Nathan Bickel
STAT 509: Section 003
Professor: Dr. Stutz
August 30, 2022

STAT 509 Homework 1b

Problem 1 You roll a die twice. Enumerate the sample space (Ω).

Solution.

Let $R = \{1, 2, 3, 4, 5, 6\}$. Let (r_1, r_2), where $r_1, r_2 \in R$, represent two rolls, where r_1 is the number rolled on the first die and r_2 is the number rolled on the second die. Then,

$$\Omega = \{(r_1, r_2) \in S^2\}.$$

Since each outcome is equally likely and $|\Omega| = 36$, the probability of each outcome in Ω is $\frac{1}{36}$.

Problem 2 Feeling lucky? I bet that at least two students in our class share a birthday (not including year). Use probability to determine if you should accept my bet or not. We have 40 students in the class and you can ignore the existence of leap years (i.e., 365 days per year).

Solution.

Let $B(n)$ be the probability of at least two people in a group of n people sharing the same birthday. For example, $B(2) = 1 - \frac{364}{365}$ because there are 364 ways the second person can not overlap with the first person's birthday and 1 way they can, and each way is equally likely. Similarly, $B(3) = 1 - \left(\frac{364}{365}\right)\left(\frac{363}{365}\right)$, because there are 364 ways for the first and second person to not overlap and 363 ways for the third person to not overlap with either of them. More generally,

$$B(n) = 1 - \left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\cdots\left(\frac{365 - n + 1}{365}\right) = 1 - \frac{365!}{(365 - n)!365^n}.$$

Using the general formula, $B(40) \approx 0.891$, so there is only around an 11% chance that a class of 40 people would have no pairs of people that share a birthday. Based on this, I wouldn't take the bet.

Problem 3 Consider two variables: one denotes whether a person truly has a disease (D) or not (D^c) and the other denotes whether a person tests positively for the disease (Pos) or not (Neg). Use the contingency table to answer the questions.

Labels	D	D^c
Pos	53	4
Neg	2	455

(a) Calculate $P(Pos \cap D)$.
(b) Calculate $P(Neg \cup D)$.
(c) Calculate $P(Pos|D)$ (this is called the sensitivity).
(d) Calculate $P(Neg|D^c)$ (this is called the specificity).

Solution.

(a) $P(Pos \cap D)$ represents the percentage of the sample space where a person has a disease and tests positive for it. This is the top left cell, which is 53. There are 514 events, so the probability is $\frac{53}{514} \approx 0.103$.

(b) $P(Neg \cup D)$ represents the cases where a person either has a disease, tests negative, or both. These three cases are the top left, bottom left, and bottom right cells, and we can add them since they are disjoint. Since $53 + 2 + 455 = 510$, the probability is $\frac{510}{514} \approx 0.992$.

(c) $P(Pos|D)$ represents the percentage of times a person with a disease tested positive for it. There were $53 + 2 = 55$ people with diseases, and 53 of those people tested positive, so the probability is $\frac{53}{55} \approx 0.963$.

(d) $P(Neg|D^c)$ represents the percentage of times without a disease tested negative for it. There were $4 + 455 = 459$ people without diseases, and 455 of those people tested negative, so the probability is $\frac{455}{459} \approx 0.991$.

Problem 4

(a) You keep rolling a set of five dice until you get a set showing either exactly one six or no sixes. You win if there is exactly one six and you lose if there are no sixes. What is the probability that you win?

(b) Ender and Harry are playing game in which they alternate making moves. When it is Ender's turn, the probability that he wins with his next move is 40%. When it is Harry's turn, the probability that he wins with his next move is p. If Ender is allowed to go first, what is the value of p such that neither player is favored to win?

Solution.

(a) The number of 6s in a roll here follows a binomial distribution, as each of the dice in the set of 5 can be thought of as a trial. Each trial will either result in a 6 or not a 6, each trial is independent from the other 4, there are always 5 trials, and the probability of success (getting a 6) remains fixed at $\frac{1}{6}$. Using the binomial distribution formula, the probability of getting no 6s when rolling a set of 5 dice is

$$\binom{5}{0}\left(\frac{1}{6}\right)^{0}\left(1-\frac{1}{6}\right)^{5-0} = \frac{5^5}{6^5}.$$

Similarly, the probability of getting exactly 1 6 when rolling a set of 5 dice is

$$\binom{5}{1}\left(\frac{1}{6}\right)^{1}\left(1-\frac{1}{6}\right)^{5-1} = \frac{5^5}{6^5}.$$

Note that these are equal. Now, assume that one of these two events happens. Then, since we win if there is exactly one 6, the chance of winning given that there are 0 or 1 6s is $\frac{1}{2}$ since the two probabilities are equal.

Since you keep rolling the set of dice indefinitely until one of these two events happen, the probability of it happening eventually approaches 1 as the number of rolls approaches infinity. Therefore, the probability of winning is $(1)\left(\frac{1}{2}\right) = \frac{1}{2}$.

Problem 5 Calculate $E(X^3)$, where X has the following distribution:

$$f(x) = \begin{cases} .5 & x = 1 \\ .3 & x = 2 \\ .2 & x = 3 \end{cases}$$

Solution.

We have that for any discrete random variable X with a PMF $f(x)$,

$$E(X) = \sum_{\text{all } x} x f(x).$$

Thus, we can use the formula to compute

$$\begin{aligned} E(X^3) &= \sum_{\text{all } x} x^3 f(x) \\ &= (1^3)(0.5) + (2^3)(0.3) + (3^3)(0.2) \\ &= 8.3. \end{aligned}$$

Problem 6 Prove that the $E(X) = \frac{1}{p}$ for a geometric distribution, where x is the number of trials. Recall that I started this in class.

We have that for any discrete random variable X with a PMF $f(x)$,

$$E(X) = \sum_{\text{all } x} x f(x).$$

For a geometric distribution, $f(x) = p(1-p)^{x-1}$ for $p \in \mathbb{R}$, $0 < p < 1$ where p is the probability of a success. The set of values x can take is $\mathbb{N} - \{0\}$. Thus, we have

$$\begin{aligned} E(X) &= \sum_{x=1}^{\infty} xp(1-p)^{x-1} \\ &= -p \sum_{x=0}^{\infty} -x(1-p)^{x-1} && \text{(adjusting bound: the } x = 0 \text{ term is simply adding 0)} \\ &= -p \sum_{x=0}^{\infty} \frac{d}{dp} \left[(1-p)^x\right] && \text{(rewriting as derivative to get rid of } x \text{ as product)} \\ &= -p \frac{d}{dp} \left[\sum_{x=0}^{\infty} (1-p)^x\right] && \text{(sum property of derivative)} \\ &= -p \frac{d}{dp} \left[\frac{1}{1-(1-p)}\right] && \text{(using geometric series since } |1-p| < 1) \\ &= -p \frac{d}{dp} \left[\frac{1}{p}\right] \\ &= -p \left(-\frac{1}{p^2}\right) && \text{(differentiating)} \\ &= \frac{1}{p} \end{aligned}$$

Problem 7 You are designing a road system leading to a city. There are 9 lights on the road and the probability that a light is red is .3. Assume that the probability that one light is red has nothing to do with any of the other lights being red (binomial).
(a) What is the probability that 4 of the lights are red? Show by hand and provide the appropriate R code.
(b) Consider that you are required to design another road leading to the city. This road has 8 lights. What must the probability of a red light be on this road so that neither road would be preferred over the other on the basis of expected number of red lights?
(c) Which road has more variability in the number of red lights?

(a) We can use a binomial distribution where X is the number of red traffic lights, so

$$P(X = 4) = \binom{9}{4}(0.3)^4(1-0.3)^{9-4} = \frac{3^4 7^5 126}{10^9} \approx 0.172.$$

```
> dbinom(4,9,0.3)
[1] 0.1715322
```

(b) We have that $E(X) = np$ for a variable X that follows a binomial distribution, where n is the number of trials and p is the probability of success. Let p_8 be the probability of a red light on the road with 8 lights. We can set the expected values equal to each other:

$$(9)(0.3) = 8p_8,$$

and thus $p_8 = \frac{27}{80} = 0.3375$.

(c) We have that $\mathrm{Var}(X) = np(1-p)$ for a variable X that follows a binomial distribution, where n is the number of trials and p is the probability of success. So the variance of the road with 9 lights is $(9)(0.3)(1-0.3) = 1.89$ and the variance of the road with 8 lights is $(8)(0.3375)(1-0.3375) = 1.78875$. Thus, the road with 9 lights has higher variability because it has a higher variance.

Problem 8 A manufacturing facility produces items with a defect rate of 1/500. Assume that defective items are produced independently of one another. Items are produced in batches of 1000. A batch is rejected if it contains more than 5 defective items.
(a) Assuming that the defect rate is constant, what is the probability that a randomly selected batch will be rejected? Show by hand (sigma notation is fine) and provide the appropriate R code. (More than one distribution could be used to yield the correct result.)
(b) You examine a batch and notice that it contains 7 defective items, but you know the company you are selling this batch to is not very thorough as they will only examine 500 of the items and reject the batch if more than 5 defectives are found. What is the probability that the purchaser fails to reject the batch? You may use R.

(a) We can use a binomial distribution where each item is a trial: an item is either defective or not defective, the trials are independent as described, the number of trials is 1000, and the probability of an item being defective is 1/500. A batch will be rejected if there are 6 or more defective items, the complement of which is that 5 or fewer items are defective. Thus, the probability of a batch being rejected is

$$1 - \sum_{n=0}^{5} \binom{1000}{n} \left(\frac{1}{500}\right)^n \left(1 - \frac{1}{500}\right)^{1000-n} \approx 0.0165.$$

```
> pbinom(5,1000,1/500,lower.tail=FALSE)
[1] 0.01645536
```

(b) There are $\binom{1000}{500}$ batches of 500 items the company can choose. Since there are 7 defective items and 993 non-defective ones, the number of possible batches with exactly 6 defective items and 494 non-defective items is $\binom{7}{6}\binom{993}{494}$, and the number of possible batches with exactly 7 defective items and 493 non-defective items is $\binom{7}{7}\binom{993}{493}$. Assuming that each item has an equal probability of being included in the batch, then, the probability of a batch containing 6 or 7 (all) of the defective items and thus being rejected is

$$\frac{\binom{7}{6}\binom{993}{494} + \binom{7}{7}\binom{993}{493}}{\binom{1000}{500}} \approx 0.0618.$$

Problem 9 The number of earthquakes that occur per week in California follows a Poisson distribution with a mean of 1.5.

(a) What is the probability that no earthquakes occur during a randomly selected week? Show by hand and provide the appropriate R code.
(b) What is the probability that more than 1 earthquake occurs? Show by hand and provide the appropriate R code.

(a) We have that for $X \sim \text{Poisson}(\lambda)$,

$$p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}.$$

Using the formula, the probability that 0 earthquakes occur with a mean of 2.5 is

$$p_X(0) = \frac{2.5^0 e^{-2.5}}{0!} \approx 0.0821.$$

```
> dpois(0,2.5)
[1] 0.082085
```

(b) The complement of this is that 0 or 1 earthquakes occur, which has a probability of

$$\frac{2.5^0 e^{-2.5}}{0!} + \frac{2.5^1 e^{-2.5}}{1!} \approx 0.287.$$

Thus, the probability that more than one earthquakes occur is $\approx 1 - 0.287 = 0.713$.

```
> ppois(1,2.5,lower.tail=FALSE)
[1] 0.7127025
```