

Solutions to Practice Problems for *Probability Review*

Many of these problems are drawn from Arnold Barnett's excellent and recent textbook *Applied Probability: Models and Intuition* (Dynamic Ideas, LLC). (You do not need access to this book for the course and in fact, it's quite expensive to obtain it here in the U.K. as it needs to be shipped from the U.S.) Questions without "Challenging!" notation are within the exam scope.

1. Independence

If A and B are independent, show that A and B^c are also independent.

Solution:

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

2. Conditional Probability

An insurance company insures an equal number of male and female drivers. In any given year the probability that a male driver has an accident involving a claim is α , independently of other years. The analogous probability for females is β . Assume the insurance company selects a driver at random. [From Jacod & Protter (2000): *Probability Essentials*]

- (a) What is the probability the selected driver makes a claim in two consecutive years?
- (b) What is the probability that a claimant is female?

Solution: (a) $(\alpha^2 + \beta^2)/2$. (b) $\beta/(\alpha + \beta)$.

3. Mean and Variance

Show that

$$\text{Var}(X) := E(X - E(X))^2 = E(X)^2 - E(X^2).$$

Solution:

$$\begin{aligned} E(X - E(X))^2 &= E(X^2 - 2XE(X) + E(X)^2) \\ &= E(X^2) - 2E(XE(X)) + E(X)^2 \\ &= E(X^2) - 2E(X)^2 + E(X)^2 \\ &= E(X)^2 - E(X^2). \end{aligned}$$

4. Mode of Binomial and Poisson Distributions

For a discrete random variable, define mode as the value with greatest probability, i.e., $\operatorname{argmax}_j \{P(X = j)\}$. Find the mode of Binomial distribution $B(n, p)$ and the mode of Poisson distribution $\operatorname{Poisson}(\lambda)$.

Solution: By calculating $\frac{P(X=k)}{P(X=k-1)}$ over $k = 1, \dots$, we can find the “turning point” which indicates the mode. The mode of $B(n, p)$ is $\lfloor (n+1)p \rfloor$ and the mode of $\operatorname{Poisson}(\lambda)$ is $\lfloor \lambda \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of x , or floor of x .

5. Continuous Random Variable

Let X be a continuous random variable with density

$$f(x) = \begin{cases} 0 & x < -1 \\ x & -1 \leq x \leq 0 \\ ae^{-bx} & x \geq 0 \end{cases}$$

and expected value $E(X) = 1$.

(a) What are a and b ?

Solution: We know $\int_{-\infty}^{\infty} f(x)dx = 1$ and $E(X) = 1$, hence solving

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(x)dx = \int_{-1}^0 xdx + a \int_0^{\infty} e^{-bx} dx = -\frac{1}{2} + \frac{a}{b}, \\ 1 &= \int_{-\infty}^{\infty} xf(x)dx = \int_{-1}^0 x^2 dx + a \int_0^{\infty} xe^{-bx} dx = \frac{1}{3} + \frac{a}{b^2}, \end{aligned}$$

we get

$$a = \frac{27}{8}, \quad b = \frac{9}{4}.$$

(b) Find $\operatorname{Var}(X)$.

Solution: Having known $E(X) = 1$, we just need $E(X^2)$.

$$E(X^2) = \int_{-1}^0 x^3 dx + \frac{27}{8} \int_0^{\infty} x^2 e^{-\frac{9}{4}x} dx = -\frac{1}{4} + \frac{8}{27}.$$

Hence

$$\operatorname{Var}(X) = E(X^2) - E(X)^2 = -\frac{1}{4} + \frac{8}{27} - 1 = -\frac{103}{108}.$$

6. Normal Distribution

Answer the following questions:

- (a) Suppose that the height (X) in inches, of 25-year-old men is a normal random variable with mean $\mu = 70$ inches. If $P(X > 79) = 0.025$, what is the standard deviation of X ?
- (b) Suppose that the weight (Y) in pounds, of 40-year-old men is a normal random variable with standard deviation $\sigma = 20$ pounds. If 5% of this population weigh less than 160 pounds, what is the mean μ of Y ?
- (c) Find an interval that covers the middle 95% of $Z \sim \mathcal{N}(64, 8)$.

Solution:

- (a) $1.96 = \frac{79-70}{\sigma} \Rightarrow \sigma = 4.59$.
- (b) $-1.645 = \frac{160-\mu}{20} \Rightarrow \mu = 192.9$.
- (c) We have 2.5% probability at each one of the two tails. Therefore

$$\pm 1.96 = \frac{x - 64}{8} \Rightarrow x = 64 \pm 1.96 \times 8 = 48.32 \text{ or } 79.68.$$

7. Exponential Distribution

Let $X \sim \exp(\lambda)$ and $Y \sim \exp(\mu)$ be two independent exponential random variables.

- (a) What is the distribution of $\min\{X, Y\}$?

Solution:

$$\begin{aligned} P(\min\{X, Y\} > x) &= P(X > x, Y > x) \\ (\text{By independence}) &= P(X > x)P(Y > x) \\ &= e^{-\lambda x} e^{-\mu x} \\ &= e^{-(\lambda+\mu)x}. \end{aligned}$$

Hence $\min\{X, Y\} \sim \exp(\lambda + \mu)$ is also an exponential random variable.

- (b) What is the probability that X is the smaller one between X and Y ? Under $\lambda = \mu$?

Solution:

$$\begin{aligned} P(X = \min\{X, Y\}) &= P(X \leq Y) = \int_0^\infty P(X \leq Y \mid Y = y) f_Y(y) dy \\ &= \int_0^\infty (1 - e^{-\lambda y}) \mu e^{-\mu y} dy \\ &= \frac{\lambda}{\lambda + \mu}. \end{aligned}$$

Clearly if $\lambda = \mu$, the probability is $1/2$, which is intuitive.

8. Getting out of the Jungle

You are lost in a jungle and just arrive at a crossroads with 3 choices. One of them will bring you back to the same point after 1 hour of walk; the second will bring you back after 6 hours of walk; the last will lead you out of the jungle after 2 hours of walk. Facing no signs on the roads you have to choose one randomly each time you are there. What is the expected time until you get out of the jungle?

Solution: Let T be the time until you get out of the jungle and R_i be the choice of road i for $i = 1, 2, 3$. Using conditional expectation, we have

$$\begin{aligned} E(T) &= E(T | R_1)P(R_1) + E(T | R_2)P(R_2) + E(T | R_3)P(R_3) \\ &= \frac{1}{3}[(1 + E(T)) + (6 + E(T)) + 2] \\ &= \frac{1}{3}[9 + 2E(T)]. \end{aligned}$$

Solving this we get $E(T) = 9$.

9. Genetics

The geneticist Gregor Mendel analyzed the inheritance of traits from parents to their offspring. The basic principles are:

- Each of the two parents has two genes on a given dimension (e.g. color), and each offspring gets one gene from each parent.
- Each of a particular parent's two genes has an equal chance of being passed on to the offspring.
- The genetic contributions from the two parents are independent.

There are two different types of genes, *dominant* and *recessive*, associated with different traits. (For example, for the colours of peas, *yellow* is dominant and *green* is recessive.) If an offspring gets one dominant gene and one recessive one, then it will exhibit the dominant trait, as it will if it gets two dominant genes from its parents. Only if both genes are recessive will the recessive trait show up.

Some questions:

- (i) Suppose that one parent pea plant has one yellow gene and one green one, i.e. is *hybrid*, and that the other has two green genes. If they “mate”, what is the probability that their offspring (named Zelda) will be green? (Actually a pea plant is self-pollinating, and can mate with itself. But, as Mendel literally did, we will create male and female pea-plants for our discussion.)

Solution: We are told the parents have genetic makeup (Y, G) and (G, G) . Moreover the offspring will be green if and only if its genetic makeup is (G, G) . This will occur with probability $1 \times 1/2 = 1/2$ given the independence assumption.

- (ii) Suppose that the overall population of pea plants is 25% all yellow, 50% hybrid yellow and green, and 25% all green in genetic makeup. If Zelda is green and picks a mate at random, what is the probability that they have a green offspring?

Solution: Note that Zelda must have genetic makeup (G, G) . We can calculate her probability of having a green offspring as

$$\begin{aligned} P(\text{green offspring}) &= P(\text{green offspring} \mid \text{mate is } GG)P(\text{mate is } GG) \\ &\quad + P(\text{green offspring} \mid \text{mate is } YG)P(\text{mate is } YG) \\ &\quad + P(\text{green offspring} \mid \text{mate is } YY)P(\text{mate is } YY) \\ &= (1 \times .25) + (.5 \times .5) + (0 \times .25) \\ &= 0.5. \end{aligned}$$

- (iii) Now suppose that Zelda is green and picks a mate at random from among those who are yellow (her favourite colour). What is the probability that their offspring will be green?

Solution: Zelda's mate must be either (Y, Y) or (Y, G) since they are the genetic makeups that yield a yellow pea. We must first compute these probabilities conditional on being yellow. We have

$$\begin{aligned} P((Y, G) \mid \text{yellow}) &= \frac{P((Y, G) \mid (Y, G) \text{ or } (Y, Y))}{P((Y, G) \text{ or } (Y, Y))} \\ &= \frac{P((Y, G) \cap \{(Y, G) \text{ or } (Y, Y)\})}{P((Y, G) \text{ or } (Y, Y))} \\ &= \frac{P((Y, G))}{P((Y, G) \text{ or } (Y, Y))} \\ &= \frac{.5}{.75} \\ &= \frac{2}{3}. \end{aligned}$$

Therefore we also have $P((Y, Y) \mid \text{yellow}) = 1/3$. We can now compute

$$\begin{aligned} P(\text{green offspring} \mid \text{yellow}) &= (0 \times 1/3) + (1/2 \times 2/3) \\ &= 1/3. \end{aligned}$$

- (iv) Assuming the 25/50/25 split among pea plants in a given generation, will green peas be less common in the next generation if mating is random. Are they on a path towards eventually dying out?

Solution: There are three routes to a green offspring:

- Both parents are (G, G) and both parents donate a green gene to the offspring. This occurs with probability $(1/4) \times (1/4) \times 1 \times 1 = 1/16$.

- One parent is (G, G) and the other parent is (Y, G) and both parents donate a green gene to the offspring. This occurs with probability $2 \times (1/4) \times (1/2) \times 1 \times (1/2) = 2/16$. (The factor of 2 arises because we can have a (Y, G) father and (G, G) mother or we can have a (Y, G) mother and (G, G) father.)
- Both parents are (Y, G) and both parents donate a green gene to the offspring. This occurs with probability $(1/2) \times (1/2) \times (1/2) \times (1/2) = 1/16$.

Because these possibilities are mutually exclusive we have

$$\begin{aligned} P(\text{green offspring}) &= \frac{1}{16} + \frac{2}{16} + \frac{1}{16} \\ &= 1/4. \end{aligned}$$

So under random mating the next generation will also be 25% green and (with a little more work) we could conclude that there is no tendency for green pea plants to die out over time.

10. Mendel's Sure Thing

Mendel has learned about a betting scheme for “double or nothing bets” that seems to ensure victory. On the first round, the player bets \$1. If he wins, he stops, having won \$2 and thus having achieved a net gain of $2-1=\$1$. If he loses, he bets \$2 the next time. If that second bet is successful, he stops, having won \$4, and thus achieving a *net* gain of $(4-2)-1 = \$1$. If the second bet also fails, he bets \$8 on the third round; if he wins then, his net gain over the three rounds is $(8-4)-(1+2) = \$1$ and he stops. If he loses on the third round, he bets \$16 on the fourth round ...

In short, the strategy is: if the player gets to the n^{th} bet (having lost the previous $n-1$ bets), he makes a bet of 2^{n-1} dollars and:

- If he wins, he collects 2^n dollars and stops.
- If he loses, he goes on to an $(n+1)^{\text{st}}$ bet and then bets 2^n dollars.

(i) If Mendel can play this game forever, is he assured of winning \$1?

Solution: If Mendel wins for the first time on the n^{th} bet then his net winnings will be

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

This means that regardless of when Mendel first wins, his net winnings will be \$1. Will he eventually win at some point? Yes! To see this note that the only way that Mendel can fail to win at some point is that if he loses infinitely many independent bets each

of which has probability $1/2$. The probability of losing n bets in a row is $(1/2)^n$ and so the probability of losing infinitely many of them is

$$\lim_{n \rightarrow \infty} \left(\frac{1}{2}\right)^n = 0.$$

(ii) Suppose that Mendel only has \$1,023, and thus cannot play forever. Under this restriction:

- What is the probability distribution for his net gain (or loss) until he stops playing?
- What is his expected gain (or loss) until he stops?

Solution: Note that $1,023 = 2^n - 1$ where $n = 10$. This means that Mendel can make at most 10 bets since if he loses all 10 of them (which will occur with probability $(1/2)^{10} = 1/1,024$) he will have lost

$$1 + 2 + \cdots + 2^9 = 2^{10} - 1 = 1,023$$

and therefore have no money left to gamble. If he wins on or before the 10^{th} bet then, as before, his net winnings will be \$1. Let U then be his net winnings. It follows that the distribution of U is

$$U = \begin{cases} -1023, & \text{w.p. } \frac{1}{1024} \\ 1, & \text{w.p. } \frac{1023}{1024}. \end{cases}$$

We can now also compute the expected gain / loss until Minerva stops. It satisfies

$$\begin{aligned} E[U] &= -1023 \times \frac{1}{1024} + 1 \times \frac{1023}{1024} \\ &= 0 \end{aligned}$$

so that on average, Minerva wins nothing!

11. Traffic Congestion

Mendel and Minerva go together to the Bureau of Motor Vehicles (BMV), where they must take care of matters concerning their vehicles and drivers' licenses. They know from experience that the time it will take to complete their business is quite unpredictable: in fact, the transaction time (entry to exit) is exponentially distributed with a mean of 20 minutes, i.e. $\lambda = 1/20$, for each of them. Because they are going on different lines and dealing with different people, their transaction times are independent. Some questions about their visit:

- (i) What is the probability that neither of them will be finished within 20 minutes?

Solution: Let us use X_1 and X_2 to denote Mendel's and Minerva's service times, respectively. Because of independence we have

$$\begin{aligned} P(\text{both times exceed } 20) &= P(X_1 > 20) \times P(X_2 > 20) \\ &= e^{-\lambda x} \times e^{-\lambda x} \\ &= e^{-2\lambda x}. \end{aligned}$$

where $x = 20$ and $\lambda = 1/20$. Simplifying we obtain $P(\text{both times exceed } 20) = e^{-2} = 0.14$.

- (ii) What is the probability that Minerva will be finished in 20 minutes but Mendel will not be done by then?

Solution: Again by independence we have (with $x = 20$ and $\lambda = 1/20$)

$$\begin{aligned} P(\text{Minerva finished in 20 but Mendel not}) &= P(X_2 \leq 20) \times P(X_1 > 20) \\ &= (1 - e^{-\lambda x}) \times e^{-\lambda x} \\ &= (1 - e^{-1}) \times e^{-1} \\ &= 0.23. \end{aligned}$$

- (iii) What is the probability that Mendel's transaction time will be at least ten minutes longer than Minerva's?

Solution: In order for this to occur Minerva must finish first (which occurs with probability $1/2$) and then Mendel must take at least 10 additional minutes. Let's define two events:

A : Minerva finishes first

B : Mendel takes at least 10 minutes longer than Minerva

We want $P(B)$ and we can calculate it as follows.

$$\begin{aligned} P(B) &= P(A \cap B) \\ &= P(B | A)P(A) \\ &= e^{-\lambda \times 10} \times \frac{1}{2} \\ &= 0.31 \end{aligned}$$

where we have used the memoryless property of the exponential distribution is obtaining that $P(B | A) = e^{-\lambda \times 10}$. (This is because at the moment Minerva finishes, Mendel's remaining time to completion has an $\text{Exp}(\lambda)$ distribution.)

- (iv) Zanzibar is driving over to pick Mendel and Minerva up after they are finished. Because no vehicle standing is allowed outside the BMV, he wants to arrive at time T when the probability is 95% that both of them are finished (and thus are waiting outside). Assuming that they reach the BMV at 1pm, what time T satisfies his requirement?

Solution: We want the number τ for which $P(X_1, X_2 \leq \tau) = 0.95$. By independence we have

$$\begin{aligned} P(X_1, X_2 \leq \tau) &= P(X_1 \leq \tau) P(X_2 \leq \tau) \\ &= \left(1 - e^{-\frac{\tau}{20}}\right)^2. \end{aligned} \tag{1}$$

We therefore set the r.h.s. of (1) equal to 0.95 and solve for τ . We find $\tau = 73.8$ minutes.

- (v) When Zanzibar arrives at that time T , what is the probability that Mendel and Minerva are waiting and both of them have waited more than 30 minutes?

Solution: This occurs with probability

$$\begin{aligned} P(X_1, X_2 \leq 43.8) &= P(X_1 \leq 43.8) P(X_2 \leq 43.8) \\ &= \left(1 - e^{-\frac{43.8}{20}}\right)^2 \\ &= 0.79. \end{aligned}$$

So there is a roughly 4 in 5 chance that *both* of them will wait at least half an hour!

- (vi) Let random variable Z be the amount of time until the first of them is finished. What is the probability distribution of Z ?

Solution: We have

$$\begin{aligned} P(Z \geq z) &= P(X_1 \geq z, X_2 \geq z) \\ &= P(X_1 \geq z) P(X_2 \geq z) \\ &= e^{-z\lambda} \times e^{-z\lambda} \\ &= e^{-z(2\lambda)} \end{aligned}$$

and so

$$P(Z \leq z) = 1 - e^{-z(2\lambda)}$$

which we recognize as the CDF of an $\text{Exp}(2\lambda)$ random variable. Therefore $Z \sim \text{Exp}(2\lambda)$.

This is a special case of a very important and powerful result:

If $X_1 \sim \text{Exp}(\lambda_1)$, $X_2 \sim \text{Exp}(\lambda_2)$ and X_1 and X_2 are independent, then $Z \sim \text{Exp}(\lambda_1 + \lambda_2)$ where $Z := \min\{X_1, X_2\}$.

- (vii) How long on average will it be until both of them are finished?

Solution: Using the previous result we know the expected time until the first one to finish will be $E[Z] = 1/(2\lambda) = 10$ minutes where $Z := \min\{X_1, X_2\}$. By the memoryless property, the second one to finish will then an additional $1/\lambda = 20$ minutes on average. So the expected completion time for both to finish is $10 + 20 = 30$ minutes.

12. California Nightmare

Some geological measurements suggest that maximum-strength earthquakes occur on the southern end of California's San Andreas fault every 160 years on average. But the individual intervals between "Big Ones" vary a bit around this mean. Suppose that an adequate model posits that, given that a huge earthquake has just occurred, the time X until the next one follows a normal distribution with mean 160 and standard deviation 30 years, i.e. $X \sim N(160, 30)$.

The last earthquake on the southern San Andreas Fault – which winds through populous Southern California – occurred in 1857. What is the probability that the next large quake will take place within the next decade? (It is mid-2014 as of this writing.)

Solution: We define two events:

E : $X > 157$, which corresponds to no earthquake from (mid) 1857 to mid-2014.

F : $157 < X < 167$, which entails a big earthquake between mid-2014 and mid-2024.

What we are seeking is $P(F | E)$. This follows from Bayes' Theorem since

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

We know $P(E | F) = 1$ (why?) and we can use R to obtain that $P(F) \approx .132$ and $P(E) \approx .540$. We conclude that $P(F | E) \approx 0.244$.

13. Airline Yield Management (Challenging!)

Yield management has been described by American Airlines as "charging the right fare to the right passenger at the right time". Many business travelers view the airlines as providing an intercity taxi service, for which they'll pay high fares in return for frequent flights that they can choose at a moment's notice. Leisure travelers, in contrast, tend to think about price above all and will purchase non-refundable tickets well in advance to get low fares. The airline's dilemma is how to trade off the two sources of demand so as to maximize ticket revenue.

The problem is illustrated by an example that is a disguised (through simplification) version of a true-life problem. Suppose that an airline offers a 5.30pm flight from San Jose to San Diego on a 100-seat Embraer 190 jet. The regular coach fare for the flight is \$190, but there is also a \$69 dollar discount fare that must be purchased more than three weeks in advance and is neither refundable nor changeable. Business travelers who pay \$190 per seat book very close to flight time; hence, at the time that discount tickets are sold, it is not known how many full-fare passengers will eventually want bookings for the flight. The question is how many of the 100 seats to make available at the \$69 fare.

Suppose that we are thinking about an upcoming Thursday afternoon flight. A sophisticated airline will have precise records about full-fare demand for the flight going back many Thursdays (as well as much other relevant information). Using its detailed statistical model, the airline estimates that X , the (unknown) number of full-fare passengers who will seek tickets for the flight, approximately follows a normal distribution with mean 90 and standard deviation 10. (More specifically, it believes that the probability that exactly k passengers seek full-fare tickets can be estimated by the chance that X falls between $k - 0.5$ and $k + 0.5$, where X is $N(90, 10)$. It is effectively rounding off customers from the normal to the nearest integer.

Suppose that every full-fare passenger who seeks a seat on the flight will in fact purchase a ticket and show up for the flight if a seat is available. If the airline *can sell as many discount seats as it wants*, how many would it choose to sell in order to maximise its expected total revenue?

Solution: Let V_i be the total revenue when the airline sells i seats at the discount price of \$69 and then sells as many of the remaining seats as possible at the regular price of \$190. Clearly we are only interested in values of $i = 0, \dots, 100$. In that case we have

$$E[V_i] = (i \times 69) + (E[\min\{100 - i, \text{round}(X)\}] \times 190) \quad (2)$$

where we recall $X \sim N(90, 10)$ and that $\min\{100 - i, \text{round}(X)\}$ is the number of regular-fare seats that the airline will sell if it has sold i seats at the discount fare. Let's simplify the expectation on the r.h.s. of (2). We have

$$\begin{aligned} E[\min\{100 - i, \text{round}(X)\}] &= 0 \times P(X \leq 0.5) \\ &+ 1 \times P(0.5 \leq X < 1.5) \\ &+ 2 \times P(1.5 \leq X < 2.5) \\ &\vdots \\ &+ (99 - i) \times P(98.5 - i \leq X < 99.5 - i) \\ &+ (100 - i) \times P(99.5 - i \leq X). \end{aligned}$$

It is easy to write a piece of R code (or use a spreadsheet) to evaluate this expression for any $i \in \{0, 1, \dots, 100\}$. Once we have done so we can determine which value of i has the largest $E[V_i]$ and therefore maximises the total expected revenue. We find the optimal value of i is $i = 6$ or $i = 7$ with a corresponding expected revenue of \$17,076.46. (To the nearest cent they have identical expected revenues.) Code for this problem can be found in the R Notebook *R_Practice_Problem_Calcs.Rmd*.

14. Correlation and Independence? (Challenging!)

Suppose that random variable X is $N(0, 1)$ and that random variable Z is equally likely to be X or $-X$.

(i) Why is it true that

$$P(z \leq Z \leq z + \Delta) = \frac{1}{2}P(z \leq X \leq z + \Delta) + \frac{1}{2}P(-z - \Delta \leq X \leq -z)?$$

Solution: We have

$$\begin{aligned} P(z \leq Z \leq z + \Delta) &= P(z \leq Z \leq z + \Delta \mid Z = X)P(Z = X) + P(z \leq Z \leq z + \Delta \mid Z = -X)P(Z = -X) \\ &= \frac{1}{2}P(z \leq X \leq z + \Delta) + \frac{1}{2}P(z \leq -X \leq z + \Delta) \\ &= \frac{1}{2}P(z \leq X \leq z + \Delta) + \frac{1}{2}P(-z \geq X \geq -z - \Delta) \end{aligned}$$

as desired.

(ii) Before X is known, does Z have the distribution $N(0, 1)$?

Solution: Yes! To see this, note that for Δ very small part (i) implies

$$\begin{aligned} P(Z \in (z, z + \Delta)) &= \frac{1}{2}P(z \leq X \leq z + \Delta) + \frac{1}{2}P(-z - \Delta \leq X \leq -z) \\ &= \frac{1}{2}f_X(z)\Delta + \frac{1}{2}f_X(-z)\Delta \end{aligned} \tag{3}$$

where $f_X(x) = e^{-x^2/2}/\sqrt{2\pi}$ is the PDF of an $N(0, 1)$ random variable. But since $f_X(z) = f_X(-z)$ we can simplify (3) to obtain

$$P(Z \in (z, z + \Delta)) = f_X(z)\Delta$$

and so f_X is also the PDF of Z , i.e. $Z \sim N(0, 1)$.

(iii) Once it is known that $X = 5$, what is the distribution of Z ?

Solution: $Z \mid X = 5$ is binary and can only take on the values 5 or -5, each w.p. 1/2.

(iv) Are X and Z independent random variables?

Solution: No, definitely not. While X and Z are both $N(0, 1)$, we have shown that the distribution of $Z \mid X$ is not $N(0, 1)$. Hence they are not independent.

(v) Find the coefficient of correlation ρ_{XZ} of X and Z .

Solution: By definition,

$$\rho_{XZ} = \frac{E[XZ] - E[X]E[Z]}{\sqrt{\text{Var}(X)\text{Var}(Z)}}$$

We already know $E[X] = E[Z] = 0$ and $\text{Var}(X) = \text{Var}(Z) = 1$. So we just need to determine $E[XZ]$. This we can use the conditional expectation formula

$$\begin{aligned} E[XZ] &= E[E[XZ | X]] \\ &= E[XE[Z | X]] \\ &= E\left[X\left(\frac{1}{2}X + \frac{1}{2}(-X)\right)\right] \\ &= 0. \end{aligned}$$

It therefore follows that $\rho_{XZ} = 0$.

- (vi) Have you just shown that uncorrelated random variables need not be independent? (Just answer yes or no.)

Solution: Yes!