MAT 3701: Assignment 4 Solutions to Selected Problems

1 Exercises from the text

The solutions to these, especially those involving multiple integration, would be easier to illustrate on the board, so ask me in class if there are any you do not understand how to do.

2 Additional Problems

2.1 Probability Spaces

The following problems are based on the axioms of a probability space. All proofs must be logical, rigorous, and complete!

1. Given a measure on $\mathbb{R}$, prove that every open interval is an event (that is, measurable) if and only if every closed interval is an event. (Hint: Use the closure of a σ-algebra under countable unions and intersections.) Note that this result extends immediately to any measure space $\mathcal{M} \in \mathbb{R}$. We will be concerned with probability spaces, of course, but the above result does not depend on the measure of the entire space being 1.

   First assume that every open interval is measurable. Let $[a, b]$ be an arbitrary closed interval. Observe that $[a, b] = \bigcap_{n=1}^{\infty}(a - \frac{1}{n}, b + \frac{1}{n})$; thus, since $[a, b]$ is the intersection of countably many measurable sets, it is measurable by the properties of a σ-algebra. Here is an even easier proof: $[a, b] = \mathbb{R} \setminus [(-\infty, a) \cup (b, \infty)]$; thus, $[a, b]$ is the complement of the union of two open intervals.

   Next, assume every closed interval is measurable, and let (a, b) be an arbitrary open interval. Since $(-\infty, a]$ and $[b, \infty)$ are generally considered to be closed intervals, a very similar simple proof works to prove (a, b) is measurable. Alternatively, observe that $(a, b) = \bigcup_{n=k}^{\infty}[a + \frac{1}{n}, b - \frac{1}{n}]$, where k is sufficiently large that $a + \frac{1}{k} < b - \frac{1}{k}$. (These observations of how to write open intervals as unions of closed intervals and closed intervals as intersections of open intervals are useful in their own right. Even though they are not needed to prove this result, we will need them for the problems below.)

In the problems that follow, let ρ be a probability measure on the positive reals, $(0, \infty)$, and define the function $F : (0, \infty) \to [0, \infty)$ by $F(t) = \rho((0, t])$. (You should recognize that if the random variable T is defined by the outcomes of this sample space, then F is the cumulative distribution function of T.)

Note that we are implicitly assuming that the intervals $(0, t]$ are measurable, since we have assigned probabilities to them. By the closure of the σ-algebra of measurable sets under complementation, we deduce that the intervals $(t_1, t_2]$ are also measurable (since $(t_1, t_2] = (0, t_2]) \setminus (0, t_1])$.

2. Prove that F must be an increasing function on $(0, \infty)$. (Recall that a function $f : (a, b) \to \mathbb{R}$ is *increasing* if, for all $x_1, x_2 \in (a, b)$, $x_1 < x_2 \Rightarrow f(x_1) \leq f(x_2)$. The function f is *strictly* increasing if the inequality between the outputs is strict. Note that the function F of this exercise need not be strictly increasing.)

   Let $t_1 < t_2$. $F(t_2) = \rho((0, t_2]) = \rho((0, t_1]) + \rho(t_1, t_2]) = F(t_1) + \rho(t_1, t_2])$, by the countable additivity of measure (and the definition of the function F). Since probability measure is non-negative, $\rho(t_1, t_2]) \geq 0$; hence, $F(t_1) < F(t_2)$.

3. Prove that F must be right continuous on $(0, \infty)$; that is, $\lim_{x \to t^+} F(x) = F(t)$. (Hint: Use the result on intersections of nested events from the previous problem set.)

   This proof must be done in two steps. We wish to calculate $\lim_{x \to t^+} F(x)$ as the probability of a *countable* intersection of nested sets, but the limit is continuous, not sequential. So first we must show that this limit exists. Once its existence is established it may be calculated as the limit of a sequence of values $F(x_1), F(x_2), F(x_3), \ldots$, where $x_n \to t^+$. The reason $\lim_{x \to t^+} F(x)$ exists is that, being a probability, $F(x)$ is bounded below by 0. Thus, the set of values $\{F(x) : x > t\}$ has a greatest lower bound. Furthermore, $F(x)$ decreases as x decreases toward t, so $F(x)$ much approach this greatest lower bound as its limit. To show the limit is $F(t)$, use the fact that $\lim_{n \to \infty} F(t + \frac{1}{n}) = \lim_{n \to \infty} \rho\left((0, t + \frac{1}{n}]\right) = \rho\left(\bigcap_{n=1}^{\infty}(0, t + \frac{1}{n}]\right) = \rho(0, t] = F(t)$. The second equality follows from the result on intersections of nested sets mentioned in the hint.

4. Prove that if F is continuous function, then $\rho((0, t)) = F(t)$ as well. (Hint: Use the result on unions of nested events from the previous problem set.) Conclude that $\rho(\{t\}) = 0$, for any $t > 0$. (Note that condition that F be continuous is not always satisfied; there are probability measures such that $F(t) = \rho((0, t]) > \rho((0, t))$ In this case, $\rho(\{t\}) > 0$.)

   $\rho((0, t)) = \rho\left(\bigcup_{k=1}^{\infty}(0, t - \frac{1}{n}]\right) = \lim_{n \to \infty} \rho\left((0, t - \frac{1}{n}]\right) = \lim_{n \to \infty} F(t - \frac{1}{n})$, by the result mentioned in the hint and the definition of F. Since F is continuous, $\lim_{n \to \infty} F(t - \frac{1}{n}) = F(t)$. We conclude that $\rho(\{t\}) = \rho((0, t] \setminus (0, t)) = \rho((0, t]) - \rho((0, t)) = F(t) - F(t) = 0$.

5. Prove the converse of the result above: if $\rho(\{t\}) = 0$ for all $t \in (0, \infty)$, then F is continuous.

   First use the fact that $\{F(x) : x < t\}$ is bounded above and the fact that $F(x)$ increases as $x \to t^-$ to show that $\lim_{x \to t^-} F(x)$ exists. Then reverse the string of equalities above to show that this limit is $F(t)$. I leave the details to you. Combining this result with the right-continuity of F (proven above) we obtain that F is continuous.

6. Prove that the probability an atom decays at any particular instant must be 0; conclude that for the measure giving the probability the atom decays during any interval of time, F as defined above is continuous. Why does this not cause a contradiction to the probability of the entire sample space, $(0, \infty)$, being 1?

As discussed in class, the probability that an atom decays in the interval $(0, t]$ is given by $F(t) = 1 - e^{-kt}$, for some positive decay constant k. This is a continuous function, so the probability the atom decays exactly at some particular time is zero, by the result of Problem 4. There is no contradiction, because the probability the atom decays in a given interval is the integral of its differential probability, $F'(t)dt$, not a discrete sum.

2.2 Random Variables and Vectors

7. Prove that if X is a discrete random variable, then $\{x \in \mathbb{R} : P(X = x) > 0\} \neq \emptyset$. (Hint: Use proof by contradiction.)

In this case we will get a contradiction, because a discrete sum of zeros, even if there are infinitely many, is zero. (Why? It is a consequence of the definition of an infinite sum as a limit of partial sums.) Here is the proof in the case that there are infinitely many values in the range of X; minor modifications give the proof in the finite case. Let $x_1, x_2, x_3, \ldots$ be the values in the range of X. Suppose, by way of contradiction, that $\{x \in \mathbb{R} : P(X = x) > 0\} = \emptyset$. For the set $\{x \in \mathbb{R} : P(X = x) > 0\}$ to be empty means that $P(X = x_i) = 0$ for all $i = 1, 2, 3, \ldots$. Then using countable additivity we obtain $1 = \rho(S) = \rho(\bigcup_{i=1}^{\infty} X^{-1}(x_i)) = \sum_{i=1}^{\infty} P(X = x_i) = \sum_{i=1}^{\infty} 0 = 0$, an obvious contradiction.

8. Suppose a fair coin is tossed 5 times. Define the random variable X to be the number of heads obtained. What is the probability that $X = x$, for $x = 0, 1, 2, 3, 4$, or 5? (This gives a formula for the probability density (distribution) of X.)

The key to this problem, as with many other discrete random variables, is to define the probability space on which X depends so that all the outcomes are equally likely. In this case, consider the set of all possible sequences of tosses, of which there are 2^5. (In general, for n tosses, there are 2^n.) These are all equally likely, so we just need to count the outcomes in $X^{-1}(x)$ for each possible value of x. The number of sequences for which the total number of heads is x is just the number of ways to choose x trials on which a head is obtained. (More precisely, to partition the 5 trials into two sets of size x and $5 - x$, the former containing heads and the latter containing tails. It is useful to remember that choosing a certain number of items out of a total is a special case of a partition.) Thus, $P(X = x) = \binom{5}{x}\frac{1}{2^5}$, for $x = 0, 1, 2, 3, 4, 5$, and 0 otherwise. (In general, for n trials, $P(X = x) = \binom{n}{x}\frac{1}{2^n}$, for $x = 0, 1, 2, 3, 4, n$, and 0 otherwise.)

Another way to think of this is to note that on each trial, the probability of obtaining a head is $\frac{1}{2}$, as is the probability of obtaining a tail. Since the trials are independent, the probability of any particular sequence of heads and tails in n trials is the product of the probabilities for each trial, thus $\left(\frac{1}{2}\right)^n$.

Yet another approach is to define independent random variables X_i, $i = 1, 2, 3, \ldots, n$ by setting $X_i = 1$ if the i^{th} trial is a head and $X_i = 0$, if the i^{th} trial is a tail, where n is the number of trials. Then $X = \sum_{i=1}^{n} X_i$. In other words, a binomial random variable for n trials is the sum of n Bernoulli random variables. This is worth noting, although as of yet we have not developed general methods for calculating the probability distribution of a sum of random variables. We will more fully develop the approach of working with random variables defined as functions of other random variables as we continue our study.

9. A fair die with six faces is tossed 5 times. Let X be the number of sixes that occur. Give a formula for the probability density (distribution) of X.

I will once again consider the various approaches we could take, since it is useful and deepens our understanding to have several ways to look at a situation. In this case, the probability of obtaining a six is $\frac{1}{6}$, whereas the probability of obtaining some other number is $\frac{5}{6}$. Since these probabilities are not equal, the approach of multiplying the probabilities of the independent trials is probably easier. Thus, the probability of any particular sequence of tosses *containing exactly x sixes* is $\left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{5-x}$, for $x = 0, 1, 2, 3, 4, 5$, and 0 otherwise. (The trials on which the sixes are obtained does not affect this probability, since multiplication is commutative. In general, for n tosses, it would be $\left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{n-x}$, for $x = 0, 1, 2, \ldots, n$, and 0 otherwise.) Thus $f(x) = P(X = x) = \binom{5}{x}\left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{5-x}$, for $x = 0, 1, 2, 3, 4, 5$, and 0 otherwise. (In general, for n tosses, $f(x) = P(X = x) = \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{n-x}$, for $x = 0, 1, 2, \ldots, n$, and 0 otherwise.)

Alternatively, it is possible to consider the sample space of all possible sequences of tosses; that is, we distinguish all of the tosses that are not six instead of lumping them all together. For n trials this gives 6^n equally likely outcomes. The number of these containing x sixes is $\binom{n}{x}5^{n-x}$, since there are x trials to choose on which the sixes are obtained and 5 possible numbers that could occur on each trial where a six is not obtained. As you can see, dividing the number of outcomes in the event that $X = x$ by the total number in the sample space gives the same result obtained above.

Finally, it is also possible to view X as the sum of n Bernoulli random variables for which the probability of "success" is $p = \frac{1}{6}$.

10. A fair coin is tossed until a head is obtained. Let X denote the number of the toss on which this head occurs.

I will discuss parts (c) and (d), in which we are to calculate the probability that the number of tosses required is even. As demonstrated in class, it is easiest to do these two calculations together, since the sum of these probabilities must be 1. Let P be the probability that X is even, and let Q be the probability that X is odd. Observe that $P = \sum_{k=1}^{\infty} \frac{1}{2^{2k}}$, and $Q = \sum_{k=1}^{\infty} \frac{1}{2^{2k-1}}$; thus, $P = \frac{1}{2}Q$. Putting this equation together with $P + Q = 1$ yields $P = \frac{1}{3}$ and $Q = \frac{2}{3}$. There is no need to sum the series! On the other hand, if you didn't think of that, it is not hard to sum the series, since they are geometric. I leave those calculations to you.

11. A fair six-sided die is tossed until a six is obtained. Let X be the number of the toss on which this six occurs.

(a) Give a formula for the probability density (distribution) of X. Using either of the approaches discussed in the solutions to earlier problems (and I suggest you try both of them), we see that $f(x) = P(X = x) = \left(\frac{5}{6}\right)^{x-1}\left(\frac{1}{6}\right)$, for $x = 1, 2, 3, \ldots$, and 0 otherwise. (You must fail $x - 1$ times and succeed on the last trial.)

(b) What is the probability that X is even?

(c) What is the probability that X is odd?

Again let us consider these two questions together, letting P be the probability that X is even and Q the probability that X is odd. Then $P = \frac{1}{6}\sum_{k=1}^{\infty}\left(\frac{5}{6}\right)^{2k-1}$ and $Q = \frac{1}{6}\sum_{k=1}^{\infty}\left(\frac{5}{6}\right)^{2k-2}$. These geometric series are not difficult to evaluate, but instead we can notice that $P = \frac{5}{6}Q$, so solving $Q + \frac{5}{6}Q = 1$ (or, alternatively, $P + \frac{6}{5}P = 1$) we obtain $P = \frac{5}{11}$ and $Q = \frac{6}{11}$. (An easy way to see the solutions is that the "total number of parts" must be $5 + 6 = 11$.)

12. Two fair six-sided dice are tossed. Let X be their sum. Give a formula for the probability density (distribution) of X.

I suggest making a 6×6 matrix of the sums; the pattern in the number of ways to obtain each sum will be apparent.

13. A fair die with six faces is rolled five times.

Remark: Although we have not yet introduced a random vector, here we are considering a multi-variate situation, since we are keeping track of several numbers.

(a) What is the probability that one of each number from one to five is rolled (but no six).

Out of the 6^5 possible rolls (or tosses, or whatever you want to call them), there are

5! that satisfy this criterion (5 possibilities for the first roll, 4 for the second, and so forth). So the probability of this event is $\frac{5!}{6^5} = \left(\frac{5}{6}\right)\left(\frac{4}{6}\right)\left(\frac{3}{6}\right)\left(\frac{2}{6}\right)\left(\frac{1}{6}\right)$.

(b) What is the probability that the five rolls are all different? Now sequences including a six are included, so the probability of this event is $\left(\frac{6}{6}\right)\left(\frac{5}{6}\right)\left(\frac{4}{6}\right)\left(\frac{3}{6}\right)\left(\frac{2}{6}\right) = \frac{5!}{6^4}$.

(c) What is the probability that two threes are rolled first, followed by three ones? There is only one outcome (specific sequence of rolls) in this event, so its probability is $\frac{1}{6^5}$.

(d) What is the probability that two threes and three ones are rolled (in any order)? Since any order is allowed for this event, we must count the ways to partition the 5 trials into sets of 2 (for the threes) and 3 (for the ones). So the probability of this event is $\binom{5}{2}\frac{1}{6^5} = \binom{5}{2,3}\frac{1}{6^5} = \binom{5}{3}\frac{1}{6^5}$. (Here I have emphasized the process of choosing as one of partition, because this will help us see how the solution to the next problem is a straightforward generalization of our reasoning that simply uses partitions with more "parts.")

14. A fair die with six faces is rolled seven times.

(a) What is the probability that two threes are rolled first, followed by one six, and then four ones? This event has one outcome out of 6^7 possible outcomes, so its probability is $\frac{1}{6^7}$.

(b) What is the probability that two threes, one six, and four ones are rolled (in any order)? We must count the number of partitions of our 7 trials into sets of 2, 1, and 4; therefore, the probability of this event is $\binom{7}{2,1,4}\frac{1}{6^7}$. (Note: Don't forget how to actually calculate the number of partitions denoted by the trinomial coefficient!)

15. A fair cubical die is tossed seven times. Let X_1 be the number of "ones" obtained among the tosses, let X_2 be the number of "twos" obtained, and so on. Let $\overrightarrow{X} = (X_1, X_2, X_3, X_4, X_5, X_6)$.

Remark: Now we have defined a random vector with 6 coordinates!

(a) Compute the probability density (distribution) of $\overrightarrow{X}$ (which is generally called the *joint probability density (distribution)* of $X_1, X_2, X_3, X_4, X_5, X_6$). $\overrightarrow{X} = (x_1, x_2, x_3, \ldots, x_6) \Leftrightarrow X_i = x_i$, for $i = 1, 2, 3, \ldots, 6$. In other words, there must be x_1 ones, x_2 twos, and so forth. To count how many outcomes are in this event is to count the number of ways to partition the 7 trials into subsets of size x_1, x_2, x_3, x_4, x_5, and x_6 where the various numbers are obtained. (Remember that each outcome is a specific sequence of rolls and has probability $\frac{1}{6^7}$.) Thus $f(x_1, x_2, x_3, x_4, x_5, x_6) = P((X_1, X_2, X_3, X_4, X_5, X_6) = (x_1, x_2, x_3, x_4, x_5, x_6)) = \binom{7}{x_1, x_2, x_3, x_4, x_5, x_6}\frac{1}{6^7}$.

(b) Compute the density of X_i for each $i = 1, 2, 3, \ldots, 7$. (X_i is generally called the i^{th} *marginal probability density (distribution)* of $\overrightarrow{X}$.) Each X_i is binomial with probability distribution $f_i(x) = \binom{7}{x}\left(\frac{1}{6}\right)^x\left(\frac{5}{6}\right)^{7-x}$.

(c) Let f denote the probability density (distribution) of $\overrightarrow{X}$, and let f_i denote the i^{th} marginal probability density (distribution) of $\overrightarrow{X}$. Show that $f_i(x) = \sum_{x_i=x} f(x_1, x_2, x_3, \ldots, x_6)$. This equation follows from the countable additivity of probability measure.

18. More generally, let l, m, and n be positive integers with $l \leq n$ and $m \leq n$. Suppose there is a population of n objects, and l of these objects are of a marked type. A sample of size m is drawn from the objects *without* replacement. Let X be the number of marked objects in the sample.

As examples, we might have n balls in an urn, of which l are red and the rest are black, as in the problem above. A sample of m balls is drawn from the urn. X is the number of red balls in the sample. In a more realistic application, there is a population of n deer in a forest, of which l have been marked with a tag in an ear. Hunters shoot m deer. X is the number of tagged deer that are shot. (This number can be used to estimate the total population, as we will see later on.)

Compute the density function of X (in terms of l, m, and n). A random variable with this density is said to be *hypergeometrically distributed.* Compare the hypergeometric distribution with the binomial distribution: if a sample of size m is drawn *with replacement* from a population of size n that contains l marked objets, then the number of marked objects obtained is binomially distributed with parameters m and $\frac{l}{n}$. The hypergeometric density can be written in two convenient forms, one entirely in terms of combinations, and the other more analogous to the formula for the binomial density. Give both forms in your answer.

Clearly X takes values between 0 and l, if $m \geq l$, and between 0 and m, if $m \leq l$. In short, X takes values between 0 and $\min(l, m)$. For each such value x, the probability that $X = x$ is $\frac{\binom{l}{x}\binom{n-l}{m-x}}{\binom{n}{m}}$. This formula was calculated by simply dividing the size of the event by the size of the sample space.

If you write out the factorials in the combination formulas, cancel, and rearrange as follows:

$$\frac{\binom{l}{x}\binom{n-l}{m-x}}{\binom{n}{m}} =$$

$$\frac{[l(l-1)(l-2)\cdots(l-x+1)][(n-l)(n-l-1)(n-l-2)\cdots(n-l-(m-x)+1)]}{n(n-1)(n-2)\cdots(n-m+1)} \cdot \frac{m!}{x!(m-x)!} =$$

$$\frac{[l(l-1)(l-2)\cdots(l-x+1)][(n-l)(n-l-1)(n-l-2)\cdots(n-l-(m-x)+1)]}{n(n-1)(n-2)\cdots(n-m+1)} \cdot \binom{m}{x},$$

you obtain the following illuminating formula:

$$f(x) = \binom{m}{x}\left(\frac{l}{n}\right)\left(\frac{l-1}{n-1}\right)\cdots\left(\frac{l-x+1}{n-x+1}\right)\left(\frac{n-l}{n-x}\right)\left(\frac{n-l-1}{n-x-1}\right)\cdots\left(\frac{n-l-(m-x)+1}{n-m+1}\right).$$

Compare this formula to the binomial density that would result from choosing m samples *with replacement* from the same population:

$$f(x) = \binom{m}{x}\left(\frac{l}{n}\right)^x\left(\frac{n-l}{n}\right)^{m-x}.$$

In both densities, $\binom{m}{x}$ is the number of ways of placing x "successes" among m choices. The factors of $\frac{l}{n}$ and $\frac{n-l}{n}$ give the probability of x successes and $m-x$ failures, which are independent. In the hypergeometric case (without replacement), the successive factors also give the probability of x successes and $m-x$ failures, but in this case each probability is conditional on the samples that have been previously removed from the population.

19. A point is picked at random from a disk of radius 1. Let R be the distance of the point from the center of the target.

(a) Compute the (cumulative) distribution function of R.

$$F(r) = P(R \leq r) = \begin{cases} 0, & \text{if } r < 0 \\ \frac{\pi r^2}{\pi} = r^2, & \text{if } 0 \leq r \leq 1 \\ 1, & \text{if } 1 < r \end{cases}$$

(b) Give a probability density for R. $f(r) = 2r$, if $0 \leq r \leq 1$, and 0, otherwise, gives a density, since integrating this function gives the cumulative distribution function.