STATS 116: Homework 6

Due: Friday, August 11, 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 7.48) Athletes compete one at a time at the high jump. Let X_j be how high the jth jumper jumped, with $X_1, X_2, \ldots,$ i.i.d. with a continuous distribution. We say that the jth jumper sets a record if X_j is greater than all of $X_{j-1}, \ldots, X_1$. Find the variance of the number of records among the first n jumpers (as a sum). What happens to the variance as $n \to \infty$?

   Let I_j be the indicator r.v. for the event the j-th jumper sets a record. By exchangeability, $\mathbb{E}(I_j) = P(I_j = 1) = 1/j$ (as all of the first j jumps are equally likely to be the largest of those jumps, and we don't need to worry about ties since the X_j are continuous). Now we show that I_i is uncorrelated with I_j for all i; j with $i < j$ (in fact, all of the indicators are independent, though we don't need to show this). To see this, note all $j!$ rankings of the first j jumpers are equally likely, so by the naive definition

   $$\mathbb{E}[I_iI_j] = P(I_i = I_j = 1) = \frac{\binom{j-1}{i} \cdot (i-1)! \cdot (j-i-1)!}{j!} = \frac{1}{ij} = \mathbb{E}[I_i]\mathbb{E}[I_j]$$

   where the numerator corresponds to putting the best of the first j jumps in position j, picking any i of the remaining jumps to fill positions 1 through i, putting them in any order with the largest one in position i, and then putting $j - i - 1$ jumps in positions $i + 1$ through $j - 1$ in any order. The variance of I_j is

   $$\mathrm{Var}(I_j) = \frac{1}{j} - \frac{1}{j^2}$$

   since $I_j \sim \mathrm{Bern}(1/j)$. Then by bilinearity of covariance and the covariance to variance formula, we have

   $$\mathrm{Var}(I_1 + \cdots + I_n) = \mathrm{Cov}(I_1 + \cdots + I_n, I_1 + \cdots + I_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(I_i, I_j) = \sum_{i=1}^{n} \mathrm{Var}(I_i) = \sum_{i=1}^{n} \frac{1}{i} - \frac{1}{i^2}$$

This goes to infinity as $n \to \infty$ since $\sum_{i=1}^{n} \frac{1}{i}$ diverges but $\sum_{i=1}^{n} \frac{1}{i^2}$ converges.

2. Suppose X_1 and X_2 are i.i.d. standard Normal random variables.

   (a) Find the PDF of X_1^2, and then use this to find the PDF of $Z = X_1^2 + X_2^2$. Show that Z has a Gamma distribution; specify the parameters.

   Note we cannot use the change of variables formula to find the PDF of X_1^2, since $g(x) = x^2$ is not one-to-one. Thus we work with the CDF directly. Let $Y = X_1^2$; then for all $y > 0$ we have

$$F_Y(y) = P(Y \le y) = P(-\sqrt{y} \le X_1 \le \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y})$$

   where Φ is the standard normal CDF. Taking derivatives gives

$$f_Y(y) = \frac{1}{\sqrt{y}}\varphi(\sqrt{y}) = \boxed{\frac{1}{\sqrt{2\pi y}} \exp\left(-\frac{y}{2}\right)}, \quad y > 0$$

   By the convolution formula, we conclude that for all $z > 0$, the PDF of Z is given by

$$\begin{aligned} f_Z(z) = \int_{-\infty}^{\infty} f_Y(y) f_Y(z-y) dy &= \int_0^z \frac{1}{\sqrt{2\pi y}} \exp\left(-\frac{y}{2}\right) \cdot \frac{1}{\sqrt{2\pi(z-y)}} \exp\left(-\frac{z-y}{2}\right) dy \\ &= \frac{1}{2\pi} \exp\left(-\frac{z}{2}\right) \int_0^z \frac{1}{\sqrt{y(z-y)}} dy \\ &= \frac{1}{2\pi z} \exp\left(-\frac{z}{2}\right) \int_0^z \left(\frac{y}{z}\right)^{-1/2} \left(1 - \frac{y}{z}\right)^{-1/2} dy \\ &= \frac{1}{2\pi} \exp\left(-\frac{z}{2}\right) \int_0^1 x^{-1/2}(1-x)^{-1/2} dx \\ &\propto \exp\left(-\frac{z}{2}\right) \end{aligned}$$

   We notice this expression is proportional to the PDF of a Gamma$(1, 1/2)$ distribution evaluated at z. Hence indeed $Z \sim$ Gamma$(1, 1/2)$ and its full PDF (with normalizing constant) is

$$f_Z(z) = \frac{1}{z\Gamma(1)} \cdot \left(\frac{z}{2}\right)^1 \exp\left(-\frac{z}{2}\right) = \frac{1}{2} \exp\left(-\frac{z}{2}\right), \quad z > 0$$

   (b) Find the joint PDF of $Z_1 = X_1 + X_2$ and $Z_2 = X_1 - X_2$.

   We have $(Z_1, Z_2) = g(X_1, X_2)$ where $\boldsymbol{g}(x_1, x_2) = (x_1+x_2, x_1-x_2)$ is invertible with $\boldsymbol{g}^{-1}(z_1, z_2) = \left(\frac{z_1+z_2}{2}, \frac{z_1-z_2}{2}\right) \equiv (x_1, x_2)$ for each $z_1, z_2 \in \mathbb{R}$. We compute

$$\frac{\partial(z_1, z_2)}{\partial(x_1, x_2)} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

and by the multivariate change of variables formula, the joint PDF of (Z_1, Z_2) at $(z_1, z_2) \in \mathbb{R}^2$ is

$$\begin{aligned}
f_{Z_1Z_2}(z_1, z_2) &= f_{X_1X_2}(x_1, x_2)\left|\frac{\partial(z_1, z_2)}{\partial(x_1, x_2)}\right|^{-1} \\
&= \frac{1}{2}\varphi(x_1)\varphi(x_2) \\
&= \frac{1}{4\pi}\exp\left(-\frac{x_1^2}{2}\right)\exp\left(-\frac{x_2^2}{2}\right) \\
&= \frac{1}{4\pi}\exp\left(-\frac{\left(\frac{z_1+z_2}{2}\right)^2 + \left(\frac{z_1-z_2}{2}\right)^2}{2}\right) \\
&= \frac{1}{4\pi}\exp\left(-\frac{z_1^2}{2}\right)\exp\left(-\frac{z_2^2}{2}\right)
\end{aligned}$$

(c) Are Z_1 and Z_2 independent? Are they uncorrelated? Explain.

The joint PDF in (b) factors into a product of two terms, one depending on z_1 only, the other depending on z_2 only. Thus Z_1 and Z_2 are independent, hence uncorrelated.

3. Suppose $\lambda \sim \text{Expo}(\theta)$ for some $\theta > 0$ and that given λ, X has a Poisson distribution with rate parameter λ.

(a) What is the marginal distribution of X?

We have by marginalization that for each $x = 0, 1, \ldots,$ we have

$$\begin{aligned}
P(X = x) &= \int_0^\infty P(X = x \mid \lambda = y) f_\lambda(y) dy \\
&= \int_0^\infty \exp(-y)\frac{y^x}{x!}\theta\exp(-\theta y) dy \\
&= \frac{\theta}{x!}\int_0^\infty \exp(-(1+\theta)y)y^x dy \\
&= \frac{\theta}{x!}(1+\theta)^{-(x+1)}\int_0^\infty \exp(-z)z^x dz \\
&= \theta(1+\theta)^{-(x+1)} = \left(\frac{\theta}{1+\theta}\right)\cdot\left(\frac{1}{1+\theta}\right)^x
\end{aligned}$$

since $\int_0^\infty \exp(-z)z^x dz = \Gamma(x+1) = x!$. By pattern matching, we conclude $X \sim \boxed{\text{Geom}\left(\frac{\theta}{1+\theta}\right)}$.

(b) Compute $\text{Cov}(X, \lambda)$. Hint: It may be helpful to use the law of iterated expectations, which will be covered in lecture on Monday, August 7. However, iterated expectations is really just LOTE written in a compact form, so the problem can be solved using LOTE.

By LOTE, we have

$$\begin{aligned}\mathbb{E}[X\lambda] &= \sum_{x=0}^{\infty} \mathbb{E}[X\lambda \mid X = x] P(X = x) \\ &= \sum_{x=0}^{\infty} x\mathbb{E}[\lambda \mid X = x] \left(\frac{\theta}{1+\theta}\right) \cdot \left(\frac{1}{1+\theta}\right)^x\end{aligned}$$

Now we derive the posterior distribution of λ:

$$\begin{aligned}f_{\lambda|X}(a \mid x) &\propto p_{X|\lambda}(x \mid a) f_\lambda(a) \\ &\propto \exp(-a)a^x \exp(-\theta a) \\ &\propto \exp(-(1+\theta)a)a^x \\ &\propto \frac{1}{a}\exp(-(1+\theta)a)((1+\theta)a)^{x+1}\end{aligned}$$

which pattern matches the Gamma$(x+1, 1+\theta)$ distribution. Hence $\mathbb{E}[\lambda \mid X = x] = \frac{x+1}{1+\theta}$ and

$$\begin{aligned}\mathbb{E}[X\lambda] &= (1+\theta)^{-1} \sum_{x=0}^{\infty} x(x+1) \left(\frac{\theta}{1+\theta}\right) \cdot \left(\frac{1}{1+\theta}\right)^x \\ &= \frac{2\left(1 - \frac{\theta}{1+\theta}\right)}{(1+\theta)\left(\frac{\theta}{1+\theta}\right)^2} \\ &= \frac{2}{\theta^2}\end{aligned}$$

where we recognize the sum as $\mathbb{E}[X(X+1)] = \mathbb{E}[X^2] + \mathbb{E}[X] = \frac{2-p}{p^2}$ for $p = \theta/(1+\theta)$, based on results we've derived for the Geometric distribution.

Alternatively, by iterated expectations we immediately have

$$\mathbb{E}[X\lambda] = \mathbb{E}[\lambda \mathbb{E}[X \mid \lambda]] = \mathbb{E}[\lambda^2] = \text{Var}(\lambda) + (\mathbb{E}[\lambda])^2 = \frac{2}{\theta^2}$$

since $X \mid \lambda \sim \text{Pois}(\lambda)$ and $\lambda \sim \text{Expo}(\theta)$. Having computed $\mathbb{E}[X\lambda]$, now we can use part (a) to conclude

$$\text{Cov}(X, \lambda) = \mathbb{E}[X\lambda] - \mathbb{E}[X]\mathbb{E}[\lambda] = \frac{2}{\theta^2} - \frac{1 - \frac{\theta}{1+\theta}}{\frac{\theta}{1+\theta}} \cdot \frac{1}{\theta} = \boxed{\frac{1}{\theta^2}}$$

4. (BH 7.71) In humans (and many other organisms), genes come in pairs. Consider a gene of interest, which comes in two types (alleles): type a and type A. The genotype of a person for that gene is the types of the two genes in the pair: AA, Aa, or aa (aA is equivalent to Aa). According to the Hardy-Weinberg law, for a population in equilibrium the frequencies of AA, Aa, aa will be p^2,

$2p(1-p)$, $(1-p)^2$ respectively, for some p with $0 < p < 1$. Suppose that the Hardy-Weinberg law holds, and that n people are drawn randomly from the population, independently. Let X_1;X_2;X_3 be the number of people in the sample with genotypes AA; Aa; aa; respectively.

(a) What is the joint PMF of X_1, X_2, X_3?

By the story of the Multinomial, $(X_1, X_2, X_3) \sim \text{Mult}_3(n, (p^2, 2pq, q^2))$ where $q = 1 - p$. The PMF is then

$$P(X_1 = n_1, X_2 = n_2, X_3 = n_3) = \frac{n!}{n_1!n_2!n_3!}p^{2n_1}(2pq)^{n_2}q^{2n_3}$$

for $n_1 + n_2 + n_3 = n$.

(b) What is the distribution of the number of people in the sample who have an A?

By the story of the Binomial (defining "success" as having an A and "failure" as not having an A), the distribution is $\text{Bin}(n, p^2 + 2pq)$.

(c) What is the distribution of how many of the $2n$ genes among the people are A's?

Let Y_j be how many A's the j-th person in the sample has. Then Y_j is 2 with probability p^2, 1 with probability $2pq$, and 0 with probability q^2, so $Y_j \sim \text{Bin}(2, p)$. The Y_j are also independent. Therefore $Y_1 + \cdots + Y_n \sim \text{Bin}(2n, p)$.

(d) Now suppose that p is unknown, and must be estimated using the observed data X_1, X_2, X_3. The maximum likelihood estimator (MLE) of p is the value of p for which the observed data are as likely as possible. Find the MLE of p. Hint: recall $\log(x)$ is an increasing function.

Let x_1, x_2, x_3 be the observed values of X_1, X_2, X_3. The MLE of p is the value of p that maximizes the function $L(p) = p^{2x_1}(pq)^{x_2}q^{2x_3} = p^{2x_1+x_2}(1-p)^{x_2+2x_3}$ (we can omit factors which are constant with respect to p, since such constants do not affect where the maximum is). Equivalently, we can maximize the log:

$$\log L(p) = (2x_1 + x_2)\log p + (x_2 + 2x_3)\log(1-p) :$$

Setting the derivative of $\log L(p)$ equal to 0, we have

$$\frac{2x_1 + x_2}{p} - \frac{x_2 + 2x_3}{1-p} = 0$$

which rearranges to

$$p = \frac{2x_1 + x_2}{2(x_1 + x_2 + x_3)} = \frac{2x_1 + x_2}{2n}$$

This value of p does maximize $\log L(p)$ since the derivative of $\log L(p)$ is positive everywhere to the left of it and is negative everywhere to the right of it. Thus, the MLE of p, which we denote by $\hat{p}$, is given by $\hat{p} = (2X_1 + X_2)/(2n)$. Note that this has an intuitive interpretation: it is the fraction of A's among the $2n$ genes.

(e) Now suppose that p is unknown, and that our observations can't distinguish between AA and Aa. So for each person in the sample, we just know whether or not that person is an aa (in genetics terms, AA and Aa have the same phenotype, and we only get to observe the phenotypes, not the genotypes). Find the MLE of p.

Let $Y \sim \text{Bin}(n, q^2)$ be the number of aa people, and let y be the observed value of Y. We need to maximize the function $L_2(q) = q^{2y}(1 - q^2)^{n-y}$ (we will maximize over q and then find the corresponding value of p). Then

$$\log L_2(q) = 2y \log(q) + (n - y) \log(1 - q^2)$$

so

$$\frac{d \log L_2(q)}{dq} = \frac{2y}{q} - \frac{2q(n - y)}{1 - q^2},$$

which simplifies to $y = q^2 n$. By looking at the sign of the derivative, we see that $\log L_2(q)$ is maximized at $q = \sqrt{y/n}$. Thus, the MLE of q is $\sqrt{Y/n}$, which shows that the MLE of p is $1 - \sqrt{Y/n}$.

5. (BH 8.12) Let T be the ratio X/Y of two i.i.d. $\mathcal{N}(0, 1)$ r.v..s X, Y. This is the *Cauchy* distribution and it has PDF

$$f_T(t) = \frac{1}{\pi(1 + t^2)}$$

(a) Show that $1/T$ has the same distribution as T using calculus, after first finding the CDF of $1/T$ in terms of the CDF F_T of T.

By rewriting the event $1/T \leq v$ for $v \in \mathbb{R}$, we see

$$F_V(v) = \begin{cases} F_T(0) + 1 - F_T(v^{-1}) & v > 0 \\ F_T(0) & v = 0 \\ F_T(0) - F_T(v^{-1}) & v < 0 \end{cases}$$

Then by the chain rule we have

$$f_V(v) = \frac{1}{v^2} f\left(\frac{1}{v}\right) = \frac{1}{v^2} \cdot \frac{1}{\pi(1+(1/v)^2)} = \frac{1}{\pi(1+v^2)} = f_T(v)$$

for all $v \neq 0$. Thus V has the same distribution as T (recall the CDF only has to be differentiable at all but countably many points, so we can safely ignore the single point $v = 0$ where this derivative does not evidently apply, although it will).

(b) Argue that $1/T$ has the same distribution as T without using calculus.

(X, Y) are i.i.d., hence exchangeable, so letting $g(x, y) = x/y$, we know $g(X, Y) = X/Y = T$ has the same distribution as $g(Y, X) = Y/X = 1/T$.

6. (BH 8.18) Let X and Y be i.i.d. $\mathcal{N}(0, 1)$ r.v.s, and (R, θ) be the polar coordinates for the point (X, Y) so $X = R\cos(\theta)$ and $Y = R\sin(\theta)$ with $R \geq 0$ and $\theta \in [0, 2\pi)$. Find the joint PDF of R^2 and θ. Also find the marginal distributions of R^2 and θ, giving their names (and parameters) if they are distributions we have studied before.

We have $X = R\cos(\theta)$, $Y = R\sin(\theta)$. Let $W = R^2$, $T = \theta$ and mirror the relationships between the capital letters via $w = r^2 = x^2 + y^2$, $x = \sqrt{w}\cos(t)$, $y = \sqrt{w}\sin(t)$. By the multivariate change of variables formula

$$f_{WT}(w, t) = f_{XY}(x, y)\left|\frac{\partial(x, y)}{\partial(w, t)}\right| = \frac{e^{-r^2/2}}{2\pi}\left|\frac{\partial(x, y)}{\partial(w, t)}\right|$$

The Jacobian matrix is

$$\frac{\partial(x, y)}{\partial(w, t)} = \begin{bmatrix} \frac{1}{2\sqrt{w}}\cos(t) & -\sqrt{w}\sin(t) \\ \frac{1}{2\sqrt{w}}\sin(t) & \sqrt{w}\cos(t) \end{bmatrix}$$

which has absolute determinant $\frac{1}{2}\cos^2(t) + \frac{1}{2}\sin^2(t) = \frac{1}{2}$. So the joint PDF of R^2 and θ is

$$f_{R^2\theta}(w, t) = \frac{1}{4\pi}e^{-w/2} = \frac{1}{2\pi} \cdot \frac{1}{2}e^{-w/2}$$

for $w > 0$ and $0 \leq t < 2\pi$ (and 0 otherwise). Thus, R^2 and θ are independent, with $R^2 \sim \text{Expo}(1/2)$ and $\theta \sim \text{Unif}(0, 2\pi)$. Note that this problem is Box-Muller in reverse.

7. Emily and Devon each independently generate a Unif$(-1, 1)$ number.

(a) Given that Emily's number is higher than Devon's, what is the expected value of Emily's number? Is this the same as the (unconditional) expected value of the maximum of Emily and

Devon's numbers?

Let X and Y be Emily and Devon's numbers, respectively. The joint PDF of (X, Y) is a constant $1/4$ on $[-1, 1]^2$, so

$$\mathbb{E}[X \mid X > Y] = \frac{\mathbb{E}[XI_{X>Y}]}{P(X > Y)} = 2\mathbb{E}[XI_{X>Y}]$$

By multivariate LOTUS, we have

$$\mathbb{E}[XI_{X>Y}] = \int_{-1}^{1}\int_{y}^{1} \frac{1}{4} x dx dy = \int_{-1}^{1} \frac{1-y^2}{8} dy = \frac{1}{6}$$

Hence $\mathbb{E}[X \mid X > Y] = \boxed{1/3}$. By LOTE we have

$$\begin{aligned}\mathbb{E}[\max(X, Y)] &= \mathbb{E}[\max(X, Y) \mid X > Y]P(X > Y) + \mathbb{E}[\max(X, Y) \mid X < Y]P(X < Y) \\ &= \frac{1}{2}\left(\mathbb{E}[X \mid X > Y] + \mathbb{E}[Y \mid Y > X]\right)\end{aligned}$$

By exchangeability/symmetry we have $\mathbb{E}[X \mid X > Y] = \mathbb{E}[Y \mid Y > X]$ so the answer is yes. Alternatively you could derive the PDF of $\max(X, Y)$.

(b) What is the covariance between A, the maximum of Emily and Devon's numbers, and B, the minimum of Emily and Devon's numbers? Are A and B independent? Explain. Hint: Think along the lines of the identity $\max(a, b) + \min(a, b) = a + b$ for any numbers a, b.

We have

$$\text{Cov}(A, B) = \mathbb{E}[AB] - \mathbb{E}[A]\mathbb{E}[B] = \mathbb{E}[XY] - \mathbb{E}[A]\mathbb{E}[B]$$

since $AB = \max(X, Y)\min(X, Y) = XY$. We know $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = 0$ since X and Y are independent, hence uncorrelated. Furthermore $\mathbb{E}[A] = 1/3$ by part (a), and since $A+B = X+Y$ we have $\mathbb{E}[B] = \mathbb{E}[X]+\mathbb{E}[Y]-\mathbb{E}[A] = -1/3$. Hence $\text{Cov}(A, B) = 1/9$. This is not 0, so A and B cannot be independent. The lack of independence can also be seen by noting that the support of B conditional on A varies based on the value of A.