Math 321 – Summer 2019

additional practice for chapters 2-3, probability theory and distributions

1. A batch of fuses produced by a particular factory are tested. There are two types of fuses. 55% of the fuses are type I, and 45% are type II. 5% of the type I fuses fail to meet quality specifications, and 4% of type II fail.

   (a) What is the probability a type I fuse satisfies quality specifications?

   Solution:

   Let T_i = the event that a fuse is type i. We use the fact that fuse type is a partition: $S = T_1 \cup T_2$. Let A = the event that a fuse satisfies quality specs. We want to calculate $P(A \mid T_1)$. We are given $P(A^c \mid T_1) = 0.05$ thus by the probability rule for complements $P(A \mid T_1) = 1 - P(A^c \mid T_1) = 0.95$.

   (b) If a fuse is randomly selected, what is the probability it is of type I and satisfied quality specs?

   Solution:

   $P(A \cap T_1) = P(A \mid T_1)P(T_1) = (0.95)(0.55) = 0.5225$

   (c) What is the probability a randomly selected fuse satisfies quality specs?

   Solution:

$$\begin{aligned} P(A) &= P(A \cap T_1) + P(A \cap T_2) \\ &= P(A \mid T_1)P(T_1) + P(A \mid T_2)P(T_2) \\ &= (0.95)(0.55) + (0.96)(0.45) \\ &= 0.9545. \end{aligned}$$

   Or in more plain language:

$$\begin{aligned} P(\text{satisfy}) &= P(\text{satisfy} \cap \text{type I}) + P(\text{satisfy} \cap \text{type II}) \\ &= P(\text{satisfy} \mid \text{type I})P(\text{type I}) + P(\text{satisfy} \mid \text{type II})P(\text{type II}) \end{aligned}$$

   (d) Given that we know a particular fuse has failed to meet quality specs, what is the probability it is type I?

   Solution:

   Now this will involve Baye's theorem since we know $P(A^c \mid T_1)$ but want to calculate $P(T_1 \mid A^c)$.

$$\begin{aligned} P(T_1 \mid A^c) &= \frac{P(A^c \cap T_1)}{P(A^c)} \\ &= \frac{P(A^c \mid T_1)P(T_1)}{P(A^c \mid T_1)P(T_1) + P(A^c \mid T_2)P(T_2)} \\ &= \frac{(0.05)(0.55)}{(0.05)(0.55) + (0.04)(0.45)} \\ &= \frac{275}{275 + 180} \\ &\approx 0.6044 \end{aligned}$$

   (e) Are fuse type and whether it passes quality specs independent?

   Solution:

   $P(A) = 0.9545$ and $P(T_1) = 0.55$ but $P(A \cap T_1) = 0.5225$. This is not the same as $P(A) \cdot P(T_1) = (0.9545) \cdot (0.55) = 0.524975$. Thus they are not independent.

   There are many ways to check this:

   $P(A) = 0.9545 \neq 0.95 = P(A \mid T_1)$.

   $P(A) = 0.9545 \neq 0.96 = P(A \mid T_2)$.

2. We roll a pair of 6-sided dice. Let $A = \{\text{sum is even}\}$, and $B = \{\text{both dice are even}\}$. Calculate $P(A)$, $P(B)$, $P(A \mid B)$, and $P(B \mid A)$. Are A and B independent?

Solution:

Let's first list out all outcomes in these events. To save on space, I will list the outcomes as two-digit numbers, with the first digit being the first die and the second digit being the second die, e.g. 63 is when the first die is 6 and the second die is 3.

$$A = \{11, 33, 55, 13, 31, 15, 51, 35, 53, 22, 44, 66, 24, 42, 26, 62, 46, 64\}$$
$$B = \{22, 44, 66, 24, 42, 26, 62, 46, 64\}$$

$P(A) = 18/36 = 0.5$, $P(B) = 9/36 = 0.25$, $P(A \mid B) = 1$ since $B \subset A$, and
$P(B \mid A) = \frac{9}{18} = 0.5 \neq P(B)$ thus they are not independent.

3. Now let's roll a single 6-sided die. Let events: $A = \{2, 4, 6\} = \{\text{even}\}$, and $B = \{1, 2\}$, $C = \{1, 2, 3\}$, $D = \{1, 2, 3, 4\}$. Is A independent of any of the other 3 events?

Solution:

$P(A) = \frac{3}{6} = 0.5$.

$P(A \mid B) = \frac{1}{2} = 0.5 = P(A)$.

$P(A \mid C) = \frac{1}{3}$.

$P(A \mid D) = \frac{2}{4} = 0.5 = P(A)$.

Thus A and D are independent, and A and B are independent. A and C are not independent. Note that independence is a property that applies to a pair of events.

4. A particular political candidate is running for office. It is known that 20% of party D voters support the candidate, that 60% of party R voters support the candidate, and that 40% of independent (no party affiliation) voters support the candidate. The general population of voters is made up of equal proportions of each party but 20% of the population are independents. Given that a randomly selected voter supports the candidate, what is the probability they are a member of party D? Are political party membership and support for the candidate independent events?

Solution:

Let events $D = \{\text{voter is member of party D}\}$, $R = \{\text{voter is member of party R}\}$, $I = \{\text{voter is independent}\}$, and $S = \{\text{voter supports the candidate}\}$. We are given that $P(S \mid D) = 0.2$, $P(S \mid R) = 0.6$, $P(D) = P(R) = 0.40$, and $P(I) = 0.2$.

$$\begin{aligned} P(D \mid S) &= \frac{P(S \mid D)P(D)}{P(S \mid D)P(D) + P(S \mid R)P(R) + P(S \mid I)P(I)} \\ &= \frac{0.2 \cdot 0.4}{0.2 \cdot 0.4 + 0.6 \cdot 0.4 + 0.4 \cdot 0.2} = \frac{8}{8 + 24 + 8} = \frac{1}{5} = 20\% \end{aligned}$$

$$\begin{aligned} P(R \mid S) &= \frac{P(S \mid R)P(R)}{P(S \mid D)P(D) + P(S \mid R)P(R) + P(S \mid I)P(I)} \\ &= \frac{0.6 \cdot 0.4}{0.2 \cdot 0.4 + 0.6 \cdot 0.4 + 0.4 \cdot 0.2} = \frac{24}{8 + 24 + 8} = \frac{3}{5} = 60\% \end{aligned}$$

$$\begin{aligned} P(I \mid S) &= \frac{P(S \mid I)P(I)}{P(S \mid D)P(D) + P(S \mid R)P(R) + P(S \mid I)P(I)} \\ &= \frac{0.4 \cdot 0.2}{0.2 \cdot 0.4 + 0.6 \cdot 0.4 + 0.4 \cdot 0.2} = \frac{8}{8 + 24 + 8} = \frac{1}{5} = 20\% \end{aligned}$$

Note that $P(D) = 40\% \neq P(D \mid S) = 20\%$, $P(R) = 40\% \neq P(R \mid S) = 60\%$, but $P(I) = 20\% = P(I \mid S)$. Thus, at least for this example, independent voters truly are *independent* in the probability theory sense!

Note that this is just a quirk of the numbers used in this problem. If you alter them slightly, the independence of independent voters goes away.

5. If you flip three fair coins, and there is at least one head, what is the probability that there are at least two heads?

   Solution:

   The only outcome excluded from {at least one head} is the all tails outcome, thus a total of $2^3 - 1 = 7$ outcomes. How many ways are there to get at least two heads? There are $\binom{3}{2}$ ways to get two heads and $\binom{3}{3}$ ways to get three heads, thus a total of 4 outcomes. All 4 of these outcomes also have at least one head. Thus the probability $P(\text{at least 2 H's} \mid \text{at least 1 H}) = \frac{4}{7}$.

6. Let's look at flipping two fair coins. Show that the coins are *independent*.

   Solution:

   That the coins are independent satisfies intuition. $S = \{HH, HT, TH, TT\}$ and we assume the outcomes are all equally likely. Let $A =$ {first coin is H}={HH,HT}, and $B =$ {second coin is H}={HH,TH}. So $P(A) = P(B) = \frac{1}{2}$ and $P(A \cap B) = \frac{1}{4} = P(A) \cdot P(B)$.

   This is enough to show they are independent. It is a rule that when A and B are independent, then so are the complementary events A^c and B, A and B^c, and A^c and B^c.

7. Consider the probability mass function given below.

x	1	2	3	4
$f(x)$	0.4	0.3	0.2	0.1

   a) Calculate the mean.
   b) Calculate the variance.
   c) Calculate $E(10 - 3X)$.

   Solution:

   a) $E(X) = 1 \cdot 0.4 + 2 \cdot 0.3 + 3 \cdot 0.2 + 4 \cdot 0.1 = 0.4 + 0.6 + 0.6 + 0.4 = 2.0$.
   b) $E(X^2) = 1 \cdot 0.4 + 2^2 \cdot 0.3 + 3^2 \cdot 0.2 + 4^2 \cdot 0.1 = 0.4 + 1.2 + 1.8 + 1.6 = 5.0$. So $Var(X) = E(X^2) - \mu^2 = 5 - 2^2 = 1.0$.
   c) $E(10 - 3X) = 10 - 3E(X) = 10 - 3 \cdot 2 = 4$.

8. A production line produces items which are defective with probability 2%. Suppose 1,500 items are produced during a day.

   a) Write the probability that x items are defective.
   b) Calculate the probability that at least 50 items are defective.
   c) Calculate the probability that at least 20 but no more than 40 items are defective.

   Solution:

   a) $X =$ the number of defective items out of the 1500 total, so $X \sim Bin(1500, 0.02)$ and has pmf given by $P(X = x) = \binom{1500}{x}(0.02)^x(0.98)^{1500-x}$
   b) $P(X \geq 50) = 1 - P(X < 50) = 1 - P(X \leq 49) =$ `1-pbinom(49,1500,0.02)`=0.0004519813
   c) $P(20 \leq X \leq 40) =$ `sum(dbinom(20:40,1500,0.02))`=0.9481877

9. Flip a fair coin 10 times. Calculate the probability that you get either 2 or 3 heads.

   Solution:

   $X =$ the number of head, so $X \sim Bin(10, 0.5)$ and $P(X = x) = \binom{10}{x}(0.5)^x(0.5)^{10-x} = \binom{10}{x}(0.5)^{10}$.
   $P(X = 2 \text{ or } 3) = P(X = 2) + P(X = 3) = \binom{10}{2}(0.5)^{10} + \binom{10}{3}(0.5)^{10} = \frac{1}{2^{10}}\left(\frac{10 \cdot 9}{2 \cdot 1} + \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1}\right) = 0.1611328$.

10. Suppose a storage closet contains 25 rock samples, 17 of them are volcanic in origin and 8 of them are sedimentary. If a geologist selects 10 rock samples randomly, what is the probability that x of them are sedimentary?

    Solution:

    Let X be the number of sedimentary rocks in the sample, so $x = 0, 1, \ldots, 7, 8$ are possible. The probability of x sedimentary rocks is given by $P(X = x) = \frac{\binom{8}{x}\binom{17}{10-x}}{\binom{25}{10}}$. This is technically called a hypergeometric random variable.

11. Suppose a manufacturer produces PVC pipe that has on average 3 defects per 100 ft of pipe. If a wastewater construction agency purchases 250 ft of pipe, what is the probability that it will have a total of less than 3 defects?

    Solution:

    X = the number of defects on the 250 ft length of pipe. The rate parameter needs to be scale to $\lambda = 3 \cdot 250/100 = 7.5$. $p(x) = \frac{e^{-7.5}(7.5)^x}{x!}$. Thus $P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{e^{-7.5}(7.5)^0}{0!} + \frac{e^{-7.5}(7.5)^1}{1!} + \frac{e^{-7.5}(7.5)^2}{2!}$. In R this can be calculated by `sum(dpois(0:2,7.5))`=0.02025672, or by `ppois(2,7.5)`.

12. Flip a biased coin with probability of heads being 0.2. Refer to getting heads as a success.

    (a) If we flip the coin 20 times, what is the probability of getting 13 heads?

    Solution:

    Let $X = \{\#\text{ of heads (successes)}\}$ so we have that $X \sim Bin(n = 20, p = 0.2)$. The pmf for binomial is:

    $$P(X = x) = f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

    Thus

    $$P(X = 13) = f(13) = \binom{20}{13} (0.2)^{13} (0.8)^7 \approx 0.00001331783.$$

    (b) Now, let's flip the coin until the first head occurs. Write the probability mass function for $X = \{\text{the total number of flips up to and including the first head}\}$. What is the probability that there are 5 tails before the first head?

    Solution:

    Now, this random variable is geometric, $X \sim Geom(p = 0.2)$. The pmf is

    $$P(X = x) = p(1-p)^{x-1}.$$

    You don't necessarily have to have this memorized if you understand that it is a sequence of $x - 1$ failures and then a single success, and that the trials are independent.

    Getting 5 tails before the first head would be 6 total trials, thus $X = 6$:

    $$P(X = 6) = 0.2(0.8)^5 \approx 0.065536.$$

13. Consider a population of animals where adult weight is normally distributed with mean 7.3 kg and standard deviation 1.8 kg.

    (a) If an individual is randomly selected, what is the probability its weight is less than 5 kg?

    Solution:

    $P(X < 5) =$ `pnorm(5,7.3,1.8)`.

(b) What is the probability that a randomly selected individual's weight will be between 8.2 and 9.5 kg?

Solution:

$P(8 < X < 9) =$ `pnorm(9.5,7.3,1.8)-pnorm(8.2,7.3,1.8)`.

(c) What is the probability that a randomly selected individual's weight will be between 5.5 and 9.1 kg?

Solution:

$7.3 - 1.8 = 5.5$ and $7.3 + 1.8 = 9.1$ so this is plus and minus 1 standard deviation. The probability of that range is approximately 68% by the 68-95-99.7 rule.

(d) What is the probability that a randomly selected individual's weight will be above 10.6 kg?

Solution:

$7.3 + 2 \cdot 1.8 = 10.6$ so this is 2 standard deviations above the mean. The probability of that range is approximately $\frac{1}{2}(1 - 95\%) = 2.5\%$ by the 68-95-99.7 rule.

Generally, when you have an exact number of standard deviations, it is ok to use this rule for approximation of probabilities.

(e) If a sample of 12 individuals is taken, what is the probability the mean weight will be less than 5 kg?

Solution:

$P(\overline{X}_{12} < 5) =$ `pnorm(5,7.3,1.8/sqrt(12))`.

(f) What is the probability that the mean weight is exactly 5 kg?

Solution:

$P(\overline{X}) = 5) = 0$. It is a continuous random variable, so the probability on each individual value is zero.

(g) If an iid sample of 25 individuals is selected, what is the probability that every individual's weight is in the range of 5 to 10 kg?

Solution:

$P(5 < X < 10) =$ `pnorm(10,7.3,1.8)-pnorm(5,7.3,1.8)` ≈ 0.83 for each individual. "iid" means they are independent so we just multiply all the probabilities: $P(\text{all between 5 and 10}) \approx (0.83)^{25} \approx 0.01$.

(h) What is the probability that $\frac{X_3 - 9}{1.8} < -1$?

Solution:

$\frac{X-\mu}{\sigma} = Z \sim N(0,1)$ thus $P(\frac{X_3-9}{1.8} < -1) \approx \frac{1}{2}(1 - 0.68)$ by the 68-95-99.7 rule.

14. Consider a manufacturer producing a thin laminate material for protective coatings of surfaces. The laminate material is produced continually with a width of 1 m and rolled onto spools that hold a total length of 200 m. Assume it is known that on average there are 3 imperfections per 40 m^2.

(a) If a particular customer is to buy a complete spool of this material, what is the probability that the total number of imperfections is greater than 25?

Solution:

The number of imperfections on the 200 m spool is Poisson with rate $\lambda = 200 \cdot 3/40 = 15$. $P(\text{more than 25 imperfections}) = P(X > 25) =$ `1-ppois(25,lambda=15)` ≈ 0.006184904.

(b) If that customer needs at least four 1 m by 10 m sections with no imperfections (a total of 1 m by 40 m), what is the probability that the first 40 m on the spool have no imperfections?

Solution:

The number of imperfections in the first 40 m is Poisson with rate $\lambda = 40 \cdot 3/10 = 12$. $P(\text{no imperfections}) = P(X = 0) =$ `dpois(0,lambda=12)` $= e^{-12}$. Very unlikely!

(c) If a customer shows up at the laminate manufacturers and needs a 1 m by 20 m section with no imperfections, what is the probability that such a section will be produced?

Solution:

Here X is the length of material produced between imperfections. The length of material produced before the next imperfection is exponentially distributed with rate $\lambda = 3/10$, $X \sim Exp(3/10)$ so $P(X > 20) = \int_{20}^{\infty} \lambda e^{-\lambda x} dx = e^{-\lambda \cdot 20} = e^{-6} \approx 0.002478752$. So the customer will likely not be happy!

15. Consider the pdf $f(x) = a(16 - x^2)$ for $0 \leq x \leq 4$.

(a) Find a.

Solution:

$\int_0^4 a(16 - x^2)dx = a(16x - \frac{1}{3}x^3)\Big|_0^4 = a(64 - 64/3)$ thus $a = \frac{3}{128}$.

(b) Find the cdf.

Solution:

$F(x) = \int_0^x a(16 - x^2)dx = a(16x - \frac{1}{3}x^3)\Big|_0^x = a(16x - x^3/3)$. So plugging in what we got for a and simplifying a bit gives $F(x) = \frac{1}{128}x(48 - x^2)$.

Note that $F(0) = 0$ and $F(4) = \frac{1}{128} \cdot 4 \cdot (48 - 4^2) = \frac{1}{32} \cdot 32 = 1$ as required for a cdf.

(c) Find $E(X)$.

Solution:

$E(X) = \int_0^4 ax(16 - x^2)dx = a(8x^2 - \frac{1}{4}x^4)\Big|_0^4 = a(128 - 4^3) = a \cdot 64 = \frac{3}{2}$

16. Consider the pdf $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$. Note that this is the exponential pdf.

(a) Find the median.

Solution:

The cdf is $F(x) = 1 - e^{-\lambda x}$. The median $\tilde{\mu}$ is such that $F(\tilde{\mu}) = 0.5$. So we set $F(x) = 0.5$ and solve for x, that will be the median.

$0.5 = 1 - e^{-\lambda x}$ gives $x = -\frac{\ln(0.5)}{\lambda} = \frac{\ln 2}{\lambda}$. Thus $\tilde{\mu} = \frac{\ln 2}{\lambda}$. Note that we have used a few logarithm properties here, notably $r \log M = \log M^r$. You should make sure you remember all properties of logs and exponentials.

(b) Find the first quartile.

Solution:

The first quartile Q_1 is the 25^{th} percentile so we set $F(x) = 0.25$ and solve for x.

$0.25 = 1 - e^{-\lambda x}$ gives $x = -\frac{\ln(0.75)}{\lambda} = \frac{\ln \frac{4}{3}}{\lambda} = \frac{\ln 4 - \ln 3}{\lambda}$.

Some further info:

Thus to find the p-percentile, we set $p = F(x)$ and solve for x:

$$\tilde{x}_p = -\frac{\ln(1 - p)}{\lambda}$$

17. Consider the pmf

x	0	10	20
$f(x)$	0.10	0.60	0.30

(a) Calculate $E(X)$ and $Var(X)$.

Solution:

$E(X) = \sum x \cdot f(x) = 0(0.1) + 10(0.6) + 20(0.3) = 0 + 6 + 6 = 12.$

We know that: $Var(X) = E(X^2) - E(X)^2$. SO we calculate $E(X^2)$.
$E(X^2) = \sum x \cdot f(x) = 0^2(0.1) + 10^2(0.6) + 20^2(0.3) = 0 + 60 + 120 = 180$.
Thus $Var(X) = 180 - 12^2 = 36$.

(b) Calculate $E(100 + 2X)$ and $Var(100 + 2X)$.

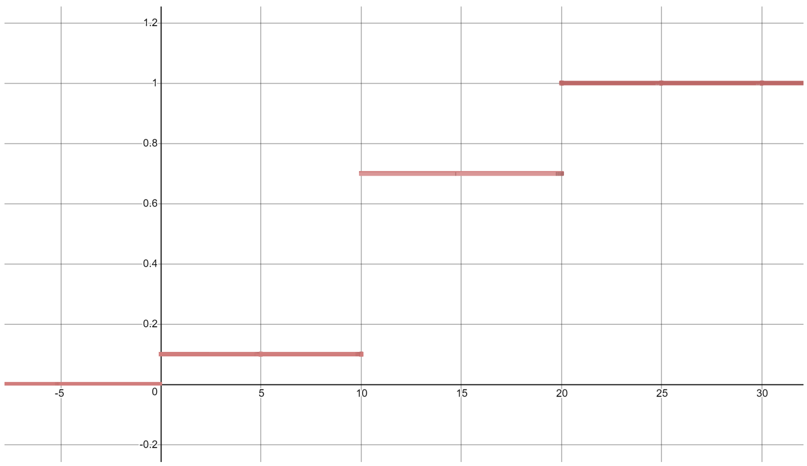
Solution:
$E(aX + b) = aE(X) + b$ and $Var(aX + b) = a^2 Var(X)$.
So $E(100 + 2X) = 100 + 2 \cdot 180 = 460$ and $Var(100 + 2X) = 4 \cdot 36 = 128$.

(c) Write the formula for and sketch a graph of the cdf.

Solution:

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ 0.1 & \text{for } 0 \leq x < 10 \\ 0.7 & \text{for } 10 \leq x < 20 \\ 1 & \text{for } 20 \leq x \end{cases}$$

Here is the graph:



Some further info:
The cdf always starts of at 0 at $x = -\infty$ and goes up to 1 at $x = \infty$. For a discrete random varianble, the cdf will be a step funciton where the steps occur at the possible values of x and the size of the jump is the probability for that x-value.

18. Consider the uniform random variable $X \sim U(0, 100)$.

(a) Write the pdf.

Solution:
$f(x) = \frac{1}{100}$ for $0 \leq x \leq 100$.

(b) Write the cdf.

Solution:
$F(x) = \int_0^x \frac{1}{100} du = \frac{x}{100}$

(c) Find $E(X)$.

Solution:
$E(x) = \int_0^{100} x \cdot \frac{1}{100} du = \frac{x^2}{200}|_0^{100} = \frac{100^2}{200} = 50$.
Some further info:
Note that for a uniform random variable, the expected value is always in the middle. $X \sim U(a, b)$ then $E(X) = \frac{a+b}{2}$.

(d) Find $P(50 < X < 70)$.

Solution:

This one we can calculate geometrically since the pdf is a constant function. This probability is the area of a rectangle with width $70 - 50 = 20$ and height $\frac{1}{100}$ thus the probability is $P(50 < X < 70) = 20 \cdot \frac{1}{100} = 20\%$.

(e) Find b so that $P(30 < X < b) = 0.6$.

Solution:

$0.6 = P(30 < X < b) = (b - 30) \cdot \frac{1}{100}$ thus $60 = b - 30$ so $b = 90$.

(f) Find the first quartile.

Solution:

$0.25 = F(x) = \frac{x}{100}$ thus $x = 25$. So we can see that $Q_1 = 25$. In general for this pdf, the $p(100)^{th}$ percentile will be $100p$.

Some further info:

In general for a uniform distribution $X \sim U(a, b)$ the cdf is $F(x) = \frac{x-a}{b-a}$ and so the $p(100)^{th}$ percentile will be given by $p = \frac{\tilde{x}_p - a}{b-a}$ thus $\tilde{x}_p = a + p(b - a)$.

Example: for $X \sim U(1, 5)$ then the 20^{th} percentile is $\tilde{x}_{0.2} = 1 + 0.2(5 - 1) = 1.8$.

19. Explain the following R code.

```
> x=c(0,2,4,6,8,10)
  p=c(0.69,0.01,0.05,0.12,0.03,0.10)
  m=sum(x*p)
  v=sum(x^2*p)-m^2
```

Solution:

This code calculates the expected value and variance for random variable X with given pmf:
`x` is a list of possible values for discrete random variable X,
`p` give the probabilities for each value,
`m` is the expected value of X, and
`v` is the variance of X.

20. Explain the following R code.

```
> x=rnorm(100,mean=25,sd=5)
  hist(x,breaks=seq(from=0,to=50,by=2))
```

Solution:

This code generates 100 random samples from a normal distribution with mean 25 and standard deviation 5, and plots a histogram.