Probability Notes with Proofs

James Marshall Reber

December 8, 2016

Contents

Chapter 1

Opening Remarks

Proof is probably too strong a word for this. In essence, this is a slightly more suped up version of Dr. Ward's notes. The goal is to hopefully give slightly more insight to what is going on behind the scenes than what would normally be given in an introduction to Probability class. However, to really be rigorous, we would need things like measure theory and more real analysis. Hopefully you find this to be a useful study guide.

Chapter 2

Preliminaries

Claim 1. *For* $a \in \mathbb{R}$, $|a| < 1$, *we have* $1 + a + a^2 + \ldots + a^r = \sum_{j=0}^{r} a^j = \frac{1-a^{r+1}}{1-a}$.

Proof. Let $S_n = \sum_{j=0}^{r} a^j$. Then we have

$$S_r = 1 + a + \ldots + a^r \to aS_n = a + a^2 + \ldots + a^{r+1}.$$

If we then take $S_r - aS_r$, we get

$$\begin{aligned} S_r - aS_r &= (1 + a + \ldots + a^r) - (a + a^2 + \ldots + a^{r+1}) \\ &= 1 + (a - a) + (a^2 - a^2) + \ldots + (a^r - a^r) - a^{r+1} \\ &= 1 - a^{r+1}. \end{aligned}$$

We can now use the distributive law on S_n, since we're treating it as though it were a variable of sorts. This gives us $S_r(1 - a) = (1 - a^{r+1})$. Dividing both sides by $1 - a$ then results in $S_r = \frac{1-a^{r+1}}{1-a}$, as required. □

Claim 2. *Using the same* a *as prior, we have* $a + a^2 + ... + a^r = \frac{a-a^{r+1}}{1-a}$.

Proof. Let's use a similar principle. Let $X_n = a + \ldots + a^r$. Then we have

$$aX_n = a^2 + \ldots + a^{r+1}.$$

We can then do

$$\begin{aligned} X_n - aX_n &= a + \ldots + a^r - (a^2 + \ldots + a^{r+1}) \\ &= a + (a^2 - a^2) + \ldots + (a^r - a^r) - a^{r+1} \\ &= a - a^{r+1}. \end{aligned}$$

Using the distributive property again leaves us with $(1 - a)X_n = (a - a^{r+1}) \leftrightarrow X_n = \frac{a-a^{r+1}}{1-a}$, as required. □

Claim 3. *Using the same* a *as prior, we have* $1 + a + \ldots = \sum_{j=0}^{\infty} a^j = \frac{1}{1-a}$.

Proof. Note that we have already derived $S_r = \sum_{j=0}^{r} a^j$. We will utilize this.
Let us take $\lim_{r\to\infty} S_r$. Using properties of limits, we can deduce that this is equivalent to asking

$$\lim_{r\to\infty} S_r = \lim_{r\to\infty} \frac{1 - a^{r+1}}{1 - a}$$

We will then use the property that $|a| < 1$. We have that $a^{r+1} \to 0$ as $r \to \infty$, since $0 < |a^r| \leq 1/r$ by principles, and we can deduce that $1/r \to 0$. By the squeeze lemma, we then know that $|a^r| \to 0$. It is then obvious that if $|a^r| \to 0$, then $a^r \to 0$ as well. So, using this fact, we have

$$\begin{aligned} \lim_{r\to\infty} S_r &= \frac{1 - (0)}{1 - a} \\ &= \frac{1}{1 - a} \end{aligned}$$

as claimed. □

Claim 4. $1 + 2 + \ldots + n = (n(n+1))/2$ *for* $n \in \mathbb{N}$.

Proof. As an aside, this is the famous theorem that is attributed to Gauss (though it is known that the Greeks utilized this, and I believe the Egyptians as well). We will proceed by induction. We must first show it holds for the base case. However, this is rather obvious, since this means $1 = (1(1+1))/2 = 1$. Now, assume it holds for n. Then we must show it holds for $n+1$. By hypothesis, we have

$$\begin{aligned} 1 + ... + n + (n+1) &= \frac{n(n+1)}{2} + (n+1) \\ &= \frac{n(n+1)}{2} + \frac{2n+2}{2} \\ &= \frac{n^2 + n + 2n + 2}{2} \\ &= \frac{n^2 + 3n + 2}{2} \\ &= \frac{(n+1)(n+2)}{2}. \end{aligned}$$

This matches the induction hypothesis, as required, and so we have that $1 + ... + n = \frac{n(n+1)}{2}$. □

Remark. *The squares version is proven essentially the same way.*

Chapter 3

Randomness

Chapter 1: Outcomes, Events, and Sample Spaces

Definition 3.0.1. Outcome: An outcome is a possible result of a probability experiment.

Definition 3.0.2. Event: An event is a set of possible outcomes.

Definition 3.0.3. Subset: A set A is a subset of a set B if for all $x \in A$, $x \in B$. It is denoted by $A \subset B$.

Definition 3.0.4. Set Equality: A set, A, is equal to another set, B, if $A \subset B$ and $B \subset A$.

Definition 3.0.5. Outcome: We say that the union of two sets, A and B, is the set of all elements x in either A or B. It is denoted by $A \cup B$.

Definition 3.0.6. Intersection: We say that the intersection of two sets, A and B, is the set of all elements x in both A and B. It is denoted by $A \cap B$.

Definition 3.0.7. Complement: Denote S to be the entire sample space. Then the complement of a set $A \subset S$, denoted A^c, is all the elements $x \in S$ such that $x \notin A$.

Definition 3.0.8. Set-Theoretic Difference: Let A and B be two sets. Then the set-theoretic difference, denoted $A \setminus B$, is the set of all the elements $x \in A$ in which $x \notin B$.

Theorem 1. *(DeMorgan's Laws) For a finite or infinite collection of events, A_1, A_2, ..., we have*

$$\big(\bigcup_j A_j\big)^c = \bigcap_j A_j^c, \tag{3.1}$$

and

$$\big(\bigcap_j A_j\big)^c = \bigcup_j A_j^c, \tag{3.2}$$

where j denotes the index of the set.

Proof. We will show it for a collection of two sets. You can use (trans)finite induction to conclude this for the uncountable infinite collection of sets.
This will be shown by using proof by contradiction. Suppose $x \in A^c \cup B^c$ but $x \notin (A \cap B)^c$. If $x \notin (A \cap B)^c$, then $x \in A \cap B$. So it then follows that $x \in A$ and $x \in B$. But this means that $x \notin A^c$ or $x \notin B^c$. So, this implies that $x \notin A^c \cup B^c$, which is a contradiction. Therefore, we have that $(A \cap B)^c \subset (A \cup B)^c$, per definition.
Now, we need to show the converse. If $x \in (A \cap B)^c$, then $x \notin A \cap B$. If $x \notin A \cap B$, then we can break this into cases. If $x \in A^c$, then this means $x \notin A \cap B$, and so $x \in A^c \cup B^c$. Likewise, if $x \in B^c$, then $x \notin A \cap B$, and so $x \in A^c \cup B^c$. So, we have that $(A \cup B)^c \subset A^c \cap B^c$, and so $(A \cup B)^c = A^c \cap B^c$, per definition.
The argument for $(A \cup B)^c = A^c \cap B^c$ is exactly the same, and is left to the reader as an exercise (because I'm lazy and it takes a lot of time). $\square$

Remark. *The set of events is also known as the Power set in set theory. The size of this set is 2^n, where n is the size of the set (generally in probability it is the sample space). It's easy to note this – there are n elements in the set, per definition of size, and either the set is in the Power set or it's not. So, we have two options for each element, and so to find the total amount of events we have $2 \times 2 \times \cdots \times 2$ n-times. In other words, we have 2^n.*

Chapter 1: Exercises

Question 1: A die is rolled, find the event that an even number is obtained.
Question 2: Two friends are playing a board game that requires each of them to roll a die. Each player uses her/his own die. If their dice are painted two different colors, what is the sample space?
Question 3: Alice, Bob, Catherine, Doug, and Edna are randomly assigned seats at a circular table in a perfectly circular room. Assume that rotations of the table do not matter, so there are exactly 24 possible outcomes in the sample space. Alice, Catherine, and Edna are sisters. Sisters love to sit together. In how many of these 24 outcomes are all three sisters sitting together (in an adjacent cluster) and are therefore happy?
Question 4: Six rocks are sitting in a straight line. We paint them, using up to three colors (say,R's, W's, and B's). If the colors of the rocks are listed, left-to-right, then one possible outcome is (R; B; B; B; W; B). Another possible outcome is (B; W; W; W; B; W). Etc. . . The sample space – consisting of all possible outcomes of the painting – has $3^6 = 729$ possible outcomes. How many of the 729 outcomes have the property that each color is used exactly two times?
Question 5: Consider a collection of 4 suite-mates. Suppose on Friday they need to buy a lot of food for the weekend, so they choose (exactly) two suite-mates to go together to the store on Friday. (You can completely ignore what happened on Wednesday.) How many outcomes are there for the pair of Friday shoppers?

Chapter 2: Probability

The Axioms of Probability:

1. Any event occurs a certain percentage of the time, and so probabilities are always between 0 and 1, inclusive. In other words, if $A \subset S$ is an event, then $0 \le P(A) \le 1$.
2. With probability 1, some outcome in the sample space occurs. In other words, if S is the sample space, then $P(S) = 1$.
3. If events have no outcome in common, then the probability of their union is the sum of the probabilities of their union is the sum of the probabilities of the individual events. In other words, if $A_i \cap A_j = \emptyset$ for all $0 \le i < j$, then $P(\cup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} P(A_j)$.

Definition 3.0.9. Disjoint: A pair of events A, B are disjoint if $A \cap B = \emptyset$.

Remark. *The empty set, $\emptyset$, is the set of no elements.*

Remark. *We will always use P to denote the probability. Also note that $P(\emptyset) = 0$.*

Theorem 2. *If $A_1, \ldots, A_n$ is a collection of finitely many disjoint events, then the probability of the union of the events equals the sum of the probabilities of the events. In other words, $P\big(\bigcup_{j=1}^n A_j\big) = \sum_{j=1}^n A_j$.*

Proof. This one is a pretty obvious one. Let all $A_j = \emptyset$ for all $j > n$. Then we have, from the axioms, $P\big(\bigcup_{j=1}^n A_j\big) = P\big(\bigcup_{j=1}^{\infty} A_j\big) = \sum_{j=1}^{\infty} A_j$. However, as we assumed, for all $j > n$, $A_j = \emptyset$ and so we have $\sum_{j=1}^n A_j$, as required. □

Theorem 3. *If a sample space S has n equally likely outcomes, then each outcome has probability $1/n$ of occurring.*

Proof. Since each probability has equal probability of occurring, let us partition the sample space into n different disjoint events. So we have $\bigcup_{j=1}^n A_j = S$. So we have, from the axioms that $1 = P(S) = P\big(\bigcup_{j=1}^n A_j\big) = \sum_{j=1}^n P(A_j)$. Since the A_j occur equally likely, then we have $\sum_{j=1}^n P(A_j) = nP(A_j)$. This finally gives us $P(A_j) = 1/n$ for all $1 \le j \le n$. □

Corollary 3.1. *If the sample space has n equally likely outcomes, and A is an event with j outcomes, then event A has probability j/n of occurring, i.e., $P(A) = j/n$.*

Proof. So, let $A_j \subset A$ such that, for all $1 \leq j \leq m$, the A_j's are disjoint and $\bigcup_{j=1}^{m} A_j = A$. Then we have $P(A) = P\big(\bigcup_{j=1}^{m} A_j\big)$, and from the axioms we have that this is $P(A) = \sum_{j=1}^{m} P(A_j)$. Since they're equally likely, this gives us $m \times P(A_j)$. From the theorem, we have $P(A_j) = 1/n$. Thus, $mP(A_j) = m/n$. □

Definition 3.0.10. Size: The number of outcomes in an event A, also called the size of A, is denoted by $|A|$.

Corollary 3.2. *If sample space S has a finite number of equally likely outcomes, then event A has probability $P(A) = |A|/|S|$, where $|S|$ and $|A|$ denotes the number of items in S and A, respectively.*

Proof. This is a direct result of the prior corollary, using the definition of the partition of a set. □

Definition 3.0.11. Partition: If a collection of nonempty events is disjoint, and the union is the entire sample space, then the collection is called a partition. If $\cup_j B_j = S$ and the B_j's are disjoint events, then the collection B_j's is called a partition.

Remark. *The probabilities of events in a partition always sums to 1.*

Theorem 4. *The complement A^c of event A has probability $P(A^c) = 1 - P(A)$.*

Proof. We know from Chapter 1 that the complement of A^c is the same thing as $S \backslash A$. Therefore, we can trivially note that $A \cup A^c = S$. Since A and A^c are disjoint, by definition, we have $P(A) + P(A^c) = P(S)$ from the axioms. Also from the axioms, $P(S) = 1$. Thus, we have $P(A) + P(A^c) = 1$, and subtracting $P(A)$ from both sides gives us $1 - P(A) = P(A^c)$. □

Theorem 5. *If $A \subset B$, then $P(A) \leq P(B)$.*

Proof. Since $A \subset B$, we can note that $A \cup (B - A) = B$. We can also note that this is, by definition, disjoint, and so using the axioms of probability we have $P(A) + P(B - A) = P(B)$. Also from the axioms, we know that $0 \leq P(B - A) \leq 1$. Thus, this means that $P(A) \leq P(B)$, by definition. □

Theorem 6. *For any finite sequence of events $A_1, ..., A_n$,*

$$P(\cup_{j=1}^{n} A_j) = \sum_{j=1}^{n} P(A_j) - \sum_{i<j} P(A_i \cap A_j) + ... + (-1)^{n+1} P(A_1 \cap ... \cap A_n).$$

Proof. We will prove this with induction. Let $n = 1$. Then we have

$$P(A_1) = P(A_1),$$

trivially. Next, assume it holds for n. We then need to show that it holds for $n + 1$. For n, we have by hypothesis

$$\begin{aligned}
P\big(\bigcup_{j=1}^{n} A_j \cup A_{n+1}\big) &= P\big(\bigcup_{j=1}^{n} A_j\big) + P(A_{n+1}) - P\big(\bigcup_{j=1}^{n} \cap A_{n+1}\big) \\
&= P\big(\bigcup_{j=1}^{n} A_j\big) + P(A_{n+1}) - P\Big(\big(\bigcup_{j=1}^{n} \cap A_{n+1}\big)\Big) \\
&= P\big(\bigcup_{j=1}^{n} A_j\big) + P(A_{n+1}) - \Big(\sum_{j=1}^{n} P(A_j \cap A_{n+1}) - \sum_{i<j} P(A_i \cap A_j \cap A_{n+1}) + \\
&\quad ... + (-1)^{n+1} P(A_1 \cap ... \cap A_n \cap A_{n+1})\Big),
\end{aligned}$$

which with some combinatorics and algebra, we get what we wanted. The steps were omitted due to requiring more theorems than were presented. For more information, see this. □

Theorem 7. *(Inclusion-Exclusion Principle (two sets)) For any two events A and B,*

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

Proof. $A \cup B = (A - B) \cup (A \cap B) \cup (B - A)$. It's trivial to note that these sets are disjoint. Therefore, applying the probability function to both sides and using the axioms of Probability gives us $P(A \cup B) = P(A - B) + P(A \cap B) + P(B - A)$. If we add and subtract $P(A \cap B)$ on the right hand side, we get $P(A \cup B) = P(A - B) + P(A \cap B) + P(B - A) + P(A \cap B) - P(A \cap B)$. Note that $P(A - B) + P(A \cap B) + P(B - A) + P(A \cap B) = P(A) + P(B)$. This is best seen when we rewrite this as $P(A - B) + P(A \cap B) + P(B - A) = P(A) + P(B) - P(A \cap B)$. Note that $P(A) + P(B)$ double counts those elements in $P(A \cap B)$, and so when we remove $P(A \cap B)$ we get $P(A \cup B)$, which, as we showed earlier, is equal to $P(A - B) + P(A \cap B) + P(B - A)$. Continuing, we have, by substitution, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, as we wanted. □

Chapter 2: Exercises

Question 6: I am out rock climbing, and the rock face has 4 easy, 7 challenging, and 3 extreme routes to get to the top. The routes are poorly marked, so I just choose on at random, with all routes equally likely. What is the probability that I do not choose an extreme route?
Question 7: A woman has 3 pairs of sneakers, 8 pairs of flip flops, 6 pairs of flats, 4 pairs of wedges, and 9 pairs of high heels. What is the probability that she selects a pair of shoes that makes her taller if she pulls a hair from her closet without looking?
Question 8: Consider the events A_1, A_2, and A_3 with the following probabilities: $P(A_i) = 1/4$ for all $i \in \{1, 2, 3\}$. $P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = 1/8$, $P(A_1 \cap A_2 \cap A_3) = 1/16$. Find the probability of $P(A_1 \cup A_2 \cup A_3)$. Do A_1, A_2, A_3 constitute as a partition of the sample space?
Question 9: Consider a collection of 3 dice. One die is red, one die is green, and one die is blue. Roll each of the dice one time. What is the probability that $R < G < B$?

Chapter 3: Independent Events

Definition 3.0.12. Event: Events A and B are called independent if $P(A \cap B) = P(A)P(B)$.

Remark. *Two events are, therefore, dependent if* $P(A \cap B) \neq P(A)P(B)$.

Remark. *Two events being independent doesn't necessarily imply that they are disjoint. Also note that the only time in which A and B are both independent and disjoint are when either* $P(A) = 0$ *or* $P(B) = 0$.

Theorem 8. *If* $A \subset B$ *and neither* $P(A) = 0$ *nor* $P(B) = 1$*, then* A, B *are dependent.*

Proof. By assumption, if $A \subset B$ then $P(A \cap B) = P(A)$. Also note that $P(B) < 1$, and so multiplying both sides by $P(A)$ preserves strict inequality. So we have $P(A)P(B) < P(A) = P(A \cap B)$. Therefore, they cannot be independent by definition. □

Theorem 9. *If neither* $P(A) = 0$ *nor* $P(A) = 1$*, then* $P(A)$ *and* $P(A^c)$ *are dependent.*

Proof. Notice that $P(A \cap A^c) = 0$ by definition of disjoint sets, however $P(A)P(A^c) = P(A) - P(A)^2 \neq 0$. Therefore, A and A^c are dependent. □

Definition 3.0.13. Independence of a finite collection of events: A finite collection of events $A_1, ..., A_n$ are called (mutually) independent if, for every sub-collection of the vents, the probability of the intersection is equal to the probabilities of the individual events in the collection.

Definition 3.0.14. Independence for an infinite collection of events: An infinite collection of events $A_1, ...$ are called independent if every finite collection of the events is independent.

Theorem 10. *When events are independent, their complements are too. In other words, if A and B are independent, then* A^c *and* B *are independent,* A *and* B^c *are independent, and* A^c *and* B^c *are independent.*

Proof. If A and B are independent then $P(A \cap B) = P(A)P(B)$. Subtracting both sides from $P(B)$ gives us

$$P(B) - P(A \cap B) = P(B) - P(A)P(B) \leftrightarrow P(B \cap A^c) = P(B)(1 - P(A))$$
$$\leftrightarrow P(B \cap A^c) = P(B)P(A^c),$$

which is what we wanted. □

Remark. *The property of independence and complements extends beyond just two events.*

Theorem 11. *Consider a sequence of independent trials, each of which can be classified as good, bad, or neutral, which happen with probabilities p, q, and $1 - p - q$ (respectively). Then the probability that something good happens before something bad happens is $\frac{p}{p+q}$.*

Proof. Since these events are independent of one another, let's consider a sequences of independent events $A_1, A_2, ...$, where the n-th event indicates that something *good* happens on the nth trial, and neither good nor bad things happen on the previous trials. We then compute $P(\cup_{n=1}^{\infty} A_n)$. The A_n's are disjoint, since something good cannot happen for the first time on two different trials. Since the A_n's are disjoint, due to properties of probability, we have

$$P\big(\bigcup_{n=1}^{j} A_n\big) = \sum_{n=1}^{\infty} P(A_n).$$

Note that if A_n denotes a good event following $n - 1$ neutral events, and so we have $P(A_n) = (1 - p - q)^{n-1}p$. Therefore, the desired probability is

$$P\big(\bigcup_{n=1}^{\infty} A_n\big) = \sum_{n=1}^{\infty}(1 - p - q)^{n-1}p$$
$$= \frac{1}{1 - (1 - p - q)}p$$
$$= \frac{p}{q + p},$$

as required. □

Chapter 3: Exercises

Question 10: Jack and Jill are independently struggling to pass their last class required for graduation. Jack needs to pass Calculus 3, but he only has a probability of 0.3 of passing. Jill needs to pass Organic Chemistry, but she only has a probability of 0.46 of passing. They work independently. What is the probability that at least one of them gets a diploma?

Question 11: A matching pair of blue gloves, a matching pair of red gloves, and one lone white right-handed glove are in a drawer. The gloves are pulled out of the drawer, one at a time. Suppose that a person is looking for the white glove. He repeatedly does the following: He pulls out a glove, checks the color, and if it is white, he stops. If it is not white, then he replaces the glove in the drawer and starts to check again, i.e., he reaches blindly into the drawer of 5 gloves. He continues to do this over and over until he finds the white glove, and then he stops. Let A denote the event that he pulls out a red glove during this process. In other words, A denotes the event that he finds a red glove before a white glove. Find P(A).

Question 12: Suppose that 11% of albums sold are country music; 15% are pop; 17% are R&B; and 29% are rock. There are several other kinds of genres not listed here. Suppose that we spoke to people about their music choices, and assume that each person has independent music preferences. If we continue talking to people until we find someone whose top music choice is one of the four genres above what is the probability that this person prefers rock?

Question 13: Consider a red 4-sided die (numbered 1, 2, 3, 4), a green 4-sided die (also 1 to 4), and a blue 6-sided die (1 to 6). Roll the three die (simultaneously) until the *sum* of the three dice equals 5, and then stop afterwards. On the final role of the dice, what is the probability that the red and green dice have the same values?

Question 14: Consider a sample space S with eight outcomes: $S = \{a, b, c, d, e, f, g, h\}$. Suppose that each outcome is equally likely to appear. Now define the events $A = \{a, b, c, d\}$ and $B = \{c, d, e, f\}$. Are A and B independent events?
Question 15: Is it possible to have two independent events A and B with the property that $P(A) + P(B) > 1$?

Chapter 4: Conditional Probability

Definition 3.0.15. Conditional Probability: The conditional probability of event A, given event B is written as $P(A|B)$. In general, if event B has nonzero probability, then the conditional probability $P(A|B)$ of A given B is defined as

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

Equivalently, $P(A \cap B) = P(B)P(A|B)$.

Theorem 12. *If $P(B) > 0$, then A and B are independent if and only if $P(A) = P(A|B)$.*

Proof. Assume $P(B) > 0$ and A and B are independent. Then we have $P(A)P(B) = P(A \cap B)$. By definition, we have $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$, since $P(B) > 0$.
Now, assume $P(B) > 0$ and $P(A) = P(A|B)$. Then, by definition, we have $P(A) = \frac{P(A \cap B)}{P(B)}$. Assuming $P(B) > 0$, we can multiply $P(B)$ on both sides, we get $P(A)P(B) = P(A \cap B)$. By definition, this means that A and B are independent. □

Theorem 13. *For any events $A_1, \ldots,$*

$$\Big(\bigcup_j A_j\Big) \cap B = \bigcup_j (A_j \cap B),$$

and,

$$\Big(\bigcap_j A_j\Big) \cup B = \bigcap_j (A_j \cup B).$$

The unions and intersections over j can be finite, written $\cup_{j=1}^{n} A_j$ or can be infinite, $\cup_{j=1}^{\infty} A_j$.

Proof. We will prove this using set equality. Assume $x \in (\cup_j A_j) \cap B$. Then by definition, $x \in \cup_j A_j$ and $x \in B$. But that means $x \in A_1$ and $x \in B$ or $x \in A_2$ and $x \in B$, and so on. However, this means that $x \in \cup_j (A_j \cap B)$. The other way is proven the exact same way. Likewise, the second part is proven the exact same way. □

Remark. *When conditioning on any event B, the conditional probabilities satisfy all three axioms of probabilities, i.e., $P(A|B)$ is a probability distribution similar to $P(A)$.*

Theorem 14. *Consider an event B with $P(B) > 0$.*

1. *For any event A,*

$$0 \le P(A|B) \le 1;$$

2. *For the sample space S,*

$$P(S|B) = 1;$$

3. *For any disjoint events $A_1, A_2, \ldots,$*

$$P\Big(\bigcup_{j=1}^{\infty} A_J | B\Big) = \sum_{j=1}^{\infty} P(A_j|B).$$

Proof. To see (1), first notice that $0 \leq P(A \cap B) \leq P(B) \leq 1$. This is since $A \cap B \subset B$. Dividing everything by $P(B)$ gives $0 \leq \frac{P(A \cap B)}{P(B)} \leq 1$.
For (2), we note that $P(S|B) = \frac{P(S \cap B)}{P(B)}$. However, $S \cap B = B$. Thus, we have $P(S|B) = \frac{P(B)}{P(B)} = 1$.
For (3), consider disjoint events $A_1, \ldots$. By definition of conditional probability, we have

$$P(\bigcup_{j=1}^{\infty} A_j|B) = \frac{P\Big(\big(\bigcup_{j=1}^{\infty} A_j\big) \cap B\Big)}{P(B)}.$$

By Theorem 12, we have $\big(\bigcup_{j=1}^{\infty} A_j\big) \cap B = \bigcup_{j=1}^{\infty}(A_j \cap B)$. Therefore, we have

$$P(\cup_{j=1}^{\infty} A_j|B) = \frac{\cup_{j=1}^{\infty}(A_j \cap B)}{P(B)}.$$

The $A_j \cap B$ are disjoint. So therefore, by axioms of probability, we have $\bigcup_{j=1}^{\infty} P(A_j \cap B) = \sum_{j=1}^{\infty} P(A_j \cap B)$. Therefore, we have,

$$\begin{aligned} P(\bigcup_{j=1}^{\infty} A_j|B) &= \frac{\sum_{j=1}^{\infty} P(A_j \cap B)}{P(B)} \\ &= \sum_{j=1}^{\infty} \frac{P(A_j \cap B)}{P(B)} \\ &= \sum_{j=1}^{\infty} P(A_j|B), \end{aligned}$$

as needed. □

Chapter 4: Exercises

Question 16: In the PGA, on par 3 holes, golfers hit the green in one shot 80% of the time. In fact, 20% of the time, they hit the green in one shot, and then need only one putt to complete the hole; so 60% of the time, they hit the green in one shot but are unsuccessful on their putt. What is the probability that a PGA golfer only needs one putt, given that she/he hits the green in one shot?
Question 17: Suppose that a drawer contains 8 marbles: 2 are red, 2 are blue, 2 are green, and 2 are yellow. The marbles are rolling around in a drawer, so that all possibilities are equally likely when they are drawn. Alice chooses 2 marbles without replacement, and then Bob chooses 2 marbles. Let A denote the event that Alice's 2 marbles have a matching color. Find $P(B|A^c)$, i.e., given that Alice's marbles do *not* have a matching color, find the probability that Bob's marbles have a matching color.
Question 18: Suppose two 6-sided dice are rolled, and the sum is 8 or larger. What is the conditional probability that at least one value of 4 appears on the dice?
Question 19: Alice, Bob, Catherine, Doug, and Edna are randomly assigned seats at a circular table in a perfectly circular room. Assume that rotations of the table do not matter, so there are exactly 24 possible outcomes in the sample space. Bob and Catherine are married. Doug and Edna are married. Given that Bob and Catherine are sitting next to each other, find the conditional probability that Doug and Edna are sitting next to each other.

Chapter 5: Bayes' Theorem

Remark. *$P(A \cap B)$ can be written in two ways according to conditional probability: $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$.*

Theorem 15. *For any two events A and B with nonzero probabilities,*

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

Proof. From the remark, we know that $P(A)P(B|A) = P(B)P(A|B)$. Divide both sides by $P(B)$ to get Bayes' Theorem. □

Remark. *We are often not given $P(B)$ directly, and we have to decompose $P(B)$ by writing:*

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

Theorem 16. *If $P(B) \neq 0$ and $0 < P(A) < 1$, then,*

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

Proof. This essentially results from the remark. □

Remark. *A generalization of the prior remark for infinite or finite cases is of the form:*

$$P(B) = \sum_{i=1}^{n} P(A_i \cap B),$$

where the $n \to \infty$. This leads to the next theorem

Theorem 17. *If $P(B) > 0$ and A_i form a partition of S with all $P(A_i) > 0$¡ then*

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$$

where $n \to \infty$.

Remark. *For any events $A_1, .., A_n$ with $A_1 \cap ... \cap A_n \neq 0$,*

$$P(A_1 \cap ... \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)...P(A_n|A_1 \cap ... \cap A_{n-1}).$$

Chapter 5: Exercises

Question 20: Roll a blue die and a red die. Given that the blue die has an odd value, what is the probability that the sum of the two dice is exactly 4?
Question 21: Suppose that a drawer contains 8 marbles: 2 are red, 2 are blue, 2 are green, and 2 are yellow. The marbles are rolling around in a drawer, so that all possibilities are equally likely when they are drawn. Alice chooses 2 marbles without replacement, and then Bob chooses 2 marbles. Let A denote the event that Alice's 2 marbles have a matching color. Let B denote the event that Bob's 2 marbles have a matching color. Find $P(A^c|B)$.
Question 22: Suppose that there are two 6-sided dice in a hat. One die has 3 white sides and 3 black sides. The other die has 2 white sides and 4 black sides. Juanita reaches into the hat and randomly pulls out a die. She rolls the chosen die and "black" appears on top. Given this condition, what is the probability that she has chosen the die that has 3 white sides and 3 black sides?
Question 23: You go to see the doctor about an ingrown toenail. The doctor selects you at random to have a blood test for swine flu, which for the purposes of this exercise we will say is currently suspected to affect 1 in 10,000 people. The test is 99% accurate. The probability of a false negative is zero. You test positive. What is the new probability that you have swine flu?

Chapter 7: Discrete Versus Continuous Random Variables

Definition (random variable): A random variable assigns a real number to each outcome in the sample space. The random variable's value is completely determined by the outcome. The random variable is a function from the sample space to the set of real numbers.

Remark. *An outcome is normally denoted by ω and $X(\omega)$ as the random variable, but often times the outcome is well-understood and the random variable is the thing we want to study. Therefore, we just write X instead of $X(\omega)$ to simplify the notation. It is implicitly understood that there is an underlying outcome ω and that X actually means $X(\omega)$.*

Definition 3.0.16. Indicator Random Variables: Random variables that are 1 when an event occurs or 0 when the event does not occur are called indicator random variables. They are also referred to as Bernoulli random variables.

Chapter 7: Exercises

Question 24: A student buys a brand new calculus textbook that has 1000 pages, each numbered with 3 digits from 000 to 999. She randomly opens the book to a page and starts to read. Assume that any of the 1000 pages are equally likely to be chosen. Let X be the page number of the chosen page. Thus, X is an integer-valued random variable between 0 and 999. Find $P(12 \leq X \leq 17)$.
Question 25: Suppose Alice rolls a 6-sided die, and Bob rolls a 4-sided die. Let X denote the minimum value on the two dice. Find $P(X = 1)$.
Question 26: Are there random variables which are neither discrete nor continuous?

Chapter 8: Probability Mass Functions and CDFs

Definition (probability mass function): If X is a random variable, the probability that X is exactly equal to x is $p_X(x) = P(X = x)$. This is called the probability mass function, or PMF. It's also sometimes referred to as the mass of X.
Definition (cumulative distribution function): If X is a random variable, the probability that X does not exceed x is written as $F_X(x) = P(X \leq x)$. This is called the cumulative distribution function, or CDF, of X.

Remark. *The mass $p_X(x) = P(X = x)$ and the CDF $F_X(x) = P(X \leq x)$ are both probabilities, and so they both take on values between 0 and 1 by axioms of probability.*

Remark. *Since discrete random variables assume only a finite or countable number of values, we can sum over all the nonzero masses, and we must get sum 1:*

$$\sum_{x:p_X(x)\neq 0} p_X(x) = 1.$$

Remark. *The CDF is a non-decreasing function. From a visual perspective, this means that as we look left to right across the plot, the CDF is always increasing or flat. So if $a \leq b$, then $F_x(a) \leq F_x(B)$.*

In other words, $\lim_{x\to\infty} F_X(x) = 1$ and similarly $\lim_{x\to-\infty} F_X(x) = 0$.

Chapter 8: Exercises

Question 27: Chris tries to throw a ball of paper in the wastebasket behind his back (without looking). He estimates that his chance of success each time, regardless of the outcome of the other attempts, is 1/3. Let X be the number of attempts required. If he is not successful within the first 5 attempts, he then quits, and he lets $X = 6$ in such a case. Find the mass of X, and also the CDF of X.
Question 28: Suppose that we draw cards from a standard 52-card deck, with replacement and shuffling in-between cards, until the first card with value 6, 7, 8, 9, or 10 appears, and then we stop. Let X be the number of flips needed. Find $F_X(x)$, the CDF of X, for integers $x \geq 1$.
Question 29: A randomly chosen song is from the blues genre with probability 330/27333; from the jazz genre with probability 537/27333; from the rock genre with probability 8286/27333; or from some other genre with probability 18180/27333. Let X be 1 if a randomly selected song is from the blues genre, 2 if from the jazz, 3 if rock, or -1 if otherwise. Find the mass and CDF of X.

Chapter 9: Independence and Conditioning

Definition 3.0.17. Joint probability mass function: The joint probability mass function of a pair of discrete random variables X and Y is $p_{X,Y}(x, y) = P(\{\omega | X(\omega) = x \text{ and } Y(\omega) = y\})$. In other words, $p_{X,Y}(x, y) = P(X = x \text{ or } Y = y)$.

Definition 3.0.18. Joint cumulative distribution function: The joint CDF of a pair of discrete random variables X and Y is $F_{X,Y}(x, y) = P(X \leq x \text{ and } Y \leq y)$.

Remark. *Since the joint mass and the joint CDF are probabilities, then they lie between 0 and 1, inclusive, by axioms of probability. The join mass, summed over all x's and y's, takes the probabilities from the whole sample space into account, so* $\sum_x \sum_y p_{X,Y}(x, y) = 1$.

Remark. *The mass of X can be calculated by summing the joint mass over all possible values of Y. In other words,* $p_X(x) = \sum_y p_{X,Y}(x, y)$. *Likewise,* $p_Y(y) = \sum_x p_{X,Y}(x, y)$.
One could also find the CDF of X, by taking the limit as $y \to \infty$ in the joint CDF. In other words $F_X(x) = \lim_{y\to\infty} F_{X,Y}(x, y)$. *Intuitively speaking, when taking $y \to \infty$ in the joint CDF, we are letting $Y < \infty$, i.e., we let Y take on any value, since we only want to extract information about X.*

Definition 3.0.19. Independent discrete random variables:

1. **Joint mass factors into a function of x times a function of y**. These functions of x and y can also be normalized so they are masses of X and Y, respectively. In other words, $p_{X,Y}(x, y) = p_X(x)p_Y(y)$, for all x and y.
2. **Joint CDF factors into a function of x times a function of y**. These functions of x and y can also be normalized so they are the CDFs of X and Y respectively. In other words, $F_{X,Y}(x, y) = F_X(x)F_Y(y)$, for all x and y.

Theorem 18. *Consider events A and B. Let X be an indicator for event A, i.e., $X = 1$ if A occurs and $X = 0$ otherwise (recall: this is a Bernoulli random variable). Let Y be an indicator for event B, i.e., $Y = 1$ if B occurs and $Y = 0$ otherwise. Then A and B are independent events if and only if X and Y are independent random variables.*

Proof. This is not much of a rigorous proof, but this can best be seen in the following equalities:

$$p_{X,Y}(1, 1) = P(A \cap B) \leftrightarrow P(A)P(B) = p_X(1)p_Y(1);$$
$$p_{X,Y}(1, 0) = P(A \cap B^c) \leftrightarrow P(A)P(B^c) = p_X(1)p_Y(0);$$
$$p_{X,Y}(0, 1) = P(A^c \cap B) \leftrightarrow P(A^c)P(B) = p_X(0)p_Y(1);$$
$$p_{X,Y}(0, 0) = P(A^c \cap B^c) \leftrightarrow P(A^c)P(B^c) = p_X(0)p_Y(0);$$

□

Remark. *All of the prior statements and definitions can be extended to a finite amount of random variables, i.e., they are not restricted to just two.*

Definition 3.0.20. Conditional probability mass function: The <u>conditional probability mass function</u> of a discrete random variable X, given the value of another discrete random variable Y, is

$$P_{X|Y}(x|y) = P(X = x|Y = y).$$

The conditional probability mass function of X, given the value of Y, is also referred to as "the conditional PMF of X given Y" or as "the conditional mass of X given Y."
The conditional mass is defined as the ratio of the joint mass divided by the mass of Y:

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

assuming that $P(Y = y) > 0$.

Remark. *For all random variables X and Y, we have* $p_{X,Y} = p_{X|Y}(x|y)p_Y(y)$.

Chapter 9: Exercises

Question 30: Flip the cards from a deck over, one at a time, until the whole deck has been flipped over. Let X denote the umber of cards until the first ace appears. Let Y denote the number of cards until the first queen appears. Are X and Y dependent or independent?
Question 31: Roll a four-sided die. Whatever value appears, flip exactly that many fair coins. Let X denote the number of heads that appear on the coins; let Y denote the number of tails that appear on the coins. Find the mass $p_{X,Y}(x, y)$ for all integers $x \geq 0$ and $y \geq 0$ that satisfy $1 \leq x + y \leq 4$.
Question 32: Let Alice roll a 6-sided die, and let X denote the result of her roll. Let Bob roll a pair of 4-sided dice and let Y denote the sum of the two values on his dice. Find $P(X < Y)$.
Question 33: Chris tries to throw a ball of paper into the wastebasket behind his back (without looking, what a cool guy). He estimates that his chance of success each time, regardless of the outcome of the other attempts, is 1/3. Let X be the number of attempts required. If he is not successful within the first 5 attempts, he then quits, and he lets $X = 6$ in such a case. Let Y indicate whether he makes the basket successfully within the first three attempts. Thus, $Y = 1$ if his first, second or third attempt is successful, and $Y = 0$ otherwise. Find the conditional mass of X given Y.
Question 34: Suppose that X and Y are random variables with joint probability mass function $P_{X,Y}(x, y) = (5/9)(1/2)^{x-1}(1/3)^{y-1}$ for integers $1 \leq x \leq y$. Find the probability mass function of X.

Chapter 10: Expected values of Discrete Random Variables

Definition 3.0.21. Expected value of a discrete random variable: A discrete random variable X that takes on values $x_1, ..., x_n$ has expected value

$$\mathbb{E}(X) = \sum_{j=1}^{n} x_j p_X(x_j).$$

If X takes on a countably infinite number of values $x_1, ...$ then the expected value of X is

$$\mathbb{E}(X) = \sum_{j=1}^{\infty} x_j p_X(x_j).$$

Remark.

1. *a sum over all of the possible values of X, each weighted by the probability of X taking on that value, or*
2. *a sum over all of the possible outcomes, taking the value of X from such an outcome, weighted by the probability of that outcome.*

Remark. *Consider a random phenomenon with possible outcomes $\omega_1, ..., \omega_n$. Suppose that the outcome ω_j causes random variable X to take on the value x_j. Then the discrete random variable X has expected value*

$$\mathbb{E}(X) = \sum_{j=1}^{n} x_j P(\{\omega_j\}).$$

If the random phenomenon can take on one of infinitely many possible outcomes $\omega_1, ...,$ and we again suppose that the outcome ω_j causes a random variable X to take on value x_j, then the expected value of X is

$$\mathbb{E}(X) = \sum_{j=1}^{\infty} x_j P(\{\omega_j\}).$$

Chapter 10: Exercises

Question 35: Chris tries to throw a ball of paper in the wastebasket behind his back (without looking). He estimates that his chance of success each time, regardless of the outcome of the other attempts, is 1/3. Let X be the number of attempts required. If he is not successful within the first 5 attempts, he then quits, and lets $X = 6$ in such a case. Find the expected value of X.

Question 36: Suppose Alice takes 3 cookies (without replacement) from a cookie jar that contains 5 cookies, 3 of which are chocolate, and the other 2 are non-chocolate. Let X be the number of chocolate cookies she gets. Find $\mathbb{E}(X)$.

Question 37: Consider two six sided dice. One die has 2 red, 2 green, and 2 blue sides. The other die has 3 red sides and 3 blue sides. Roll both dice, and let X denote the number of red sides that appear. Let X denote the event of a red side appearing on the dice. Find $\mathbb{E}(X)$.

Chapter 11: Expected Values of Sums of Random Variables

Theorem 19. *If $X_1, ..., X_n$ are discrete random variables with finite expected value and $a_1, .., a_n$ are constant numbers then $\mathbb{E}(a_1X_1 + ... + a_nX_n) = a_1\mathbb{E}(X_1) + ... + a_n\mathbb{E}(X_n)$.*

Proof. Consider any finite collection of discrete random variables $X_1, ..., X_n$ with finite expected values. Let $a_1, ..., a_n$ be any constants. Let S denote the underlying sample space. Let $X = a_1X_1 + ... + a_nX_n$. Then $X(\omega) = a_1X(\omega) + ... + a_nX_n(\omega)$ for each outcome $\omega \in S$. So

$$\begin{aligned}\mathbb{E}(a_1X_1 + ... + a_nX_n) &= \sum_{\omega \in S}(a_1X_1(\omega) + ... + a_nX_n(\omega))P(\{\omega\}) \\ &= \sum_{j=1}^{n} a_j \sum_{\omega \in S} X_j(\omega)P(\{\omega\}) \\ &= \sum_{j=1}^{n} a_j\mathbb{E}(X_j),\end{aligned}$$

as required. □

Corollary 19.1. *For any random variable X and any constants a and b, we have $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$.*

Proof. By definition, $\mathbb{E}(X) = \sum_x xp_X(x)$. Therefore, $\mathbb{E}(aX+b) = \sum_x(ax+b)p_X(x)$. Due to properties of sums, however, we can break this apart into $\sum_x axp_X(x)+\sum_x bp_X(x)$. By the axioms of probability, we know $\sum_x p_X(x) = 1$, and so we have $\sum_x axp_X(x) + b$. Now, we can take the a out to get $a\sum_x xp_X(x) + b$. By definition of expectation, we then have $a\mathbb{E}(X) + b$, as required. □

Theorem 20. *If X is an indicator random variable for event A, i.e., $X = 1$ if event A occurs and $X = 0$ otherwise, then the expected value of X is equal to the probability that event A occurs. In other words, $\mathbb{E}(X) = P(A)$.*

Proof. If X is an indicator random variable, then the only case where $X = 1$ is when A occurs. By definition of expectation, we have $\sum_x xP(X = x)$. However, we then have, for all $X \neq A$, that $P(X) = 0$, and so therefore we have then $(1)P(A) = P(A)$. Therefore, $\mathbb{E}(X) = P(A)$. □

Theorem 21. *If $X_1, ..., X_n$ are identically distributed discrete random variables, then the expected value of $X_1 + ... + X_n$ is equal to one of the expected values multiplied by n, i.e,*

$$\mathbb{E}(X_1 + ... + X_n) = n\mathbb{E}(X_1).$$

Proof. This follows directly from **Theorem 18** and **Theorem 19**. □

Chapter 11: Exercises

Question 38: Flip a coin three times. Let X denote the number of heads. Find $\mathbb{E}(X)$, where X is the number of heads.

Question 39: A standard deck of 52 cards has 13 hearts. The cards in such a deck are shuffled, and the top five cards are dealt to a player. What is the expected number of hearts that the player receives?

Question 40: A student shuffles a deck of cards thoroughly (one time) and then selects cards from the deck *without replacement* until the ace of spades appears. How many cards does the student expect to draw?
Question 41: When rolling a die, a "high value" is a 5 or 6. Roll seven dice. Let X denote the number of "high values" obtained on the seven dice altogether. Find $\mathbb{E}(X)$.
Question 42 (The meaning of life, the universe, and everything): The Infinite Improbability Drive has a probability of going where we want it to 1/100 of the time. If we were to use it seven times, what's the expected amount of times that we'd reach the destination that we'd want?

Chapter 12: Variance of Discrete Random Variables

Definition 3.0.22. Expected value of a function of a discrete random variable: If g is any function, and X is a discrete random variable that takes on values $x_1, ..., x_n$, then the expected value of $g(X)$ is

$$\mathbb{E}(g(X)) = \sum_{j=1}^{n} g(x_j)P(X = x_j).$$

Definition 3.0.23. Expected value of a function of discrete random variables: This is essentially the same as the last definition, but is a different way of thinking about it. Consider a random phenomenon with possible outcomes $\omega_1, ..., \omega_n$. Suppose that the outcome ω_j causes random variable X to take on value x_j. Then $g(X)$ has expected value

$$\mathbb{E}(g(X)) = \sum_{j=1}^{n} g(x_j)P(\{\omega_j\}).$$

Definition 3.0.24. Variance: The <u>variance</u>, σ_x^2, of a random variable X, with expected value $\mu_X = \mathbb{E}(X)$, is

$$Var(X) = \mathbb{E}((X - \mu_X)^2).$$

Theorem 22. *For any random variable X, the variance of X is non-negative:*

$$Var(X) = \mathbb{E}((X - \mu_X)^2) \geq 0.$$

Proof. Note that $(X - \mu_X)^2 = X^2 - 2X\mu_X + \mu_X^2$. Using the linearity of expected values, this gives us $Var(X) = \mathbb{E}(X^2 - 2\mu_X X + \mu_X^2) = \mathbb{E}(X) - 2\mu_X\mathbb{E}(X) + \mu_X^2$. Since $\mu_X = \mathbb{E}(X)$, this gives us $\mathbb{E}(X^2) - \mathbb{E}(X)^2$. This therefore implies that the variance can never be less than 0. □

Remark. *The equation derived in the prior proof is often used as an alternative definition to the variance.*

Definition 3.0.25. The standard deviation: For a random variable X with variance $Var(X)$, the <u>standard deviation</u> of X is $\sigma_X = \sqrt{Var(X)}$.

Remark. *For any random variable X and any constants a and b, we have $Var(aX+b) = a^2Var(X)$.*

****<u>WARNING</u>**: The theorems and equations upcoming all require the variables to be <u>independent</u>.**

Theorem 23. *If X and Y are independent random variables, and g and h are any two functions, then $\mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X))\mathbb{E}(h(Y))$.*

Proof. By definition of expected value, we have $\mathbb{E}(g(X)h(Y)) = \sum_x \sum_y g(x)h(y)P(X = x \text{ and } Y = y)$. Since they are independent, we have $\sum_x \sum_y g(x)h(y)P(X = x)(Y = y)$. By properties of sums, we can separate this to give us $\sum_x g(x)P(X = X) \sum_y h(y)P(Y = y)$. This, however, by definition, can be written as $\mathbb{E}(g(X))\mathbb{E}(h(Y))$. □

Corollary 23.1. *If X and Y are independent random variables, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.*

Proof. The proof is nearly the same as before, just take $g(X) = X$ and $h(Y) = Y$. □

Theorem 24. *If $X_1, ..., X_n$ are independent random variables, and $a_1, ..., a_n$ are constants, then $Var(a_1X_1 + ... + a_nX_n) = a_1^2Var(X_1) + ... + a_n^2Var(X_n)$.*

Proof. This follows by computation. We have $Var(a_1X_1+...+a_nX_n) = Var(\sum_{i=1}^{n} a_iX_i) = \mathbb{E}((\sum_{i=1}^{n} a_iX_i)^2) - (\mathbb{E}(\sum_{i=1}^{n} a_iX_i))^2$. Due to time constraints, the ending is simply $Var(a_1X_1+...+a_nX_n) = \sum_{i=1}^{n} a_ia_i\mathbb{E}(X_i^2) - \sum_{i=1}^{n} a_ia_i\mathbb{E}(X_i)\mathbb{E}(X_i) = \sum_{i=1}^{n} a_i^2Var(X_i)$. □

Corollary 24.1. *If X is any random variable, and if a and B are any constants, then $Var(aX+b) = a^2Var(X)$.*

Proof. Use the final equation given in the prior proof, and let $n = 2$, $a_1 = a$, $X_1 = X$, and $X_2 = b$ (a constant). Then the equation follows. □

Corollary 24.2. *If $X_1, ..., X_n$ are independent random variables, then $Var(X_1+...+X_n) = Var(X_1)+ ...Var(X_n)$.*

Proof. It results from the prior Theorem. A better proof will come later. □

Chapter 12: Exercises

Question 43: At a certain college, 40% of students live in a residence hall, and the other 60% of the students live off-campus. Suppose 6 students are independently selected at random, and X denotes the number of those students who live in a residence hall. Find $\mathbb{E}(X^2)$.
Question 44: Consider a collection of 9 bears. There is a family of red bears, consisting of one father bear, one mother bear, and one baby bear. There is a similar green bear family, and a similar blue bear family. We draw 5 consecutive times from this collection without replacement. We keep track of the kind of bears that we get. Let X denote the number of red bears selected. Let $Y_i = 1$ denote which bear is selected, i.e., let father bear be 1, mother bear be 2, and baby bear be 3. Find $\mathbb{E}(X^2)$.
Question 45: Suppose Alice rolls a 6-sided die, and Bob rolls a 4-sided die. Let X denote the maximum of the two dice. Find $\mathbb{E}(X^2)$.

Chapter 4

Discrete Random Variables

Chapter 14: Bernoulli Random Variables

Definition 4.0.1. Bernoulli random variable: A Bernoulli random variable is always 1 or 0 to indicate the respective probability of success or failure. It is sometimes referred to as an indicator random variable. The parameters for it are
p: the probability that the outcome is a success;
$q = 1 - p$: the probability that the outcome is a failure.
The expected value for the Bernoulli is $\mathbb{E}(X) = p$ and the variance is $Var(X) = pq$.

Deriving the Mean and Variance: By definition, we have $\mathbb{E}(X) = 0p_X(0) + 1p_X(1)$. In this case, $p_X(0) = q$ and $p_X(1) = p$, by definition, which gives us $\mathbb{E}(X) = 0 * q + 1 * p = p$. For the variance, we have $\mathbb{E}(X^2) = 0^2q + 1^2p = p$, and $\mathbb{E}(X)^2 = p^2$. So, by the alternative definition, this gives us $\mathbb{E}(X^2) - \mathbb{E}(X)^2 = p - p^2 = p(1-p) = pq$.

Chapter 14: Exercises

Question 46: Suppose that a person wins a game of chance with probability 0.4, and loses otherwise. If he wins, he earns 5 dollars, and if he loses, then he loses 4 dollars. What is his expected gain or loss?
Question 47: Consider a circular table and a collection of 6 bears. There is a pair of red bears, blue bears, and green bears, each consisting of a mother bear and a father bear. A bear couple is happy if it is sitting with it's respective mate. Find $\mathbb{E}(X^2)$.
Question 48: At a certain local restaurant, students are known to prefer Japanese pan noodles 40% of the time. Let X be an indicator for whether a randomly selected student order Japanese pan noodles. Find $\mathbb{E}(X)$ and $Var(X)$.
Question 49: Let X be the number of nights you study in a month. Assume that there are 30 days in a month. Assume that you study on a given night, with probability 0.65, independent of the other nights. What is $\mathbb{E}(X)$ and $Var(X)$?

Chapter 15: Binomial Random Variables

Remark. *A binomial random variable is a sum of Bernoulli random variables.*

Definition 4.0.2. Binomial random variable: A binomial random variable is the number of successes in a fixed number of independent trials with the same probability of success on each trial. The parameters are
n: the total number of trials
p: the probability that a given trial is a success
$q = 1 - p$: the probability that a given trial is a failure.
The mass function is $P(X = x) = \binom{n}{x}p^xq^{n-x}$. The expected value is $\mathbb{E}(X) = np$, and the variance is $Var(X) = npq$.

Remark. *The motivation for the binomial random variable is simple. Imagine we had n Bernoulli random variables, and we wanted to see what the probability was for 8 of them. Then we have tons of different arrangements of these Bernoulli random variables such that 8 of them are one – in fact, we have* $\binom{n}{8}$ *of them. Since eight are success, we have* p^8*, since they're independent, and* $n-8$ *of them are failures, so we have* q^{n-8}*. Putting this together, we can see that this gives us the probability mass function for a binomial random variable.*

Deriving the Mean and Variance: By definition, $\mathbb{E}(X) = \mathbb{E}(X_1) + ... + \mathbb{E}(X_n)$. From prior, we know that these are i.i.d Bernoulli random variables, and so this gives us $\mathbb{E}(X) = np$. Likewise, for the variance we have $Var(X) = Var(X_1) + ... + Var(X_n)$. Since these are i.i.d, we then have $Var(X) = npq$.

Chapter 15: Exercises

Question 50: Consider a deck of 15 cards containing 5 blue cards, 5 red cards, and 5 green cards. Shuffle the cards and deal all 15 of the cards out around a circular table, with one per seat. A card is isolated if it's color does not agree with either of the neighboring colors. Let X denote the number of isolated cards. Find $\mathbb{E}(X)$.
Question 51: Suppose that a person wins a game of chance with probability 0.4, and loses otherwise. If she/he wins, she/he earns 5 dollars, and if she/he loses, then she/he loses 4 dollars. Assume that he plays ten games independently. Let X denote the number of games she/he wins. What is the probability that he wins \$32 or more during the ten games?
Question 52: You draw seven cards, without replacement, from a shuffled, standard deck of 52 playing cards. Let X be the number of hearts that are selected. What is the expected number of hearts?
Question 53: A spinner in a certain game lands on 0 70% of the time and lands on 1 30% of the time. How many times must a player spin so that the probability of having at least one result of 1 exceeds 95%?

Chapter 16: Geometric Random Variables

Remark. *A Geometric random variable is the number of independent trials needed until the first success occurs. An equivalent interpretation is that a Geometric random variable is the number of independent Bernoulli random variables we need to check until the first one that indicates success.*

Definition 4.0.3. Geometric random variable: The <u>Geometric random variable</u> is the number of independent trials, each with the same probability of success, until the first success occurs. The parameters are
p: the probability that a given trial is a success
$q = 1-p$: the probability that a given trial is a failure.
A quick note is that the values of p and q must be the same for every trial. The mass for a Geometric random variable is $p_X(x) = q^{x-1}p$, the expected value is $1/p$, and the variance is q/p^2.

Derivations of the mean: By definition, the mean is $\sum_{x=1}^{\infty} x p_X(x)$. Rewriting this using the mass function, we have $p\sum_{x=1}^{\infty} x q^{x-1}$. However, we can note here that this is the derivative of q^x, and so we can rewrite this as $p\frac{d}{dq}\sum_{x=1}^{\infty} q^x$. We know from preliminaries that this is $p\frac{d}{dq}\frac{q}{1-q}$, which results in $p\frac{1}{(1-q)^2}$. However, we know that $p = 1-q$, and so we get $p/p^2 = 1/p$. The variance is left as an exercise.

Remark. *$P(X \geq j) = P(X > (j-1))$ for Geometric random variables.*

Remark. *The Geometric random variable is memoryless – which means that if we had a conditional, for example, then we can rewrite it in such a way that it is nicer. So let's say we have $P(X > 8|X > 2)$. Using conditional probability, we have $P(X > 8|X > 2) = \frac{P(X>8, X>2)}{P(X>2)}$. However, the numerator can be rewritten in the form of $P(X > 8)$, since this is equivalent. So we have $P(X > 8|X > 2) = \frac{P(X>8)}{P(X>2)}$. If $P(X > 8)$, then we know we had 8 failures, and so we have q^8 on the numerator. If $P(X > 2)$, then we know we had at least 2 failures, and so q^2 is in denominator. Therefore, we have $P(X > 8|X > 2) = \frac{q^8}{q^2} = q^{8-2} = q^6 = P(X > 6)$.*

Chapter 16: Exercises

Question 54: Matilda rolls a die until the first occurrence of 1, and then she stops. Let X denote the number of rolls until, and including, that first occurrence of 1. Find $\mathbb{E}(X)$.
Question 55: Let X be a geometric random variable with $\mathbb{E}(X) = 1/p$. Let a and b be fixed positive integers with $a < b$. Find the probability that $P(X \leq b | X > a)$.
Question 56: Suppose Alice rolls a six-sided die until she gets her first occurrence of 1 and then she stops. Let X denote the number of rolls until (and including) that first occurrence of 1. Suppose Bob flips a fair coin until he gets his first occurrence of heads, and then he stops. Let Y denote the number of flips until (and including) the first occurrence of heads. Find $P(X \geq Y)$.
Question 57: Let X be a Geometric random variable with $\mathbb{E}(X) = 3$. Let A denote the event that X is even, i.e., is a multiple of 2. Find $P(A)$.

Chapter 17: Negative Binomial Random Variables

Definition 4.0.4. Negative Binomial random variables: The Negative Binomial random variable is the random variable which measures the number of independent trials required until a certain number of successes have occurred. It can be interpreted as the sum of independent Geometric random variables. The parameters for it are
r: the desired number of successes
p: the probability that a given trial is a success
$q = 1 - p$: the probability that a given trial is a failure.
The mass for it is $P(X = x) = \binom{x-1}{r-1} q^{x-r} p^r$, the expected value is $\mathbb{E}(X) = r/p$, and the variance is $Var(X) = qr/p^2$. These are easy to derive if we take the Geometric random variables to be independent and have the same parameter.

Remark. *Let $X_1, ..., X_n$ be independent Negative Binomial random variables. Let the number of successes be, respectively, $r_1, ..., r_n$. Then $X = X_1 + ... + X_n$ is a Negative Binomial random variable with p being the same as the other variables, and $r = \sum_{i=1}^{n} r_i$. The sum has probability mass $p_X(x) = \binom{x-1}{R-1} q^{y-R} p^R$.*

Chapter 17: Exercises

Question 58: Hermionie is frustrated because she is extremely good at spells, but she is struggling to learn how to fly on her broomstick. She repeatedly tries to fly on the broomstick. Assume that her trials are independent, and a trial succeeds with probability 0.15. She conducts trials until her 4th success, and then she stops. Let X denote the number of trials that are required until (and including) her 4th success with the broomstick. What is the probability mass function?
Question 59: Suppose Jessica picks homework problems at random to practice for her midterm exam. She practices until she has solved 5 worthwhile questions, and then she quits after that. Her selections of problems are independent, each with probability 0.9 of being worthwhile. Find the probability that she solves 8 or fewer questions.
Question 60: A political company calls people to see whether they plan to watch the Clinton/Trump debate tonight. Suppose that each person they talk to has an 80% chance of watching the debate, and that these decisions are independent, from person to person. They continue making such calls until they have found 10 people who plan to watch the debate. Find the probability that they need to call 12 or more people, to achieve their goal.
Question 61: You vow to replay a tough level in Super Mario World until you win 3 times. Assume that you have a 0.25 chance of winning each time you play, and each round is independent. Let X denote the number of times you have to play. What is the probability it will take you more than an hour to win 3 times?

Chapter 18: Poisson Random Variable

Definition 4.0.5. Poisson Random Variable: The Poisson random variable counts the number of events that *actually* occurs. The parameter is λ, which is the average rate of events that occur during the specified period. The mass for the Poisson is $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$, the expected value is $\mathbb{E}(X) = \lambda$, and the variance is $Var(X) = \lambda$.

Derivation of Mean: By definition, we have

$$\mathbb{E}(X) = \sum_x xP(X = x) = \sum_x \frac{e^{-\lambda}\lambda^x}{x!}.$$

The $x = 0$ term is 0, so we drop it and this leaves

$$\mathbb{E}(X) = \sum_{x\geq 1} x\frac{e^{-\lambda}\lambda^x}{x!} = \lambda e^{-\lambda} \sum_{x\geq 1} \frac{\lambda^x}{x!}.$$

We now recognize the fact that $\sum_{x\geq 1} \frac{\lambda^x}{x!} = e^{\lambda}$. This then gives us $\lambda e^{-\lambda}e^{\lambda} = \lambda$.

Remark. *The sum of independent Poisson random variables is Poisson as well. The parameters add together (i.e. the parameter for $X = X_1 + ... + X_n$ where $X_1, ..., X_n$ are Poisson with parameters $\lambda_1, ..., \lambda_n$ has parameter $\lambda = \lambda_1 + ... + \lambda_n$.*

Remark. *If X is a Binomial random variable, we can approximate it with a Poisson random variable if n is very large and npq is relatively close to 1. Usually, we say npq should be within a factor of 10 away from 1.*

Chapter 18: Exercises

Question 62: Catherine watches raindrops hit the window. The number of raindrops that fall in a fixed period of time is Poisson with an average of 6 per minute. What is the probability that exactly 5 raindrops fall during the next one minute?
Question 63: Suppose that, during a given week, $5,000,000$ people play a lottery game. If their chances to win the lottery are independent, and if each person has probably $1/2,000,000$ of winning the lottery, find the probability that there are exactly 4 winners.
Question 64: Cars pass an intersection at an average rate of 4 cars per 5-minute interval. Let X be the number of cars that pass in the next 5 minutes. What is the pmf?
Question 65: Suppose that the number of people who get food at the salad bar has a Poisson distribution with an average of 2 people per minute. Find the probability that at least 4 people get food at the salad bar during the next 3 minutes.

Chapter 19: Hypergeometric Random Variables

Definition 4.0.6. Hypergeometric Random Variable: A hypergeometric random variable is the number of desirable items we pick when selecting some items without replacement form a mixed collection of desirable and undesirable items. It has parameters:
N: the total number of items available
M: the total number of items that are desirable
$N - M$: the total number of items that are undesirable
n: the number of items that are selected.
We have that this distribution has mass $P(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$. The expected value is $\mathbb{E}(X) = n\frac{M}{N}$ and the variance is $Var(X) = n\frac{M}{N}(1 - \frac{M}{N})\frac{N-n}{N-1}$.

Remark. *This is the first distribution we've seen so far which uses dependent trials.*

Remark. *If X is Hypergeometric, then X can be viewed as the sum of Bernoulli random variables that are dependent, where $X_j = 1$ if the jth item selected is desirable, and $X_j = 0$ otherwise. Thus,*

$$X = X_1 + ... + X_n$$

Remark. *The Binomial distribution can be used as an approximation to the Hypergeometric distribution when N is really big and n is relatively small, compared to N.*

Chapter 19: Exercises

Question 66: In a state lottery, one million tickets are issued, out of which only 50,000 are winning tickets. If a man buys 10 tickets, what is the probability that exactly one of them is a winning ticket?
Question 67: You have 5 freshmen, 6 sophomores, 10 juniors, and 2 seniors in your probability class. The teacher randomly selects 3 students to go to the board to do problems. What is the probability that all 3 students will be juniors?
Question 68: A professor estimates that, among 15 students in a seminar, 5 of them enjoyed the seminar that day, and the other 10 did not enjoy it. He interviews 3 of the people in the class, selected at random, and without replacement, and he lets X denote the number of people who enjoyed the seminar. What is $\mathbb{E}(X)$?
Question 69: Suppose X and Y are independent Geometric random variables, with $\mathbb{E}(X) = 4$ and $\mathbb{E}(Y) = 3/2$. Find the probability that X and Y are equal.
Question 70: Suppose that X and Y are independent Hypergeometric random variables that each have parameters $N = 6$, $M = 3$, and $n = 2$. What is the probability that X and Y are equal?

Chapter 20: Discrete Uniform Random Variables

Definition 4.0.7. Discrete Uniform Random Variables: The core idea behind the discrete uniform random variable is that we select a single item out of a collection, and all possibilities within this collection are equally likely. The only parameter for this random variable is N, which is the total number of possible outcomes. The mass is simply $P(X = x) = \frac{1}{N}$, the expected value is $\mathbb{E}(X) = \frac{N+1}{2}$ and the variance is $Var(X) = \frac{N^2-1}{12}$.

Chapter 22: Introduction to Counting

Remark. *One thing that generally is not explained is the intuition behind using the factorial for the different permutations of objects. The best way, in my opinion, to understand the factorial expression is when imagining people sitting in chairs.*
Let's take four people in four chairs, and consider the question of how many different ways we can seat people. Imagine that these chairs were in a line, and we take one of the four people and place them in this chair arbitrarily. In that case, we have four different combinations for the first seat. However, we are now missing a person. So, when it comes to the second chair, we have three different ways of placing a person in this seat. So, in total, we're up to $4 \cdot 3 = 12$ *different ways to arrange people in the first two seats. For the next seat, we are now missing two people, and so there are two different ways to place a person in this seat. Finally, when we got to the last seat, we only have one person left and so there's no way to really arrange these people. So, in total, we have* $4 \cdot 3 \cdot 2 \cdot 1 = 24$ *different combinations, or* $4!$ *different combinations.*

Example 1. *Let's say we have nine people and nine chairs. There are five specific people that will go in the five chairs to the left, and the remaining four will go in the other four chairs. So, we have* $5!$ *ways of arranging the five specific people, and* $4!$ *ways of arranging the remaining four. In total, there are* $5! \cdot 4! = 2,880$ *ways of arranging the people.*

Corollary 24.3. *If a sample space S has n equally likely outcomes, and A is an event with j outcomes, then event A has probability j/n of occurring, i.e., $P(A) = j/n$.*

Proof. See Chapter 2 for proof. □

Theorem 25. *Suppose j processes are happening, and the first one has n_1 possibilities, and for each such possibility there are n_2 possibilities for the second process, and so on and so forth. Then there are $n_1 ... n_j$ possibilities altogether.*

Proof. This one seems pretty obvious to me, but I'll leave this open to adding a proof to it later. □

Remark. *For consistency, we will use n as the number of objects to choose from and r as the number of objects we actually choose.*

	Sampling with replacement	Sampling without replacement
Order matters	n^r	$\frac{n!}{(n-r)!}$
Order does not matter	$\binom{n+r-1}{r}$	$\binom{n}{r}$

Chapter 20 and 22 Exercises:

Question 71: If I have 10 students, how many unique ways can 4 different students go to the board to do 4 different homework problems?
Question 72: If n couples sit in a row of chairs, what is the probability that each of the n couples sits together?
Question 73: If Alice, Barbara, and Christine (three women) and Alan, Bob, and Charlie (three men) sit in a row of chairs, what is the probability that the women all sit together (the men may or may not be in a group)?

Chapter 5

Continuous Random Variables

Chapter 24: Continuous Random Variables and PDFs

Definition 5.0.1. Continuous Random Variable: A continuous random variable takes on values within an interval of finite or infinite length.

Definition 5.0.2. Density of a Continuous Random Variable: If X is a continuous random variable and $f_X(x)$ is the density of X, then:

1. The density is always nonnegative, i.e., $f_X(x) \geq 0$ for all real numbers x.
2. The density, integrated from a to b, gives the probability that X is between a and b, i.e.,

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

Remark. *If X is a continuous random variable, and if $f_X(x)$ is the density of X, then*

$$P(-\infty < X < \infty) = 1,$$

so the integral over the whole real line must be 1:

$$\int_{-\infty}^{\infty} f_X(x)dx = 1.$$

Remark. *If X is a continuous random variable, and if a is any particular value, then X has probability 0 of being equal to that particular value. To see this, we just write*

$$P(X = a) = P(a \leq X \leq a) = \int_a^a f_X(x)dx = 0.$$

Note that this results from properties of integrals.

Remark. *The following values are the same:*

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

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

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

$$P(a < X < b) = \int_a^b f_X(x)dx.$$

This follows from the fact that $P(X = c) = 0$ for any arbitrary c.

Definition 5.0.3. The CDF: The CDF is the integral of the density, i.e.,

$$F_x(a) = \int_{-\infty}^{a} f_X(x)dx,$$

so the density is the derivative of the CDF:

$$f_X(x) = \frac{d}{dx}F_X(x).$$

The second equation only holds at x for which the CDF is differentiable.

Chapter 24: Exercises

Question 74: If X is a continuous random variable with density

$$f_X(x) = \begin{cases} \frac{1}{26}(4x+1), & 2 \le x \le 4, \\ 0 & otherwise, \end{cases}$$

what is $P(1 \le X \le 3)$?
Question 75: If X is a continuous random variable with CDF

$$F_X(x) = \begin{cases} 1 - e^{-x/4} & if x > 0 \\ 0 & otherwise, \end{cases}$$

what is the density of $f_X(x)$?
Question 76: Suppose that a random variable X has density

$$f_X(x) = \frac{3}{8}(x)(2-x)(-x)$$

for $0 \le x \le 2$, and $f_X(x) = 0$ otherwise. Find $P(X \le 1)$.

Chapter 25: Joint Densities

Definition 5.0.4. Joint probability density function: The joint probability density function – also called a joint density – of a pair of continuous random variables X and Y is $f_{X,Y}(x, y)$ and it has the following properties:

1. The joint denisty is always nonnegative, i.e.,

$$f_{X,Y}(x, y) \ge 0 \text{ for all } x, y$$

2. The joint density can be integrated to get probabilities, i.e., if A and B are sets of real numbers, then

$$P(X \in A \text{ and } Y \in B) = \int_A \int_B f_{X,Y}(x, y)dydx.$$

Definition 5.0.5. Joint cumulative distribution function: If X and Y are random variables, then the joint cumulative distribution function (also called the joint CDF) of X and Y is the probability that X does not exceed x and Y does not exceed y, i.e.,

$$F_{X,Y}(x, y) = P(X \le x \text{ and } Y \le y)$$

Remark. *If X and Y are a pair of continuous random variables, with joint density $f_{X,Y}(x, y)$, then X and Y have joint CDF*

$$F_{X,Y}(a, b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{X,Y}(x, y)dydx.$$

Remark. *If $f(x,y)$ is any nonnegative function, and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dydx = 1$, then $f(x,y)$ is the joint density for a pair of continuous random variables. We can write $f(x,y) = f_{X,Y}(x,y)$ and then associated random variables X and Y have the property that*

$$P(a \le X \le b \text{ and } c \le Y \le d) = \int_a^b \int_c^d f_{X,Y}(x,y)dydx.$$

Theorem 26. *If two continuous random variables X and Y have joint density $f_{X,Y}(x,y)$, the density of X can be retrieved by integrating over all y's:*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy.$$

Similarly, for the density of Y we have

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

Chapter 25: Exercises

Question 77: Suppose that X and Y have the joint density

$$f_{X,Y}(x,y) = 15e^{-3x-5y} \text{ for } x > 0 \text{ and } y > 0,\ f_{X,Y}(x,y) = 0 \text{ otherwise.}$$

Check that $f_{X,Y}$ is a joint density and find $P(X \le 1/2 \text{ and } Y \le 1/4)$.

Chapter 26: Independent Continuous Random Variables

Definition 5.0.6. Independent continuous random variables:

1. Joint density can be factored into a function of x times a function of y. In other words, we have $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, for all x and y.
2. Joint CDF can be factored into a function of x times a function of y. In other words, we have $F_{X,Y}(x,y) = F_X(x)F_Y(y)$, for all x and y.

Remark. *The joint density of $f_{X,Y}(x,y)$ must be defined on a rectangle, or on several rectangles arranged in a grid shape (why?). Also, we need $f_{X,Y}(x,y)$ to be factored into a product, with all of the x's in one part and all of the y's in the other part.*

Chapter 26: Exercises

Question 78: The lifetime (in years) of a music player – before it permanently fails – is a random variable X with density

$$f_X(x) = \frac{1}{3}e^{-x/3}, \text{ for x ¿ 0}$$

, and $f_X(x) = 0$ otherwise.
Consider 2 music players that are assumed to have independent lifetimes X and Y repsectively. Then find the joint density for X and Y.
Question 79: If two random variables X and Y have joint density

$$f_{X,Y}(x,y) = \frac{12}{7}(xy + x^2), \text{ for } 0 \le x \le 1, 0 \le y \le 1$$

,
and $f_{X,Y}(x,y) = 0$ otherwise, are X and Y independent? Why or why not?

Chapter 27: Conditional Distributions

Definition 5.0.7. Condition density: The conditional density of a continuous random variable X, given $Y = y$ is defined as the joint density of X and Y divided by the density of Y:

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

Equivalently, we have $f_{X,Y}(x,y) = f_Y(y)f_{X|Y}(x|y)$. In both cases, we need values for which $f_Y(y) > 0$.

Remark. *The conditional probability for independent continuous random variables operates essentially the same as in the discrete case. It's left to the reader to show why this is the case (Hint: use the definitions of indepdence and conditional probability to see this).*

Chapter 27: Exercises

Question 80: Consider a pair of random variables X, Y with constant joint density on the triangle with vertices at $(0,0), (3,0)$, and $(0,3)$. For $0 \leq y \leq 3$, find the conditional density $f_{X|Y}(x|y)$ of X, given $Y = y$.

Chapter 28: Expected Values of Continuous Random Variables

Definition 5.0.8. Expected Value of a continuous random variable A continuous random variable X with density $f_X(x)$ has expected value

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

.

Theorem 27. *If $a \leq X \leq b$, then we must also have $a \leq \mathbb{E}(X) \leq b$ too.*

Proof. We have $\mathbb{E}(X) = \int_a^b x f_X(x)dx$. Note that $\int_a^b b f_X(x)dx$, since $x \leq b$ on the interval $[a,b]$, and so $b\int_a^b f_X(x)dx$. By definition, $\int_a^b f_X(x)dx = 1$, so we have $\mathbb{E}(X) \leq b$. Preform a similar procedure for the lower bound to find $a \leq \mathbb{E}(X)$. □

Theorem 28. *If X is a continuous random variable with density*

$$f_X(x) = \frac{1}{b-a} \text{ for } a \leq x \leq b$$

, and $f_X(x) = 0$ otherwise, then

$$\mathbb{E}(x) = \frac{a+b}{2}.$$

Proof. We will just do this using the definition. We have $\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)dx = \int_a^b x(\frac{1}{b-a})dx = \frac{x^2}{2(b-a)}\Big|_{x=a}^{b}$. Substituting the values in then gives us $\mathbb{E}(X) = \frac{a+b}{2}$. □

Theorem 29. *If, for some constant $\lambda > 0$, the random variable X has density*

$$f_X(x) = \lambda e^{-\lambda x} \text{ for } x > 0$$

, and $f_X(x) = 0$ otherwise, then

$$\mathbb{E}(X) = 1/\lambda$$

.

Proof. In order to derive the expected value, we once again use the definition to get

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^{\infty} x\lambda e^{-\lambda x} dx$$

. Use integration by parts – let $u = x$, $dv = \lambda e^{-\lambda x} dx$. Then we have $du = dx$, $v = -e^{-\lambda x}$. Multiplying it out gives us

$$\mathbb{E}(X) = -xe^{-\lambda x}\Big|_{x=0}^{\infty} - \int_0^{\infty} -e^{-\lambda x} dx = \frac{e^{-\lambda x}}{-\lambda}\Big|_{x=0}^{\infty} = 1/\lambda$$

. □

Chapter 28: Exercises

Question 81: Let X be the time (in minutes) that Maxine waits for a traffice light to turn green, and let Y be the time (in minutes, at a different intersection) that Daniella waits for a traffic light to turn green. Suppose that X and Y have joint density

$$f_{X,Y}(x, y) = 15e^{-3x-5y}, \text{ for } x > 0 \text{ and } y > 0$$

, and $f_{X,Y}(x, y) = 0$ otherwise. Find the expected time that each of them wait at their lights.

Question 82: A student models the number of revolutions X on a pencil sharpener that are needed to fully sharpen his pencil, by using density

$$f_X(x) = \begin{cases} \frac{2}{3}(x - 19) \text{ for } 19 \leq x \leq 20; \\ -\frac{1}{3}(x - 22) \text{ for } 20 \leq x \leq 22; \end{cases}$$

and $f_X(x) = 0$ otherwise. Find the expected value of X.

Question 83: Let X and Y have joint density

$$f_{X,Y}(x, y) = \frac{3}{80}(x^2 + y), \text{ for } 0 \leq x \leq 2 \text{ and } 0 \leq y \leq 4$$

, and $f_{X,Y}(x, y) = 0$ otherwise. Find $\mathbb{E}(X)$.

Chapter 29: Variance of Continuous Random Variables

Definition 5.0.9. Variance: For a random variable X with expected value $\mu_X = \mathbb{E}(X)$, the variance of X is

$$Var(X) = \mathbb{E}((X - \mu_X)^2).$$

Remark. *For X with expected value $\mu_X = \mathbb{E}(X)$, the variance of X is always*

$$Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

.

Definition 5.0.10. Standard Deviation: FOr a random variable X with variance $\text{Var}(X)$, the standard deviation of X is

$$\sigma_X = \sqrt{Var(X)}$$

.

Definition 5.0.11. Expected value of the square of a continuous random variable: If X is a continuous random variable X with density $f_X(x)$, then the expected value of X^2 is

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

.

Definition 5.0.12. Expected value of a function of a continuous random variable: If X is a continuous random variable X with density $f_X(x)$, and if g is an arbitrary function, then the expected value of $g(X)$ is

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

.

Theorem 30. *If X is a continuous random variable X with density $f_X(x)$, and if a and b are constants, then the expected value of $aX + b$ is*

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$$

.

Proof. This is relatively clear; expanding out $\mathbb{E}(aX + b)$ we have

$$\mathbb{E}(aX + b) = \int_{-\infty}^{\infty} (ax + b) f_X(x) dx = a \int_{-\infty}^{\infty} x f_X(x) dx + b \int_{-\infty}^{\infty} f_X(x) dx = a\mathbb{E}(X) + b.$$

□

Definition 5.0.13. Expected values of functions of continuous random variables: If X and Y are continuous random variables with joint density $f_{X,Y}(x, y)$, and if g is any function of two variables, then

$$\mathbb{E}(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dy dx$$

.

Remark. *If X and Y are continuous random variables with joint density $f_{X,Y}(x, y)$, then the expected value of X is*

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X,Y}(x, y) dy dx$$

.

Remark. *If X and Y are continuous random variables with joint density $f_{X,Y}(x, y)$, then the expected value of $X + Y$ is*

$$\mathbb{E}(X+Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x+y) f_{X,Y}(x, y) dy dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X,Y} dy dx + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{X,Y}(x, y) dy dx = \mathbb{E}(X) + \mathbb{E}(Y)$$

Theorem 31. *If X and Y are continuous random variables with joint density $f_{X,Y}(x, y)$, and if a and b are any constants, then the expected value of $aX + bY$ is*

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$$

.

Proof. Using the definition, we have

$$\mathbb{E}(aX + bY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (ax + by) f_{X,Y}(x, y) dy dx$$

$$= a \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X,Y}(x, y) dy dx + b \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{X,Y}(x, y) dy dx$$

$$= a\mathbb{E}(X) + b\mathbb{E}(Y)$$

. □

Corollary 31.1. *If $X_1, ..., X_n$ are continuous random variables and if $a_1, ..., a_n$ are any constants, then the expected value of $a_1 X_1 + ... + a_n X_N$ is*

$$\mathbb{E}(a_1 X_1 + ... + a_n X_N) = a_1 \mathbb{E}(X_1) + ... + a_n \mathbb{E}(X_n).$$

Theorem 32. *If X and Y are* <u>independent</u> *random variables, and g and h are any two functions, then*

$$\mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X))\mathbb{E}(h(Y)).$$

Proof. Applying the definition, we have

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

. □

Corollary 32.1. *If X and Y are independent random variables, then*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Theorem 33. *If $X_1, ..., X_n$ are independent random variables, and $a_1, ..., a_n$ are constants, then*

$$Var(a_1X_1 + \cdots + a_nX_n) = a_1^2Var(X_1) + \cdots + a_n^2Var(X_n).$$

Proof. It's left as an exercise to do this directly from the definition.s □

Corollary 33.1. *If X is any random variable, and if a and b are any constants, then*

$$Var(aX + b) = a^2Var(X).$$

Corollary 33.2. *If $X_1, ..., X_n$ are independent random variables, then*

$$Var(X_1 + \cdots + X_n) = Var(X_1) + \cdots + Var(X_n).$$

Chapter 29: Exercises

Question 84: Upon arriving at a restaurant and finding that no table is available, Diana and Markus have a waiting time X (in hours) with density

$$f_X(x) = 2e^{-2x}, \text{ for } 0 < x,$$

and $f_X(x) = 0$ otherwise. Find the variance of the waiting time.
Question 85: Let X, Y have joint density

$$f_{X,Y}(x,y) = \frac{1}{36}, \text{ for } 0 \le x \le 4, 0 \le y \le 9,$$

and $f_{X,Y}(x,y) = 0$ otherwise. Let $g(x,y) = xy$. Find $\mathbb{E}(g(X,Y))$.
Question 86: Let X and Y correspond to the horizontal and vertical coordinates in the triangle with corners at $(2,0), (0,2)$, and $(0,0)$. Let $f_{X,Y}(x,y) = \frac{15}{28}(xy^2 + y)$ for (x,y) inside the triangle and 0 otherwise. Find $\mathbb{E}(XY)$.

Chapter 30: More exercises

Question 87: What is the constant k that makes the following a function a valid density?

$$f_X(x) = \begin{cases} kx^2(1-x)^7 \text{ if } 0 \le x \le 1, \\ 0 \text{ otherwise} \end{cases}$$

Question 88: Using the following probability density function, find $P(X > 3)$.

$$f_X(x) = \begin{cases} \frac{1}{x^2} \text{ if } x \ge 1, \\ 0 \text{ otherwise} \end{cases}$$

Question 89: Suppose X and Y have joint probability density function

$$f_{X,Y}(x,y) = 70e^{-3x-7y}$$

for $0 < x < y$; and $f_{X,Y}(x,y) = 0$ otherwise. For $x > 0$ find the density $f_X(x)$ of X.

Chapter 6

Named Continuous Random Variables

Chapter 31: Continuous Uniform Random Variables

Definition 6.0.1. Continuous Uniform Random Variable: We say that a random variable is a continuous uniform random variable, or often referred to as uniform, if it is constant on some interval. The variable, X is an exact position or in some cases an arrival times. The two parameters it takes are a and b, denoting the beginning and the end of the interval (respectively). The density is often given in the form $f_X(x) = \frac{1}{b-a}$ for $a \leq x \leq b$ and $f_X(x) = 0$ otherwise.

Theorem 34. *The CDF, $F_X(x)$, of a uniform random variable is*

$$F_X(x) = \begin{cases} 0, \ x < a, \\ \frac{x-a}{b-a} \ , a \leq x \leq b \\ 1, \ b < x \end{cases}$$

Proof. By definition, we have that the CDF of a function is $\int_{-\infty}^{c} f_X(x)dx = \int_{a}^{c} f_X(x)dx$. Applying this here, we have

$$\int_a^c \frac{1}{b-a} dx = \frac{c-a}{b-a},$$

as required. Thus, if $c \geq b$, then we have $\frac{b-a}{b-a} = 1$, if $c \leq a$, then we have $\frac{a-a}{b-a} = 0$, and for all inbetween we have $\frac{c-a}{b-a}$. □

Theorem 35. *The expected value of the uniform random variable is $\mathbb{E}(X) = \frac{a+b}{2}$.*

Proof. From the definition, we have

$$\mathbb{E}(X) = \int_a^b \frac{x}{b-a} dx = \frac{x^2}{2(b-a)}\bigg|_{x=a}^{b} = \frac{b^2 - a^2}{2(b-a)} = \frac{(b+a)(b-a)}{2(b-a)} = \frac{b+a}{2}.$$

□

Theorem 36. *The variance of the uniform random variable is $Var(X) = \frac{(b-a)^2}{12}$.*

Remark. *The reader should attempt to prove this themselves. This just follows from the definition.*

Chapter 31: Exercises

Question 90: In a certain manufacturing process, an automated quality control computer checks 10 yards of rope at a time. If no defects are detected in that 10-yard section, that portion of the rop is passed on. However, if there is a defect detected, a person will have to check the rope over more carefully to determine where the defect is. Determine the probability distribution for finding a defect, and find $P(X > 8)$.

Question 91: Suppose that the length X of a randomly selected passenger's trip in a cab is Uniformly distributed between 5 and 30 miles. The charge induced for such a trip, in dollars, is $Y = 2.50X + 3.00$. Find the expected amount of money a randomly selected customer would pay.
Question 92: Let X_1, X_2, and X_3 be independent random variables with continuous uniform distribution over the interval $[0, 1]$. Then find $P(X_1 < X_2 < X_3)$.

Chapter 32: Exponential Random Variables

Definition 6.0.2. Exponential random variable: Exponential random variables represent the waiting time for the next event to occur. The variable for this is X = time until the next event occurs. X must always be greater than or equal to 0. The parameter for the exponential random variable is λ, or the average rate. The density is

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0 \text{ otherwise} \end{cases}$$

Remark. *In essence, one can think of the Exponential random variable as the continuous version of the Geometric random variable.*

Theorem 37. *The CDF of the exponential random variable is*

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0 \ \textit{otherwise} \end{cases}$$

Proof. By definition, the CDF is

$$\int_{-\infty}^{a} f_X(x)dx = \int_0^a \lambda e^{-\lambda x} dx = -e^{-\lambda x}\Big|_{x=0}^{a} = 1 - e^{-\lambda a} = F_X(a).$$

□

Theorem 38. *The expected value is* $\mathbb{E}(X) = \frac{1}{\lambda}$

Proof. By definition, we have that the expected value is

$$\int_{-\infty}^{\infty} x f_X(x)dx = \int_0^{\infty} x\lambda e^{-\lambda x} dx.$$

Let $u = x$, $dv = \lambda e^{-\lambda x}dx$. Then we have $du = dx$ and $v = -e^{-\lambda x}$. Substituting this into the definition of integration by parts, we have

$$-xe^{-\lambda x}\Big|_{x=0}^{\infty} + \int_0^{\infty} e^{-\lambda x} dx.$$

However, note that $-xe^{-\lambda x}\Big|_{x=0}^{\infty} = 0$, and so we have

$$\int_0^{\infty} e^{-\lambda x} dx = \frac{-e^{-\lambda x}}{\lambda}\Big|_{x=0}^{\infty} = 1/\lambda,$$

as required. □

Theorem 39. *The variance is* $Var(X) = \frac{1}{\lambda^2}$.

Remark. *The proof is left as an exercise. Use the definition of variance, and the fact that* $Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

Theorem 40. *If X is an Exponential random variable with parameter λ, then for any positive constants a and b,*

$$P(X > a + b | X > a) = P(X > b).$$

Proof. By definition, $P(X > a+b|X > a) = \frac{P(X>a+b \text{ and } X>a)}{P(X>a)} = \frac{P(X>a+b)}{P(X>a)} = \frac{e^{-(a+b)\lambda}}{e^{-a\lambda}} = e^{-b\lambda} = P(X > b)$. □

Theorem 41. *If $X_1, ..., X_n$ are independent Exponential random variables with parameters $\lambda_1, ..., \lambda_n$, respectively, and we define $Z = min(X_1, ..., X_n)$, then Z is also an Exponential random variable, with parameter $\lambda_1 + ... + \lambda_n$.*

Proof. One can use induction to prove this. I will show it for the case $n = 2$.
Let $Z = min(X_1, X_2)$. Then we have $F_Z(z) = P(Z \le z) = 1 - P(Z > z) = 1 - P(X_1 > z$ and $X_2 > z) = 1 - P(X_1 > z)(X_2 > z)$ (by independence) $= 1 - e^{-\lambda_1 z}e^{-\lambda_2 z} = 1 - e^{-(\lambda_1+\lambda_2)z}$. □

Chapter 32: Exercises

Question 93: The waiting time for rides at an amusement park has an Exponential distribution with average waiting time of 1/2 an hour. Find the time 't' such that .8 of the people have waiting time t or less.
Question 94: Lily estmates that her time to fall asleep each night is approximately Exponential, with an average time of 30 minutes until she falls asleep. What is the probability that it takes her less than 10 minutes to fall asleep?
Question 95: Consider an Exponential random variable X with parameter $\lambda > 0$. Let $Y = \lfloor X \rfloor$. Find an expression for the mass of Y.

Chapter 33: Gamma Random Variables

Definition 6.0.3. Gamma random variable: The Gamma random variable is the continuous analogy to the negative binomial random variable. In short, it is the sum of r independent Exponential random variables. Heurestically, it is the waiting time until the rth event occurs. The variable is X = time until the rth event occurs, $X \ge 0$. The parameters are r =total number of events that you are waiting for and λ = the average rate. The density is defined to be

$$f_X(x) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} x^{r-1}e^{-\lambda x}, \ x > 0, \\ 0 \text{ otherwise} \end{cases}$$

where $\Gamma(r) = (r-1)!$.

Theorem 42. *The CDF is*

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}\sum_{j=0}^{r-1} \frac{(\lambda x)^j}{j!} \ \ x > 0, \\ 0 \ \textit{otherwise} \end{cases}$$

Theorem 43. *The expected value is (as expected)* $\mathbb{E}(X) = r/\lambda$.

Theorem 44. *The variance is (also as expected)* r/λ^2.

Remark. *I will not be proving these, since integration is somewhat tricky. I may come back and edit this in later.*

Chapter 33: Exercises

Question 96: The time (in minutes) until a person's flight departs at an airport has density

$$f_X(x) = \frac{1}{45}e^{-x/45}, \text{ for } x > 0,$$

and $f_X(x) = 0$ otherwise. What is the expected total waiting time of seven passengers on seven different flights?
Question 97: The time that teach customer spends in a grocery store line is Exponential with average waiting time 3 minutes. If I am the 7th customer in line, how long do I expect to wait until all 7 of us have had our groceries processed?

Question 98: An astronomer watches a meteor shower. He believes that the density of the time in between each meteor is

$$f_X(x) = 10e^{-10x}, \text{ for } x > 0$$

, otherwise, $f_X(x) = 0$. He has also estimated that there are 500 meteors in the shower. How many minutes does he expect to be watching this meteor shower?

Chapter 34: Beta Random Variables

Definition 6.0.4. Beta random variable:
The Beta distribution deals with percents, proportions, or fractions. The parameters for it are α and β, which will be given. The density is defined to be

$$f_X(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \text{ for } 0 \leq x \leq 1 \\ 0 \text{ otherwise} \end{cases}$$

Note that there is no nice shortcut for the CDF.

Theorem 45. *The expected value is* $\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}$.

Theorem 46. *The varaince is* $Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

Chapter 34: Exercises

Question 99: The proportion of people who pass a professional qualifying exam on the first try has a Beta distribution with $\alpha = 3$ and $\beta = 4$. What is the expected value of this distribution?

Chapter 35: Normal Random Variables

Definition 6.0.5. Normal random variable: The normal distribution is a theoretical frequency distribution for a random variable, and is often characterized by it's bell-shape curve. The variable is X = the actual height, weight, volume, or whatever quantity you're measuring. The parameters are $\mu_X = \mathbb{E}(X)$ = the expected value and σ_X^2 = the variance. Sometimes, however, the standard deviation is given instead.

Remark. *We say that Z is a standard Normal random variable if it has parameters 0 and 1 respectively. A standard normal random variable Z has density $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ for $-\infty < z < \infty$. However, we do not really need the density for a Normal random variable – we will, instead, normalize the variables and use a z-table to find the corresponding probability for the values $P(Z \leq z)$. Note that for continuous random variables $P(Z = z) = 0$.*

Instead of working through the theorems, it may be more beneficial to instead note how exactly one normalizes a random variable. In essence, all of the theorems in this chapter revolve around how we are able to do this.

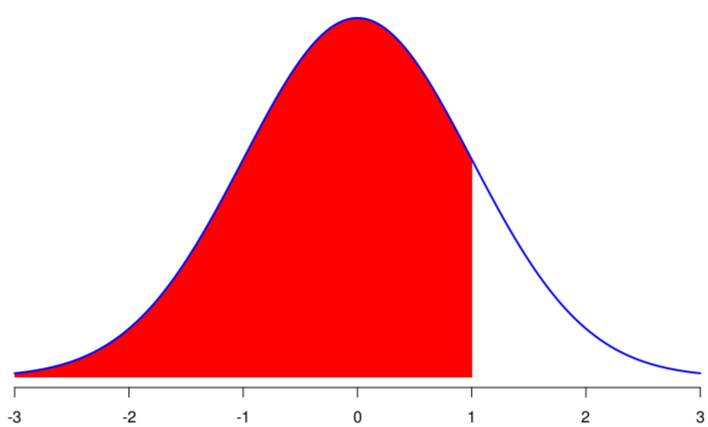
Remark. *Say you have a non-standard normal random variable, X, with mean μ_X and variance σ_X^2. If we want to find the probability that $P(X \leq a)$, we will need to first normalize it. Let Z denote the normalized version of X. Then we need to, correspondingly, normalize a. In order to do that, we preform the operation $P(Z \leq \frac{a-\mu_X}{\sqrt{\sigma_X^2}}$. We can then look for the value $\frac{a-\mu_X}{\sqrt{\sigma_X^2}}$ on a z-table, and thus we have found our probability.*

Remark. *For normal random variables, if we have a probability of the form $P(a \leq X \leq b)$ for $a < b \in \mathbb{R}$, then we can rewrite the probability to be $P(X \leq b) - P(X \leq a)$.*

Following is a z-table to refer to on the exercises:

Table 6.1: Z-table

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999



To generate a z-table like this, I used R and the following code:

```
library(xtable)
u=seq(0,3.69,by=0.01)
p=pnorm(u)
m=matrix(p,ncol=10,byrow=TRUE)
newm=xtable(m, digits=4)
print.xtable(newm, type="latex", file="nor1.tex")
```

Chapter 35: Exercises

Question 100: The quantity of sugar X (measured in grams) in a randomly selected piece of candy is Normally distributed, with expected value $\mathbb{E}(X) = \mu_X = 22$ and varaince $Var(X) = \sigma_X^2 = 8$. FInd the probability that a randomly selected piece of candy has less than 20 grams of sugar.
Question 101: Suppose that X is a Normal random variable with $\mathbb{E}(X) = 3$ and $Var(X) = 4$. Define $Y = 5X + 1$. Find $P(10 < Y < 20)$.
Question 102: Assume that the heights of college females are Normally distributed, with expected height of 64 inches and standard deviation of 4.8 inches. What is the probability that a college female is 67 inches tall or taller?

Chapter 36: Sums of Independent Normal Random Variables

Theorem 47. *If $X_1, ..., X_n$ are independent Normal random variables, with expected values $\mu_1, ..., \mu_n$, respectively, and with variance $\sigma_1^2, ..., \sigma_n^2$, respectively, then*

$$X_1, ..., X_n \text{ is also a Normal random variable}$$

with expected value $\mu_1 + ... + \mu_n$ and variance $\sigma_1^2 + ... + \sigma_n^2$.

Corollary 47.1. *If $X_1, X_2, ..., X_n$ are independent Normal random variables, which each have expected value μ and variance σ^2, then*

$$X_1 + ... + X_n \text{ is also a Normal random variable}$$

with expected value $n\mu$ and variance $n\sigma^2$.

Corollary 47.2. *If $X_1, .., X_n$ are independent Normal random variables with expected values $\mu_1, ..., \mu_n$, respectively, and with variances $\sigma_1^2, ..., \sigma_n^2$, respectively, then*

$$\frac{X_1 + \cdots + X_n - (\mu_1 + \cdots + \mu_n)}{\sqrt{\sigma_1^2 + \cdots + \sigma_n^2}} \text{ is a standard Normal random variable}$$

Corollary 47.3. *If $X_1, ..., X_n$ are independent Normal random variables, which each have expected vlue μ and variance σ^2, then*

$$\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}} \text{ is a standard Normal random variable.}$$

Chapter 36: Exercises

Question 103: A student has 23 candy sticks in a bag, with lengths that are Normally distributed. Each stick is, on average 1.8 cm long, with standard deviation 0.5 cm. What is the probability that the total length of the candy is less than 40 cm?
Question 104: A certain kind of apple weighs 150 grams, on average. Suppose that the standard deviation of this type of apple is 20 grams. When picking 66 such apples, what is the probability that they weight (altogether) 9966 grams or more?

Chapter 37: Central Limit Theorem

Remark. *Most of the proofs in this text were more of a heuristic approach to the proof. For it to be actually rigorous, we would need to utilize measure theory. For the proof of the Central limit theorem, we would actually need a good understanding of measure theory to even begin approaching a proof of it. Because of this, I will just present the theorems and results of the theorem instead of proving anything (akin to the normal random variable section).*

Theorem 48. *(The Weak Law of Large Numbers) Consider a sequence of independent random variablex $X_1, X_2 ...$, that each have finite expected value μ. Also consider any possitive number $\epsilon > 0$. The weak law of large numbers states that the average of the first n of the X_j's will not be too far from μ. More specifically, the probability that the average of $X_1, ..., X_n$ is more than ϵ away from μ will converge to 0, as $n \to \infty$. In other words*

$$\lim_{n\to\infty} P\Big(\Big|\frac{X_1 + \cdots + X_n}{n} - \mu\Big| < \epsilon\Big) = 1,$$

or equivalently,

$$\lim_{n\to\infty} P\Big(\Big|\frac{X_1 + \cdots + X_n}{n} - \mu\Big| \geq \epsilon\Big) = 0.$$

So if we fix a small ϵ, then the probability of the average $\frac{X_1+\cdots+X_n}{n}$ of the X_j's being more than ϵ away from μ will converge to 0 as $n \to \infty$.

Theorem 49. *Consider a sequence of independnet random variables $X_1, X_2, ...$, that each have finite expected value μ. The strong law of large numbers states that the average of the first n of the X_j's will converge as $n \to \infty$ to μ with probability 1. In other words, the probability that the average of $X_1, ..., X_n$ converges to μ as $n \to \infty$ is 100%. In other words*

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

Theorem 50. *Consider a sequence of independent random variables $X_1, X_2, ...$, that each have finite expected value μ and finite variance σ^2. Let $Z\ N(0,1)$ be a standard Normal random variable. THe central limit theorem states that the probability of the average $X_1, ..., X_n$, properly scaled (i.e., with subtraction of $n\mu$ and then divison by $\sqrt{n\sigma^2}$), being less than "a," will converge to the cumulative distribution function of a standard Normal random variable evaluated at a. In other words*

$$\lim_{n\to\infty} P\Big(\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}} \leq a\Big) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz,$$

or equivalently,

$$\lim_{n\to\infty} P\Big(\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}} \leq a\Big) = F_z(a)$$

Remark. *(Continuity correction rule) We can use the central limit theorem to estimate discrete random variables. The general guide is as follows:*

Strictly less than	Subtract 0.5
Less than or equal to	Add 0.5
Greater than or equal to	Subtract 0.5
Strictly greater than	Add 0.5

Table 6.2: Continuity Correction Table

Remark. *If X is a Binomial random variable with parameters n and p such that $np(1-p)$ is large, then X behaves approximately like a Normal random variable with expected value np and variance $np(1-p)$ and standard deviation $\sqrt{np(1-p)}$. Thus*

$$Z \approx \frac{X - \mu_X}{\sigma_X} = \frac{X - np}{\sqrt{np(1-p)}},$$

i.e., we approximately have a standard Normal random variable, when we standardize the X. In other words,

$$P\Big(\frac{X - np}{\sqrt{np(1-p)}} \leq a\Big) \approx P(Z \leq a) = F_Z(a).$$

Such approximations should also include continuity correction.

Remark. *If X is a Poisson random variable with parameter λ sufficiently large, e.g., $\lambda \geq 10$, then X behaves approximately like a Normal random variable with expected value λ, variance λ, and standard deviation $\sqrt{\lambda}$. Thus*

$$\frac{X - \mu_X}{\sigma_X} = \frac{X - \lambda}{\sqrt{\lambda}}$$

behaves approximately like a standard Normal random variable Z. In other words,

$$P\Big(\frac{X - \lambda}{\sqrt{\lambda}} \leq a\Big) \approx P(Z \leq a) = F_Z(a).$$

As with Binomial random variables, these approximations of Poisson random variables should also include continuity correction, resulting in a slight adjustment to the value of a.

Chapter 37: Exercise

Question 105: If the amount of rainfall in a given region is assumed to be Uniform on the interval $[0.2, 4.0]$ (in inches) each month, what is the approximate probability that there are 53 or more inches of rain during a 24-month period?

Question 106: Consider a group of students, who are assigned to work a random number of hours. Their hours per student, per week, are modeled by a Binomial random variable, with $n = 20$ and $p = 0.8$ (Each hour assigned to work will count as a "success" in the Binomial model.) If there are 100 students in the fraternity, and the number of hours spent working are independent, find an estimate for the probability that they work between 1580 and 1620 hours altogether during a given week.

Question 107: Twenty-five students want burritos, but the dining hall only has one burrito maker. Each burrito takes an average of 72.5 seconds to cook, with standard deviation 3.2 seconds. Estimate the probability that all twenty-five students can cook their burritos in half an hour or less, if we ignore the time in between the consecutive students.

Chapter 7

Extensions to Theory

Chapter 39: Variance of Sums; Covariance; Correlation

The motivation for this part is that we have such a nice fact for expected values – $\mathbb{E}(X_1 + \cdots + X_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n)$. However, this doesn't hold for variance unless it's independent. In order to solve this problem generally, we explore the topic of covariance.

Definition 7.0.1. Covariance of two random variables: The covariance of two random variables X and Y is defined as

$$\text{Cov}(X,Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))).$$

Theorem 51. *The covariance of two random variables X and Y can be expressed as*

$$Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Proof. From the definition, we have $\text{Cov}(X,Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$. Note that expanding this we have

$$\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY - Y\mathbb{E}(X) - X\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y))$$

$$= \mathbb{E}(XY) - \mathbb{E}(Y\mathbb{E}(X)) - \mathbb{E}(X\mathbb{E}(Y)) + \mathbb{E}(\mathbb{E}(X)\mathbb{E}(Y)).$$

Note that $\mathbb{E}(X\mathbb{E}(Y)) = \mathbb{E}(X)\mathbb{E}(Y)$. So, we obtain

$$\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(Y)\mathbb{E}(X) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y).$$

This simplifies to

$$\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y),$$

as required. □

Theorem 52. *If $X_1, ..., X_n$ are random variables (that are not necessarily independent), then*

$$Var(X_1 + \cdots + X_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i X_j).$$

Remark. *Note that if we set $n = 1$, we have $Cov(X,X) = \mathbb{E}((X - \mathbb{E}(X))^2) = Var(X)$.*

Corollary 52.1. *If $X_1, \ldots, X_n$ are random variables (that are not necessarily independent), then*

$$Var(X_1 + \cdots + X_n) = \sum_{i=1}^{n} Var(X_i) + \sum_{i=1}^{n}\sum_{j \neq i} Cov(X_i, X_j).$$

Theorem 53. *The covariance of X and Y equals the covariance of Y and X.*

Corollary 53.1. *If $X_1, \dots, X_n$ are random varaibles (that are not necessarily independent), then*

$$Var(X_1 + \cdots + X_n) = \sum_{i=1}^{n} Var(X_i) + 2\sum\sum_{i<j} Cov(X - i, X_j).$$

Theorem 54. *The covariance of two independent random variables X and Y is 0:*

$$Cov(X, Y) = 0.$$

Remark. *If two random variables X and Y are independent, then $Cov(X, Y) = 0$. The converse is not necessarily true, however. For instance, it is possible to have two random variables X and Y such that $Cov(X, Y) = 0$ but X and Y are not independent.*

Theorem 55. *If $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ are random variables (not necessarily independent), and $a_1, \dots, a_n$ and $b_1, \dots, b_m$ are constants, tehn*

$$Cov(a_1X_1 + \cdots + a_nX_n, b_1Y_1 + \cdots + b_mY_m) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_j\, Cov(X_i, Y_j).$$

Proof. (rough) Just brute compute this –

$$\begin{aligned}
&\mathrm{Cov}(a_1X_1 + \cdots + a_nX_n, b_1Y_1 + \cdots + b_mY_m) \\
&= \mathbb{E}(\sum_{i=1}^{n} a_iX_i \sum_{j=1}^{m} b_jY_j) - \mathbb{E}(\sum_{i=1}^{n} a_iX_i)\mathbb{E}(\sum_{j=1}^{m} b_jY_j) \\
&= \mathbb{E}(\sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_jX_iY_j) - \sum_{i=1}^{n} a_i\mathbb{E}(X_i)\sum_{j=1}^{m} b_j\mathbb{E}(Y_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_j\mathbb{E}(X_iY_j) - \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_j\mathbb{E}(X_i)\mathbb{E}(Y_j)
\end{aligned}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_j(\mathbb{E}(X + iY_j) - \mathbb{E}(X_i)\mathbb{E}(Y_j)) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_j\mathrm{Cov}(X_i, Y_j).$$

□

Corollary 55.1. *If $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ are random variables (not necessarily independent), then*

$$Cov(X_1 + \cdots + X_n, Y_1 + \cdots + Y_m) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j).$$

Corollary 55.2. *If X and Y are random variables (not necessarily independent), and a and b are coefficients, then*

$$Cov(aX, bY) = ab\,Cov(X, Y)$$

We can use covariance to define correlation.

Definition 7.0.2. Correlation: The correlation of two random variables X and Y, usually written as ρ, is defined as

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

Note: The variance is only zero when a random variable is constant. So, as long as X and Y are not constant, then the correlation between them is well-defined.

Remark. *Correlation lies between -1 and 1.*

Chapter 39: Exercises

Question 108: Consider a tray with 8 lemonades and 3 raspberry juices. Alice and Bob each take 1 drink from the tray without replacement. Assume that all of their choices are equally likely. Let X be the number of lemonades that Alice and Bob get. (Note: X is either 0, 1, or 2.) Find the variance of X.

Question 109: Suppose that X is a continuous random variable that is Uniformly distributed on $[10, 14]$, and suppose $Y = 2X + 2$. Find $\text{Cov}(X, Y)$, i.e., the covariance of X and Y.

Question 110: Suppose that X and Y have a constant joint density on the quadrilateral with vertices located at the points $(0,0), (3,0), (5,2), (0,2)$. What is the correlation?

Chapter 40: Conditional Expectation

The idea of this chapter is that given some value of a random variable Y, e.g. $Y = y$, then we want to find the conditional expectation of another random variable, e.g. X.

Definition 7.0.3. Conditional Expectation for Continuous Random Variables: If X is a continuous random variable, then we define the conditional expectation to be

$$\mathbb{E}(X \mid Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx.$$

Definition 7.0.4. Conditional Expectation for Discrete Random Variables: If X is a discrete random variable, then we define the conditional expectation to be

$$\mathbb{E}(X \mid Y = y) = \sum_x x p_{X|Y}(x|y).$$

Remark. *As another application of this formula, let A denote any event, and we define X to be a random variable such that $X = 1$ if the event is true, and $X = 0$ if the event A is false. Then we have*

$$\mathbb{E}(X) = P(A) + (1 - P(A)) \cdot 0 = P(A).$$

That is, the expected value of this random variable X is equal to the probability of the event A. Now, suppose that W is any other random variable such that observing the value of W might cause us to revise our beliefs about the probability of A. Then the conditionally expected value of X given the value of W would be similarly equal to the conditional probability of A given the value of W. That is, for any number w that is a possible value of the random variable W, we have

$$\mathbb{E}(X \mid W = w) = P(A \mid W = w).$$

Thus, the general equation

$$\mathbb{E}(X) = \sum_w w\mathbb{E}(X \mid W = w)$$

gives us the following probability equation:

$$P(A) = \sum_w wP(A \mid W = w).$$

What this means is that given the probability of A, as we assess it given our current information, must equal the current expected value of what we would think is the probability of A after learning the value of W. Learning W might cause us to revise our assessment of the probability of A positively or negatively, but the weighted average of these possible revisions, weighted by their likelihoods, must be equal to our current assessed probability of A.

Remark. *If X and Y are independent, then we have $p_{X|Y}(x|y) = p_X(x)$ and $f_{X|Y}(x|y) = f_X(x)$. Therefore, using the formula, we have*

$$\mathbb{E}(X \mid Y = y) = \sum_x x p_{X|Y}(x|y) = \sum_x x p_X(x),$$

if X is discrete, and

$$\mathbb{E}(X \mid Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) = \int_{-\infty}^{\infty} x f_X(x),$$

as expected.

Chapter 40: Exercises

Question 111: Let X and Y have a joint uniform distribution on the triangle with corners at $(0, 2)$, $(2, 0)$, and the origin (i.e. $(0, 0)$). Find $\mathbb{E}(Y \mid X = 1/2)$.
Question 112: Suppose X and Y have joint probability density function

$$f_{X,Y}(x, y) = 70e^{-3x-7y}$$

for $0 < x < y$; and $f_{X,Y}(x, y) = 0$ otherwise. Find $f_{Y|X}(y|x)$.
Question 113: Roll two 6-sided dice. Let X denote the maximum value, and let Y denote the minimum value. Find $\mathbb{E}(X|Y = 5)$.

Chapter 41: Markov and Chebychev Inequalities

Theorem 56. *(Markov's Inequality) If X is a nonnegative random variable and $a > 0$. then*

$$P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Proof. Let I be the indicator random variable for $X \geq a$. Then we have $I_{X\geq a} = 1$ if $X \geq a$ and $I_{X\geq a} = 0$ otherwise. Then, given $a > 0$, we have

$$aI_{X\geq a} \leq X$$

which will be shown by cases.
Case 1: Let $X \geq a$. Then we have $X \geq a(1)$. However, since $X \geq a$, $I_{X\geq a} = 1$. Therefore, $X \geq aI_{X\geq a}$.
Case 2: Let $X < a$. Then we have $X \geq 0$, since by assumptions X is nonnegative. However, $I_{X\geq a} = 0$. So we have $X \geq 0 = I_{X\geq a}a$.
Since we have that $\mathbb{E}$ is monotonically increasing, taking the expected value of both sides gives us

$$\mathbb{E}(aI_{X\geq a}) \leq \mathbb{E}(X).$$

Using the linearity of expectance, we have

$$a\mathbb{E}(I_{X\geq a}) \leq \mathbb{E}(X)$$

and by definition

$$a(1 \cdot P(X \geq a) + 0 \cdot P(X < a)) \leq \mathbb{E}(X).$$

This reduces to

$$aP(X \geq a) \leq \mathbb{E}(X) \leftrightarrow P(X \geq a) \leq \frac{\mathbb{E}(X)}{a},$$

as expected. □

Theorem 57. *(Chebychev's Inequality) Let X be a random variable with finite expected value μ and finite non-zero variance σ^2. Then for any real number $k > 0$,*

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Proof. Using Markov's inequality, let $Y = (X - \mu)^2$ be the random variable and let $a = (k\sigma)^2$.

Remark. *If it isn't clear at first, recall $\mathbb{E}((X - \mu)^2) = \sigma^2$.*

□

Remark. *We equivalently have*

$$P(|X - \mu| \leq k\sigma) \geq 1 - \frac{1}{k^2}.$$

What exactly does Chebychev's inequality say? It guarantees that, for a wide class of probability distributions, "nearly all" values are close to the mean. In other words, a minimum of just 75% of values must lie within two standard deviations and 89% within three standard deviations (in contrast to the 68-95-99.7 rule).

Chapter 41: Exercises

Question 114: Suppose that the average salary of a university professor on a given campus is \$82,000 a year. What is the probability that a randomly chosen professor's salary exceds \$95,000 a year?
Question 115: Suppose that in a certain course, the expected value of a student's grade is 0.80, and the standard deviation is 0.05. Find a bound on the probability that the student's grade is in the range between 0.73 and 0.87.
Question 116: Juanita carefully studies the distribution of the time between consecutive emails arriving. Suppose that the average time between consecutive emails has expected value 2.5 minutes, with a standard deviation of 1.5 minutes. Find a bound on the probability that the waiting time between two of her emails is 2 to 3 minutes.

Chapter 42: Order Statistics

Order statistics is a fundamental tool in non-parametric statistics and inference. In essence, order statistics are about understanding samples – if we have $X_{(1)}, \ldots, X_{(n)}$, then we have n samples or observations, and we organize it by smallest to largest. More info can be found here.

Definition 7.0.5. Continuous Order Statistic: Assume we have a collection of continuous random variables $X_1, \ldots, X_n$ where all X_i have the same distribution. When we talk about the order statistics, we have $X_{(i)}$, where i denotes the ith order statistic, or the ith smallest (resp. largest) statistics of the group. The general formula for the order statistics of n independent, continuous random variables that each have a common density $f_X(x)$ is

$$f_{X_{(1)}}, \ldots, X_{(n)}(x_1, \ldots, x_n) = n! f(x_1) f(x_2) \cdots f(x_n) \text{ for } x_1 < x_2 < \ldots < x_n, \text{ 0 otherwise.}$$

Definition 7.0.6. General Formula for the Density of $X_{(j)}$: The general formula is as follows:

$$f_{X_{(j)}}(x) = \binom{n}{j-1, 1, n-j} f_X(x)(F_X(x))^{j-1}(1 - F_X(x))^{n-j}.$$

Theorem 58. *The general formula for an exponential random variable is*

$$f_{X_{(j)}}(x) = \binom{n}{j-1, 1, n-j} (\lambda e^{-\lambda x})(1 - e^{-\lambda x})^{j-1}(e^{-\lambda x})^{n-j} = n\lambda e^{-n\lambda x} \text{ for } x > 0.$$

Proof. Substitute $f_X(x) = \lambda e^{-\lambda x}$ and $F_X(x) = 1 - e^{-\lambda x}$, per definition. □

Theorem 59. *The general formula for a uniform random variable is*

$$f_{X_{(j)}}(x) = \binom{n}{j-1, 1, n-j} \left(\frac{1}{b-a}\right)\left(\frac{x-a}{b-a}\right)^{j-1}\left(1 - \frac{x-a}{b-a}\right)^{n-j} \text{ for } a < x < b.$$

Proof. Substitute $f_X(x) = \frac{1}{b-a}$ and $F_X(x) = \frac{x-a}{b-a}$, per definition. □

Remark.

$$\binom{n}{j-1, 1, n-j} = \frac{n!}{(j-1)!1!(n-j)!}$$

Chapter 42: Exercises

Question 117: Consider three independent Exponential random variables X_1, X_2, X_3 each with mean 1. Find the density of $X_{(1)}$.
Question 118: Consider five independent uniform random variables $U_1, \ldots, U_5$, each uniformly distributed on $[0, 10]$. Find $U_{(4)}$.
Question 119: Consider 3 independent random variables X_1, X_2, X_3, each of which has probability density function $x/18$ for $0 < x < 6$ and 0 otherwise. Find the density of $X_{(1)}$.

Chapter 43: Moment Generating Functions

Given a random variable X, let $f(x)$ be its pdf. The quantity

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

is called the k-th moment of X. The moment generating function gives us a nice way of collecting together all the moments of a random variable X into a single power series in the variable t.

Definition 7.0.7. Moment Generating Function: We define the moment generating function to be

$$M_X := \mathbb{E}(e^{Xt}) = \mathbb{E}\Big(\sum_{k=0}^{\infty} \frac{X^k t}{k!}\Big) = \sum_{k=0}^{\infty} \frac{\mathbb{E}(X^k)t^k}{k!}.$$

Remark. *If we take the kth derivative of the moment generating function, and set $t = 0$, then the result is the kth moment, i.e.*

$$\Big(\frac{d}{dt}\Big)^k M_X(0) = \mathbb{E}(X^k).$$

Theorem 60. *If X is a binomial random variable with parameters n and p, then*

$$M_X(t) = (e^t p + 1 - p)^n$$

Proof. Note that

$$M_X(t) := \mathbb{E}(e^{tx}) = \sum_{x=0}^{n} e^{tx}\binom{n}{x}p^x(1-p)^{n-x} = \sum_{x=0}^{n}\binom{n}{x}(e^t p)^x(1-p)^{n-x} = (e^t p + 1 - p)^n.$$

□

Theorem 61. *If X is a Poisson random variable with $\mu = \lambda$, then*

$$M_X(t) = e^{(e^t-1)\lambda}.$$

Proof. Note that

$$M_X(t) := \mathbb{E}(e^{tx}) = \sum_{x=0}^{\infty} e^{tx}\frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(e^t\lambda)^x}{x!} = e^{-\lambda}e^{e^t\lambda}$$
$$= e^{(e^t-1)\lambda}.$$

□

Chapter 43: Exercises

Question 120: Suppose that X is an Exponential random variable with $\mathbb{E}(X) = 1/3$. Find the moment generating function of X.
Question 121: If X is a Geometric random variable with probability of success $p = 1/5$ on each trial, find the moment generating function of X.
Question 122: Suppose that the number of errors a student makes on her/his exam has a Poisson distribution, with an average of 3. Let X denote the number of errors. Find the moment generating function of X.

Chapter 44: Transformations of One or Two Random Variables

Remark. *(How to find the CDF and density of a function of a continuous random variable) If X is a continuous random variable and g is a function, then $Y = g(X)$ is a function of a continuous random variable. Often we know the distribution (e.g., either the density or the CDF) of X, but not the distribution of (e.g. either the density or CDF) of Y.*
To find the CDF of Y, follow these basic steps:

1. *Determine where Y is defined. In other words, find where $f_Y(y) = 0$ and where $f_Y(y) > 0$.*
2. *Focus on y in the region where Y is defined. Find $F_Y(y) = P(Y \leq y)$ by using the inverse function g^{-1} if one exists, or use a similar technique if g does not have a unique inverse.*
3. *Differentiate $F_Y(y)$ with respect to y to get the density $f_Y(y)$ of Y.*
4. *If desired, check to see that $f_Y(y)$ is really a density, i.e., that integrating $f_Y(y)$ over all possible y's gives a result of 1.*

The Distributions of Functions of Two Random Variables

Remark. *I didn't see this on Dr. Ward's website, but it was in the chapter on this so I'll leave it here just in case.*

In this section, we push the prior concepts further. If U, V are random variables that are each functions of another pair of random variables, X and Y, then we study how to get the joint density $f_{U,V}(u, v)$ of U and V from the joint density of $f_{X,Y}(x, y)$ of X and Y. The procedure is relatively straightforward every time, but solving several examples wil help improve our understanding.
First, consider two random variables, X and Y, for which the joint density $f_{X,Y}$ is known. Write U as a function of X and Y, i.e.,

$$U = g(X, Y)$$

and write V as a function of X and Y, i.e.,

$$V = h(X, Y).$$

(The functions g and h should already be given, so the above two equations do not require any work! Some exxamples of functions g and h are given in teh examples later in this section.)
Compute four partial derivatives of g and h, with respect to x and y:

$$\frac{\partial}{\partial x}g(x, y), \frac{\partial}{\partial y}g(x, y), \frac{\partial}{\partial x}h(x, y), \frac{\partial}{\partial y}h(x, y),$$

These partial derivatives should be straightforward to compute.
Write the joint density of U and V:

$$f_{U,V}(u, v) = \frac{f_{X,Y}(x, y)}{\left|\frac{\partial}{\partial x}g(x, y)\frac{\partial}{\partial y}h(x, y) - \frac{\partial}{\partial y}g(x, y)\frac{\partial}{\partial x}h(x, y)\right|}$$

.
Notice that the left-hand side has u's and v's, but the right hand side has x's and y's (because, as we noted, the x's and y's depend on u and v). So we still need to make a substitution. Before we can use such a result, we will have to solve for X and Y in terms of U and V, and substitute appropriate combinations of u's and v's for every x and y.

Example 2. *Consider random variables U, V, X, Y such that*

$$U = 1 - X - Y$$

and

$$V = X - Y$$

Suppose that the joint density $f_{X,Y}(x, y)$ of X and Y is known. Find the joint density $f_{U,V}(u, v)$ of U and V.

Notice that, in this example, the new random variables are just linear combinations of the old random variables in this example.
We write

$$U = g(X, Y) = 1 - X - Y$$

and

$$V = h(X, Y) = X - Y.$$

Then we compute four partial derivatives of g and h, with respect to x and y:

$$\frac{\partial}{\partial x} g(x, y) = -1, \frac{\partial}{\partial y} g(x, y) = -1, \frac{\partial}{\partial x} h(x, y) = 1, \frac{\partial}{\partial y} h(x, y) = -1$$

So we get

$$f_{U,V}(u, v) = \frac{f_{X,Y}(x, y)}{\left|\frac{\partial}{\partial x} g(x, y) \frac{\partial}{\partial y} h(x, y) - \frac{\partial}{\partial y} g(x, y) \frac{\partial}{\partial x} h(x, y)\right|}$$
$$= \frac{f_{X,Y}(x, y)}{|(-1)(-1) - (1)(-1)|} = \frac{1}{2} f_{X,Y}(x, y)$$

.
Finally, we need to substitute u's and v's instead of x's and y's on the right hand side of the equation. THis means that – when possible – we can just solve for the values of x and y in the equations analogous to the definitions of U and V:

$$U = 1 - X - Y$$

and

$$V = X - Y$$

. For instance, in this case, if we add U and V, we get $U + V = 1 - 2Y$, so $Y = \frac{1-U-V}{2}$. Likewise, we have $U - V = 1 - 2X$, and so $X = \frac{1-U+V}{2}$.
So, we conclude that

$$f_{U,V}(u, v) = \frac{1}{2} f_{X,Y}\left(\frac{1 - u + v}{2}, \frac{1 - u - v}{2}\right)$$

Example 3. *Consider the random variables U, V, X, Y such that*

$$U = XY$$

and

$$V = X/Y$$

. Suppose that the joint density $f_{X,Y}(x, y)$ of X and Y is known. Find the joint density $f_{U,V}(u, v)$ of U and V.
Notice that the new random variables are not just linear combinations of the old random variables in this example.
We write

$$U = g(X, Y) = XY$$

and

$$V = h(X, Y) = X/Y$$

. Then we compute four partial derivative of g and h with respect to x and Y:

$$\frac{\partial}{\partial x} g(x, y) = y, \frac{\partial}{\partial y} g(x, y) = x, \frac{\partial}{\partial x} h(x, y) = 1/y, \frac{\partial}{\partial y} h(x, y) = -xy^{-2}$$

.
So we get

$$f_{U,V}(u, v) = \frac{f_{X,Y}(x, y)}{\left|\frac{\partial}{\partial x} g(x, y) \frac{\partial}{\partial y} h(x, y) - \frac{\partial}{\partial y} g(x, y) \frac{\partial}{\partial x} h(x, y)\right|}$$
$$= \frac{f_{X,Y}(x, y)}{|(y)(-xy^{-2}) - (x)(1/y)|} = \frac{y}{2x} f_{X,Y}(x, y)$$

.

Finally, we need to substitute u's and v's instead of x's and y's on the right hand side of the equation. The fraction $\frac{y}{2x}$ is just $\frac{1}{2v}$. Also $uv = x^2$, so $\sqrt{uv} = x$ and $u/v = y^2$, so $\sqrt{u/v} = y$.
So we conclude that

$$f_{U,V}(u,v) = \frac{1}{2v} f_{X,Y}(\sqrt{uv}, \sqrt{u/v}$$

.

Chapter 44: Exercises

Question 123: A certain type of cylindrical bottle always has height 14 cm. During the manufacturing process however, the radius of the bottom is Uniformly distributed between 2.3 cm and 2.7 cm. What is the probability that the bottle has a volume of less than 275 cm^3.
Question 124: If the amount of cookie dough, X, used in a cookie is Uniformly distributed between 1.7 and 2.6 tablespoons, then the height Y of the cookie is

$$Y = \frac{3}{10} X^{1/3}$$

. Find $P(Y > 0.4)$.
Question 125: Jim cuts wood planks of length X for customers, where X is uniformly distributed between 10 and 14 feet. The price of a piece of wood is \$2 per foot, plus an at-rate surcharge of \$2 for Jim's services. So Y = 2X + 2 is the amount he charges for a piece of wood. Find $f_Y(y)$.

Chapter 8

Reference Tables

Name	Density	Domain	Expected Value	Variance	Parameters	When Used
Uniform	$1/(b-a)$	$a \leq X \leq b$	$(a+b)/2$	$(b-a)^2/12$	a, b	Over intervals
Exponential	$\lambda e^{-\lambda x}$	$x \geq 0$	$1/\lambda$	$1/\lambda^2$	λ is average number of successes	Wait time until 1st event
Gamma	$\frac{\lambda e^{-\lambda x}(\lambda x)^{r-1}}{\Gamma(r)}$	$x \geq 0$	r/λ	r/λ^2	λ is average number of successes, r is number of things	Wait time until rth event
Beta	$\frac{\Gamma(\alpha+\beta)x^{\alpha-1}(1-x)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)}$	$0 \leq x \leq 1$	$\alpha/(\alpha+\beta)$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	Usually given constraints	Bayesian statistics
Normal	$\frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}$	$-\infty < x < \infty$	μ	σ^2	μ =expected val, σ^2 = variance	Central Limit theorem and applic.

Table 8.1: Named Continuous Random Variables

Note: $\Gamma(r) = (r-1)!$.

Name	Mass	Expected Value	Variance	When Used
Bernoulli	$p_X(1) = p$, $p_X(0) = q$	p	pq	If there is 1 success or failure.
Binomial	$\binom{n}{x}p^x q^{n-x}$	np	npq	If we're measuring the amount of successes in n trials.
Geometric	$q^{x-1}p$	$1/p$	q/p^2	Measuring the amount of trials until the first success
Negative Binomial	$\binom{x-1}{r-1}q^{x-r}p^r$	r/p	qr/p^2	Measuring the amount of trials until the r-th success.
Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ	Measuring the number of events in a period.
Hypergeometric	$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$	$n\frac{M}{N}$	$n\frac{M}{N}(1-\frac{M}{N})\frac{N-n}{N-1}$	Measuring the number of good things selected.
Discrete Uniform	$\frac{1}{N}$	$\frac{N+1}{2}$	$\frac{N^2-1}{12}$	If everything is equally likely.

Table 8.2: Named Discrete Random Variables

Table 8.3: Counting equations

	Sampling with replacement	Sampling without replacement
Order matters	n^r	$\frac{n!}{(n-r)!}$
Order does not matter	$\binom{n+r-1}{r}$	$\binom{n}{r}$

Table 8.4: Random Variable Facts

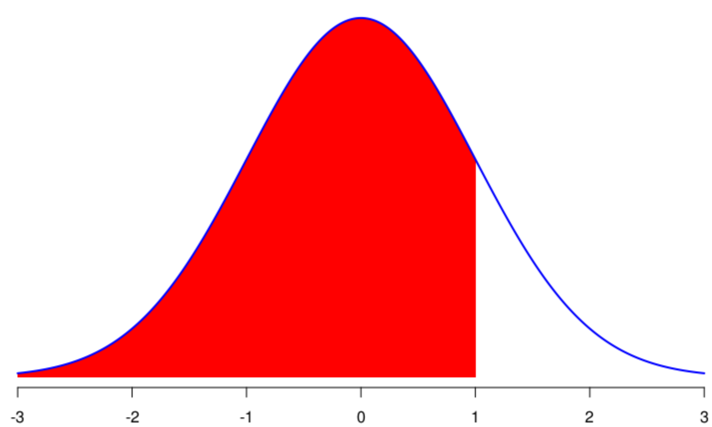
	Discrete	Continuous
Probability Function	Mass (probability mass function; PMF)	Density (probability density function; PDF)
	$0 \leq p_X(x) \leq 1$	$0 \leq f_X(x)$
	$\sum_x p_X(x) = 1$	$\int_{-\infty}^{\infty} f_X(x)dx = 1$
	$P(0 \leq X \leq 1) = P(X = 0) + P(X = 1)$ if X is integer valued	$P(0 \leq X \leq 1) = \int_0^1 f_X(x)dx$
	$P(X \leq 3) \neq P(X < 3)$ when $P(X = 3) \neq 0$	$P(X \leq 3) = P(X < 3)$
cumulative distribution function (CDF)	$F_X(a) = P(X \leq a) = \sum_{x \leq a} P(X = a)$	$F_x(a) = P(X \leq a) = \int_{-\infty}^{a} f_X(x)dx$
named distributions	Bernoulli, Binomial, Geometric, Negative, Binomial, Poisson, Hypergeometric, Discrete Uniform	Continuous Uniform, Exponential, Gamma, Beta, Normal
expected value	$\mathbb{E}(X) = \sum_x xp_X(x)$, $\mathbb{E}(g(X)) = \sum_x g(x)p_X(x)$	$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf_X(x)dx$, $\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x)f_X(x)dx$
variance	Var(X) = $\mathbb{E}(X^2) - (\mathbb{E}(X))^2$	Var(X) = $\mathbb{E}(X^2) - (\mathbb{E}(X))^2$

Strictly less than	Subtract 0.5
Less than or equal to	Add 0.5
Greater than or equal to	Subtract 0.5
Strictly greater than	Add 0.5

Table 8.5: Continuity Correction Table

Table 8.6: Z-table

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999



Chapter 9

Solutions

Solution 1: For now, we will assume (unless otherwise stated) that a die means a six-sided die (aside: there are other dice in the world). So, we need the sample space for this experiment. The sample space is $S = \{1, 2, 3, 4, 5, 6\}$. Now, an even number, by definition, is of the form $2k$ where $k \in \mathbb{Z}$. We could possibly complicate this even more by using some sort of an equivalence relation where a number is 0 iff it is equal to 0 mod(2), however in this case we can just point out that it's easier to just say the event is $E = \{2, 4, 6\}$.
Solution 2: We need to categorize the events of each die. Let x be player one's die, and let y be player two's die. Then we can note that, since it's a six-sided die, we have $1 \leq x \leq 6$ and $1 \leq y \leq 6$. It is also important to note that $x, y \in \mathbb{Z}$. Therefore, the sample space is $\{x, y — 1 \leq x, y \leq y$ and $x, y \in \mathbb{Z}\}$.
Solution 3: There are many ways to approach this problem. Let's arbitrarily choose Doug to be our man, and place Doug in a seat. Then we have four remaining seats. In order for Alice, Edna, and Catherine to sit together, we need Doug and Bob to be together. So, there are two seats next to Doug, and there are four total remaining seats. We therefore have a 1/2 chance of placing Bob in the right seat. So, we therefore have $1/2 * 24 = 12$ different events in which the sisters are happy.
Solution 4: Here, we need to utilize counting. The binomial operator is often used to count the amount of occurrences in a set. So, let's say for example you know that there are 5 vertices, and that two are red. In how many occurrences, then, do we have the two red? We would then do $\binom{5}{2}$. This can be read aloud as "5 choose 2," which is often how I remember what to do. For a general formula, we have $\binom{a}{b}$, or "a choose b", is equal to $\frac{a!}{b!(a-b)!}$. Another quick remark: the sign ! represents factorial. If we had $x \geq 0$, where $x \in \mathbb{N}$, then we define the factorial to be $x! = 1 * 2 * ... * (x - 1) * x$. We also define, for convenience, $0! = 1$.
With this, we can now properly do the problem. So we have six rocks total, and we need to make 2 of them red, let's say. Then we have $\binom{6}{2}$. Now, we need to make the remaining 4 to be blue. We then have $\binom{4}{2}$. Finally, we need the final 2 to be white. We then have $\binom{2}{2} = 1$. We then multiply these altogether to get $\binom{6}{2} * \binom{4}{2} = (15)(6) = 90$ total outcomes. This works since we arbitrarily chose the colors.
Solution 5: This is another counting problem. So we have 4 total suite-mates, and we want to select 2. To find the amount of all possibilities, we simply take $\binom{4}{2} = 6$ outcomes.
Solution 6: Let A be the event that we choose an easy route, B be the event that we choose a challenging route, and C be the event that we choose an extreme route. The questions is asking for us to find $P(C^c)$. From Theorem 4, we know that $P(C^c) = 1 - P(C)$. From Corollary 2, we can note that $P(C) = \frac{|C|}{|S|} = \frac{3}{4+7+3} = \frac{3}{14}$. Therefore, $P(C^c) = 1 - \frac{3}{14} = \frac{11}{14}$. So, the probability that I do not select an extreme route is 11/14.
Solution 7: I don't know much about shoes, but based on what Sydney tells me wedges and high heels make you taller. In that case, let A be the event she selects high heels, and B the event she selects wedges. Then we're looking for $P(A \cup B)$. Based on Corollary 2, we can note that $P(A \cup B) = \frac{|A|+|B|}{|S|}$. This is trivial to note, since we're adding the events of either selecting heels or wedges to include in the probability of either heels or wedges. This then gives us $\frac{13}{30}$ as our total probability.
Solution 8: First, let's utilize the inclusion exclusion property. So we therefore have $P(A_1 \cup A_2 \cup A_3) = \sum_{i=0}^{3} A_i - \sum_{i<j} P(A_i \cap A_j) + P(A_1 \cap A_2 \cap A_3)$. So we have $3/4 - 3/8 + 1/16 = 7/16$. They do not

constitute as a partition of the sample space, since for them to do so their union would have to be 1.
Solution 9: First, let's figure out the size of our sample space. Since we have three dice with six values each, the total number of possibilities is 6^3. Next, we have that there are (6)(5)(4) different times where $R < G < B$. So, we have $\frac{6*5*4}{6^3} = 5/9$.
Solution 10: Let A be the event that Jack passes, and let B be the event that Jill passes. Then we first note that we are looking for $P(A \cup B)$. By the inclusion-exclusion principle, we know that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. From the problem, we know that $P(A) = 0.3$ and $P(B) = 0.46$. Due to the definition of independence, we have that $P(A \cap B) = P(A)P(B)$. So, in total we have $P(A \cup B) = P(A) + P(B) - P(A)P(B)$. Putting in numbers, this gives us

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A)P(B) \\ &= 0.3 + 0.46 - (0.3)(0.46) \\ &= 0.622. \end{aligned}$$

In other words, there is 0.622 chance that at least one of them passes.
Solution 11: We have a *bad* event, a *good* event, and a *neutral* event, which are (respectively) finding a white glove, finding a red glove, and finding a blue glove. Using Corollary 2, the probability of him finding a red glove is 2/5 and the probability of finding a white glove is 1/5 (in this context, we don't care about the probability of finding a blue glove, but if we were to it would be 2/5). Using Theorem 10, then, we have $\frac{2/5}{2/5+1/5} = \frac{2/5}{3/5} = \frac{2}{3}$.
Solution 12: Let's take our good event to be if someone likes rock, our neutral event to be anything that's not one of the genres we are choosing from, and our bad event to be one of the other three genres. Then we have, using Theorem 10, that the probability of someone preferring rock is $\frac{.29}{(.11+.15+.17+.29)} = 0.402\bar{7} \approx 0.4028$.
Solution 13: We need to consider the good, bad, and neutral events. A good event is when the sum of the three dice equals 5 and the red and green dice have the same values, a bad event is when the sum of the three dice equals 5 but the red and green dice do not have the same values, and a neutral event is when the dice do not have a sum of 5. So, the red and green dice have the same values in the case of $\{(1,1,3),(2,2,1)\}$. So, in total there is a probability of 2/96 for the good event. For the bad events, we have $P(\{(1,3,1),(3,1,1),(2,1,2),(1,2,2)\}) = 4/96$. Therefore, using Theorem 10, we have $\frac{2/96}{(4/96+2/96)} = \frac{2/96}{6/96} = \frac{1}{3}$.
Solution 14: Since each outcome is equally likely, using Corollary 2 we have $P(A) = \frac{4}{8} = 1/2$, $P(B) = \frac{4}{8} = 1/2$, and $P(A \cap B) = \frac{2}{8} = \frac{1}{4}$. So, by definition, A and B are independent, since $P(A)P(B) = (\frac{1}{2})(\frac{1}{2}) = \frac{1}{4} = P(A \cap B)$.
Solution 15: Yes, for example let $P(A) = 0.7$, $P(B) = 0.4$, $P(A \cap B) = 0.28$, and $P(A \cup B) = 0.82$.
Solution 16: Let B be the event that she/he hits the green in one shot, and A be the event that she/he only needs one putt. Then we have $P(A|B)$. By definition, this is equivalent to $P(A|B) = \frac{P(A \cap B)}{P(B)}$. $P(B)$ is just the probability that she/he hits the green in one shot, so from the question we know that this is $P(B) = 0.8$. Also from the question, we know that $P(A \cap B) = 0.2$. Therefore, we have $\frac{0.2}{0.8} = \frac{1}{4}$.
Solution 17: This one will not follow directly from the definition, but rather from an intuitive point of view. We have that, if event A^c occurs, there remains four marbles which have pairs. Therefore, Bob has a 4/6 chance of selecting one of those marbles. Then, given that he's selected this marble, there is one other marble which matches the original one, and so there is a 1/5 chance of selecting that one. So, in total, $P(B|A^c) = (4/6)(1/5) = 2/15$.
Solution 18: Let A be the event that at least one value of 4 appears, and let B be the event that the sum is 8 or larger. Then we have $P(A|B) = \frac{P(A \cap B)}{P(B)}$, by definition. It's then left to the reader to find that $P(B) = 15/36$. Note that if the sum is greater than or equal to 8, then we have that four appears five times in these sums, and so we have $P(A \cap B) = 5/36$. So, using the definition, we have $P(A|B) = \frac{5/36}{15/36} = \frac{5}{15} = \frac{1}{3}$.
Solution 19: Let D denote Doug, E denote Edna, and A denote Alice. We can therefore note that we have six different combinations: ADE, AED, DEA, DAE, EAD, EDA. So, in four of the six, we have that Doug and Edna are sitting together, and so we have that the probability is 2/3.
Solution 20: Let A be the event that the sum of the two dice is exactly 4 and B be the event that the blue die has an odd value. Then, by definition of conditional probability, we have $P(A|B) = \frac{P(A \cap B)}{P(B)}$. $P(B)$ is trivial to find – using Corollary 2, we have $P(B) = \frac{|B|}{|S|}$, and we have that $|S| = 6$ and

$B = \{1, 3, 5\}$, so $P(B) = \frac{3}{6} = \frac{1}{2}$. Next, we need to find $P(A \cap B)$. If the blue die is odd, then we have $x + 1 = 4$ and $x + 3 = 4$. This gives us only two values to care about, if the red die is three and the blue die one, and the red die is one and the blue die is three. So, we have, by Corollary 2, $P(A \cap B) = \frac{2}{36} = \frac{1}{18}$. Therefore, by substitution we get $\frac{1/18}{1/2} = \frac{1}{9}$.

Solution 21: By definition of Bayes' Theorem, we have $P(A^c|B) = \frac{P(A^c)P(B|A^c)}{P(B)}$. Let Alice pick one marble arbitrarily. Then the probability of her not getting a marble that matches is 6/7. We have that Bob needs to get two matching marbles, and there are 6 remaining marbles, two of which do not have matching pairs because of Alice. So, we therefore have that Bob has a probability of getting a matching marble is $(4/6)(1/5) = P(B|A^c)$.

Bob goes second, and so the probability of him getting two matching marbles is $P(B) = (1/7)(1/5) + (4/6)(1/5)(6/7)$. So, by substitution, we have $P(A^c|B) = \frac{(6/7)(4/6)(1/5)}{(1/7)(1/5)+(4/6)(1/5)(6/7)} = 4/5$.

Solution 22: Let A be the event of selecting the die with 3 white sides and 3 black sides, and let B be the event of rolling a black. Then we have $P(A|B) = \frac{P(A \cap B)}{P(B)}$. $P(A \cap B)$ is, by definition, $(3/6)(1/2)$, and $P(B)$ is the probability of getting a black side on both dice, added together. So we have $P(B) = (\frac{1}{2})(\frac{1}{2}) + (\frac{1}{2})(\frac{2}{3})$. By substitution, this gives $P(A|B) = \frac{(\frac{3}{6})(\frac{1}{2})}{(\frac{1}{2})(\frac{1}{2})+(\frac{1}{2})(\frac{2}{3})} = 3/7$.

Solution 23: Let A be the event that you have swine flu, and B be the event that you test positive. Then we have $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$ by Bayes' Theorem. This can be rewritten to be $\frac{P(A)P(B|A)}{P(B|A)P(A)+P(B|A^c)P(A^c)}$, by the remark. So, from the question, we have that $P(A) = \frac{1}{10000}$. By a theorem, we know $P(A^c) = 1 - \frac{1}{10000} = \frac{9999}{10000}$. $P(B|A)$ is the probability that we test positive given that we have swine flu, which from the question is 0.99. $P(B|A^c)$ is the probability of a false positive, which we know is 0.01. So, by substitution, we have

$$\frac{P(A)P(B|A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} = \frac{(\frac{1}{10000})(0.99)}{(\frac{1}{10000})(0.99) + (\frac{9999}{10000})(0.01)} \approx 0.01.$$

Solution 24: Since each page is equally likely, we have the probability of selecting anywhere from page 12 to page 17. There are 6 pages, then, and so therefore there is a 6/1000 chance of selecting a page in this range.

Solution 25: In this case, we need to utilize Corollary 2 again. If we have two dice, then the probability of 1 being the minimum is, essentially, the probability of rolling at least one on either die. So, we have A as the event that we roll at least one, and $A = \{(1, 1), (1, 2), (2, 1), (1, 3), (3, 1), (1, 4), (4, 1), (1, 5), (1, 6)\}$. There are 7 events here, and the total number of events, S, is 36 (6(4)), and so we have $P(A) = 7/24$.

Solution 26: Yes, $X(\omega) = 0$ if $\omega < 1$, $X(\omega) = x/2$ if $1 \leq \omega < 2$, and $X(\omega) = 1$ if $x \geq 2$. (Why?)

Solution 27: First, let's find the mass. For the $P(X = 1)$, we have the Chris makes it on his first try. From the problem we then know $P(X = 1) = 1/3$. Next, if $P(X = 2)$, then we know Chris took two attempts to make it, and so he failed on his first attempt. Thus $P(X = 2) = (2/3)(1/3) = 2/9$. Following this procedure, we then have $P(X = 3) = 4/27$, $P(X = 4) = 8/81$, $P(X = 5) = 16/243$, and $P(X = 6) = 1 - P(X = 1) - P(X = 2) - P(X = 3) - P(X = 4) - P(X = 5) = 1 - (1/3) - (2/9) - (4/27) - (8/81) - (16/243) = 32/243$.

Solution 28: First, we should note that there are 20 good cards in total. So we have 20/52 chance of getting a "good" card. Next, we have a $1 - 20/52 = 32/52$ chance of getting a "bad" card. Since we stop after getting a good card, we can note that this is a **geometric random variable**, though this isn't necessary really. So, we therefore have $p_X(x) = (\frac{32}{52})^{x-1}(\frac{20}{52})$. Now, we need the CDF function, per the question. The CDF is defined to be $\sum_{i=1}^{x} p_X(i) = \sum_{i=1}^{x} (\frac{32}{52})^{i-1}(\frac{20}{52})$. Now, note that we can pull out the 20/52 to get $20/52 * \sum_{i=1}^{x} (\frac{32}{52})^{i-1}$. We can now shift the index backwards one to makes this $20/52 * \sum_{i=0}^{x-1} (\frac{32}{52})^{i}$. From the **Preliminaries** section, we know that this is just $20/52 * \frac{1-(32/52)^{(x-1)+1}}{1-(32/52)} = 20/52 * \frac{1-(32/52)^{x}}{1-(32/52)} = 1 - (32/52)^x$.

Solution 29: We have $p_X(1) = 330/27333$, $p_X(2) = 537/27333$, $p_X(3) = 8286/27333$, and $p_X(-1) = 18180/27333$. We then have $F_X(-1) = 18180/27333$, $F_X(1) = 18180/27333 + 330/27333 = 6170/9111$, $F_X(2) = 6170/9111 + 537/27333 = 6349/9111$, and $F_X(3) = 6349/9111 + 8286/27333 = 1$.

Solution 30: First, notice that $p_Y(1) > 0$, because it is possible that the first card is a queen. On the other hand, if $X = 1$, or in other words the first card is an ace, then the first card is not a queen, and so we have $p_{Y|X}(1|1) = 0$. Thus, $p_Y(y) \neq p_{Y|X}(y|x)$, and so X and Y are not independent.

Solution 31: We have $p_{X,Y}(0,1) = p_{X,Y}(1,0) = (1/4)(1/2)$, $p_{X,Y}(0,2) = p_{X,Y}(2,0) = (1/4)(1/2)^2$, $p_{X,Y}(1,1) = (1/4)(1/2)$, $p_{X,Y}(3,0) = p_{X,Y}(0,3) = (1/4)(1/2)^3$, $p_{X,Y}(2,1) = p_{X,Y}(1,2) = (1/4)(3/8)$, $p_{X,Y}(0,4) = p_{X,Y}(4,0) = (1/4)(1/2)^4$, $p_{X,Y}(1,3) = p_{X,Y}(3,1) = (1/4)(4/16)$, and $p_{X,Y}(2,2) = (1/4)(6/16)$.
Solution 32: First, let's note that X and Y are independent. The minimum value of Y is 2. So, we have that the probability of $X < 2$ is 1/6. Next, $Y = 3$. So, we have $X < 3 \leftrightarrow X = 1 or X = 2$, and so $P(X < 3) = 2/6$. Following suit, we end up with: $P(X < 2, Y = 2) + P(X < 3, Y = 3) + P(X < 4, Y = 4) + P(X < 5, Y = 5) + P(X < 6, Y = 6) + P(Y = 7) + P(Y = 8) \leftrightarrow P(X < 2)P(Y = 2) + P(X < 3)P(Y = 3) + P(X < 4)P(Y = 4) + P(X < 5)P(Y = 5) + P(X < 6)P(Y = 6) + P(Y = 7) + P(Y = 8)$ $\leftrightarrow (1/6)(1/16) + (2/6)(2/16) + (3/6)(3/16) + (4/6)(4/16) + (5/6)(3/16) + (2/16) + (1/16) = 21/32$.
Solution 33: So, we need to find all 12 values. Lets begin:

$$p_{X|Y}(1,1) = \frac{P(X = 1 \cap Y = 1)}{P(Y = 1)} = \frac{1/3}{(1/3) + (1/3)(2/3) + (1/3)(2/3)^2} = 9/19.$$

The rest of the values are found in a similar manner.
Solution 34: From a remark from earlier, we have that $p_X(x) = \sum_y p_{X,Y}(x,y)$. So, in this case, we have $\sum_{y=1}^{\infty}(5/9)(1/3)^{y-1}(1/2)^{x-1}$. Since we're not dealing with the x's, we can take that and the $(5/9)$ out to get $(5/9)(1/2)^{x-1}\sum_{y=x}^{\infty}(1/3)^{y-1}$. We can then note that $\sum_{y=x}^{\infty}(1/3)^{y-1} = \frac{(1/3)^{x-1}}{1-(1/3)} = (3/2)(1/3)^{x-1}$. Putting this altogether gives us $(5/6)(1/6)^{x-1}$.
Solution 35: We have the probability mass function from **Question 27**. So we then have, by definition of expectation, $(1)\mathbb{E}(X(1)) + (2)\mathbb{E}(X(2)) + 3\mathbb{E}(X(3)) + 4\mathbb{E}(X(4)) + 5\mathbb{E}(X(5)) + (6)\mathbb{E}(X(6)) \leftrightarrow (1)(1/3) + (2)(2/9) + (3)(4/27) + (4)(8/81) + (5)(16/243) + (6)(32/243) = 665/243$.
Solution 36: So we have $P(X = 1) = (3/5)(2/4)(1/3) + (2/5)(3/4)(1/3) + (2/5)(1/4)(1) = \frac{3}{10}$. Next, we have $P(X = 2) = (3/5)(2/4)(2/3) + (2/5)(3/4)(2/3) + (3/5)(2/4)(2/3) = \frac{3}{5}$. Finally, we have $P(X = 3) = (3/5)(2/4)(1/3) = \frac{1}{10}$. Using the definition of expectation we then have $(1)(3/10) + (2)(3/5) + (3)(1/10) = \frac{9}{5}$.
Solution 37: We have $P(X = 0) = (4/6)(3/6) = 12/36 = 1/3$, $P(X = 1) = (4/6)(3/6) + (2/6)(3/6) = 1/2$, $P(X = 2) = (2/6)(3/6) = 1/6$. So, using the definition of expectation, we have $\mathbb{E}(X) = (0)(1/3) + (1)(1/2) + (2)(1/6) = \frac{5}{6}$.
Solution 38: There are two main ways of doing this, but because we are in Chapter 11 we'll go about using the Bernoulli random variable. Let us break this into three main events − $X = X_1 + X_2 + X_3$, where X_i represents the event of getting a heads on the i-th flip. Then we have $\mathbb{E}(X) = \mathbb{E}(X_1 + X_2 + X_3)$. However, these are independent events which are equivalent (the term for this is i.i.d., though I believe Dr. Ward has not discussed this.) So, we can therefore use **Theorem 18** to break this up into $\mathbb{E}(X) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \mathbb{E}(X_3)$. We have $\mathbb{E}(X_i) = P(A_i) = 1/2$ for each flip, and so by **Theorem 20** we have $3 * 1/2 = 3/2$ as our final answer.
Solution 39: Let A_j be the event that the j-th card is a heart. So we have $P(A_j) = 1/4$ for each j. Let X_j be the indicator for A_j. Then we have, by prior theorems, that $\mathbb{E}(X_j) = P(A_j) = 1/4$. Also by prior theorems, since the X_j's are i.i.d, we have $\mathbb{E}(X) = 5 * \mathbb{E}(X_1) = 5 * 1/4 = 5/4$.
Solution 40: There are three major ways of doing this, but we're going to do the easiest way − with indicator random variables. Let X_j be the indicator for the j-th card being the ace of spades. So we therefore have $X_1 + ... + X_{51} + 1$ (the 1 is because once reach the 52nd card, and we haven't gotten the ace of spades, then it for sure must be the last card). Then the expected number of cards to draw until the ace of spades appears is $\mathbb{E}(X_1 + ... + X_{51} + 1) = \mathbb{E}(X_1) + ... + \mathbb{E}(X_{51}) + 1$. The individual $\mathbb{E}(X_i) = \frac{1}{2}$ (why?), and so we have $51/2 + 1 = 53/2$ as our final answer.
Solution 41: Break up X into seven events, $X_1 + ... + X_7$. Then we have that $\mathbb{E}(X) = \mathbb{E}(X_1) + ... + \mathbb{E}(X_7)$. For any arbitrary $\mathbb{E}(X_i)$, we have that this is equal to $2/6 = 1/3$. So since the means are i.i.d, we have $7 * 1/3 = 7/3$.
Solution 42: We have $\mathbb{E}(X) = \mathbb{E}(X_1) + ... + \mathbb{E}(X_7)$. For any arbitrary $\mathbb{E}(X_i)$, we have that it is equal to 1/100. So, using prior theorems, we have $\mathbb{E}(X) = 7/100$.
Solution 43: We have $p_X(j) = \binom{6}{j}(.4)^j(.6)^{6-j}$. Therefore, we have $\sum_{j=0}^{6}(j^2)(\binom{6}{j}(0.4)^j(0.6)^{j-1}) = 7.2$. Another way to do this is to use indicator random variables. So, we have the event A_j occurs when the j-th student lives in a residence hall. Therefore, $\mathbb{E}(X_j) = 0.4$. Now, we need to square $\mathbb{E}(X_1) + ... \mathbb{E}(X_6)$.

Doing so gives is $6 * \mathbb{E}(X_1) + (6^2 - 6) * \mathbb{E}(X_1 X_2)$, since the variables are i.i.d. Note that $\mathbb{E}(X_i X_j)$ occurs when both X_i and X_j occurs. Since the students are independent, this is simply $0.4^2 = 0.16$. So we therefore have $\mathbb{E}(X^2) = (6)(0.4) + (30)(0.16) = 7.2$, which matches our previous answer.
Solution 44: There are two ways to go about this. Let's do it by finding the probability mass function first. We can note that this is sampling without replacement, which means that hyper-geometric is the best route. We have 3 good options, and we have 6 bad options, and we are selecting five. So the probability mass function would be $\frac{\binom{3}{x}\binom{6}{5-x}}{\binom{9}{5}}$. Therefore, $\mathbb{E}(X^2) = \sum_{x=0}^{5}(x^2)(\frac{\binom{3}{x}\binom{6}{5-x}}{\binom{9}{5}}) = \frac{10}{3}$.
Next, we can do it by defining $X = Y_1 + Y_2 + Y_3$. So, we have $X^2 = 3 * Y_k + 6 * (Y_i * Y_j)$, $0 \leq i < j \leq 5$ and $0 \leq k \leq 5$. For the first part, the probability of one of the bears appearing is $Y_i = 5/9$ and the probability of two bears appearing is $Y_i * Y_j = (5/9)(4/8)$. So we therefore get $3*5/9+6*(5/9)(4/8) = 10/3$.
Solution 45: There are two ways of doing this. We can use the probability mass function, which is $p_X(1) = (1/24)$, $p_X(2) = (3/24)$, $p_X(3) = (5/24)$, $p_X(4) = (7/24)$, $p_X(5) = 1/6$, and $p_X(6) = 1/6$. So, we have $1(1/24) + (4)(3/24) + (9)(5/24) + (16)(7/24) + (25)(1/6) + (36)(1/6) = 69/4$.
Using indicator random variables, let X_i denote the event of the i-th value or more being the maximum of the two dice. Then we have $X^2 = (X_1 + ... + X_6)^2$. However, these variables are dependent, and so we can't use the i.i.d trick. So, we need to consider all the cases:

- If we have just X_i for $1 \leq i \leq 4$, then the probability of it occurring is $1/24$.
- If we have just X_i for $5 \leq i \leq 6$, then the probability of it occurring is $1/6$.
- If we have $X_i X_6$, then we have simply X_6. This happens for 10 items.
- If we have $X_i X_5$, then we have simply X_5. This happens for 8 items.
- If we have $X_i X_4$, then we have X_4 for 6 items.
- If we have $X_i X_3$, then we have X_3 for 4 items.
- If we have $X_i X_2$, then we have X_2 for 2 items.

So, we therefore have $(11)(1/6) + (9)(2/6) + (7)(15/24) + (5)(20/24) + (3)(23/24) + (1)(1) = 69/4$
Solution 46: His expected gain, in this case, would be $\mathbb{E}(9X - 4)$. Therefore, due to the linearity of the mean, we have $9\mathbb{E}(X) - 4$, which is simply $9(0.4) - 4 = -0.4$.
Solution 47: The most efficient answer uses indicator random variables. Let X_i denote that the i-th pair is happy. Then we have $X = X_1 + X_2 + X_3$. We expand $(X_1 + X_2 + X_3)^2$ to get $X_1 + X_2 + X_3 + X_1X_2 + X_1X_3 + X_2X_3 + X_2X_1 + X_3X_2 + X_3X_1$. Note that $X_iX_j = X_jX_i$, and so we have $X_1 + X_2 + X_3 + 2X_1X_2 + 2X_1X_3 + 2X_2X_3$. We therefore have X_iX_j occurring if two of the bear couples are happy. So, we have $X_1X_2 = (2/5)(1/2)$, and $X_1 = 2/5$. This is because if we arbitrarily throw a bear down, there is a $2/5$ chance that we get one bear couple happy, and consequently a $1/2$ chance of getting two bear couples happy. Since these variables are i.i.d, we can rewrite this as $3 * (2/5) + (6)(2/5)(1/2) = 6/5 + 12/10 = 24/10$.
Solution 48: We can utilize a Bernoulli random variable here. We therefore have $p = 0.4$ and $q = 0.6$. So, by definition, $\mathbb{E}(X) = 0.4$ and $Var(X) = (0.4)(0.6) = (0.24)$.
Solution 49: So we have $X = X_1 + ... + X_30$. They are independent, so using the linearity of the expected value we get $\mathbb{E}(X) = 30 * \mathbb{E}(X_1) = 30 * 0.65 = 19.5$. We can then use the theorem about the independence of variance to get $Var(X) = 30 * Var(X_1) = 30 * (0.65)(0.35) = 6.825$.
Solution 50: Pick a card around the table arbitrarily, and let that be denoted by X_i. Then the probability that it is isolated is $\mathbb{E}(X_i) = (10/14)(9/13) = (45/91)$. Since we have 15 cards, and this is a sum of Bernoulli random variables (hence, binomial) we have $\mathbb{E}(X) = np = (15)(45/91) = (675/91)$.
Solution 51: We have a sum of Bernoulli random variables, and hence a binomial random variable by definition. Since X denotes the amount of games he wins, we have that he needs to win $9X - 40 = 32 \leftrightarrow X = 8$, or he needs to win 8 or more games. The probability mass function is $p_X(x) = \binom{10}{x}(0.4)^x(0.6)^{10-x}$, and so we therefore have $\sum_{x=8}^{10}\binom{10}{x}(0.4)^x(0.6)^{10-x} = 0.01229....$
Solution 52: Let $X = X_1 + ... + X_7$. Then the probability that the X_i card is a heart is $13/52$. So we therefore have $\mathbb{E}(X) = 7 * \mathbb{E}(X_1) = 7 * (13/52) = 7/4$.

Solution 53: Since it's at least one, we want to do $\sum_{x=1}^{n}\binom{n}{x}(0.3)^x(0.7)^{n-x} = 0.95$, and we're solving for n. The best way to do this is to really just cycle through all possible answers for n until we get $n = 9$.
Solution 54: It is trivial to note that p remains the same throughout, and so this must be a Geometric random variable. So we have, then, that $p = 1/6$ and $\mathbb{E}(X) = 1/p = 6$.
Solution 55: We can use the fact that $P(A^c|B) = 1 - P(A|B)$. So we have $P(X \leq b|X > a) = 1 - P(X > b|X > a)$. Due to memoryless-ness, we then get that this is $1 - q^{b-a}$.
Solution 56: We have $\sum_{y=1}^{\infty}\sum_{x=y}^{\infty}(1/2)(1/2)^{y-1}(1/6)(5/6)^{x-1}$. Rearranging variables gives us

$$(1/12)\sum_{y=1}^{\infty}(1/2)^{y-1}\sum_{x=y}^{\infty}(5/6)^{x-1}.$$

This then gives us

$$\begin{aligned}(1/12)\sum_{y=1}^{\infty}(1/2)^{y-1}\frac{(5/6)^{y-1}}{1-(5/6)} &= (6/12)\sum_{y=1}^{\infty}(5/12)^{y-1} \\ &= (1/2)\frac{1}{1-(5/12)} = (1/2)(12/7) \\ &= (12/14).\end{aligned}$$

Solution 57: For X to be even, we must have that there is always an odd amount of failures leading up to the success. Since $\mathbb{E}(X) = 3$, we know that $p = 1/3$. Therefore, $q = 2/3$. So we have that $P(A) = \sum_{x=1}^{\infty}(2/3)^{2x-1}(1/3) = (1/3)(3/2)\sum_{x=1}^{\infty}(4/9)^x = (1/2)\frac{4/9}{5/9} = 2/5$.
Solution 58: It is good to check that the assumptions for a negative binomial random variable. However, we're gonna assume that it is one. We then have $p = 0.15$, $r = 4$, and so $p_X(x) = \binom{x-1}{3}(0.85)^{x-4}(0.15)^4$.
Solution 59: We have $p = 0.9$, $r = 5$. So we have $\binom{4}{4}(0.9)^5(0.1)^0 + \binom{5}{4}(0.9)^5(0.1)^1 + \binom{6}{4}(0.9)^5(0.1)^2 + \binom{7}{4}(0.9)^5(0.1)^3$.
Solution 60: We have $p = 0.8$, $r = 10$. So we have $P(X \geq 12) = 1 - P(X < 12)$, or $1 - \binom{9}{9}(0.8)^{10}(0.2)^0 - \binom{10}{9}(0.8)^{10}(0.2)^1 = 0.6779$.
Solution 61: We have $r = 3$, and $p = 0.25$. An hour consists of 12 different attempts. So we have $P(X > 12) = 1 - P(X \leq 12)$. So we therefore have $1 - \sum_{x=3}^{12}\binom{x-1}{2}(0.25)^3(0.75)^{x-3} = 0.3907$.
Solution 62: Let $\lambda = 6$. Then we have $P(X = 5) = \frac{e^{-6}(6)^5}{5!} = 0.1606$.
Solutions 63: This is technically Binomial, but since the numbers are very big we can approximate it with Poisson (because I'm lazy). So we have $\lambda = \mathbb{E}(X) = (5000000)(1/2000000) = 5/2$. So, we therefore have $P(X = 4) = \frac{e^{-5/2}(5/2)^4}{4!} = 0.1336....$
Solution 64: We have $\lambda = 4$, and so $P(X = x) = \frac{e^{-4}(4)^x}{x!}$.
Solution 65: Note that $\lambda \neq 2$. We're measuring in terms of 3 minutes, and so it's actually $\lambda = (3)(2) = 6$. Since we want $P(X \geq 4)$, we can instead find $P1 - P(X < 4)$. This is $1 - \sum_{x=0}^{3}\frac{e^{-6}6^x}{x!} = 0.8488$.
Solution 66: Here, we have $N = 1000000$, $M = 50000$, and $n = 10$. So therefore we have $P(X = 1) = \frac{\binom{50000}{1}\binom{950000}{9}}{\binom{1000000}{10}}$. However, this is hard to do on a calculator. We can use the fact that N is really really big and n is relatively small. So, we then have $P(X = 1) = \binom{10}{1}(0.05)^1(0.95)^9 = 0.3151$.
Solution 67: Here, we can note that $M = 10$, $N = 23$, $N - M = 13$, and $x = 3$. So, using the probability mass function, we have $P(X = 3) = \frac{\binom{10}{3}\binom{13}{0}}{\binom{23}{3}} = 0.06775$.
Solution 68: We have $M = 5$, $N = 15$, and so by the formula given in the previous chapter we have $\mathbb{E}(X) = n\frac{M}{N} = (3)\frac{5}{15} = 15/15 = 1$.
Solution 69: Since these are Geometric random variables, we have that $\mathbb{E}(X) = 1/p_X$, and so this implies that $p_X = 1/4$. Likewise, we have $p_Y = 2/3$. By definition, if $X = Y$, then we have $p_{X=Y}(x) = (1/4)(3/4)^{x-1}(2/3)(1/3)^{x-1}$. Rewriting this gives us $(2/12)(3/12)^{x-1}$, and so we need to solve $(2/12)\sum_{x=1}^{\infty}(3/12)^{x-1}$. By preliminaries, we know that this is $\frac{(2/12)}{1-(3/12)} = \frac{(2/12)}{(9/12)} = 2/9$.

Solution 70: We need to cycle through $P(X=Y=0)+P(X=Y=1)+P(X=Y=2)$. Calculating this out, we have $P(X=0)=\frac{\binom{3}{0}\binom{3}{2}}{\binom{6}{2}}=1/5$, $P(X=1)=\frac{\binom{3}{1}\binom{3}{1}}{\binom{6}{2}}=3/5$, and $P(X=2)=\frac{\binom{3}{2}\binom{3}{0}}{\binom{6}{2}}=1/5$. Since the variables are independent, we can multiply them together to get $(1/5)^2+(3/5)^2+(1/5)^2=11/25$.
Solution 71: This is without replacement, and order does matter. Referring to the table in Chapter 22, we see that the equation is $\frac{n!}{(n-r)!}$, where n is the total amount and r is the amount we choose. So we have $n=10$ and $r=4$, and by substitution we get $\frac{10!}{(10-4)!}=5040$.
Solution 72: If there are n couples, then that means there's $(2n)!$ total combinations that we're dealing with. Since there's a man and a woman in a couple, we have 2 ways of arranging the couple for each couple. In total, this means that there are 2^n ways of arranging the couples. Finally, we have $n!$ ways of arranging the couples, and so we have a probability of $\frac{2^n n!}{(2n)!}$.
Solution 73: There are 3! ways of arranging the girls in their group. Imagine that we have the girls sitting on a couch and the guys are sitting on a chair. Note there are 4 ways of arranging this couch. Next, there are 3! ways of arranging the guys on their chairs. Finally, there's a total of 6! ways of arranging everyone. So, we have that the probability is $\frac{(3!)(4)(3!)}{6!}=\frac{1}{5}$.
Solution 74: Since our function starts when $2\leq x$, we have

$$\int_1^3 f_X(x)dx=\int_2^3 f_X(x)dx.$$

Substituting the function gives

$$\int_2^3 \frac{1}{26}(4x+1)dx=\frac{1}{26}(2x^2+x)\Big|_2^3=\frac{11}{26}=0.4231$$

Solution 75: From the definition of CDF, we have that $f_X(x)=\frac{d}{dx}F_X(x)$. Therefore, taking the derivative, we get $\frac{1}{4}e^{-x/4}$.
Solution 76: We compute

$$P(X\leq 1)=\int_{-\infty}^1 f_X(x)dx=\int_0^1 \frac{3}{8}(x)(2-x)(3-x)=19/32$$

Solution 77: To check that it's a joint density, we need to check

$$\int_0^\infty\int_0^\infty 15e^{-3x-5y}dydx=1.$$

First, let's examine the inner integral. We have

$$\int_0^\infty 15e^{-3x-5y}dy.$$

Let's treat the x's as though they were constants. Doing so gives us

$$-3e^{-3x-5y}\Big|_{x=0}^\infty=0-(-3e^{-3x})=3e^{-3x}.$$

We can now substitute that in to the original integral to get

$$\int_0^\infty 3e^{-3x}dx.$$

Solving this integral gives us

$$-e^{-3x}\Big|_{x=0}^\infty=0-(-e^0)=1,$$

as required.
Next, we need to solve for when $P(X\leq 1/2 \text{ and } Y\leq 1/4)$. This is equivalent to asking

$$\int_0^{1/2}\int_0^{1/4} 15e^{-3x-5y}dydx.$$

We can seperate the integrals to get

$$\int_0^{1/2} e3^{-3x}dx * \int_0^{1/4} 5e^{-5y}dy = 0.5543$$

Solution 78: Since they are independent, we simply multiply them together. So we ahve $f_{X,Y}(x,y) = f_X(x)f_Y(y) = 1/3e^{-x/3}1/3e^{-y/3} = 1/9e^{-(x+y)/3}$.
Solution 79: They are not. Note that we cannot factor the polynomial $xy + x^2$ into two seperate function of x's and y's.
Solution 80: $f_Y(y) = \frac{2}{9}(3-y)$. Note that $f_{X,Y}(x,y) = 2/9$. So, by definition, we have $f_{X|Y}(x|y) = \frac{2/9}{(2/9)(3-y)} = \frac{1}{3-y}$ for $0 \le x \le 3-y$, and it's 0 otherwise.
Solution 81: Let's note that these distributions are independent. We then have $f_X(x) = 3e^{-3x}$ and $f_Y(y) = 5e^{-5y}$. Using the formula, we then have $\mathbb{E}(X) = 1/3$ and $\mathbb{E}(Y) = 1/5$.
Solution 82: Using the definition, we have that this is

$$\int_{19}^{20} \frac{2}{3}x(x-19)dx + \int_{20}^{22} -\frac{1}{3}x(x-22)dx$$

. Factoring out the constants and distributing the variable, x, gives us:

$$\frac{2}{3}\int_{19}^{20}(x^2-19x)dx - \frac{1}{3}\int_{20}^{22}(x^2-22x)dx$$

. Solving this then gives us 61/3.
Solution 83: We need to first find the function for f_X. By definition, we do

$$\int_0^4 (3/80)(x^2+y)dy = (3/20)(x^2+2) = f_X(x)$$

. Thus, we now use the definition to get

$$\int_0^2 (3/20)(x^3+2x)dx = \frac{6}{5}$$

.
Solution 84: We can do this directly from the definition. We need to find $\mathbb{E}(X)$ and $\mathbb{E}(X^2)$. To find the mean, we have

$$\int_0^\infty 2xe^{-2x}dx.$$

Let $u = x$ and $dv = 2e^{-2x}dx$. Then we have $du = dx$ and $v = -e^{-2x}dx$. Substituting this into the integration by parts formula, we have

$$-xe^{-2x}|_{x=0}^{\infty} + \int_0^\infty e^{-2x}dx$$

. Note that $-xe^{-2x}\Big|_{x=0}^{\infty} = 0$, and so we have $\int_0^\infty e^{-2x}dx = \frac{-e^{-2x}}{2}\Big|_{x=0}^{\infty}$. Evaluating this gives us 1/2. Next, we need to evaluate

$$\int_0^\infty 2x^2e^{-2x}dx.$$

We will have to preform integration by parts twice to get this. Let $u = x^2$ and $dv = 2e^{-2x}dx$. Then $du = 2xdx$ and $v = -e^{-2x}$. Substituting this in gives

$$-x^2e^{-2x}|_{x=0}^{\infty} + \int_0^\infty 2xe^{-2x}dx$$

. Note that $-x^2e^{-2x}|_{x=0}^{\infty} = 0$. This leaves $\int_0^\infty 2xe^{-2x}dx$. We know this evaluates out to 1/2 from prior. Thus, by definition, we have that $Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = 1/2 - 1/4 = 1/4$.

Solution 85: By definition, $\mathbb{E}(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dydx$. Thus, substituting the values in, we have

$$\int_0^4\int_0^9 (xy)/36dydx.$$

Solving the first integral, we have $\int_0^9 xy/36dy = xy^2/72\Big|_{y=0}^{9} = 81x/72$. Substituting this in gives us

$$\int_0^4 81x/72dx = 81x^2/144\Big|_{x=0}^{4} = 1296/144 = 9.$$

Solution 86: We need to first establish the boundaries. We have $0 \le x \le 2$ and $0 \le y \le 2-x$, since X and Y aren't independent (why?). Therefore, our integral is

$$\int_0^2\int_0^{2-x} xy(\frac{15}{28}(xy^2+y))dydx = \frac{15}{28}\int_0^2\int_0^{2-x}(x^2y^3+xy^2)dydx.$$

Solving the integral for y first, we have $\int_0^{2-x} x^2y^3+xy^2dy = x^2y^4/4+xy^3/3\Big|_{y=0}^{2-x} = x^2(2-x)^4/4+x(2-x)^3/3$. Substituting this in, we have

$$\frac{15}{28}\int_0^2(x^2(2-x)^4/4+x(2-x)^3/3)dx = 22/49.$$

Solution 87: We need to check for what value k does $\int_0^1 kx^2(1-x)^7dx = 1$. Let $u = 1-x$. Then we have $x = 1-u$. Taking the derivative of u gives us $du = -dx$. So, we have

$$\int_1^0 k(1-u)^2u^7du.$$

Now we are able to easily expand $(1-u)^2$ to get $1-2u+u^2$. Multiplying this by u^7 gives $u^7-2u^8+u^9$. Note that the k is constant, and so we can pull it out to get

$$k\int_1^0 u^7-2u^8+u^9du.$$

This is just a sum of polynomials, and so we have $k(u^8/8-2u^9/9+u^10/10)\Big|_{u=1}^{0} = 1$. Substituting the values in gives us $-k(1/8-2/9+1/10) = 1$, or $k = -\frac{1}{1/8-2/9+1/10} = -360$.
Solution 88: We can do this by preforming the integral $\int_3^\infty \frac{dx}{x^2}$. We can rewrite this to get $\int_3^\infty x^{-2}dx = -x^{-1}\Big|_{x=3}^{\infty} = \frac{1}{3}$.
Solution 89: We need to sum over all possible y's in order to find the probability density function of x. In order to do so, we have

$$\int_x^\infty 70e^{-3x-7y}dy = -10e^{-3x-7y}\Big|_{y=x}^{\infty} = 10e^{-10x},$$

for $x > 0$. For $x \le 0$, we have $f_X(x) = 0$.
Solution 90: Assuming that the probability or getting a defect is uniform over this 10 yard rope, we have that the probability for the location is $f_X(x) = \frac{1}{10}$ for $0 \le x \le 10$, and 0 otherwise. Thus, for $P(X > 8)$, we have $\int_8^{10}\frac{1}{10}dx = \frac{10-8}{10} = \frac{1}{5}$.
Solution 91: We want to find $\mathbb{E}(Y)$. Rewriting this, we have $\mathbb{E}(Y) = \mathbb{E}(2.50X+3.00)$. However, recall that the mean is a linear operator, and so we have $2.50\mathbb{E}(X)+3.00$. From the theorem in the chapter, we have that $\mathbb{E}(X) = \frac{35}{2} = 17.5$, and so we have $2.50(17.5)+3.00 = \$46.75$
Solution 92: The trick here is to realize that we have $P(X_1 < X_2 < X_3)+P(X_1 < X_3 < X_2)+P(X_2 < X_1 < X_3)+P(X_2 < X_3 < X_1)+...=1$, where the ... represents all possible permutations of X_1, X_2, and X_3. Note that $P(X_i = X_j < X_k) = 0$ for all possible i, j, k, and $P(X_i = X_j = X_k)$ for all possible i, j, k, $i \ne j \ne k$ in both cases (why is that?). It then makes sense to say that the summation of all these probabilities is 0, since we are summing over all possible combinations of disjoint probabilities on

the interval. Now, note that there are $3! = 6$ different possible combinations of X_1, X_2, and X_3 such that their probability is not zero. Then we have $6P(X_1 < X_2 < X_3) = 1$, or $P(X_1 < X_2 < X_3) = \frac{1}{6}$.
Solution 93: We want to find t such that $P(X \leq t) = 0.8$, where X is the waiting time in hours; in other words, the t such that $F_X(t) = 0.8$. By the theorem, we have that the CDF is $1 - e^{-\lambda t} = 0.8$. Rewriting this, we have $e^{-\lambda t} = 0.2 \leftrightarrow -\lambda t = ln(0.2) \leftrightarrow t = -ln(0.2)/\lambda$. Recall from the definition that λ is one over the mean, and so we have $t = -ln(0.2)/2 \approx 0.804719....$
Solution 94: Note that $\mathbb{E}(X) = 30$, and so $\lambda = 1/30$. We then need to find $P(X < 10)$, where X is the amount of time it takes for her to fall asleep. In other words, we need to find $F_X(10)$. Using the theorem, we have $1 - e^{-(1/3)} \approx 0.2835...$ hours.
Solution 95: Note that $Y = \lfloor X \rfloor \leftrightarrow P(Y = y) = P(y \leq X \leq y + 1)$. So, we have

$$\int_y^{y+1} \lambda e^{-\lambda x} dx = -e^{-\lambda x}\Big|_{x=y}^{y+1} = e^{-\lambda y} - e^{-\lambda(y+1)}.$$

As an aside, note that this is a Geometric random variable.
Solution 96: Since they are on different flights, we can assume independence. So we have $r = 7$ and $\lambda_i = \frac{1}{1/45}$. Using the theorem, then, we have $\mathbb{E}(X) = \frac{7}{(1/45)} = 315$.
Solution 97: Assume each value is independent. Then we have that $\mathbb{E}(X_i) = 3$ and $r = 7$, and so applying the theorem we have $\mathbb{E}(Z) = 3 * 7 = 21$ minutes, where Z represents the amount of time for 7 people in line.
Solution 98: Assuming independence, we have that $r = 500$, $\lambda = 10$, so $\mathbb{E}(X_i) = 1/10$, and by the theorem $\mathbb{E}(Z) = 500 * 1/10 = 50$ minutes total, where Z is the amount of time in between each meteor for 500 meteors.
Solution 99: Using the theorem above, we have that the expected value is $\frac{3}{7}$.
Solution 100: We are looking for $P(X < 20)$. Let Z denote the normalized random variable. Normalizing 20, we have $\frac{20-22}{\sqrt{8}} = -0.71$. We now have $P(Z < -0.71)$. Referring to the z-table, this gives us $1 - P(Z < 0.71) = 1 - 0.7611 = 0.2389$.
Solution 101: We need to find $\mathbb{E}(Y)$ and $Var(Y)$. First, we have $\mathbb{E}(Y) = 5\mathbb{E}(X) + 1 = 16$. Next, we have $Var(Y) = 25 * 4 = 100$. So, letting Z represent the normalized Y, we have

$$P(\frac{10-16}{\sqrt{100}} < Z < \frac{20-16}{\sqrt{100}}) = P(-0.6 < Z < 0.4) = P(Z < 0.4) - P(Z < -0.6).$$

Referring to the z-table, this then gives $0.6554 - (1 - 0.7257) = 0.3811$.
Solution 102: Let X denote the heights of college females. Then we are looking for $P(X \geq 67)$. Let Z denote the normalized X. Then we have $P(Z \geq \frac{67-64}{4.8}) = P(Z \geq 0.63) = 1 - P(Z \leq 0.63)$. Referring to the z-table, we have that this is equivalent to $1 - 0.7357 = 0.2643$.
Solution 103: We are looking for $P(X < 40)$. Normalizing this, we have $P(Z < \frac{40-23(1.8)}{\sqrt{23(0.5^2)}} = P(Z < -0.58)$. Referring to the z-table, we have $1 - P(Z < 0.58) = 1 - 0.7190 = 0.281$.
Solution 104: We have $P(X > 9966)$. In other words, this is the same thing as asking $1 - P(X < 9966)$. Normalizing this, we have $1 - P(Z < \frac{9966-66(150)}{\sqrt{66(20^2)}}) = 1 - P(Z < 0.40)$. Referring to the z-table we have $1 - 0.6554 = 0.3446$.
Solution 105: We need to find the expected value and the variance, in order to utilize the Central limit theorem. So we have that the expected value for a month is $\frac{4.0+0.2}{2} = 2.1$ and the variance is $\frac{(4.0-0.2)^2}{12} = 1.203$. So, we want to find $P(X > 53)$. Normalizing this we have $P(Z > \frac{53-24(2.1)}{\sqrt{24(1.203}}) = P(Z > 0.48) = 1 - P(Z < 0.48) = 1 - 0.6844 = 0.3156$.
Solution 106: Notice that it's not inclusive, and so by the continuity correction table we are looking for $P(1580.5 < X < 1619.5)$. Normalizing this, we have $P(\frac{1580.5-(20)(0.8)(100)}{\sqrt{20(0.8)(0.2)(100)}} < Z < \frac{1619.5-(20)(0.8)(100)}{\sqrt{20(0.8)(0.2)(100)}} = P(-1.09 < Z < 1.09) = P(Z < 1.09) - (1 - P(Z < 1.09)) = 2P(Z < 1.09) - 1 = 2(0.8621) - 1 = 0.7242$.
Solution 107: Note that a half an hour is 1800 seconds. So, we want to look for $P(X < 1800)$. Normalizing this, we have $P(Z < \frac{1800-(72.5)(25)}{\sqrt{25(3.2)^2}}) = P(Z < -0.78) = 1 - P(Z < 0.78) = 1 - 0.7823 = 0.2177$.
Solution 108: We have a hypergeometric random variable, where $M = 8$, $N = 11$, and $n = 2$. Using the formula for variance, we have $(2)(\frac{8}{11})(\frac{3}{11})(\frac{9}{10}) = \frac{216}{605}$

Solution 109: By definition, the covariance is $\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$. To find $\mathbb{E}(XY)$, we take

$$\int_{10}^{14} \frac{xy}{4}dx = \int_{10}^{14} \frac{x(2x+2)}{4}dx = \int_{10}^{14} \frac{2x^2+2x}{4}dx = \frac{944}{3}.$$

Note that by the equation, we have $\mathbb{E}(X) = 12$, and by the linearity of expected values, $\mathbb{E}(Y) = 2(12) + 2 = 26$. Therefore, we have that $\text{Cov}(X,Y) = 944/3 - 12(26) = 8/3$.

Solution 110: We have $\mathbb{E}(X) = \int_0^3 \int_0^{3-y}(x)(2/9)dxdy = 1$, $\mathbb{E}(Y) = \int_0^3 \int_0^{3-y}(y)(2/9)dxdy = 1$, and $\mathbb{E}(XY) = \int_0^3 \int_0^{3-y}(xy)(2/9)dxdy = 3/4$. So, $\text{Cov}(X,Y) = 3/4 - 1 = -1/4$. To find the variance, we find $\mathbb{E}(X^2) = \int_0^3 \int_0^{3-y}(x^2)(2/9)dxdy = 3/2$, and by symmetry $\mathbb{E}(Y^2) = 3/2$. Therefore, Var(X) $= 3/2 - 1^2 = 1/2$, Var(Y) $= 1/2$, and so we have that the correlation is $\rho_{X,Y} = \frac{-1/4}{\sqrt{(1/2)(1/2)}} = -1/2$.

Solution 111: By definition, we have $f_{Y|X}(y|x) = \frac{f_{Y,X}(y,x)}{f_X(x)}$. Note that $f_{Y,X}(y,x) = \frac{1}{2}$, and $f_X(x) = \int_y f_{Y,X}(y,x)dy$. We have that $y \in [0, 3/2]$ and so $f_X(x) = \int_0^{3/2} 1/2dy = 3/4$. So, substituting our values in, we have $f_{Y|X}(y|x) = 3/4$.

Solution 112: By definition, we have $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$. So, we need to first find $f_X(x)$. However, by definition, we have $f_X(x) = \int_x^\infty 70e^{-3x-7y}dy = 10e^{-10x}$. Therefore, we substitute it in to get

$$f_{Y|X}(y|x) = \frac{70e^{-3x-7y}}{10e^{-10x}} = 7e^{10x-3x-7y} = 7e^{7x-7y}$$

for $y > x$, and $f_{Y|X}(y|x) = 0$ otherwise.

Solution 113: If the minimum value is 5, then our options are $(5,5), (5,6)$, and $(6,5)$. Therefore, we have $P(X = 5|Y = 5) = 1/3$ and $P(X = 6|Y = 5) = 2/3$. So, we then have $\mathbb{E}(X|Y = 5) = 5(1/3) + 6(2/3) = 17/3$.

Solution 114: Use Markov's inequality, with $a = 95000$. Then we have $\mathbb{E}(X) = 82000$, and so $P(X \geq 95000) \leq \frac{82000}{95000} = 0.8632$.

Solution 115: Note that $\sigma = 0.05$ and $\mu = 0.8$. Then by Chebychev we have $P(|X - \mu| \leq 0.07)$. Let $k = 0.07/0.05$. Then we have by the remark that $P(|X - \mu| \leq 0.07) \geq 1 - \frac{1}{k^2} = 0.49$.

Solution 116: We have $\mu = 2.5$, and $\sigma = 1.5$. We are therefore looking for $P(|X - \mu| \leq 0.5)$, and so $k = 0.5/1.5 = 1/3$. Using the remark, we then have $P(|X - \mu| \leq 0.5) \geq 1 - \frac{1}{(1/3)^2} = -8$. This tells us nothing, since we already know $P(|X - \mu| \leq 0.5) \geq 0$. However, this can happen with Chebychev.

Solution 117: Using the theorem, we have $3e^{-3x}$.

Solution 118: Using the theorem, we have $U_{(4)}(x) = \binom{5}{3,1,2}\left(\frac{1}{10}\right)\left(\frac{x}{10}\right)^3\left(1 - \frac{x}{10}\right)^2 = 2(x/10)^3 - 2(x/10)^4$.

Solution 119: By definition, we have $n = 3$, $j = 1$, and we need to find $f_X(x)$ and $F_X(x)$. However, $f_X(x) = x/18$, and $F_X(a) = \int_0^A x/18dx = x^2/36\Big|_{x=0}^{a} = a^2/36$. Substituting these values in, we have

$$\binom{3}{0,1,2}(x/18)(x^2/36)^0(1 - x^2/36)^2 = x/6 - x^3/108 + x^5/7776,$$

for $0 < x < 6$.

Soluton 120: We have $f_X(x) = 3e^{-3x}$. So, we are looking for $\int_0^\infty 3e^{x(t-3)}dx$. Letting $u = t - 3$, we find that we have $\int_0^\infty 3e^{x(t-3)}dx = 3/(3 - t)$.

Solution 121: We are looking for, by definition $\sum_{x=0}^\infty e^{tx}(4/5)^{x-1}(1/5)$. Note that $(4/5)^{-1} = 5/4$, and so rewriting this we have $(1/4)\sum_{x=0}^\infty e^{tx}(4/5)^x$, or $(1/4)\sum_{x=0}^\infty(\frac{4e^t}{5})^x$. Using the geometric sum formula and some algebra, we have $e^t(5 - 4e^t)$.

Solution 122: Using the theorem, we have $e^{(e^t-1)3}$.

Solution 123: Since we have that the height is fixed, we have $Y = 14\pi X^2$, where $X \in [2.3, 2.7]$. We then are looking for $P(Y \leq 275)$. However, this is equivalent to $P(14\pi X^2 \leq 275) \leftrightarrow P(X \leq \sqrt{\frac{275}{14\pi}}) = \frac{\sqrt{\frac{275}{14\pi}} - 2.3}{2.7 - 2.3} = 0.5013$.

Solution 124: We are looking for $P(Y > 0.4) = P(\frac{3}{10}X^{1/3} > 0.4) = P(X > (0.4 \cdot \frac{10}{3})^3)) = 1 - P(X < (0.4 \cdot \frac{10}{3})^3)) = 1 - \frac{(0.4(\frac{10}{3}))^3 - 1.7}{2.6 - 1.7} = 0.2551$.

Solution 125: Since X is uniform, then so is Y. Note that $2(10) + 2 < Y < 2(14) + 2 \rightarrow 22 < Y < 30$. Therefore, we have $f_Y(y) = \frac{1}{30-22} = \frac{1}{8}$ for $22 < Y < 30$, and 0 otherwise.