

# Stat 110 Strategic Practice 7, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

## 1 Joint, Conditional, and Marginal Distributions

1. A random point  $(X, Y, Z)$  is chosen uniformly in the ball  $B = \{(x, y, z) : x^2 + y^2 + z^2 \leq 1\}$ .
  - (a) Find the joint PDF of  $X, Y, Z$ .
  - (b) Find the joint PDF of  $X, Y$ .
  - (c) Find an expression for the marginal PDF of  $X$ , as an integral.
2. Let  $X$  and  $Y$  be i.i.d.  $\text{Unif}(0, 1)$ . Find the expected value and the standard deviation of the distance between  $X$  and  $Y$ .
3. Let  $U_1, U_2, U_3$  be i.i.d.  $\text{Unif}(0, 1)$ , and let  $L = \min(U_1, U_2, U_3)$ ,  $M = \max(U_1, U_2, U_3)$ .
  - (a) Find the marginal CDF and marginal PDF of  $M$ , and the joint CDF and joint PDF of  $L, M$ .

Hint: for the latter, start by considering  $P(L \geq l, M \leq m)$ .
  - (b) Find the conditional PDF of  $M$  given  $L$ .
4. A group of  $n \geq 2$  people decide to play an exciting game of Rock-Paper-Scissors. As you may recall, Rock smashes Scissors, Scissors cuts Paper, and Paper covers Rock (despite Bart Simpson saying “Good old rock, nothing beats that!”).

Usually this game is played with 2 players, but it can be extended to more players as follows. If exactly 2 of the 3 choices appear when everyone reveals their choice, say  $a, b \in \{\text{Rock}, \text{Paper}, \text{Scissors}\}$  where  $a$  beats  $b$ , the game is decisive: the players who chose  $a$  win, and the players who chose  $b$  lose. Otherwise, the game is indecisive and the players play again.

For example, with 5 players, if one player picks Rock, two pick Scissors, and two pick Paper, the round is indecisive and they play again. But if 3 pick Rock and 2 pick Scissors, then the Rock players win and the Scissors players lose the game.

Assume that the  $n$  players independently and randomly choose between Rock, Scissors, and Paper, with equal probabilities. Let  $X, Y, Z$  be the number of players who pick Rock, Scissors, Paper, respectively in one game.

(a) Find the joint PMF of  $X, Y, Z$ .

(b) Find the probability that the game is decisive. Simplify your answer (it should not involve a sum of many terms).

(c) What is the probability that the game is decisive for  $n = 5$ ? What is the limiting probability that a game is decisive as  $n \rightarrow \infty$ ? Explain briefly why your answer makes sense.

5. A chicken lays  $n$  eggs. Each egg independently does or doesn't hatch, with probability  $p$  of hatching. For each egg that hatches, the chick does or doesn't survive (independently of the other eggs), with probability  $s$  of survival. Let  $N \sim \text{Bin}(n, p)$  be the number of eggs which hatch,  $X$  be the number of chicks which survive, and  $Y$  be the number of chicks which hatch but don't survive (so  $X + Y = N$ ). Find the marginal PMF of  $X$ , and the joint PMF of  $X$  and  $Y$ . Are they independent?

# Stat 110 Strategic Practice 7 Solutions, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

## 1 Joint, Conditional, and Marginal Distributions

1. A random point  $(X, Y, Z)$  is chosen uniformly in the ball  $B = \{(x, y, z) : x^2 + y^2 + z^2 \leq 1\}$ .

(a) Find the joint PDF of  $X, Y, Z$ .

Just as in 2 dimensions uniform in a region means that probability is proportional to length, here probability is proportional to volume. That is,

$$P((X, Y, Z) \in A) = c \cdot \text{Volume}(A)$$

if  $A$  is contained in  $B$ , where  $c$  is a constant. Letting  $A = B$ , we have that  $\frac{1}{c}$  is  $\frac{4}{3}\pi$ , the volume of the ball. So the joint PDF of  $(X, Y, Z)$  is

$$f(x, y, z) = \begin{cases} \frac{3}{4\pi}, & \text{if } x^2 + y^2 + z^2 \leq 1; \\ 0, & \text{otherwise.} \end{cases}$$

(b) Find the joint PDF of  $X, Y$ .

We just need to “integrate out” the  $z$  from the joint PDF of  $X, Y, Z$ . The limits of integration are found by noting that for any  $(x, y)$ , we need to have  $z$  satisfy  $x^2 + y^2 + z^2 \leq 1$ .

$$\begin{aligned} f_{X,Y}(x, y) &= \int_{-\infty}^{\infty} f(x, y, z) dz \\ &= \frac{3}{4\pi} \int_{-\sqrt{1-x^2-y^2}}^{\sqrt{1-x^2-y^2}} dz \\ &= \frac{3}{2\pi} \sqrt{1-x^2-y^2}, \end{aligned}$$

for  $x^2 + y^2 \leq 1$  (and the PDF is 0 otherwise).

(c) Find an expression for the marginal PDF of  $X$ , as an integral.

We can integrate out  $y, z$  from the joint PDF of  $X, Y, Z$ , or integrate out  $y$  from the joint PDF of  $X, Y$ . Using the result of (b), we have for  $-1 \leq x \leq 1$  that the marginal PDF of  $X$  is

$$f_X(x) = \frac{3}{2\pi} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \sqrt{1-x^2-y^2} dy.$$

2. Let  $X$  and  $Y$  be i.i.d.  $\text{Unif}(0, 1)$ . Find the expected value and the standard deviation of the distance between  $X$  and  $Y$ .

Let  $W = |X - Y|$ . By 2-D LOTUS,

$$E(W) = \int_0^1 \int_0^1 |x - y| dx dy.$$

Split this into two parts (to get rid of the absolute values):  $x < y$  and  $x \geq y$  (i.e., break the square into two triangles). By symmetry the integral over  $x < y$  equals the integral over  $x > y$ , so

$$E(W) = 2 \int_0^1 \int_0^y (y - x) dx dy = 2 \int_0^1 \frac{y^2}{2} dy = \frac{1}{3}.$$

Next, we find  $E(W^2)$ . This can either be done by computing the double integral

$$E(W^2) = \int_0^1 \int_0^1 (x - y)^2 dx dy,$$

or by writing

$$E(W^2) = E(X - Y)^2 = EX^2 + EY^2 - 2E(XY),$$

which is

$$2E(X^2) - 2(EX)^2 = 2\text{Var}(X) = \frac{1}{6},$$

since  $E(XY) = E(X)E(Y)$  for  $X, Y$  independent, and  $E(X) = E(Y)$  and  $E(X^2) = E(Y^2)$  (as  $X$  and  $Y$  have the same distribution). Thus,  $E(W) = 1/3$ ,

$$\text{Var}(W) = E(W^2) - (E(W))^2 = \frac{1}{18},$$

and the standard deviation of the distance between  $X$  and  $Y$  is  $\frac{1}{\sqrt{18}} = \frac{1}{3\sqrt{2}}$ .

3. Let  $U_1, U_2, U_3$  be i.i.d.  $\text{Unif}(0, 1)$ , and let  $L = \min(U_1, U_2, U_3)$ ,  $M = \max(U_1, U_2, U_3)$ .
- (a) Find the marginal CDF and marginal PDF of  $M$ , and the joint CDF and joint PDF of  $L, M$ .

Hint: for the latter, start by considering  $P(L \geq l, M \leq m)$ .

The event  $M \leq m$  is the same as the event that all 3 of the  $U_j$  are at most  $m$ , so the CDF of  $M$  is  $F_M(m) = m^3$  and the PDF is  $f_M(m) = 3m^2$ , for  $0 \leq m \leq 1$ .

The event  $L \geq l, M \leq m$  is the same as the event that all 3 of the  $U_j$  are between  $l$  and  $m$  (inclusive), so

$$P(L \geq l, M \leq m) = (m - l)^3$$

for  $m \geq l$  with  $m, l \in [0, 1]$ . By the axioms of probability, we have

$$P(M \leq m) = P(L \leq l, M \leq m) + P(L > l, M \leq m).$$

So the joint CDF is

$$P(L \leq l, M \leq m) = m^3 - (m - l)^3,$$

for  $m \geq l$  with  $m, l \in [0, 1]$ . The joint PDF is obtained by differentiating this with respect to  $l$  and then with respect to  $m$  (or vice versa):

$$f(l, m) = 6(m - l),$$

for  $m \geq l$  with  $m, l \in [0, 1]$ . As a check, note that getting the marginal PDF of  $M$  by finding  $\int_0^m f(l, m) dl$  does recover the PDF of  $M$  (the limits of integration are from 0 to  $m$  since the min can't be more than the max).

- (b) Find the conditional PDF of  $M$  given  $L$ .

The marginal PDF of  $L$  is  $f_L(l) = 3(1 - l)^2$  for  $0 \leq l \leq 1$  since  $P(L > l) = P(U_1 > l, U_2 > l, U_3 > l) = (1 - l)^3$  (alternatively, use the PDF of  $M$  together with the symmetry that  $1 - U_j$  has the same distribution as  $U_j$ , or integrate out  $m$  in the joint PDF of  $L, M$ ). So the conditional PDF of  $M$  given  $L$  is

$$f_{M|L}(m|l) = \frac{f(l, m)}{f_L(l)} = \frac{2(m - l)}{(1 - l)^2},$$

for all  $m, l \in [0, 1]$  with  $m \geq l$ .

4. A group of  $n \geq 2$  people decide to play an exciting game of Rock-Paper-Scissors. As you may recall, Rock smashes Scissors, Scissors cuts Paper, and Paper covers Rock (despite Bart Simpson saying “Good old rock, nothing beats that!”).

Usually this game is played with 2 players, but it can be extended to more players as follows. If exactly 2 of the 3 choices appear when everyone reveals their choice, say  $a, b \in \{Rock, Paper, Scissors\}$  where  $a$  beats  $b$ , the game is decisive: the players who chose  $a$  win, and the players who chose  $b$  lose. Otherwise, the game is indecisive and the players play again.

For example, with 5 players, if one player picks Rock, two pick Scissors, and two pick Paper, the round is indecisive and they play again. But if 3 pick Rock and 2 pick Scissors, then the Rock players win and the Scissors players lose the game.

Assume that the  $n$  players independently and randomly choose between Rock, Scissors, and Paper, with equal probabilities. Let  $X, Y, Z$  be the number of players who pick Rock, Scissors, Paper, respectively in one game.

- (a) Find the joint PMF of  $X, Y, Z$ .

The joint distribution of  $X, Y, Z$  is

$$P(X = a, Y = b, Z = c) = \frac{n!}{a!b!c!} \left(\frac{1}{3}\right)^{a+b+c}$$

where  $a, b, c$  are any nonnegative integers with  $a + b + c = n$ , since  $(1/3)^{a+b+c}$  is the probability of any specific configuration of choices for each player with the right numbers in each category, and the coefficient in front counts the number of distinct ways to permute such a configuration.

Alternatively, we can write the joint PMF as

$$P(X = a, Y = b, Z = c) = P(X = a)P(Y = b|X = a)P(Z = c|X = a, Y = b),$$

where for  $a + b + c = n$ ,  $P(X = a)$  can be found from the  $\text{Bin}(n, 1/3)$  PMF,  $P(Y = b|X = a)$  can be found from the  $\text{Bin}(n - a, 1/2)$  PMF, and  $P(Z = c|X = a, Y = b) = 1$ .

This is a Multinomial( $n, (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ ). distribution.

(b) Find the probability that the game is decisive. Simplify your answer (it should not involve a sum of many terms).

Hint: using symmetry, the probability can be written as 3 times a certain sum. To do the summation, use the binomial theorem or the fact that  $\sum_{k=0}^n \binom{n}{k} = 2^n$ .

The game is decisive if and only if exactly one of  $X, Y, Z$  is 0. These cases are disjoint so by symmetry, the probability is 3 times the probability that  $X$  is zero and  $Y$  and  $Z$  are nonzero. Note that if  $X = 0$  and  $Y = k$ , then  $Z = n - k$ . This gives

$$\begin{aligned} P(\text{decisive}) &= 3 \sum_{k=1}^{n-1} \frac{n!}{0!k!(n-k)!} \left(\frac{1}{3}\right)^n \\ &= 3 \left(\frac{1}{3}\right)^n \sum_{k=1}^{n-1} \binom{n}{k} \\ &= \frac{2^n - 2}{3^{n-1}} \end{aligned}$$

since  $\sum_{k=1}^{n-1} \binom{n}{k} = -1 - 1 + \sum_{k=0}^n \binom{n}{k} = 2^n - 2$  (by the binomial theorem or the fact that a set with  $n$  elements has  $2^n$  subsets). As a check, when  $n = 2$  this reduces to  $2/3$ , which makes sense since for 2 players, the game is decisive if and only if the two players do not pick the same choice.

(c) What is the probability that the game is decisive for  $n = 5$ ? What is the limiting probability that a game is decisive as  $n \rightarrow \infty$ ? Explain briefly why your answer makes sense.

For  $n = 5$ , the probability is  $(2^5 - 2)/3^4 = 30/81 \approx 0.37$ . As  $n \rightarrow \infty$ ,  $(2^n - 2)/3^{n-1} \rightarrow 0$ , which make sense since if the number of players is very large, it is very likely that there will be at least one of each of Rock, Paper, and Scissors.

5. A chicken lays  $n$  eggs. Each egg independently does or doesn't hatch, with probability  $p$  of hatching. For each egg that hatches, the chick does or doesn't survive (independently of the other eggs), with probability  $s$  of survival. Let  $N \sim \text{Bin}(n, p)$  be the number of eggs which hatch,  $X$  be the number of chicks which survive, and  $Y$  be the number of chicks which hatch but don't survive (so  $X + Y = N$ ). Find the marginal PMF of  $X$ , and the joint PMF of  $X$  and  $Y$ . Are they independent?

Marginally we have  $X \sim \text{Bin}(n, ps)$ , as shown on a previous homework problem using a story proof (the eggs can be thought of as independent Bernoulli trials with probability  $ps$  of success for each). Here  $X$  and  $Y$  are *not* independent, unlike in the chicken-egg problem from class (where  $N$  was Poisson). This follows immediately from thinking about an *extreme case*: if  $X = n$ , then clearly  $Y = 0$ . So they are not independent:  $P(Y = 0) < 1$ , while  $P(Y = 0|X = n) = 1$ .

To find the joint distribution, condition on  $N$  and note that only the  $N = i + j$  term is nonzero: for any nonnegative integers  $i, j$  with  $i + j \leq n$ ,

$$\begin{aligned} P(X = i, Y = j) &= P(X = i, Y = j|N = i + j)P(N = i + j) \\ &= P(X = i|N = i + j)P(N = i + j) \\ &= \binom{i + j}{i} s^i (1 - s)^j \binom{n}{i + j} p^{i + j} (1 - p)^{n - i - j} \\ &= \frac{n!}{i!j!(n - i - j)!} (ps)^i (p(1 - s))^j (1 - p)^{n - i - j}. \end{aligned}$$

If we let  $Z$  be the number of eggs which don't hatch, then from the above we have that  $(X, Y, Z)$  has a Multinomial( $n, (ps, p(1 - s), 1 - p)$ ) distribution, which makes sense intuitively since each egg independently falls into 1 of 3 categories: hatch-and-survive, hatch-and-don't-survive, and don't-hatch, with probabilities  $ps, p(1 - s), 1 - p$  respectively.



## Stat 110 Homework 7, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

1. (a) A stick is broken into three pieces by picking two points independently and uniformly along the stick, and breaking the stick at those two points. What is the probability that the three pieces can be assembled into a triangle?

Hint: a triangle can be formed from 3 line segments of lengths  $a, b, c$  if and only if  $a, b, c \in (0, 1/2)$ . The probability can be interpreted geometrically as proportional to an area in the plane, avoiding all calculus, but make sure for that approach that the distribution of the random point in the plane is Uniform over some region.

(b) Three legs are positioned uniformly and independently on the perimeter of a round table. What is the probability that the table will stand?

2. Let  $(X, Y)$  be a uniformly random point in the triangle in the plane with vertices  $(0, 0), (0, 1), (1, 0)$ . Find the joint PDF of  $X$  and  $Y$ , the marginal PDF of  $X$ , and the conditional PDF of  $X$  given  $Y$ .

3. Let  $X, Y$  be i.i.d.  $\text{Expo}(\lambda)$ . Find  $E|X - Y|$  in two ways: (a) using a 2-D LOTUS and (b) using the memoryless property without any calculus.

4. Two students,  $A$  and  $B$ , are working independently on homework (not necessarily for the same class). Student  $A$  takes  $Y_1 \sim \text{Expo}(\lambda_1)$  hours to finish his or her homework, while  $B$  takes  $Y_2 \sim \text{Expo}(\lambda_2)$  hours.

(a) Find the CDF and PDF of  $\frac{Y_1}{Y_2}$ , the ratio of their problem-solving times.

(b) Find the probability that  $A$  finishes his or her homework before  $B$  does.

5. The bus company from Blissville decides to start service in Blotchville, sensing a promising business opportunity. Meanwhile, Fred has moved back to Blotchville, inspired by a close reading of *I Had Trouble in Getting to Solla Sollew*. Now when Fred arrives at the bus stop, either of two independent bus lines may come by (both of which take him home). The Blissville company's bus arrival times are exactly 10 minutes apart, whereas the time from one Blotchville company bus to the next is  $\text{Expo}(\frac{1}{10})$ . Fred arrives at a uniformly random time on a certain day.

(a) What is the probability that the Blotchville company bus arrives first?

Hint: one good way is to use the continuous Law of Total Probability.

(b) What is the CDF of Fred's waiting time for a bus?

6. Emails arrive in an inbox according to a Poisson process with rate  $\lambda$  (so the number of emails in a time interval of length  $t$  is distributed as  $\text{Pois}(\lambda t)$ , and the numbers of emails arriving in disjoint time intervals are independent). Let  $X, Y, Z$  be the numbers of emails that arrive from 9 am to noon, noon to 6 pm, and 6 pm to midnight (respectively) on a certain day.

(a) Find the joint PMF of  $X, Y, Z$ .

(b) Find the conditional joint PMF of  $X, Y, Z$  given that  $X + Y + Z = 36$ .

(c) Find the conditional PMF of  $X + Y$  given that  $X + Y + Z = 36$ , and find  $E(X + Y | X + Y + Z = 36)$  and  $\text{Var}(X + Y | X + Y + Z = 36)$  (conditional expectation and conditional variance given an event are defined in the same way as expectation and variance, using the conditional distribution given the event in place of the unconditional distribution).

7. Shakespeare wrote a total of 884647 words in his known works. Of course, many words are used more than once, and the number of distinct words in Shakespeare's known writings is 31534 (according to one computation). This puts a lower bound on the size of Shakespeare's vocabulary, but it is likely that Shakespeare knew words which he did not use in these known writings.

More specifically, suppose that a new poem of Shakespeare were uncovered, and consider the following (seemingly impossible) problem: give a good prediction of the number of words in the new poem that do not appear anywhere in Shakespeare's previously known works.

The statisticians Ronald Thisted and Bradley Efron studied this problem in a paper called "Did Shakespeare write a newly-discovered poem?", which performed statistical tests to try to determine whether Shakespeare was the author of a poem discovered by a Shakespearean scholar in 1985. A simplified version of their method is developed in the problem below. The method was originally invented by Alan Turing (the founder of computer science) and I.J. Good as part of the effort to break the German Enigma code during World War II.

Let  $N$  be the number of distinct words that Shakespeare knew, and assume these words are numbered from 1 to  $N$ . Suppose for simplicity that Shakespeare wrote only two plays,  $A$  and  $B$ . The plays are reasonably long and they are of the same length. Let  $X_j$  be the number of times that word  $j$  appears in play  $A$ , and  $Y_j$  be the number of times it appears in play  $B$ , for  $1 \leq j \leq N$ .

(a) Explain why it is reasonable to model  $X_j$  as being Poisson, and  $Y_j$  as being Poisson with the same parameter as  $X_j$ .

(b) Let the numbers of occurrences of the word “eyeball” (which was coined by Shakespeare) in the two plays be independent  $\text{Pois}(\lambda)$  r.v.s. Show that the probability that “eyeball” is used in play  $B$  but not in play  $A$  is

$$e^{-\lambda}(\lambda - \lambda^2/2! + \lambda^3/3! - \lambda^4/4! + \dots).$$

(c) Now assume that  $\lambda$  from (b) is unknown and is itself taken to be a random variable to reflect this uncertainty. So let  $\lambda$  have a PDF  $f_0$ . Let  $X$  be the number of times the word “eyeball” appears in play  $A$  and  $Y$  be the corresponding value for play  $B$ . Assume that the conditional distribution of  $X, Y$  given  $\lambda$  is that they are independent  $\text{Pois}(\lambda)$  r.v.s. Show that the probability that “eyeball” is used in play  $B$  but not in play  $A$  is the alternating series

$$P(X = 1) - P(X = 2) + P(X = 3) - P(X = 4) + \dots$$

Hint: condition on  $\lambda$  and use (b).

(d) Assume that every word’s numbers of occurrences in  $A$  and  $B$  are distributed as in (c), where  $\lambda$  may be different for different words but  $f_0$  is fixed. Let  $W_j$  be the number of words that appear exactly  $j$  times in play  $A$ . Show that the expected number of distinct words appearing in play  $B$  but not in play  $A$  is

$$E(W_1) - E(W_2) + E(W_3) - E(W_4) + \dots$$

(This shows that  $W_1 - W_2 + W_3 - W_4 + \dots$  is an *unbiased* predictor of the number of distinct words appearing in play  $B$  but not in play  $A$ : on average it is correct. Moreover, it can be computed just from having seen play  $A$ , without needing to know  $f_0$  or any of the  $\lambda_j$ . This method can be extended in various ways to give predictions for unobserved plays based on observed plays.)

## Stat 110 Homework 7 Solutions, Fall 2011

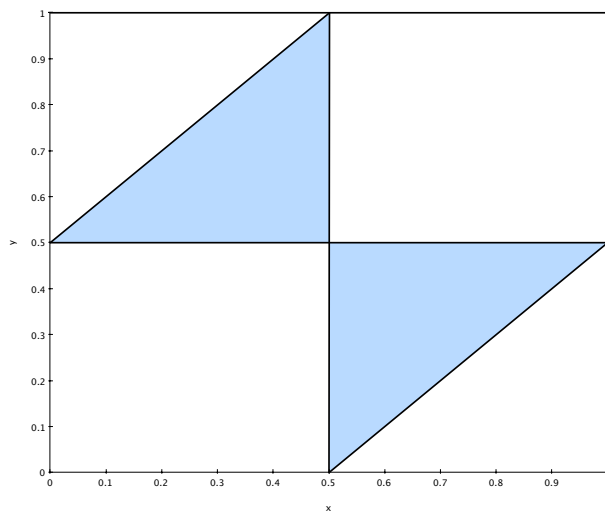
Prof. Joe Blitzstein (Department of Statistics, Harvard University)

1. (a) A stick is broken into three pieces by picking two points independently and uniformly along the stick, and breaking the stick at those two points. What is the probability that the three pieces can be assembled into a triangle?

Hint: a triangle can be formed from 3 line segments of lengths  $a, b, c$  if and only if  $a, b, c \in (0, 1/2)$ . The probability can be interpreted geometrically as proportional to an area in the plane, avoiding all calculus, but make sure for that approach that the distribution of the random point in the plane is Uniform over some region.

We can assume the length is 1 (in some choice of units, the length will be 1, and the choice of units for length does not affect whether a triangle can be formed). So let  $X, Y$  be i.i.d.  $\text{Unif}(0,1)$  random variables. Let  $x$  and  $y$  be the observed values of  $X$  and  $Y$  respectively. If  $x < y$ , then the side lengths are  $x, y - x$ , and  $1 - y$ , and a triangle can be formed if and only if  $y > \frac{1}{2}, y < x + \frac{1}{2}, x < \frac{1}{2}$ . Similarly, if  $x > y$ , then a triangle can be formed if and only if  $x > \frac{1}{2}, x < y + \frac{1}{2}, y < \frac{1}{2}$ .

Since  $(X, Y)$  is Uniform over the square  $0 \leq x \leq 1, 0 \leq y \leq 1$ , the probability of a subregion is proportional to its area. The region given by  $y > 1/2, y < x + 1/2, x < 1/2$  is a triangle with area  $1/8$ , as is the region given by  $x > 1/2, x < y + 1/2, y < 1/2$ , as illustrated in the picture below. Thus, the probability that a triangle can be formed is  $1/8 + 1/8 = 1/4$ .



Note that the idea of interpreting probabilities as areas works here because  $(X, Y)$

is *Uniform* on the square. For other distributions, in general we would need to find the joint PDF of  $X, Y$  and integrate over the appropriate region.

(b) Three legs are positioned uniformly and independently on the perimeter of a round table. What is the probability that the table will stand?

Think of the legs as points on a circle, chosen randomly one at a time, and choose units so that the circumference of the circle is 1. Let  $A, B, C$  be the arc lengths from one point to the next (clockwise, starting with the first point chosen). Then

$$\begin{aligned} P(\text{table falls}) &= P(\text{the 3 legs are all contained in some semicircle}) \\ &= P(\text{at least one of } A, B, C \text{ is greater than } 1/2) = 3/4, \end{aligned}$$

by Part (a). So the probability that the table will stand is  $1/4$ .

Alternatively, let  $C_j$  be the clockwise semicircle starting from the  $j$ th of the 3 points. Let  $A_j$  be the event that  $C_j$  contains all 3 points. Then  $P(A_j) = 1/4$  and with probability 1, at most one  $A_j$  occurs. So  $P(A_1 \cup A_2 \cup A_3) = 3/4$ , which again shows that the probability that the table will stand is  $1/4$ .

2. Let  $(X, Y)$  be a uniformly random point in the triangle in the plane with vertices  $(0, 0), (0, 1), (1, 0)$ . Find the joint PDF of  $X$  and  $Y$ , the marginal PDF of  $X$ , and the conditional PDF of  $X$  given  $Y$ .

The area of the triangle is  $\frac{1}{2}$ , so the joint PDF of  $(X, Y)$  is 2 inside the triangle and 0 outside the triangle. The triangle is given by  $x \geq 0, y \geq 0, x + y \leq 1$ , so the marginal PDF of  $X$  is  $\int_0^{1-x} 2dy = 2(1-x)$ , for  $x \in [0, 1]$  (note that this is nonnegative and integrates to 1). The conditional PDF of  $X$  given  $Y$  is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{2}{2(1-y)} = \frac{1}{1-y},$$

for  $(x, y)$  in the triangle (and 0 otherwise). Since  $\frac{1}{1-y}$  is constant with respect to  $x$ , we have  $X|Y \sim \text{Unif}(0, 1 - Y)$ .

3. Let  $X, Y$  be i.i.d.  $\text{Expo}(\lambda)$ . Find  $E|X - Y|$  in two ways: (a) using a 2-D LOTUS and (b) using the memoryless property without any calculus.

(a) First consider the case  $\lambda = 1$ . By LOTUS,

$$\begin{aligned}
 E|X - Y| &= \int_0^\infty \int_0^\infty |x - y| e^{-x} e^{-y} dx dy \\
 &= 2 \int_0^\infty \int_y^\infty (x - y) e^{-x} e^{-y} dx dy \\
 &= 2 \int_0^\infty e^{-y} \int_y^\infty (x e^{-x} - y e^{-x}) dx dy \\
 &= 2 \int_0^\infty e^{-y} (-e^{-x}(x + 1) + y e^{-x}) \Big|_y^\infty dy \\
 &= 2 \int_0^\infty e^{-y} e^{-y} dy = 2 \int_0^\infty e^{-2y} dy = 1.
 \end{aligned}$$

For general  $\lambda$ , this and the fact that  $\lambda X, \lambda Y$  are i.i.d.  $\text{Expo}(1)$  yield  $E|X - Y| = 1/\lambda$ .

(b) Write  $|X - Y| = \max(X, Y) - \min(X, Y)$ . By the memoryless property, this is  $\text{Expo}(\lambda)$ , similarly to the “three students working on a pset” problem from HW 6. So  $E|X - Y| = \frac{1}{\lambda}$ , which agrees with (a). (This also shows  $\text{Var}(|X - Y|) = \frac{1}{\lambda^2}$ .)

4. Two students,  $A$  and  $B$ , are working independently on homework (not necessarily for the same class). Student  $A$  takes  $Y_1 \sim \text{Expo}(\lambda_1)$  hours to finish his or her homework, while  $B$  takes  $Y_2 \sim \text{Expo}(\lambda_2)$  hours.

(a) Find the CDF and PDF of  $\frac{Y_1}{Y_2}$ , the ratio of their problem-solving times.

Let  $t > 0$ . The CDF of the ratio is

$$\begin{aligned}
 F(t) &= P\left(\frac{Y_1}{Y_2} \leq t\right) = P(Y_1 \leq tY_2) \\
 &= \int_0^\infty \left(\int_0^{ty_2} \lambda_1 e^{-\lambda_1 y_1} dy_1\right) \lambda_2 e^{-\lambda_2 y_2} dy_2 \\
 &= \int_0^\infty (1 - e^{-\lambda_1 t y_2}) \lambda_2 e^{-\lambda_2 y_2} dy_2 \\
 &= 1 - \int_0^\infty \lambda_2 e^{-(\lambda_1 t + \lambda_2) y_2} dy_2 \\
 &= 1 - \frac{\lambda_2}{t\lambda_1 + \lambda_2} \\
 &= \frac{t\lambda_1}{t\lambda_1 + \lambda_2}.
 \end{aligned}$$

Of course,  $F(t) = 0$  for  $t \leq 0$ . The PDF of the ratio is

$$f(t) = \frac{d}{dt} \left( \frac{t\lambda_1}{t\lambda_1 + \lambda_2} \right) = \frac{\lambda_1\lambda_2}{(\lambda_1 t + \lambda_2)^2}, \text{ for } t > 0.$$

(b) Find the probability that  $A$  finishes his or her homework before  $B$  does.

Plugging in  $t = 1$ , we have  $P(Y_1 < Y_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$ . This result makes intuitive sense in terms of waiting times. For example, let  $Y_1$  be the waiting time for the first green car to pass by a certain point and  $Y_2$  be the waiting time for the first red car to pass by a certain point, and say  $\lambda_1 = 3$  is the rate at which green cars pass by (measured in cars per hour) and  $\lambda_2 = 7$  is the rate at which red cars pass by: then on average 10 green or red cars pass by per hour, of which 3 are green (on average), so it makes sense that the probability that the next car to pass by is green is  $3/10$ .

As another check, note that for  $\lambda_1 = \lambda_2$ ,  $P(Y_1 < Y_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$  simplifies to  $1/2$ , which must be true by symmetry. Also by symmetry, we have  $P(Y_2 < Y_1) = \frac{\lambda_2}{\lambda_2 + \lambda_1}$ , which makes sense since  $1 - \frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{\lambda_2}{\lambda_1 + \lambda_2}$ .

5. The bus company from Blissville decides to start service in Blotchville, sensing a promising business opportunity. Meanwhile, Fred has moved back to Blotchville, inspired by a close reading of *I Had Trouble in Getting to Solla Sollew*. Now when Fred arrives at the bus stop, either of two independent bus lines may come by (both of which take him home). The Blissville company's bus arrival times are exactly 10 minutes apart, whereas the time from one Blotchville company bus to the next is  $\text{Expo}(\frac{1}{10})$ . Fred arrives at a uniformly random time on a certain day.

(a) What is the probability that the Blotchville company bus arrives first?

Hint: one good way is to use the continuous Law of Total Probability.

Let  $U \sim \text{Unif}(0, 10)$  be the arrival time of the next Blissville company bus, and  $X \sim \text{Expo}(\frac{1}{10})$  be the arrival time of the next Blotchville company bus (the latter is  $X \sim \text{Expo}(\frac{1}{10})$  by the memoryless property). Then

$$\begin{aligned} P(X < U) &= \int_0^{10} P(X < U | U = u) \frac{1}{10} du \\ &= \frac{1}{10} \int_0^{10} P(X < u | U = u) du \\ &= \frac{1}{10} \int_0^{10} (1 - e^{-u/10}) du = \frac{1}{e}. \end{aligned}$$

(b) What is the CDF of Fred's waiting time for a bus?

Let  $T = \min(X, U)$  be the waiting time. Then

$$P(T > t) = P(X > t, U > t) = P(X > t)P(U > t).$$

So the CDF of  $T$  is

$$P(T \leq t) = 1 - P(X > t)P(U > t) = 1 - e^{-t/10}(1 - t/10),$$

for  $0 < t < 10$  (and 0 for  $t \leq 0$ , and 1 for  $t \geq 10$ ).

6. Emails arrive in an inbox according to a Poisson process with rate  $\lambda$  (so the number of emails in a time interval of length  $t$  is distributed as  $\text{Pois}(\lambda t)$ , and the numbers of emails arriving in disjoint time intervals are independent). Let  $X, Y, Z$  be the numbers of emails that arrive from 9 am to noon, noon to 6 pm, and 6 pm to midnight (respectively) on a certain day.

(a) Find the joint PMF of  $X, Y, Z$ .

Since  $X \sim \text{Pois}(3\lambda), Y \sim \text{Pois}(6\lambda), Z \sim \text{Pois}(6\lambda)$  independently, the joint PMF is

$$P(X = x, Y = y, Z = z) = \frac{e^{-3\lambda}(3\lambda)^x}{x!} \frac{e^{-6\lambda}(6\lambda)^y}{y!} \frac{e^{-6\lambda}(6\lambda)^z}{z!},$$

for any nonnegative integers  $x, y, z$ .

(b) Find the conditional joint PMF of  $X, Y, Z$  given that  $X + Y + Z = 36$ .

Let  $T = X + Y + Z \sim \text{Pois}(15\lambda)$ , and suppose that we observe  $T = t$ . The conditional PMF is 0 for  $x + y + z \neq t$ . For  $x + y + z = t$ ,

$$\begin{aligned} P(X = x, Y = y, Z = z | T = t) &= \frac{P(T = t | X = x, Y = y, Z = z)P(X = x, Y = y, Z = z)}{P(T = t)} \\ &= \frac{\frac{e^{-3\lambda}(3\lambda)^x}{x!} \frac{e^{-6\lambda}(6\lambda)^y}{y!} \frac{e^{-6\lambda}(6\lambda)^z}{z!}}{\frac{e^{-15\lambda}(15\lambda)^t}{t!}} \\ &= \frac{t!}{x!y!z!} \left(\frac{3}{15}\right)^x \left(\frac{6}{15}\right)^y \left(\frac{6}{15}\right)^z. \end{aligned}$$

Thus,  $(X, Y, Z)$  is conditionally Multinomial given  $T = t$ , and we have that  $(X, Y, Z)$  is conditionally Multinomial $(36, (\frac{1}{5}, \frac{2}{5}, \frac{2}{5}))$  given  $T = 36$ .

(c) Find the conditional PMF of  $X + Y$  given that  $X + Y + Z = 36$ , and find  $E(X + Y | X + Y + Z = 36)$  and  $\text{Var}(X + Y | X + Y + Z = 36)$  (conditional expectation



and conditional variance given an event are defined in the same way as expectation and variance, using the conditional distribution given the event in place of the unconditional distribution).

Let  $W = X + Y$  and  $T = X + Y + Z$ . Using the story of the Multinomial and Part (b), we can merge the categories “9 am to noon” and “noon to 6 pm” to get

$$W|T = 36 \sim \text{Bin}\left(36, \frac{9}{15}\right).$$

Therefore,  $E(W|T = 36) = 36 \cdot \frac{9}{15} = 21.6$  and  $\text{Var}(W|T = 36) = 36 \cdot \frac{9}{15} \cdot \frac{6}{15} = 8.64$ .

7. Shakespeare wrote a total of 884647 words in his known works. Of course, many words are used more than once, and the number of distinct words in Shakespeare’s known writings is 31534 (according to one computation). This puts a lower bound on the size of Shakespeare’s vocabulary, but it is likely that Shakespeare knew words which he did not use in these known writings.

More specifically, suppose that a new poem of Shakespeare were uncovered, and consider the following (seemingly impossible) problem: give a good prediction of the number of words in the new poem that do not appear anywhere in Shakespeare’s previously known works.

The statisticians Ronald Thisted and Bradley Efron studied this problem in a paper called “Did Shakespeare write a newly-discovered poem?”, which performed statistical tests to try to determine whether Shakespeare was the author of a poem discovered by a Shakespearean scholar in 1985. A simplified version of their method is developed in the problem below. The method was originally invented by Alan Turing (the founder of computer science) and I.J. Good as part of the effort to break the German Enigma code during World War II.

Let  $N$  be the number of distinct words that Shakespeare knew, and assume these words are numbered from 1 to  $N$ . Suppose for simplicity that Shakespeare wrote only two plays,  $A$  and  $B$ . The plays are reasonably long and they are of the same length. Let  $X_j$  be the number of times that word  $j$  appears in play  $A$ , and  $Y_j$  be the number of times it appears in play  $B$ , for  $1 \leq j \leq N$ .

(a) Explain why it is reasonable to model  $X_j$  as being Poisson, and  $Y_j$  as being Poisson with the same parameter as  $X_j$ .

It is reasonable to model  $X_j$  and  $Y_j$  as Poisson, because this distribution is used to describe the number of “events” (such as emails received) happening at some average rate in a fixed interval or volume. The Poisson paradigm applies here: each

individual word in a play has some very small probability of being word  $j$ , and the words are weakly dependent. Here an event means using word  $j$ , the average rate is determined by how frequently Shakespeare uses that word overall. It is reasonable to assume that the average rate of occurrence of a particular word is the same for two plays by the same author, so we take  $\lambda$  to be the same for  $X_j$  and  $Y_j$ .

(b) Let the numbers of occurrences of the word “eyeball” (which was coined by Shakespeare) in the two plays be independent  $\text{Pois}(\lambda)$  r.v.s. Show that the probability that “eyeball” is used in play  $B$  but not in play  $A$  is

$$e^{-\lambda}(\lambda - \lambda^2/2! + \lambda^3/3! - \lambda^4/4! + \dots).$$

Let  $X$  be the number of times that “eyeball” is used in play  $A$ , and  $Y$  be the number of times that it is used in play  $B$ . Since  $X$  and  $Y$  are independent  $\text{Pois}(\lambda)$ ,

$$\begin{aligned} P(X = 0, Y > 0) &= P(X = 0)(1 - P(Y = 0)) = e^{-\lambda}(1 - e^{-\lambda}) \\ &= e^{-\lambda} \left( 1 - \left( 1 - \lambda + \frac{\lambda^2}{2!} - \frac{\lambda^3}{3!} + \frac{\lambda^4}{4!} - \dots \right) \right) \\ &= e^{-\lambda} \left( \lambda - \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} - \frac{\lambda^4}{4!} + \dots \right). \end{aligned}$$

(c) Now assume that  $\lambda$  from (b) is unknown and is itself taken to be a random variable to reflect this uncertainty. So let  $\lambda$  have a PDF  $f_0$ . Let  $X$  be the number of times the word “eyeball” appears in play  $A$  and  $Y$  be the corresponding value for play  $B$ . Assume that the conditional distribution of  $X, Y$  given  $\lambda$  is that they are independent  $\text{Pois}(\lambda)$  r.v.s. Show that the probability that “eyeball” is used in play  $B$  but not in play  $A$  is the alternating series

$$P(X = 1) - P(X = 2) + P(X = 3) - P(X = 4) + \dots$$

Hint: condition on  $\lambda$  and use (b).

Now  $\lambda$  is a random variable. Given  $\lambda$ , the calculation from (b) holds. By the law of total probability,

$$\begin{aligned} P(X = 0, Y > 0) &= \int_0^\infty P(X = 0, Y > 0 \mid \lambda) f_0(\lambda) d\lambda \\ &= \int_0^\infty P(X = 1 \mid \lambda) f_0(\lambda) d\lambda - \int_0^\infty P(X = 2 \mid \lambda) f_0(\lambda) d\lambda \\ &\quad + \int_0^\infty P(X = 3 \mid \lambda) f_0(\lambda) d\lambda - \int_0^\infty P(X = 4 \mid \lambda) f_0(\lambda) d\lambda + \dots \\ &= P(X = 1) - P(X = 2) + P(X = 3) - P(X = 4) + \dots \end{aligned}$$

(d) Assume that every word's numbers of occurrences in  $A$  and  $B$  are distributed as in (c), where  $\lambda$  may be different for different words but  $f_0$  is fixed. Let  $W_j$  be the number of words that appear exactly  $j$  times in play  $A$ . Show that the expected number of distinct words appearing in play  $B$  but not in play  $A$  is

$$E(W_1) - E(W_2) + E(W_3) - E(W_4) + \dots$$

(This shows that  $W_1 - W_2 + \dots$  is an *unbiased* prediction of the number of distinct words appearing in play  $B$  but not in play  $A$ : on average it is correct. Moreover, it can be computed just from having seen play  $A$ , without needing to know  $f_0$  or any of the  $\lambda_j$ . This method can be extended in various ways to give predictions for unobserved plays based on observed plays.)

Let  $X_j$  be the number of times word  $j$  appears in play  $A$  and let  $W$  be the number of distinct words that appear in play  $B$  but not in  $A$ . Then  $W = \sum_{j=1}^N I_j$ , where  $I_j$  is the indicator r.v. of the event that word  $j$  appears in play  $B$  but not in play  $A$ , and  $N$  is the total number of words. By (c), for  $1 \leq j \leq N$ ,

$$EI_j = \sum_{i=1}^{\infty} (-1)^{i+1} P(X_j = i).$$

Also, note that the number of words that appear exactly  $i$  times in play  $A$  is

$$W_i = I(X_1 = i) + I(X_2 = i) + I(X_3 = i) + \dots + I(X_N = i),$$

where  $I(X_j = i)$  is the indicator of word  $j$  appearing exactly  $i$  times in play  $A$ . So

$$EW_i = \sum_{j=1}^N EI(X_j = i) = \sum_{j=1}^N P(X_j = i).$$

Then

$$\begin{aligned} EW &= \sum_{j=1}^N EI_j = \sum_{j=1}^N \sum_{i=1}^{\infty} (-1)^{i+1} P(X_j = i) \\ &= \sum_{i=1}^{\infty} (-1)^{i+1} \sum_{j=1}^N P(X_j = i) \\ &= \sum_{i=1}^{\infty} (-1)^{i+1} EW_i \\ &= EW_1 - EW_2 + EW_3 - EW_4 + \dots \end{aligned}$$