STATS 116: Homework 4

Due: Thursday, July 27, 2023 at 10:00 pm PDT on Gradescope

There are 7 problems on this assignment, each worth 8 points, although subparts within a problem may not be equally weighted. Credit will be assigned primarily based on reasoning and work, not the final answer. You do not need to simplify arithmetic expressions unless otherwise noted. While you may discuss the problems on this assignment other students, you must write up your own solutions. As per the syllabus, you may occasionally use the Internet or other public resources to clarify concepts with citation when this information is used as part of your own solution to a homework problem. However, you may not search for direct solutions to any problems assigned for homework or exams. For example, you can ask ChatGPT to clarify a particular concept from lecture that may be related to a problem, but you cannot feed it any part of a course assignment or a substantively similar version.

1. The Chinese card game *tuo la ji* involves four players and two shuffled decks of cards. Each deck consists of the standard 52 cards along with a black joker and a red joker, so the total number of cards being drawn from is 108. At the beginning of the game, each player draws 25 cards at random from these two decks (without replacement). The 8 remaining cards are placed face down and form what is known as the kitty.

   (a) On average, how many spades can a player expect to draw?

   Note there are $13 \cdot 2 = 26$ spades among the two decks. Number them $1, \ldots, 26$. Let $I_i, i = 1, \ldots, 26$ be the indicator r.v. for the event A_i that the player draws the i-th spade. Then X, the number of spades drawn by the player, is equal to the sum $I_1 + \cdots + I_{26}$. Note $\mathbb{E}[I_i] = P(A_i) = 25/108$ (WLOG the player draws the first 25 cards; any given card has probability $25/108$ lying among the first 25). Then by linearity

   $$\mathbb{E}[X] = \sum_{i=1}^{26} P(A_i) = \boxed{\frac{25 \cdot 26}{108}} \approx 6.02$$

   (b) On average, how many pairs can a player expect to draw? A pair is defined as two identical looking cards, e.g. two queens of spades, or two black jokers.

   There are 54 pairs. Let $I_i, i = 1, \ldots, 54$ be the indicator random variable for the event A_i that the player draws both of the cards in pair i. Then if X is the number of pairs the player draws, we have $X = I_1 + \cdots + I_{54}$. We compute

   $$P(A_i) = \frac{\binom{106}{23}}{\binom{108}{25}} = \frac{\frac{106!}{23!83!}}{\frac{108!}{25!83!}} = \frac{25 \cdot 24}{108 \cdot 107}$$

Thus by linearity

$$\mathbb{E}[X] = \sum_{i=1}^{54} P(A_i) = \boxed{54 \cdot \frac{25 \cdot 24}{108 \cdot 107}} \approx 2.80$$

2. Suppose n computers are attacked simultaneously by k hackers. Each hacker randomly selects one (and only one) computer to break into, independently from the other hackers, and each hacker successfully breaks into the selected computer with probability p, regardless of which computer was chosen. On average, how many computers end up being broken into?

Let $I_i, i = 1, \ldots, n$ be the indicator r.v. of the event A_i that computer i is (successfully) broken into. Then if X is the number of computers that are broken into, then $X = I_1 + \cdots + I_k$. Let B_{ij} be the event that hacker $j = 1, \ldots, k$ successfully breaks into computer i. Note $P(B_{ij}) = p/n$ since the hacker chooses computer i with probability $1/n$, then (independently) has probably p of succeeding. Then by independence,

$$P(A_i^c) = P(\cap_{j=1}^k B_{ij}^c) = \prod_{j=1}^k P(B_{ij}^c) = \prod_{j=1}^k \left(1 - \frac{p}{n}\right) = \left(1 - \frac{p}{n}\right)^k$$

Hence $P(A_i) = 1 - \left(1 - \frac{p}{n}\right)^k$, and so by linearity

$$\mathbb{E}[X] = \sum_{i=1}^n P(A_i) = \boxed{n\left(1 - \left(1 - \frac{p}{n}\right)^k\right)}$$

3. (BH 4.48) Suppose X is Hypergeometric with parameters w, b, n.

(a) Find $\mathbb{E}\left[\binom{X}{2}\right]$. Hint: It is not necessary to do much algebra with binomial coefficients. Also, note $\binom{0}{2} = \binom{1}{2} = 0$.

Consider the story of the Hypergeometric. $\binom{X}{2}$ is the number of different pairs of white balls among the n balls that are sampled. Number all pairs of white balls in the bag $i = 1, \ldots, \binom{w}{2}$ and let I_i be the indicator r.v. for the event A_i that both of the (white) balls in pair i are selected. Similar to problem 1(b), we have

$$P(A_i) = \frac{\binom{w+b-2}{n-2}}{\binom{w+b}{n}} = \frac{\frac{(w+b-2)!}{(n-2)!(w+b-n)!}}{\frac{(w+b)!}{n!(w+b-n)!}} = \frac{n(n-1)}{(w+b)(w+b-1)}$$

With $\binom{X}{2} = \sum_{i=1}^{\binom{w}{2}} I_i$, we have

$$\mathbb{E}\left[\binom{X}{2}\right] = \binom{w}{2} \frac{n(n-1)}{(w+b-1)(w+b-2)} = \boxed{\frac{wn(w-1)(n-1)}{2(w+b)(w+b-1)}}$$

Alternatively, you could number all pairs in the selected sample of n balls from 1 through $\binom{n}{2}$. For each such pair, you could define an indicator for the event that both balls in the pair are white.

(b) Use part (a) to prove that $\text{Var}(X) = \frac{N-n}{N-1} np(1-p)$, where $N = w + b$ and $p = w/N$.

Recall $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ and that $\mathbb{E}[X] = nw/(w+b) = np$. Note that the identity $\binom{X}{2} = \frac{X(X-1)}{2}$ holds for all $X \in \{0, 1, \dots, \}$, hence

$$\frac{wn(w-1)(n-1)}{(w+b)(w+b-1)} = 2\mathbb{E}\left[\binom{X}{2}\right] = \mathbb{E}[X^2] - \mathbb{E}[X]$$

We conclude

$$\begin{aligned}
\text{Var}(X) &= \frac{wn(w-1)(n-1)}{(w+b)(w+b-1)} + \mathbb{E}[X] - (\mathbb{E}[X])^2 = \frac{wn(w-1)(n-1)}{(w+b)(w+b-1)} + np - n^2p^2 \\
&= np\frac{(w-1)(n-1)}{N-1} + np - n^2p^2 \\
&= \frac{np}{N-1}[(w-1)(n-1) + (N-1) - np(N-1)] \\
&= \frac{np}{N-1}[-Np - n + N + np] \\
&= \frac{N-n}{N-1} np(1-p)
\end{aligned}$$

4. (BH 5.13) A stick of length 1 is broken at a uniformly random point, yielding two pieces. Let X and Y be the lengths of the shorter and longer pieces, respectively, and let $R = X/Y$ be the ratio of the lengths X and Y.

(a) Find the CDF and PDF of R.

Let $U \sim \text{Unif}(0,1)$ be the breakpoint, so $X = \min(U, 1-U)$. Note $\text{supp}(R) = (0,1)$, and for $r \in (0,1)$ we have

$$F_R(r) = P(R \leq r) = P(X \leq r(1-X)) = P\left(X \leq \frac{r}{1+r}\right)$$

Note $r \in (0,1)$ implies $r/(1+r) \in (0, 1/2)$. For any $x \in [0, 1/2]$, we have

$$F_X(x) = P(X \leq x) = 1 - P(X > x) = 1 - P(U > x, 1-U > x) = 1 - P(x < U < 1-x) = 1 - (1-2x) = 2x$$

Hence

$$F_R(r) = \begin{cases} 0 & r < 0 \\ \frac{2r}{1+r} & 0 \leq r \leq 1 \\ 1 & r > 1 \end{cases}$$

Finally, to obtain the PDF we differentiate:

$$f_R(r) = f_R(r) = \frac{(1+r)\cdot 2 - 2r}{(1+r)^2} = \boxed{\frac{2}{(1+r)^2}}, \quad 0 < r < 1$$

(b) Find the expected value of R (if it exists).

By LOTUS

$$\mathbb{E}(R) = 2\int_0^1 \frac{r}{(1+r)^2}dr = 2\int_1^2 \frac{t-1}{t^2}dt = 2\int_1^2 \frac{1}{t}dt - 2\int_1^2 \frac{1}{t^2}dt = \boxed{2\log(2) - 1}$$

(c) Find the expected value of $1/R$ (if it exists).

$$\mathbb{E}(1/R) = 2\int_0^1 \frac{1}{r(1+r)^2}dr = \infty$$

since $1/(r(1+r)^2) \geq 1/(4r) > 0$ for $r \in (0,1)$ but $\int_0^1 1/(4r)dr$ diverges.

5. Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$. Find the CDF and PDF of $Y = X^2$ in terms of Φ, the standard Normal CDF, and φ, the standard Normal PDF.

We have for $y \geq 0$ that

$$\begin{aligned} F_Y(y) &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= P(X \leq \sqrt{y}) - P(X \leq -\sqrt{y}) \\ &= P\left(\frac{X-\mu}{\sigma} \leq \frac{\sqrt{y}-\mu}{\sigma}\right) - P\left(\frac{X-\mu}{\sigma} \leq \frac{-\sqrt{y}-\mu}{\sigma}\right) \\ &= \boxed{\Phi\left(\frac{\sqrt{y}-\mu}{\sigma}\right) - \Phi\left(\frac{-\sqrt{y}-\mu}{\sigma}\right)} \text{ since } \frac{X-\mu}{\sigma} \sim N(0,1) \end{aligned}$$

Taking the derivative and using the chain rule gives

$$\begin{aligned} f_Y(y) &= \varphi\left(\frac{\sqrt{y}-\mu}{\sigma}\right) \cdot \frac{1}{2\sigma\sqrt{y}} + \varphi\left(\frac{-\sqrt{y}-\mu}{\sigma}\right) \cdot \frac{1}{2\sigma\sqrt{y}} \\ &= \boxed{\frac{1}{2\sigma\sqrt{y}}\left(\phi\left(\frac{\sqrt{y}-\mu}{\sigma}\right) + \phi\left(\frac{-\sqrt{y}-\mu}{\sigma}\right)\right)}, \quad y > 0 \end{aligned}$$

6. (BH 5.16) Suppose $U \sim \text{Unif}(0,1)$ and let $X = \log(U/(1-U))$. Then X has the Logistic distribution.

(a) Write down (but do not compute) a definite integral whose value is equal to $\mathbb{E}(X^2)$.

By LOTUS,

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

(b) Find $\mathbb{E}(X)$ without using calculus. Hint: What is the distribution of $1-U$?

First we note $1 - U \sim \text{Unif}(0,1)$; for any $u \in (0,1)$ we have

$$P(1-U \leq u) = P(U \geq 1-u) = 1 - P(U \leq 1-u) = 1-(1-u) = u$$

which matches the Unif(0,1) CDF. Hence $\log(U)$ and $\log(1-U)$ must have the same distribution, therefore the same expectation. But by linearity

$$\mathbb{E}[X] = \mathbb{E}\left[\log\left(\frac{U}{1-U}\right)\right] = \mathbb{E}[\log(U) - \log(1-u)] = \mathbb{E}[\log(U)] - \mathbb{E}[\log(1-U)] = \boxed{0}$$

7. (BH 5.26) Walter and Carl both often need to travel from Location A to Location B. Walter walks, and his travel time is Normal with mean w minutes and standard deviation σ minutes (travel time can't be negative without using a tachyon beam, but assume that w is so much larger than σ that the chance of a negative travel time is negligible). Carl drives his car, and his travel time is Normal with mean c minutes and standard deviation 2σ minutes (the standard deviation is larger for Carl due to variability in traffic conditions). Walter's travel time is independent of Carl's. On a certain day, Walter and Carl leave from Location A to Location B at the same time.

(a) Find the probability that Carl arrives first (in terms of Φ and the parameters). For this you can use the important fact, proven later in the course, that if X_1 and X_2 are independent with $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, then $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Let W be Walter's travel time and C be Carl's. Then $C - W \sim \mathcal{N}(c-w, 5\sigma^2)$ by the given fact. Using standardization,

$$P(C < W) = P(C - W < 0) = P\left(\frac{C - W - (c-w)}{\sigma\sqrt{5}} < \frac{-(c-w)}{\sigma\sqrt{5}}\right) = \boxed{\Phi\left(\frac{w-c}{\sigma\sqrt{5}}\right)}$$

(b) Give a fully simplified criterion (not in terms of Φ), such that Carl has more than a 50% chance of arriving first if and only if the criterion is satisfied.

Since Φ is a strictly increasing function with $\Phi(0) = 1/2$, $P(C < W) > 1/2$ if and only if $w > c$.

(c) Walter and Carl want to make it to a meeting at Location B that is scheduled to begin $w + 10$ minutes after they depart from Location A. Give a fully simplified criterion (not in terms of Φ) such that Carl is more likely than Walter to make it on time for the meeting if and only if the criterion is satisfied.

For Walter, the probability of making it on time is

$$P(W \leq w + 10) = P\left(\frac{W - w}{\sigma} \leq \frac{10}{\sigma}\right) = \Phi\left(\frac{10}{\sigma}\right)$$

For Carl, the probability is

$$P(C \leq w + 10) = P\left(\frac{C - c}{2\sigma} \leq \frac{w - c + 10}{2\sigma}\right) = \Phi\left(\frac{w - c + 10}{2\sigma}\right)$$

The latter is bigger than the former if and only if $(w - c + 10)/(2\sigma) > 10/\sigma$, which simplifies to $w > c + 10$.