Fall 2022, MATH 3215-J, Final (60 pts)

Name:

GT ID:

- Regular exam time is 170 min.
- Open-book/notes. Calculators are allowed. No communication in any form.
- All the solutions are supposed to be short and can fit in the given space. Please read the problem statements carefully.
- Consider skipping a part if you get stuck somewhere.
- If there is a mistake in a problem statement making it unsolvable, skip it and you will be awarded full points for that problem.

1. In a 5-game series played between two teams, the first team to win a total of 3 games is the winner. Suppose that the outcomes of games are independent, and that team A wins each game with probability 0.8 and team B wins each game with probability 0.2.

(a) (2 pts) Conditional on the event that team A leads by 2 to 0, what is the probability that team A wins the series?

(b) (2 pts) Conditional on the event that team A leads by 1 to 0, what is the probability that team B wins the series?

The final answers (numerical values) should be computed explicitly and exactly.

Let the probability that team A wins a game be $p = 0.8$.

(a) If team B wins the series, it must win 3 consecutive games, which happens with probability $(1-p)^3$. Therefore, the answer is $1-(1-p)^3 = 0.992$.

(b) Let E be the event that team A leads by 1 to 0, and let F be the event that team B wins the series. For the event $E \cap F$, the series can have winners ABBB, AABBB, ABABB, or ABBAB. Therefore,

$$\mathcal{P}(F \mid E) = \frac{\mathcal{P}(E \cap F)}{\mathcal{P}(E)} = \frac{p(1-p)^3 + 3p^2(1-p)^3}{p} = (1-p)^3 + 3p(1-p)^3 = 0.0272.$$

2. (2 pts) Suppose that the total weight of products produced daily at a factory is X^3 lbs where $X \sim \mathcal{N}(21, 0.25)$. What is the probability that the total weight exceeds 8,000 lbs in a day? Use the CDF Φ of $\mathcal{N}(0,1)$ to express the answer.

The probability is

$$\mathbb{P}\{X^3 > 8000\} = \mathbb{P}\{X > 20\} = \mathbb{P}\left\{\frac{X-21}{0.5} > \frac{20-21}{0.5}\right\} = 1 - \Phi(-2) = \Phi(2).$$

3. Suppose that the number of accidents on Highway A in a year has the Poisson distribution Poi(100), and that number on Highway B has distribution Poi(300), independently from the number on Highway A. Use the CDF Φ of the standard normal $\mathcal{N}(0,1)$ to approximate the probabilities of the following events (no need to worry about correction for continuity or strict vs. non-strict inequalities):

(a) (2 pts) The number of accidents on Highway A is at least 110.

(b) (2 pts) The total number of accidents on Highway A and Highway B is at most 410.

(c) (2 pts) There are more accidents on Highway B than on Highway A.

(a) Let X be the number of accidents on Highway A. Then $X \sim \mathrm{Poi}(100)$, so $\mathbb{E}[X] = 100$ and $\mathrm{Var}(X) = 100$. It follows that

$$\mathbb{P}\{X \geq 110\} = \mathbb{P}\left\{\frac{X-100}{10} \geq \frac{110-100}{10}\right\} \approx 1 - \Phi(1) = \Phi(-1).$$

(b) Let Y be the number of accidents on Highway B. Then $Y \sim \mathrm{Poi}(300)$, so $\mathbb{E}[Y] = 300$ and $\mathrm{Var}(Y) = 300$. Hence, $X + Y$ is approximately $\mathcal{N}(400, 400)$. It follows that

$$\mathbb{P}\{X + Y \leq 410\} = \mathbb{P}\left\{\frac{X+Y-400}{20} \leq \frac{410-400}{20}\right\} \approx \Phi(0.5).$$

(c) We have that $Y - X$ is approximately $\mathcal{N}(200, 400)$, so

$$\mathbb{P}\{Y > X\} = \mathbb{P}\left\{\frac{Y-X-200}{20} > \frac{-200}{20}\right\} \approx 1 - \Phi(-10) = \Phi(10) \approx 1.$$

4. Let $\mathsf{Exp}(3)$ denote the exponential distribution with PDF $f(x) = \lambda e^{-3x}$ for $x \geq 0$ and $f(x) = 0$ for $x < 0$, where λ is a fixed constant. Let X and Y be i.i.d. $\mathsf{Exp}(3)$ random variables.

(a) (2 pts) What is λ?

(b) (1 pt) Denote the joint PDF of (X, Y) by $\tilde{f}(x, y)$. What is $\tilde{f}(x, y)$? (The domain of the function needs to be clearly indicated.)

(c) (2 pts) Denote the CDF of $X + Y$ by $F(t)$ for $t \in \mathbb{R}$. What is $F(t)$ for $t \geq 0$?

(d) (2 pts) Denote the PDF of $X + Y$ by $f(t)$ for $t \in \mathbb{R}$. What is $f(t)$ for $t \geq 0$?

(a) We have $1 = \int_0^\infty \lambda e^{-3x}\, dx = \lambda/3$, so $\lambda = 3$.

(b) We have

$$\tilde{f}(x, y) = 9e^{-3(x+y)}$$

for $x, y \geq 0$, and $\tilde{f}(x, y) = 0$ otherwise.

(c) For $t \geq 0$, we have

$$\begin{aligned} F(t) = \mathbb{P}\{X + Y \leq t\} &= \int_0^t \int_0^{t-y} 9e^{-3(x+y)}\, dx\, dy \\ &= \int_0^t -3e^{-3(x+y)}\Big|_{x=0}^{t-y}\, dy \\ &= \int_0^t 3(e^{-3y} - e^{-3t})\, dy \\ &= -e^{-3y}\Big|_{y=0}^{t} - 3te^{-3t} \\ &= 1 - e^{-3t} - 3te^{-3t}. \end{aligned}$$

(d) For $t \geq 0$, we have

$$f(t) = F'(t) = 3e^{-3t} - 3e^{-3t} + 9te^{-3t} = 9te^{-3t}.$$

5. A plane is missing and is equally likely to have gone down in one of three possible regions. For $i = 1, 2, 3$, let p_i be the probability that the plane can be found upon a search of the ith region when the plane is in fact in that region.

(a) (1 pt) What is the probability that a search in the 2nd region is successful?

(b) (2 pts) Conditional on the event that a search of the 2nd region is unsuccessful, what is the probability that the plane is in the 1st region?

(c) (2 pts) Conditional on the event that a search of the 2nd region is unsuccessful, what is the probability that the plane is in the 2nd region?

(Express the answers in terms of p_1, p_2, and p_3.)

Let X be the region the plane is in. Let Y be the indicator that the search in the 2nd region is successful (i.e., $Y = 1$, successful; $Y = 0$, unsuccessful).

(a) We have

$$\mathbb{P}\{Y = 1\} = \sum_{i=1}^{3} \mathbb{P}\{Y = 1 \,|\, X = i\} \cdot \mathbb{P}\{X = i\} = \frac{p_2}{3}.$$

(b) We have

$$\mathbb{P}\{X = 1 \,|\, Y = 0\} = \frac{\mathbb{P}\{Y = 0 \,|\, X = 1\} \cdot \mathbb{P}\{X = 1\}}{\mathbb{P}\{Y = 0\}} = \frac{1/3}{1 - p_2/3} = \frac{1}{3 - p_2}.$$

(c) We have

$$\mathbb{P}\{X = 2 \,|\, Y = 0\} = \frac{\mathbb{P}\{Y = 0 \,|\, X = 2\} \cdot \mathbb{P}\{X = 2\}}{\mathbb{P}\{Y = 0\}} = \frac{(1 - p_2)/3}{1 - p_2/3} = \frac{1 - p_2}{3 - p_2}.$$

6. Let X have PDF $f_X(x) = x/4$ for $x \in [1, 3]$ and $f_X(x) = 0$ otherwise. Conditional on $X = x$, let Y have the uniform distribution over $[0, x]$.

(a) (1 pt) What is the conditional PDF of Y given $X = x \in [1, 3]$?

(b) (2 pts) What is the joint PDF $f(x, y)$ of (X, Y)?

(c) (2 pts) What is the marginal PDF $f_Y(y)$ of Y?

The domains of the above functions (where a function is nonzero) should be clearly specified.

(a) We have

$$f_{Y|X}(y|x) = 1/x, \qquad y \in [0, x].$$

(b) We have

$$f(x, y) = f_X(x) \cdot f_{Y|X}(y|x) = 1/4, \qquad x \in [1, 3],\ y \in [0, x].$$

(c) For $y \in [0, 1]$,

$$f_Y(y) = \int_1^3 \frac{1}{4}\, dx = \frac{1}{2},$$

and for $y \in (1, 3]$,

$$f_Y(y) = \int_y^3 \frac{1}{4}\, dx = \frac{3 - y}{4}.$$

7. Suppose that we are given i.i.d. $X_1, \dots, X_n$ from the Pareto distribution with PDF $f(x) = \lambda\theta^{\lambda}x^{-(\lambda+1)}$ for $x \geq \theta$ and $f(x) = 0$ otherwise, where $\theta > 0$ and $\lambda > 0$ are both unknown parameters.

(a) (2 pts) What is the joint PDF of $(X_1, \dots, X_n)$?
(Write it as a function of $x_1, \dots, x_n$ with the domain clearly specified.)

(b) (1 pt) What is the MLE $\hat{\theta}$ of θ?

(c) (1 pt) Now fix $\theta = \hat{\theta}$ (no need to plug in its formula from the previous part). What is the log-likelihood for λ?

(d) (2 pts) What is the MLE $\hat{\lambda}$ of λ?

(a) The joint PDF is $\lambda^n\theta^{n\lambda}(x_1 \cdots x_n)^{-(\lambda+1)}$ if $x_i \leq \theta$ for all $i = 1, \dots, n$ and is zero otherwise.

(b) The likelihood is the same as the joint PDF. To maximize this function over θ, we should take $\theta = \min_{1\leq i\leq n} x_i$. Therefore, the MLE of θ is

$$\hat{\theta} = \min_{1\leq i\leq n} X_i.$$

(c) The log-likelihood is

$$n \log \lambda + n\lambda \log \hat{\theta} - (\lambda + 1) \log(x_1 \cdots x_n).$$

(d) Differentiating the log-likelihood with respect to λ and setting the result to zero, we obtain

$$\frac{n}{\lambda} + n \log \hat{\theta} - \log(x_1 \cdots x_n) = 0.$$

Therefore, the MLE of λ is

$$\hat{\lambda} = \frac{n}{\log(X_1 \cdots X_n) - n \log(\hat{\theta})}.$$

8. The mean breaking strength of a type of fiber is believed to be at least 200 psi. Suppose that the breaking strength is normally distributed with standard deviation equal to 2 psi. A sample of 16 pieces of fiber yielded average breaking strength 190 psi.

(a) (1 pt) If we would like to show with high confidence that this type of fiber is not as strong as it is believed to be, what is the null hypothesis H_0 and what is alternative hypothesis H_1? (Write the hypotheses in mathematical terms like $H_0 : \mu \ldots$.)

(b) (1 pt) What is the test statistic that has the standard normal distribution? Compute its numerical value.

(c) (1 pt) What is the test at significance level 0.05? Express the test as rejecting H_0 if some condition holds (in terms of the above statistic and an explicit numerical threshold). (Recall that $\Phi(1.96) \approx 0.975$ and $\Phi(1.645) \approx 0.95$ where Φ is the CDF of $\mathcal{N}(0, 1)$).

(d) (1 pt) Do we reject H_0 or not?

(e) (1 pt) What is the p-value, expressed using Φ?

(f) (1 pt) What is the p-value approximately? Choose your answer from: (i) 0.1; (ii) 0.05; (iii) 0.01; (iv) 0.005; (v) 0.

(a) $H_0 : \mu \geq 200$; $H_1 : \mu < 200$

(b) $\frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0) = \frac{\sqrt{16}}{2}(190 - 200) = -20$

(c) Reject H_0 if $\frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0) < -z_\alpha = \Phi^{-1}(0.95) \approx -1.645$

(d) Reject H_0. Not acceptable.

(e) $\mathbb{P}\{Z < -20\} = \Phi(-20)$

(f) 0

9. Consider two independent samples of sizes n_1 and n_2 from $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$ respectively. Let S_1^2 and S_2^2 be the respective sample variances. Let $\mathcal{F}$ denote the F-distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom. Recall that

$$R := \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim \mathcal{F}.$$

Note that R takes values in $(0, \infty)$. Let G denote the CDF of $\mathcal{F}$. For $\alpha \in (0,1)$, let $g_\alpha := G^{-1}(1-\alpha)$ be the quantile of order $1 - \alpha$.

(a) (1 pt) Give a test statistic (in terms of S_1^2 and S_2^2 only) that has distribution $\mathcal{F}$ if $\sigma_1^2 = 2\sigma_2^2$.

(b) (2 pts) Suppose that we test $H_0 : \sigma_1^2 \leq 2\sigma_2^2$ (or $\sigma_1^2 = 2\sigma_2^2$) against $H_1 : \sigma_1^2 > 2\sigma_2^2$. What is the one-sided test at significance level $\alpha \in (0,1)$? Express the test as rejecting H_0 if some condition holds (in terms of the above statistic and a quantile).

(c) (2 pts) What is the p-value associated with the above test (i.e., associated with the statistic from part (a))? Express it in terms of S_1^2, S_2^2, and the CDF G.

(d) (1 pt) Suppose that we test $H_0 : \sigma_1^2 \geq 2\sigma_2^2$ (or $\sigma_1^2 = 2\sigma_2^2$) against $H_1 : \sigma_1^2 < 2\sigma_2^2$. What is the one-sided test at significance level $\alpha \in (0,1)$? Express the test as rejecting H_0 if some condition holds (in terms of S_1^2, S_2^2, and a quantile).

(a) If $\sigma_1^2 = 2\sigma_2^2$, then

$$\frac{S_1^2}{2S_2^2} = \frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2} = R \sim \mathcal{F}.$$

(b) We have $\mathbb{P}\left\{\frac{S_1^2}{2S_2^2} > g_\alpha\right\} = \alpha$, so we reject H_0 if $S_1^2 > 2g_\alpha S_2^2$.

(c) The p-value is

$$\mathbb{P}\left\{L > \frac{S_1^2}{2S_2^2}\right\} = 1 - G\Big(\frac{S_1^2}{2S_2^2}\Big),$$

where the probability is with respect to $L \sim \mathcal{F}$.

(d) We reject H_0 if $S_1^2 < 2g_{1-\alpha}S_2^2$.

10. Suppose that we observe i.i.d. $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ where σ^2 is unknown. Consider testing $H_0 : \mu \le \mu_0$ against $H_1 : \mu > \mu_0$. Let S^2 denote the sample variance. For $\alpha \in (0,1)$, let $t_\alpha := F^{-1}(1-\alpha)$, where F denotes the CDF of the t-distribution with $n-1$ degrees of freedom.

(a) (2 pts) Find a statistic B (a quantity expressed in terms of the quantile t_α and the data) such that the following holds: The test that rejects H_0 if $B > \mu_0$ achieves significance level α exactly.

(b) (2 pts) Find a p-value P (a quantity expressed in terms of F, μ_0, and the data) such that the following holds: The test that rejects H_0 if $P < \alpha$ achieves significance level α exactly.

(c) (1 pt) Are the above two tests equivalent?

(d) (1 pt) If $\mu = \mu_0$, what is the CDF $\tilde{F}(t)$ of the P? (Recall that the p-value is expressed in terms of the random data, so it is a random variable and therefore has its own CDF.)

(a) The t-test at significance level α rejects H_0 if $\frac{\sqrt{n}}{S}(\bar{X} - \mu_0) > t_\alpha$. The rejection rule is equivalent to $\bar{X} - \mu_0 > t_\alpha \frac{S}{\sqrt{n}}$, i.e., $\mu_0 < \bar{X} - t_\alpha \frac{S}{\sqrt{n}}$. Therefore, it suffices to set $B = \bar{X} - t_\alpha \frac{S}{\sqrt{n}}$.

(b) The p-value is $1 - F\big(\frac{\sqrt{n}}{S}(\bar{X} - \mu_0)\big)$.

(c) They are equivalent because, by (a) and (b),

$$\mu_0 < B \quad \Longleftrightarrow \quad \frac{\sqrt{n}}{S}(\bar{X} - \mu_0) > t_\alpha \quad \Longleftrightarrow \quad F\Big(\frac{\sqrt{n}}{S}(\bar{X} - \mu_0)\Big) > F(t_\alpha) = 1 - \alpha,$$

where the last condition is equivalent to p-value smaller than α.

(d) The test at significance level α rejects H_0 if $P < \alpha$. This literally means that if $\mu = \mu_0$, then $\mathbb{P}\{P < \alpha\} = \alpha$. Therefore, $\tilde{F}(t) = t$ for $t \in (0,1)$, $\tilde{F}(t) = 0$ for $t \le 0$, and $\tilde{F}(t) = 1$ for $t \ge 1$.

11. Consider the linear regression model

$$Y_i = \beta x_i + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where β is the parameter to be estimated, x_i are fixed covariates, and the noise terms ε_i are i.i.d. $\mathcal{N}(0, 1)$ random variables.

(a) (1 pt) What is the distribution of Y_i?
(The name and parameters of the distribution all need to be specified.)

(b) (1 pt) What is the joint PDF of $(Y_1, \ldots, Y_n)$?

(c) (1 pt) What is the log-likelihood at β?

(d) (2 pts) What is the MLE $\hat{\beta}$ of β?

(e) (2 pts) What is the mean of $\hat{\beta}$?

(a) We have $Y_i \sim \mathcal{N}(\beta x_i, 1)$.

(b) The joint PDF is

$$f(y_1, \ldots, y_n) = \frac{1}{(2\pi\sigma)^{n/2}} \exp\Bigg(-\frac{1}{2} \sum_{i=1}^{n} (y_i - \beta x_i)^2 \Bigg).$$

(c) The log-likelihood is

$$-\frac{n}{2} \log(2\pi\sigma) - \sum_{i=1}^{n} (y_i - \beta x_i)^2.$$

(d) To find the MLE, we need to minimize

$$\sum_{i=1}^{n} (Y_i - \beta x_i)^2,$$

take its derivative with respect to β and set the result to zero:

$$2 \sum_{i=1}^{n} x_i (Y_i - \beta x_i) = 0.$$

This yields

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}.$$

(e) The mean of $\hat{\beta}$ is

$$\frac{\sum_{i=1}^{n} x_i \, \mathbb{E}[Y_i]}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} x_i \beta x_i}{\sum_{i=1}^{n} x_i^2} = \beta.$$