PhD Preliminary Exam in Probability & Statistics, Fall 2021

Answer all 7 questions. Each part of each question is worth 5 points. Give numerical answers whenever possible.

The exam duration is 4 hours. The exam is closed notes. The students are allowed to use a graphing calculator.

Normal, t and χ^2 tables are attached to the exam.

Problem	1		2				3			4	
	a	b	a	b	c	d	a	b	c	a	b
Earned											
Possible	5	5	5	5	5	5	5	5	5	5	5

Problem	5				6	7		Total
	a	b	c	d	a	a	b	
Earned								
Possible	5	5	5	5	5	5	5	90

1. Let $X_1, X_2, \ldots, X_n$ be independent random variables with PDF

$$f(x \mid \theta) = \begin{cases} \frac{1}{\theta} x^{\frac{1-\theta}{\theta}} & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

with the parameter $\theta > 0$.

(a) Find the maximum likelihood estimator (MLE) for θ, call it $\hat{\theta}$. Calculate the estimate numerically for $n = 4$ and $X_1 = 0.10$, $X_2 = 0.22$,$X_3 = 0.54$ and $X_4 = 0.36$.

(b) Find the method of moments estimator for θ, call it $\tilde{\theta}$. Calculate the estimate numerically for $n = 4$ and $X_1 = 0.10$, $X_2 = 0.22$,$X_3 = 0.54$ and $X_4 = 0.36$.

2. Consider the following joint density for random variables X and Y:

$$f(x,y) = \begin{cases} 6xy & \text{for } 0 < x < 1, 0 < y < \sqrt{x} \\ 0 & \text{elsewhere} \end{cases}$$

(a) Find marginal densities $f_X(x)$ and $f_Y(y)$. Are X, Y independent?

(b) Find the conditional density of X given $Y = y$.

(c) Find $\mathbb{E}(X \mid Y = y)$

(d) Find $Var(X \mid Y = y)$

3. Let $X_1, X_2, \ldots, X_n$ be independent random variables following Poisson distribution with the unknown mean θ. The prior distribution for θ is $Gamma(\alpha, \beta)$ with the PDF

$$\frac{\theta^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} e^{-\theta/\beta}, \qquad \theta > 0, \alpha > 0, \beta > 0$$

(a) Show that the posterior distribution of θ is again a gamma distribution with parameters $\alpha^* = \alpha + \sum X_i$ and $\beta^* = \frac{\beta}{1+n\beta}$

(b) What is the Bayes estimator (under the square loss) for θ?

(c) Is the Bayes estimator for θ consistent?

4. A baseball player will go to the plate six times during a game. 20% of the time that the player goes to the plate, he gets a walk, and thus cannot get a hit. The other 80% of the time, the player gets an official "at bat". For each "at bat", there is a 30% chance of getting a hit.

(a) Use conditioning to determine the player's expected number of hits per game.

(b) Use conditioning to find the probability that the player will get no hits in a game.

5. Let $Y_1, Y_2, \ldots, Y_n$ be independent and identically distributed with probability density function given by

$$f(y) = \begin{cases} \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{y^2}{2\theta}}, & \text{for } -\infty < y < \infty,\ \theta > 0 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Find the Cramer-Rao lower bound for an unbiased estimator of θ.

(b) Find the MLE $\hat{\theta}$ of θ.

(c) Is $\hat{\theta}$ an unbiased estimate of θ? Why or why not ?

(d) Find the MLE of $\ln \theta$ and justify your answer.

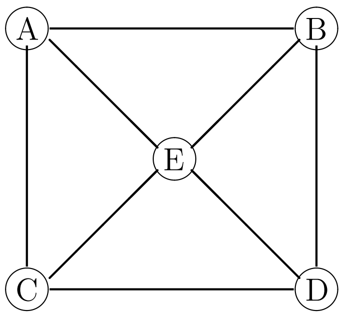
6. Let $X(t)$ be a pure birth process with initial value $X(0) = 1$ and the birth rate $\lambda_n = \lambda n$. Let $P_n(t) = P(X(t) = n)$. Find a system of differential equations for $P_n(t)$ and show that their solution is

$$P_n(t) = e^{-\lambda t}(1 - e^{-\lambda t})^{n-1}, n \geq 1.$$

7. A Markov chain is defined by a random walk on the graph pictured below. From a given node, you are equally likely to go to any neighboring node.

(a) Specify the transition matrix and find the stationary distribution for this Markov chain.

(b) Find the expected time it takes, when starting from A, to visit D.



Answers

1. (a) The likelihood function is

$$L(\theta; X_1, ..., X_n) = \prod_{i=1}^{n} \frac{1}{\theta} X_i^{1/\theta-1}, \quad \implies \quad \ln L(\theta) = -n\ln\theta + (1/\theta - 1)\sum_{i=1}^{n} \ln X_i$$

$$\frac{\partial \ln L}{\partial \theta} = -\frac{n}{\theta} - \frac{1}{\theta^2}\sum_{i=1}^{n} \ln X_i = 0 \quad \implies \quad \hat{\theta} = \frac{\sum_{i=1}^{n} \ln X_i}{n},$$

numerically $\hat{\theta} \approx 1.36$

(b) for M.O.M., find $\mathbb{E}(X)$ and equate it to $\overline{X}$.

$$\mathbb{E}(X) = \int x f(x)\, dx = \int_0^1 x * \frac{1}{\theta} x^{1/\theta-1}\, dx = \int_0^1 \frac{1}{\theta} x^{1/\theta}\, dx = \frac{1}{\theta} * \frac{x^{1/\theta+1}}{1/\theta+1}\Bigg]_0^1 = \frac{1}{1+\theta},$$

Thus $\dfrac{1}{1+\tilde{\theta}} = \overline{X} = 0.305 \quad \implies \quad \tilde{\theta} = \frac{1}{\overline{X}} - 1 \approx 2.28$

2. (a)

$$f_X(x) = \int f(x,y)\, dy = \int_0^{\sqrt{x}} 6xy\, dy = 6x\frac{y^2}{2}\Bigg]_{y=0}^{\sqrt{x}} = 3x^2, 0 < x < 1.$$

$$f_Y(y) = \int f(x,y)\, dx = \int_{y^2}^{1} 6xy\, dy = 6y\frac{x^2}{2}\Bigg]_{x=y^2}^{1} = 3y(1-y^4), \quad 0 < y < 1$$

Since $f(x,y) \neq f_X(x) f_Y(y)$, X and Y are not independent. Alternatively, notice that the region boundary is $y < \sqrt{x}$, therefore X, Y cannot be independent.

(b)

$$f_{X|Y=y}(x) = \frac{f(x,y)}{f_Y(y)} = \frac{6xy}{3y(1-y^4)} = \frac{2x}{1-y^4}, y^2 < x < 1$$

(c)

$$\mathbb{E}[X|Y=y] = \int x f_{X|Y=y}(x)\, dx = \int_{y^2}^{1} \frac{x * 2x}{1-y^4}\, dx =$$

$$= \frac{2}{1-y^4}\int_{y^2}^{1} x^2\, dx = \frac{2}{1-y^4} * \frac{x^3}{3}\Bigg]_{x=y^2}^{1} = \frac{2(1-y^6)}{3(1-y^4)}$$

(d)

$$Var[X|Y=y] = \mathbb{E}\left[X^2|Y=y\right] - \left(\mathbb{E}\left[X|Y=y\right]\right)^2,$$

$$\mathbb{E}\left[X^2|Y=y\right] = \int x^2 f_{X|Y=y}(x)\,dx = \int_{y^2}^{1} \frac{x^2 * 2x}{1-y^4}\,dx =$$

$$= \frac{2}{1-y^4}\int_{y^2}^{1} x^3\,dx = \frac{2}{1-y^4} * \frac{x^4}{4}\Bigg]_{x=y^2}^{1} = \frac{1-y^8}{2(1-y^4)},$$

Hence, $Var[X|Y=y] = \dfrac{1-y^8}{2(1-y^4)} - \left[\dfrac{2(1-y^6)}{3(1-y^4)}\right]^2$

3. (a) `posterior` $\propto$ `prior` $\times$ `likelihood`

$$f(\theta\,|\,X_1,...,X_n) \propto \theta^{\alpha-1}e^{-\theta/\beta}\prod_{i=1}^{n} e^{-\theta}\frac{\theta^{X_i}}{X_i!} \propto \theta^{\alpha-1+\sum X_i}e^{-\frac{\theta}{\beta}-n\theta},$$

which is the Gamma density with $\alpha^* = \alpha + \sum X_i$ and $\dfrac{1}{\beta^*} = \dfrac{1}{\beta} + n$, therefore $\beta^* = \dfrac{\beta}{1+n\beta}$

(b) Under the square loss, the Bayes estimate is the posterior mean $\mathbb{E}\left[\theta\,|\,X_1,...,X_n\right]$, for the Gamma distribution above it's

$$\hat{\theta} = \alpha^*\beta^* = \frac{(\alpha+\sum X_i)\beta}{1+n\beta}$$

(c)

$$\lim_{n\to\infty}\frac{(\alpha+\sum X_i)\beta}{1+n\beta} = \lim\frac{\alpha\beta}{1+n\beta} + \lim\frac{\sum X_i}{n} * \lim\frac{n\beta}{1+n\beta} = \lim\frac{\sum X_i}{n},$$

and due to the Law of Large Numbers, $\dfrac{\sum X_i}{n} \to_P \theta$, therefore $\hat{\theta}$ is consistent.
Alternatively, you can quote the theorem of consitency for MLE, and the fact that the Bayes estimate approaches MLE as $n \to \infty$.

4. Let A be the number of at bats the player gets. This is a binomial random variable with n=6 and p=0.8. Let H be the number of hits the player gets. This is a binomial random variable with $n = A$, and $p = 0.3$. Then

$$E[H] = \sum_{k=0}^{6} E[H|A = k]P(A = k).$$

Since $E[H|A = 0] = 0$, we can simplify this to

$$E[H] = \sum_{k=1}^{6} E[H|A = k]P(A = k).$$

Given k at bats, the expected number of hits is $E[H|A = k] = 0.3k$. The probability that $A = k$ is a binomial probability

$$P(A = k) = \binom{6}{k} 0.8^k 0.2^{(6-k)}$$

For $k = 0, 1, 2, \ldots, 6$, these probabilities are $6.4 \times 10^{-5}, 1.5 \times 10^{-3}, 1.54 \times 10^{-2}, 8.192 \times 10^{-2}, 0.2458, 0.3932, 0.2621$.

$$E[H] = \sum_{k=1}^{6} 0.3k \binom{6}{k} 0.8^k 0.2^{(6-k)} = 1.440$$

The probability that the player gets no hits in the game is

$$P(H = 0) = \sum_{k=0}^{6} P(H = 0|A = k)P(A = k).$$

$$P(H = 0) = \sum_{k=0}^{6} 0.7^k \binom{6}{k} 0.8^k 0.2^{(6-k)} = 0.1927.$$

Alt. solution:

We can notice that the probability of a hit on any given at bat is $p = 0.8 * 0.3 = 0.24$. Thus, the total number of hits H is Binomial with $n = 6$ and $p = 0.24$. Thus,

$$\mathbb{E}(H) = np = 1.44 \text{ and } P(0 \texttt{ hits}) = (1 - 0.24)^6 \approx 0.1927$$

5. (a) CRLB states that, for an unbiased estimate $\hat{\theta}$, $Var(\hat{\theta}) \geq \frac{1}{I_n(\theta)}$, where $I_n(\theta) = nI(\theta) = n\mathbb{E}\left(-\frac{\partial^2}{\partial\theta^2}\ln f(\theta; X)\right)$ is the Fisher information.

We have

$$\ln f(\theta; X) = const - \frac{1}{2}\ln\theta - \frac{X^2}{\theta} \quad \Longrightarrow \quad \frac{\partial}{\partial\theta}\ln f(\theta; X) = -\frac{1}{2\theta} + \frac{X^2}{2\theta^2}$$

$$\Longrightarrow \quad \frac{\partial^2}{\partial\theta^2}\ln f(\theta; X) = \frac{1}{2\theta^2} - \frac{X^2}{\theta^3}$$

$$\Longrightarrow \quad I(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ln f\right] = -\frac{1}{2\theta^2} + \frac{\mathbb{E}(X^2)}{\theta^3} = \frac{1}{2\theta^2}$$

because $\mathbb{E}(X^2) = Var(X) + (\mathbb{E}(X))^2 = \theta + 0$ for the given Normal distribution.

Thus, $Var(\hat{\theta}) \geq \frac{2\theta^2}{n}$

(b) The likelihood function is

$$L(\theta; X_1, ..., X_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{X_i^2}{2\theta}}, \quad \Longrightarrow \quad \ln L(\theta) = const - \frac{n}{2}\ln\theta - \frac{\sum_{i=1}^n X_i^2}{2\theta}$$

$$\Longrightarrow \quad \frac{\partial \ln L}{\partial\theta} = -\frac{n}{2\theta} + \frac{\sum_{i=1}^n X_i^2}{2\theta^2} = 0 \quad \Longrightarrow \quad \frac{\sum_{i=1}^n X_i^2}{2\theta^2} = \frac{n}{2\theta}$$

Therefore, the MLE is $\hat{\theta} = \frac{\sum_{i=1}^n X_i^2}{n}$

(c) Unbiased: need $\mathbb{E}(\hat{\theta}) = \theta$, this follows because

$$\mathbb{E}(\hat{\theta}) = \frac{\sum_{i=1}^n \mathbb{E}[X_i^2]}{n} = \frac{n\theta}{n} = \theta.$$

(d) by MLE invariance (or equivariance) property, if $\hat{\theta}$ is the MLE for θ, then for some function g, $g(\hat{\theta})$ is the MLE for $g(\theta)$. Thus, the MLE for $\ln\theta$ equals $\ln\hat{\theta} = \ln\left(\frac{\sum_{i=1}^n X_i^2}{n}\right)$

6. Using Forward Kolmogorov equations,

$$P'_{i,j}(t) = \lambda_{j-1}P_{i,j-1}(t) + \mu_{j+1}P_{i,j+1}(t) - (\lambda_j + \mu_j)P_{i,j}(t)$$

let $i = 1$ and $j = n$, then it follows

$$P'_n(t) = (n-1)\lambda P_{n-1}(t) - n\lambda P_{n-1}(t) \quad (n \geq 2) \tag{1}$$

and $P_1'(t) = -\lambda P_1(t) \implies P_1(t) = e^{-\lambda t}$, which also satisfies the initial condition $P_1(0) = 1$. Plugging the given expression $P_n(t) = e^{-\lambda t}(1 - e^{-\lambda t})^{n-1}$ into (1), we need to check

$$\left[e^{-\lambda t}(1 - e^{-\lambda t})^{n-1}\right]' = (n-1)\lambda e^{-\lambda t}(1 - e^{-\lambda t})^{n-2} - n\lambda e^{-\lambda t}(1 - e^{-\lambda t})^{n-1},$$

which can be verified after simplification.

7. (a) The transition probability matrix

$$\mathbb{P} = \begin{array}{c} \\ \mathbf{A} \\ \mathbf{B} \\ \mathbf{C} \\ \mathbf{D} \\ \mathbf{E} \end{array} \begin{Vmatrix} \mathbf{A} & \mathbf{B} & \mathbf{C} & \mathbf{D} & \mathbf{E} \\ \hline 0 & 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 0 & 0 & 1/3 & 1/3 \\ 1/3 & 0 & 0 & 1/3 & 1/3 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 \end{Vmatrix}$$

Let the stationary distribution $\boldsymbol{\pi} = [\pi_A, \pi_B, \pi_C, \pi_D, \pi_E]'$.
Solve the equations $\boldsymbol{\pi}' = \mathbb{P}\boldsymbol{\pi}$ and $\sum \pi_i = 1$.

Due to symmetry, $\pi_A = \pi_B = \pi_C = \pi_D$, so let $x = \pi_E$ and $y = \pi_A$. Then we get the system

$$\begin{cases} \pi_A = \pi_B/3 + \pi_C/3 + \pi_E/4 \\ \pi_A + \pi_B + \pi_C + \pi_D + \pi_E = 1 \end{cases} \implies$$

$$\begin{cases} x = \frac{4}{3}y \\ x + 4y = 1 \end{cases}$$

The solutions are $x = 4/16, y = 3/16$. Then $\boldsymbol{\pi}' = [3/16, 3/16, 3/16, 3/16, 4/16]$

(b)

Let $v_i = \mathbb{E}\left[\texttt{time to visit } D \,|\, X(0) = i\right]$. Using the first step analysis, we get the system

$$\begin{cases} v_A = 1 + \frac{1}{3}(v_B + v_C + v_E) \\ v_B = 1 + \frac{1}{3}(v_A + v_D + v_E) = v_C \\ v_E = 1 + \frac{1}{3}(v_B + v_C + v_E) \\ v_D = 0 \end{cases}$$

Solving, we obtain $v_A = \dfrac{16}{3}$, $v_B = v_C = \dfrac{64}{15}$ and $v_E = \dfrac{67}{15}$