therefore the desired probability is $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) = 1/3$.

Question 4 - Let S be the sum of the value of the two dice, and A be the event that at least one of the die lands on a 6. To compute the conditional probabilities, as the dice rolls are equally likely, it suffices to calculate the proportion

$$\frac{\text{number of dice rolls which sum to } S = i \text{ and contain one six}}{\text{number of dice rolls which sum to } S = i}.$$

These can be calculated simply by writing out the possible dice roll combinations which sum to $S = i$, and then counting the total number and the number which contain at least one six. The desired probabilities are then given as follows:

- $\mathbb{P}(A|S = i) = 0$ for $2 \leq i \leq 6$;
- $\mathbb{P}(A|S = 7) = 1/3$;
- $\mathbb{P}(A|S = 8) = 2/5$;
- $\mathbb{P}(A|S = 9) = 1/2$;
- $\mathbb{P}(A|S = 10) = 2/3$;
- $\mathbb{P}(A|S = i) = 1$ for $11 \leq i \leq 12$.

Question 5 - On the first pick, we have a probability of 6/15 of picking a white ball. Then for the second pick, we have 5 white balls and 9 black balls, so the probability of picking a white ball now is 5/14. For the third pick, we have 4 white balls and 9 black balls, so the probability of picking a black ball is 9/13. Finally, for the last pick, we have 4 white balls and 8 black balls, so the probability of picking a black ball is 8/12. Multiplying these together gives the desired probability, which is equal to 6/91 after some simplification.

Question 6 - Given that the sample drawn contains exactly 3 white balls (and so only 1 black ball), the black ball is equally likely to be in any of the 4 positions. In other words, conditional on the sample drawn, the probability that the i-th ball is white is equal to that of whether it is black, and so the answer (in both cases) is $1/2$.

Question 7 - Here we suppose that the two children are male or female with equal probability and independently of each other. Now, given what we know, the probability of interest is

$$\mathbb{P}(\text{one boy, one girl} \mid \text{at least one boy}) = \frac{\mathbb{P}(\text{one boy, one girl})}{\mathbb{P}(\text{at least one boy})} = \frac{1/2}{3/4} = \frac{2}{3}.$$

Question 10 - This can either be done by using the definition of conditional probability, or (as we do) by employing a symmetry argument. This allows us to say that it is equivalent to consider the conditional probability as if the second and third draws from the deck were actually the first and second, and the first draw as being the third after two spades were drawn. If you do not believe this immediately, let A_i be the event that the i-th draw from the deck is a spade, and note that

$$\mathbb{P}(A_1|A_2, A_3) = \frac{\mathbb{P}(A_1 \cap A_2 \cap A_3)}{\mathbb{P}(A_2 \cap A_3)} = \frac{\mathbb{P}(A_1 \cap A_2 \cap A_3)}{\mathbb{P}(A_1 \cap A_2)} = \mathbb{P}(A_3|A_1, A_2).$$

The latter probability is then given by $11/50$, as there are 11 spades remaining from 50 cards.

Question 14 - a) The probability of the first ball selected being black is $5/12$. Afterwards, there are 7 black balls and 7 white balls, meaning the probability of the second ball selected being black is $7/14$. Then there are 9 white balls and 7 white balls, meaning the probability of the third ball selected being white is $7/16$. Finally, there are 9 white balls and 9 black balls, so the probability that the fourth and last ball selected being white is $9/18$. Multiplying these four probabilities gives the desired result of $35/768$.

b) There are two ways of approaching this problem. Letting W represent a white ball, and B a black ball, we can recognize that the probability we are interested in is equal to

$$\mathbb{P}(WWBB) + \mathbb{P}(WBWB) + \mathbb{P}(BWBW) + \mathbb{P}(BWWB) + \mathbb{P}(WBBW) + \mathbb{P}(BBWW).$$

We can then either compute each of these probabilities by hand and note that they are all equal to $35/768$, or argue by symmetry that they are all equal and so by part a) they are all equal to $35/768$. In either case, we see that the desired probability is $210/768 = 0.273$.

Question 15 - Let E be the event that a pregnant women has an ectopic pregnancy, and S be the even that they are a smoker. Extracting information from the question, we see that $\mathbb{P}(E|S) = 2\mathbb{P}(E|S^c)$ and $\mathbb{P}(S) = 0.32$. Then by Bayes' theorem, we find that

$$\begin{aligned}\mathbb{P}(S|E) &= \frac{\mathbb{P}(E|S)\mathbb{P}(S)}{\mathbb{P}(E|S)\mathbb{P}(S) + \mathbb{P}(E|S^c)\mathbb{P}(S^c)} \\ &= \frac{2\mathbb{P}(S)}{2\mathbb{P}(S) + 1 - \mathbb{P}(S)} = \frac{0.64}{1.32} = \frac{32}{66} = 0.4848.\end{aligned}$$

Question 19 - a) Let A be the event that a person attends the party, W be the event that this person is a women, and $M = W^c$ be the event that this person is a man. Then by Bayes' theorem,

$$\begin{aligned}\mathbb{P}(W|A) &= \frac{\mathbb{P}(A|W)\mathbb{P}(W)}{\mathbb{P}(A|W)\mathbb{P}(W) + \mathbb{P}(A|M)\mathbb{P}(M)} \\ &= \frac{0.48 \cdot 0.38}{0.48 \cdot 0.38 + 0.37 \cdot 0.62} = 0.443.\end{aligned}$$

As the question asks for us to report a percentage, the answer is that 44.3% of the attendees at the party were women.

b) By the law of total probability, we have that

$$\mathbb{P}(A) = \mathbb{P}(A|W)\mathbb{P}(W) + \mathbb{P}(A|M)\mathbb{P}(M) = 0.48 \cdot 0.38 + 0.37 \cdot 0.62 = 0.412,$$

and so 41.2% of the class attended the party.

GU4203 - Introduction to Probability

Homework 4 Solutions

1 Problems

Question 23 - a) Let R be the event that a red ball is transferred from urn I to urn II, and W be the event that it is a white ball instead. Then we know that $\mathbb{P}(R) = 2/3$ and $\mathbb{P}(W) = 1/3$. Let A be the event that the ball selected from urn II is white. Then by the law of total probability, we have that

$$\mathbb{P}(A) = \mathbb{P}(A \mid W)\mathbb{P}(W) + \mathbb{P}(A \mid R)\mathbb{P}(R) = \frac{2}{3}\ \frac{1}{3} + \frac{1}{3}\ \frac{2}{3} = \frac{4}{9}.$$

b) By Bayes' theorem, we see that

$$\mathbb{P}(W \mid A) = \frac{\mathbb{P}(A \mid W)\mathbb{P}(W)}{\mathbb{P}(A)} = \frac{\frac{2}{3}\ \frac{1}{3}}{\frac{4}{9}} = \frac{1}{2}.$$

Question 27 - Honestly, this question is worded really badly. On a brief philosophical note, either method could be used to estimate this quantity; the point is that one is better than the other. I could roll a die and tell you that the face-up value is an "estimate" of the average number of workers in a car. Of course, as the two processes are completely independent, this would be a useless estimate.

Anyways, the second method is the better way of estimating this quantity. As we are interested in the average number of workers per car, we want to sample from the cars as the number of workers inside it is a property of the cars. The first method is bad as it has the possibility from over-sampling from cars with a large number of workers in them; if we sample multiple people from the same car with a large number of workers inside of it, we will over-estimate the average proportion of workers per car.

Question 32 - Let E be the event that the eldest child in the family is chosen, and let F_j be the event that the family selected has j children. Then by Bayes' theorem, we have that

$$\mathbb{P}(F_j \mid E) = \frac{\mathbb{P}(E \mid F_j)\mathbb{P}(F_j)}{\sum_{i=1}^{4} \mathbb{P}(E \mid F_i)\mathbb{P}(F_i)} = \frac{p_j/j}{\sum_{i=1}^{4} p_i/i}.$$

Using this formula then gives the answers a) 0.24, b) 0.18; nothing changes when repeating the calculation for when the randomly selected child is the youngest, so the answers are the same again.

Question 43 - By Bayes' theorem, the probability $\mathbb{P}(\text{two headed coin} \mid \text{heads})$ is given by

$$\frac{\mathbb{P}(\text{heads} \mid \text{two headed})\mathbb{P}(\text{two headed})}{\mathbb{P}(\text{heads} \mid \text{two headed})\mathbb{P}(\text{two headed}) + \mathbb{P}(\text{heads} \mid \text{fair})\mathbb{P}(\text{fair}) + \mathbb{P}(\text{heads} \mid \text{biased})\mathbb{P}(\text{biased})},$$

so substituting in the various quantities gives us that the solution is

$$\frac{\frac{1}{3} \cdot 1}{\frac{1}{3} \cdot 1 + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{3}{4}} = \frac{4}{9}.$$

Question 44 - This problem is similar to the Monty Hall problem. The jailer's reasoning is faulty, as by revealing that one of his fellow prisoners is to be set free, the probability of the other fellow prisoner being executed has risen to 2/3. However, his own probability has stayed the same at 1/3.

Question 45 - By Bayes' theorem,

$$\mathbb{P}(\text{fifth coin} \mid H) = \frac{\mathbb{P}(H \mid \text{fifth coin})\mathbb{P}(\text{fifth coin})}{\sum_{i=1}^{10} \mathbb{P}(H \mid i\text{-th coin})\mathbb{P}(i\text{-th coin})} = \frac{\frac{5}{10} \cdot \frac{1}{10}}{\sum_{i=1}^{10} \frac{i}{10} \cdot \frac{1}{10}} = \frac{1}{11}.$$

Question 49 - Let C be the event that the patient has cancer, and let E be the event that the test indicates an elevated PSA level. Writing $p := \mathbb{P}(C)$, we obtain by Bayes' theorem that

$$\mathbb{P}(C \mid E) = \frac{\mathbb{P}(E \mid C)\mathbb{P}(C)}{\mathbb{P}(E \mid C)\mathbb{P}(C) + \mathbb{P}(E \mid C^c)\mathbb{P}(C^c)} = \frac{0.268p}{0.268p + 0.135(1-p)}$$

$$\mathbb{P}(C \mid E^c) = \frac{\mathbb{P}(E^c \mid C)\mathbb{P}(C)}{\mathbb{P}(E^c \mid C)\mathbb{P}(C) + \mathbb{P}(E^c \mid C^c)\mathbb{P}(C^c)} = \frac{0.732p}{0.732p + 0.865(1-p)}.$$

Therefore if $p = 0.7$, we have a) 0.8224, b) 0.6638; if $p = 0.3$, we have a) 0.4597, b) 0.2661.

Question 51 - a) Let R be the event that the worker receives a job offer. Then by the law of total probability, we have that

$$\begin{aligned}\mathbb{P}(R) &= \mathbb{P}(R \mid \text{strong})\mathbb{P}(\text{strong}) + \mathbb{P}(R \mid \text{moderate})\mathbb{P}(\text{moderate}) + \mathbb{P}(R \mid \text{weak})\mathbb{P}(\text{weak}) \\ &= 0.8 \cdot 0.7 + 0.4 \cdot 0.2 + 0.1 \cdot 0.1 = 0.65.\end{aligned}$$

b), c) These problems boil down to using Bayes' theorem twice by noting that

$$\mathbb{P}(A \mid R) = \frac{\mathbb{P}(R \mid A)\mathbb{P}(A)}{\mathbb{P}(R)}, \qquad \mathbb{P}(A \mid R^c) = \frac{(1 - \mathbb{P}(R \mid A))\mathbb{P}(A)}{1 - \mathbb{P}(R)}$$

for $A \in \{\text{strong}, \text{moderate}, \text{weak}\}$. The desired probabilities are then, in order for which they are asked for by the textbook, 56/65, 8/65, 1/65, 14/35, 12/35, 9/35.

Question 55 - Let x be the number of sophomore girls present. Then as the class and sex only take two values each, it is in fact equivalent to check that the events of being a boy (B) and being a first-year (F) are independent. Then as

$$\mathbb{P}(B, F) = \frac{4}{16 + x}, \qquad \mathbb{P}(B) = \frac{10}{16 + x}, \qquad \mathbb{P}(F) = \frac{10}{16 + x},$$

it follows that B and F are independent if and only if

$$4 = \frac{100}{16+x} \iff x = 9.$$

Question 64 - Strategy a) gives a probability p of getting the correct answer. For strategy b), we have by the law of total probability that

$$\mathbb{P}(\text{win}) = \mathbb{P}(\text{win}|\text{both correct})p^2 + \mathbb{P}(\text{win}|\text{only one correct})2p(1-p) + \mathbb{P}(\text{win}|\text{neither correct})(1-p)^2 = p^2 + p(1-p) = p.$$

Therefore both strategies have the same chance of being successful. Here's a fun thing to think about - is there a better strategy than either a) or b)?

Question 70 - Let C be the event that the queen is a carrier, and A be the event that the three princes do not have the disease. Then by Bayes' theorem,

$$\mathbb{P}(C|A) = \frac{\mathbb{P}(A|C)\mathbb{P}(C)}{\mathbb{P}(A|C)\mathbb{P}(C) + \mathbb{P}(A|C^c)\mathbb{P}(C^c)} = \frac{\frac{1}{8}\cdot\frac{1}{2}}{\frac{1}{8}\cdot\frac{1}{2} + 1\cdot\frac{1}{2}} = \frac{1}{9}.$$

Then if there is a fourth prince, if the queen is a carrier (which is probability 1/9 using our knowledge), then there is a 1/2 probability of the prince having haemophilia; if the queen is not a carrier, then there is zero chance of the prince having haemophilia. Therefore by the law of total probability, the probability that the fourth prince has haemophilia is 1/18.

10 Exam Review

Go back to Table of Contents. Please click TOC

A few words about exam-taking philosophy and TA responsibility.

Exam-taking Philosophy.

Coming from an environment that promotes test-bank strategy (a strategy based on quantity of tests students take) and currently trained in an environment that promotes test-type strategy (an opposite kind of strategy as test-bank strategy that based on types of questions students take), I think all strategies can be laid out in the following spectrum.

Left (	)	Right
Test-bank Strategy (	)	Test-type Strategy
a.k.a Focus on Quantity (	)	Focus on Quality

That being said, a student can be on the left or right or middle of the spectrum. To cover this spectrum (e.g. to ful ll everyone's needs as much as possible), I will point out the guidelines rst (covering what types of problems are fair games in the exam). Then I will lead discussion of a few problems in each type. There will be more problems students can do, however, I will have to leave that to students as the class has limited amount of time.

10.1 1st Midterm

The following notions are important for this midterm.

1. Chapter 1. Counting Principle, Permutation, Combinations, Binomial Theorem (and its Propositions).
2. Chapter 2. Sample Space, Union, Intersection, Complement, Mutually Exclusive, Inclusion-Exclusion Identity
3. Chapter 3. Conditional Probability, Multiplication Rule of Probability, Bayes's Formula, Denominator of Bayes's Formula by Law of Total Probability.

Permutation

There are n people sitting in a row and this will give us $n!$ di erent arrangements. Then on top of this the problem can build up premises:

A classical example (Homework #1) is the following.

Consider delegates from 10 countries with R, F, E, U and rest of the 6 more countries sitting in a row. We want to satisfy two premises (1) F and E are sitting together, i.e. FE or EF. (2) R and U are not sitting together.

Answer. The key is to work out (1) rst and then subtract the compliment of (2) to obtain the answer.

(1) We want FE- - - - - - - -, - FE - - - - - - -, ..., - - - - - - - - FE so this is total of 9 possible outcomes and with FE and EF being di erent arrangement. The rest of the 8 countries may sit in di erent arrangement. That is, (2)(9)(8!) possibilities.

(2) We want the compliment of (2). (2) says R and U are not sitting together. The compliment of this event is referring to the situation while R and U are indeed sitting together. While satisfying (1) the same time, we have compliment of (2) to be FERU, FEUR, EFRU, EFUR. so this is 2 2 and the rest 8 countries may sit in di erent arrangements.

Thus, we have final answer (1) minus compliment of (2). This gives us

$$(2)(9)(8!) - (2)(2)(8!) = 564,440$$

□

Another classical example is the grid problem from homework and also textbook. One can refer to the following.

Consider a grid that has size 3 by 4. We want to move from A to B that requires 3 ups and 4 rights. It does not matter which move you go first.

Answer. Thus, we have

$$\binom{3+4}{3} = \binom{7}{3} = \binom{7}{4} = 35$$

□

Positive Solution

Given an equation in the form of

$$\sum_{i=1}^{r} x_i = n$$

you need to be familiar how to solve this type of problems. Be aware of both premises (1) assume positive solution, and (2) assume non-negative solution for X_i's.

You should recall formula

$$\binom{n-1}{r-1} \text{ for positive integer-valued vector } (x_1, \dots, x_r)$$

and

$$\binom{n+r-1}{r-1} \text{ for non-negative integer-valued vector } (x_1, \dots, x_r)$$

A classical example is from Homework 1. Given 8 identical blackboards to be identically distributed among 4 schools, we want to find out

(a) How many distributions are possible? This is

$$X_1 + X_2 + X_3 + X_4 = 8$$

while allowing "0" so we have

Answer.

$$\binom{8+4-1}{4-1} = \binom{11}{3} = 165$$

□

(b) How many if at least 1 is distributed to each school? This is not allowing "0". Hence, we have

Answer.

$$\binom{8-1}{4-1} = \binom{7}{3} = 35$$

□

Positive Solution

Bayes' Formula is definitely within the scope of this exam. You should be familiar with all sorts of formulas in this arena. Let us recall Bayes' Formula

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Please also be aware of the identity (aka multiplication rule of probability).

$$P(E) = P(E|F)P(F) + P(E|F^c)P(F^c)$$

You should also be aware of the related formulas as one can always derive Bayes' formula in fancy ways depending on the problem.

A classical example is from Homework 3. An ectopic pregnancy is twice as likely to develop when the pregnant woman is a smoker as it is a non-smoker. If 32 percent of women of childbearing age are smokers, what percentage of women having ectopic pregnancies are smokers?

Answer. Let E be the event that a pregnant women has an ectopic pregnancy, and S be the event that they are smokers. We know that $P(E|S) = 2P(E|S^c)$ and $P(S) = 0.32$. Then by Bayes' Theorem, we find

$$\begin{aligned}
P(S|E) &= \frac{P(E|S)P(S)}{P(E|S)P(S) + P(E|S^c)P(S^c)} \\
&= \frac{2P(E|S^c)(0.32)}{2P(E|S^c)(0.32) + P(E|S^c)(1 - 0.32)} \\
&= \frac{0.64}{0.64 + 0.68} \\
&= \frac{0.64}{1.32} = \frac{32}{66}
\end{aligned}$$

□

One can also refer to another problem in Homework 3. A total of 48 percent of the women and 37 percent of the men who took a certain "quit smoking" class remained nonsmokers for at least one year after completing the class. These people then attended a success party at the end of a year. If 62 percent of the original class was male,

1. what percentage of those attending the party were women?
2. what percentage of the original class attended the party?

Answer. We answer the question in the following

1. Let A be the event that a person attends a party. Let W be the event that the person is a woman, and $M = W^c$ be the event that this person is a man. Then by Bayes' Theorem

$$\begin{aligned}
P(W|A) &= \frac{P(A|W)P(W)}{P(A|W)P(W) + P(A|M)P(M)} \\
&= \frac{0.48 \times 0.38}{0.48 \times 0.38 + 0.37 \times 0.62} \\
&= 0.44
\end{aligned}$$

2. By the law of total probability, we have that

$$P(A) = P(A|W)P(W) + P(A|M)P(M) = 0.48 \cdot 0.38 + 0.37 \cdot 0.62 = 0.41$$

□

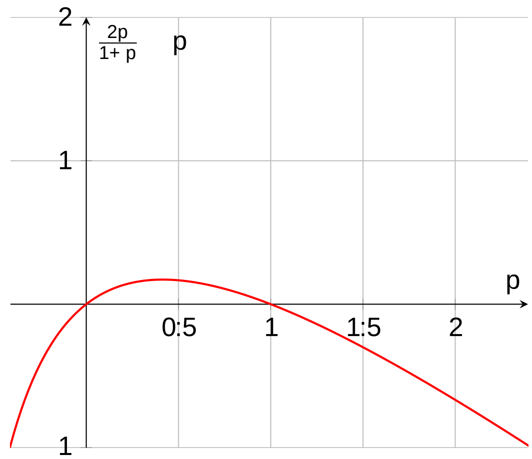
Female chimp gave birth. It is not certain which of two male chimps is the father. Before genetic analysis, it is believed that the probability that male number 1 is the father is p and the probability that male number 2 is the father is $1 - p$. DNA obtained from the mother, male number 1, and male number 2 indicates that on one specific location of the genome, the mother has the gene pair (A,A), male number 1 has gene pair (a,a), and male number 2 has the gene pair (A,a). If a DNA test shows that the baby chimp has the gene pair (A,a), what is the probability that male number 1 is the father?

Answer. Let M_i be the event that male number i is the father. Let $B_{A,a}$ be the event that baby chimp has the gene pair (A,a). Then $P(M_1|B_{A,a})$ is obtained:

$$\begin{aligned} P(M_1|B_{A,a}) &= \frac{P(M_1 B_{A,a})}{P(B_{A,a})} \\ &= \frac{P(B_{A,a}|M_1)P(M_1)}{P(B_{A,a}|M_1)P(M_1) + P(B_{A,a}|M_2)P(M_2)} \\ &= \frac{1 \cdot p}{1 \cdot p + (1/2)(1-p)} \\ &= \frac{2p}{1+p} \end{aligned}$$

Now let us compare result with p

Figure 3: The figure presents the graph of $\frac{2p}{1+p} - p$.



Hence, we arrived the inequality

$$\frac{2p}{1+p} > p$$

We conclude the information that the baby's gene pair is (A,a) increases the probability that male number 1 is the father. □

10.2 2nd Midterm

The following notions are important for this midterm.

1. Chapter 1. Counting Principle, Permutation, Combinations, Binomial Theorem (and its Propositions).
2. Chapter 2. Sample Space, Union, Intersection, Complement, Mutually Exclusive, Inclusion-Exclusion Identity
3. Chapter 3. Conditional Probability, Multiplication Rule of Probability, Bayes's Formula, Denominator of Bayes's Formula by Law of Total Probability.
4. Chapter 4. Density function (PDF), Distribution function (CDF), Expectation (Mean), Variance, Famous distributions (binomial, Poisson, geometric, negative binomial).
5. Chapter 5. Uniform. Normal. Exponential. Memoryless Property. Application on Memoryless Property.

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

The expected value of X is a weighted average of the possible values that X can take on, each value being weighted by the probability that X assumes it.

Example 10.2.1. Find $E[X]$, where X is the outcome when we roll a fair die.

Answer. Since $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = 1/6$, we obtain

$$E[X] = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = 7/2$$

□

Example 10.2.2. Calculate $\text{Var}(X)$ if X represents the outcome when a fair die is rolled.

Answer. You can easily find $E[X] = \frac{7}{2}$. Now, we find

$$\begin{aligned} E[X^2] &= 1^2(1/6) + 2^2(1/6) + 3^2(1/6) + 4^2(1/6) + 5^2(1/6) + 6^2(1/6) \\ &= (91)(1/6) \end{aligned}$$

and thus we have variance

$$\text{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

□

Example 10.2.3. Suppose X is a continuous random variable whose probability density function is

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{else} \end{cases}$$

1. What is the value of C?
2. Find $P(X > 1)$.

Answer. We have the following

1. Since f is a probability density function, we must have

$$\int_{1}^{1} f(x)dx = 1$$

and we can solve $C\int_0^2 (4x \quad 2x^2)dx = 1$. After integration, we have $C(2x^2 \quad \frac{2x^3}{3})|_{x=0}^2 = 1$ and we have result $C = \frac{3}{8}$.

2. $P(X > \quad 1) = \int_1^1 f(x)dx = \frac{3}{8}\int_1^2 (4x \quad 2x^2)dx = \frac{1}{2}$.

□

Example 10.2.4. The amount of time in hours that a computer functions before breaking down is a continuous random variable with probability density function given by

$$f(x) = \begin{cases} e^{\quad x= 100} & x \quad 0 \\ 0 & x < 0 \end{cases}$$

What is the probability that

1. a computer will function between 50 and 150 hours before breaking down?
2. it will function for fewer than 100 hours?

Answer. We solve the parts accordingly

1. Since $1 = \int_1^1 f(x)dx = \int_0^1 e^{\quad x= 100}dx$, we can take integral and obtain $1 = \quad (100)e^{\quad x= 100}|_0^1 = 100 \quad$. We can solve for $\quad = \frac{1}{100}$. Then we can proceed to nd the probability

$$\begin{aligned} P(50 < X < \quad 1500) &= \int_{50}^{150} \frac{1}{100} e^{\quad x= 100}dx \\ &= \quad e^{\quad x= 100}|_{50}^{150} \\ &= e^{\quad 1=2} \quad e^{\quad 3=2} \\ &= 0:383 \end{aligned}$$

2. I will leave this to you as an exercise.

□

In general, we say that X is a uniform random variable on the interval $(\quad ; \quad)$ if the probability density function of X is given by

$$f(x) = \begin{cases} \frac{1}{\quad} & \text{if} \quad < x < \\ 0 & \text{else} \end{cases}$$

Since $F(a) = \int_1^a f(x)dx$, it follows that

$$f(x) = \begin{cases} 0 & a \\ \frac{1}{\quad} & \text{if} \quad < x < \\ 1 & a \end{cases}$$

Example 10.2.5. Let X be uniformly distributed over $(\quad ; \quad)$. Find (a) $E[X]$ and (b) $Var(X)$.

Answer. We proceed accordingly

1. Compute

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha} dx \\ &= \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} \\ &= \frac{\beta + \alpha}{2} \end{aligned}$$

2. To find Var(X), first calculate $E[X^2]$.

$$\begin{aligned} E[X^2] &= \int_{\alpha}^{\beta} \frac{1}{\beta - \alpha} x^2 dx \\ &= \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)} \\ &= \frac{\beta^2 + \alpha\beta + \alpha^2}{3} \end{aligned}$$

Hence,

$$\text{Var}(X) = \frac{\beta^2 + \alpha\beta + \alpha^2}{3} - \frac{(\alpha + \beta)^2}{4} = \frac{(\beta - \alpha)^2}{12}$$

□

Example 10.2.6. If X is uniformly distributed over $(0, 10)$, calculate the probability that $X < 3$.

Answer. Compute $P(X < 3) = \int_0^3 \frac{1}{10} dx = \frac{3}{10}$. □

We say that X is a normal random variable, or simply that X is normally distributed, with parameters μ and σ^2 if the density of X is given by

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

for $-\infty < x < \infty$. The density function is a bell-shaped curve that is symmetric about μ.

Example 10.2.7. Find $E[X]$ and Var(X) when X is a normal random variable with parameters μ and σ^2.

Answer. Let us start by finding the mean and variance of the standard normal random variable $Z = (X - \mu)/\sigma$. We have

$$\begin{aligned} E[Z] &= \int_{-\infty}^{\infty} x f_Z(x) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} dx \\ &= -\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \Big|_{-\infty}^{\infty} \\ &= 0 \end{aligned}$$

Thus,

$$\begin{aligned}
\mathrm{Var}(Z) &= E[Z^2] \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx, \text{ IBP: let } \mu = x \text{ and } dv = xe^{-x^2/2} \\
&= \frac{1}{\sqrt{2\pi}} \left(-xe^{-x^2/2}\Big|_{-\infty}^{\infty} + \underbrace{\int_{-\infty}^{\infty} e^{-x^2/2} dx}_{=1} \right) \\
&= 1
\end{aligned}$$

Because $X = \mu + \sigma Z$, the preceding yields the results

$$E[X] = \mu + \sigma E[Z] = \mu$$

and

$$\mathrm{Var}(X) = \sigma^2 \mathrm{Var}(Z) = \sigma^2$$

□

Example 10.2.8. If X, the gain from an investment, is a normal random variable with mean μ and variance σ^2, then because the loss is equal to the negative of the gain, the VAR of such an investment is that value v such that

$$0.01 = P(-X > v)$$

We compute the following

$$\begin{aligned}
0.01 &= P\left(\frac{-X+\mu}{\sigma} > \frac{v+\mu}{\sigma} \right) \\
&= 1 - \Phi(\frac{v+\mu}{\sigma})
\end{aligned}$$

and from table we know $\Phi(2.33) = 0.99$ so we know $\frac{v+\mu}{\sigma} = 2.33$. That is, $v = \mathrm{VAR} = 2.33\sigma - \mu$. Consequently, among set of investments all of whose gains are normally distributed, the investment having the smallest VAR is the one having the largest value of $\mu - 2.33\sigma$.

Remark 10.2.9. Please repeat the above analysis for

1. Discrete: Binomial, Poisson, Geometric.
2. Continuous: Uniform, Normal, Exponential.

Remark 10.2.10. Resources:

1. StatLect Website: `https://www.statlect.com/probability-distributions/`
2. Univariate Distribution Relationships: `http://www.math.wm.edu/~leemis/chart/UDR/UDR.html`

10.3 Final Exam

The following notions are important for this final exam.

1. Chapter 1. Counting Principle, Permutation, Combinations, Binomial Theorem (and its Propositions).
2. Chapter 2. Sample Space, Union, Intersection, Complement, Mutually Exclusive, Inclusion-Exclusion Identity
3. Chapter 3. Conditional Probability, Multiplication Rule of Probability, Bayes's Formula, Denominator of Bayes's Formula by Law of Total Probability.
4. Chapter 4. Density function (PDF), Distribution function (CDF), Expectation (Mean), Variance, Famous distributions (binomial, Poisson, geometric, negative binomial).
5. Chapter 5. Uniform. Normal. Exponential. Memoryless Property. Application on Memoryless Property.
6. Chapter 6. Joint Cumulative Probability Distribution Function, Joint Probability Mass Function
7. Chapter 7. MGF, Expectation, Variance.
8. Chapter 8. Markov, Chebyshev.

Please be aware of the following:

- Exam is cumulative from Chapter 1 to Chapter 8.
- Midterm I & II are very good reference of the final. For problems asking topics discussed in Chapter 5 and before, Midterm I & II provide very good insight.
- For problems asking topics discussed in Chapter 6 and after, please refer to sample exam.
- Yellow highlight from "This is important." appear in the text can be valuable reference.

Index

References

[1] Ross, S. "A First Course in Probability", 9th Edition.