Stats 116 Problem Set 4

Due: Wednesday, April 27 5:00 p.m. on Gradescope

Please show your work for each exercise. If you collaborate with someone else—this is fine—be sure to note that in your homework submission. You must each write up separate answer sets. Questions are either from Ross's *A First Course in Probability* or our are home-cooked.

Question 4.1 (Variant and extension of of Ross problem 4.30, the *St. Petersburg paradox*)**:** A person tosses a fair coin until a tail appears for the first time. If the tail appears on the nth flip, the person wins 2^n dollars. Let X denote the player's winnings.

(a) Show that $\mathbb{E}[X] = +\infty$

(b) Would you be willing to pay \$1 million to play this game once?

(c) Would you be willing to pay \$1 million for each game if you could play for as long as you liked and only had to settle up when you stopped playing?

(d) Suppose that the player is in the more realistic scenario that the casino against whom they play has a finite budget B. (For simplicity, assume B is a multiple of 2, i.e. $B = 2^k$ for some $k \in \mathbb{N}$.) In this case, the casino simply pays out the entire budget B once the player's winnings exceed B and the game ends. Let $Y = \min\{X, B\}$. Show that

$$\mathbb{E}[Y] = \log_2 B + 1.$$

(i) How much would you be willing to pay to play against the TAs, whose yearly salary (and hence wealth B) we assume are $2^{14} = \$16{,}384$?

(ii) How much would you be willing to pay to play against Jeff Bezos, whose approximate wealth is $2^{37} \approx \$137$ billion?

(iii) How about against the U.S. government, which collects around $2^{40} \approx \$1$ trillion in taxes?

Question 4.2 (Ross problem 4.40)**:** A card is drawn at random from an ordinary deck of 52 playing cards. After the card is drawn, it is replaced. The deck is reshuffled and another card is drawn at random. This process goes on indefinitely. What is the probability that exactly 3 out of the first 5 cards that have been drawn are red?

Question 4.3 (Ross theoretical exercise 4.15)**:** A family has n children with probability αp^n, $n \geq 1$, where $\alpha \leq \frac{1-p}{p}$ and $0 < p < 1$.

(a) What proportion of families has no children?

(b) If each child is equally likely to be a boy or a girl (independently of each other), what proportion of families consist of k boys (and any number of girls)?

Question 4.4 (Cheating your friends)**:** Find a coin: any coin will do, but we will assume you have access to (one of) a penny, a nickel, a dime, or a quarter. We will evaluate how fair your coin is and whether you can use a coin in a way undectable to your friends to frequently win gambling contests with them.

(a) Note the type of coin you have (e.g., U.S. quarter, or Canadian dollar piece, Euro, etc.).

(b) Flip your coin $n = 30$ times, letting it land either in your hand or on the floor (note which you did). How many times was it heads? How many tails?

(c) Spin your coin on your desk, flat table, or other smooth surface $n = 30$ times (that is, stand it on its side, then use your thumb and another finger to spin it). Any time it falls off the surface or hits an obstruction before it falls naturally, spin it again. Record the number of heads and tails you achieve.

(d) Assume that both modes of randomizing your coin are fair (flipping and spinning). Are your results within what you would expect? More precisely, let X be the number of times your coin lands heads after flipping, and Y the number of times your coin lands heads after spinning it. If both are fair, we expect $\mathbb{E}[X] = \mathbb{E}[Y] = \frac{n}{2}$, and $\mathrm{Var}(X) = \mathrm{Var}(Y) = \frac{n}{4}$. Compute (for your realizations) the values

$$(X - \mathbb{E}[X])^2 \quad \text{and} \quad (Y - \mathbb{E}[Y])^2.$$

Are the values consistent with your coin being fair? Explain in one or two sentences.

Question 4.5 (Ross theoretical exercise 4.23, more or less)**:** An urn contains $2n$ balls, of which two are numbered 1, two are numbered 2, two are numbered 3, etc., and two are numbered n. (That is, for each $i = 1, \ldots, n$, two balls have number i.) One draws balls successively 2 at a time without replacement. Let T denote the first selection in which the balls withdrawn have the same number (where $T = \infty$ if no pair has identical numbers). We wish to show that for $0 \le \alpha < 1$ under "appropriate" approximations,

$$\lim_{n \to \infty} P(T > \alpha n) = \exp\left(-\frac{\alpha}{2}\right).$$

To verify the preceding formula, let M_k denote the number of pairs withdrawn in the first k selections, $k = 1, \ldots, n$. You may write X_i as the indicator that at the ith draw, the balls drawn match, and you may assume (this is obviously false, but that's fine) that the X_i are independent and identically distributed. Then

$$M_k = \sum_{i=1}^{k} X_i.$$

(a) Show that $P(X_1) = \frac{1}{2n-1} \approx \frac{1}{2n}$.

(b) Approximate $P(M_k = 0)$ when n is large using the Poisson approximation to the binomial.

(c) Write the event $\{T > \alpha n\}$ in terms of the value of one of the variables M_k.

(d) Verify the limiting probability given for $P(T > \alpha n)$ (again, using the Poisson approximation).

Question 4.6* (**Extra credit:** validity of the Poisson approximation)**:** In the Ross textbook, Ross writes that if $X \sim \mathsf{Binomial}(n, p)$, and np is "not too large" and i is "not too large," then a good approximation to $\mathbb{P}(X = i)$ is that

$$P(X = i) \approx P(Y = i) \tag{4.1}$$

for $Y \sim \mathsf{Poisson}(\lambda)$, where $\lambda = np$. In this question, we will make this approximation rigorous in the sense that if $np \ll \sqrt{n}$ and $i \ll \sqrt{n}$, this approximation will be accurate to within a small multiplicative factor.

(a) Show that if $f(x) = \log(1-x)$ and $g(x) = -x - x^2$, then $f(x) \geq g(x)$ for all $x \in [0, \frac{1}{4}]$. Conclude that

$$\exp(-x - x^2) \leq 1 - x \leq e^{-x},$$

where the first equality is valid when $0 \leq x \leq \frac{1}{4}$, while the second is valid for all x (you do not need to prove the second inequality).

(b) Show that if $\lambda = pn$, then

$$P(X = i) = \frac{n(n-1)\cdots(n-i+1)}{n^i}\left(1 - \frac{\lambda}{n}\right)^{n-i} \cdot \frac{\lambda^i}{i!}.$$

(c) Assume that for some $\epsilon > 0$, we have $\frac{\lambda i}{n} \leq \epsilon$. Show that

$$P(X = i) \leq \frac{e^{-\lambda}\lambda^i}{i!} \cdot e^{\epsilon}.$$

(d) Assume that for some $0 < \epsilon \leq \frac{1}{4}$, we have $\frac{\lambda}{n} \leq \epsilon$, $\frac{i^2}{n} \leq \epsilon$, and $\frac{\lambda^2}{n} \leq \epsilon$. Show that

$$P(X = i) \geq \frac{e^{-\lambda}\lambda^i}{i!} \cdot e^{-2\epsilon - \epsilon^2}.$$

(e) Combining the preceding two results, conclude that for any sequences of random variables $X_{n,p} \sim \mathsf{Binomial}(n, p)$ and $Y_{n,p} \sim \mathsf{Poisson}(n \cdot p)$ and any sequence ϵ_n satisfying $\epsilon_n \to 0$ as $n \to \infty$,

$$\max_i \max_p \left\{\left|\log \frac{P(X_{n,p} = i)}{P(Y_{n,p} = i)}\right| \text{ such that } 0 \leq p \leq \frac{\epsilon_n}{\sqrt{n}},\ 0 \leq i \leq \epsilon_n \sqrt{n}\right\} \to 0.$$

Conclude that (roughly) as long as p smaller than $1/\sqrt{n}$ and i is smaller than $\sqrt{n}$, the approximation (4.1) is valid for $X \sim \mathsf{Binomial}(n, p)$ and $Y \sim \mathsf{Poisson}(np)$.

Question 4.7* (**Extra credit:** Ross theoretical exercise 4.23, modified; the textbook author and publisher couldn't be bothered to do this.)**:** An urn contains $2n$ balls, of which two are numbered 1, two are numbered 2, two are numbered 3, etc., and two are numbered n. (That is, for each $i = 1, \ldots, n$, two balls have number i.) One draws balls successively 2 at a time without replacement. Let T denote the first selection in which the balls withdrawn have the same number (where $T = \infty$ if no pair has identical numbers). We wish to show that for α near 0 and the appropriate approximations,

$$\lim_{n \to \infty} P(T > \alpha n) = \sqrt{1 - \alpha}.$$

To verify the preceding formula, let M_k denote the number of pairs withdrawn in the first k selections, $k = 1, \ldots, n$.

(a) Argue that when n is large, we may approximate M_k as the number of successes in k (approximately) independent trials. (Note: you may wish to skip this part of the problem and do parts (b)–(d), as this part is a bit tedious to make rigorous.) To do so, employ the following steps:

(i) Let X_i be the indicator that at the ith draw, the balls drawn match. Show that

$$P(X_1 = 1) = \frac{1}{2n-1} \approx \frac{1}{2n}.$$

(ii) Let R_i be the collection ball numbers for which two balls remain after i draws. For example, $R_0 = \{1, 2, \ldots, n\}$, while if the first draw consists of balls numbered $\{1, 2\}$, then $R_1 = \{3, 4, \ldots, n\}$, and if the first two draws are $\{1, 1\}$ and $\{2, 5\}$, then $R_2 = \{3, 4, 6, 7, \ldots, n\}$. Let B_{i1} and B_{i2} denote the indices of the ith balls drawn from the bin (so that if the first draw is $\{1, 2\}$, then $B_{i1} = 1$ and $B_{i2} = 2$). After i draws of pairs of balls from the urn, show that the cardinality

$$|R_i| = n - i - \frac{s}{2},$$

where s is the number of singleton indices after i draws, that is, the number of indices that appear only once in the list of all balls drawn,

$$\{B_{11}, B_{12}, B_{21}, B_{22}, \ldots, B_{i1}, B_{i2}\}. \tag{4.2}$$

(iii) Show that if the values of the draws B_{j1}, B_{j2}, $j = 1, \ldots, i$, are such that there are s singletons in the list (4.2),

$$P(X_{i+1} = 1 \mid B_{11}, B_{12}, \ldots, B_{i1}, B_{i2}) = \frac{n - i - s/2}{(n-i)(2n - 2i - 1)} \approx \frac{1}{2(n-i)} - \frac{s}{4(n-i)^2},$$

where the approximation is valid when $i \le cn$ for any constant $c < 1$. *Hint:* write the event $X_{i+1} = 1$ as the event that $B_{i+1,1} = j$ and $B_{i+1,2} = j$ for some $j \in R_i$.

(iv) Let $x_1, \ldots, x_i \in \{0, 1\}$. Conclude by Bayes' rule that if $B(x_1, \ldots, x_i)$ is the set of lists (4.2) for which we have matching indicators $X_1 = x_1, X_2 = x_2, \ldots, X_i = x_i$ (i.e. the pair j matches if $x_j = 1$), then

$$\begin{aligned} P(X_{i+1} = 1 \mid X_1 = x_1, \ldots, X_i = x_i) &= \frac{\sum_{b \in B(x_1,\ldots,x_i)} P(X_{i+1} = 1, b)}{\sum_{b \in B(x_1,\ldots,x_i)} P(b)} \\ &\approx \frac{1}{2(n-i)} \frac{\sum_{b \in B(x_1,\ldots,x_i)} P(b)}{\sum_{b \in B(x_1,\ldots,x_i)} P(b)} = \frac{1}{2(n-i)} \end{aligned}$$

From these four steps, conclude that $P(X_{i+1} = 1 \mid X_1, \ldots, X_i) \approx P(X_{i+1} = 1)$ and so the X_i are *nearly* independent and each is approximately Bernoulli with $X_i \sim \mathsf{Bernoulli}(\frac{1}{2(n-i)})$.

(b) Approximate $P(M_k = 0)$ when n is large and $k \le cn$ for some constant $c < 1$. In particular, show that

$$P(M_k = 0) \approx \left(1 - \frac{1}{2n}\right) \cdot \left(1 - \frac{1}{2(n-1)}\right) \cdots \left(1 - \frac{1}{2(n-k+1)}\right).$$

Hint. Write $M_k = \sum_{i=1}^{k} X_i$ and assume the X_i are independent and distributed as in part (b). Then $M_k = 0$ if and only if what?

(c) Show that as long as $k \le cn$ for some constant $c < 1$,

$$\log P(M_k = 0) \approx \frac{1}{2}\log\left(1 - \frac{k-1}{n}\right) + \mathrm{error}(n) \approx \frac{1}{2}\log\left(1 - \frac{k}{n}\right) + \mathrm{error}(n)$$

where the term error(n) is an error term satisfying $|\mathrm{error}(n)| \le \frac{C}{n}$ for some C (you don't need to specify this). You may use the approximations that

$$\log(1-x) \approx -x - \frac{x^2}{2},$$

valid for $|x| \le \frac{1}{2}$, and that for any $m < n$, we have the Harmonic series identity

$$\sum_{i=m+1}^{n} \frac{1}{i} = \log\frac{n}{m} + \frac{1}{2n} - \frac{1}{2m} + \varepsilon_n - \varepsilon_m,$$

where ε_n is an error term satisfying $0 \le \varepsilon_n \le \frac{1}{8n^2}$.

(d) Write the event $\{T > \alpha n\}$ in terms of the value of one of the variables M_k.

(e) Verify the limiting probability given for $P(T > \alpha n)$.