STATS 116: Homework 7

Due: Thursday, August 17, 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. (BH 9.20) Let $\boldsymbol{X} \sim \text{Mult}_5(n, \boldsymbol{p})$.

   (a) Find $\mathbb{E}(X_1 \mid X_2)$ and $\text{Var}(X_1 \mid X_2)$.

   By our known results on the conditional distributions of Multinomials, we have

   $$(X_1, X_3, X_4, X_5) \mid X_2 \sim \text{Mult}_4(n - X_2, (p_1/(1-p_2), p_3/(1-p_2), p_4/(1-p_2), p_5(1-p_2)))$$

   Hence $X_1 \mid X_2 \sim \text{Bin}(n - X_2, p_1/(1-p_2))$ and

   $$\mathbb{E}(X_1 \mid X_2) = \boxed{(n - X_2) \cdot \frac{p_1}{1-p_2}}, \quad \text{Var}(X_1 \mid X_2) = \boxed{(n - X_2) \cdot \frac{p_1}{1-p_2} \cdot \left(1 - \frac{p_1}{1-p_2}\right)}$$

   (b) Find $\mathbb{E}(X_1 \mid X_2 + X_3)$.

   By Multinomial lumping, $(X_1, X_2 + X_3, X_4, X_5) \sim \text{Mult}_4(n, (p_1, p_2 + p_3, p_4, p_5))$. By the conditional result, we then have

   $$(X_1, X_4, X_5) \mid X_2 + X_3 \sim \text{Mult}_3\left(n - X_2 - X_3, \left(\frac{p_1}{1-p_2-p_3}, \frac{p_4}{1-p_2-p_3}, \frac{p_5}{1-p_2-p_3}\right)\right)$$

   Hence $X_1 \mid X_2 + X_3 \sim \text{Bin}(n - X_2 - X_3, p_1/(1-p_2-p_3))$ and so

   $$\mathbb{E}[X_1 \mid X_2 + X_3] = \boxed{(n - X_2 - X_3) \cdot \frac{p_1}{1-p_2-p_3}}$$

2. Suppose the numbers of moles that different inhabitants in a very large city A have are i.i.d. with mean μ_A and variance σ_A^2, independent of the numbers of moles that the different inhabitants in

a very large city B have, which are i.i.d. with mean μ_B and variance σ_B^2. Ciara is interested in estimating $\mu_A - \mu_B$, the difference in means between the two cities. To that end, she has the budget to sample a total of $n+2$ individuals from both cities. First, she samples 1 individual from each city, to guarantee she always has at least one individual from each city. Then she flips a fair coin n times. Each time the coin lands heads, she samples from city A; otherwise she samples from city B. At the end of her sampling process, she computes $\bar{X}_A$, the sample mean in the number of moles for the inhabitants from city A, and subtracts $\bar{X}_B$, the sample mean in the number of moles for the inhabitants from city B. Compute $\mathbb{E}[\bar{X}_A - \bar{X}_B]$ and $\mathrm{Var}(\bar{X}_A - \bar{X}_B)$. Simplify.

Let N be the number of individuals sampled from city A. We have $N - 1 \sim \mathrm{Bin}(n, 1/2)$. We can write $\bar{X}_A = N^{-1}\sum_{i=1}^N X_i$ where $X_1, X_2, \dots$ are i.i.d. from the city A distribution. Then

$$\mathbb{E}(\bar{X}_A \mid N = k) = k^{-1}\sum_{i=1}^{k}\mathbb{E}(X_i \mid N = k) = \mathbb{E}(X_i) = \mu_A$$

since the X_i are independent of N by the setup of the problem. Hence $\mathbb{E}(\bar{X}_A \mid N) = \mu_A$ and

$$\mathbb{E}[\bar{X}_A] = \mathbb{E}[\mathbb{E}(\bar{X}_A \mid N)] = \mathbb{E}[\mu_A] = \mu_A$$

Similarly $\mathbb{E}[\bar{X}_B] = \mu_B$, so $\mathbb{E}[\bar{X}_A - \bar{X}_B] = \boxed{\mu_A - \mu_B}$ by linearity. Next by the law of total variance we have

$$\mathrm{Var}(\bar{X}_A - \bar{X}_B) = \mathbb{E}[\mathrm{Var}(\bar{X}_A - \bar{X}_B \mid N)] + \mathrm{Var}(\mathbb{E}[\bar{X}_A - \bar{X}_B \mid N])$$

For $Y_1, Y_2, \dots$ i.i.d. from the city B distribution, we have

$$\begin{aligned}
\mathrm{Var}(\bar{X}_A - \bar{X}_B \mid N = k) &= \mathrm{Var}\left(N^{-1}\sum_{i=1}^{N} X_i - (n+2-N)^{-1}\sum_{j=1}^{n+2-N} Y_j \mid N = k\right)\\
&= \mathrm{Var}\left(k^{-1}\sum_{i=1}^{k} X_i - (n+2-k)^{-1}\sum_{j=1}^{n+2-k} Y_j \mid N = k\right)\\
&= \frac{\sigma_A^2}{k} + \frac{\sigma_B^2}{n+2-k}
\end{aligned}$$

since the X_i and Y_j are all independent. Then $\mathrm{Var}(\bar{X}_A - \bar{X}_B \mid N) = \frac{\sigma_A^2}{N} + \frac{\sigma_B^2}{n+2-N}$ so

$$\mathrm{Var}(\bar{X}_A - \bar{X}_B) = \mathbb{E}\left[\frac{\sigma_A^2}{N} + \frac{\sigma_B^2}{n+2-N}\right] + \mathrm{Var}(\mu_a) = \mathbb{E}\left[\frac{\sigma_A^2}{N} + \frac{\sigma_B^2}{n+2-N}\right]$$

We have

$$\begin{aligned}
\mathbb{E}\left[\frac{\sigma_A^2}{N}\right] &= \sigma_A^2 \sum_{k=1}^{n+1} \frac{1}{k} P(N=k) \\
&= \sigma_A^2 \sum_{k=1}^{n+1} \frac{1}{k} (P(N-1=k-1) \\
&= \sigma_A^2 \sum_{k=1}^{n+1} \frac{1}{k} \cdot \binom{n}{k-1} \left(\frac{1}{2}\right)^n \text{ by the Binomial PMF} \\
&= \frac{\sigma_A^2}{2^n(n+1)} \sum_{k=1}^{n+1} \binom{n+1}{k} \\
&= \frac{\sigma_A^2}{2^n(n+1)} (2^{n+1}-1) \text{ by considering the Bin}(n+1, 1/2) \text{ PMF}
\end{aligned}$$

With $n+2-N$ having the same distribution as N by symmetry, we have

$$\mathbb{E}\left[\frac{\sigma_B^2}{n+2-N}\right] = \frac{\sigma_B^2}{2^n(n+1)}(2^{n+1}-1)$$

so the final answer is

$$\text{Var}(\bar{X}_A - \bar{X}_B) = \boxed{\frac{\sigma_A^2 + \sigma_B^2}{2^n(n+1)}(2^{n+1}-1)}$$

3. (BH 9.48) Paul and n other runners compete in a marathon. Their times are independent continuous r.v.s with CDF F.

(a) For $j = 1, 2, \ldots, n$, let A_j be the event that anonymous runner j completes the race faster than Paul. Explain whether the events A_j are independent, and whether they are conditionally independent given Paul's time to finish the race.

The A_j are not independent. If runners 1 through $n-1$ all complete the race faster than Paul, then this is evidence that Paul was slow, which in turn increases the chance that runner n ran faster than Paul. But the A_j are conditionally independent given Paul's time, since the runners' times are independent.

(b) For the rest of this problem, let N be the number of runners who finish faster than Paul. Find $\mathbb{E}(N)$.

Write $N = I_1 + \cdots + I_n$ where I_j is the indicator of runner j being faster than Paul. By symmetry, $\mathbb{E}(I_j) = 1/2$, so by linearity $\mathbb{E}(N) = \boxed{n/2}$.

(c) Find the conditional distribution of N, given that Paul's time to finish the marathon is t.

Given that Paul's time is t, each of the other runners has a better time than Paul with probability $F(t)$, independently. So the conditional distribution is $\text{Bin}(n, F(t))$.

(d) Find $\text{Var}(N)$.

Let T be Paul's time. By the law of total variance,

$$\begin{aligned}
\text{Var}(N) &= \mathbb{E}(\text{Var}(N \mid T)) + \text{Var}(\mathbb{E}(N \mid T)) \\
&= \mathbb{E}(nF(T)(1 - F(T))) + \text{Var}(nF(T)) \\
&= \frac{n}{2} - \frac{n}{3} + \frac{n^2}{12} = \boxed{\frac{n}{6} + \frac{n^2}{12}}
\end{aligned}$$

since $F(T) \sim \text{Unif}(0,1)$ by universality of the Uniform.

4. (BH 10.13) Let X and Y be i.i.d. positive random variables. Assume that the various expressions below exist. Write the most appropriate of $\leq$, $\geq$, $=$, or ? in the blank for each part (where "?" means that no relation holds in general). Give a brief justification for each answer.

(a) $\mathbb{E}(e^{X+Y})$ _____ $e^{2\mathbb{E}(X)}$

Answer: $\boxed{\geq}$. By Jensen's inequality, we have $\mathbb{E}(e^{X+Y}) \geq e^{\mathbb{E}[X+Y]} = e^{2\mathbb{E}[X]}$.

(b) $\mathbb{E}(X^2 e^X)$ _____ $\sqrt{\mathbb{E}(X^4)\mathbb{E}(e^{2X})}$

Answer: $\boxed{\leq}$ by applying Cauchy-Schwarz to the random variables X^2 and e^X.

(c) $\mathbb{E}(X \mid 3X)$ _____ $\mathbb{E}(X \mid 2X)$

Answer: $\boxed{=}$. Let $Y = 3X$; given $Y = y$ we have $X = y/3$ for sure, hence $e(X \mid Y) = Y/3 = X$. Similarly $\mathbb{E}(X \mid 2X) = X$ as well.

(d) $\mathbb{E}(X^7 Y)$ _____ $\mathbb{E}(X^7 \mathbb{E}(Y \mid X))$

Answer: $\boxed{=}$. By the law of iterated expectations, we have $\mathbb{E}(X^7Y) = \mathbb{E}(\mathbb{E}(X^7Y \mid X)) = \mathbb{E}(X^7\mathbb{E}(Y \mid X))$.

(e) $\mathbb{E}\left(\frac{X}{Y} + \frac{Y}{X}\right)$ _____ 2

Answer: $\boxed{\geq}$. We have $\mathbb{E}(X/Y) = \mathbb{E}(X)\mathbb{E}(1/Y) \geq \frac{\mathbb{E}(X)}{\mathbb{E}(Y)} = 1$ by Jensen's inequality. By exchangeability, X/Y and Y/X have the same distribution, so $\mathbb{E}(Y/X) \geq 1$ as well.

(f) $P(|X - Y| > 2)$ _____ $\frac{\mathrm{Var}(X)}{2}$

Answer: $\boxed{\leq}$. By Markov's inequality, we have

$$P(|X-Y| > 2) = P((X-Y)^2 > 4) \leq \frac{\mathbb{E}[(X-Y)^2]}{4} = \frac{\mathrm{Var}(X-Y)}{4} = \frac{\mathrm{Var}(X) + \mathrm{Var}(Y)}{4} = \frac{\mathrm{Var}(X)}{2}$$

where the second equality uses the fact that $\mathbb{E}[X - Y] = \mathbb{E}[X] - \mathbb{E}[Y] = 0$.

5. Let $X_1, \ldots, X_n$ be i.i.d. random variables with mean μ taking on values in the interval $[0, 1]$, and take $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$ to be the sample mean. Show that $P(\bar{X}_n - \mu \geq t) \leq \exp\left(-2nt^2\right)$ for all $t > 0$, given the following fact, called Hoeffding's Lemma: For any random variable X with $P(0 \leq X \leq 1) = 1$ and $\lambda \in \mathbb{R}$, we have

$$\mathbb{E}[\exp(\lambda(X - \mathbb{E}[X]))] \leq \exp\left(\frac{\lambda^2}{8}\right)$$

Hint: Use a Chernoff bound to upper bound $P(\bar{X}_n - \mu \geq t)$ in terms of $\mathbb{E}[\exp(\lambda(\bar{X}_n - \mu))]$. Use the fact that the X_i are i.i.d. to simplify $\mathbb{E}[\exp(\lambda(\bar{X}_n - \mu))]$ in terms of $\mathbb{E}[\exp(\lambda(X_1 - \mu))]$ and apply Hoeffding's Lemma to the latter. Finally, optimize the upper bound over $\lambda > 0$.

For any $\lambda > 0$, we have

$$\begin{aligned}
P(\bar{X}_n - \mu \geq t) &= P(\exp(\lambda(\bar{X}_n - \mu)) \geq \exp(\lambda t)) \\
&\leq \frac{\mathbb{E}[\exp(\lambda(\bar{X}_n - \mu))]}{\exp(\lambda t)} \text{ by Markov's inequality} \\
&= \frac{\mathbb{E}\left[\exp(\frac{\lambda}{n}\sum_{i=1}^n (X_i - \mu))\right]}{\exp(\lambda t)} \\
&= \exp(-\lambda t)\prod_{i=1}^n \left[\exp\left(\frac{\lambda}{n}(X_i - \mu)\right)\right] \\
&= \exp\left(\frac{\lambda^2}{8n} - \lambda t\right) \text{ by Hoeffding's Lemma with } \lambda/n
\end{aligned}$$

The RHS is minimized by $\lambda = 4nt$, giving $\exp(-2nt^2)$.

6. Suppose Tammyville has a population of n individuals in a city with heights $h_1, \ldots, h_n$. The City Council takes a census of Tammyville, but due to staffing limitations they only look at the heights of $m < n$ individuals, sampled randomly from the city without replacement. Let $X_1, \ldots, X_m$ be these heights. Assume that all heights lie within a finite interval $[a, b]$ (probably a reasonable assumption!).

   (a) Let $\bar{X}_m = m^{-1} \sum_{i=1}^m X_i$ be the sample average. Show that $\mathbb{E}(\bar{X}_m) = \mu_n$ and $\mathrm{Var}(\bar{X}_m) = \frac{n-m}{m(n-1)} \sigma_n^2$ where

$$\mu_n = \frac{1}{n} \sum_{i=1}^n h_i, \quad \sigma_n^2 = \frac{1}{n} \sum_{i=1}^n (h_i - \mu)^2 = \frac{1}{n} \sum_{i=1}^n h_i^2 - \mu^2$$

   Hint: It may be helpful to define indicator r.v.'s for whether each individual in Tammyville was chosen.

   Let $I_i, i = 1, \ldots, n$ be the indicator for the event that individual i is chosen. Then $\bar{X}_m = m^{-1} \sum_{i=1}^n h_i I_i$, so by linearity

$$\mathbb{E}(\bar{X}_m) = m^{-1} \sum_{i=1}^n h_i \mathbb{E}(I_i) = m^{-1} \sum_{i=1}^n h_i \cdot \frac{m}{n} = \boxed{\mu_n}$$

   Next

$$\begin{aligned}
\mathrm{Var}(\bar{X}_m) &= \mathrm{Cov}(\bar{X}_m, \bar{X}_m) \\
&= \mathrm{Cov}\left(m^{-1} \sum_{i=1}^n h_i I_i, m^{-1} \sum_{i=1}^n h_i I_i \right) \\
&= m^{-2} \left(\sum_{i=1}^n h_i^2 \mathrm{Var}(I_i) + \sum_{1 \le i \ne j \le n} h_i h_j \mathrm{Cov}(I_i, I_j) \right)
\end{aligned}$$

   For $i \neq j$ we have

$$\mathrm{Cov}(I_i, I_j) = \mathbb{E}[I_i I_j] - \mathbb{E}[I_i]\mathbb{E}[I_j] = \frac{m(m-1)}{n(n-1)} - \frac{m^2}{n^2} = \frac{m(m-1)n - m^2(n-1)}{n^2(n-1)} = \frac{m(m-n)}{n^2(n-1)}$$

With $\text{Var}(I_i) = \frac{m}{n}\left(1 - \frac{m}{n}\right) = \frac{m(n-m)}{n^2}$, we have

$$\begin{aligned}
\text{Var}(\bar{X}_m) &= m^{-2}\left(\frac{m(n-m)}{n^2}\sum_{i=1}^{n} h_i^2 + \frac{m(m-n)}{n^2(n-1)}\sum_{1\le i\ne j\le n} h_i h_j\right)\\
&= \frac{n-m}{mn^2}\left(\sum_{i=1}^{n} h_i^2 - \frac{1}{n-1}\sum_{1\le i\ne j\le n} h_i h_j\right)\\
&= \frac{n-m}{mn^2}\left(\frac{n^2}{n-1}\cdot\frac{1}{n}\sum_{i=1}^{n} h_i^2 - \frac{n^2}{n-1}\mu^2\right)\\
&= \frac{n-m}{m(n-1)}\sigma_n^2
\end{aligned}$$

where the penultimate equality uses the fact that

$$\mu^2 = \frac{1}{n^2}\left(\sum_{i=1}^{n} h_i^2 + \sum_{i=1}^{n} h_i h_j\right)$$

(b) Show a weak law of large numbers in an asymptotic regime where $m \to \infty$ as $n \to \infty$, for some $p \in (0,1)$. That is, given any sequence of heights $h_1, h_2, \ldots \in [a,b]$, show that for any $\epsilon > 0$, we have $P(|\bar{X}_m - \mu_n| > \epsilon) \to 0$ as $n \to \infty$, if we have $m \to \infty$ whenever $n \to \infty$.

By Chebyshev's inequality, we have

$$P(|\bar{X}_m - \mu_n| > \epsilon) \le \frac{\text{Var}(\bar{X}_m)}{\epsilon^2} = \frac{n-m}{\epsilon^2 m(n-1)}\sigma_n^2 \le \frac{\sigma_n^2}{m\epsilon^2} \le \frac{(b-a)^2}{m\epsilon^2}$$

where the last inequality follows since all the h_i and hence the μ_n are contained in $[a,b]$, so that $\sigma_n^2 \le (b-a)^2$ for all n. The result follows by letting $n \to \infty$ on both sides.

7. (BH 10.25)

(a) Let $Y = e^X$, with $X \sim \text{Expo}(3)$. Find the mean and variance of Y.

By LOTUS we have

$$\begin{aligned}
\mathbb{E}(Y) &= \int_0^\infty \exp(x)\cdot 3\exp(-3x)dx = 3\int_0^\infty \exp(-2x)dx = 3\lim_{L\to\infty}\left[-(1/2)\exp(-2x)\right]\Big|_{x=0}^{x=L} = \boxed{\frac{3}{2}}\\
\mathbb{E}(Y^2) &= \int_0^\infty \exp(2x)\cdot 3\exp(-3x)dx = 3\int_0^\infty \exp(-x)dx = 3
\end{aligned}$$

and so $\text{Var}(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 = \boxed{\frac{3}{4}}$.

(b) For $Y_1, \ldots, Y_n$ i.i.d. with the same distribution as Y from (a), what is the approximate distribution of the sample mean $\bar{Y}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j$ when n is large?

By the CLT, we know $\bar{Y}_n$ is approximately Normal when n is large, with mean parameter $\mathbb{E}(Y_1) = \frac{3}{2}$ and variance parameter $\mathrm{Var}(Y_1)/n = \frac{3}{4n}$. That is, $\bar{Y}_n$ is approximately $\mathcal{N}(3/2, 3/(4n))$.