Solutions to Exercises Marked with Ⓢ
from the book
Introduction to Probability by
Joseph K. Blitzstein and Jessica Hwang
© Chapman & Hall/CRC Press, 2014

Joseph K. Blitzstein and Jessica Hwang
Departments of Statistics, Harvard University and Stanford University

Chapter 1: Probability and counting

Counting

8. Ⓢ (a) How many ways are there to split a dozen people into 3 teams, where one team has 2 people, and the other two teams have 5 people each?

   (b) How many ways are there to split a dozen people into 3 teams, where each team has 4 people?

   Solution:

   (a) Pick any 2 of the 12 people to make the 2 person team, and then any 5 of the remaining 10 for the first team of 5, and then the remaining 5 are on the other team of 5; this overcounts by a factor of 2 though, since there is no designated "first" team of 5. So the number of possibilities is $\binom{12}{2}\binom{10}{5}/2 = 8316$. Alternatively, politely ask the 12 people to line up, and then let the first 2 be the team of 2, the next 5 be a team of 5, and then last 5 be a team of 5. There are 12! ways for them to line up, but it does not matter which order they line up in *within* each group, nor does the order of the 2 teams of 5 matter, so the number of possibilities is $\frac{12!}{2!5!5!\cdot 2} = 8316$.

   (b) By either of the approaches above, there are $\frac{12!}{4!4!4!}$ ways to divide the people into a Team A, a Team B, and a Team C, if we care about which team is which (this is called a *multinomial coefficient*). Since here it doesn't matter which team is which, this over counts by a factor of 3!, so the number of possibilities is $\frac{12!}{4!4!4!3!} = 5775$.

9. Ⓢ (a) How many paths are there from the point $(0, 0)$ to the point $(110, 111)$ in the plane such that each step either consists of going one unit up or one unit to the right?

   (b) How many paths are there from $(0, 0)$ to $(210, 211)$, where each step consists of going one unit up or one unit to the right, and the path has to go through $(110, 111)$?

   Solution:

   (a) Encode a path as a sequence of U's and R's, like $URURURUUUR \dots UR$, where U and R stand for "up" and "right" respectively. The sequence must consist of 110 R's and 111 U's, and to determine the sequence we just need to specify where the R's are located. So there are $\binom{221}{110}$ possible paths.

   (b) There are $\binom{221}{110}$ paths to $(110, 111)$, as above. From there, we need 100 R's and 100 U's to get to $(210, 211)$, so by the multiplication rule the number of possible paths is $\binom{221}{110} \cdot \binom{200}{100}$.

Story proofs

15. Ⓢ Give a story proof that $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.

   Solution: Consider picking a subset of n people. There are $\binom{n}{k}$ choices with size k, on the one hand, and on the other hand there are 2^n subsets by the multiplication rule.

16. Ⓢ Show that for all positive integers n and k with $n \geq k$,

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

doing this in two ways: (a) algebraically and (b) with a story, giving an interpretation for why both sides count the same thing.

Hint for the story proof: Imagine $n+1$ people, with one of them pre-designated as "president".

Solution:

(a) For the algebraic proof, start with the definition of the binomial coefficients in the left-hand side, and do some algebraic manipulation as follows:

$$\begin{aligned}\binom{n}{k} + \binom{n}{k-1} &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k-1)!(n-k+1)!} \\ &= \frac{(n-k+1)n! + (k)n!}{k!(n-k+1)!} \\ &= \frac{n!(n+1)}{k!(n-k+1)!} \\ &= \binom{n+1}{k}.\end{aligned}$$

(b) For the story proof, consider $n+1$ people, with one of them pre-designated as "president". The right-hand side is the number of ways to choose k out of these $n+1$ people, with order not mattering. The left-hand side counts the same thing in a different way, by considering two cases: the president is or isn't in the chosen group.

The number of groups of size k which include the president is $\binom{n}{k-1}$, since once we fix the president as a member of the group, we only need to choose another $k-1$ members out of the remaining n people. Similarly, there are $\binom{n}{k}$ groups of size k that don't include the president. Thus, the two sides of the equation are equal.

18. Ⓢ (a) Show using a story proof that

$$\binom{k}{k} + \binom{k+1}{k} + \binom{k+2}{k} + \cdots + \binom{n}{k} = \binom{n+1}{k+1},$$

where n and k are positive integers with $n \geq k$. This is called the *hockey stick identity*.

Hint: Imagine arranging a group of people by age, and then think about the oldest person in a chosen subgroup.

(b) Suppose that a large pack of Haribo gummi bears can have anywhere between 30 and 50 gummi bears. There are 5 delicious flavors: pineapple (clear), raspberry (red), orange (orange), strawberry (green, mysteriously), and lemon (yellow). There are 0 non-delicious flavors. How many possibilities are there for the composition of such a pack of gummi bears? You can leave your answer in terms of a couple binomial coefficients, but not a sum of lots of binomial coefficients.

Solution:

(a) Consider choosing $k+1$ people out of a group of $n+1$ people. Call the oldest person in the subgroup "Aemon." If Aemon is also the oldest person in the full group, then there are $\binom{n}{k}$ choices for the rest of the subgroup. If Aemon is the second oldest in the full group, then there are $\binom{n-1}{k}$ choices since the oldest person in the full group can't be

chosen. In general, if there are j people in the full group who are younger than Aemon, then there are $\binom{j}{k}$ possible choices for the rest of the subgroup. Thus,

$$\sum_{j=k}^{n} \binom{j}{k} = \binom{n+1}{k+1}.$$

(b) For a pack of i gummi bears, there are $\binom{5+i-1}{i} = \binom{i+4}{i} = \binom{i+4}{4}$ possibilities since the situation is equivalent to getting a sample of size i from the $n = 5$ flavors (with replacement, and with order not mattering). So the total number of possibilities is

$$\sum_{i=30}^{50} \binom{i+4}{4} = \sum_{j=34}^{54} \binom{j}{4}.$$

Applying the previous part, we can simplify this by writing

$$\sum_{j=34}^{54} \binom{j}{4} = \sum_{j=4}^{54} \binom{j}{4} - \sum_{j=4}^{33} \binom{j}{4} = \binom{55}{5} - \binom{34}{5}.$$

(This works out to 3200505 possibilities!)

Naive definition of probability

22. Ⓢ A certain family has 6 children, consisting of 3 boys and 3 girls. Assuming that all birth orders are equally likely, what is the probability that the 3 eldest children are the 3 girls?

Solution: Label the girls as $1, 2, 3$ and the boys as $4, 5, 6$. Think of the birth order is a permutation of $1, 2, 3, 4, 5, 6$, e.g., we can interpret 314265 as meaning that child 3 was born first, then child 1, etc. The number of possible permutations of the birth orders is 6!. Now we need to count how many of these have all of $1, 2, 3$ appear before all of $4, 5, 6$. This means that the sequence must be a permutation of $1, 2, 3$ followed by a permutation of $4, 5, 6$. So with all birth orders equally likely, we have

$$P(\text{the 3 girls are the 3 eldest children}) = \frac{(3!)^2}{6!} = 0.05.$$

Alternatively, we can use the fact that there are $\binom{6}{3}$ ways to choose where the girls appear in the birth order (without taking into account the ordering of the girls amongst themselves). These are all equally likely. Of these possibilities, there is only 1 where the 3 girls are the 3 eldest children. So again the probability is $\frac{1}{\binom{6}{3}} = 0.05$.

23. Ⓢ A city with 6 districts has 6 robberies in a particular week. Assume the robberies are located randomly, with all possibilities for which robbery occurred where equally likely. What is the probability that some district had more than 1 robbery?

Solution: There are 6^6 possible configurations for which robbery occurred where. There are 6! configurations where each district had exactly 1 of the 6, so the probability of the complement of the desired event is $6!/6^6$. So the probability of some district having more than 1 robbery is

$$1 - 6!/6^6 \approx 0.9846.$$

Note that this also says that if a fair die is rolled 6 times, there's over a 98% chance that some value is repeated!

26. Ⓢ A college has 10 (non-overlapping) time slots for its courses, and blithely assigns courses to time slots randomly and independently. A student randomly chooses 3 of the courses to enroll in. What is the probability that there is a conflict in the student's schedule?

Solution: The probability of no conflict is $\frac{10\cdot 9\cdot 8}{10^3} = 0.72$. So the probability of there being at least one scheduling conflict is 0.28.

27. Ⓢ For each part, decide whether the blank should be filled in with $=, <$, or $>$, and give a clear explanation.

(a) (probability that the total after rolling 4 fair dice is 21) ____ (probability that the total after rolling 4 fair dice is 22)

(b) (probability that a random 2-letter word is a palindrome[1]) ____ (probability that a random 3-letter word is a palindrome)

Solution:

(a) $\boxed{>}$. All *ordered* outcomes are equally likely here. So for example with two dice, obtaining a total of 9 is more likely than obtaining a total of 10 since there are two ways to get a 5 and a 4, and only one way to get two 5's. To get a 21, the outcome must be a permutation of $(6,6,6,3)$ (4 possibilities), $(6,5,5,5)$ (4 possibilities), or $(6,6,5,4)$ ($4!/2 = 12$ possibilities). To get a 22, the outcome must be a permutation of $(6,6,6,4)$ (4 possibilities), or $(6,6,5,5)$ ($4!/2^2 = 6$ possibilities). So getting a 21 is more likely; in fact, it is exactly twice as likely as getting a 22.

(b) $\boxed{=}$. The probabilities are equal, since for both 2-letter and 3-letter words, being a palindrome means that the first and last letter are the same.

29. Ⓢ Elk dwell in a certain forest. There are N elk, of which a simple random sample of size n are captured and tagged ("simple random sample" means that all $\binom{N}{n}$ sets of n elk are equally likely). The captured elk are returned to the population, and then a new sample is drawn, this time with size m. This is an important method that is widely used in ecology, known as *capture-recapture*. What is the probability that exactly k of the m elk in the new sample were previously tagged? (Assume that an elk that was captured before doesn't become more or less likely to be captured again.)

Solution: We can use the naive definition here since we're assuming all samples of size m are equally likely. To have exactly k be tagged elk, we need to choose k of the n tagged elk, and then $m-k$ from the $N-n$ untagged elk. So the probability is

$$\frac{\binom{n}{k}\cdot\binom{N-n}{m-k}}{\binom{N}{m}},$$

for k such that $0 \le k \le n$ and $0 \le m-k \le N-n$, and the probability is 0 for all other values of k (for example, if $k > n$ the probability is 0 since then there aren't even k tagged elk in the entire population!). This is known as a *Hypergeometric* probability; we will encounter it again in Chapter 3.

31. Ⓢ A jar contains r red balls and g green balls, where r and g are fixed positive integers. A ball is drawn from the jar randomly (with all possibilities equally likely), and then a second ball is drawn randomly.

[1] A *palindrome* is an expression such as "A man, a plan, a canal: Panama" that reads the same backwards as forwards (ignoring spaces, capitalization, and punctuation). Assume for this problem that all words of the specified length are equally likely, that there are no spaces or punctuation, and that the alphabet consists of the lowercase letters a,b,...,z.

(a) Explain intuitively why the probability of the second ball being green is the same as the probability of the first ball being green.

(b) Define notation for the sample space of the problem, and use this to compute the probabilities from (a) and show that they are the same.

(c) Suppose that there are 16 balls in total, and that the probability that the two balls are the same color is the same as the probability that they are different colors. What are r and g (list all possibilities)?

Solution:

(a) This is true by *symmetry*. The first ball is equally likely to be any of the $g+r$ balls, so the probability of it being green is $g/(g+r)$. But the second ball is also equally likely to be any of the $g+r$ balls (there aren't certain balls that enjoy being chosen second and others that have an aversion to being chosen second); once we know whether the first ball is green we have information that affects our uncertainty about the second ball, but before we have this information, the second ball is equally likely to be any of the balls.

Alternatively, intuitively it shouldn't matter if we pick one ball at a time, or take one ball with the left hand and one with the right hand at the same time. By symmetry, the probabilities for the ball drawn with the left hand should be the same as those for the ball drawn with the right hand.

(b) Label the balls as $1, 2, \ldots, g+r$, such that $1, 2, \ldots, g$ are green and $g+1, \ldots, g+r$ are red. The sample space can be taken to be the set of all pairs (a, b) with $a, b \in \{1, \ldots, g+r\}$ and $a \neq b$ (there are other possible ways to define the sample space, but it is important to specify all possible outcomes using clear notation, and it make sense to be as richly detailed as possible in the specification of possible outcomes, to avoid losing information). Each of these pairs is equally likely, so we can use the naive definition of probability. Let G_i be the event that the ith ball drawn is green.

The denominator is $(g+r)(g+r-1)$ by the multiplication rule. For G_1, the numerator is $g(g+r-1)$, again by the multiplication rule. For G_2, the numerator is also $g(g+r-1)$, since in counting favorable cases, there are g possibilities for the second ball, and for each of those there are $g+r-1$ favorable possibilities for the first ball (note that the multiplication rule doesn't require the experiments to be listed in chronological order!); alternatively, there are $g(g-1)+rg = g(g+r-1)$ favorable possibilities for the second ball being green, as seen by considering 2 cases (first ball green and first ball red). Thus,

$$P(G_i) = \frac{g(g+r-1)}{(g+r)(g+r-1)} = \frac{g}{g+r},$$

for $i \in \{1, 2\}$, which concurs with (a).

(c) Let A be the event of getting one ball of each color. In set notation, we can write $A = (G_1 \cap G_2^c) \cup (G_1^c \cap G_2)$. We are given that $P(A) = P(A^c)$, so $P(A) = 1/2$. Then

$$P(A) = \frac{2gr}{(g+r)(g+r-1)} = \frac{1}{2},$$

giving the quadratic equation

$$g^2 + r^2 - 2gr - g - r = 0,$$

i.e.,

$$(g-r)^2 = g + r.$$

But $g+r = 16$, so $g-r$ is 4 or -4. Thus, either $g = 10, r = 6$, or $g = 6, r = 10$.

32. Ⓢ A random 5-card poker hand is dealt from a standard deck of cards. Find the probability of each of the following possibilities (in terms of binomial coefficients).

(a) A flush (all 5 cards being of the same suit; do not count a royal flush, which is a flush with an ace, king, queen, jack, and 10).

(b) Two pair (e.g., two 3's, two 7's, and an ace).

Solution:

(a) A flush can occur in any of the 4 suits (imagine the tree, and for concreteness suppose the suit is Hearts); there are $\binom{13}{5}$ ways to choose the cards in that suit, except for one way to have a royal flush in that suit. So the probability is

$$\frac{4\left(\binom{13}{5}-1\right)}{\binom{52}{5}}.$$

(b) Choose the two ranks of the pairs, which specific cards to have for those 4 cards, and then choose the extraneous card (which can be any of the $52-8$ cards not of the two chosen ranks). This gives that the probability of getting two pairs is

$$\frac{\binom{13}{2}\cdot\binom{4}{2}^2\cdot 44}{\binom{52}{5}}.$$

40. Ⓢ A *norepeatword* is a sequence of at least one (and possibly all) of the usual 26 letters a,b,c,...,z, with repetitions not allowed. For example, "course" is a norepeatword, but "statistics" is not. Order matters, e.g., "course" is not the same as "source".
A norepeatword is chosen randomly, with all norepeatwords equally likely. Show that the probability that it uses all 26 letters is very close to $1/e$.

Solution: The number of norepeatwords having all 26 letters is the number of ordered arrangements of 26 letters: 26!. To construct a norepeatword with $k \leq 26$ letters, we first select k letters from the alphabet ($\binom{26}{k}$ selections) and then arrange them into a word ($k!$ arrangements). Hence there are $\binom{26}{k}k!$ norepeatwords with k letters, with k ranging from 1 to 26. With all norepeatwords equally likely, we have

$$\begin{aligned} P(\text{norepeatword having all 26 letters}) &= \frac{\#\text{ norepeatwords having all 26 letters}}{\#\text{ norepeatwords}} \\ &= \frac{26!}{\sum_{k=1}^{26}\binom{26}{k}k!} = \frac{26!}{\sum_{k=1}^{26}\frac{26!}{k!(26-k)!}k!} \\ &= \frac{1}{\frac{1}{25!}+\frac{1}{24!}+\ldots+\frac{1}{1!}+1}. \end{aligned}$$

The denominator is the first 26 terms in the Taylor series $e^x = 1 + x + x^2/2! + \ldots$, evaluated at $x = 1$. Thus the probability is approximately $1/e$ (this is an *extremely* good approximation since the series for e converges very quickly; the approximation for e differs from the truth by less than 10^{-26}).

Axioms of probability

46. Ⓢ Arby has a belief system assigning a number $P_{\text{Arby}}(A)$ between 0 and 1 to every event A (for some sample space). This represents Arby's degree of belief about how likely A is to occur. For any event A, Arby is willing to pay a price of $1000 \cdot P_{\text{Arby}}(A)$ dollars to buy a certificate such as the one shown below:

Certificate

The owner of this certificate can redeem it for \$1000 if A occurs. No value if A does not occur, except as required by federal, state, or local law. No expiration date.

Likewise, Arby is willing to sell such a certificate at the same price. Indeed, Arby is willing to buy or sell any number of certificates at this price, as Arby considers it the "fair" price.

Arby stubbornly refuses to accept the axioms of probability. In particular, suppose that there are two *disjoint* events A and B with

$$P_{\text{Arby}}(A \cup B) \neq P_{\text{Arby}}(A) + P_{\text{Arby}}(B).$$

Show how to make Arby go bankrupt, by giving a list of transactions Arby is willing to make that will *guarantee* that Arby will lose money (you can assume it will be known whether A occurred and whether B occurred the day after any certificates are bought/sold).

Solution: Suppose first that

$$P_{\text{Arby}}(A \cup B) < P_{\text{Arby}}(A) + P_{\text{Arby}}(B).$$

Call a certificate like the one show above, with any event C in place of A, a C-certificate. Measuring money in units of thousands of dollars, Arby is willing to pay $P_{\text{Arby}}(A) + P_{\text{Arby}}(B)$ to buy an A-certificate and a B-certificate, and is willing to sell an $(A \cup B)$-certificate for $P_{\text{Arby}}(A \cup B)$. In those transactions, Arby loses $P_{\text{Arby}}(A) + P_{\text{Arby}}(B) - P_{\text{Arby}}(A \cup B)$ and will not recoup any of that loss because if A or B occurs, Arby will have to pay out an amount equal to the amount Arby receives (since it's impossible for both A and B to occur).

Now suppose instead that

$$P_{\text{Arby}}(A \cup B) > P_{\text{Arby}}(A) + P_{\text{Arby}}(B).$$

Measuring money in units of thousands of dollars, Arby is willing to sell an A-certificate for $P_{\text{Arby}}(A)$, sell a B-certificate for $P_{\text{Arby}}(B)$, and buy a $(A \cup B)$-certificate for $P_{\text{Arby}}(A \cup B)$. In so doing, Arby loses $P_{\text{Arby}}(A \cup B) - (P_{\text{Arby}}(A) + P_{\text{Arby}}(B))$, and Arby won't recoup any of this loss, similarly to the above. (In fact, in this case, even if A and B are not disjoint, Arby will not recoup any of the loss, and will lose more money if both A and B occur.)

By buying/selling a sufficiently large number of certificates from/to Arby as described above, you can guarantee that you'll get all of Arby's money; this is called an *arbitrage opportunity*. This problem illustrates the fact that the axioms of probability are not arbitrary, but rather are *essential* for coherent thought (at least the first axiom, and the second with finite unions rather than countably infinite unions).

Arbitrary axioms allow arbitrage attacks; principled properties and perspectives on probability potentially prevent perdition.

Inclusion-exclusion

48. Ⓢ A card player is dealt a 13-card hand from a well-shuffled, standard deck of cards. What is the probability that the hand is void in at least one suit ("void in a suit" means having no cards of that suit)?

Solution: Let S, H, D, C be the events of being void in Spades, Hearts, Diamonds, Clubs, respectively. We want to find $P(S \cup D \cup H \cup C)$. By inclusion-exclusion and symmetry,

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

The probability of being void in a specific suit is $\frac{\binom{39}{13}}{\binom{52}{13}}$. The probability of being void in 2 specific suits is $\frac{\binom{26}{13}}{\binom{52}{13}}$. The probability of being void in 3 specific suits is $\frac{1}{\binom{52}{13}}$. And the last term is 0 since it's impossible to be void in everything. So the probability is

$$4\frac{\binom{39}{13}}{\binom{52}{13}} - 6\frac{\binom{26}{13}}{\binom{52}{13}} + \frac{4}{\binom{52}{13}} \approx 0.051.$$

52. Ⓢ Alice attends a small college in which each class meets only once a week. She is deciding between 30 non-overlapping classes. There are 6 classes to choose from for each day of the week, Monday through Friday. Trusting in the benevolence of randomness, Alice decides to register for 7 randomly selected classes out of the 30, with all choices equally likely. What is the probability that she will have classes every day, Monday through Friday? (This problem can be done either directly using the naive definition of probability, or using inclusion-exclusion.)

Solution: We will solve this both by direct counting and using inclusion-exclusion.

Direct Counting Method: There are two general ways that Alice can have class every day: either she has 2 days with 2 classes and 3 days with 1 class, or she has 1 day with 3 classes, and has 1 class on each of the other 4 days. The number of possibilities for the former is $\binom{5}{2}\binom{6}{2}^2 6^3$ (choose the 2 days when she has 2 classes, and then select 2 classes on those days and 1 class for the other days). The number of possibilities for the latter is $\binom{5}{1}\binom{6}{3}6^4$. So the probability is

$$\frac{\binom{5}{2}\binom{6}{2}^2 6^3 + \binom{5}{1}\binom{6}{3}6^4}{\binom{30}{7}} = \frac{114}{377} \approx 0.302.$$

Inclusion-Exclusion Method: We will use inclusion-exclusion to find the probability of the complement, which is the event that she has at least one day with no classes. Let $B_i = A_i^c$. Then

$$P(B_1 \cup B_2 \cdots \cup B_5) = \sum_i P(B_i) - \sum_{i<j} P(B_i \cap B_j) + \sum_{i<j<k} P(B_i \cap B_j \cap B_k)$$

(terms with the intersection of 4 or more B_i's are not needed since Alice must have classes on at least 2 days). We have

$$P(B_1) = \frac{\binom{24}{7}}{\binom{30}{7}}, P(B_1 \cap B_2) = \frac{\binom{18}{7}}{\binom{30}{7}}, P(B_1 \cap B_2 \cap B_3) = \frac{\binom{12}{7}}{\binom{30}{7}}$$

and similarly for the other intersections. So

$$P(B_1 \cup \cdots \cup B_5) = 5\frac{\binom{24}{7}}{\binom{30}{7}} - \binom{5}{2}\frac{\binom{18}{7}}{\binom{30}{7}} + \binom{5}{3}\frac{\binom{12}{7}}{\binom{30}{7}} = \frac{263}{377}.$$

Therefore,

$$P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) = \frac{114}{377} \approx 0.302.$$

Mixed practice

59. Ⓢ There are 100 passengers lined up to board an airplane with 100 seats (with each seat assigned to one of the passengers). The first passenger in line crazily decides to sit in a randomly chosen seat (with all seats equally likely). Each subsequent passenger takes his or her assigned seat if available, and otherwise sits in a random available seat. What is the probability that the last passenger in line gets to sit in his or her assigned seat? (This is a common interview problem, and a beautiful example of the power of symmetry.)

    Hint: Call the seat assigned to the jth passenger in line "seat j" (regardless of whether the airline calls it seat 23A or whatever). What are the possibilities for which seats are available to the last passenger in line, and what is the probability of each of these possibilities?

    Solution: The seat for the last passenger is either seat 1 or seat 100; for example, seat 42 can't be available to the last passenger since the 42nd passenger in line would have sat there if possible. Seat 1 and seat 100 are equally likely to be available to the last passenger, since the previous 99 passengers view these two seats symmetrically. So the probability that the last passenger gets seat 100 is $1/2$.

Chapter 2: Conditional probability

Conditioning on evidence

1. Ⓢ A spam filter is designed by looking at commonly occurring phrases in spam. Suppose that 80% of email is spam. In 10% of the spam emails, the phrase "free money" is used, whereas this phrase is only used in 1% of non-spam emails. A new email has just arrived, which does mention "free money". What is the probability that it is spam?

   Solution: Let S be the event that an email is spam and F be the event that an email has the "free money" phrase. By Bayes' rule,

   $$P(S|F) = \frac{P(F|S)P(S)}{P(F)} = \frac{0.1 \cdot 0.8}{0.1 \cdot 0.8 + 0.01 \cdot 0.2} = \frac{80/1000}{82/1000} = \frac{80}{82} \approx 0.9756.$$

2. Ⓢ A woman is pregnant with twin boys. Twins may be either identical or fraternal (non-identical). In general, 1/3 of twins born are identical. Obviously, identical twins must be of the same sex; fraternal twins may or may not be. Assume that identical twins are equally likely to be both boys or both girls, while for fraternal twins all possibilities are equally likely. Given the above information, what is the probability that the woman's twins are identical?

   Solution: By Bayes' rule,

   $$P(\text{identical}|BB) = \frac{P(BB|\text{identical})P(\text{identical})}{P(BB)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{2}{3}} = 1/2.$$

22. Ⓢ A bag contains one marble which is either green or blue, with equal probabilities. A green marble is put in the bag (so there are 2 marbles now), and then a random marble is taken out. The marble taken out is green. What is the probability that the remaining marble is also green?

   Solution: Let A be the event that the initial marble is green, B be the event that the removed marble is green, and C be the event that the remaining marble is green. We need to find $P(C|B)$. There are several ways to find this; one natural way is to condition on whether the initial marble is green:

   $$P(C|B) = P(C|B, A)P(A|B) + P(C|B, A^c)P(A^c|B) = 1P(A|B) + 0P(A^c|B).$$

   To find $P(A|B)$, use Bayes' Rule:

   $$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{1/2}{P(B|A)P(A) + P(B|A^c)P(A^c)} = \frac{1/2}{1/2 + 1/4} = \frac{2}{3}.$$

   So $P(C|B) = 2/3$.

   Historical note: This problem was first posed by Lewis Carroll in 1893.

23. Ⓢ Let G be the event that a certain individual is guilty of a certain robbery. In gathering evidence, it is learned that an event E_1 occurred, and a little later it is also learned that another event E_2 also occurred. Is it possible that individually, these pieces of evidence

increase the chance of guilt (so $P(G|E_1) > P(G)$ and $P(G|E_2) > P(G)$), but together they decrease the chance of guilt (so $P(G|E_1, E_2) < P(G)$)?

Solution: Yes, this is possible. In fact, it is possible to have two events which separately provide evidence in favor of G, yet which together preclude G! For example, suppose that the crime was committed between 1 pm and 3 pm on a certain day. Let E_1 be the event that the suspect was at a specific nearby coffeeshop from 1 pm to 2 pm that day, and let E_2 be the event that the suspect was at the nearby coffeeshop from 2 pm to 3 pm that day. Then $P(G|E_1) > P(G), P(G|E_2) > P(G)$ (assuming that being in the vicinity helps show that the suspect had the opportunity to commit the crime), yet $P(G|E_1 \cap E_2) < P(G)$ (as being in the coffeehouse from 1 pm to 3 pm gives the suspect an alibi for the full time).

25. Ⓢ A crime is committed by one of two suspects, A and B. Initially, there is equal evidence against both of them. In further investigation at the crime scene, it is found that the guilty party had a blood type found in 10% of the population. Suspect A does match this blood type, whereas the blood type of Suspect B is unknown.

(a) Given this new information, what is the probability that A is the guilty party?

(b) Given this new information, what is the probability that B's blood type matches that found at the crime scene?

Solution:

(a) Let M be the event that A's blood type matches the guilty party's and for brevity, write A for "A is guilty" and B for "B is guilty". By Bayes' Rule,

$$P(A|M) = \frac{P(M|A)P(A)}{P(M|A)P(A) + P(M|B)P(B)} = \frac{1/2}{1/2 + (1/10)(1/2)} = \frac{10}{11}.$$

(We have $P(M|B) = 1/10$ since, given that B is guilty, the probability that A's blood type matches the guilty party's is the same probability as for the general population.)

(b) Let C be the event that B's blood type matches, and condition on whether B is guilty. This gives

$$P(C|M) = P(C|M, A)P(A|M) + P(C|M, B)P(B|M) = \frac{1}{10} \cdot \frac{10}{11} + \frac{1}{11} = \frac{2}{11}.$$

26. Ⓢ To battle against spam, Bob installs two anti-spam programs. An email arrives, which is either legitimate (event L) or spam (event L^c), and which program j marks as legitimate (event M_j) or marks as spam (event M_j^c) for $j \in \{1, 2\}$. Assume that 10% of Bob's email is legitimate and that the two programs are each "90% accurate" in the sense that $P(M_j|L) = P(M_j^c|L^c) = 9/10$. Also assume that given whether an email is spam, the two programs' outputs are conditionally independent.

(a) Find the probability that the email is legitimate, given that the 1st program marks it as legitimate (simplify).

(b) Find the probability that the email is legitimate, given that both programs mark it as legitimate (simplify).

(c) Bob runs the 1st program and M_1 occurs. He updates his probabilities and then runs the 2nd program. Let $\tilde{P}(A) = P(A|M_1)$ be the updated probability function after running the 1st program. Explain briefly in words whether or not $\tilde{P}(L|M_2) = P(L|M_1 \cap M_2)$: is conditioning on $M_1 \cap M_2$ in one step equivalent to first conditioning on M_1, then updating probabilities, and then conditioning on M_2?

Solution:

(a) By Bayes' rule,

$$P(L|M_1) = \frac{P(M_1|L)P(L)}{P(M_1)} = \frac{\frac{9}{10}\cdot\frac{1}{10}}{\frac{9}{10}\cdot\frac{1}{10}+\frac{1}{10}\cdot\frac{9}{10}} = \frac{1}{2}.$$

(b) By Bayes' rule,

$$P(L|M_1,M_2) = \frac{P(M_1,M_2|L)P(L)}{P(M_1,M_2)} = \frac{(\frac{9}{10})^2\cdot\frac{1}{10}}{(\frac{9}{10})^2\cdot\frac{1}{10}+(\frac{1}{10})^2\cdot\frac{9}{10}} = \frac{9}{10}.$$

(c) Yes, they are the same, since Bayes' rule is coherent. The probability of an event given various pieces of evidence does not depend on the order in which the pieces of evidence are incorporated into the updated probabilities.

Independence and conditional independence

30. Ⓢ A family has 3 children, creatively named A, B, and C.

(a) Discuss intuitively (but clearly) whether the event "A is older than B" is independent of the event "A is older than C".

(b) Find the probability that A is older than B, given that A is older than C.

Solution:

(a) They are not independent: knowing that A is older than B makes it more likely that A is older than C, as the if A is older than B, then the only way that A can be younger than C is if the birth order is CAB, whereas the birth orders ABC and ACB are both compatible with A being older than B. To make this more intuitive, think of an extreme case where there are 100 children instead of 3, call them $A_1, \ldots, A_{100}$. Given that A_1 is older than all of $A_2, A_3, \ldots, A_{99}$, it's clear that A_1 is very old (relatively), whereas there isn't evidence about where A_{100} fits into the birth order.

(b) Writing $x > y$ to mean that x is older than y,

$$P(A > B|A > C) = \frac{P(A > B, A > C)}{P(A > C)} = \frac{1/3}{1/2} = \frac{2}{3}$$

since $P(A > B, A > C) = P(A$ is the eldest child$) = 1/3$ (unconditionally, any of the 3 children is equally likely to be the eldest).

31. Ⓢ Is it possible that an event is independent of itself? If so, when is this the case?

Solution: Let A be an event. If A is independent of itself, then $P(A) = P(A\cap A) = P(A)^2$, so $P(A)$ is 0 or 1. So this is only possible in the extreme cases that the event has probability 0 or 1.

32. Ⓢ Consider four nonstandard dice (the *Efron dice*), whose sides are labeled as follows (the 6 sides on each die are equally likely).

A: $4, 4, 4, 4, 0, 0$

B: $3, 3, 3, 3, 3, 3$

C: $6, 6, 2, 2, 2, 2$

D: $5, 5, 5, 1, 1, 1$

These four dice are each rolled once. Let A be the result for die A, B be the result for die B, etc.

(a) Find $P(A > B), P(B > C), P(C > D)$, and $P(D > A)$.

(b) Is the event $A > B$ independent of the event $B > C$? Is the event $B > C$ independent of the event $C > D$? Explain.

Solution:

(a)

$$P(A > B) = P(A = 4) = 2/3$$
$$P(B > C) = P(C = 2) = 2/3$$
$$P(C > D) = P(C = 6) + P(C = 2, D = 1) = 2/3$$
$$P(D > A) = P(D = 5) + P(D = 1, A = 0) = 2/3$$

(b) The event $A > B$ is independent of the event $B > C$ since $A > B$ is the same thing as $A = 4$, knowledge of which gives no information about $B > C$ (which is the same thing as $C = 2$). On the other hand, $B > C$ is *not* independent of $C > D$ since $P(C > D|C = 2) = 1/2 \neq 1 = P(C > D|C \neq 2)$.

35. Ⓢ You are going to play 2 games of chess with an opponent whom you have never played against before (for the sake of this problem). Your opponent is equally likely to be a beginner, intermediate, or a master. Depending on which, your chances of winning an individual game are 90%, 50%, or 30%, respectively.

(a) What is your probability of winning the first game?

(b) Congratulations: you won the first game! Given this information, what is the probability that you will also win the second game (assume that, given the skill level of your opponent, the outcomes of the games are independent)?

(c) Explain the distinction between assuming that the outcomes of the games are independent and assuming that they are conditionally independent given the opponent's skill level. Which of these assumptions seems more reasonable, and why?

Solution:

(a) Let W_i be the event of winning the ith game. By the law of total probability,

$$P(W_1) = (0.9 + 0.5 + 0.3)/3 = 17/30.$$

(b) We have $P(W_2|W_1) = P(W_2, W_1)/P(W_1)$. The denominator is known from (a), while the numerator can be found by conditioning on the skill level of the opponent:

$$P(W_1, W_2) = \frac{1}{3}P(W_1, W_2|\text{beginner}) + \frac{1}{3}P(W_1, W_2|\text{intermediate}) + \frac{1}{3}P(W_1, W_2|\text{expert}).$$

Since W_1 and W_2 are conditionally independent given the skill level of the opponent, this becomes

$$P(W_1, W_2) = (0.9^2 + 0.5^2 + 0.3^2)/3 = 23/60.$$

So

$$P(W_2|W_1) = \frac{23/60}{17/30} = 23/34.$$

(c) Independence here means that knowing one game's outcome gives no information about the other game's outcome, while conditional independence is the same statement where all probabilities are conditional on the opponent's skill level. Conditional independence given the opponent's skill level is a more reasonable assumption here. This is because winning the first game gives information about the opponent's skill level, which in turn gives information about the result of the second game.

That is, if the opponent's skill level is treated as fixed and known, then it may be reasonable to assume independence of games given this information; with the opponent's skill level random, earlier games can be used to help infer the opponent's skill level, which affects the probabilities for future games.

Monty Hall

38. Ⓢ (a) Consider the following 7-door version of the Monty Hall problem. There are 7 doors, behind one of which there is a car (which you want), and behind the rest of which there are goats (which you don't want). Initially, all possibilities are equally likely for where the car is. You choose a door. Monty Hall then opens 3 goat doors, and offers you the option of switching to any of the remaining 3 doors.

Assume that Monty Hall knows which door has the car, will always open 3 goat doors and offer the option of switching, and that Monty chooses with equal probabilities from all his choices of which goat doors to open. Should you switch? What is your probability of success if you switch to one of the remaining 3 doors?

(b) Generalize the above to a Monty Hall problem where there are $n \geq 3$ doors, of which Monty opens m goat doors, with $1 \leq m \leq n-2$.

Solution:

(a) Assume the doors are labeled such that you choose door 1 (to simplify notation), and suppose first that you follow the "stick to your original choice" strategy. Let S be the event of success in getting the car, and let C_j be the event that the car is behind door j. Conditioning on which door has the car, we have

$$P(S) = P(S|C_1)P(C_1) + \cdots + P(S|C_7)P(C_7) = P(C_1) = \frac{1}{7}.$$

Let M_{ijk} be the event that Monty opens doors i, j, k. Then

$$P(S) = \sum_{i,j,k} P(S|M_{ijk})P(M_{ijk})$$

(summed over all i, j, k with $2 \leq i < j < k \leq 7$.) By symmetry, this gives

$$P(S|M_{ijk}) = P(S) = \frac{1}{7}$$

for all i, j, k with $2 \leq i < j < k \leq 7$. Thus, the conditional probability that the car is behind 1 of the remaining 3 doors is $6/7$, which gives $2/7$ for each. So you should switch, thus making your probability of success $2/7$ rather than $1/7$.

(b) By the same reasoning, the probability of success for "stick to your original choice" is $\frac{1}{n}$, both unconditionally and conditionally. Each of the $n-m-1$ remaining doors has conditional probability $\frac{n-1}{(n-m-1)n}$ of having the car. This value is greater than $\frac{1}{n}$, so you should switch, thus obtaining probability $\frac{n-1}{(n-m-1)n}$ of success (both conditionally and unconditionally).

39. Ⓢ Consider the Monty Hall problem, except that Monty enjoys opening door 2 more than he enjoys opening door 3, and if he has a choice between opening these two doors, he opens door 2 with probability p, where $\frac{1}{2} \leq p \leq 1$.

To recap: there are three doors, behind one of which there is a car (which you want), and behind the other two of which there are goats (which you don't want). Initially, all possibilities are equally likely for where the car is. You choose a door, which for concreteness we assume is door 1. Monty Hall then opens a door to reveal a goat, and offers you the option of switching. Assume that Monty Hall knows which door has the car, will always open a goat door and offer the option of switching, and as above assume that if Monty Hall has a choice between opening door 2 and door 3, he chooses door 2 with probability p (with $\frac{1}{2} \leq p \leq 1$).

(a) Find the unconditional probability that the strategy of always switching succeeds (unconditional in the sense that we do not condition on which of doors 2 or 3 Monty opens).

(b) Find the probability that the strategy of always switching succeeds, given that Monty opens door 2.

(c) Find the probability that the strategy of always switching succeeds, given that Monty opens door 3.

Solution:

(a) Let C_j be the event that the car is hidden behind door j and let W be the event that we win using the switching strategy. Using the law of total probability, we can find the unconditional probability of winning:

$$\begin{aligned} P(W) &= P(W|C_1)P(C_1) + P(W|C_2)P(C_2) + P(W|C_3)P(C_3) \\ &= 0 \cdot 1/3 + 1 \cdot 1/3 + 1 \cdot 1/3 = 2/3. \end{aligned}$$

(b) A tree method works well here (delete the paths which are no longer relevant after the conditioning, and reweight the remaining values by dividing by their sum), or we can use Bayes' rule and the law of total probability (as below).

Let D_i be the event that Monty opens Door i. Note that we are looking for $P(W|D_2)$, which is the same as $P(C_3|D_2)$ as we first choose Door 1 and then switch to Door 3. By Bayes' rule and the law of total probability,

$$\begin{aligned} P(C_3|D_2) &= \frac{P(D_2|C_3)P(C_3)}{P(D_2)} \\ &= \frac{P(D_2|C_3)P(C_3)}{P(D_2|C_1)P(C_1) + P(D_2|C_2)P(C_2) + P(D_2|C_3)P(C_3)} \\ &= \frac{1 \cdot 1/3}{p \cdot 1/3 + 0 \cdot 1/3 + 1 \cdot 1/3} \\ &= \frac{1}{1+p}. \end{aligned}$$

(c) The structure of the problem is the same as part (b) (except for the condition that $p \geq 1/2$, which was not needed above). Imagine repainting doors 2 and 3, reversing which is called which. By part (b) with $1-p$ in place of p, $P(C_2|D_3) = \frac{1}{1+(1-p)} = \frac{1}{2-p}$.

First-step analysis and gambler's ruin

42. Ⓢ A fair die is rolled repeatedly, and a running total is kept (which is, at each time, the total of all the rolls up until that time). Let p_n be the probability that the running total is ever *exactly* n (assume the die will always be rolled enough times so that the running total will eventually exceed n, but it may or may not ever equal n).

(a) Write down a recursive equation for p_n (relating p_n to earlier terms p_k in a simple way). Your equation should be true for all positive integers n, so give a definition of p_0 and p_k for $k < 0$ so that the recursive equation is true for small values of n.

(b) Find p_7.

(c) Give an intuitive explanation for the fact that $p_n \to 1/3.5 = 2/7$ as $n \to \infty$.

Solution:

(a) We will find something to condition on to reduce the case of interest to earlier, simpler cases. This is achieved by the useful strategy of *first step anaysis*. Let p_n be the probability that the running total is ever *exactly* n. Note that if, for example, the first

throw is a 3, then the probability of reaching n exactly is p_{n-3} since starting from that point, we need to get a total of $n-3$ exactly. So

$$p_n = \frac{1}{6}(p_{n-1} + p_{n-2} + p_{n-3} + p_{n-4} + p_{n-5} + p_{n-6}),$$

where we define $p_0 = 1$ (which makes sense anyway since the running total is 0 before the first toss) and $p_k = 0$ for $k < 0$.

(b) Using the recursive equation in (a), we have

$$p_1 = \frac{1}{6}, \quad p_2 = \frac{1}{6}(1 + \frac{1}{6}), \quad p_3 = \frac{1}{6}(1 + \frac{1}{6})^2,$$
$$p_4 = \frac{1}{6}(1 + \frac{1}{6})^3, \quad p_5 = \frac{1}{6}(1 + \frac{1}{6})^4, \quad p_6 = \frac{1}{6}(1 + \frac{1}{6})^5.$$

Hence,

$$p_7 = \frac{1}{6}(p_1 + p_2 + p_3 + p_4 + p_5 + p_6) = \frac{1}{6}\left((1 + \frac{1}{6})^6 - 1\right) \approx 0.2536.$$

(c) An intuitive explanation is as follows. The average number thrown by the die is (total of dots)/6, which is $21/6 = 7/2$, so that every throw adds on an average of $7/2$. We can therefore expect to land on 2 out of every 7 numbers, and the probability of landing on any particular number is $2/7$. A mathematical derivation (which was not requested in the problem) can be given as follows:

$$\begin{aligned}
&p_{n+1} + 2p_{n+2} + 3p_{n+3} + 4p_{n+4} + 5p_{n+5} + 6p_{n+6} \\
=&p_{n+1} + 2p_{n+2} + 3p_{n+3} + 4p_{n+4} + 5p_{n+5} \\
&+ p_n + p_{n+1} + p_{n+2} + p_{n+3} + p_{n+4} + p_{n+5} \\
=&p_n + 2p_{n+1} + 3p_{n+2} + 4p_{n+3} + 5p_{n+4} + 6p_{n+5} \\
=&\cdots \\
=&p_{-5} + 2p_{-4} + 3p_{-3} + 4p_{-2} + 5p_{-1} + 6p_0 = 6.
\end{aligned}$$

Taking the limit of the lefthand side as n goes to ∞, we have

$$(1 + 2 + 3 + 4 + 5 + 6) \lim_{n\to\infty} p_n = 6,$$

so $\lim_{n\to\infty} p_n = 2/7$.

44. Ⓢ Calvin and Hobbes play a match consisting of a series of games, where Calvin has probability p of winning each game (independently). They play with a "win by two"" rule: the first player to win two games more than his opponent wins the match. Find the probability that Calvin wins the match (in terms of p), in two different ways:

(a) by conditioning, using the law of total probability.

(b) by interpreting the problem as a gambler's ruin problem.

Solution:

(a) Let C be the event that Calvin wins the match, $X \sim \text{Bin}(2, p)$ be how many of the first 2 games he wins, and $q = 1 - p$. Then

$$P(C) = P(C|X = 0)q^2 + P(C|X = 1)(2pq) + P(C|X = 2)p^2 = 2pqP(C) + p^2,$$

so $P(C) = \frac{p^2}{1-2pq}$. This can also be written as $\frac{p^2}{p^2+q^2}$, since $p + q = 1$.

Sanity check: Note that this should (and does) reduce to 1 for $p = 1$, 0 for $p = 0$, and

$\frac{1}{2}$ for $p = \frac{1}{2}$. Also, it makes sense that the probability of Hobbes winning, which is $1 - P(C) = \frac{q^2}{p^2+q^2}$, can also be obtained by swapping p and q.

(b) The problem can be thought of as a gambler's ruin where each player starts out with \$2. So the probability that Calvin wins the match is

$$\frac{1-(q/p)^2}{1-(q/p)^4} = \frac{(p^2-q^2)/p^2}{(p^4-q^4)/p^4} = \frac{(p^2-q^2)/p^2}{(p^2-q^2)(p^2+q^2)/p^4} = \frac{p^2}{p^2+q^2},$$

which agrees with the above.

Simpson's paradox

49. Ⓢ (a) Is it possible to have events A, B, C such that $P(A|C) < P(B|C)$ and $P(A|C^c) < P(B|C^c)$, yet $P(A) > P(B)$? That is, A is less likely than B given that C is true, and also less likely than B given that C is false, yet A is more likely than B if we're given no information about C. Show this is impossible (with a short proof) or find a counterexample (with a story interpreting A, B, C).

(b) If the scenario in (a) is possible, is it a special case of Simpson's paradox, equivalent to Simpson's paradox, or neither? If it is impossible, explain intuitively why it is impossible even though Simpson's paradox is possible.

Solution:

(a) It is *not* possible, as seen using the law of total probability:

$$P(A) = P(A|C)P(C) + P(A|C^c)P(C^c) < P(B|C)P(C) + P(B|C^c)P(C^c) = P(B).$$

(b) In Simpson's paradox, using the notation from the chapter, we can expand out $P(A|B)$ and $P(A|B^c)$ using LOTP to condition on C, but the inequality can flip because of the weights such as $P(C|B)$ on the terms (e.g., Dr. Nick performs a lot more Band-Aid removals than Dr. Hibbert). In this problem, the weights $P(C)$ and $P(C^c)$ are the same in both expansions, so the inequality is preserved.

50. Ⓢ Consider the following conversation from an episode of *The Simpsons*:

> Lisa: *Dad, I think he's an ivory dealer! His boots are ivory, his hat is ivory, and I'm pretty sure that check is ivory.*
>
> Homer: *Lisa, a guy who has lots of ivory is* less *likely to hurt Stampy than a guy whose ivory supplies are low.*

Here Homer and Lisa are debating the question of whether or not the man (named Blackheart) is likely to hurt Stampy the Elephant if they sell Stampy to him. They clearly disagree about how to use their observations about Blackheart to learn about the probability (conditional on the evidence) that Blackheart will hurt Stampy.

(a) Define clear notation for the various events of interest here.

(b) Express Lisa's and Homer's arguments (Lisa's is partly implicit) as conditional probability statements in terms of your notation from (a).

(c) Assume it is true that someone who has a lot of a commodity will have less desire to acquire more of the commodity. Explain what is wrong with Homer's reasoning that the evidence about Blackheart makes it less likely that he will harm Stampy.

Solution:

(a) Let H be the event that the man will hurt Stampy, let L be the event that a man has lots of ivory, and let D be the event that the man is an ivory dealer.

(b) Lisa observes that L is true. She suggests (reasonably) that this evidence makes D more likely, i.e., $P(D|L) > P(D)$. Implicitly, she suggests that this makes it likely that the man will hurt Stampy, i.e.,

$$P(H|L) > P(H|L^c).$$

Homer argues that

$$P(H|L) < P(H|L^c).$$

(c) Homer does not realize that observing that Blackheart has so much ivory makes it much more likely that Blackheart is an ivory dealer, which in turn makes it more likely that the man will hurt Stampy. This is an example of Simpson's paradox. It may be true that, *controlling for whether or not Blackheart is a dealer*, having high ivory supplies makes it less likely that he will harm Stampy: $P(H|L, D) < P(H|L^c, D)$ and $P(H|L, D^c) < P(H|L^c, D^c)$. However, this does not imply that $P(H|L) < P(H|L^c)$.

53. Ⓢ The book *Red State, Blue State, Rich State, Poor State* by Andrew Gelman [13] discusses the following election phenomenon: within any U.S. state, a wealthy voter is more likely to vote for a Republican than a poor voter, yet the wealthier states tend to favor Democratic candidates! In short: rich individuals (in any state) tend to vote for Republicans, while states with a higher percentage of rich people tend to favor Democrats.

(a) Assume for simplicity that there are only 2 states (called Red and Blue), each of which has 100 people, and that each person is either rich or poor, and either a Democrat or a Republican. Make up numbers consistent with the above, showing how this phenomenon is possible, by giving a 2×2 table for each state (listing how many people in each state are rich Democrats, etc.).

(b) In the setup of (a) (not necessarily with the numbers you made up there), let D be the event that a randomly chosen person is a Democrat (with all 200 people equally likely), and B be the event that the person lives in the Blue State. Suppose that 10 people move from the Blue State to the Red State. Write P_{old} and P_{new} for probabilities before and after they move. Assume that people do not change parties, so we have $P_{\text{new}}(D) = P_{\text{old}}(D)$. Is it possible that *both* $P_{\text{new}}(D|B) > P_{\text{old}}(D|B)$ and $P_{\text{new}}(D|B^c) > P_{\text{old}}(D|B^c)$ are true? If so, explain how it is possible and why it does not contradict the law of total probability $P(D) = P(D|B)P(B) + P(D|B^c)P(B^c)$; if not, show that it is impossible.

Solution:

(a) Here are two tables that are as desired:

Red	Dem	Rep	**Total**
Rich	5	25	30
Poor	20	50	70
Total	25	75	100

Blue	Dem	Rep	**Total**
Rich	45	15	60
Poor	35	5	40
Total	80	20	100

In these tables, within each state a rich person is more likely to be a Republican than a poor person; but the richer state has a higher percentage of Democrats than the poorer state. Of course, there are many possible tables that work.

The above example is a form of Simpson's paradox: aggregating the two tables seems to give different conclusions than conditioning on which state a person is in. Letting D, W, B be the events that a randomly chosen person is a Democrat, wealthy, and from the Blue State (respectively), for the above numbers we have $P(D|W, B) < P(D|W^c, B)$ and $P(D|W, B^c) < P(D|W^c, B^c)$ (controlling for whether the person is in the Red State or the Blue State, a poor person is more likely to be a Democrat than a rich person),

but $P(D|W) > P(D|W^c)$ (stemming from the fact that the Blue State is richer and more Democratic).

(b) Yes, it is possible. Suppose with the numbers from (a) that 10 people move from the Blue State to the Red State, of whom 5 are Democrats and 5 are Republicans. Then $P_{\text{new}}(D|B) = 75/90 > 80/100 = P_{\text{old}}(D|B)$ and $P_{\text{new}}(D|B^c) = 30/110 > 25/100 = P_{\text{old}}(D|B^c)$. Intuitively, this makes sense since the Blue State has a higher percentage of Democrats initially than the Red State, and the people who move have a percentage of Democrats which is between these two values.

This result does not contradict the law of total probability since the weights $P(B), P(B^c)$ also change: $P_{\text{new}}(B) = 90/200$, while $P_{\text{old}}(B) = 1/2$. The phenomenon could not occur if an equal number of people also move from the Red State to the Blue State (so that $P(B)$ is kept constant).

Chapter 3: Random variables and their distributions

PMFs and CDFs

6. Ⓢ *Benford's law* states that in a very large variety of real-life data sets, the first digit approximately follows a particular distribution with about a 30% chance of a 1, an 18% chance of a 2, and in general

$$P(D=j) = \log_{10}\left(\frac{j+1}{j}\right), \text{ for } j \in \{1,2,3,\ldots,9\},$$

where D is the first digit of a randomly chosen element. Check that this is a valid PMF (using properties of logs, not with a calculator).

Solution: The function $P(D=j)$ is nonnegative and the sum over all values is

$$\sum_{j=1}^{9} \log_{10}\frac{j+1}{j} = \sum_{j=1}^{9}(\log_{10}(j+1) - \log_{10}(j)).$$

All terms cancel except $\log_{10} 10 - \log_{10} 1 = 1$ (this is a *telescoping series*). Since the values add to 1 and are nonnegative, $P(D=j)$ is a PMF.

11. Ⓢ Let X be an r.v. whose possible values are $0, 1, 2, \ldots$, with CDF F. In some countries, rather than using a CDF, the convention is to use the function G defined by $G(x) = P(X < x)$ to specify a distribution. Find a way to convert from F to G, i.e., if F is a known function, show how to obtain $G(x)$ for all real x.

Solution: Write

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

If x is not a nonnegative integer, then $P(X=x) = 0$ so $G(x) = F(x)$. For x a nonnegative integer,

$$P(X = x) = F(x) - F(x - 1/2)$$

since the PMF corresponds to the lengths of the jumps in the CDF. (The $1/2$ was chosen for concreteness; we also have $F(x-1/2) = F(x-a)$ for any $a \in (0,1]$.) Thus,

$$G(x) = \begin{cases} F(x) \text{ if } x \notin \{0,1,2,\ldots\} \\ F(x-1/2) \text{ if } x \in \{0,1,2,\ldots\}. \end{cases}$$

More compactly, we can also write $G(x) = \lim_{t \to x^-} F(t)$, where the $-$ denotes taking a limit from the left (recall that F is right continuous), and $G(x) = F(\lceil x \rceil - 1)$, where $\lceil x \rceil$ is the ceiling of x (the smallest integer greater than or equal to x).

Named distributions

18. Ⓢ (a) In the World Series of baseball, two teams (call them A and B) play a sequence of games against each other, and the first team to win four games wins the series. Let

p be the probability that A wins an individual game, and assume that the games are independent. What is the probability that team A wins the series?

(b) Give a clear intuitive explanation of whether the answer to (a) depends on whether the teams always play 7 games (and whoever wins the majority wins the series), or the teams stop playing more games as soon as one team has won 4 games (as is actually the case in practice: once the match is decided, the two teams do not keep playing more games).

Solution:

(a) Let $q = 1 - p$. First let us do a direct calculation:

$$\begin{aligned} P(\text{A wins}) &= P(\text{A wins in 4 games}) + P(\text{A wins in 5 games}) \\ &\quad + P(\text{A wins in 6 games}) + P(\text{A wins in 7 games}) \\ &= p^4 + \binom{4}{3}p^4 q + \binom{5}{3}p^4 q^2 + \binom{6}{3}p^4 q^3. \end{aligned}$$

To understand how these probabilities are calculated, note for example that

$$\begin{aligned} P(\text{A wins in 5}) &= P(\text{A wins 3 out of first 4}) \cdot P(\text{A wins 5th game}|\text{A wins 3 out of first 4}) \\ &= \binom{4}{3}p^3 qp. \end{aligned}$$

(Each of the 4 terms in the expression for $P(\text{A wins})$ can also be found using the PMF of a distribution known as the *Negative Binomial*, which is introduced in Chapter 4.)

An neater solution is to use the fact (explained in the solution to Part (b)) that we can assume that the teams play all 7 games no matter what. Let X be the number of wins for team A, so that $X \sim \text{Bin}(7, p)$. Then

$$\begin{aligned} P(X \geq 4) &= P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7) \\ &= \binom{7}{4}p^4 q^3 + \binom{7}{5}p^5 q^2 + \binom{7}{6}p^6 q + p^7, \end{aligned}$$

which looks different from the above but is actually identical as a function of p (as can be verified by simplifying both expressions as polynomials in p).

(b) The answer to (a) does not depend on whether the teams play all seven games no matter what. Imagine telling the players to continue playing the games even after the match has been decided, just for fun: the outcome of the match won't be affected by this, and this also means that the probability that A wins the match won't be affected by assuming that the teams always play 7 games!

21. Ⓢ Let $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(m, p)$, independent of X. Show that $X - Y$ is *not* Binomial.

Solution: A Binomial can't be negative, but $X - Y$ is negative with positive probability.

25. Ⓢ Alice flips a fair coin n times and Bob flips another fair coin $n + 1$ times, resulting in independent $X \sim \text{Bin}(n, \frac{1}{2})$ and $Y \sim \text{Bin}(n + 1, \frac{1}{2})$.

(a) Show that $P(X < Y) = P(n - X < n + 1 - Y)$.

(b) Compute $P(X < Y)$.

Hint: Use (a) and the fact that X and Y are integer-valued.

Solution:

(a) Note that $n-X \sim \text{Bin}(n, 1/2)$ and $n+1-Y \sim \text{Bin}(n+1, 1/2)$ (we can interpret this by thinking of counting Tails rather than counting Heads), with $n - X$ and $n + 1 - Y$ independent. So $P(X < Y) = P(n - X < n + 1 - Y)$, since both sides have exactly the same structure.

(b) We have

$$P(X < Y) = P(n - X < n + 1 - Y) = P(Y < X + 1) = P(Y \leq X)$$

since X and Y are integer-valued (e.g., $Y < 5$ is equivalent to $Y \leq 4$). But $Y \leq X$ is the complement of $X < Y$, so $P(X < Y) = 1 - P(X < Y)$. Thus, $P(X < Y) = 1/2$.

28. Ⓢ There are n eggs, each of which hatches a chick with probability p (independently). Each of these chicks survives with probability r, independently. What is the distribution of the number of chicks that hatch? What is the distribution of the number of chicks that survive? (Give the PMFs; also give the names of the distributions and their parameters, if applicable.)

Solution:

Let H be the number of eggs that hatch and X be the number of hatchlings that survive. Think of each egg as a Bernoulli trial, where for H we define "success" to mean hatching, while for X we define "success" to mean surviving. For example, in the picture above, where ☺ denotes an egg that hatches with the chick surviving, x denotes an egg that hatched but whose chick died, and ▢ denotes an egg that didn't hatch, the events $H = 7$, $X = 5$ occurred. By the story of the Binomial, $H \sim \text{Bin}(n, p)$, with PMF

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

for $k = 0, 1, \ldots, n$.

The eggs independently have probability pr each of hatching a chick that survives. By the story of the Binomial, we have $X \sim \text{Bin}(n, pr)$, with PMF

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

for $k = 0, 1, \ldots, n$.

29. Ⓢ A sequence of n independent experiments is performed. Each experiment is a success with probability p and a failure with probability $q = 1 - p$. Show that conditional on the number of successes, all valid possibilities for the list of outcomes of the experiment are equally likely.

Solution:

Let X_j be 1 if the jth experiment is a success and 0 otherwise, and let $X = X_1 + \cdots + X_n$ be the total number of successes. Then for any k and any $a_1, \ldots, a_n \in \{0, 1\}$ with $a_1 + \cdots + a_n = k$,

$$\begin{aligned} P(X_1 = a_1, \ldots, X_n = a_n | X = k) &= \frac{P(X_1 = a_1, \ldots, X_n = a_n, X = k)}{P(X = k)} \\ &= \frac{P(X_1 = a_1, \ldots, X_n = a_n)}{P(X = k)} \\ &= \frac{p^k q^{n-k}}{\binom{n}{k} p^k q^{n-k}} \\ &= \frac{1}{\binom{n}{k}}. \end{aligned}$$

This does not depend on $a_1, \ldots, a_n$. Thus, for n independent Bernoulli trials, given that there are exactly k successes, the $\binom{n}{k}$ possible sequences consisting of k successes and $n-k$ failures are equally likely. Interestingly, the conditional probability above also does not depend on p (this is closely related to the notion of a *sufficient statistic*, which is an important concept in statistical inference).

35. Ⓢ Players A and B take turns in answering trivia questions, starting with player A answering the first question. Each time A answers a question, she has probability p_1 of getting it right. Each time B plays, he has probability p_2 of getting it right.

(a) If A answers m questions, what is the PMF of the number of questions she gets right?

(b) If A answers m times and B answers n times, what is the PMF of the total number of questions they get right (you can leave your answer as a sum)? Describe exactly when/whether this is a Binomial distribution.

(c) Suppose that the first player to answer correctly wins the game (with no predetermined maximum number of questions that can be asked). Find the probability that A wins the game.

Solution:

(a) The r.v. is Bin(m, p_1), so the PMF is $\binom{m}{k} p_1^k (1-p_1)^{m-k}$ for $k \in \{0, 1, \ldots, m\}$.

(b) Let T be the total number of questions they get right. To get a total of k questions right, it must be that A got 0 and B got k, or A got 1 and B got $k-1$, etc. These are disjoint events so the PMF is

$$P(T=k) = \sum_{j=0}^{k} \binom{m}{j} p_1^j (1-p_1)^{m-j} \binom{n}{k-j} p_2^{k-j} (1-p_2)^{n-(k-j)}$$

for $k \in \{0, 1, \ldots, m+n\}$, with the usual convention that $\binom{n}{k}$ is 0 for $k > n$.

This is the Bin$(m+n, p)$ distribution if $p_1 = p_2 = p$ (using the story for the Binomial, or using Vandermonde's identity). For $p_1 \neq p_2$, it's not a Binomial distribution, since the trials have different probabilities of success; having some trials with one probability of success and other trials with another probability of success isn't equivalent to having trials with some "effective" probability of success.

(c) Let $r = P(\text{A wins})$. Conditioning on the results of the first question for each player, we have

$$r = p_1 + (1-p_1)p_2 \cdot 0 + (1-p_1)(1-p_2)r,$$

which gives $r = \frac{p_1}{1-(1-p_1)(1-p_2)} = \frac{p_1}{p_1+p_2-p_1p_2}$.

37. Ⓢ A message is sent over a noisy channel. The message is a sequence $x_1, x_2, \ldots, x_n$ of n bits ($x_i \in \{0, 1\}$). Since the channel is noisy, there is a chance that any bit might be corrupted, resulting in an error (a 0 becomes a 1 or vice versa). Assume that the error events are independent. Let p be the probability that an individual bit has an error ($0 < p < 1/2$). Let $y_1, y_2, \ldots, y_n$ be the received message (so $y_i = x_i$ if there is no error in that bit, but $y_i = 1 - x_i$ if there is an error there).

To help detect errors, the nth bit is reserved for a parity check: x_n is defined to be 0 if $x_1 + x_2 + \cdots + x_{n-1}$ is even, and 1 if $x_1 + x_2 + \cdots + x_{n-1}$ is odd. When the message is received, the recipient checks whether y_n has the same parity as $y_1 + y_2 + \cdots + y_{n-1}$. If the parity is wrong, the recipient knows that at least one error occurred; otherwise, the recipient assumes that there were no errors.

(a) For $n = 5, p = 0.1$, what is the probability that the received message has errors which go undetected?

(b) For general n and p, write down an expression (as a sum) for the probability that the received message has errors which go undetected.

(c) Give a simplified expression, not involving a sum of a large number of terms, for the probability that the received message has errors which go undetected.

Hint for (c): Letting

$$a = \sum_{k \text{ even}, k \geq 0} \binom{n}{k} p^k (1-p)^{n-k} \text{ and } b = \sum_{k \text{ odd}, k \geq 1} \binom{n}{k} p^k (1-p)^{n-k},$$

the binomial theorem makes it possible to find simple expressions for $a + b$ and $a - b$, which then makes it possible to obtain a and b.

Solution:

(a) Note that $\sum_{i=1}^n x_i$ is even. If the number of errors is even (and nonzero), the errors will go undetected; otherwise, $\sum_{i=1}^n y_i$ will be odd, so the errors will be detected.
The number of errors is $\text{Bin}(n, p)$, so the probability of undetected errors when $n = 5, p = 0.1$ is

$$\binom{5}{2} p^2 (1-p)^3 + \binom{5}{4} p^4 (1-p) \approx 0.073.$$

(b) By the same reasoning as in (a), the probability of undetected errors is

$$\sum_{k \text{ even}, k \geq 2} \binom{n}{k} p^k (1-p)^{n-k}.$$

(c) Let a, b be as in the hint. Then

$$a + b = \sum_{k \geq 0} \binom{n}{k} p^k (1-p)^{n-k} = 1,$$

$$a - b = \sum_{k \geq 0} \binom{n}{k} (-p)^k (1-p)^{n-k} = (1-2p)^n.$$

Solving for a and b gives

$$a = \frac{1 + (1-2p)^n}{2} \text{ and } b = \frac{1 - (1-2p)^n}{2}.$$

$$\sum_{k \text{ even}, k \geq 0} \binom{n}{k} p^k (1-p)^{n-k} = \frac{1 + (1-2p)^n}{2}.$$

Subtracting off the possibility of no errors, we have

$$\sum_{k \text{ even}, k \geq 2} \binom{n}{k} p^k (1-p)^{n-k} = \frac{1 + (1-2p)^n}{2} - (1-p)^n.$$

Sanity check: Note that letting $n = 5, p = 0.1$ here gives 0.073, which agrees with (a); letting $p = 0$ gives 0, as it should; and letting $p = 1$ gives 0 for n odd and 1 for n even, which again makes sense.

Independence of r.v.s

42. Ⓢ Let X be a random day of the week, coded so that Monday is 1, Tuesday is 2, etc. (so X takes values $1, 2, \ldots, 7$, with equal probabilities). Let Y be the next day after X (again represented as an integer between 1 and 7). Do X and Y have the same distribution? What is $P(X < Y)$?

Solution: Yes, X and Y have the same distribution, since Y is also equally likely to represent any day of the week. However, X is likely to be less than Y. Specifically,

$$P(X < Y) = P(X \neq 7) = \frac{6}{7}.$$

In general, if Z and W are *independent* r.v.s with the same distribution, then $P(Z < W) = P(W < Z)$ by symmetry. Here though, X and Y are *dependent*, and we have $P(X < Y) = 6/7, P(X = Y) = 0, P(Y < X) = 1/7$.

Mixed practice

45. Ⓢ A new treatment for a disease is being tested, to see whether it is better than the standard treatment. The existing treatment is effective on 50% of patients. It is believed initially that there is a 2/3 chance that the new treatment is effective on 60% of patients, and a 1/3 chance that the new treatment is effective on 50% of patients. In a pilot study, the new treatment is given to 20 random patients, and is effective for 15 of them.

(a) Given this information, what is the probability that the new treatment is better than the standard treatment?

(b) A second study is done later, giving the new treatment to 20 new random patients. Given the results of the first study, what is the PMF for how many of the new patients the new treatment is effective on? (Letting p be the answer to (a), your answer can be left in terms of p.)

Solution:

(a) Let B be the event that the new treatment is better than the standard treatment and let X be the number of people in the study for whom the new treatment is effective. By Bayes' rule and LOTP,

$$\begin{aligned} P(B|X=15) &= \frac{P(X=15|B)P(B)}{P(X=15|B)P(B) + P(X=15|B^c)P(B^c)} \\ &= \frac{\binom{20}{15}(0.6)^{15}(0.4)^5(\frac{2}{3})}{\binom{20}{15}(0.6)^{15}(0.4)^5(\frac{2}{3}) + \binom{20}{15}(0.5)^{20}(\frac{1}{3})}. \end{aligned}$$

(b) Let Y be how many of the new patients the new treatment is effective for and $p = P(B|X = 15)$ be the answer from (a). Then for $k \in \{0, 1, \ldots, 20\}$,

$$\begin{aligned} P(Y=k|X=15) &= P(Y=k|X=15,B)P(B|X=15) + P(Y=k|X=15,B^c)P(B^c|X=15) \\ &= P(Y=k|B)P(B|X=15) + P(Y=k|B^c)P(B^c|X=15) \\ &= \binom{20}{k}(0.6)^k(0.4)^{20-k}p + \binom{20}{k}(0.5)^{20}(1-p). \end{aligned}$$

(This distribution is *not* Binomial. As in the coin with a random bias problem, the individual outcomes are conditionally independent but not independent. Given the true probability of effectiveness of the new treatment, the pilot study is irrelevant and the distribution is Binomial, but without knowing that, we have a mixture of two different Binomial distributions.)

Chapter 4: Expectation

Expectations and variances

13. Ⓢ Are there discrete random variables X and Y such that $E(X) > 100E(Y)$ but Y is greater than X with probability at least 0.99?

Solution: Yes. Consider what happens if we make X usually 0 but on rare occasions, X is extremely large (like the outcome of a lottery); Y, on the other hand, can be more moderate. For a simple example, let X be 10^6 with probability $1/100$ and 0 with probability $99/100$, and let Y be the constant 1 (which is a degenerate r.v.).

Named distributions

17. Ⓢ A couple decides to keep having children until they have at least one boy and at least one girl, and then stop. Assume they never have twins, that the "trials" are independent with probability $1/2$ of a boy, and that they are fertile enough to keep producing children indefinitely. What is the expected number of children?

Solution: Let X be the number of children needed, starting with the 2nd child, to obtain one whose gender is not the same as that of the firstborn. Then $X - 1$ is Geom(1/2), so $E(X) = 2$. This does not include the firstborn, so the expected total number of children is $E(X + 1) = E(X) + 1 = 3$.

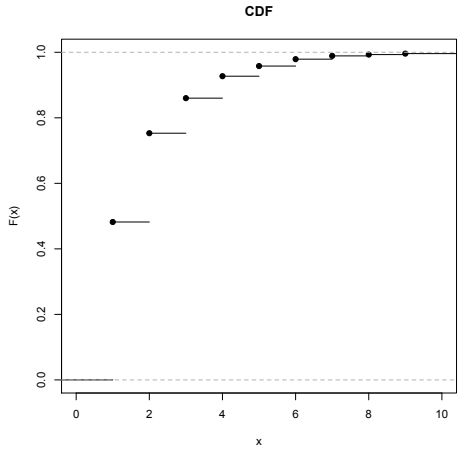
Sanity check: An answer of 2 or lower would be a miracle since the couple always needs to have at least 2 children, and sometimes they need more. An answer of 4 or higher would be a miracle since 4 is the expected number of children needed such that there is a boy and a girl with the boy older than the girl.

18. Ⓢ A coin is tossed repeatedly until it lands Heads for the first time. Let X be the number of tosses that are required (including the toss that landed Heads), and let p be the probability of Heads, so that $X \sim \text{FS}(p)$. Find the CDF of X, and for $p = 1/2$ sketch its graph.

Solution: By the story of the Geometric, we have $X - 1 \sim \text{Geometric}(p)$. Using this or directly, the PMF is $P(X = k) = p(1-p)^{k-1}$ for $k \in \{1, 2, 3, \dots\}$ (and 0 otherwise). The CDF can be obtained by adding up the PMF (from $k = 1$ to $k = \lfloor x \rfloor$, where $\lfloor x \rfloor$ is the greatest integer less than or equal to x). We can also see directly that

$$P(X \leq x) = 1 - P(X > x) = 1 - (1-p)^{\lfloor x \rfloor}$$

for $x \geq 1$, since $X > x$ says that the first $\lfloor x \rfloor$ flips land tails. The CDF is 0 for $x < 1$. For a fair coin, the CDF is $F(x) = 1 - \frac{1}{2^{\lfloor x \rfloor}}$ for $x \geq 1$, and $F(x) = 0$ for $x < 1$, as illustrated below.



20. Ⓢ Let $X \sim \text{Bin}(n, \frac{1}{2})$ and $Y \sim \text{Bin}(n+1, \frac{1}{2})$, independently. (This problem has been revised from that in the first printing of the book, to avoid overlap with Exercise 3.25.)

(a) Let $V = \min(X, Y)$ be the smaller of X and Y, and let $W = \max(X, Y)$ be the larger of X and Y. So if X crystallizes to x and Y crystallizes to y, then V crystallizes to $\min(x, y)$ and W crystallizes to $\max(x, y)$. Find $E(V) + E(W)$.

(b) Show that $E|X - Y| = E(W) - E(V)$, with notation as in (a).

(c) Compute $\text{Var}(n - X)$ in two different ways.

Solution:

(a) Note that $V + W = X + Y$ (since adding the smaller and the larger of two numbers is the same as adding both numbers). So by linearity,

$$E(V) + E(W) = E(V + W) = E(X + Y) = E(X) + E(Y) = (2n + 1)/2 = n + \frac{1}{2}.$$

(b) Note that $|X - Y| = W - V$ (since the absolute difference between two numbers is the larger number minus the smaller number). So

$$E|X - Y| = E(W - V) = E(W) - E(V).$$

(c) We have $n - X \sim \text{Bin}(n, 1/2)$, so $\text{Var}(n - X) = n/4$. Alternatively, by properties of variance we have $\text{Var}(n - X) = \text{Var}(n) + \text{Var}(-X) = \text{Var}(X) = n/4$.

21. Ⓢ Raindrops are falling at an average rate of 20 drops per square inch per minute. What would be a reasonable distribution to use for the number of raindrops hitting a particular region measuring 5 inches^2 in t minutes? Why? Using your chosen distribution, compute the probability that the region has no rain drops in a given 3-second time interval.

Solution: A reasonable choice of distribution is $\text{Pois}(\lambda t)$, where $\lambda = 20 \cdot 5 = 100$ (the average number of raindrops per minute hitting the region). Assuming this distribution,

$$P(\text{no raindrops in } 1/20 \text{ of a minute}) = e^{-100/20}(100/20)^0/0! = e^{-5} \approx 0.0067.$$

22. Ⓢ Alice and Bob have just met, and wonder whether they have a mutual friend. Each has 50 friends, out of 1000 other people who live in their town. They think that it's unlikely that they have a friend in common, saying "each of us is only friends with 5% of the people here, so it would be very unlikely that our two 5%'s overlap."

    Assume that Alice's 50 friends are a random sample of the 1000 people (equally likely to be any 50 of the 1000), and similarly for Bob. Also assume that knowing who Alice's friends are gives no information about who Bob's friends are.

    (a) Compute the expected number of mutual friends Alice and Bob have.

    (b) Let X be the number of mutual friends they have. Find the PMF of X.

    (c) Is the distribution of X one of the important distributions we have looked at? If so, which?

    Solution:

    (a) Let I_j be the indicator r.v. for the jth person being a mutual friend. Then

    $$E\left(\sum_{j=1}^{1000} I_j\right) = 1000E(I_1) = 1000P(I_1 = 1) = 1000 \cdot \left(\frac{5}{100}\right)^2 = 2.5.$$

    (b) Condition on who Alice's friends are, and then count the number of ways that Bob can be friends with exactly k of them. This gives

    $$P(X = k) = \frac{\binom{50}{k}\binom{950}{50-k}}{\binom{1000}{50}}$$

    for $0 \leq k \leq 50$ (and 0 otherwise).

    (c) Yes, it is the Hypergeometric distribution, as shown by the PMF from (b) or by thinking of "tagging" Alice's friends (like the elk) and then seeing how many tagged people there are among Bob's friends.

24. Ⓢ Calvin and Hobbes play a match consisting of a series of games, where Calvin has probability p of winning each game (independently). They play with a "win by two" rule: the first player to win two games more than his opponent wins the match. Find the expected number of games played.

    Hint: Consider the first two games as a pair, then the next two as a pair, etc.

    Solution: Think of the first 2 games, the 3rd and 4th, the 5th and 6th, etc. as "mini-matches." The match ends right after the first mini-match which isn't a tie. The probability of a mini-match not being a tie is $p^2 + q^2$, so the number of mini-matches needed is 1 plus a Geom$(p^2 + q^2)$ r.v. Thus, the expected number of games is $\frac{2}{p^2+q^2}$.

    Sanity check: For $p = 0$ or $p = 1$, this reduces to 2. The expected number of games is maximized when $p = \frac{1}{2}$, which makes sense intuitively. Also, it makes sense that the result is symmetric in p and q.

26. Ⓢ Let X and Y be Pois(λ) r.v.s, and $T = X + Y$. Suppose that X and Y are *not* independent, and in fact $X = Y$. Prove or disprove the claim that $T \sim$ Pois(2λ) in this scenario.

    Solution: The r.v. $T = 2X$ is *not* Poisson: it can only take even values $0, 2, 4, 6, \ldots$, whereas any Poisson r.v. has positive probability of being any of $0, 1, 2, 3, \ldots$.
    Alternatively, we can compute the PMF of $2X$, or note that $\text{Var}(2X) = 4\lambda \neq 2\lambda = E(2X)$, whereas for any Poisson r.v. the variance equals the mean.

29. Ⓢ A discrete distribution has the *memoryless property* if for X a random variable with that distribution, $P(X \geq j+k|X \geq j) = P(X \geq k)$ for all nonnegative integers j, k.

(a) If X has a memoryless distribution with CDF F and PMF $p_i = P(X = i)$, find an expression for $P(X \geq j+k)$ in terms of $F(j), F(k), p_j, p_k$.

(b) Name a discrete distribution which has the memoryless property. Justify your answer with a clear interpretation in words or with a computation.

Solution:
(a) By the memoryless property,

$$P(X \geq k) = P(X \geq j+k|X \geq j) = \frac{P(X \geq j+k, X \geq j)}{P(X \geq j)} = \frac{P(X \geq j+k)}{P(X \geq j)},$$

so

$$P(X \geq j+k) = P(X \geq j)P(X \geq k) = (1 - F(j) + p_j)(1 - F(k) + p_k).$$

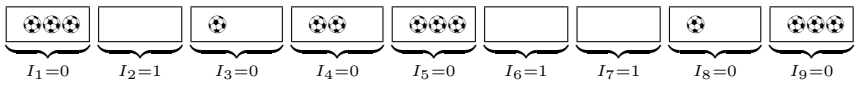
(b) The Geometric distribution is memoryless (in fact, it turns out to be essentially the *only* discrete memoryless distribution!). This follows from the story of the Geometric: consider Bernoulli trials, waiting for the first success (and defining waiting time to be the number of failures before the first success). Say we have already had j failures without a success. Then the additional waiting time from that point forward has the same distribution as the original waiting time (the Bernoulli trials neither are conspiring against the experimenter nor act as if he or she is due for a success: the trials are independent). A calculation agrees: for $X \sim \text{Geom}(p)$,

$$P(X \geq j+k|X \geq j) = \frac{P(X \geq j+k)}{P(X \geq j)} = \frac{q^{j+k}}{q^j} = q^k = P(X \geq k).$$

Indicator r.v.s

30. Ⓢ Randomly, k distinguishable balls are placed into n distinguishable boxes, with all possibilities equally likely. Find the expected number of empty boxes.

Solution:



Let I_j be the indicator random variable for the jth box being empty, so $I_1 + \cdots + I_n$ is the number of empty boxes (the above picture illustrates a possible outcome with 3 empty boxes, for $n = 9, k = 13$). Then $E(I_j) = P(I_j = 1) = (1 - 1/n)^k$. By linearity,

$$E\left(\sum_{j=1}^{n} I_j\right) = \sum_{j=1}^{n} E(I_j) = n(1 - 1/n)^k.$$

31. Ⓢ A group of 50 people are comparing their birthdays (as usual, assume their birthdays are independent, are not February 29, etc.). Find the expected number of pairs of people with the same birthday, and the expected number of days in the year on which at least two of these people were born.

Solution: Creating an indicator r.v. for each pair of people, we have that the expected number of pairs of people with the same birthday is $\binom{50}{2}\frac{1}{365}$ by linearity. Now create an indicator r.v. for each day of the year, taking the value 1 if at least two of the people were born that day (and 0 otherwise). Then the expected number of days on which at least two people were born is $365\left(1 - (\frac{364}{365})^{50} - 50 \cdot \frac{1}{365} \cdot (\frac{364}{365})^{49}\right)$.

32. Ⓢ A group of $n \geq 4$ people are comparing their birthdays (as usual, assume their birthdays are independent, are not February 29, etc.). Let I_{ij} be the indicator r.v. of i and j having the same birthday (for $i < j$). Is I_{12} independent of I_{34}? Is I_{12} independent of I_{13}? Are the I_{ij} independent?

Solution: The indicator I_{12} is independent of the indicator I_{34} since knowing the birthdays of persons 1 and 2 gives us no information about the birthdays of persons 3 and 4. Also, I_{12} is independent of I_{13} since even though both of these indicators involve person 1, knowing that persons 1 and 2 have the same birthday gives us no information about whether persons 1 and 3 have the same birthday (this relies on the assumption that the 365 days are equally likely). In general, the indicator r.v.s here are pairwise independent. But they are *not* independent since, for example, if person 1 has the same birthday as person 2 and person 1 has the same birthday as person 3, then persons 2 and 3 must have the same birthday.

33. Ⓢ A total of 20 bags of Haribo gummi bears are randomly distributed to 20 students. Each bag is obtained by a random student, and the outcomes of who gets which bag are independent. Find the average number of bags of gummi bears that the first three students get in total, and find the average number of students who get at least one bag.

Solution: Let X_j be the number of bags of gummi bears that the jth student gets, and let I_j be the indicator of $X_j \geq 1$. Then $X_j \sim \text{Bin}(20, \frac{1}{20})$, so $E(X_j) = 1$. So $E(X_1 + X_2 + X_3) = 3$ by linearity.
The average number of students who get at least one bag is

$$E(I_1 + \cdots + I_{20}) = 20E(I_1) = 20P(I_1 = 1) = 20\left(1 - \left(\frac{19}{20}\right)^{20}\right).$$

40. Ⓢ There are 100 shoelaces in a box. At each stage, you pick two random ends and tie them together. Either this results in a longer shoelace (if the two ends came from different pieces), or it results in a loop (if the two ends came from the same piece). What are the expected number of steps until everything is in loops, and the expected number of loops after everything is in loops? (This is a famous interview problem; leave the latter answer as a sum.)

Hint: For each step, create an indicator r.v. for whether a loop was created then, and note that the number of free ends goes down by 2 after each step.

Solution: Initially there are 200 free ends. The number of free ends decreases by 2 each time since either two separate pieces are tied together, or a new loop is formed. So exactly 100 steps are always needed. Let I_j be the indicator r.v. for whether a new loop is formed at the jth step. At the time when there are n unlooped pieces (so $2n$ ends), the probability of forming a new loop is $\frac{n}{\binom{2n}{2}} = \frac{1}{2n-1}$ since any 2 ends are equally likely to be chosen, and there are n ways to pick both ends of 1 of the n pieces. By linearity, the expected number of loops is

$$\sum_{n=1}^{100} \frac{1}{2n-1}.$$

44. Ⓢ Let X be Hypergeometric with parameters w, b, n.

(a) Find $E\binom{X}{2}$ *by thinking*, without any complicated calculations.

(b) Use (a) to find the variance of X. You should get

$$\text{Var}(X) = \frac{N-n}{N-1}npq,$$

where $N = w + b, p = w/N, q = 1 - p$.

Solution:

(a) In the story of the Hypergeometric, $\binom{X}{2}$ is the number of pairs of draws such that both balls are white. Creating an indicator r.v. for each pair, we have

$$E\binom{X}{2} = \binom{n}{2}\frac{w}{w+b}\frac{w-1}{w+b-1}.$$

(b) By (a),

$$EX^2 - EX = E(X(X-1)) = n(n-1)p\frac{w-1}{N-1},$$

so

$$\begin{aligned}
\mathrm{Var}(X) &= E(X^2) - (EX)^2 \\
&= n(n-1)p\frac{w-1}{N-1} + np - n^2p^2 \\
&= np\left(\frac{(n-1)(w-1)}{N-1} + 1 - np\right) \\
&= np\left(\frac{nw - w - n + N}{N-1} - \frac{nw}{N}\right) \\
&= np\left(\frac{Nnw - Nw - Nn + N^2 - Nnw + nw}{N(N-1)}\right) \\
&= np\left(\frac{(N-n)(N-w)}{N(N-1)}\right) \\
&= \frac{N-n}{N-1}npq.
\end{aligned}$$

47. Ⓢ A hash table is being used to store the phone numbers of k people, storing each person's phone number in a uniformly random location, represented by an integer between 1 and n (see Exercise 25 from Chapter 1 for a description of hash tables). Find the expected number of locations with no phone numbers stored, the expected number with exactly one phone number, and the expected number with more than one phone number (should these quantities add up to n?).

Solution: Let I_j be an indicator random variable equal to 1 if the j^{th} location is empty, and 0 otherwise, for $1 \leq j \leq n$. Then $P(I_j = 1) = (1 - 1/n)^k$, since the phone numbers are stored in independent random locations. Then $I_1 + \cdots + I_n$ is the number of empty locations. By linearity of expectation, we have

$$E\left(\sum_{j=1}^{n} I_j\right) = \sum_{j=1}^{n} E(I_j) = n(1-1/n)^k.$$

Similarly, the probability of a specific location having exactly 1 phone number stored is $\frac{k}{n}(1-\frac{1}{n})^{k-1}$, so the expected number of such locations is $k(1-1/n)^{k-1}$. By linearity, the sum of the three expected values is n, so the expected number of locations with more than one phone number is $n - n(1-1/n)^k - k(1-1/n)^{k-1}$.

50. Ⓢ Consider the following algorithm, known as *bubble sort*, for sorting a list of n distinct numbers into increasing order. Initially they are in a random order, with all orders equally likely. The algorithm compares the numbers in positions 1 and 2, and swaps them if needed, then it compares the new numbers in positions 2 and 3, and swaps them if needed, etc., until it has gone through the whole list. Call this one "sweep" through the list. After the first sweep, the largest number is at the end, so the second sweep (if needed) only needs to work with the first $n-1$ positions. Similarly, the third sweep (if needed) only needs to work with the first $n-2$ positions, etc. Sweeps are performed until $n-1$ sweeps have been completed or there is a swapless sweep.

For example, if the initial list is 53241 (omitting commas), then the following 4 sweeps are performed to sort the list, with a total of 10 comparisons:

$$53241 \to 35241 \to 32541 \to 32451 \to 32415.$$

$$32415 \to 23415 \to 23415 \to 23145.$$

$$23145 \to 23145 \to 21345.$$

$$21345 \to 12345.$$

(a) An *inversion* is a pair of numbers that are out of order (e.g., 12345 has no inversions, while 53241 has 8 inversions). Find the expected number of inversions in the original list.

(b) Show that the expected number of comparisons is between $\frac{1}{2}\binom{n}{2}$ and $\binom{n}{2}$.

Hint: For one bound, think about how many comparisons are made if $n-1$ sweeps are done; for the other bound, use Part (a).

Solution:

(a) There are $\binom{n}{2}$ pairs of numbers, each of which is equally likely to be in either order. So by symmetry, linearity, and indicator r.v.s, the expected number of inversions is $\frac{1}{2}\binom{n}{2}$.

(b) Let X be the number of comparisons and V be the number of inversions. On the one hand, $X \geq V$ since every inversion must be repaired. So $E(X) \geq E(V) = \frac{1}{2}\binom{n}{2}$. On the other hand, there are $n-1$ comparisons needed in the first sweep, $n-2$ in the second sweep (if needed), ..., and 1 in the $(n-1)$st sweep (if needed). So

$$X \leq (n-1) + (n-2) + \cdots + 2 + 1 = \frac{n(n-1)}{2} = \binom{n}{2}.$$

Hence, $\frac{1}{2}\binom{n}{2} \leq E(X) \leq \binom{n}{2}$.

52. Ⓢ An urn contains red, green, and blue balls. Balls are chosen randomly with replacement (each time, the color is noted and then the ball is put back). Let r, g, b be the probabilities of drawing a red, green, blue ball, respectively ($r+g+b=1$).

(a) Find the expected number of balls chosen before obtaining the first red ball, not including the red ball itself.

(b) Find the expected number of different *colors* of balls obtained before getting the first red ball.

(c) Find the probability that at least 2 of n balls drawn are red, given that at least 1 is red.

Solution:

(a) The distribution is Geom(r), so the expected value is $\frac{1-r}{r}$.

(b) Use indicator random variables: let I_1 be 1 if green is obtained before red, and 0 otherwise, and define I_2 similarly for blue. Then

$$E(I_1) = P(\text{green before red}) = \frac{g}{g+r}$$

since "green before red" means that the first nonblue ball is green. Similarly, $E(I_2) = b/(b+r)$, so the expected number of colors obtained before getting red is

$$E(I_1 + I_2) = \frac{g}{g+r} + \frac{b}{b+r}.$$

(c) By definition of conditional probability,

$$P(\text{at least 2 red} | \text{ at least 1 red}) = \frac{P(\text{at least 2 red})}{P(\text{at least 1 red})} = \frac{1-(1-r)^n - nr(1-r)^{n-1}}{1-(1-r)^n}.$$

53. Ⓢ Job candidates $C_1, C_2, \ldots$ are interviewed one by one, and the interviewer compares them and keeps an updated list of rankings (if n candidates have been interviewed so far, this is a list of the n candidates, from best to worst). Assume that there is no limit on the number of candidates available, that for any n the candidates $C_1, C_2, \ldots, C_n$ are equally likely to arrive in any order, and that there are no ties in the rankings given by the interview.

Let X be the index of the first candidate to come along who ranks as better than the very first candidate C_1 (so C_X is better than C_1, but the candidates after 1 but prior to X (if any) are worse than C_1. For example, if C_2 and C_3 are worse than C_1 but C_4 is better than C_1, then $X = 4$. All 4! orderings of the first 4 candidates are equally likely, so it could have happened that the first candidate was the best out of the first 4 candidates, in which case $X > 4$.

What is $E(X)$ (which is a measure of how long, on average, the interviewer needs to wait to find someone better than the very first candidate)?

Hint: Find $P(X > n)$ by interpreting what $X > n$ says about how C_1 compares with other candidates, and then apply the result of Theorem 4.4.8.

Solution: For $n \geq 2$, $P(X > n)$ is the probability that none of $C_2, C_3, \ldots, C_n$ are better candidates than C_1, i.e., the probability that the first candidate is the highest ranked out of the first n. Since any ordering of the first n candidates is equally likely, each of the first n is equally likely to be the highest ranked of the first n, so $P(X > n) = 1/n$. For $n = 0$ or $n = 1$, $P(X > n) = 1$ (note that it does not make sense to say the probability is $1/n$ when $n = 0$). By Theorem 4.4.8,

$$E(X) = \sum_{n=0}^{\infty} P(X > n) = P(X > 0) + \sum_{n=1}^{\infty} P(X > n) = 1 + \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

since the series is the *harmonic series*, which diverges.

How can the average waiting time to find someone better than the first candidate be infinite? In the real world, there are always only finitely many candidates so the expected waiting time is finite, just as in the St. Petersburg paradox there must in reality be an upper bound on the number of rounds. The harmonic series diverges very slowly, so even with millions of job candidates the average waiting time would not be very large.

LOTUS

56. Ⓢ For $X \sim \text{Pois}(\lambda)$, find $E(X!)$ (the average factorial of X), if it is finite.

Solution: By LOTUS,

$$E(X!) = e^{-\lambda} \sum_{k=0}^{\infty} k! \frac{\lambda^k}{k!} = \frac{e^{-\lambda}}{1-\lambda},$$

for $0 < \lambda < 1$ since this is a geometric series (and $E(X!)$ is infinite if $\lambda \geq 1$).

59. Ⓢ Let $X \sim \text{Geom}(p)$ and let t be a constant. Find $E(e^{tX})$, as a function of t (this is known as the *moment generating function*; we will see in Chapter 6 how this function is useful).

Solution: Letting $q = 1 - p$, we have

$$E(e^{tX}) = p \sum_{k=0}^{\infty} e^{tk} q^k = p \sum_{k=0}^{\infty} (qe^t)^k = \frac{p}{1 - qe^t},$$

for $qe^t < 1$ (while for $qe^t \geq 1$, the series diverges).

60. Ⓢ The number of fish in a certain lake is a Pois(λ) random variable. Worried that there might be no fish at all, a statistician adds one fish to the lake. Let Y be the resulting number of fish (so Y is 1 plus a Pois(λ) random variable).

(a) Find $E(Y^2)$.

(b) Find $E(1/Y)$.

Solution:

(a) We have $Y = X + 1$ with $X \sim \text{Pois}(\lambda)$, so $Y^2 = X^2 + 2X + 1$. So

$$E(Y^2) = E(X^2 + 2X + 1) = E(X^2) + 2E(X) + 1 = (\lambda + \lambda^2) + 2\lambda + 1 = \lambda^2 + 3\lambda + 1,$$

since $E(X^2) = \text{Var}(X) + (EX)^2 = \lambda + \lambda^2$.

(b) By LOTUS,

$$E\left(\frac{1}{Y}\right) = E\left(\frac{1}{X+1}\right) = \sum_{k=0}^{\infty} \frac{1}{k+1} e^{-\lambda} \frac{\lambda^k}{k!}.$$

Using the identity $k!(k+1) = (k+1)!$ and the Taylor series for e^λ, this becomes

$$e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{(k+1)!} = \frac{e^{-\lambda}}{\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{(k+1)!} = \frac{e^{-\lambda}}{\lambda}(e^\lambda - 1) = \frac{1}{\lambda}(1 - e^{-\lambda}).$$

61. Ⓢ Let X be a Pois(λ) random variable, where λ is fixed but unknown. Let $\theta = e^{-3\lambda}$, and suppose that we are interested in estimating θ based on the data. Since X is what we observe, our estimator is a function of X, call it $g(X)$. The *bias* of the estimator $g(X)$ is defined to be $E(g(X)) - \theta$, i.e., how far off the estimate is on average; the estimator is *unbiased* if its bias is 0.

(a) For estimating λ, the r.v. X itself is an unbiased estimator. Compute the bias of the estimator $T = e^{-3X}$. Is it unbiased for estimating θ?

(b) Show that $g(X) = (-2)^X$ is an unbiased estimator for θ. (In fact, it turns out to be the only unbiased estimator for θ.)

(c) Explain intuitively why $g(X)$ is a silly choice for estimating θ, despite (b), and show how to improve it by finding an estimator $h(X)$ for θ that is always at least as good as $g(X)$ and sometimes strictly better than $g(X)$. That is,

$$|h(X) - \theta| \leq |g(X) - \theta|,$$

with the inequality sometimes strict.

Solution:

(a) The estimator is biased, with bias given by

$$E(e^{-3X}) - \theta = \sum_{k=0}^{\infty} e^{-3k} \frac{\lambda^k}{k!} e^{-\lambda} - e^{-3\lambda}$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^{-3}\lambda)^k}{k!} - e^{-3\lambda}$$

$$= e^{-\lambda} e^{e^{-3}\lambda} - e^{-3\lambda}$$

$$= e^{-3\lambda}(e^{(2+e^{-3})\lambda} - 1) \neq 0.$$

(b) The estimator $g(X) = (-2)^X$ is unbiased since

$$E(-2)^X - \theta = \sum_{k=0}^{\infty} (-2)^k \frac{\lambda^k}{k!} e^{-\lambda} - e^{-3\lambda}$$
$$= e^{-\lambda} e^{-2\lambda} - e^{-3\lambda} = 0.$$

(c) The estimator $g(X)$ is silly in the sense that it is sometimes negative, whereas $e^{-3\lambda}$ is positive. One simple way to get a better estimator is to modify $g(X)$ to make it nonnegative, by letting $h(X) = 0$ if $g(X) < 0$ and $h(X) = g(X)$ otherwise.
Better yet, note that $e^{-3\lambda}$ is between 0 and 1 since $\lambda > 0$, so letting $h(X) = 0$ if $g(X) < 0$ and $h(X) = 1$ if $g(X) > 0$ is clearly more sensible than using $g(X)$.

Poisson approximation

62. Ⓢ Law school courses often have assigned seating to facilitate the Socratic method. Suppose that there are 100 first-year law students, and each takes the same two courses: Torts and Contracts. Both are held in the same lecture hall (which has 100 seats), and the seating is uniformly random and independent for the two courses.

(a) Find the probability that no one has the same seat for both courses (exactly; you should leave your answer as a sum).

(b) Find a simple but accurate approximation to the probability that no one has the same seat for both courses.

(c) Find a simple but accurate approximation to the probability that at least two students have the same seat for both courses.

Solution:

(a) Let N be the number of students in the same seat for both classes. The problem has essentially the same structure as the matching problem. Let E_j be the event that the jth student sits in the same seat in both classes. Then

$$P(N = 0) \quad = \quad 1 - P\left(\bigcup_{j=1}^{100} E_j\right).$$

By symmetry, inclusion-exclusion gives

$$P\left(\bigcup_{j=1}^{100} E_j\right) \quad = \quad \sum_{j=1}^{100} (-1)^{j-1} \binom{100}{j} P\left(\bigcap_{k=1}^{j} E_k\right).$$

The j-fold intersection represents j particular students sitting pat throughout the two lectures, which occurs with probability $(100-j)!/100!$. So

$$P\left(\bigcup_{j=1}^{100} E_j\right) \quad = \quad \sum_{j=1}^{100} (-1)^{j-1} \binom{100}{j} \frac{(100-j)!}{100!} = \sum_{j=1}^{100} (-1)^{j-1}/i!$$
$$P(N = 0) \quad = \quad 1 - \sum_{j=1}^{100} \frac{(-1)^{j-1}}{j!} = \sum_{j=0}^{100} \frac{(-1)^j}{j!}.$$

(b) Define I_i to be the indicator for student i having the same seat in both courses, so that $N = \sum_{i=1}^{100} I_i$. Then $P(I_i = 1) = 1/100$, and the I_i are weakly dependent because

$$P((I_i = 1) \cap (I_j = 1)) = \left(\frac{1}{100}\right)\left(\frac{1}{99}\right) \quad \approx \quad \left(\frac{1}{100}\right)^2 = P(I_i = 1)P(I_j = 1).$$

So N is close to Pois(λ) in distribution, where $\lambda = E(N) = 100E(I_1) = 1$. Thus,

$$P(N = 0) \approx e^{-1}1^0/0! = e^{-1} \approx 0.37.$$

This agrees with the result of (a), which we recognize as the Taylor series for e^x, evaluated at $x = -1$.

(c) Using a Poisson approximation, we have

$$P(N \geq 2) = 1 - P(N = 0) - P(N = 1) \approx 1 - e^{-1} - e^{-1} = 1 - 2e^{-1} \approx 0.26.$$

63. Ⓢ A group of n people play "Secret Santa" as follows: each puts his or her name on a slip of paper in a hat, picks a name randomly from the hat (without replacement), and then buys a gift for that person. Unfortunately, they overlook the possibility of drawing one's own name, so some may have to buy gifts for themselves (on the bright side, some may like self-selected gifts better). Assume $n \geq 2$.

(a) Find the expected value of the number X of people who pick their own names.

(b) Find the expected number of pairs of people, A and B, such that A picks B's name and B picks A's name (where $A \neq B$ and order doesn't matter).

(c) Let X be the number of people who pick their own names. What is the *approximate* distribution of X if n is large (specify the parameter value or values)? What does $P(X = 0)$ converge to as $n \to \infty$?

Solution:

(a) Let I_j be the indicator r.v. for the jth person picking his or her own name. Then $E(I_j) = P(I_j = 1) = \frac{1}{n}$. By linearity, the expected number is $n \cdot E(I_j) = 1$.

(b) Let I_{ij} the the indicator r.v. for the ith and jth persons having such a swap (for $i < j$). Then $E(I_{ij}) = P(i \text{ picks } j)P(j \text{ picks } i | i \text{ picks } j) = \frac{1}{n(n-1)}$.

Alternatively, we can get this by counting: there are $n!$ permutations for who picks whom, of which $(n-2)!$ have i pick j and j pick i, giving $\frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$. So by linearity, the expected number is $\binom{n}{2} \cdot \frac{1}{n(n-1)} = \frac{1}{2}$.

(c) By the Poisson paradigm, X is approximately Pois(1) for large n. As $n \to \infty$, $P(X = 0) \to 1/e$, which is the probability of a Pois(1) r.v. being 0.

65. Ⓢ Ten million people enter a certain lottery. For each person, the chance of winning is one in ten million, independently.

(a) Find a simple, good approximation for the PMF of the number of people who win the lottery.

(b) Congratulations! You won the lottery. However, there may be other winners. Assume now that the number of winners other than you is $W \sim$ Pois(1), and that if there is more than one winner, then the prize is awarded to one randomly chosen winner. Given this information, find the probability that you win the prize (simplify).

Solution:

(a) Let X be the number of people who win. Then

$$E(X) = \frac{10^7}{10^7} = 1.$$

A Poisson approximation is very good here since X is the number of "successes" for a

very large number of independent trials where the probability of success on each trial is very low. So X is approximately Pois(1), and for k a nonnegative integer,

$$P(X = k) \approx \frac{1}{e \cdot k!}.$$

(b) Let A be the event that you win the prize, and condition on W:

$$P(A) = \sum_{k=0}^{\infty} P(A|W = k)P(W = k) = \frac{1}{e}\sum_{k=0}^{\infty} \frac{1}{k+1}\frac{1}{k!} = \frac{1}{e}\sum_{k=0}^{\infty} \frac{1}{(k+1)!} = \frac{e-1}{e} = 1 - \frac{1}{e}.$$

*Existence

68. Ⓢ Each of 111 people names his or her 5 favorite movies out of a list of 11 movies.

(a) Alice and Bob are 2 of the 111 people. Assume *for this part only* that Alice's 5 favorite movies out of the 11 are random, with all sets of 5 equally likely, and likewise for Bob, independently. Find the expected number of movies in common to Alice's and Bob's lists of favorite movies.

(b) Show that there are 2 movies such that at least 21 of the people name both of these movies as favorites.

Solution:

(a) Let I_j be the indicator for the jth movie being on both lists, for $1 \leq j \leq 11$. By symmetry and linearity, the desired expected value is

$$11\left(\frac{5}{11}\right)^2 = \frac{25}{11}.$$

(b) Choose 2 *random* movies (one at a time, without replacement). Let X be the number of people who name both of these movies. Creating an indicator r.v. for each person,

$$E(X) = 111P(\text{Alice names both random movies}) = 111\left(\frac{5}{11}\cdot\frac{4}{10}\right) = \left(\frac{111}{110}\right)20 > 20,$$

since the first chosen movie has a 5/11 chance of being on Alice's list and given that it is, the second chosen movie has a 4/10 chance of being on the list (or we can use the Hypergeometric PMF after "tagging" Alice's favorite movies).

Thus, there must exist 2 movies such that at least 21 of the people name both of them as favorites.

69. Ⓢ The circumference of a circle is colored with red and blue ink such that 2/3 of the circumference is red and 1/3 is blue. Prove that no matter how complicated the coloring scheme is, there is a way to inscribe a square in the circle such that at least three of the four corners of the square touch red ink.

Solution: Consider a random square, obtained by picking a uniformly random point on the circumference and inscribing a square with that point a corner; say that the corners are $U_1, \ldots, U_4$, in clockwise order starting with the initial point chosen. Let I_j be the indicator r.v. of U_j touching red ink. By symmetry, $E(I_j) = 2/3$ so by linearity, the expected number of corners touching red ink is 8/3. Thus, there must exist an inscribed square with at least 8/3 of its corners touching red ink. Such a square must have at least 3 of its corners touching red ink.

70. Ⓢ A hundred students have taken an exam consisting of 8 problems, and for each problem at least 65 of the students got the right answer. Show that there exist two students who collectively got everything right, in the sense that for each problem, at least one of the two got it right.

Solution: Say that the "score" of a pair of students is how many problems at least one of them got right. The expected score of a random pair of students (with all pairs equally likely) is at least $8(1-0.35^2) = 7.02$, as seen by creating an indicator r.v. for each problem for the event that at least one student in the pair got it right. (We can also improve the 0.35^2 to $\frac{35}{100} \cdot \frac{34}{99}$ since the students are sampled without replacement.) So some pair of students must have gotten a score of at least 7.02, which means that they got a score of at least 8.

71. Ⓢ Ten points in the plane are designated. You have ten circular coins (of the same radius). Show that you can position the coins in the plane (without stacking them) so that all ten points are covered.

Hint: Consider a *honeycomb tiling* of the plane (this is a way to divide the plane into hexagons). You can use the fact from geometry that if a circle is inscribed in a hexagon then the ratio of the area of the circle to the area of the hexagon is $\frac{\pi}{2\sqrt{3}} > 0.9$.

Solution: Take a uniformly random honeycomb tiling (to do this, start with any honeycomb tiling and then shift it horizontally and vertically by uniformly random amounts; by periodicity there is an upper bound on how large the shifts need to be). Choose the tiling so that a circle the same size as one of the coins can be inscribed in each hexagon. Then inscribe a circle in each hexagon, and let I_j be the indicator r.v. for the jth point being contained inside one the circles. We have $E(I_j) > 0.9$ by the geometric fact mentioned above, so by linearity $E(I_1 + \cdots + I_{10}) > 9$. Thus, there is a positioning of the honeycomb tiling such that all 10 points are contained inside the circles. Putting coins on top of the circles containing the points, we can cover all ten points.

72. Ⓢ Let S be a set of binary strings $a_1 \dots a_n$ of length n (where juxtaposition means concatenation). We call S *k-complete* if for any indices $1 \leq i_1 < \cdots < i_k \leq n$ and any binary string $b_1 \dots b_k$ of length k, there is a string $s_1 \dots s_n$ in S such that $s_{i_1} s_{i_2} \dots s_{i_k} = b_1 b_2 \dots b_k$. For example, for $n = 3$, the set $S = \{001, 010, 011, 100, 101, 110\}$ is 2-complete since all 4 patterns of 0's and 1's of length 2 can be found in any 2 positions. Show that if $\binom{n}{k} 2^k (1-2^{-k})^m < 1$, then there exists a k-complete set of size at most m.

Solution: Generate m random strings of length n independently, using fair coin flips to determine each bit. Let S be the resulting random set of strings. If we can show that the probability that S is k-complete is *positive*, then we know that a k-complete set of size at most m must *exist*. Let A be the event that S is k-complete. Let $N = \binom{n}{k} 2^k$ and let $A_1, \dots, A_N$ be the events of the form "S contains a string which is $b_1 \dots b_k$ at coordinates $i_1 < \cdots < i_k$," in any fixed order. For example, if $k = 3$ then A_1 could be the event "S has an element which is 110 at positions $1, 2, 3$." Then $P(A) > 0$ since

$$P(A^c) = P(\cup_{j=1}^N A_j^c) \leq \sum_{j=1}^N P(A_j^c) = N(1-2^{-k})^m < 1.$$

Mixed practice

75. Ⓢ A group of 360 people is going to be split into 120 teams of 3 (where the order of teams and the order within a team don't matter).

(a) How many ways are there to do this?

(b) The group consists of 180 married couples. A random split into teams of 3 is chosen,

with all possible splits equally likely. Find the expected number of teams containing married couples.

Solution:

(a) Imagine lining the people up and saying the first 3 are a team, the next 3 are a team, etc. This overcounts by a factor of $(3!)^{120} \cdot 120!$ since the order within teams and the order of teams don't matter. So the number of ways is

$$\frac{360!}{6^{120} \cdot 120!}.$$

(b) Let I_j be the indicator for the jth team having a married couple (taking the teams to be chosen one at a time, or with respect to a random ordering). By symmetry and linearity, the desired quantity is $120E(I_1)$. We have

$$E(I_1) = P(\text{first team has a married couple}) = \frac{180 \cdot 358}{\binom{360}{3}},$$

since the first team is equally likely to be any 3 of the people, and to have a married couple on the team we need to choose a couple and then any third person. So the expected value is

$$\frac{120 \cdot 180 \cdot 358}{\binom{360}{3}}.$$

(This simplifies to $\frac{120 \cdot 180 \cdot 358}{360 \cdot 359 \cdot 358/6} = \frac{360}{359}$. Another way to find the probability that the first team has a married couple is to note that any particular pair in the team has probability $\frac{1}{359}$ of being married to each other, so since there are 3 disjoint possibilities the probability is $\frac{3}{359}$.)

76. Ⓢ The gambler de Méré asked Pascal whether it is more likely to get at least one six in 4 rolls of a die, or to get at least one double-six in 24 rolls of a pair of dice. Continuing this pattern, suppose that a group of n fair dice is rolled $4 \cdot 6^{n-1}$ times.

(a) Find the expected number of times that "all sixes" is achieved (i.e., how often among the $4 \cdot 6^{n-1}$ rolls it happens that all n dice land 6 simultaneously).

(b) Give a simple but accurate approximation of the probability of having at least one occurrence of "all sixes", for n large (in terms of e but not n).

(c) de Méré finds it tedious to re-roll so many dice. So after one normal roll of the n dice, in going from one roll to the next, with probability 6/7 he leaves the dice in the same configuration and with probability 1/7 he re-rolls. For example, if $n = 3$ and the 7th roll is $(3, 1, 4)$, then 6/7 of the time the 8th roll remains $(3, 1, 4)$ and 1/7 of the time the 8th roll is a new random outcome. Does the expected number of times that "all sixes" is achieved stay the same, increase, or decrease (compared with (a))? Give a short but clear explanation.

Solution:

(a) Let I_j be the indicator r.v. for the event "all sixes" on the jth roll. Then $E(I_j) = 1/6^n$, so the expected value is $4 \cdot 6^{n-1}/6^n = 2/3$.

(b) By a Poisson approximation with $\lambda = 2/3$ (the expected value from (a)), the probability is approximately $1 - e^{-2/3}$.

(c) The answer stays the same, by the same reasoning as in (a), since linearity of expectation holds even for dependent r.v.s.

77. Ⓢ Five people have just won a \$100 prize, and are deciding how to divide the \$100 up between them. Assume that whole dollars are used, not cents. Also, for example, giving \$50 to the first person and \$10 to the second is different from vice versa.

(a) How many ways are there to divide up the \$100, such that each gets at least \$10?

(b) Assume that the \$100 is randomly divided up, with all of the possible allocations counted in (a) equally likely. Find the expected amount of money that the first person receives.

(c) Let A_j be the event that the jth person receives more than the first person (for $2 \leq j \leq 5$), when the \$100 is randomly allocated as in (b). Are A_2 and A_3 independent?

Solution:

(a) Give each person \$10, and then distribute the remaining \$50 arbitrarily. By Bose-Einstein (thinking of people as boxes and dollars as balls!), the number of ways is

$$\binom{5+50-1}{50} = \binom{54}{50} = \binom{54}{4}.$$

(b) Let X_j be the amount that j gets. By symmetry, $E(X_j)$ is the same for all j. But $X_1 + \cdots + X_5 = 100$, so by linearity $100 = 5EX_1$. Thus, EX_1 is \$20.

(c) The events A_2 and A_3 are not independent since knowing that A_2 occurred makes it more likely that person 1 received a low percentage of the money, which in turn makes it more likely that A_3 occurred.

78. Ⓢ Joe's iPod has 500 different songs, consisting of 50 albums of 10 songs each. He listens to 11 random songs on his iPod, with all songs equally likely and chosen independently (so repetitions may occur).

(a) What is the PMF of how many of the 11 songs are from his favorite album?

(b) What is the probability that there are 2 (or more) songs from the same album among the 11 songs he listens to?

(c) A pair of songs is a *match* if they are from the same album. If, say, the 1st, 3rd, and 7th songs are all from the same album, this counts as 3 matches. Among the 11 songs he listens to, how many matches are there on average?

Solution:

(a) The distribution is $\text{Bin}(n, p)$ with $n = 11, p = \frac{1}{50}$ (thinking of getting a song from the favorite album as a "success"). So the PMF is

$$\binom{11}{k}\left(\frac{1}{50}\right)^k\left(\frac{49}{50}\right)^{11-k}, \text{ for } 0 \leq k \leq 11.$$

(b) This is a version of the birthday problem. We have

$$P(\text{at least 1 match}) = 1 - P(\text{no matches}) = 1 - \frac{50 \cdot 49 \cdot \cdots \cdot 40}{50^{11}} = 1 - \frac{49!}{39! \cdot 50^{10}}.$$

(c) Defining an indicator r.v. I_{jk} for the event that the jth and kth songs match, we have $E(I_{jk}) = P(I_{jk} = 1) = 1/50$, so the expected number of matches is

$$\binom{11}{2}\frac{1}{50} = \frac{11 \cdot 10}{2 \cdot 50} = \frac{110}{100} = 1.1.$$

79. Ⓢ In each day that the Mass Cash lottery is run in Massachusetts, 5 of the integers from 1 to 35 are chosen (randomly and without replacement).

(a) When playing this lottery, find the probability of guessing exactly 3 numbers right, given that you guess at least 1 of the numbers right.

(b) Find an exact expression for the expected number of days needed so that all of the $\binom{35}{5}$ possible lottery outcomes will have occurred.

(c) Approximate the probability that after 50 days of the lottery, every number from 1 to 35 has been picked at least once.

Solution:

(a) The distribution is Hypergeometric (think of capture-recapture, "tagging" the numbers you choose). So

$$\begin{aligned} P(\text{exactly 3 right}|\text{at least 1 right}) &= \frac{P(\text{exactly 3 right})}{1-P(\text{none right})} \\ &= \frac{\binom{5}{3}\binom{30}{2}/\binom{35}{5}}{1-\binom{5}{0}\binom{30}{5}/\binom{35}{5}}. \end{aligned}$$

(b) Let $n = \binom{35}{5}$. By the coupon collector problem (or directly by linearity, writing the expected number of days as a sum of T_j's with $T_j - 1$ a Geometric), the expected value is

$$n\left(\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{2} + 1\right).$$

(c) Let A_j be the event that j doesn't get picked, so

$$P(A_j) = (30/35)^{50} = (6/7)^{50}.$$

Let X be the number of A_j that occur. A Poisson approximation for X is reasonable since these events are rare and weakly dependent. This gives

$$P(X = 0) \approx e^{-35\cdot(6/7)^{50}} \approx 0.98.$$

Chapter 5: Continuous random variables

Uniform and universality of the Uniform

11. Ⓢ Let U be a Uniform r.v. on the interval $(-1, 1)$ (be careful about minus signs).

(a) Compute $E(U)$, $\text{Var}(U)$, and $E(U^4)$.

(b) Find the CDF and PDF of U^2. Is the distribution of U^2 Uniform on $(0, 1)$?

Solution:

(a) We have $E(U) = 0$ since the distribution is symmetric about 0. By LOTUS,

$$E(U^2) = \frac{1}{2}\int_{-1}^{1} u^2 du = \frac{1}{3}.$$

So $\text{Var}(U) = E(U^2) - (EU)^2 = E(U^2) = \frac{1}{3}$. Again by LOTUS,

$$E(U^4) = \frac{1}{2}\int_{-1}^{1} u^4 du = \frac{1}{5}.$$

(b) Let $G(t)$ be the CDF of U^2. Clearly $G(t) = 0$ for $t \leq 0$ and $G(t) = 1$ for $t \geq 1$, because $0 \leq U^2 \leq 1$. For $0 < t < 1$,

$$G(t) = P(U^2 \leq t) = P(-\sqrt{t} \leq U \leq \sqrt{t}) = \sqrt{t},$$

since the probability of U being in an interval in $(-1, 1)$ is proportional to its length. The PDF is $G'(t) = \frac{1}{2}t^{-1/2}$ for $0 < t < 1$ (and 0 otherwise). The distribution of U^2 is *not* Uniform on $(0, 1)$ as the PDF is not a constant on this interval (it is an example of a *Beta distribution*, an important distribution that is introduced in Chapter 8).

12. Ⓢ A stick is broken into two pieces, at a uniformly random break point. Find the CDF and average of the length of the longer piece.

Solution: We can assume the units are chosen so that the stick has length 1. Let L be the length of the longer piece, and let the break point be $U \sim \text{Unif}(0, 1)$. For any $l \in [1/2, 1]$, observe that $L < l$ is equivalent to $\{U < l \text{ and } 1 - U < l\}$, which can be written as $1 - l < U < l$. We can thus obtain L's CDF as

$$F_L(l) = P(L < l) = P(1 - l < U < l) = 2l - 1,$$

so $L \sim \text{Unif}(1/2, 1)$. In particular, $E(L) = 3/4$.

16. Ⓢ Let $U \sim \text{Unif}(0, 1)$, and

$$X = \log\left(\frac{U}{1-U}\right).$$

Then X has the Logistic distribution, as defined in Example 5.1.6.

(a) Write down (but do not compute) an integral giving $E(X^2)$.

(b) Find $E(X)$ without using calculus.

Hint: A useful symmetry property here is that $1 - U$ has the same distribution as U.

Solution:

(a) By LOTUS,

$$E(X^2) = \int_0^1 \left(\log\left(\frac{u}{1-u}\right)\right)^2 du.$$

(b) By the symmetry property mentioned in the hint, $1-U$ has the same distribution as U. So by linearity,

$$E(X) = E(\log U - \log(1-U)) = E(\log U) - E(\log(1-U)) = 0.$$

19. Ⓢ Let F be a CDF which is continuous and strictly increasing. Let μ be the mean of the distribution. The quantile function, F^{-1}, has many applications in statistics and econometrics. Show that the area under the curve of the quantile function from 0 to 1 is μ.

Hint: Use LOTUS and universality of the Uniform.

Solution: We want to find $\int_0^1 F^{-1}(u)du$. Let $U \sim \text{Unif}(0,1)$ and $X = F^{-1}(U)$. By universality of the Uniform, $X \sim F$. By LOTUS,

$$\int_0^1 F^{-1}(u)du = E(F^{-1}(U)) = E(X) = \mu.$$

Equivalently, make the substitution $u = F(x)$, so $du = f(x)dx$, where f is the PDF of the distribution with CDF F. Then the integral becomes

$$\int_{-\infty}^{\infty} F^{-1}(F(x))f(x)dx = \int_{-\infty}^{\infty} xf(x)dx = \mu.$$

Sanity check: For the simple case that F is the Unif$(0,1)$ CDF, which is $F(u) = u$ on $(0,1)$, we have $\int_0^1 F^{-1}(u)du = \int_0^1 udu = 1/2$, which is the mean of a Unif$(0,1)$.

Normal

32. Ⓢ Let $Z \sim \mathcal{N}(0,1)$ and let S be a random sign independent of Z, i.e., S is 1 with probability $1/2$ and -1 with probability $1/2$. Show that $SZ \sim \mathcal{N}(0,1)$.

Solution: Condition on S to find the CDF of SZ:

$$\begin{aligned} P(SZ \le x) &= P(SZ \le x|S=1)\frac{1}{2} + P(SZ \le x|S=-1)\frac{1}{2} \\ &= P(Z \le x)\frac{1}{2} + P(Z \ge -x)\frac{1}{2} \\ &= P(Z \le x)\frac{1}{2} + P(Z \le x)\frac{1}{2} \\ &= \Phi(x), \end{aligned}$$

where the penultimate equality is by symmetry of the Normal.

33. Ⓢ Let $Z \sim \mathcal{N}(0,1)$. Find $E(\Phi(Z))$ *without* using LOTUS, where Φ is the CDF of Z.

Solution: By universality of the Uniform, $F(X) \sim \text{Unif}(0,1)$ for any continuous random variable X with CDF F. Therefore, $E(\Phi(Z)) = 1/2$.

34. Ⓢ Let $Z \sim \mathcal{N}(0,1)$ and $X = Z^2$. Then the distribution of X is called *Chi-Square with 1 degree of freedom*. This distribution appears in many statistical methods.

(a) Find a good numerical approximation to $P(1 \leq X \leq 4)$ using facts about the Normal distribution, without querying a calculator/computer/table about values of the Normal CDF.

(b) Let Φ and φ be the CDF and PDF of Z, respectively. Show that for any $t > 0$, $I(Z > t) \leq (Z/t)I(Z > t)$. Using this and LOTUS, show that $\Phi(t) \geq 1 - \varphi(t)/t$.

Solution:

(a) By symmetry of the Normal,

$$P(1 \leq Z^2 \leq 4) = P(-2 \leq Z \leq -1 \text{ or } 1 \leq Z \leq 2) = 2P(1 \leq Z \leq 2) = 2(\Phi(2) - \Phi(1)).$$

By the 68-95-99.7% Rule, $P(-1 \leq Z \leq 1) \approx 0.68$. This says that 32% of the area under the Normal curve is outside of $[-1, 1]$, which by symmetry says that 16% is in $(1, \infty)$. So $\Phi(1) \approx 1 - 0.16 = 0.84$. Similarly, $P(-2 \leq Z \leq 2) \approx 0.95$ gives $\Phi(2) \approx 0.975$. In general, symmetry of the Normal implies that for any $t > 0$,

$$P(-t \leq Z \leq t) = \Phi(t) - \Phi(-t) = 2\Phi(t) - 1.$$

(b) The inequality $I(Z > t) \leq (Z/t)I(Z > t)$ is true since if the indicator is 0 then both sides are 0, and if it is 1 then $Z/t > 1$. So

$$E(I(Z > t)) \leq \frac{1}{t}E(ZI(Z > t)) = \frac{1}{t}\int_{-\infty}^{\infty} zI(z > t)\varphi(z)dz = \frac{1}{t}\int_{t}^{\infty} z\varphi(z)dz.$$

The integral can be done using a substitution: letting $u = z^2/2$, we have

$$\int ze^{-z^2/2}dz = \int e^{-u}du = -e^{-u} + C = -e^{-z^2/2} + C.$$

Thus,

$$P(Z > t) = E(I(Z > t)) \leq \varphi(t)/t,$$

which proves the desired bound on $\Phi(t)$.

36. Ⓢ Let $Z \sim \mathcal{N}(0,1)$. A measuring device is used to observe Z, but the device can only handle positive values, and gives a reading of 0 if $Z \leq 0$; this is an example of *censored data*. So assume that $X = ZI_{Z>0}$ is observed rather than Z, where $I_{Z>0}$ is the indicator of $Z > 0$. Find $E(X)$ and $\text{Var}(X)$.

Solution: By LOTUS,

$$E(X) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} I_{z>0}ze^{-z^2/2}dz = \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} ze^{-z^2/2}dz.$$

Letting $u = z^2/2$, we have

$$E(X) = \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} e^{-u}du = \frac{1}{\sqrt{2\pi}}.$$

To obtain the variance, note that

$$E(X^2) = \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} z^2e^{-z^2/2}dz = \frac{1}{2}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^2e^{-z^2/2}dz = \frac{1}{2},$$

since a $\mathcal{N}(0,1)$ r.v. has variance 1. Thus,

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

Note that X is neither purely discrete nor purely continuous, since $X = 0$ with probability $1/2$ and $P(X = x) = 0$ for $x \neq 0$. So X has neither a PDF nor a PMF; but LOTUS still works, allowing us to work with the PDF of Z to study expected values of functions of Z.

Sanity check: The variance is positive, as it should be. It also makes sense that the variance is substantially less than 1 (which is the variance of Z), since we are reducing variability by making the r.v. 0 half the time, and making it nonnegative rather than roaming over the entire real line.

Exponential

38. Ⓢ A post office has 2 clerks. Alice enters the post office while 2 other customers, Bob and Claire, are being served by the 2 clerks. She is next in line. Assume that the time a clerk spends serving a customer has the Exponential(λ) distribution.

(a) What is the probability that Alice is the last of the 3 customers to be done being served?

Hint: No integrals are needed.

(b) What is the expected total time that Alice needs to spend at the post office?

Solution:

(a) Alice begins to be served when either Bob or Claire leaves. By the memoryless property, the additional time needed to serve whichever of Bob or Claire is still there is Expo(λ). The time it takes to serve Alice is also Expo(λ), so by symmetry the probability is $1/2$ that Alice is the last to be done being served.

(b) The expected time spent waiting in line is $\frac{1}{2\lambda}$, since the minimum of two independent Expo(λ) r.v.s is Expo(2λ) (by Example 5.6.3). The expected time spent being served is $\frac{1}{\lambda}$. So the expected total time is

$$\frac{1}{2\lambda} + \frac{1}{\lambda} = \frac{3}{2\lambda}.$$

41. Ⓢ Fred wants to sell his car, after moving back to Blissville (where he is happy with the bus system). He decides to sell it to the first person to offer at least \$15,000 for it. Assume that the offers are independent Exponential random variables with mean \$10,000.

(a) Find the expected number of offers Fred will have.

(b) Find the expected amount of money that Fred will get for the car.

Solution:

(a) The offers on the car are i.i.d. $X_i \sim \text{Expo}(1/10^4)$. So the number of offers that are too low is Geom(p) with $p = P(X_i \geq 15000) = e^{-1.5}$. Including the successful offer, the expected number of offers is thus $(1-p)/p + 1 = 1/p = e^{1.5}$.

(b) Let N be the number of offers, so the sale price of the car is X_N. Note that

$$E(X_N) = E(X|X \geq 12000)$$

for $X \sim \text{Expo}(1/10^4)$, since the successful offer is an Exponential for which our information is that the value is at least \$15,000. To compute this, remember the memoryless

property! For any $a > 0$, if $X \sim \text{Expo}(\lambda)$ then the distribution of $X - a$ given $X > a$ is itself $\text{Expo}(\lambda)$. So

$$E(X|X \geq 15000) = 15000 + E(X) = 25000,$$

which shows that Fred's expected sale price is \$25,000.

44. Ⓢ Joe is waiting in continuous time for a book called *The Winds of Winter* to be released. Suppose that the waiting time T until news of the book's release is posted, measured in years relative to some starting point, has an Exponential distribution with $\lambda = 1/5$.

Joe is not so obsessive as to check multiple times a day; instead, he checks the website *once* at the end of each day. Therefore, he observes the day on which the news was posted, rather than the exact time T. Let X be this measurement, where $X = 0$ means that the news was posted within the first day (after the starting point), $X = 1$ means it was posted on the second day, etc. (assume that there are 365 days in a year). Find the PMF of X. Is this a named distribution that we have studied?

Solution: The event $X = k$ is the same as the event $k \leq 365T < k+1$, i.e., $X = \lfloor 365T \rfloor$, where $\lfloor t \rfloor$ is the floor function of t (the greatest integer less than or equal to t). The CDF of T is $F_T(t) = 1 - e^{-t/5}$ for $t > 0$ (and 0 for $t \leq 0$). So

$$P(X = k) = P\left(\frac{k}{365} \leq T < \frac{k+1}{365}\right) = F_T\left(\frac{k+1}{365}\right) - F_T\left(\frac{k}{365}\right) = e^{-k/1825} - e^{-(k+1)/1825}.$$

This factors as $\left(e^{-1/1825}\right)^k (1 - e^{-1/1825})$, which shows that $X \sim \text{Geom}(1 - e^{-1/1825})$.

Sanity check: A Geometric distribution is plausible for a waiting time, and does take values $0, 1, 2, \ldots$. The parameter $p = 1 - e^{-1/1825} \approx 0.0005$ is very small, which reflects both the fact that there are a lot of days in a year (so each day is unlikely) and the fact that the author is not known for the celerity of his writing.

50. Ⓢ Find $E(X^3)$ for $X \sim \text{Expo}(\lambda)$, using LOTUS and the fact that $E(X) = 1/\lambda$ and $\text{Var}(X) = 1/\lambda^2$, and integration by parts at most once. In the next chapter, we'll learn how to find $E(X^n)$ for all n.

Solution: By LOTUS,

$$E(X^3) = \int_0^\infty x^3 \lambda e^{-\lambda x} dx = -x^3 e^{-\lambda x}\Big|_0^\infty + \frac{3}{\lambda}\int_0^\infty x^2 \lambda e^{-\lambda x} dx$$

$$= \frac{3}{\lambda}E(X^2) = \frac{3}{\lambda}(Var(X) + (EX)^2) = \frac{6}{\lambda^3},$$

where the second equality uses integration by parts, letting $u = x^3$ and $dv = \lambda e^{-\lambda x} dx$ and we multiply the second term by 1 written as λ/λ.

51. Ⓢ The *Gumbel distribution* is the distribution of $-\log X$ with $X \sim \text{Expo}(1)$.

(a) Find the CDF of the Gumbel distribution.

(b) Let $X_1, X_2, \ldots$ be i.i.d. Expo(1) and let $M_n = \max(X_1, \ldots, X_n)$. Show that $M_n - \log n$ converges in distribution to the Gumbel distribution, i.e., as $n \to \infty$ the CDF of $M_n - \log n$ converges to the Gumbel CDF.

Solution:

(a) Let G be Gumbel and $X \sim \text{Expo}(1)$. The CDF is

$$P(G \leq t) = P(-\log X \leq t) = P(X \geq e^{-t}) = e^{-e^{-t}}$$

for all real t.

(b) The CDF of $M_n - \log n$ is

$$P(M_n - \log n \leq t) = P(X_1 \leq t + \log n, \ldots, X_n \leq t + \log n) = P(X_1 \leq t + \log n)^n.$$

Using the Expo CDF and the fact that $(1 + \frac{x}{n})^n \to e^x$ as $n \to \infty$, this becomes

$$(1 - e^{-(t+\log n)})^n = (1 - \frac{e^{-t}}{n})^n \to e^{-e^{-t}}.$$

Mixed practice

55. Ⓢ Consider an experiment where we observe the value of a random variable X, and estimate the value of an unknown constant θ using some random variable $T = g(X)$ that is a function of X. The r.v. T is called an *estimator*. Think of X as the data observed in the experiment, and θ as an unknown parameter related to the distribution of X.

For example, consider the experiment of flipping a coin n times, where the coin has an unknown probability θ of Heads. After the experiment is performed, we have observed the value of $X \sim \text{Bin}(n, \theta)$. The most natural estimator for θ is then X/n.

The *bias* of an estimator T for θ is defined as $b(T) = E(T) - \theta$. The *mean squared error* is the average squared error when using $T(X)$ to estimate θ:

$$\text{MSE}(T) = E(T - \theta)^2.$$

Show that

$$\text{MSE}(T) = \text{Var}(T) + (b(T))^2.$$

This implies that for fixed MSE, lower bias can only be attained at the cost of higher variance and vice versa; this is a form of the *bias-variance tradeoff*, a phenomenon which arises throughout statistics.

Solution: Using the fact that adding a constant does not affect variance, we have

$$\begin{aligned} \text{Var}(T) &= \text{Var}(T - \theta) \\ &= E(T - \theta)^2 - (E(T - \theta))^2 \\ &= \text{MSE}(T) - (b(T))^2, \end{aligned}$$

which proves the desired identity.

56. Ⓢ (a) Suppose that we have a list of the populations of every country in the world.

Guess, without looking at data yet, what percentage of the populations have the digit 1 as their first digit (e.g., a country with a population of 1,234,567 has first digit 1 and a country with population 89,012,345 does not).

(b) After having done (a), look through a list of populations and count how many start with a 1. What percentage of countries is this? *Benford's law* states that in a very large variety of real-life data sets, the first digit approximately follows a particular distribution with about a 30% chance of a 1, an 18% chance of a 2, and in general

$$P(D = j) = \log_{10}\left(\frac{j+1}{j}\right), \text{ for } j \in \{1, 2, 3, \ldots, 9\},$$

where D is the first digit of a randomly chosen element. (Exercise from Chapter 3 asks for a proof that this is a valid PMF.) How closely does the percentage found in the data agree with that predicted by Benford's law?

(c) Suppose that we write the random value in some problem (e.g., the population of a random country) in scientific notation as $X \times 10^N$, where N is a nonnegative integer and $1 \leq X < 10$. Assume that X is a continuous r.v. with PDF

$$f(x) = c/x, \text{ for } 1 \leq x \leq 10,$$

and 0 otherwise, with c a constant. What is the value of c (be careful with the bases of logs)? Intuitively, we might hope that the distribution of X does not depend on the choice of units in which X is measured. To see whether this holds, let $Y = aX$ with $a > 0$. What is the PDF of Y (specifying where it is nonzero)?

(d) Show that if we have a random number $X \times 10^N$ (written in scientific notation) and X has the PDF f from (c), then the first digit (which is also the first digit of X) has Benford's law as PMF.

Hint: What does $D = j$ correspond to in terms of the values of X?

Solution:

(a) What did you guess?

(b) According to Wikipedia (as of October 15, 2011), 63 out of 225 countries have a total population whose first digit is 1, which is 28%. (This depends slightly on whether certain territories included in the Wikipedia list should be considered as "countries" but the purpose of this problem is not to delve into the sovereignty or nation-status of territories). It is striking that 28% of the countries have first digit 1, as this is so much higher than one would expect from guessing that the first digit is equally likely to be any of $1, 2, \ldots, 9$. This is an example of Benford's law; similar phenomena have been observed in many different settings (such as with lengths of rivers, physical constants, and stock prices).

(c) The PDF ($f(x) = c/x, 1 \leq x \leq 10$) must integrate to one, by definition; therefore

$$1 = c\int_1^{10} \frac{dx}{x} = c(\ln 10 - \ln 1) = c \ln 10.$$

So the constant of proportionality $c = 1/\ln 10 = \log_{10} e$). If $Y = aX$ (a *change in scale*), then Y has pdf c/y with the same value of c as before, except now $a \leq y \leq 10a$ rather than $1 \leq x \leq 10$. So the PDF takes the same form for aX as for X, but over a different range.

(d) The first digit $D = d$ when $d \leq X < d + 1$. The probability of this is

$$P(D = d) = P(d \leq X < d + 1) = \int_d^{d+1} \frac{1}{x \ln 10} dx,$$

which is $\log_{10}(d + 1) - \log_{10}(d)$, as desired.

57. Ⓢ (a) Let $X_1, X_2, \ldots$ be independent $\mathcal{N}(0, 4)$ r.v.s., and let J be the smallest value of j such that $X_j > 4$ (i.e., the index of the first X_j exceeding 4). In terms of Φ, find $E(J)$.

(b) Let f and g be PDFs with $f(x) > 0$ and $g(x) > 0$ for all x. Let X be a random variable with PDF f. Find the expected value of the ratio

$$R = \frac{g(X)}{f(X)}.$$

Such ratios come up very often in statistics, when working with a quantity known as a *likelihood ratio* and when using a computational technique known as *importance sampling*.

(c) Define

$$F(x) = e^{-e^{-x}}.$$

This is a CDF and is a continuous, strictly increasing function. Let X have CDF F, and define $W = F(X)$. What are the mean and variance of W?

Solution:

(a) We have $J - 1 \sim \text{Geom}(p)$ with $p = P(X_1 > 4) = P(X_1/2 > 2) = 1 - \Phi(2)$, so $E(J) = 1/(1 - \Phi(2))$.

(b) By LOTUS,

$$E\frac{g(X)}{f(X)} = \int_{-\infty}^{\infty} \frac{g(x)}{f(x)} f(x)dx = \int_{-\infty}^{\infty} g(x)dx = 1.$$

(c) By universality of the Uniform, $W \sim \text{Unif}(0, 1)$. Alternatively, we can compute directly that the CDF of W is

$$P(W \leq w) = P(F(X) \leq w) = P(X \leq F^{-1}(w)) = F(F^{-1}(w)) = w$$

for $0 < w < 1$, so again we have $W \sim \text{Unif}(0, 1)$. Thus, $E(W) = 1/2$ and $\text{Var}(W) = 1/12$.

59. Ⓢ As in Example 5.7.3, athletes compete one at a time at the high jump. Let X_j be how high the jth jumper jumped, with $X_1, X_2, \ldots$ i.i.d. with a continuous distribution. We say that the jth jumper is "best in recent memory" if he or she jumps higher than the previous 2 jumpers (for $j \geq 3$; the first 2 jumpers don't qualify).

(a) Find the expected number of best in recent memory jumpers among the 3rd through nth jumpers.

(b) Let A_j be the event that the jth jumper is the best in recent memory. Find $P(A_3 \cap A_4), P(A_3)$, and $P(A_4)$. Are A_3 and A_4 independent?

Solution:

(a) Let I_j be the indicator of the jth jumper being best in recent memory, for each $j \geq 3$. By symmetry, $E(I_j) = 1/3$ (similarly to examples from class and the homework). By linearity, the desired expected value is $(n - 2)/3$.

(b) The event $A_3 \cap A_4$ occurs if and only if the ranks of the first 4 jumps are $4, 3, 2, 1$ or $3, 4, 2, 1$ (where 1 denotes the best of the first 4 jumps, etc.). Since all orderings are equally likely,

$$P(A_3 \cap A_4) = \frac{2}{4!} = \frac{1}{12}.$$

As in (a), we have $P(A_3) = P(A_4) = 1/3$. So $P(A_3 \cap A_4) \neq P(A_3)P(A_4)$, which shows that A_3 and A_4 are not independent.

Chapter 6: Moments

Moment generating functions

13. Ⓢ A fair die is rolled twice, with outcomes X for the first roll and Y for the second roll. Find the moment generating function $M_{X+Y}(t)$ of $X+Y$ (your answer should be a function of t and can contain unsimplified finite sums).

Solution: Since X and Y are i.i.d., LOTUS gives

$$M_{X+Y}(t) = E(e^{t(X+Y)}) = E(e^{tX})E(e^{tY}) = \left(\frac{1}{6}\sum_{k=1}^{6} e^{kt}\right)^2.$$

14. Ⓢ Let $U_1, U_2, \ldots, U_{60}$ be i.i.d. Unif$(0,1)$ and $X = U_1 + U_2 + \cdots + U_{60}$. Find the MGF of X.

Solution: The MGF of U_1 is $E(e^{tU_1}) = \int_0^1 e^{tu}du = \frac{1}{t}(e^t - 1)$ for $t \neq 0$ (and the value is 1 for $t = 0$). Thus, the MGF of X is

$$E(e^{tX}) = E(e^{t(U_1+\cdots+U_{60})}) = \left(E(e^{tU_1})\right)^{60} = \frac{(e^t-1)^{60}}{t^{60}},$$

for $t \neq 0$ (and the value is 1 for $t = 0$).

20. Ⓢ Let $X \sim$ Pois(λ), and let $M(t)$ be the MGF of X. The *cumulant generating function* is defined to be $g(t) = \log M(t)$. Expanding $g(t)$ as a Taylor series

$$g(t) = \sum_{j=1}^{\infty} \frac{c_j}{j!}t^j$$

(the sum starts at $j = 1$ because $g(0) = 0$), the coefficient c_j is called the jth *cumulant* of X. Find the jth cumulant of X, for all $j \geq 1$.

Solution: Using the Taylor series for e^t,

$$g(t) = \lambda(e^t - 1) = \sum_{j=1}^{\infty} \lambda\frac{t^j}{j!},$$

so $c_j = \lambda$ for all $j \geq 1$.

21. Ⓢ Let $X_n \sim$ Bin(n, p_n) for all $n \geq 1$, where np_n is a constant $\lambda > 0$ for all n (so $p_n = \lambda/n$). Let $X \sim$ Pois(λ). Show that the MGF of X_n converges to the MGF of X (this gives another way to see that the Bin(n,p) distribution can be well-approximated by the $Pois(\lambda)$ when n is large, p is small, and $\lambda = np$ is moderate).

Solution: Using the fact that $(1 + x/n)^n \to e^x$ as $n \to \infty$ (the compound interest limit, which is reviewed in the math appendix), we have

$$E(e^{tX_n}) = (1 - p_n + p_n e^t)^n = (1 + \lambda(e^t - 1)/n)^n \quad \to \quad e^{\lambda(e^t-1)} = E(e^{tX}).$$

Chapter 7: Joint distributions

Joint, marginal, and conditional distributions

14. Ⓢ (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?

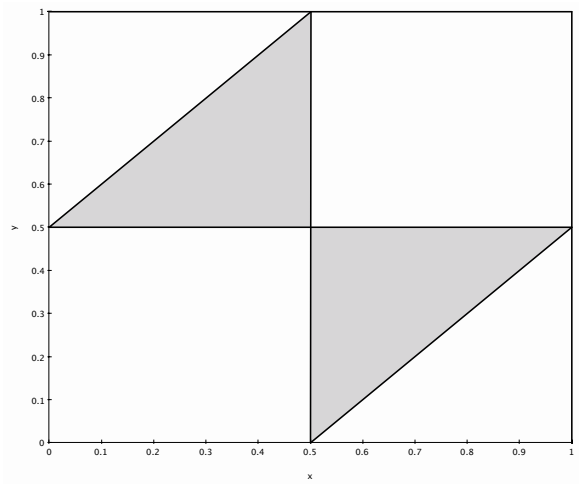
Hint: A triangle can be formed from 3 line segments of lengths a, b, c if and only if $a, b, c \in (0, 1/2)$. The probability can be interpreted geometrically as proportional to an area in the plane, avoiding all calculus, but make sure for that approach that the distribution of the random point in the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

Solution:

(a) 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. 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) 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 jth 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 $1/4$.

18. Ⓢ 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.

Solution: 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 $X|Y \sim \text{Unif}(0, 1 - Y)$.

19. Ⓢ 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.

Solution:

(a) Just as in 2 dimensions uniform in a region means that probability is proportional to area, in 3 dimensions 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) 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) 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

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

20. Ⓢ Let U_1, U_2, U_3 be i.i.d. 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.

Solution:

(a) 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) 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 $m \geq l$.

24. Ⓢ 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 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.

Solution:

(a) 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) Plugging in $t = 1$ above, we have

$$P(Y_1 < Y_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

Alternatively, we can get the same result by applying Example 7.1.23. (The result can also be derived without using calculus by thinking about Poisson processes, as shown in Chapter 13.)

26. Ⓢ The bus company from Blissville decides to start service in Blotchville, sensing a promising business opportunity. Meanwhile, Fred has moved back to Blotchville. 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 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?

Solution:

(a) 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) 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 $t \geq 10$).

2D LOTUS

31. Ⓢ Let X and Y be i.i.d. Unif$(0, 1)$. Find the standard deviation of the distance between X and Y.

Solution: 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 $\frac{1}{\sqrt{18}} = \frac{1}{3\sqrt{2}}$.

32. Ⓢ Let X, Y be i.i.d. Expo(λ). Find $E|X - Y|$ in two different ways: (a) using 2D LOTUS and (b) using the memoryless property without any calculus.

Solution:

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

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

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

(b) Write $|X - Y| = \max(X, Y) - \min(X, Y)$. By the memoryless property, this is Expo(λ), as in Example 7.3.6. So $E|X - Y| = \frac{1}{\lambda}$, which agrees with (a). (This also shows $\text{Var}(|X - Y|) = \frac{1}{\lambda^2}$.)

Covariance

38. Ⓢ Let X and Y be r.v.s. Is it correct to say "$\max(X,Y) + \min(X,Y) = X + Y$"? Is it correct to say "$\text{Cov}(\max(X,Y), \min(X,Y)) = \text{Cov}(X,Y)$ since either the max is X and the min is Y or vice versa, and covariance is symmetric"? Explain.

Solution: The identity $\max(x,y) + \min(x,y) = x + y$ is true for all numbers x and y. The random variable $M = \max(X,Y)$ is *defined* by $M(s) = \max(X(s), Y(s))$; this just says to perform the random experiment, observe the numerical values of X and Y, and take their maximum. It follows that

$$\max(X,Y) + \min(X,Y) = X + Y$$

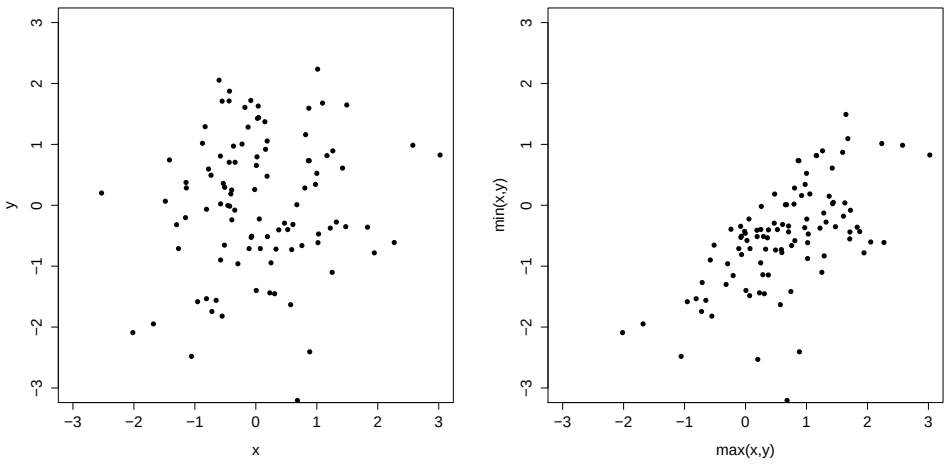
for all r.v.s X and Y, since whatever the outcome s of the random experiment is, we have

$$\max(X(s), Y(s)) + \min(X(s), Y(s)) = X(s) + Y(s).$$

In contrast, the covariance of two r.v.s is a number, not a r.v.; it is *not* defined by observing the values of the two r.v.s and then taking their covariance (that would be a useless quantity, since the covariance between two numbers is 0). It is wrong to say "$\text{Cov}(\max(X,Y), \min(X,Y)) = \text{Cov}(X,Y)$ since either the max is X and the min is Y or vice versa, and covariance is symmetric" since the r.v. X does not equal the r.v. $\max(X,Y)$, nor does it equal the r.v. $\min(X,Y)$.

To gain more intuition into this, consider a repeated sampling interpretation, where we independently repeat the same experiment many times and observe pairs $(x_1, y_1), \ldots, (x_n, y_n)$, where (x_j, y_j) is the observed value of (X,Y) for the $j$the experiment. Suppose that X and Y are independent non-constant r.v.s (and thus they are uncorrelated). Imagine a *scatter plot* of the observations (which is just a plot of the points (x_j, y_j)). Since X and Y are independent, there should be no trend in the plot.

On the other hand, imagine a scatter plot of the $(\max(x_j, y_j), \min(x_j, y_j))$ points. Here we'd expect to see a clear increasing trend (since the max is always bigger than or equal to the min, so having a large value of the min (relative to its mean) should make it more likely that we'll have a large value of the max (relative to its mean). So it makes sense that $\max(X,Y)$ and $\min(X,Y)$ should be positive correlated. This is illustrated in the plots below, in which we generated $(X_1, Y_1), \ldots, (X_{100}, Y_{100})$ with the X_i's and Y_j's i.i.d. $\mathcal{N}(0,1)$.



The simulation was done in R, using the following code:

```
x <- rnorm(100); y <- rnorm(100)
plot(x,y, xlim=c(-3,3),ylim=c(-3,3), pch=16)
plot(pmax(x,y),pmin(x,y), xlim=c(-3,3),ylim=c(-3,3), xlab="max(x,y)",
ylab = "min(x,y)", pch=16)
```

39. Ⓢ Two fair six-sided dice are rolled (one green and one orange), with outcomes X and Y respectively for the green and the orange.

(a) Compute the covariance of $X+Y$ and $X-Y$.

(b) Are $X+Y$ and $X-Y$ independent?

Solution:

(a) We have

$$\text{Cov}(X+Y, X-Y) = \text{Cov}(X,X) - \text{Cov}(X,Y) + \text{Cov}(Y,X) - \text{Cov}(Y,Y) = 0.$$

(b) They are not independent: information about $X+Y$ may give information about $X-Y$, as shown by considering an *extreme example*. Note that if $X+Y=12$, then $X=Y=6$, so $X-Y=0$. Therefore, $P(X-Y=0|X+Y=12) = 1 \neq P(X-Y=0)$, which shows that $X+Y$ and $X-Y$ are not independent. Alternatively, note that $X+Y$ and $X-Y$ are both even or both odd, since the sum $(X+Y)+(X-Y)=2X$ is even.

41. Ⓢ Let X and Y be standardized r.v.s (i.e., marginally they each have mean 0 and variance 1) with correlation $\rho \in (-1,1)$. Find a,b,c,d (in terms of ρ) such that $Z = aX+bY$ and $W = cX+dY$ are uncorrelated but still standardized.

Solution: Let us look for a solution with $Z=X$, finding c and d to make Z and W uncorrelated:

$$\text{Cov}(Z,W) = \text{Cov}(X, cX+dY) = \text{Cov}(X,cX) + \text{Cov}(X,dY) = c + d\rho = 0.$$

Also, $\text{Var}(W) = c^2 + d^2 + 2cd\rho = 1$. Solving for c,d gives

$$a = 1, b = 0, c = -\rho/\sqrt{1-\rho^2}, d = 1/\sqrt{1-\rho^2}.$$

42. Ⓢ Let X be the number of distinct birthdays in a group of 110 people (i.e., the number of days in a year such that at least one person in the group has that birthday). Under the usual assumptions (no February 29, all the other 365 days of the year are equally likely, and the day when one person is born is independent of the days when the other people are born), find the mean and variance of X.

Solution: Let I_j be the indicator r.v. for the event that at least one of the people was born on the jth day of the year, so $X = \sum_{j=1}^{365} I_j$ with $I_j \sim \text{Bern}(p)$, where $p = 1-(364/365)^{110}$. The I_j's are dependent but by linearity, we still have

$$E(X) = 365p \approx 95.083.$$

By symmetry, the variance is

$$\text{Var}(X) = 365\text{Var}(I_1) + 2\binom{365}{2}\text{Cov}(I_1,I_2).$$

To get the covariance, note that $\text{Cov}(I_1,I_2) = E(I_1I_2) - E(I_1)E(I_2) = E(I_1I_2) - p^2$, and $E(I_1I_2) = P(I_1I_2=1) = P(A_1 \cap A_2)$, where A_j is the event that at least one person was born on the jth day of the year. The probability of the complement is

$$P(A_1^c \cup A_2^c) = P(A_1^c) + P(A_2^c) - P(A_1^c \cap A_2^c) = 2\left(\frac{364}{365}\right)^{110} - \left(\frac{363}{365}\right)^{110},$$

so $\text{Var}(X) = 365p(1-p) + 365 \cdot 364 \cdot (1 - (2\left(\frac{364}{365}\right)^{110} - \left(\frac{363}{365}\right)^{110}) - p^2) \approx 10.019$.

47. Ⓢ Athletes compete one at a time at the high jump. Let X_j be how high the jth jumper jumped, with $X_1, X_2, \ldots$ i.i.d. with a continuous distribution. We say that the jth jumper sets a *record* if X_j is greater than all of $X_{j-1}, \ldots, X_1$.

Find the variance of the number of records among the first n jumpers (as a sum). What happens to the variance as $n \to \infty$?

Solution: Let I_j be the indicator r.v. for the jth jumper setting a record. By symmetry, $E(I_j) = P(I_j = 1) = 1/j$ (as all of the first j jumps are equally likely to be the largest of those jumps). It was shown on Example 5.7.3 that I_{110} and I_{111} are independent. Similarly, I_i is independent of I_j for all i, j with $i < j$ (in fact, they are independent, not just pairwise independent). To see this, note that by symmetry, learning that the jth jumper sets a record gives no information whatsoever about how the first i jumpers rank among themselves, or compute

$$P(I_i = I_j = 1) = \frac{\binom{j-1}{j-i-1}(j-i-1)!(i-1)!}{j!} = \frac{(i-1)!(j-1)!}{i!j!} = \frac{1}{ij} = P(I_1 = 1)P(I_2 = 1),$$

where the numerator corresponds to putting the best of the first j jumps in position j, picking any $j - 1 + 1$ of the remaining jumps to fill positions $i + 1$ through $j - 1$ and putting them in any order, putting the best of the remaining i jumps in position i, and then putting the remaining $i - 1$ jumps in any order.

The variance of I_j is $\text{Var}(I_j) = E(I_j^2) - (EI_j)^2 = \frac{1}{j} - \frac{1}{j^2}$. Since the I_j are pairwise independent (and thus uncorrelated), the variance of $I_1 + \cdots + I_n$ is

$$\sum_{j=1}^{n} \left(\frac{1}{j} - \frac{1}{j^2} \right),$$

which goes to ∞ as $n \to \infty$ since $\sum_{j=1}^{n} \frac{1}{j}$ diverges and $\sum_{j=1}^{n} \frac{1}{j^2}$ converges (to $\pi^2/6$, as it turns out).

48. Ⓢ A chicken lays a Pois(λ) number N of eggs. Each egg hatches a chick with probability p, independently. Let X be the number which hatch, so $X|N = n \sim \text{Bin}(n, p)$.

Find the correlation between N (the number of eggs) and X (the number of eggs which hatch). Simplify; your final answer should work out to a simple function of p (the λ should cancel out).

Solution: By the chicken-egg story, X is independent of Y, with $X \sim \text{Pois}(\lambda p)$ and $Y \sim \text{Pois}(\lambda q)$, for $q = 1 - p$. So

$$\text{Cov}(N, X) = \text{Cov}(X + Y, X) = \text{Cov}(X, X) + \text{Cov}(Y, X) = \text{Var}(X) = \lambda p,$$

giving

$$\text{Corr}(N, X) = \frac{\lambda p}{SD(N)SD(X)} = \frac{\lambda p}{\sqrt{\lambda \lambda p}} = \sqrt{p}.$$

52. Ⓢ A drunken man wanders around randomly in a large space. At each step, he moves one unit of distance North, South, East, or West, with equal probabilities. Choose coordinates such that his initial position is $(0, 0)$ and if he is at (x, y) at some time, then one step later he is at $(x, y+1), (x, y-1), (x+1, y)$, or $(x-1, y)$. Let (X_n, Y_n) and R_n be his position and distance from the origin after n steps, respectively.

General hint: Note that X_n is a sum of r.v.s with possible values $-1, 0, 1$, and likewise for Y_n, but be careful throughout the problem about independence.

(a) Determine whether or not X_n is independent of Y_n.

(b) Find $\text{Cov}(X_n, Y_n)$.

(c) Find $E(R_n^2)$.

Solution:

(a) They are *not* independent, as seen by considering an *extreme case* such as the event that the drunk headed East for the entire time: note that $P(Y_n = 0|X_n = n) = 1$.

(b) Write $X_n = \sum_{i=1}^n Z_i$ and $Y_n = \sum_{j=1}^n W_j$, where Z_i is -1 if his ith step is Westward, 1 if his ith step is Eastward, and 0 otherwise, and similarly for W_j. Then Z_i is independent of W_j for $i \neq j$. But Z_i and W_i are highly dependent: exactly one of them is 0 since he moves in one direction at a time. Then $\text{Cov}(Z_i, W_i) = E(Z_iW_i) - E(Z_i)E(W_i) = 0$ since Z_iW_i is always 0, and Z_i and W_i have mean 0. So

$$\text{Cov}(X_n, Y_n) = \sum_{i,j} \text{Cov}(Z_i, W_j) = 0.$$

(c) We have $R_n^2 = X_n^2 + Y_n^2$, and $E(Z_iZ_j) = 0$ for $i \neq j$. So

$$E(R_n^2) = E(X_n^2) + E(Y_n^2) = 2E(X_n^2) = 2nE(Z_1^2) = n,$$

since $Z_1^2 \sim \text{Bern}(1/2)$.

53. Ⓢ A scientist makes two measurements, considered to be independent standard Normal r.v.s. Find the correlation between the larger and smaller of the values.

Hint: Note that $\max(x, y) + \min(x, y) = x + y$ and $\max(x, y) - \min(x, y) = |x - y|$.

Solution: Let X and Y be i.i.d $\mathcal{N}(0, 1)$ and $M = \max(X, Y), L = \min(X, Y)$. By the hint,

$$E(M) + E(L) = E(M + L) = E(X + Y) = E(X) + E(Y) = 0,$$
$$E(M) - E(L) = E(M - L) = E|X - Y| = \frac{2}{\sqrt{\pi}},$$

where the last equality was shown in Example 7.2.3. So $E(M) = 1/\sqrt{\pi}$, and

$$\text{Cov}(M, L) = E(ML) - E(M)E(L) = E(XY) + (EM)^2 = (EM)^2 = \frac{1}{\pi},$$

since $ML = XY$ has mean $E(XY) = E(X)E(Y) = 0$. To obtain the correlation, we also need $\text{Var}(M)$ and $\text{Var}(L)$. By symmetry of the Normal, $(-X, -Y)$ has the same distribution as (X, Y), so $\text{Var}(M) = \text{Var}(L)$; call this v. Then

$$E(X - Y)^2 = \text{Var}(X - Y) = 2, \text{ and also}$$

$$E(X - Y)^2 = E(M - L)^2 = EM^2 + EL^2 - 2E(ML) = 2v + \frac{2}{\pi}.$$

So $v = 1 - \frac{1}{\pi}$ (alternatively, we can get this by taking the variance of both sides of $\max(X, Y) + \min(X, Y) = X + Y$). Thus,

$$\text{Corr}(M, L) = \frac{\text{Cov}(M, L)}{\sqrt{\text{Var}(M)\text{Var}(L)}} = \frac{1/\pi}{1 - 1/\pi} = \frac{1}{\pi - 1}.$$

55. Ⓢ Consider the following method for creating a *bivariate Poisson* (a joint distribution for two r.v.s such that both marginals are Poissons). Let $X = V + W, Y = V + Z$ where V, W, Z are i.i.d. Pois(λ) (the idea is to have something borrowed and something new but not something old or something blue).

(a) Find $\text{Cov}(X, Y)$.

(b) Are X and Y independent? Are they conditionally independent given V?

(c) Find the joint PMF of X, Y (as a sum).

Solution:

(a) Using the properties of covariance, we have

$$\text{Cov}(X,Y) = \text{Cov}(V,V) + \text{Cov}(V,Z) + \text{Cov}(W,V) + \text{Cov}(W,Z) = \text{Var}(V) = \lambda.$$

(b) Since X and Y are correlated (with covariance $\lambda > 0$), they are not independent. Alternatively, note that $E(Y) = 2\lambda$ but $E(Y|X=0) = \lambda$ since if $X = 0$ occurs then $V = 0$ occurs. But X and Y are conditionally independent given V, since the conditional joint PMF is

$$\begin{aligned} P(X=x, Y=y|V=v) &= P(W=x-v, Z=y-v|V=v) \\ &= P(W=x-v, Z=y-v) \\ &= P(W=x-v)P(Z=y-v) \\ &= P(X=x|V=v)P(Y=y|V=v). \end{aligned}$$

This makes sense intuitively since if we observe that $V = v$, then X and Y are the independent r.v.s W and Z, shifted by the constant v.

(c) By (b), a good strategy is to condition on V:

$$\begin{aligned} P(X=x, Y=y) &= \sum_{v=0}^{\infty} P(X=x, Y=y|V=v)P(V=v) \\ &= \sum_{v=0}^{\min(x,y)} P(X=x|V=v)P(Y=y|V=v)P(V=v) \\ &= \sum_{v=0}^{\min(x,y)} e^{-\lambda}\frac{\lambda^{x-v}}{(x-v)!}e^{-\lambda}\frac{\lambda^{y-v}}{(y-v)!}e^{-\lambda}\frac{\lambda^{v}}{v!} \\ &= e^{-3\lambda}\lambda^{x+y}\sum_{v=0}^{\min(x,y)} \frac{\lambda^{-v}}{(x-v)!(y-v)!v!}, \end{aligned}$$

for x and y nonnegative integers. Note that we sum only up to $\min(x, y)$ since we know for sure that $V \leq X$ and $V \leq Y$.

Sanity check: Note that $P(X = 0, Y = 0) = P(V = 0, W = 0, Z = 0) = e^{-3\lambda}$.

Chicken-egg

59. Ⓢ A Pois(λ) number of people vote in a certain election. Each voter votes for candidate A with probability p and for candidate B with probability $q = 1 - p$, independently of all the other voters. Let V be the difference in votes, defined as the number of votes for A minus the number for B.

(a) Find $E(V)$.

(b) Find $\text{Var}(V)$.

Solution:

(a) Let X and Y be the number of votes for A and B respectively, and let $N = X + Y$. Then $X|N \sim \text{Bin}(N, p)$ and $Y|N \sim \text{Bin}(N, q)$. By Adam's Law or the chicken-egg story, $E(X) = \lambda p$ and $E(Y) = \lambda q$. So

$$E(V) = E(X - Y) = E(X) - E(Y) = \lambda(p - q).$$

(b) By the chicken-egg story, $X \sim \text{Pois}(\lambda p)$ and $Y \sim \text{Pois}(\lambda q)$ are independent. So

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) = \lambda p + \lambda q = \lambda.$$

Multinomial

64. Ⓢ Let $(X_1, \ldots, X_k)$ be Multinomial with parameters n and $(p_1, \ldots, p_k)$. Use indicator r.v.s to show that $\text{Cov}(X_i, X_j) = -np_ip_j$ for $i \neq j$.

Solution: First let us find $\text{Cov}(X_1, X_2)$. Consider the story of the Multinomial, where n objects are being placed into categories $1, \ldots, k$. Let I_i be the indicator r.v. for object i being in category 1, and let J_j be the indicator r.v. for object j being in category 2. Then $X_1 = \sum_{i=1}^n I_i, X_2 = \sum_{j=1}^n J_j$. So

$$\begin{aligned}\text{Cov}(X_1, X_2) &= \text{Cov}(\sum_{i=1}^n I_i, \sum_{j=1}^n J_j) \\ &= \sum_{i,j} \text{Cov}(I_i, J_j).\end{aligned}$$

All the terms here with $i \neq j$ are 0 since the ith object is categorized independently of the jth object. So this becomes

$$\sum_{i=1}^n \text{Cov}(I_i, J_i) = n\text{Cov}(I_1, J_1) = -np_1p_2,$$

since

$$\text{Cov}(I_1, J_1) = E(I_1J_1) - (EI_1)(EJ_1) = -p_1p_2.$$

By the same method, we have $\text{Cov}(X_i, X_j) = -np_ip_j$ for all $i \neq j$.

65. Ⓢ Consider the birthdays of 100 people. Assume people's birthdays are independent, and the 365 days of the year (exclude the possibility of February 29) are equally likely. Find the covariance and correlation between how many of the people were born on January 1 and how many were born on January 2.

Solution: Let X_j be the number of people born on January j. Then

$$\text{Cov}(X_1, X_2) = -\frac{100}{365^2},$$

using the result about covariances in a Multinomial. Since $X_j \sim \text{Bin}(100, 1/365)$, we then have

$$\text{Corr}(X_1, X_2) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = -\frac{100/365^2}{100(1/365)(364/365)} = -\frac{1}{364}.$$

67. Ⓢ 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.

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

Solution:

(a) The joint PMF 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 $\text{Mult}_3(n, (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}))$ distribution.

(b) 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) For $n = 5$, the probability is $(2^5 - 2)/3^4 = 30/81 \approx 0.37$. As $n \to \infty$, $(2^n - 2)/3^{n-1} \to 0$, 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.

68. Ⓢ Emails arrive in an inbox according to a Poisson process with rate λ (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).

Solution:

(a) 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) 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 $\text{Mult}_3(36, (\frac{1}{5}, \frac{2}{5}, \frac{2}{5}))$ given $T = 36$.

(c) 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 $\text{Var}(W|T = 36) = 36 \cdot \frac{9}{15} \cdot \frac{6}{15} = 8.64$.

Multivariate Normal

71. Ⓢ Let (X,Y) be Bivariate Normal, with X and Y marginally $\mathcal{N}(0,1)$ and with correlation ρ between X and Y.

(a) Show that $(X+Y, X-Y)$ is also Bivariate Normal.

(b) Find the joint PDF of $X+Y$ and $X-Y$ (without using calculus), assuming $-1 < \rho < 1$.

Solution:

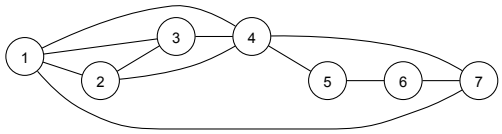
(a) The linear combination $s(X+Y)+t(X-Y) = (s+t)X+(s-t)Y$ is also a linear combination of X and Y, so it is Normal, which shows that $(X+Y, X-Y)$ is MVN.

(b) Since $X+Y$ and $X-Y$ are uncorrelated (as $\text{Cov}(X+Y, X-Y) = \text{Var}(X)-\text{Var}(Y) = 0$) and $(X+Y, X-Y)$ is MVN, they are independent. Marginally, $X+Y \sim \mathcal{N}(0, 2+2\rho)$ and $X-Y \sim \mathcal{N}(0, 2-2\rho)$. Thus, the joint PDF is

$$f(s,t) = \frac{1}{4\pi\sqrt{1-\rho^2}} e^{-\frac{1}{4}(s^2/(1+\rho)+t^2/(1-\rho))}.$$

Mixed practice

84. Ⓢ A *network* consists of n *nodes*, each pair of which may or may not have an *edge* joining them. For example, a social network can be modeled as a group of n nodes (representing people), where an edge between i and j means they know each other. Assume the network is undirected and does not have edges from a node to itself (for a social network, this says that if i knows j, then j knows i and that, contrary to Socrates' advice, a person does not know himself or herself). A *clique* of size k is a set of k nodes where every node has an edge to every other node (i.e., within the clique, everyone knows everyone). An *anticlique* of size k is a set of k nodes where there are no edges between them (i.e., within the anticlique, no one knows anyone else). For example, the picture below shows a network with nodes labeled $1, 2, \ldots, 7$, where $\{1, 2, 3, 4\}$ is a clique of size 4, and $\{3, 5, 7\}$ is an anticlique of size 3.



(a) Form a random network with n nodes by independently flipping fair coins to decide for each pair $\{x, y\}$ whether there is an edge joining them. Find the expected number of cliques of size k (in terms of n and k).

(b) A *triangle* is a clique of size 3. For a random network as in (a), find the variance of the number of triangles (in terms of n).

Hint: Find the covariances of the indicator random variables for each possible clique. There are $\binom{n}{3}$ such indicator r.v.s, some pairs of which are dependent.

*(c) Suppose that $\binom{n}{k} < 2^{\binom{k}{2}-1}$. Show that there is a network with n nodes containing no cliques of size k or anticliques of size k.

Hint: Explain why it is enough to show that for a random network with n nodes, the probability of the desired property is positive; then consider the complement.

Solution:

(a) Order the $\binom{n}{k}$ subsets of people of size k in some way (i.e., give each subset of size k a code number), and let X_i be the indicator. Since $X_1 + X_2 + \cdots + X_{\binom{n}{k}}$ is the number of cliques of size k, the expected number is

$$E(X_1 + X_2 + \cdots + X_{\binom{n}{k}}) = \binom{n}{k} E(X_1) = \binom{n}{k} P(X_1 = 1) = \frac{\binom{n}{k}}{2^{\binom{k}{2}}}.$$

(b) Let $k = 3$ and the X_i be as in (a). Then

$$\begin{aligned} \mathrm{Var}(X_1 + \cdots + X_{\binom{n}{3}}) &= \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_{\binom{n}{3}}) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j) \\ &= \binom{n}{3} \mathrm{Var}(X_1) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j), \end{aligned}$$

with

$$\mathrm{Var}(X_1) = 2^{-\binom{3}{2}}(1 - 2^{-\binom{3}{2}}) = \frac{7}{64}.$$

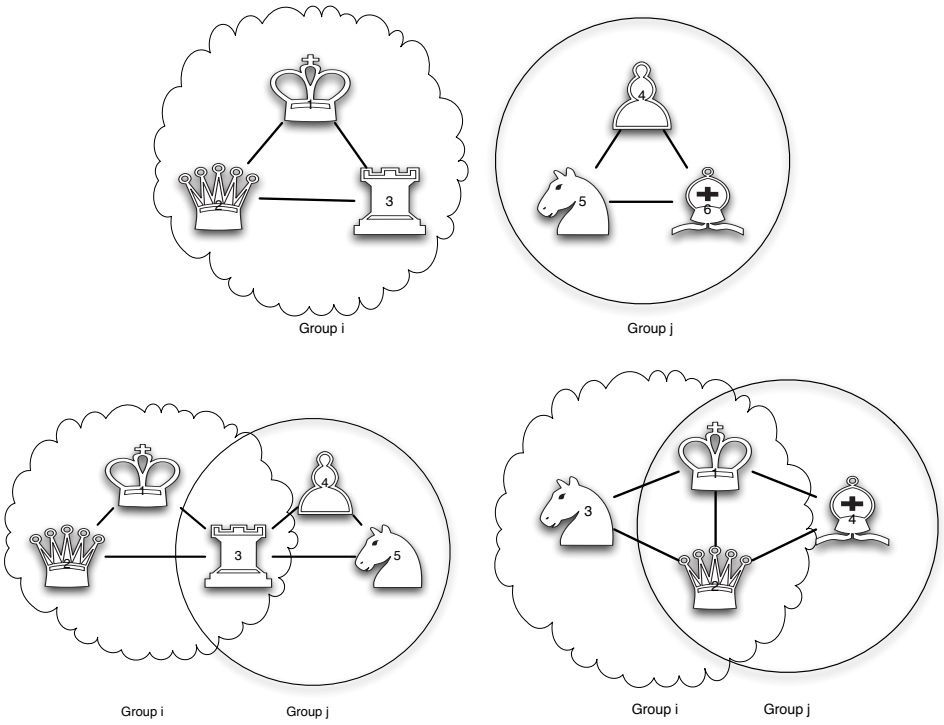


FIGURE 1
Two groups with 0 (upper), 1 (lower left), 2 (lower right) people in common.

To compute $\text{Cov}(X_i, X_j)$ for $i < j$, consider how many people are in common for group i and group j. If the number of people in common is 0 or 1 (as shown in the upper and lower left cases in the above figure, respectively), then the $\text{Cov}(X_i, X_j) = 0$ since the coin flips used to determine whether Group i is a clique are independent of those used for Group j. If there are 2 people in common (as shown in the lower right case of Figure 1), then

$$\text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j) = \frac{1}{2^5} - (\frac{1}{2^3})^2 = \frac{1}{64},$$

since 5 distinct pairs of people must know each other to make $X_i X_j$ equal to 1.

There are $\binom{n}{4}\binom{4}{2} = 6\binom{n}{4}$ pairs of groups $\{i, j\}$ $(i \neq j)$ with 1 pair of people in common (choose 4 people out of the n, then choose which 2 of the 4 are the overlap of the groups). The remaining pairs of groups have covariance 0. Thus, the variance of the number of cliques is

$$\frac{7}{64}\binom{n}{3} + 2 \cdot 6\binom{n}{4} \cdot \frac{1}{64} = \frac{7}{64}\binom{n}{3} + \frac{3}{16}\binom{n}{4}.$$

(c) We will prove the existence of a network with the desired property by showing that the probability is positive that a random network has the property is positive (this strategy is explored in the starred Section 4.9). Form a random network as in (a), and let A_i be the event that the ith group of k people (in any fixed ordering) is neither a clique nor an anticlique. We have

$$P(\bigcup_{i=1}^{\binom{n}{k}} A_i^c) \leq \sum_{i=1}^{\binom{n}{k}} P(A_i^c) = \binom{n}{k} 2^{-\binom{k}{2}+1} < 1,$$

which shows that

$$P(\bigcap_{i=1}^{\binom{n}{k}} A_i) = 1 - P(\bigcup_{i=1}^{\binom{n}{k}} A_i^c) > 0,$$

as desired. Alternatively, let C be the number of cliques of size k and A be the number of anticliques of size k, and write $C + A = T$. Then

$$E(T) = E(C) + E(A) = \binom{n}{k} 2^{-\binom{k}{2}+1} < 1,$$

by the method of Part (a). So $P(T = 0) > 0$, since $P(T \geq 1) = 1$ would imply $E(T) \geq 1$. This again shows that there must be a network with the desired property.

85. Ⓢ 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.

Ronald Thisted and Bradley Efron studied this problem in the papers [9] and [10], developing theory and methods and then applying the methods 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 Pois(λ) r.v.s. Show that the probability that "eyeball" is used in play B but not in play A is

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

(c) Now assume that λ from (b) is unknown and is itself taken to be a random variable to reflect this uncertainty. So let λ 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 λ is that they are independent Pois(λ) r.v.s. Show that the probability that "eyeball" is used in play B but not in play A is the alternating series

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

Hint: Condition on λ and use (b).

(d) Assume that every word's numbers of occurrences in A and B are distributed as in

(c), where λ 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

$$E(W_1) - E(W_2) + E(W_3) - E(W_4) + \ldots.$$

(This shows that $W_1 - W_2 + W_3 - W_4 + \ldots$ 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 λ_j. This method can be extended in various ways to give predictions for unobserved plays based on observed plays.)

Solution:

(a) 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 λ to be the same for X_j and Y_j.

(b) 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 Pois(λ),

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

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

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

(d) 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 $1 \le j \le N$,

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

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

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

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

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

Chapter 8: Transformations

Change of variables

4. Ⓢ Find the PDF of Z^4 for $Z \sim \mathcal{N}(0,1)$.

Solution: Let $Y = Z^4$. For $y > 0$, the CDF of Y is

$$P(Y \leq y) = P(Z^4 \leq y) = P(-y^{1/4} \leq Z \leq y^{1/4}) = \Phi(y^{1/4}) - \Phi(-y^{1/4}) = 2\Phi(y^{1/4}) - 1.$$

So the PDF is

$$f_Y(y) = \frac{2}{4} y^{-3/4} \varphi(y^{1/4}) = \frac{1}{2\sqrt{2\pi}} y^{-3/4} e^{-y^{1/2}/2},$$

for $y > 0$, where φ is the $\mathcal{N}(0,1)$ PDF.

6. Ⓢ Let $U \sim \text{Unif}(0,1)$. Find the PDFs of U^2 and $\sqrt{U}$.

Solution:
(PDF of U^2.) Let $Y = U^2$, $0 < u < 1$, and $y = u^2$, so $u = \sqrt{y}$. The absolute Jacobian determinant is $\left|\frac{du}{dy}\right| = \left|\frac{1}{2\sqrt{y}}\right| = \frac{1}{2\sqrt{y}}$ for $0 < y < 1$. The PDF of Y for $0 < y < 1$ is

$$f_Y(y) = f_U(u) \left|\frac{du}{dy}\right| = \frac{1}{2\sqrt{y}},$$

with $f_Y(y) = 0$ otherwise. This is the $\text{Beta}(\frac{1}{2}, 1)$ PDF, so $Y = U^2 \sim \text{Beta}(\frac{1}{2}, 1)$.

(PDF of $\sqrt{U}$.) Now let $Y = U^{1/2}$, $0 < u < 1$, and $y = u^{1/2}$, so $u = y^2$. The absolute Jacobian determinant is $\left|\frac{du}{dy}\right| = |2y| = 2y$ for $0 < y < 1$. The PDF of Y for $0 < y < 1$ is

$$f_Y(y) = f_U(u) \left|\frac{du}{dy}\right| = 2y,$$

with $f_Y(y) = 0$ otherwise. This says that Y has a Beta(2,1) distribution.

In general, the same method shows that $U^{\frac{1}{\alpha}} \sim \text{Beta}(\alpha, 1)$ for any $\alpha > 0$.

16. Ⓢ Let X, Y be continuous r.v.s with a *spherically symmetric* joint distribution, which means that the joint PDF is of the form $f(x,y) = g(x^2+y^2)$ for some function g. Let (R, θ) be the polar coordinates of (X, Y), so $R^2 = X^2 + Y^2$ is the squared distance from the origin and θ is the angle (in $[0, 2\pi)$), with $X = R\cos\theta, Y = R\sin\theta$.

(a) Explain intuitively why R and θ are independent. Then prove this by finding the joint PDF of (R, θ).

(b) What is the joint PDF of (R, θ) when (X, Y) is Uniform in the unit disk $\{(x,y) : x^2 + y^2 \leq 1\}$?

(c) What is the joint PDF of (R, θ) when X and Y are i.i.d. $\mathcal{N}(0,1)$?

Solution:

(a) Intuitively, this makes sense since the joint PDF of X, Y at a point (x, y) only

depends on the distance from (x, y) to the origin, not on the angle, so knowing R gives no information about θ. The absolute Jacobian is r (as shown on the math review handout), so

$$f_{R,\theta}(r,t) = f_{X,Y}(x,y)r = r \cdot g(r^2)$$

for all $r \geq 0, t \in [0, 2\pi)$. This factors as a function of r times a (constant) function of t, so R and θ are independent with $\theta \sim \text{Unif}(0, 2\pi)$.

(b) We have $f_{X,Y}(x,y) = \frac{1}{\pi}$ for $x^2 + y^2 \leq 1$, so $f_{R,\theta}(r,t) = \frac{r}{\pi}$ for $0 \leq r \leq 1, t \in [0, 2\pi)$ (and the PDF is 0 otherwise). This says that R and θ are independent with marginal PDFs $f_R(r) = 2r$ for $0 \leq r \leq 1$ and $f_\theta(t) = \frac{1}{2\pi}$ for $0 \leq t < 2\pi$.

(c) The joint PDF of X, Y is $\frac{1}{2\pi}e^{-(x^2+y^2)/2}$, so $g(r^2) = \frac{1}{2\pi}e^{-r^2/2}$ and the joint PDF of (R, θ) is $\frac{1}{2\pi}re^{-r^2/2}$. This says that R and θ are independent with marginal PDFs $f_R(r) = re^{-r^2/2}$ for $r \geq 0$ and $f_\theta(t) = \frac{1}{2\pi}$ for $0 \leq t < 2\pi$. (The distribution of R is an example of a *Weibull*; note that it is the distribution of $W^{1/2}$ for $W \sim \text{Expo}(1/2)$.)

Convolutions

24. Ⓢ Let X and Y be independent positive r.v.s, with PDFs f_X and f_Y respectively, and consider the product $T = XY$. When asked to find the PDF of T, Jacobno argues that "it's like a convolution, with a product instead of a sum. To have $T = t$ we need $X = x$ and $Y = t/x$ for some x; that has probability $f_X(x)f_Y(t/x)$, so summing up these possibilities we get that the PDF of T is $\int_0^\infty f_X(x)f_Y(t/x)dx$." Evaluate Jacobno's argument, while getting the PDF of T (as an integral) in 2 ways:

(a) using the continuous version of the law of total probability to get the CDF, and then taking the derivative (you can assume that swapping the derivative and integral is valid);

(b) by taking the log of both sides of $T = XY$ and doing a convolution (and then converting back to get the PDF of T).

Solution:

(a) By the law of total probability (conditioning on X),

$$\begin{aligned} P(T \leq t) &= \int_0^\infty P(XY \leq t | X = x) f_X(x) dx \\ &= \int_0^\infty P(Y \leq t/x | X = x) f_X(x) dx \\ &= \int_0^\infty F_Y(t/x) f_X(x) dx, \end{aligned}$$

which has derivative

$$f_T(t) = \int_0^\infty f_X(x) f_Y(t/x) \frac{dx}{x}.$$

This is *not* the same as Jacobno claimed: there is an extra x in the denominator. This stems from the fact that the transformation (X, Y) to (XY, X) is nonlinear, in contrast to the transformation (X, Y) to $(X+Y, X)$ considered in SP 8 # 2.3. Jacobno is ignoring the distinction between probabilities and probability *densities*, and is implicitly (and incorrectly) assuming that there is no Jacobian term.

(b) Let $Z = \log(T), W = \log(X), V = \log(Y)$, so $Z = W + V$. The PDF of Z is

$$f_Z(z) = \int_{-\infty}^\infty f_W(w) f_V(z - w) dw,$$

where by change of variables $f_W(w) = f_X(e^w)e^w$, $f_V(v) = f_Y(e^v)e^v$. So

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(e^w)e^w f_Y(e^{z-w})e^{z-w}dw = e^z \int_{-\infty}^{\infty} f_X(e^w)f_Y(e^{z-w})dw.$$

Transforming back to T, we have

$$f_T(t) = f_Z(\log t)\frac{1}{t} = \int_{-\infty}^{\infty} f_X(e^w)f_Y(e^{\log(t)-w})dw = \int_0^{\infty} f_X(x)f_Y(t/x)\frac{dx}{x},$$

letting $x = e^w$. This concurs with (a): Jacobno is missing the x in the denominator.

Beta and Gamma

29. Ⓢ Let $B \sim \text{Beta}(a, b)$. Find the distribution of $1 - B$ in two ways: (a) using a change of variables and (b) using a story proof. Also explain why the result makes sense in terms of Beta being the conjugate prior for the Binomial.

Solution:

(a) Let $W = 1 - B$. The function $g(t) = 1 - t$ is strictly decreasing with absolute derivative $|-1| = 1$, so the PDF of W is

$$f_W(w) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}(1-w)^{a-1}w^{b-1},$$

for $0 < w < 1$, which shows that $W \sim \text{Beta}(b, a)$.

(b) Using the bank–post office story, we can represent $B = \frac{X}{X+Y}$ with $X \sim \text{Gamma}(a, 1)$ and $Y \sim \text{Gamma}(b, 1)$ independent. Then $1 - B = \frac{Y}{X+Y} \sim \text{Beta}(b, a)$ by the same story.

This result makes sense intuitively since if we use $\text{Beta}(a, b)$ as the prior distribution for the probability p of success in a Binomial problem, interpreting a as the number of prior successes and b as the number of prior failures, then $1 - p$ is the probability of failure and, interchanging the roles of "success" and "failure," it makes sense to have $1 - p \sim \text{Beta}(b, a)$.

30. Ⓢ Let $X \sim \text{Gamma}(a, \lambda)$ and $Y \sim \text{Gamma}(b, \lambda)$ be independent, with a and b integers. Show that $X + Y \sim \text{Gamma}(a + b, \lambda)$ in three ways: (a) with a convolution integral; (b) with MGFs; (c) with a story proof.

Solution:

(a) The convolution integral is

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(x)f_Y(t-x)dx = \int_0^t \frac{1}{\Gamma(a)}\frac{1}{\Gamma(b)}(\lambda x)^a(\lambda(t-x))^b e^{-\lambda x}e^{-\lambda(t-x)}\frac{1}{x}\frac{1}{t-x}dx,$$

where we integrate from 0 to t since we need $x > 0$ and $t - x > 0$. This is

$$\lambda^{a+b}\frac{e^{-\lambda t}}{\Gamma(a)\Gamma(b)}\int_0^t x^{a-1}(t-x)^{b-1}dx = \lambda^{a+b}\frac{e^{-\lambda t}}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}t^{a+b-1} = \frac{1}{\Gamma(a+b)}(\lambda t)^{a+b}e^{-\lambda t}\frac{1}{t},$$

using a Beta integral (after letting $u = x/t$ so that we can integrate from 0 to 1 rather than 0 to t). Thus, $X + Y \sim \text{Gamma}(a + b, \lambda)$.

(b) The MGF of $X + Y$ is $M_X(t)M_Y(t) = \frac{\lambda^a}{(\lambda-t)^a}\frac{\lambda^b}{(\lambda-t)^b} = \frac{\lambda^{a+b}}{(\lambda-t)^{a+b}} = M_{X+Y}(t)$, which again shows that $X + Y \sim \text{Gamma}(a + b, \lambda)$.

(c) Interpret X as the time of the ath arrival in a Poisson process with rate λ, and Y as the time needed for b more arrivals to occur (which is independent of X since the times between arrivals are independent Expo(λ) r.v.s). Then $X + Y$ is the time of the $(a + b)$th arrival, so $X + Y \sim \text{Gamma}(a + b, \lambda)$.

32. Ⓢ Fred waits $X \sim \text{Gamma}(a, \lambda)$ minutes for the bus to work, and then waits $Y \sim \text{Gamma}(b, \lambda)$ for the bus going home, with X and Y independent. Is the ratio X/Y independent of the total wait time $X + Y$?

Solution: As shown in the bank–post office story, $W = \frac{X}{X+Y}$ is independent of $X + Y$. So any function of W is independent of any function of $X + Y$. And we have that X/Y is a function of W, since

$$\frac{X}{Y} = \frac{\frac{X}{X+Y}}{\frac{Y}{X+Y}} = \frac{W}{1-W},$$

so X/Y is independent of $X + Y$.

33. Ⓢ The F-test is a very widely used statistical test based on the $F(m, n)$ distribution, which is the distribution of $\frac{X/m}{Y/n}$ with $X \sim \text{Gamma}(\frac{m}{2}, \frac{1}{2}), Y \sim \text{Gamma}(\frac{n}{2}, \frac{1}{2})$. Find the distribution of $mV/(n + mV)$ for $V \sim F(m, n)$.

Solution: Let $X \sim \text{Gamma}(\frac{m}{2}, \frac{1}{2}), Y \sim \text{Gamma}(\frac{n}{2}, \frac{1}{2})$, and $V = \frac{n}{m}\frac{X}{Y}$. Then

$$mV/(n + mV) = \frac{nX/Y}{n + nX/Y} = \frac{X}{X + Y} \sim \text{Beta}\left(\frac{m}{2}, \frac{n}{2}\right).$$

34. Ⓢ Customers arrive at the Leftorium store according to a Poisson process with rate λ customers per hour. The true value of λ is unknown, so we treat it as a random variable. Suppose that our prior beliefs about λ can be expressed as $\lambda \sim \text{Expo}(3)$. Let X be the number of customers who arrive at the Leftorium between 1 pm and 3 pm tomorrow. Given that $X = 2$ is observed, find the posterior PDF of λ.

Solution: It follows from Story 8.4.5 (Gamma-Poisson conjugacy) that the posterior distribution of λ given the data is $\text{Gamma}(3, 5)$. Equivalently, we can use Bayes' rule directly. Writing f_0 for the prior PDF and f_1 for the posterior PDF, we have

$$f_1(\lambda|x) = \frac{P(X = x|\lambda)f_0(\lambda)}{P(X = x)},$$

where $f_0(\lambda) = 3e^{-3\lambda}$ for $\lambda > 0$, and $P(X = x|\lambda)$ is obtained from the $\text{Pois}(2\lambda)$ PMF. For $x = 2$, the numerator is

$$\frac{e^{-2\lambda}(2\lambda)^2}{2!} \cdot 3e^{-3\lambda} = 6\lambda^2 e^{-5\lambda}.$$

The denominator does not depend on λ, so it serves as a normalizing constant for the posterior PDF. So the posterior PDF is proportional to $\lambda^2 e^{-5\lambda}$, which shows that the posterior distribution is $\text{Gamma}(3, 5)$. Including the normalizing constant for the Gamma distribution, we have

$$f_1(\lambda|2) = \frac{5^3}{\Gamma(3)}\lambda^2 e^{-5\lambda} = \frac{125}{2}\lambda^2 e^{-5\lambda},$$

for $\lambda > 0$.

35. Ⓢ Let X and Y be independent, positive r.v.s. with finite expected values.

(a) Give an example where $E(\frac{X}{X+Y}) \neq \frac{E(X)}{E(X+Y)}$, computing both sides exactly.

Hint: Start by thinking about the simplest examples you can think of!

(b) If X and Y are i.i.d., then is it necessarily true that $E(\frac{X}{X+Y}) = \frac{E(X)}{E(X+Y)}$?

(c) Now let $X \sim \text{Gamma}(a, \lambda)$ and $Y \sim \text{Gamma}(b, \lambda)$. Show *without using calculus* that

$$E\left(\frac{X^c}{(X + Y)^c}\right) = \frac{E(X^c)}{E((X + Y)^c)}$$

for every real $c > 0$.

Solution:

(a) As a simple example, let X take on the values 1 and 3 with probability 1/2 each, and let Y take on the values 3 and 5 with probability 1/2 each. Then $E(X)/E(X+Y) = 2/(2+4) = 1/3$, but $E(X/(X+Y)) = 31/96$ (the average of the 4 possible values of $X/(X+Y)$, which are equally likely). An even simpler example is to let X be the constant 1 (a degenerate r.v.), and let Y be 1 or 3 with probability 1/2 each. Then $E(X)/E(X+Y) = 1/(1+2) = 1/3$, but $E(X/(X+Y)) = 3/8$.

(b) Yes, since by symmetry $E(\frac{X}{X+Y}) = E(\frac{Y}{X+Y})$ and by linearity

$$E\left(\frac{X}{X+Y}\right) + E\left(\frac{Y}{X+Y}\right) = E\left(\frac{X+Y}{X+Y}\right) = 1,$$

so $E(\frac{X}{X+Y}) = 1/2$, while on the other hand

$$\frac{E(X)}{E(X+Y)} = \frac{E(X)}{E(X)+E(Y)} = \frac{E(X)}{E(X)+E(X)} = 1/2.$$

(c) The equation we need to show can be paraphrased as the statement that $X^c/(X+Y)^c$ and $(X+Y)^c$ are uncorrelated. By the bank–post office story, $X/(X+Y)$ is independent of $X+Y$. So $X^c/(X+Y)^c$ is independent of $(X+Y)^c$, which shows that they are uncorrelated.

Order statistics

41. Ⓢ Let $X \sim \text{Bin}(n,p)$ and $B \sim \text{Beta}(j, n-j+1)$, where n is a positive integer and j is a positive integer with $j \leq n$. Show using a story about order statistics that

$$P(X \geq j) = P(B \leq p).$$

This shows that the CDF of the continuous r.v. B is closely related to the CDF of the discrete r.v. X, and is another connection between the Beta and Binomial.

Solution: Let $U_1, \ldots, U_n$ be i.i.d. Unif(0,1). Think of these as Bernoulli trials, where U_j is defined to be "successful" if $U_j \leq p$ (so the probability of success is p for each trial). Let X be the number of successes. Then $X \geq j$ is the same event as $U_{(j)} \leq p$, so $P(X \geq j) = P(U_{(j)} \leq p)$.

45. Ⓢ Let X and Y be independent Expo(λ) r.v.s and $M = \max(X,Y)$. Show that M has the same distribution as $X + \frac{1}{2}Y$, in two ways: (a) using calculus and (b) by remembering the memoryless property and other properties of the Exponential.

Solution:

(a) The CDF of M is

$$F_M(x) = P(M \leq x) = P(X \leq x, Y \leq x) = (1-e^{-\lambda x})^2,$$

and the CDF of $X + \frac{1}{2}Y$ is

$$F_{X+\frac{1}{2}Y}(x) = P(X + \frac{1}{2}Y \leq x) = \iint\limits_{s+\frac{1}{2}t \leq x} \lambda^2 e^{-\lambda s - \lambda t} ds dt$$

$$= \int_0^{2x} \lambda e^{-\lambda t} dt \int_0^{x-\frac{1}{2}t} \lambda e^{-\lambda s} ds$$

$$= \int_0^{2x} (1 - e^{-\lambda x - \frac{1}{2}t})\lambda e^{-\lambda t} dt = (1 - e^{-\lambda x})^2.$$

Thus, M and $X + \frac{1}{2}Y$ have the same CDF.

(b) As in Example 7.3.6, imagine that two students are independently trying to solve a problem. Suppose that X and Y are the times required. Let $L = \min(X, Y)$, and write $M = L + (M - L)$. $L \sim \text{Expo}(2\lambda)$ is the time it takes for the first student to solve the problem and then by the memoryless property, the additional time until the second student solves the problem is $M - L \sim \text{Expo}(\lambda)$, independent of L. Since $\frac{1}{2}Y \sim \text{Expo}(2\lambda)$ is independent of $X \sim \text{Expo}(\lambda)$, $M = L + (M - L)$ has the same distribution as $\frac{1}{2}Y + X$.

46. Ⓢ (a) If X and Y are i.i.d. continuous r.v.s with CDF $F(x)$ and PDF $f(x)$, then $M = \max(X, Y)$ has PDF $2F(x)f(x)$. Now let X and Y be discrete and i.i.d., with CDF $F(x)$ and PMF $f(x)$. Explain in words why the PMF of M is *not* $2F(x)f(x)$.

(b) Let X and Y be independent Bern(1/2) r.v.s, and let $M = \max(X, Y)$, $L = \min(X, Y)$. Find the joint PMF of M and L, i.e., $P(M = a, L = b)$, and the marginal PMFs of M and L.

Solution:

(a) The PMF is not $2F(x)f(x)$ in the discrete case due to the problem of ties: there is a nonzero chance that $X = Y$. We can write the PMF as $P(M = a) = P(X = a, Y < a) + P(Y = a, X < a) + P(X = Y = a)$ since $M = a$ means that at least one of X, Y equals a, with neither greater than a. The first two terms together become $2f(a)P(Y < a)$, but the third term may be nonzero and also $P(Y < a)$ may not equal $F(a) = P(Y \leq a)$.

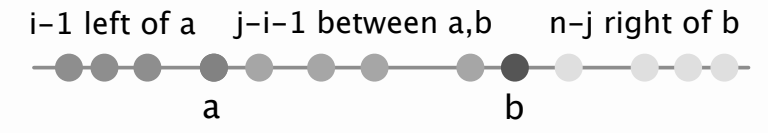
(b) In order statistics notation, $L = X_{(1)}, M = X_{(2)}$. Marginally, we have $X_{(1)} \sim \text{Bern}(1/4), X_{(2)} \sim \text{Bern}(3/4)$. The joint PMF is

$$\begin{aligned} P(X_{(1)} = 0, X_{(2)} = 0) &= 1/4 \\ P(X_{(1)} = 0, X_{(2)} = 1) &= 1/2 \\ P(X_{(1)} = 1, X_{(2)} = 0) &= 0 \\ P(X_{(1)} = 1, X_{(2)} = 1) &= 1/4. \end{aligned}$$

Note that these values are nonnegative and sum to 1, and that $X_{(1)}$ and $X_{(2)}$ are dependent.

48. Ⓢ Let $X_1, X_2, \ldots, X_n$ be i.i.d. r.v.s with CDF F and PDF f. Find the joint PDF of the order statistics $X_{(i)}$ and $X_{(j)}$ for $1 \leq i < j \leq n$, by drawing and thinking about a picture.

Solution:



To have $X_{(i)}$ be in a tiny interval around a and $X_{(j)}$ be in a tiny interval around b, where $a < b$, we need to have 1 of the X_k's be almost exactly at a, another be almost exactly at b, $i - 1$ of them should be to the left of a, $n - j$ should be to the right of b, and the remaining $j - i - 1$ should be between a and b, as shown in the picture. This gives that the PDF is

$$f_{(i),(j)}(a, b) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} F(a)^{i-1} f(a)(F(b) - F(a))^{j-i-1} f(b)(1 - F(b))^{n-j},$$

for $a < b$. The coefficient in front counts the number of ways to put the X_k's into the 5 categories "left of a," "at a," "between a and b," "at b," "right of b" with the desired number in each category (which is the same idea used to find the coefficient in front of the Multinomial PMF). Equivalently, we could write the coefficient as $n(n-1)\binom{n-2}{i-1}\binom{n-i-1}{j-i-1}$, since there are n choices for which X_k is at a, then $n-1$ choices for which is at b, etc.

49. Ⓢ Two women are pregnant, both with the same due date. On a timeline, define time 0 to be the instant when the due date begins. Suppose that the time when the woman gives birth has a Normal distribution, centered at 0 and with standard deviation 8 days. Assume that the two birth times are i.i.d. Let T be the time of the first of the two births (in days).

(a) Show that

$$E(T) = \frac{-8}{\sqrt{\pi}}.$$

Hint: For any two random variables X and Y, we have $\max(X,Y)+\min(X,Y) = X+Y$ and $\max(X,Y) - \min(X,Y) = |X-Y|$. Example 7.2.3 derives the expected distance between two i.i.d. $\mathcal{N}(0,1)$ r.v.s.

(b) Find $\text{Var}(T)$, in terms of integrals. You can leave your answers unsimplified for this problem, but it can be shown that the answer works out to

$$\text{Var}(T) = 64\left(1 - \frac{1}{\pi}\right).$$

Solution: Let $T = \min(T_1, T_2)$, with T_1 and T_2 the i.i.d. birth times. Standardizing, let

$$X = \frac{T_1}{8}, Y = \frac{T_2}{8}, L = \frac{T}{8},$$

so X and Y are i.i.d. $\mathcal{N}(0,1)$, and $L = \min(X,Y)$. Also, let

$$M = \max(X,Y), S = X+Y, W = |X-Y|.$$

We have $M + L = S$ and $M - L = W$, so $M = \frac{1}{2}(S+W)$ and $L = \frac{1}{2}(S-W)$. Then

$$E(S) = 0 \text{ and } E(W) = \frac{2}{\sqrt{\pi}},$$

by Example 7.2.3. Thus,

$$E(M) = \frac{1}{2}E(W) = \frac{1}{\sqrt{\pi}}, E(L) = \frac{-1}{2}E(W) = \frac{-1}{\sqrt{\pi}}.$$

It follows that

$$E(T) = 8E(L) = \frac{-8}{\sqrt{\pi}}.$$

(b) We can find the PDF of T using order statistics results, or directly using

$$P(T > t) = P(T_1 > t, T_2 > t) = (1 - \Phi(t/8))^2.$$

So the PDF of T is

$$f(t) = \frac{1}{4}\left(1 - \Phi(t/8)\right)\varphi(t/8),$$

where $\varphi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$ is the $\mathcal{N}(0,1)$ PDF. Thus,

$$\text{Var}(T) = \int_{-\infty}^{\infty} t^2 f(t)dt - \left(\int_{-\infty}^{\infty} t f(t)dt\right)^2,$$

with f as above.

To get the variance in closed form (which was *not* requested in the problem), note that (with notation as above) $X+Y$ and $X-Y$ are independent since they are uncorrelated and $(X+Y, X-Y)$ is MVN. So $S = X+Y$ and $W = |X-Y|$ are independent. Thus,

$$\text{Var}(L) = \frac{1}{4}(\text{Var}(S) + \text{Var}(W)).$$

We have $\text{Var}(S) = 2$, and

$$\text{Var}(W) = E(X-Y)^2 - (E(W))^2 = \text{Var}(X-Y) - (E(W))^2 = 2 - \frac{4}{\pi}.$$

Therefore,

$$\text{Var}(L) = 1 - \frac{1}{\pi},$$

which shows that

$$\text{Var}(T) = 64\left(1 - \frac{1}{\pi}\right).$$

Mixed practice

53. Ⓢ A DNA sequence can be represented as a sequence of letters, where the "alphabet" has 4 letters: A,C,T,G. Suppose such a sequence is generated randomly, where the letters are independent and the probabilities of A,C,T,G are p_1, p_2, p_3, p_4 respectively.

(a) In a DNA sequence of length 115, what is the expected number of occurrences of the expression "CATCAT" (in terms of the p_j)? (Note that, for example, the expression "CATCATCAT" counts as 2 occurrences.)

(b) What is the probability that the first A appears earlier than the first C appears, as letters are generated one by one (in terms of the p_j)?

(c) For this part, assume that the p_j are unknown. Suppose we treat p_2 as a Unif$(0,1)$ r.v. before observing any data, and that then the first 3 letters observed are "CAT". Given this information, what is the probability that the next letter is C?

Solution:

(a) Let I_j be the indicator r.v. of "CATCAT" appearing starting at position j, for $1 \leq j \leq 110$. Then $E(I_j) = (p_1p_2p_3)^2$, so the expected number is $110(p_1p_2p_3)^2$.

Sanity check: The number of occurrences is between 0 and 110, so the expected value must also be between 0 and 110. If any of the letters C, A, or T is very rare, then "CATCAT" will be even more rare; this is reflected in the p_j^2 factors, which will make the expected number small if any of p_1, p_2, p_3 is small.

(b) Consider the first letter which is an A or a C (call it X; alternatively, condition on the first letter of the sequence). This gives

$$P(\text{A before C}) = P(X \text{ is A} | X \text{ is A or C}) = \frac{P(X \text{ is A})}{P(X \text{ is A or C})} = \frac{p_1}{p_1+p_2}.$$

Sanity check: The answer should be $1/2$ for $p_1 = p_2$, should go to 0 as $p_1 \to 0$, should be increasing in p_1 and decreasing in p_2, and finding $P(\text{A before C})$ by $1 - P(\text{A before C})$ should agree with finding it by swapping p_1, p_2.

(c) Let X be the number of C's in the data (so $X = 1$ is observed here). The prior is $p_2 \sim \text{Beta}(1,1)$, so the posterior is $p_2|X=1 \sim \text{Beta}(2,3)$ (by the connection between Beta and Binomial, or by Bayes' Rule). Given p_2, the indicator of the next letter being

C is Bern(p_2). So given X (but not given p_2), the probability of the next letter being C is $E(p_2|X) = \frac{2}{5}$.

Sanity check: It makes sense that the answer should be strictly in between 1/2 (the mean of the prior distribution) and 1/3 (the observed frequency of C's in the data).

54. Ⓢ Consider independent Bernoulli trials with probability p of success for each. Let X be the number of failures incurred before getting a total of r successes.

(a) Determine what happens to the distribution of $\frac{p}{1-p}X$ as $p \to 0$, using MGFs; what is the PDF of the limiting distribution, and its name and parameters if it is one we have studied?

Hint: Start by finding the Geom(p) MGF. Then find the MGF of $\frac{p}{1-p}X$, and use the fact that if the MGFs of r.v.s Y_n converge to the MGF of an r.v. Y, then the CDFs of the Y_n converge to the CDF of Y.

(b) Explain intuitively why the result of (a) makes sense.

Solution:

(a) Let $q = 1 - p$. For $G \sim \text{Geom}(p)$, the MGF is

$$E(e^{tG}) = p\sum_{k=0}^{\infty} e^{tk}q^k = p\sum_{k=0}^{\infty}(qe^t)^k = \frac{p}{1-qe^t},$$

for $qe^t < 1$. So the NBin(r, p) MGF is $\frac{p^r}{(1-qe^t)^r}$ for $qe^t < 1$. Then the MGF of $\frac{p}{1-p}X$ is

$$E(e^{\frac{tp}{q}X}) = \frac{p^r}{(1-qe^{tp/q})^r}$$

for $qe^{tp/q} < 1$. Let us first consider the limit for $r = 1$. As $p \to 0$, the numerator goes to 0 and so does the denominator (since $qe^{tp/q} \to 1e^0 = 1$). By L'Hôpital's Rule,

$$\lim_{p\to 0}\frac{p}{1-(1-p)e^{tp/(1-p)}} = \lim_{p\to 0}\frac{1}{e^{tp/(1-p)} - (1-p)t\left(\frac{1-p+p}{(1-p)^2}\right)e^{tp/(1-p)}} = \frac{1}{1-t}.$$

So for any fixed $r > 0$, as $p \to 0$ we have

$$E\left(e^{\frac{tp}{q}X}\right) = \frac{p^r}{(1-qe^{tp/q})^r} \to \frac{1}{(1-t)^r}.$$

This is the Gamma(r, 1) MGF for $t < 1$ (note also that the condition $qe^{tp/q} < 1$ is equivalent to $t < -\frac{1-p}{p}\log(1-p)$, which converges to the condition $t < 1$ since again by L'Hôpital's Rule, $\frac{-p}{\log(1-p)} \to 1$). Thus, the scaled Negative Binomial $\frac{p}{1-p}X$ converges to Gamma(r, 1) in distribution as $p \to 0$.

(b) The result of (a) makes sense intuitively since the Gamma is the continuous analogue of the Negative Binomial, just as the Exponential is the continuous analog of the Geometric. To convert from discrete to continuous, imagine performing many, many trials where each is performed very, very quickly and has a very, very low chance of success. To balance the rate of trials with the chance of success, we use the scaling $\frac{p}{q}$ since this makes $E(\frac{p}{q}X) = r$, matching the Gamma(r, 1) mean.

Chapter 9: Conditional expectation

Conditional expectation given an event

7. Ⓢ You get to choose between two envelopes, each of which contains a check for some positive amount of money. Unlike in the two-envelope paradox, it is not given that one envelope contains twice as much money as the other envelope. Instead, assume that the two values were generated independently from some distribution on the positive real numbers, with no information given about what that distribution is.

   After picking an envelope, you can open it and see how much money is inside (call this value x), and then you have the option of switching. As no information has been given about the distribution, it may seem impossible to have better than a 50% chance of picking the better envelope. Intuitively, we may want to switch if x is "small" and not switch if x is "large", but how do we define "small" and "large" in the grand scheme of all possible distributions? [The last sentence was a rhetorical question.]

   Consider the following strategy for deciding whether to switch. Generate a threshold $T \sim \text{Expo}(1)$, and switch envelopes if and only if the observed value x is less than the value of T. Show that this strategy succeeds in picking the envelope with more money with probability strictly greater than $1/2$.

   Hint: Let t be the value of T (generated by a random draw from the Expo(1) distribution). First explain why the strategy works very well if t happens to be in between the two envelope values, and does no harm in any case (i.e., there is no case in which the strategy succeeds with probability strictly less than $1/2$).

   Solution: Let a be the smaller value of the two envelopes and b be the larger value (assume $a < b$ since in the case $a = b$ it makes no difference which envelope is chosen!). Let G be the event that the strategy succeeds and A be the event that we pick the envelope with a initially. Then $P(G|A) = P(T > a) = 1 - (1 - e^{-a}) = e^{-a}$, and $P(G|A^c) = P(T \leq b) = 1 - e^{-b}$. Thus, the probability that the strategy succeeds is

$$\frac{1}{2}e^{-a} + \frac{1}{2}(1 - e^{-b}) = \frac{1}{2} + \frac{1}{2}(e^{-a} - e^{-b}) > \frac{1}{2},$$

   because $e^{-a} - e^{-b} > 0$.

10. Ⓢ A coin with probability p of Heads is flipped repeatedly. For (a) and (b), suppose that p is a known constant, with $0 < p < 1$.

    (a) What is the expected number of flips until the pattern HT is observed?

    (b) What is the expected number of flips until the pattern HH is observed?

    (c) Now suppose that p is unknown, and that we use a Beta(a, b) prior to reflect our uncertainty about p (where a and b are known constants and are greater than 2). In terms of a and b, find the corresponding answers to (a) and (b) in this setting.

    Solution:

    (a) This can be thought of as "Wait for Heads, then wait for the first Tails after the first Heads," so the expected value is $\frac{1}{p} + \frac{1}{q}$, with $q = 1 - p$.

(b) Let X be the waiting time for HH and condition on the first toss, writing H for the event that the first toss is Heads and T for the complement of H:

$$E(X) = E(X|H)p + E(X|T)q = E(X|H)p + (1 + EX)q.$$

To find $E(X|H)$, condition on the second toss:

$$E(X|H) = E(X|HH)p + E(X|HT)q = 2p + (2 + EX)q.$$

Solving for $E(X)$, we have

$$E(X) = \frac{1}{p} + \frac{1}{p^2}.$$

Sanity check: This gives $E(X) = 6$ when $p = 1/2$, in agreement with Example 9.1.9.

(c) Let X and Y be the number of flips until HH and until HT, respectively. By (a), $E(Y|p) = \frac{1}{p} + \frac{1}{1-p}$. So $E(Y) = E(E(Y|p)) = E(\frac{1}{p}) + E(\frac{1}{1-p})$. Likewise, by (b), $E(X) = E(E(X|p)) = E(\frac{1}{p}) + E(\frac{1}{p^2})$. By LOTUS,

$$E\left(\frac{1}{p}\right) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 p^{a-2}(1-p)^{b-1}dp = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a-1)\Gamma(b)}{\Gamma(a+b-1)} = \frac{a+b-1}{a-1},$$

$$E\left(\frac{1}{1-p}\right) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 p^{a-1}(1-p)^{b-2}dp = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a)\Gamma(b-1)}{\Gamma(a+b-1)} = \frac{a+b-1}{b-1},$$

$$E\left(\frac{1}{p^2}\right) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 p^{a-3}(1-p)^{b-1}dp = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a-2)\Gamma(b)}{\Gamma(a+b-2)} = \frac{(a+b-1)(a+b-2)}{(a-1)(a-2)}.$$

Thus,

$$E(Y) = \frac{a+b-1}{a-1} + \frac{a+b-1}{b-1},$$

$$E(X) = \frac{a+b-1}{a-1} + \frac{(a+b-1)(a+b-2)}{(a-1)(a-2)}.$$

Conditional expectation given a random variable

13. Ⓢ Let X_1, X_2 be i.i.d., and let $\bar{X} = \frac{1}{2}(X_1 + X_2)$ be the sample mean. In many statistics problems, it is useful or important to obtain a conditional expectation given $\bar{X}$. As an example of this, find $E(w_1X_1 + w_2X_2|\bar{X})$, where w_1, w_2 are constants with $w_1 + w_2 = 1$.

Solution: By symmetry $E(X_1|\bar{X}) = E(X_2|\bar{X})$ and by linearity and taking out what's known, $E(X_1|\bar{X}) + E(X_2|\bar{X}) = E(X_1 + X_2|\bar{X}) = X_1 + X_2$. So $E(X_1|\bar{X}) = E(X_2|\bar{X}) = \bar{X}$ (see also Example 9.3.6). Thus,

$$E(w_1X_1 + w_2X_2|\bar{X}) = w_1E(X_1|\bar{X}) + w_2E(X_2|\bar{X}) = w_1\bar{X} + w_2\bar{X} = \bar{X}.$$

15. Ⓢ Consider a group of n roommate pairs at a college (so there are $2n$ students). Each of these $2n$ students independently decides randomly whether to take a certain course, with probability p of success (where "success" is defined as taking the course).

Let N be the number of students among these $2n$ who take the course, and let X be the number of roommate pairs where both roommates in the pair take the course. Find $E(X)$ and $E(X|N)$.

Solution: Create an indicator r.v. I_j for the jth roommate pair, equal to 1 if both take the course. The expected value of such an indicator r.v. is p^2, so $E(X) = np^2$ by symmetry and linearity. Similarly, $E(X|N) = nE(I_1|N)$. We have

$$E(I_1|N) = \frac{N}{2n} \frac{N-1}{2n-1}$$

since given that N of the $2n$ students take the course, the probability is $\frac{N}{2n}$ that any particular student takes Stat 110 (the p no longer matters), and given that one particular student in a roommate pair takes the course, the probability that the other roommate does is $\frac{N-1}{2n-1}$. Or write $E(I_1|N) = \frac{\binom{N}{2}}{\binom{2n}{2}}$, since given N, the number of students in the first roommate pair who are in the course is Hypergeometric! Thus,

$$E(X|N) = nE(I_1|N) = \frac{N(N-1)}{2}\frac{1}{2n-1}.$$

Historical note: an equivalent problem was first solved in the 1760s by Daniel Bernoulli, a nephew of Jacob Bernoulli. (The Bernoulli distribution is named after Jacob Bernoulli.)

16. Ⓢ Show that $E((Y - E(Y|X))^2|X) = E(Y^2|X) - (E(Y|X))^2$, so these two expressions for $\mathrm{Var}(Y|X)$ agree.

Solution: This is the conditional version of the fact that

$$\mathrm{Var}(Y) = E((Y - E(Y))^2) = E(Y^2) - (E(Y))^2,$$

and so must be true since conditional expectations *are* expectations, just as conditional probabilities are probabilities. Algebraically, letting $g(X) = E(Y|X)$ we have

$$E((Y-E(Y|X))^2|X) = E(Y^2-2Yg(X)+g(X)^2|X) = E(Y^2|X)-2E(Yg(X)|X)+E(g(X)^2|X),$$

and $E(Yg(X)|X) = g(X)E(Y|X) = g(X)^2, E(g(X)^2|X) = g(X)^2$ by taking out what's known, so the righthand side above simplifies to $E(Y^2|X) - g(X)^2$.

22. Ⓢ Let X and Y be random variables with finite variances, and let $W = Y - E(Y|X)$. This is a *residual*: the difference between the true value of Y and the predicted value of Y based on X.

(a) Compute $E(W)$ and $E(W|X)$.

(b) Compute $\mathrm{Var}(W)$, for the case that $W|X \sim \mathcal{N}(0, X^2)$ with $X \sim \mathcal{N}(0, 1)$.

Solution:

(a) Adam's law, taking out what's known, and linearity give

$$\begin{aligned} E(W) &= EY - E(E(Y|X)) = EY - EY = 0, \\ E(W|X) &= E(Y|X) - E(E(Y|X)|X) = E(Y|X) - E(Y|X) = 0. \end{aligned}$$

(b) Eve's Law gives

$$\mathrm{Var}(W) = \mathrm{Var}(E(W|X)) + E(\mathrm{Var}(W|X)) = \mathrm{Var}(0) + E(X^2) = 0 + 1 = 1.$$

23. Ⓢ One of two identical-looking coins is picked from a hat randomly, where one coin has probability p_1 of Heads and the other has probability p_2 of Heads. Let X be the number of Heads after flipping the chosen coin n times. Find the mean and variance of X.

Solution: The distribution of X is a *mixture* of two Binomials; this is *not* Binomial unless $p_1 = p_2$. Let I be the indicator of having the p_1 coin. Then

$$E(X) = E(X|I = 1)P(I = 1) + E(X|I = 0)P(I = 0) = \frac{1}{2}n(p_1 + p_2).$$

Alternatively, we can represent X as $X = IX_1 + (1 - I)X_2$ with $X_j \sim \mathrm{Bin}(n, p_j)$, and I, X_1, X_2 independent. Then

$$E(X) = E(E(X|I)) = E(Inp_1 + (1 - I)np_2) = \frac{1}{2}n(p_1 + p_2).$$

For the variance, note that it is *not* valid to say "$\text{Var}(X) = \text{Var}(X|I=1)P(I=1) + \text{Var}(X|I=0)P(I=0)$"; an extreme example of this mistake would be claiming that "$\text{Var}(I) = 0$ since $\text{Var}(I|I=1)P(I=1) + \text{Var}(I|I=0)P(I=0) = 0$"; of course, $\text{Var}(I) = \frac{1}{4}$). Instead, we can use Eve's Law:

$$\text{Var}(X) = E(\text{Var}(X|I)) + \text{Var}(E(X|I)),$$

where $\text{Var}(X|I) = Inp_1(1-p_1) + (1-I)np_2(1-p_2)$ is $np_1(1-p_1)$ with probability $1/2$ and $np_2(1-p_2)$ with probability $1/2$, and $E(X|I) = Inp_1 + (1-I)np_2$ is np_1 or np_2 with probability $\frac{1}{2}$ each, so

$$\text{Var}(X) = \frac{1}{2}\left(np_1(1-p_1) + np_2(1-p_2)\right) + \frac{1}{4}n^2(p_1-p_2)^2.$$

27. Ⓢ We wish to estimate an unknown parameter θ, based on an r.v. X we will get to observe. As in the Bayesian perspective, assume that X and θ have a joint distribution. Let $\hat{\theta}$ be the estimator (which is a function of X). Then $\hat{\theta}$ is said to be *unbiased* if $E(\hat{\theta}|\theta) = \theta$, and $\hat{\theta}$ is said to be the *Bayes procedure* if $E(\theta|X) = \hat{\theta}$.

(a) Let $\hat{\theta}$ be unbiased. Find $E(\hat{\theta}-\theta)^2$ (the average squared difference between the estimator and the true value of θ), in terms of marginal moments of $\hat{\theta}$ and θ.

Hint: Condition on θ.

(b) Repeat (a), except in this part suppose that $\hat{\theta}$ is the *Bayes procedure* rather than assuming that it is unbiased.

Hint: Condition on X.

(c) Show that it is *impossible* for $\hat{\theta}$ to be both the Bayes procedure and unbiased, except in silly problems where we get to know θ perfectly by observing X.

Hint: If Y is a nonnegative r.v. with mean 0, then $P(Y=0) = 1$.

Solution:

(a) Conditioning on θ, we have

$$\begin{aligned} E(\hat{\theta}-\theta)^2 &= E(E(\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2|\theta)) \\ &= E(E(\hat{\theta}^2|\theta)) - E(E(2\hat{\theta}\theta|\theta)) + E(E(\theta^2|\theta)) \\ &= E(\hat{\theta}^2) - 2E(\theta E(\hat{\theta}|\theta)) + E(\theta^2) \\ &= E(\hat{\theta}^2) - 2E(\theta^2) + E(\theta^2) \\ &= E(\hat{\theta}^2) - E(\theta^2). \end{aligned}$$

(b) By the same argument as for (a) except now conditioning on X, we have

$$\begin{aligned} E(\hat{\theta}-\theta)^2 &= E(E(\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2|X)) \\ &= E(E(\hat{\theta}^2|X)) - E(E(2\hat{\theta}\theta|X)) + E(E(\theta^2|X)) \\ &= E(\hat{\theta}^2) - 2E(\hat{\theta}^2) + E(\theta^2) \\ &= E(\theta^2) - E(\hat{\theta}^2). \end{aligned}$$

(c) Suppose that $\hat{\theta}$ is both the Bayes procedure and unbiased. By the above, we have $E(\hat{\theta}-\theta)^2 = a$ and $E(\hat{\theta}-\theta)^2 = -a$, where $a = E(\hat{\theta}^2) - E(\theta^2)$. But that implies $a = 0$, which means that $\hat{\theta} = \theta$ (with probability 1). That can only happen in the extreme situation where the observed data reveal the true θ *perfectly*; in practice, nature is much more elusive and does not reveal its deepest secrets with such alacrity.

30. Ⓢ Emails arrive one at a time in an inbox. Let T_n be the time at which the nth email arrives (measured on a continuous scale from some starting point in time). Suppose that the waiting times between emails are i.i.d. Expo(λ), i.e., $T_1, T_2 - T_1, T_3 - T_2, \ldots$ are i.i.d. Expo(λ).

Each email is non-spam with probability p, and spam with probability $q = 1 - p$ (independently of the other emails and of the waiting times). Let X be the time at which the first non-spam email arrives (so X is a continuous r.v., with $X = T_1$ if the 1st email is non-spam, $X = T_2$ if the 1st email is spam but the 2nd one isn't, etc.).

(a) Find the mean and variance of X.

(b) Find the MGF of X. What famous distribution does this imply that X has (be sure to state its parameter values)?

Hint for both parts: Let N be the number of emails until the first non-spam (including that one), and write X as a sum of N terms; then condition on N.

Solution:

(a) Write $X = X_1 + X_2 + \cdots + X_N$, where X_j is the time from the $(j-1)$th to the jth email for $j \geq 2$, and $X_1 = T_1$. Then $N - 1 \sim \text{Geom}(p)$, so

$$E(X) = E(E(X|N)) = E(N\frac{1}{\lambda}) = \frac{1}{p\lambda}.$$

And

$$\text{Var}(X) = E(\text{Var}(X|N)) + \text{Var}(E(X|N)) = E(N\frac{1}{\lambda^2}) + \text{Var}(N\frac{1}{\lambda}),$$

which is

$$\frac{1}{p\lambda^2} + \frac{1-p}{p^2\lambda^2} = \frac{1}{p^2\lambda^2}.$$

(b) Again conditioning on N, the MGF is

$$E(e^{tX}) = E(E(e^{tX_1}e^{tX_2}\cdots e^{tX_N}|N)) = E\left(E(e^{tX_1}|N)E(e^{tX_2}|N)\ldots E(e^{tX_N}|N)\right) = E(M_1(t)^N),$$

where $M_1(t)$ is the MGF of X_1 (which is $\frac{\lambda}{\lambda - t}$ for $t < \lambda$). By LOTUS, this is

$$p\sum_{n=1}^{\infty} M_1(t)^n q^{n-1} = \frac{p}{q}\sum_{n=1}^{\infty}(qM_1(t))^n = \frac{p}{q}\frac{qM_1(t)}{1 - qM_1(t)} = \frac{\frac{p\lambda}{\lambda - t}}{1 - \frac{q\lambda}{\lambda - t}} = \frac{p\lambda}{p\lambda - t}$$

for $t < p\lambda$ (as we need $qM_1(t) < 1$ for the series to converge). This is the Expo($p\lambda$) MGF, so $X \sim \text{Expo}(p\lambda)$.

33. Ⓢ Judit plays in a total of $N \sim \text{Geom}(s)$ chess tournaments in her career. Suppose that in each tournament she has probability p of winning the tournament, independently. Let T be the number of tournaments she wins in her career.

(a) Find the mean and variance of T.

(b) Find the MGF of T. What is the name of this distribution (with its parameters)?

Solution:

(a) We have $T|N \sim \text{Bin}(N, p)$. By Adam's Law,

$$E(T) = E(E(T|N)) = E(Np) = p(1-s)/s.$$

By Eve's Law,

$$\begin{aligned}\mathrm{Var}(T) &= E(\mathrm{Var}(T|N)) + \mathrm{Var}(E(T|N)) \\ &= E(Np(1-p)) + \mathrm{Var}(Np) \\ &= p(1-p)(1-s)/s + p^2(1-s)/s^2 \\ &= \frac{p(1-s)(s+(1-s)p)}{s^2}.\end{aligned}$$

(b) Let $I_j \sim \mathrm{Bern}(p)$ be the indicator of Judit winning the jth tournament. Then

$$\begin{aligned}E(e^{tT}) &= E(E(e^{tT}|N)) \\ &= E((pe^t+q)^N) \\ &= s\sum_{n=0}^{\infty}(pe^t+1-p)^n(1-s)^n \\ &= \frac{s}{1-(1-s)(pe^t+1-p)}.\end{aligned}$$

This is reminiscent of the Geometric MGF, which was derived in Example 6.4.3. If $T \sim \mathrm{Geom}(\theta)$, we have $\theta = \frac{s}{s+p(1-s)}$, as found by setting $E(T) = \frac{1-\theta}{\theta}$ or by finding $\mathrm{Var}(T)/E(T)$. Writing the MGF of T as

$$E(e^{tT}) = \frac{s}{s+(1-s)p-(1-s)pe^t} = \frac{\frac{s}{s+(1-s)p}}{1-\frac{(1-s)p}{s+(1-s)p}e^t},$$

we see that $T \sim \mathrm{Geom}(\theta)$, with $\theta = \frac{s}{s+(1-s)p}$. Note that this is consistent with (a).

The distribution of T can also be obtained by a story proof. Imagine that just before each tournament she may play in, Judit retires with probability s (if she retires, she does not play in that or future tournaments). Her tournament history can be written as a sequence of W (win), L (lose), R (retire), ending in the first R, where the probabilities of W, L, R are $(1-s)p, (1-s)(1-p), s$ respectively. For calculating T, the losses can be ignored: we want to count the number of W's before the R. The probability that a result is R given that it is W or R is $\frac{s}{s+(1-s)p}$, so we again have

$$T \sim \mathrm{Geom}(\frac{s}{s+(1-s)p}).$$

36. Ⓢ A certain stock has low volatility on some days and high volatility on other days. Suppose that the probability of a low volatility day is p and of a high volatility day is $q = 1-p$, and that on low volatility days the percent change in the stock price is $\mathcal{N}(0, \sigma_1^2)$, while on high volatility days the percent change is $\mathcal{N}(0, \sigma_2^2)$, with $\sigma_1 < \sigma_2$.

Let X be the percent change of the stock on a certain day. The distribution is said to be a *mixture* of two Normal distributions, and a convenient way to represent X is as $X = I_1X_1 + I_2X_2$ where I_1 is the indicator r.v. of having a low volatility day, $I_2 = 1-I_1$, $X_j \sim \mathcal{N}(0, \sigma_j^2)$, and I_1, X_1, X_2 are independent.

(a) Find $\mathrm{Var}(X)$ in two ways: using Eve's law, and by calculating $\mathrm{Cov}(I_1X_1 + I_2X_2, I_1X_1 + I_2X_2)$ directly.

(b) Recall from Chapter 6 that the *kurtosis* of an r.v. Y with mean μ and standard deviation σ is defined by

$$\mathrm{Kurt}(Y) = \frac{E(Y-\mu)^4}{\sigma^4} - 3.$$

Find the kurtosis of X (in terms of $p, q, \sigma_1^2, \sigma_2^2$, fully simplified). The result will show that

even though the kurtosis of any Normal distribution is 0, the kurtosis of X is positive and in fact can be very large depending on the parameter values.

Solution:

(a) By Eve's Law,

$\mathrm{Var}(X) = E(\mathrm{Var}(X|I_1))+\mathrm{Var}(E(X|I_1)) = E(I_1^2\sigma_1^2+(1-I_1)^2\sigma_2^2)+\mathrm{Var}(0) = p\sigma_1^2+(1-p)\sigma_2^2$,

since $I_1^2 = I_1, I_2^2 = I_2$. For the covariance method, expand

$\mathrm{Var}(X) = \mathrm{Cov}(I_1X_1+I_2X_2, I_1X_1+I_2X_2) = \mathrm{Var}(I_1X_1)+\mathrm{Var}(I_2X_2)+2\mathrm{Cov}(I_1X_1, I_2X_2)$.

Then $\mathrm{Var}(I_1X_1) = E(I_1^2X_1^2) - (E(I_1X_1))^2 = E(I_1)E(X_1^2) = p\mathrm{Var}(X_1)$ since $E(I_1X_1) = E(I_1)E(X_1) = 0$. Similarly, $\mathrm{Var}(I_2X_2) = (1-p)\mathrm{Var}(X_2)$. And

$$\mathrm{Cov}(I_1X_1, I_2X_2) = E(I_1I_2X_1X_2) - E(I_1X_1)E(I_2X_2) = 0,$$

since I_1I_2 always equals 0. So again we have $\mathrm{Var}(X) = p\sigma_1^2 + (1-p)\sigma_2^2$.

(b) Note that $(I_1X_1 + I_2X_2)^4 = I_1X_1^4 + I_2X_2^4$ since the cross terms disappear (because I_1I_2 is always 0) and any positive power of an indicator r.v. is that indicator r.v.! So

$$E(X^4) = E(I_1X_1^4 + I_2X_2^4) = 3p\sigma_1^4 + 3q\sigma_2^4.$$

Alternatively, we can use $E(X^4) = E(X^4|I_1 = 1)p + E(X^4|I_1 = 0)q$ to find $E(X^4)$. The mean of X is $E(I_1X_1) + E(I_2X_2) = 0$, so the kurtosis of X is

$$\mathrm{Kurt}(X) = \frac{3p\sigma_1^4 + 3q\sigma_2^4}{(p\sigma_1^2 + q\sigma_2^2)^2} - 3.$$

This becomes 0 if $\sigma_1 = \sigma_2$, since then we have a Normal distribution rather than a mixture of two different Normal distributions. For $\sigma_1 < \sigma_2$, the kurtosis is positive since

$$p\sigma_1^4 + q\sigma_2^4 > (p\sigma_1^2 + q\sigma_2^2)^2,$$

as can be seen by interpreting this as saying $E(Y^2) > (EY)^2$, where Y is σ_1^2 with probability p and σ_2^2 with probability q.

Mixed practice

43. Ⓢ Empirically, it is known that 49% of children born in the U.S. are girls (and 51% are boys). Let N be the number of children who will be born in the U.S. in March of next year, and assume that N is a Pois(λ) random variable, where λ is known. Assume that births are independent (e.g., don't worry about identical twins).
Let X be the number of girls who will be born in the U.S. in March of next year, and let Y be the number of boys who will be born then.

(a) Find the joint distribution of X and Y. (Give the joint PMF.)

(b) Find $E(N|X)$ and $E(N^2|X)$.

Solution:

(a) By the chicken-egg story, X and Y are independent with $X \sim \mathrm{Pois}(0.49\lambda), Y \sim \mathrm{Pois}(0.51\lambda)$. The joint PMF is

$$P(X = i, Y = j) = (e^{-0.49\lambda}(0.49\lambda)^i/i!)(e^{-0.51\lambda}(0.51\lambda)^j/j!).$$

(b) Since X and Y are independent,

$$E(N|X) = E(X+Y|X) = X + E(Y|X) = X + EY = X + 0.51\lambda,$$

$$E(N^2|X) = E(X^2 + 2XY + Y^2|X) = X^2 + 2XE(Y) + E(Y^2) = (X + 0.51\lambda)^2 + 0.51\lambda.$$

44. Ⓢ Let X_1, X_2, X_3 be independent with $X_i \sim \text{Expo}(\lambda_i)$ (so with possibly different rates). Recall from Chapter 7 that

$$P(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

(a) Find $E(X_1 + X_2 + X_3 | X_1 > 1, X_2 > 2, X_3 > 3)$ in terms of $\lambda_1, \lambda_2, \lambda_3$.

(b) Find $P\left(X_1 = \min(X_1, X_2, X_3)\right)$, the probability that the first of the three Exponentials is the smallest.

Hint: Restate this in terms of X_1 and $\min(X_2, X_3)$.

(c) For the case $\lambda_1 = \lambda_2 = \lambda_3 = 1$, find the PDF of $\max(X_1, X_2, X_3)$. Is this one of the important distributions we have studied?

Solution:

(a) By linearity, independence, and the memoryless property, we get

$$E(X_1|X_1 > 1) + E(X_2|X_2 > 2) + E(X_3|X_3 > 3) = \lambda_1^{-1} + \lambda_2^{-1} + \lambda_3^{-1} + 6.$$

(b) The desired probability is $P(X_1 \leq \min(X_2, X_3))$. Noting that $\min(X_2, X_3) \sim \text{Expo}(\lambda_2 + \lambda_3)$ is independent of X_1, we have

$$P(X_1 \leq \min(X_2, X_3)) = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3}.$$

(c) Let $M = \max(X_1, X_2, X_3)$. Using the order statistics results from class or by directly computing the CDF and taking the derivative, for $x > 0$ we have

$$f_M(x) = 3(1 - e^{-x})^2 e^{-x}.$$

This is not one of the important distributions we have studied. The form is reminiscent of a Beta, but a Beta takes values between 0 and 1, while M can take any positive real value (in fact, $B \sim \text{Beta}(1, 3)$ if we make the transformation $B = e^{-M}$.)

45. Ⓢ A task is randomly assigned to one of two people (with probability 1/2 for each person). If assigned to the first person, the task takes an $\text{Expo}(\lambda_1)$ length of time to complete (measured in hours), while if assigned to the second person it takes an $\text{Expo}(\lambda_2)$ length of time to complete (independent of how long the first person would have taken). Let T be the time taken to complete the task.

(a) Find the mean and variance of T.

(b) Suppose instead that the task is assigned to *both* people, and let X be the time taken to complete it (by whoever completes it first, with the two people working independently). It is observed that after 24 hours, the task has not yet been completed. Conditional on this information, what is the expected value of X?

Solution: Write $T = IX_1 + (1 - I)X_2$, with $I \sim \text{Bern}(1/2), X_1 \sim \text{Expo}(\lambda_1), X_2 \sim \text{Expo}(\lambda_2)$ independent. Then

$$ET = \frac{1}{2}(\lambda_1^{-1} + \lambda_2^{-1}),$$

$$\begin{aligned}
\text{Var}(T) &= E(\text{Var}(T|I)) + \text{Var}(E(T|I)) \\
&= E(I^2 \frac{1}{\lambda_1^2} + (1 - I)^2 \frac{1}{\lambda_2^2}) + \text{Var}\left(\frac{I}{\lambda_1} + \frac{1 - I}{\lambda_2}\right) \\
&= E(I \frac{1}{\lambda_1^2} + (1 - I)\frac{1}{\lambda_2^2}) + \text{Var}\left(I(\frac{1}{\lambda_1} - \frac{1}{\lambda_2})\right) \\
&= \frac{1}{2}(\frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2}) + \frac{1}{4}(\frac{1}{\lambda_1} - \frac{1}{\lambda_2})^2.
\end{aligned}$$

Sanity check: For $\lambda_1 = \lambda_2$, the two people have the same distribution so randomly assigning the task to one of the two should be equivalent to just assigning it to the first person (so the mean and variance should agree with those of an Expo(λ_1) r.v.). It makes sense that the mean is the average of the two means, as we can condition on whether $I = 1$ (though the variance is *greater* than the average of the two variances, by Eve's Law). Also, the results should be (and are) the same if we swap λ_1 and λ_2.

(b) Here $X = \min(X_1, X_2)$ with $X_1 \sim \text{Expo}(\lambda_1), X_2 \sim \text{Expo}(\lambda_2)$ independent. Then $X \sim \text{Expo}(\lambda_1 + \lambda_2)$ (since $P(X > x) = P(X_1 > x)P(X_2 > x) = e^{-(\lambda_1+\lambda_2)x}$, or by results on order statistics). By the memoryless property,

$$E(X|X > 24) = 24 + \frac{1}{\lambda_1 + \lambda_2}.$$

Sanity check: The answer should be greater than 24 and should be very close to 24 if λ_1 or λ_2 is very large. Considering a Poisson process also helps make this intuitive.

47. Ⓢ A certain genetic characteristic is of interest. It can be measured numerically. Let X_1 and X_2 be the values of the genetic characteristic for two twin boys. If they are identical twins, then $X_1 = X_2$ and X_1 has mean 0 and variance σ^2; if they are fraternal twins, then X_1 and X_2 have mean 0, variance σ^2, and correlation ρ. The probability that the twins are identical is 1/2. Find $\text{Cov}(X_1, X_2)$ in terms of ρ, σ^2.

Solution: Since the means are 0, $\text{Cov}(X_1, X_2) = E(X_1X_2) - (EX_1)(EX_2) = E(X_1X_2)$. Now condition on whether the twins are identical or fraternal:

$$E(X_1X_2) = E(X_1X_2|\text{identical})\frac{1}{2} + E(X_1X_2|\text{fraternal})\frac{1}{2} = E(X_1^2)\frac{1}{2} + \rho\sigma^2\frac{1}{2} = \frac{\sigma^2}{2}(1+\rho).$$

48. Ⓢ The Mass Cash lottery randomly chooses 5 of the numbers from $1, 2, \ldots, 35$ each day (without repetitions within the choice of 5 numbers). Suppose that we want to know how long it will take until all numbers have been chosen. Let a_j be the average number of additional days needed if we are missing j numbers (so $a_0 = 0$ and a_{35} is the average number of days needed to collect all 35 numbers). Find a recursive formula for the a_j.

Solution: Suppose we are missing j numbers (with $0 \leq j \leq 35$), and let T_j be the additional number of days needed to complete the collection. Condition on how many "new" numbers appear the next day; call this N. This gives

$$E(T_j) = \sum_{n=0}^{5} E(T_j|N = n)P(N = n).$$

Note that N is Hypergeometric (imagine tagging the numbers that we don't already have in our collection)! Letting $a_k = 0$ for $k < 0$, we have

$$a_j = 1 + \sum_{n=0}^{5} \frac{a_{j-n}\binom{j}{n}\binom{35-j}{5-n}}{\binom{35}{5}}.$$

Chapter 10: Inequalities and limit theorems

Inequalities

1. Ⓢ In a national survey, a random sample of people are chosen and asked whether they support a certain policy. Assume that everyone in the population is equally likely to be surveyed at each step, and that the sampling is with replacement (sampling without replacement is typically more realistic, but with replacement will be a good approximation if the sample size is small compared to the population size). Let n be the sample size, and let $\hat{p}$ and p be the proportion of people who support the policy in the sample and in the entire population, respectively. Show that for every $c > 0$,

$$P(|\hat{p} - p| > c) \leq \frac{1}{4nc^2}.$$

Solution: We can write $\hat{p} = X/n$ with $X \sim \text{Bin}(n, p)$. So $E(\hat{p}) = p, \text{Var}(\hat{p}) = p(1-p)/n$. Then by Chebyshev's inequality,

$$P(|\hat{p} - p| > c) \leq \frac{\text{Var}(\hat{p})}{c^2} = \frac{p(1-p)}{nc^2} \leq \frac{1}{4nc^2},$$

where the last inequality is because $p(1-p)$ is maximized at $p = 1/2$.

2. Ⓢ For i.i.d. r.v.s $X_1, \ldots, X_n$ with mean μ and variance σ^2, give a value of n (as a specific number) that will ensure that there is at least a 99% chance that the sample mean will be within 2 standard deviations of the true mean μ.

Solution: We have to find n such that

$$P(|\overline{X}_n - \mu| > 2\sigma) \leq 0.01.$$

By Chebyshev's inequality (in the form $P(|Y - EY| > c) \leq \frac{\text{Var}(Y)}{c^2}$), we have

$$P(|\overline{X}_n - \mu| > 2\sigma) \leq \frac{\text{Var}\bar{X}_n}{(2\sigma)^2} = \frac{\frac{\sigma^2}{n}}{4\sigma^2} = \frac{1}{4n}.$$

So the desired inequality holds if $n \geq 25$.

3. Ⓢ Show that for any two positive r.v.s X and Y with neither a constant multiple of the other,

$$E(X/Y)E(Y/X) > 1.$$

Solution: The r.v. $W = Y/X$ is positive and non-constant, so Jensen's inequality yields

$$E(X/Y) = E(1/W) > 1/E(W) = 1/E(Y/X).$$

4. Ⓢ The famous *arithmetic mean-geometric mean* inequality says that for any positive numbers $a_1, a_2, \ldots, a_n$,

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}.$$

Show that this inequality follows from Jensen's inequality, by considering $E\log(X)$ for an r.v. X whose possible values are $a_1, \ldots, a_n$ (you should specify the PMF of X; if you want, you can assume that the a_j are distinct (no repetitions), but be sure to say so if you assume this).

Solution: Assume that the a_j are distinct, and let X be a random variable which takes values from $a_1, a_2, ..., a_n$ with equal probability (the case of repeated a_j's can be handled similarly, letting the probability of $X = a_j$ be m_j/n, where m_j is the number of times a_j appears in the list $a_1, \ldots, a_n$). Jensen's inequality gives $E(\log X) \leq \log(EX)$, since the log function is concave. The left-hand side is $\frac{1}{n}\sum_{i=1}^{n} \log a_i$, while the right hand-side is $\log \frac{a_1+a_2+\cdots+a_n}{n}$. So we have the following inequality:

$$\log \frac{a_1 + a_2 + \cdots + a_n}{n} \geq \frac{1}{n}\sum_{i=1}^{n} \log a_i$$

Thus,

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq e^{\frac{1}{n}\sum_{i=1}^{n} \log a_i} = e^{\frac{\log(a_1 \cdots a_n)}{n}} = (a_1 \cdots a_n)^{1/n}.$$

5. Ⓢ Let X be a discrete r.v. whose distinct possible values are $x_0, x_1, \ldots$, and let $p_k = P(X = x_k)$. The entropy of X is $H(X) = \sum_{k=0}^{\infty} p_k \log_2(1/p_k)$.

(a) Find $H(X)$ for $X \sim \text{Geom}(p)$.

Hint: Use properties of logs, and interpret part of the sum as an expected value.

(b) Let X and Y be i.i.d. discrete r.v.s. Show that $P(X = Y) \geq 2^{-H(X)}$.

Hint: Consider $E(\log_2(W))$, where W is an r.v. taking value p_k with probability p_k.

Solution:

(a) We have

$$\begin{aligned} H(X) &= -\sum_{k=0}^{\infty} (pq^k) \log_2(pq^k) \\ &= -\log_2(p) \sum_{k=0}^{\infty} pq^k - \log_2(q) \sum_{k=0}^{\infty} kpq^k \\ &= -\log_2(p) - \frac{q}{p}\log_2(q), \end{aligned}$$

with $q = 1 - p$, since the first series is the sum of a Geom(p) PMF and the second series is the expected value of a Geom(p) r.v.

(b) Let W be as in the hint. By Jensen, $E(\log_2(W)) \leq \log_2(EW)$. But

$$E(\log_2(W)) = \sum_k p_k \log_2(p_k) = -H(X),$$

$$EW = \sum_k p_k^2 = P(X = Y),$$

so $-H(X) \leq \log_2 P(X = Y)$. Thus, $P(X = Y) \geq 2^{-H(X)}$.

Fill-in-the-blank inequalities

7. Ⓢ Let X and Y be i.i.d. positive r.v.s, and let $c > 0$. For each part below, fill in the appropriate equality or inequality symbol: write $=$ if the two sides are always equal,

$\leq$ if the left-hand side is less than or equal to the right-hand side (but they are not necessarily equal), and similarly for $\geq$. If no relation holds in general, write ?.

(a) $E(\ln(X))$ ____ $\ln(E(X))$

(b) $E(X)$ ____ $\sqrt{E(X^2)}$

(c) $E(\sin^2(X)) + E(\cos^2(X))$ ____ 1

(d) $E(|X|)$ ____ $\sqrt{E(X^2)}$

(e) $P(X > c)$ ____ $\frac{E(X^3)}{c^3}$

(f) $P(X \leq Y)$ ____ $P(X \geq Y)$

(g) $E(XY)$ ____ $\sqrt{E(X^2)E(Y^2)}$

(h) $P(X + Y > 10)$ ____ $P(X > 5 \text{ or } Y > 5)$

(i) $E(\min(X, Y))$ ____ $\min(EX, EY)$

(j) $E(X/Y)$ ____ $\frac{EX}{EY}$

(k) $E(X^2(X^2 + 1))$ ____ $E(X^2(Y^2 + 1))$

(l) $E\left(\frac{X^3}{X^3+Y^3}\right)$ ____ $E(\frac{Y^3}{X^3+Y^3})$

Solution:

(a) $E(\ln(X)) \leq \ln(E(X))$ (by Jensen: logs are concave)

(b) $E(X) \leq \sqrt{E(X^2)}$ (since $\text{Var}(X) \geq 0$, or by Jensen)

(c) $E(\sin^2(X)) + E(\cos^2(X)) = 1$ (by linearity, trig identity)

(d) $E(|X|) \leq \sqrt{E(X^2)}$ (by (b) with $|X|$ in place of X; here $|X| = X$ anyway)

(e) $P(X > c) \leq \frac{E(X^3)}{c^3}$ (by Markov, after cubing both sides of $X > c$)

(f) $P(X \leq Y) = P(X \geq Y)$ (by symmetry, as X, Y are i.i.d.)

(g) $E(XY) \leq \sqrt{E(X^2)E(Y^2)}$ (by Cauchy-Schwarz)

(h) $P(X + Y > 10) \leq P(X > 5 \text{ or } Y > 5)$ (if $X + Y > 10$, then $X > 5$ or $Y > 5$)

(i) $E(\min(X, Y)) \leq \min(EX, EY)$ (since $\min(X, Y) \leq X$ gives $E\min(X, Y) \leq EX$, and similarly $E\min(X, Y) \leq EY$)

(j) $E(X/Y) \geq \frac{EX}{EY}$ (since $E(X/Y) = E(X)E(\frac{1}{Y})$, with $E(\frac{1}{Y}) \geq \frac{1}{EY}$ by Jensen)

(k) $E(X^2(X^2 + 1)) \geq E(X^2(Y^2 + 1))$ (since $E(X^4) \geq (EX^2)^2 = E(X^2)E(Y^2) = E(X^2Y^2)$, because X^2 and Y^2 are i.i.d. and independent implies uncorrelated)

(l) $E(\frac{X^3}{X^3+Y^3}) = E(\frac{Y^3}{X^3+Y^3})$ (by symmetry!)

8. Ⓢ Write the most appropriate of $\leq$, $\geq$, $=$, or ? in the blank for each part (where "?" means that no relation holds in general).

In (c) through (f), X and Y are i.i.d. (independent identically distributed) positive random variables. Assume that the various expected values exist.

(a) (probability that a roll of 2 fair dice totals 9) ____ (probability that a roll of 2 fair dice totals 10)

(b) (probability that at least 65% of 20 children born are girls) ___ (probability that at least 65% of 2000 children born are girls)

(c) $E(\sqrt{X})$ ___ $\sqrt{E(X)}$

(d) $E(\sin X)$ ___ $\sin(EX)$

(e) $P(X+Y>4)$ ___ $P(X>2)P(Y>2)$

(f) $E\left((X+Y)^2\right)$ ___ $2E(X^2)+2(EX)^2$

Solution:

(a) (probability that a roll of 2 fair dice totals 9) $\geq$ (probability that a roll of 2 fair dice totals 10)

The probability on the left is 4/36 and that on the right is 3/36 as there is only one way for both dice to show 5's.

(b) (probability that at least 65% of 20 children born are girls) $\geq$ (probability that at least 65% of 2000 children born are girls)

With a large number of births, by LLN it becomes likely that the fraction that are girls is close to 1/2.

(c) $E(\sqrt{X}) \leq \sqrt{E(X)}$

By Jensen's inequality (or since $\mathrm{Var}(\sqrt{X}) \geq 0$).

(d) $E(\sin X)$? $\sin(EX)$

The inequality can go in either direction. For example, let X be 0 or π with equal probabilities. Then $E(\sin X) = 0$, $\sin(EX) = 1$. But if we let X be $\pi/2$ or $5\pi/2$ with equal probabilities, then $E(\sin X) = 1$, $\sin(EX) = -1$.

(e) $P(X+Y>4) \geq P(X>2)P(Y>2)$

The righthand side is $P(X>2, Y>2)$ by independence. The $\geq$ then holds since the event $X>2, Y>2$ is a subset of the event $X+Y>4$.

(f) $E\left((X+Y)^2\right) = 2E(X^2)+2(EX)^2$

The lefthand side is

$$E(X^2)+E(Y^2)+2E(XY) = E(X^2)+E(Y^2)+2E(X)E(Y) = 2E(X^2)+2(EX)^2$$

since X and Y are i.i.d.

10. Ⓢ Let X and Y be positive random variables, *not necessarily independent*. Assume that the various expected values below exist. Write the most appropriate of $\leq$, $\geq$, $=$, or ? in the blank for each part (where "?" means that no relation holds in general).

(a) $(E(XY))^2$ ___ $E(X^2)E(Y^2)$

(b) $P(|X+Y|>2)$ ___ $\frac{1}{10}E((X+Y)^4)$

(c) $E(\ln(X+3))$ ___ $\ln(E(X+3))$

(d) $E(X^2e^X)$ ___ $E(X^2)E(e^X)$

(e) $P(X+Y=2)$ ___ $P(X=1)P(Y=1)$

(f) $P(X+Y=2)$ ___ $P(\{X\geq 1\}\cup\{Y\geq 1\})$

Solution:

(a) $(E(XY))^2 \leq E(X^2)E(Y^2)$ (by Cauchy-Schwarz)

(b) $P(|X+Y|>2) \leq \frac{1}{10}E((X+Y)^4)$ (by Markov's inequality)

(c) $E(\ln(X+3)) \leq \ln(E(X+3))$ (by Jensen)

(d) $E(X^2e^X) \geq E(X^2)E(e^X)$ (since X^2 and e^X are positively correlated)

(e) $P(X+Y=2)$? $P(X=1)P(Y=1)$ (What if X, Y are independent? What if $X \sim \text{Bern}(1/2)$ and $Y = 1-X$?)

(f) $P(X+Y=2) \leq P(\{X \geq 1\} \cup \{Y \geq 1\})$ (the left event is a subset of the right event)

11. Ⓢ Let X and Y be positive random variables, *not necessarily independent*. Assume that the various expected values below exist. Write the most appropriate of $\leq$, $\geq$, $=$, or ? in the blank for each part (where "?" means that no relation holds in general).
(a) $E(X^3)$ ____ $\sqrt{E(X^2)E(X^4)}$

(b) $P(|X+Y|>2)$ ____ $\frac{1}{16}E((X+Y)^4)$

(c) $E(\sqrt{X+3})$ ____ $\sqrt{E(X+3)}$

(d) $E(\sin^2(X)) + E(\cos^2(X))$ ____ 1

(e) $E(Y|X+3)$ ____ $E(Y|X)$

(f) $E(E(Y^2|X))$ ____ $(EY)^2$

Solution:

(a) $E(X^3) \leq \sqrt{E(X^2)E(X^4)}$ (by Cauchy-Schwarz)

(b) $P(|X+Y|>2) \leq \frac{1}{16}E((X+Y)^4)$ (by Markov, taking 4th powers first)

(c) $E(\sqrt{X+3}) \leq \sqrt{E(X+3)}$ (by Jensen with a concave function)

(d) $E(\sin^2(X)) + E(\cos^2(X)) = 1$ (by linearity)

(e) $E(Y|X+3) = E(Y|X)$ (since knowing $X+3$ is equivalent to knowing X)

(f) $E(E(Y^2|X)) \geq (EY)^2$ (by Adam's law and Jensen)

12. Ⓢ Let X and Y be positive random variables, *not necessarily independent*. Assume that the various expressions below exist. Write the most appropriate of $\leq$, $\geq$, $=$, or ? in the blank for each part (where "?" means that no relation holds in general).
(a) $P(X+Y>2)$ ____ $\frac{EX+EY}{2}$

(b) $P(X+Y>3)$ ____ $P(X>3)$

(c) $E(\cos(X))$ ____ $\cos(EX)$

(d) $E(X^{1/3})$ ____ $(EX)^{1/3}$

(e) $E(X^Y)$ ____ $(EX)^{EY}$

(f) $E(E(X|Y) + E(Y|X))$ ____ $EX + EY$

Solution:

(a) $P(X+Y>2) \leq \frac{EX+EY}{2}$ (by Markov and linearity)

(b) $P(X+Y>3) \geq P(X>3)$ (since $X>3$ implies $X+Y>3$ since $Y>0$)

(c) $E(\cos(X))$?$\cos(EX)$ (e.g., let $W \sim \text{Bern}(1/2)$ and $X = aW + b$ for various a, b)

(d) $E(X^{1/3}) \leq (EX)^{1/3}$ (by Jensen)

(e) $E(X^Y)$?$(EX)^{EY}$ (take X constant or Y constant as examples)

(f) $E\left(E(X|Y) + E(Y|X)\right) = EX + EY$ (by linearity and Adam's law)

13. Ⓢ Let X and Y be i.i.d. positive random variables. Assume that the various expressions below exist. Write the most appropriate of $\leq$, $\geq$, $=$, or ? in the blank for each part (where "?" means that no relation holds in general).
(a) $E(e^{X+Y})$ ____ $e^{2E(X)}$

(b) $E(X^2e^X)$ ____ $\sqrt{E(X^4)E(e^{2X})}$

(c) $E(X|3X)$ ____ $E(X|2X)$

(d) $E(X^7Y)$ ____ $E(X^7E(Y|X))$

(e) $E(\frac{X}{Y} + \frac{Y}{X})$ ____ 2

(f) $P(|X - Y| > 2)$ ____ $\frac{\text{Var}(X)}{2}$

Solution:

(a) $E(e^{X+Y}) \geq e^{2E(X)}$ (write $E(e^{X+Y}) = E(e^Xe^Y) = E(e^X)E(e^Y) = E(e^X)E(e^X)$ using the fact that X, Y are i.i.d., and then apply Jensen)

(b) $E(X^2e^X) \leq \sqrt{E(X^4)E(e^{2X})}$ (by Cauchy-Schwarz)

(c) $E(X|3X) = E(X|2X)$ (knowing $2X$ is equivalent to knowing $3X$)

(d) $E(X^7Y) = E(X^7E(Y|X))$ (by Adam's law and taking out what's known)

(e) $E(\frac{X}{Y} + \frac{Y}{X}) \geq 2$ (since $E(\frac{X}{Y}) = E(X)E(\frac{1}{Y}) \geq \frac{EX}{EY} = 1$, and similarly $E(\frac{Y}{X}) \geq 1$)

(f) $P(|X - Y| > 2) \leq \frac{\text{Var}(X)}{2}$ (by Chebyshev, applied to the r.v. $W = X - Y$, which has variance $2\text{Var}(X)$: $P(|W - E(W)| > 2) \leq \text{Var}(W)/4 = \text{Var}(X)/2$)

14. Ⓢ Let X and Y be i.i.d. Gamma($\frac{1}{2}, \frac{1}{2}$), and let $Z \sim \mathcal{N}(0, 1)$ (note that X and Z may be dependent, and Y and Z may be dependent). For (a),(b),(c), write the most appropriate of $<$, $>$, $=$, or ? in each blank; for (d),(e),(f), write the most appropriate of $\leq$, $\geq$, $=$, or ? in each blank.
(a) $P(X < Y)$ ____ 1/2

(b) $P(X = Z^2)$ ____ 1

(c) $P(Z \geq \frac{1}{X^4+Y^4+7})$ ____ 1

(d) $E(\frac{X}{X+Y})E((X + Y)^2)$ ____ $E(X^2) + (E(X))^2$

(e) $E(X^2Z^2)$ ____ $\sqrt{E(X^4)E(X^2)}$

(f) $E((X + 2Y)^4)$ ____ 3^4

Solution:

(a) $P(X < Y) = 1/2$

This is since X and Y are i.i.d. continuous r.v.s.

(b) $P(X = Z^2)$? 1

This is since the probability is 0 if X and Z are independent, but it is 1 if X and Z^2 are the same r.v., which is possible since $Z^2 \sim \chi_1^2$, so $Z^2 \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2})$.

(c) $P(Z \geq \frac{1}{X^4+Y^4+7}) < 1$

This is since Z may be negative, and $\frac{1}{X^4+Y^4+7}$ is positive.

(d) $E(\frac{X}{X+Y})E((X+Y)^2) = E(X^2) + (E(X))^2$

By the bank–post office story, $X/(X+Y)$ and $(X+Y)^2$ are independent (and thus uncorrelated). So since X and Y are i.i.d., the lefthand side becomes

$$E(X(X+Y)) = E(X^2+XY) = E(X^2) + E(XY) = E(X^2) + (E(X))^2.$$

(e) $E(X^2Z^2) \leq \sqrt{E(X^4)E(X^2)}$

By Cauchy-Schwarz, $E(X^2Z^2) \leq \sqrt{E(X^4)E(Z^4)}$. And $E(Z^4) = E(X^2)$ since X and Z^2 are χ_1^2, or since $E(Z^4) = 3$ (as shown in Chapter 6) and $E(X^2) = \text{Var}(X) + (E(X))^2 = 2 + 1 = 3$.

(f) $E((X+2Y)^4) \geq 3^4$

This is true by Jensen's inequality, since $E(X+2Y) = 1 + 2 = 3$.

LLN and CLT

17. Ⓢ Let $X_1, X_2, \ldots$ be i.i.d. positive random variables with mean 2. Let $Y_1, Y_2, \ldots$ be i.i.d. positive random variables with mean 3. Show that

$$\frac{X_1 + X_2 + \cdots + X_n}{Y_1 + Y_2 + \cdots + Y_n} \to \frac{2}{3}$$

with probability 1. Does it matter whether the X_i are independent of the Y_j?

Solution: By the law of large numbers,

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to 2$$

with probability 1 and

$$\frac{Y_1 + Y_2 + \cdots + Y_n}{n} \to 3$$

with probability 1, as $n \to \infty$. Note that if two events A and B both have probability 1, then the event $A \cap B$ also has probability 1. So with probability 1, *both* the convergence involving the X_i and the convergence involving the Y_j occur. Therefore,

$$\frac{X_1 + X_2 + \cdots + X_n}{Y_1 + Y_2 + \cdots + Y_n} = \frac{(X_1 + X_2 + \cdots + X_n)/n}{(Y_1 + Y_2 + \cdots + Y_n)/n} \to \frac{2}{3} \text{ with probability 1}$$

as $n \to \infty$. It was not necessary to assume that the X_i are independent of the Y_j because of the pointwise with probability 1 convergence.

18. Ⓢ Let $U_1, U_2, \ldots, U_{60}$ be i.i.d. Unif(0,1) and $X = U_1 + U_2 + \cdots + U_{60}$.

(a) Which important distribution is the distribution of X very close to? Specify what the parameters are, and state which theorem justifies your choice.

(b) Give a simple but accurate approximation for $P(X > 17)$. Justify briefly.

Solution:

(a) By the central limit theorem, the distribution is approximately $\mathcal{N}(30, 5)$ since $E(X) = 30, \text{Var}(X) = 60/12 = 5$.

(b) We have

$$P(X > 17) = 1 - P(X \leq 17) = 1 - P\left(\frac{X-30}{\sqrt{5}} \leq \frac{-13}{\sqrt{5}}\right) \approx 1 - \Phi\left(\frac{-13}{\sqrt{5}}\right) = \Phi\left(\frac{13}{\sqrt{5}}\right).$$

Since $13/\sqrt{5} > 5$, and we already have $\Phi(3) \approx 0.9985$ by the 68-95-99.7% rule, the value is extremely close to 1.

19. Ⓢ Let $V_n \sim \chi^2_n$ and $T_n \sim t_n$ for all positive integers n.

(a) Find numbers a_n and b_n such that $a_n(V_n - b_n)$ converges in distribution to $\mathcal{N}(0, 1)$.

(b) Show that $T_n^2/(n + T_n^2)$ has a Beta distribution (without using calculus).

Solution:

(a) By definition of χ^2_n, we can take $V_n = Z_1^2 + \cdots + Z_n^2$, where $Z_j \sim \mathcal{N}(0, 1)$ independently. We have $E(Z_1^2) = 1$ and $E(Z_1^4) = 3$, so $\text{Var}(Z_1^2) = 2$. By the CLT, if we standardize V_n it will go to $\mathcal{N}(0, 1)$:

$$\frac{Z_1^2 + \cdots + Z_n^2 - n}{\sqrt{2n}} \to \mathcal{N}(0, 1) \text{ in distribution.}$$

So we can take $a_n = \frac{1}{\sqrt{2n}}$, $b_n = n$.

(b) We can take $T_n = Z_0/\sqrt{V_n/n}$, with $Z_0 \sim \mathcal{N}(0, 1)$ independent of V_n. Then we have $T_n^2/(n+T_n^2) = Z_0^2/(Z_0^2+V_n)$, with $Z_0^2 \sim \text{Gamma}(1/2, 1/2), V_n \sim \text{Gamma}(n/2, 1/2)$. By the bank–post office story, $Z_0^2/(Z_0^2 + V_n) \sim \text{Beta}(1/2, n/2)$.

20. Ⓢ Let $T_1, T_2, \ldots$ be i.i.d. Student-t r.v.s with $m \geq 3$ degrees of freedom. Find constants a_n and b_n (in terms of m and n) such that $a_n(T_1 + T_2 + \cdots + T_n - b_n)$ converges to $\mathcal{N}(0, 1)$ in distribution as $n \to \infty$.

Solution: First let us find the mean and variance of each T_j. Let $T = \frac{Z}{\sqrt{V/m}}$ with $Z \sim \mathcal{N}(0, 1)$ independent of $V \sim \chi^2_m$. By LOTUS, for $G \sim \text{Gamma}(a, \lambda)$, $E(G^r)$ is $\lambda^{-r}\Gamma(a + r)/\Gamma(a)$ for $r > -a$, and does not exist for $r \leq -a$. So

$$\begin{aligned}
E(T) &= E(Z)E(\frac{1}{\sqrt{V/m}}) = 0, \\
\text{Var}(T) = E(T^2) - (ET)^2 &= mE(Z^2)E\left(\frac{1}{V}\right) \\
&= m\frac{(1/2)\Gamma(m/2-1)}{\Gamma(m/2)} \\
&= \frac{m\Gamma(m/2-1)}{2\Gamma(m/2)} \\
&= \frac{m/2}{m/2-1} = \frac{m}{m-2}.
\end{aligned}$$

By the CLT, this is true for

$$\begin{aligned}
b_n &= E(T_1) + \ldots + E(T_n) = 0, \\
a_n &= \frac{1}{\sqrt{\text{Var}(T_1) + \cdots + \text{Var}(T_n)}} = \sqrt{\frac{m-2}{mn}}.
\end{aligned}$$

21. Ⓢ (a) Let $Y = e^X$, with $X \sim \text{Expo}(3)$. Find the mean and variance of Y.

(b) For $Y_1, \ldots, Y_n$ i.i.d. with the same distribution as Y from (a), what is the approximate distribution of the sample mean $\bar{Y}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j$ when n is large?

Solution:

(a) By LOTUS,

$$E(Y) = \int_0^\infty e^x (3e^{-3x}) dx = \frac{3}{2},$$
$$E(Y^2) = \int_0^\infty e^{2x} (3e^{-3x}) dx = 3.$$

So $E(Y) = 3/2, \text{Var}(Y) = 3 - 9/4 = 3/4$.

(b) By the CLT, $\bar{Y}_n$ is approximately $\mathcal{N}(\frac{3}{2}, \frac{3}{4n})$ for large n.

22. Ⓢ (a) Explain why the $\text{Pois}(n)$ distribution is approximately Normal if n is a large positive integer (specifying what the parameters of the Normal are).

(b) Stirling's formula is an amazingly accurate approximation for factorials:

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n,$$

where in fact the ratio of the two sides goes to 1 as $n \to \infty$. Use (a) to give a quick heuristic derivation of Stirling's formula by using a Normal approximation to the probability that a $\text{Pois}(n)$ r.v. is n, with the continuity correction: first write $P(N = n) = P(n - \frac{1}{2} < N < n + \frac{1}{2})$, where $N \sim \text{Pois}(n)$.

Solution:

(a) Let $S_n = X_1 + \cdots + X_n$, with $X_1, X_2, \ldots$ i.i.d. $\sim \text{Pois}(1)$. Then $S_n \sim \text{Pois}(n)$ and for n large, S_n is approximately $\mathcal{N}(n, n)$ by the CLT.

(b) Let $N \sim \text{Pois}(n)$ and $X \sim \mathcal{N}(n, n)$. Then

$$P(N = n) \approx P\left(n - \frac{1}{2} < X < n + \frac{1}{2}\right) = \frac{1}{\sqrt{2\pi n}} \int_{n-1/2}^{n+1/2} e^{\frac{-(x-n)^2}{2n}} dx.$$

The integral is approximately 1 since the interval of integration has length 1 and for large n the integrand is very close to 1 throughout the interval. So

$$e^{-n} n^n / n! \approx (2\pi n)^{-1/2}.$$

Rearranging this gives exactly Stirling's formula.

23. Ⓢ (a) Consider i.i.d. $\text{Pois}(\lambda)$ r.v.s $X_1, X_2, \ldots$. The MGF of X_j is $M(t) = e^{\lambda(e^t - 1)}$. Find the MGF $M_n(t)$ of the sample mean $\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$.

(b) Find the limit of $M_n(t)$ as $n \to \infty$. (You can do this with almost no calculation using a relevant theorem; or you can use (a) and the fact that $e^x \approx 1 + x$ if x is very small.)

Solution:

(a) The MGF is

$$E(e^{\frac{t}{n}(X_1 + \cdots + X_n)}) = \left(E(e^{\frac{t}{n}X_1})\right)^n = e^{n\lambda(e^{t/n} - 1)},$$

since the X_j are i.i.d. and $E(e^{\frac{t}{n}X_1})$ is the MGF of X_1 evaluated at t/n.

(b) By the law of large numbers, $\bar{X}_n \to \lambda$ with probability 1. The MGF of the constant λ (viewed as an r.v. that always equals λ) is $e^{t\lambda}$. Thus, $M_n(t) \to e^{t\lambda}$ as $n \to \infty$.

Mixed practice

31. Ⓢ Let X and Y be independent standard Normal r.v.s and let $R^2 = X^2 + Y^2$ (where $R > 0$ is the distance from (X, Y) to the origin).

(a) The distribution of R^2 is an example of three of the important distributions we have seen. State which three of these distributions R^2 is an instance of, specifying the parameter values.

(b) Find the PDF of R.

Hint: Start with the PDF $f_W(w)$ of $W = R^2$.

(c) Find $P(X > 2Y + 3)$ in terms of the standard Normal CDF Φ.

(d) Compute $\text{Cov}(R^2, X)$. Are R^2 and X independent?

Solution:

(a) It is χ^2_2, Expo(1/2), and Gamma(1,1/2).

(b) Since $R = \sqrt{W}$ with $f_W(w) = \frac{1}{2}e^{-w/2}$, we have

$$f_R(r) = f_W(w)|dw/dr| = \frac{1}{2}e^{-w/2}2r = re^{-r^2/2}, \text{ for } r > 0.$$

(This is the *Rayleigh distribution*, which was seen in Example 5.1.7.)

(c) We have

$$P(X > 2Y + 3) = P(X - 2Y > 3) = 1 - \Phi\left(\frac{3}{\sqrt{5}}\right)$$

since $X - 2Y \sim \mathcal{N}(0, 5)$.

(d) They are not independent since knowing X gives information about R^2, e.g., X^2 being large implies that R^2 is large. But R^2 and X are uncorrelated:

$$\text{Cov}(R^2, X) = \text{Cov}(X^2+Y^2, X) = \text{Cov}(X^2, X)+\text{Cov}(Y^2, X) = E(X^3)-(EX^2)(EX)+0 = 0.$$

32. Ⓢ Let $Z_1, \ldots, Z_n \sim \mathcal{N}(0, 1)$ be i.i.d.

(a) As a function of Z_1, create an Expo(1) r.v. X (your answer can also involve the standard Normal CDF Φ).

(b) Let $Y = e^{-R}$, where $R = \sqrt{Z_1^2 + \cdots + Z_n^2}$. Write down (but do not evaluate) an integral for $E(Y)$.

(c) Let $X_1 = 3Z_1 - 2Z_2$ and $X_2 = 4Z_1 + 6Z_2$. Determine whether X_1 and X_2 are independent (be sure to mention which results you're using).

Solution:

(a) Use Z_1 to get a Uniform and then the Uniform to get X: we have $\Phi(Z_1) \sim \text{Unif}(0, 1)$, and we can then take $X = -\log(1-\Phi(Z_1))$. By symmetry, we can also use $-\log(\Phi(Z_1))$.

Sanity check: $0 < \Phi(Z_1) < 1$, so $-\ln(\Phi(Z_1))$ is well-defined and positive.

(b) Let $W = Z_1^2 + \cdots + Z_n^2 \sim \chi^2_n$, so $Y = e^{-\sqrt{W}}$. We will use LOTUS to write $E(Y)$ using the PDF of W (there are other possible ways to use LOTUS here, but this is simplest since we get a single integral and we know the χ^2_n PDF). This gives

$$E(Y) = \int_0^\infty e^{-\sqrt{w}} \frac{1}{2^{n/2}\Gamma(n/2)} w^{n/2-1} e^{-w/2} dw.$$

(c) They are uncorrelated:

$$\text{Cov}(X_1, X_2) = 12\text{Var}(Z_1) + 10\text{Cov}(Z_1, Z_2) - 12\text{Var}(Z_2) = 0.$$

Also, (X_1, X_2) is Multivariate Normal since any linear combination of X_1, X_2 can be written as a linear combination of Z_1, Z_2 (and thus is Normal since the sum of two independent Normals is Normal). So X_1 and X_2 are independent.

33. Ⓢ Let $X_1, X_2, \ldots$ be i.i.d. positive r.v.s. with mean μ, and let $W_n = \frac{X_1}{X_1+\cdots+X_n}$.

(a) Find $E(W_n)$.

Hint: Consider $\frac{X_1}{X_1+\cdots+X_n} + \frac{X_2}{X_1+\cdots+X_n} + \cdots + \frac{X_n}{X_1+\cdots+X_n}$.

(b) What random variable does nW_n converge to (with probability 1) as $n \to \infty$?

(c) For the case that $X_j \sim \text{Expo}(\lambda)$, find the distribution of W_n, preferably without using calculus. (If it is one of the named distributions, state its name and specify the parameters; otherwise, give the PDF.)

Solution:

(a) The expression in the hint equals 1, and by linearity and symmetry its expected value is $nE(W_n)$. So $E(W_n) = 1/n$.

Sanity check: in the case that the X_j are actually constants, $\frac{X_1}{X_1+\cdots+X_n}$ reduces to $\frac{1}{n}$. Also in the case $X_j \sim \text{Expo}(\lambda)$, Part (c) shows that the answer should reduce to the mean of a Beta$(1, n-1)$ (which is $\frac{1}{n}$).

(b) By LLN, with probability 1 we have

$$nW_n = \frac{X_1}{(X_1 + \cdots + X_n)/n} \to \frac{X_1}{\mu} \text{ as } n \to \infty.$$

Sanity check: the answer should be a random variable since it's asked what random variable nW_n converges to. It should *not* depend on n since we let $n \to \infty$.

(c) Recall that $X_1 \sim \text{Gamma}(1)$ and $X_2 + \cdots + X_n \sim \text{Gamma}(n-1)$. By the connection between Beta and Gamma (i.e., the bank–post office story), $W_n \sim \text{Beta}(1, n-1)$.

Sanity check: The r.v. W_n clearly always takes values between 0 and 1, and the mean should agree with the answer from (a).

Chapter 11: Markov chains

Markov property

1. Ⓢ Let $X_0, X_1, X_2, \ldots$ be a Markov chain. Show that $X_0, X_2, X_4, X_6, \ldots$ is also a Markov chain, and explain why this makes sense intuitively.

   Solution: Let $Y_n = X_{2n}$; we need to show $Y_0, Y_1, \ldots$ is a Markov chain. By the definition of a Markov chain, we know that $X_{2n+1}, X_{2n+2}, \ldots$ ("the future" if we define the "present" to be time $2n$) is conditionally independent of $X_0, X_1, \ldots, X_{2n-2}, X_{2n-1}$ ("the past"), given X_{2n}. So given Y_n, we have that $Y_{n+1}, Y_{n+2}, \ldots$ is conditionally independent of $Y_0, Y_1, \ldots, Y_{n-1}$. Thus,

   $$P(Y_{n+1} = y | Y_0 = y_0, \ldots, Y_n = y_n) = P(Y_{n+1} = y | Y_n = y_n).$$

2. Ⓢ Let $X_0, X_1, X_2, \ldots$ be an irreducible Markov chain with state space $\{1, 2, \ldots, M\}$, $M \geq 3$, transition matrix $Q = (q_{ij})$, and stationary distribution $\mathbf{s} = (s_1, \ldots, s_M)$. Let the initial state X_0 follow the stationary distribution, i.e., $P(X_0 = i) = s_i$.

   (a) On average, how many of $X_0, X_1, \ldots, X_9$ equal 3? (In terms of $\mathbf{s}$; simplify.)

   (b) Let $Y_n = (X_n - 1)(X_n - 2)$. For $M = 3$, find an example of Q (the transition matrix for the *original* chain $X_0, X_1, \ldots$) where $Y_0, Y_1, \ldots$ is Markov, and another example of Q where $Y_0, Y_1, \ldots$ is not Markov. In your examples, make $q_{ii} > 0$ for at least one i and make sure it is possible to get from any state to any other state eventually.

   Solution:

   (a) Since X_0 has the stationary distribution, all of $X_0, X_1, \ldots$ have the stationary distribution. By indicator random variables, the expected value is $10s_3$.

   (b) Note that Y_n is 0 if X_n is 1 or 2, and Y_n is 2 otherwise. So the Y_n process can be viewed as merging states 1 and 2 of the X_n-chain into one state. Knowing the history of Y_n's means knowing when the X_n-chain is in state 3, without being able to distinguish state 1 from state 2.

   If $q_{13} = q_{23}$, then Y_n is Markov since given Y_n, even knowing the past $X_0, \ldots, X_n$ does not affect the transition probabilities. But if $q_{13} \neq q_{23}$, then the Y_n past history can give useful information about X_n, affecting the transition probabilities. So one example (not the only possible example!) is

   $$Q_1 = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \text{ (Markov)} \qquad Q_2 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 1 & 0 & 0 \end{pmatrix} \text{ (not Markov).}$$

   (c) The stationary distribution is uniform over all states:

   $$\mathbf{s} = (1/M, 1/M, \ldots, 1/M).$$

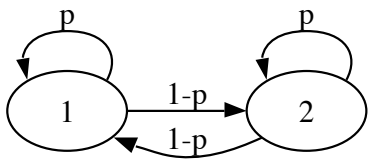
   This is because

   $$\begin{pmatrix} 1/M & 1/M & \ldots & 1/M \end{pmatrix} Q = \frac{1}{M} \begin{pmatrix} 1 & 1 & \ldots & 1 \end{pmatrix} Q = \begin{pmatrix} 1/M & 1/M & \ldots & 1/M \end{pmatrix},$$

   where the matrix multiplication was done by noting that multiplying a row vector of 1's times Q gives the column sums of Q.

Stationary distribution

4. Ⓢ Consider the Markov chain shown below, where $0 < p < 1$ and the labels on the arrows indicate transition probabilities.



(a) Write down the transition matrix Q for this chain.

(b) Find the stationary distribution of the chain.

(c) What is the limit of Q^n as $n \to \infty$?

Solution:

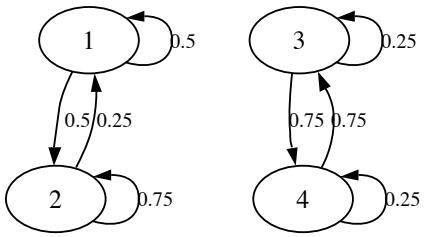
(a) The transition matrix is

$$Q = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}$$

(b) Because Q is symmetric, the stationary distribution for the chain is the uniform distribution $(1/2, 1/2)$.

(c) The limit of Q^n as $n \to \infty$ is the matrix with the limit distribution $(1/2, 1/2)$ as each row, i.e., $\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$.

5. Ⓢ Consider the Markov chain shown below, with state space $\{1, 2, 3, 4\}$ and the labels on the arrows indicate transition probabilities.



(a) Write down the transition matrix Q for this chain.

(b) Which states (if any) are recurrent? Which states (if any) are transient?

(c) Find two different stationary distributions for the chain.

Solution:

(a) The transition matrix is

$$Q = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.25 & 0.75 & 0 & 0 \\ 0 & 0 & 0.25 & 0.75 \\ 0 & 0 & 0.75 & 0.25 \end{pmatrix}$$

(b) All of the states are recurrent. Starting at state 1, the chain will go back and forth between states 1 and 2 forever (sometimes lingering for a while). Similarly, for any starting state, the probability is 1 of returning to that state.

(c) Solving

$$(a \quad b)\begin{pmatrix} 0.5 & 0.5 \\ 0.25 & 0.75 \end{pmatrix} = (a \quad b)$$

$$(c \quad d)\begin{pmatrix} 0.25 & 0.75 \\ 0.75 & 0.25 \end{pmatrix} = (c \quad d)$$

shows that $(a, b) = (1/3, 2/3)$, and $(c, d) = (1/2, 1/2)$ are stationary distributions on the 1, 2 chain and on the 3, 4 chain respectively, viewed as separate chains. It follows that $(1/3, 2/3, 0, 0)$ and $(0, 0, 1/2, 1/2)$ are both stationary for Q (as is any mixture $p(1/3, 2/3, 0, 0) + (1-p)(0, 0, 1/2, 1/2)$ with $0 \leq p \leq 1$).

6. Ⓢ Daenerys has three dragons: Drogon, Rhaegal, and Viserion. Each dragon independently explores the world in search of tasty morsels. Let X_n, Y_n, Z_n be the locations at time n of Drogon, Rhaegal, Viserion respectively, where time is assumed to be discrete and the number of possible locations is a finite number M. Their paths $X_0, X_1, X_2, \ldots$; $Y_0, Y_1, Y_2, \ldots$; and $Z_0, Z_1, Z_2, \ldots$ are independent Markov chains with the same stationary distribution $\mathbf{s}$. Each dragon starts out at a random location generated according to the stationary distribution.

(a) Let state 0 be home (so s_0 is the stationary probability of the home state). Find the expected number of times that Drogon is at home, up to time 24, i.e., the expected number of how many of $X_0, X_1, \ldots, X_{24}$ are state 0 (in terms of s_0).

(b) If we want to track all 3 dragons simultaneously, we need to consider the vector of positions, (X_n, Y_n, Z_n). There are M^3 possible values for this vector; assume that each is assigned a number from 1 to M^3, e.g., if $M = 2$ we could encode the states $(0,0,0), (0,0,1), (0,1,0), \ldots, (1,1,1)$ as $1, 2, 3, \ldots, 8$ respectively. Let W_n be the number between 1 and M^3 representing (X_n, Y_n, Z_n). Determine whether $W_0, W_1, \ldots$ is a Markov chain.

(c) Given that all 3 dragons start at home at time 0, find the expected time it will take for all 3 to be at home again at the same time.

Solution:

(a) By definition of stationarity, at each time Drogon has probability s_0 of being at home. By linearity, the desired expected value is $25s_0$.

(b) Yes, $W_0, W_1, \ldots$ is a Markov chain, since given the entire past history of the X, Y, and Z chains, only the most recent information about the whereabouts of the dragons should be used in predicting their vector of locations. To show this algebraically, let A_n be the event $\{X_0 = x_0, \ldots, X_n = x_n\}$, B_n be the event $\{Y_0 = y_0, \ldots, Y_n = y_n\}$, C_n be the event $\{Z_0 = z_0, \ldots, Z_n = z_n\}$, and $D_n = A_n \cap B_n \cap C_n$. Then

$$P(X_{n+1} = x, Y_{n+1} = y, Z_{n+1} = z | D_n)$$
$$= P(X_{n+1} = x | D_n) P(Y_{n+1} = y | X_{n+1} = x, D_n) P(Z_{n+1} = z | X_{n+1} = x, Y_{n+1} = y, D_n)$$
$$= P(X_{n+1} = x | A_n) P(Y_{n+1} = y | B_n) P(Z_{n+1} = z | C_n)$$
$$= P(X_{n+1} = x | X_n = x_n) P(Y_{n+1} = y | Y_n = y_n) P(Z_{n+1} = z | Z_n = z_n).$$

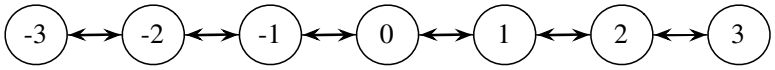
(c) The stationary probability for the W-chain of the state with Drogon, Rhaegal, Viserion being at locations x, y, z is $s_x s_y s_z$, since if (X_n, Y_n, Z_n) is drawn from this distribution, then marginally each dragon's location is distributed according to its stationary distribution, so

$$P(X_{n+1} = x, Y_{n+1} = y, Z_{n+1} = z) = P(X_{n+1} = x) P(Y_{n+1} = y) P(Z_{n+1} = z) = s_x s_y s_z.$$

So the expected time for all 3 dragons to be home at the same time, given that they all start at home, is $1/s_0^3$.

Reversibility

7. Ⓢ A Markov chain $X_0, X_1, \ldots$ with state space $\{-3,-2,-1,0,1,2,3\}$ proceeds as follows. The chain starts at $X_0 = 0$. If X_n is not an endpoint (-3 or 3), then X_{n+1} is $X_n - 1$ or $X_n + 1$, each with probability 1/2. Otherwise, the chain gets reflected off the endpoint, i.e., from 3 it always goes to 2 and from -3 it always goes to -2. A diagram of the chain is shown below.



(a) Is $|X_0|, |X_1|, |X_2|, \ldots$ also a Markov chain? Explain.

Hint: For both (a) and (b), think about whether the past and future are conditionally independent given the present; don't do calculations with a 7 by 7 transition matrix!

(b) Let sgn be the sign function: $\text{sgn}(x) = 1$ if $x > 0$, $\text{sgn}(x) = -1$ if $x < 0$, and $\text{sgn}(0) = 0$. Is $\text{sgn}(X_0), \text{sgn}(X_1), \text{sgn}(X_2), \ldots$ a Markov chain? Explain.

(c) Find the stationary distribution of the chain $X_0, X_1, X_2, \ldots$.

(d) Find a simple way to modify some of the transition probabilities q_{ij} for $i \in \{-3, 3\}$ to make the stationary distribution of the modified chain uniform over the states.

Solution:

(a) Yes, $|X_0|, |X_1|, |X_2|, \ldots$ is also a Markov Chain. It can be viewed as the chain on state space $0, 1, 2, 3$ that moves left or right with equal probability, except that at 0 it bounces back to 1 and at 3 it bounces back to 2. Given that $|X_n| = k$, we know that $X_n = k$ or $X_n = -k$, and being given information about $X_{n-1}, X_{n-2}, \ldots$ does not affect the conditional distribution of $|X_{n+1}|$.

(b) No, this is not a Markov chain because knowing that the chain was at 0 recently affects how far the chain can be from the origin. For example,

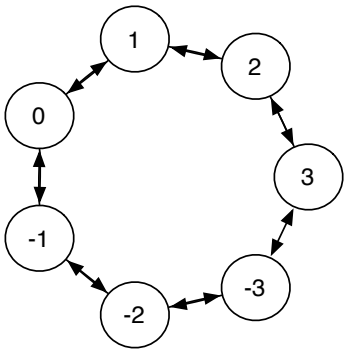
$$P(\text{sgn}(X_2) = 1 | \text{sgn}(X_1) = 1) > P(\text{sgn}(X_2) = 1 | \text{sgn}(X_1) = 1, \text{sgn}(X_0) = 0)$$

since the conditioning information on the righthand side implies $X_1 = 1$, whereas the conditioning information on the lefthand side says exactly that X_1 is $1, 2$, or 3.

(c) Using the result about the stationary distribution of a random walk on an undirected network, the stationary distribution is proportional to the degree sequence, $(1, 2, 2, 2, 2, 2, 1)$. Thus, the stationary distribution is $\frac{1}{12}(1, 2, 2, 2, 2, 2, 1)$.

(d) The uniform distribution will be the stationary distribution if we modify the transition matrix to make it symmetric. Connecting state -3 to state 3 so that the states are arranged in a circle gives the desired symmetry, as illustrated below.

8. Ⓢ Let G be an undirected network with nodes labeled $1, 2, \ldots, M$ (edges from a node to itself are not allowed), where $M \geq 2$ and random walk on this network is irreducible. Let d_j be the degree of node j for each j. Create a Markov chain on the state space $1, 2, \ldots, M$, with transitions as follows. From state i, generate a proposal j by choosing a uniformly random j such that there is an edge between i and j in G; then go to j with probability $\min(d_i/d_j, 1)$, and stay at i otherwise.



(a) Find the transition probability q_{ij} from i to j for this chain, for all states i, j.

(b) Find the stationary distribution of this chain.

Solution:

(a) First let $i \neq j$. If there is no $\{i, j\}$ edge, then $q_{ij} = 0$. If there is an $\{i, j\}$ edge, then

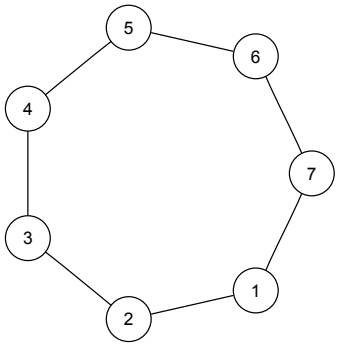
$$q_{ij} = (1/d_i)\min(d_i/d_j, 1) = \begin{cases} 1/d_i \text{ if } d_i \geq d_j, \\ 1/d_j \text{ if } d_i < d_j \end{cases},$$

since the proposal to go to j must be made and then accepted. For $i = j$, we have $q_{ii} = 1 - \sum_{j \neq i} q_{ij}$ since each row of the transition matrix must sum to 1.

(b) Note that $q_{ij} = q_{ji}$ for all states i, j. This is clearly true if $i = j$ or $q_{ij} = 0$, so assume $i \neq j$ and $q_{ij} > 0$. If $d_i \geq d_j$, then $q_{ij} = 1/d_i$ and $q_{ji} = (1/d_j)(d_j/d_i) = 1/d_i$, while if $d_i < d_j$, then $q_{ij} = (1/d_i)(d_i/d_j) = 1/d_j$ and $q_{ji} = 1/d_j$.

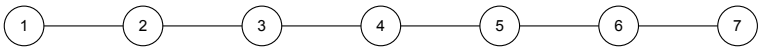
Thus, the chain is reversible with respect to the uniform distribution over the states, and the stationary distribution is uniform over the states, i.e., state j has stationary probability $1/M$ for all j. (This is an example of the *Metropolis algorithm*, a Monte Carlo method explored in Chapter 12.)

9. Ⓢ (a) Consider a Markov chain on the state space $\{1, 2, \ldots, 7\}$ with the states arranged in a "circle" as shown below, and transitions given by moving one step clockwise or counterclockwise with equal probabilities. For example, from state 6, the chain moves to state 7 or state 5 with probability 1/2 each; from state 7, the chain moves to state 1 or state 6 with probability 1/2 each. The chain starts at state 1.



Find the stationary distribution of this chain.

(b) Consider a new chain obtained by "unfolding the circle". Now the states are arranged as shown below. From state 1 the chain always goes to state 2, and from state 7 the chain always goes to state 6. Find the new stationary distribution.



Solution:

(a) The symmetry of the chain suggests that the stationary distribution should be uniform over all the states. To verify this, note that the reversibility condition is satisfied. So the stationary distribution is $(1/7, 1/7, \ldots, 1/7)$.

(b) By the result about random walk on an undirected network, the stationary probabilities are proportional to the degrees. So we just need to normalize $(1, 2, 2, 2, 2, 2, 1)$, obtaining $(1/12, 1/6, 1/6, 1/6, 1/6, 1/6, 1/12)$.

10. Ⓢ Let X_n be the price of a certain stock at the start of the nth day, and assume that $X_0, X_1, X_2, \ldots$ follows a Markov chain with transition matrix Q. (Assume for simplicity that the stock price can never go below 0 or above a certain upper bound, and that it is always rounded to the nearest dollar.)

(a) A lazy investor only looks at the stock once a year, observing the values on days $0, 365, 2 \cdot 365, 3 \cdot 365, \ldots$. So the investor observes $Y_0, Y_1, \ldots$, where Y_n is the price after n years (which is $365n$ days; you can ignore leap years). Is $Y_0, Y_1, \ldots$ also a Markov chain? Explain why or why not; if so, what is its transition matrix?

(b) The stock price is always an integer between \$0 and \$28. From each day to the next, the stock goes up or down by \$1 or \$2, all with equal probabilities (except for days when the stock is at or near a boundary, i.e., at \$0, \$1, \$27, or \$28).
If the stock is at \$0, it goes up to \$1 or \$2 on the next day (after receiving government bailout money). If the stock is at \$28, it goes down to \$27 or \$26 the next day. If the stock is at \$1, it either goes up to \$2 or \$3, or down to \$0 (with equal probabilities); similarly, if the stock is at \$27 it either goes up to \$28, or down to \$26 or \$25. Find the stationary distribution of the chain.

Solution:

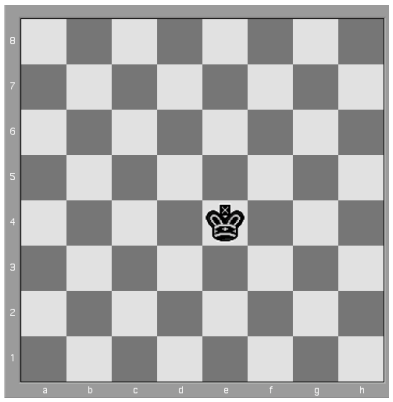
(a) Yes, it is a Markov chain: given the whole past history $Y_0, Y_1, \ldots, Y_n$, only the most recent information Y_n matters for predicting Y_{n+1}, because $X_0, X_1, \ldots$ is Markov. The transition matrix of $Y_0, Y_1, \ldots$ is Q^{365}, since the kth power of Q gives the k-step transition probabilities.

(b) This is an example of random walk on an undirected network, so we know the stationary probability of each node is proportional to its degree. The degrees are $(2, 3, 4, 4, \ldots, 4, 4, 3, 2)$, where there are $29 - 4 = 25$ 4's. The sum of these degrees is 110 (coincidentally?). Thus, the stationary distribution is

$$\left(\frac{2}{110}, \frac{3}{110}, \frac{4}{110}, \frac{4}{110}, \ldots, \frac{4}{110}, \frac{4}{110}, \frac{3}{110}, \frac{2}{110}\right),$$

with 25 $\frac{4}{110}$'s.

11. Ⓢ In chess, the king can move one square at a time in any direction (horizontally, vertically, or diagonally).
For example, in the diagram, from the current position the king can move to any of 8 possible squares. A king is wandering around on an otherwise empty 8×8 chessboard, where for each move all possibilities are equally likely. Find the stationary distribution

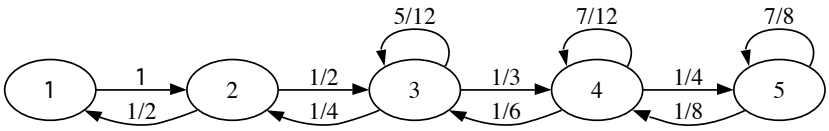


of this chain (of course, don't list out a vector of length 64 explicitly! Classify the 64 squares into types and say what the stationary probability is for a square of each type).

Solution: There are 4 corner squares, 24 edge squares, and 36 normal squares, where by "edge" we mean a square in the first or last row or column, excluding the 4 corners, and by "normal" we mean a square that's not on the edge or in a corner. View the chessboard as an undirected network, where there is an edge between two squares if the king can walk from one to the other in one step.

The stationary probabilities are proportional to the degrees. Each corner square has degree 3, each edge square has degree 5, and each normal square has degree 8. The total degree is $420 = 3 \cdot 4 + 24 \cdot 5 + 36 \cdot 8$ (which is also twice the number of edges in the network). Thus, the the stationary probability is $\frac{3}{420}$ for a corner square, $\frac{5}{420}$ for an edge square, and $\frac{8}{420}$ for a normal square.

13. Ⓢ Find the stationary distribution of the Markov chain shown below, *without using matrices*. The number above each arrow is the corresponding transition probability.



Solution: We will show that this chain is reversible by solving for $\mathbf{s}$ (which will work out nicely since this is a birth-death chain). Let q_{ij} be the transition probability from i to j, and solve for $\mathbf{s}$ in terms of s_1. Noting that $q_{ij} = 2q_{ji}$ for $j = i+1$ (when $1 \leq i \leq 4$), we have that

$$s_1 q_{12} = s_2 q_{21} \text{ gives } s_2 = 2s_1.$$
$$s_2 q_{23} = s_3 q_{32} \text{ gives } s_3 = 2s_2 = 4s_1.$$
$$s_3 q_{34} = s_4 q_{43} \text{ gives } s_4 = 2s_3 = 8s_1.$$
$$s_4 q_{45} = s_5 q_{54} \text{ gives } s_5 = 2s_4 = 16s_1.$$

The other reversibility equations are automatically satisfied since here $q_{ij} = 0$ unless $|i-j| \leq 1$. Normalizing, the stationary distribution is

$$\left(\frac{1}{31}, \frac{2}{31}, \frac{4}{31}, \frac{8}{31}, \frac{16}{31}\right).$$

Sanity check: This chain "likes" going from left to right more than from right to left, so

the stationary probabilities should be increasing from left to right. We also know that $s_j = \sum_i s_i q_{ij}$ (since if the chain is in the stationary distribution at time n, then it is also in the stationary distribution at time $n+1$), so we can check, for example, that $s_1 = \sum_i s_i q_{i1} = \frac{1}{2} s_2$.

Mixed practice

17. Ⓢ A cat and a mouse move independently back and forth between two rooms. At each time step, the cat moves from the current room to the other room with probability 0.8. Starting from room 1, the mouse moves to room 2 with probability 0.3 (and remains otherwise). Starting from room 2, the mouse moves to room 1 with probability 0.6 (and remains otherwise).

(a) Find the stationary distributions of the cat chain and of the mouse chain.

(b) Note that there are 4 possible (cat, mouse) states: both in room 1, cat in room 1 and mouse in room 2, cat in room 2 and mouse in room 1, and both in room 2. Number these cases $1, 2, 3, 4$, respectively, and let Z_n be the number representing the (cat, mouse) state at time n. Is $Z_0, Z_1, Z_2, \ldots$ a Markov chain?

(c) Now suppose that the cat will eat the mouse if they are in the same room. We wish to know the expected time (number of steps taken) until the cat eats the mouse for two initial configurations: when the cat starts in room 1 and the mouse starts in room 2, and vice versa. Set up a system of two linear equations in two unknowns whose solution is the desired values.

Solution:

(a) The cat chain has transition matrix

$$Q_{\text{cat}} = \begin{pmatrix} \frac{2}{10} & \frac{8}{10} \\ \frac{8}{10} & \frac{2}{10} \end{pmatrix}.$$

The uniform distribution $(\frac{1}{2}, \frac{1}{2})$ is stationary since the transition matrix is symmetric.

The mouse chain has transition matrix

$$Q_{\text{mouse}} = \begin{pmatrix} \frac{7}{10} & \frac{3}{10} \\ \frac{6}{10} & \frac{4}{10} \end{pmatrix}.$$

The stationary distribution is proportional to (x, y) with $7x + 6y = 10x, 3x + 4y = 10y$. This reduces to $x = 2y$. So the stationary distribution is $(\frac{2}{3}, \frac{1}{3})$.

(b) Yes, it is a Markov chain. Given the current (cat, mouse) state, the past history of where the cat and mouse were previously are irrelevant for computing the probabilities of what the next state will be.

(c) Let a and b be the expected values for the two initial configurations, respectively. Conditioning on the first move of the cat and the first move of the mouse, we have

$$a = \underbrace{(0.2)(0.6)}_{\text{both in room 1}} + \underbrace{(0.8)(0.4)}_{\text{both in room 2}} + \underbrace{(0.2)(0.4)(1+a)}_{\text{cat in room 1, mouse in room 2}} + \underbrace{(0.8)(0.6)(1+b)}_{\text{cat in room 2, mouse in room 1}},$$

$$b = \underbrace{(0.8)(0.7)}_{\text{both in room 1}} + \underbrace{(0.2)(0.3)}_{\text{both in room 2}} + \underbrace{(0.8)(0.3)(1+a)}_{\text{cat in room 1, mouse in room 2}} + \underbrace{(0.2)(0.7)(1+b)}_{\text{cat in room 2, mouse in room 1}}.$$

(The solution to this system works out to $a = 335/169, b = 290/169$.)