STAT 244 (Statistical Theory/Method I), Winter 2016

Homework 6 solutions

February 22, 2016

100 points total. Marks are assigned based on progress made.

Rice 8.30 (10 points)

(a) The probability of observing a measurement larger than 10.0 is $\mathbb{P}(X > 10) = 1 - F(10) = e^{-10\lambda}$. Assuming the measurements are independent of one another, the likelihood is given by

$$\ell(\lambda) = f(x_1 \mid \lambda) \times f(x_2 \mid \lambda) \times \mathbb{P}(X_3 > 10) = (\lambda e^{-\lambda x_1})(\lambda e^{-\lambda x_2})(e^{-10\lambda}) = \lambda^2 e^{-\lambda(x_1+x_2+10)} = \boxed{\lambda^2 e^{-18\lambda}.} \tag{1}$$

(b) Solving $\frac{\mathrm{d}\ell}{\mathrm{d}\lambda} = \frac{2}{\lambda} - 18 = 0$ gives $\boxed{\hat{\lambda} = \frac{1}{9}}$ as the maximum likelihood estimator of λ.

6 points for (a), 4 points for (b).

Rice 8.52 (20 points)

(b) The log likelihood is

$$\ell(\theta) = \log\left(\prod_{i=1}^{n}(\theta+1)x_i^\theta\right) = n\log(\theta+1) + \theta\sum_{i=1}^{n}\log(x_i), \tag{2}$$

and solving $\frac{\mathrm{d}\ell}{\mathrm{d}\theta} = \frac{n}{\theta+1} + \sum_{i=1}^{n}\log(x_i) = 0$ gives $\boxed{\hat{\theta} = -n \Big/ \sum_{i=1}^{n}\log(x_i) - 1}$ as the maximum likelihood estimator of θ. Since $\frac{{}^2\ell}{\mathrm{d}\theta^2} = -\frac{n}{(\theta+1)^2} < 0$, this is indeed a maximum.

(c) According to Fisher's approximation in Section 5.6 of Stigler's notes, the asymptotic variance of $\hat{\theta}$ is given by $\frac{\tau^2(\theta)}{n}$, where

$$\frac{1}{\tau^2(\theta)} = -\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2}\log\left(f(x \mid \theta)\right)\right) = -\mathbb{E}\left(\frac{\partial}{\partial\theta}\left(\frac{1}{\theta+1}\right)\right) = -\mathbb{E}\left(-\frac{1}{(\theta+1)^2}\right) = \frac{1}{(\theta+1)^2}. \tag{3}$$

Hence the asymptotic variance of $\hat{\theta}$ is $\boxed{\frac{(\theta+1)^2}{n}.}$

(d) Note that for $\boxed{T(x_1, \ldots, x_n) := \prod_{i=1}^{n} x_i}$, we have the trivial factorization

$$f(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^{n}(\theta+1)x_i^\theta = \underbrace{(\theta+1)^n\left(\prod_{i=1}^{n}x_i\right)^\theta}_{g(T(x_1,\ldots,x_n),\theta)} \times \underbrace{1}_{h(x_1,\ldots,x_n)}, \tag{4}$$

so T is a sufficient statistic by Theorem A on page 306 of the textbook.

6 points for (b) and (c), 8 points for (d).

Rice 8.60 (25 points)

(a) The log likelihood is

$$\ell(\tau) = \log\left(\prod_{i=1}^{n} \frac{1}{\tau} e^{-\frac{x_i}{\tau}}\right) = -n\log(\tau) - \frac{1}{\tau}\sum_{i=1}^{n} x_i, \tag{5}$$

and solving $\frac{\mathrm{d}\ell}{\mathrm{d}\tau} = -\frac{n}{\tau} + \frac{1}{\tau^2}\sum_{i=1}^{n} x_i$ gives the sample average $\boxed{\hat{\tau} = \sum_{i=1}^{n} x_i \Big/ n}$ as the maximum likelihood estimator of τ.

(b) If $X_1 \sim \mathrm{Gamma}(\alpha_1, \beta)$ and $X_2 \sim \mathrm{Gamma}(\alpha_2, \beta)$ are independent then $X_1 + X_2 \sim \mathrm{Gamma}(\alpha_1 + \alpha_2, \beta)$, and since the exponential distribution with parameter τ is precisely the gamma distribution with parameters 1 and τ, the sum of n exponential random variables with parameter τ follows a gamma distribution with parameters n and τ, so $\boxed{\hat{\tau} \sim \mathrm{Gamma}(n, \frac{\tau}{n})}$ after scaling by n.

(c) Since $\hat{\tau}$ is the sample average of independent and identically distributed random variables with mean $\mathbb{E}(X_i) = \tau$ and variance $\mathrm{Var}(X_i) = \tau^2$, the central limit theorem implies $\boxed{\hat{\tau} \sim \mathcal{N}(\tau, \frac{\tau^2}{n})}$ for large n.

(d) According to part (b), the mean of $\hat{\tau}$ is the mean of a $\mathrm{Gamma}(n, \frac{\tau}{n})$-distributed random variable, which is $\boxed{\mathbb{E}(\hat{\tau}) = n \times \frac{\tau}{n} = \tau}$. Hence $\hat{\tau}$ is unbiased. Similarly, we have $\boxed{\mathrm{Var}(\hat{\tau}) = n \times \left(\frac{\tau}{n}\right)^2 = \frac{\tau^2}{n}.}$

(e) Since $f(x \mid \tau)$ is smooth, it follows from the Cramér-Rao inequality (Theorem A on page 300 of the textbook) that the variance of any unbiased estimator of τ is bounded below by $\frac{\tau^2(\tau)}{n}$, where

$$\begin{aligned}\frac{1}{\tau^2(\tau)} &= -\mathbb{E}\left(\frac{\partial^2}{\partial\tau^2}\log\left(f(x \mid \tau)\right)\right) = -\mathbb{E}\left(\frac{\partial}{\partial\tau}\left(-\frac{1}{\tau} + \frac{1}{\tau^2}\sum_{i=1}^{n} X_i\right)\right) = -\mathbb{E}\left(\frac{1}{\tau^2} - \frac{2}{\tau^3}\sum_{i=1}^{n} X_i\right) \\ &= -\frac{1}{\tau^2} + \frac{2}{\tau^3}\sum_{i=1}^{n}\mathbb{E}(X_i) = -\frac{1}{\tau^2} + \frac{2}{\tau^3}(n\tau) = \frac{1}{\tau^2},\end{aligned} \tag{6}$$

so the answer is no, the MLE $\hat{\tau} = \sum_{i=1}^{n} x_i \Big/ n$ achieves the smallest possible variance among all unbiased estimators.

5 points for each part.

Rice 8.68 (25 points)

(a) The joint density of $X_1, \ldots, X_n$ and $T(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i$ is

$$\begin{aligned}f(x_1, \ldots, x_n, t) &= f\left(x_1, \ldots, x_n = t - \sum_{i=1}^{n-1} x_i\right) \\ &= \prod_{i=1}^{n-1} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \times \frac{\lambda^{t - \sum_{i=1}^{n-1} x_i}}{\left(t - \sum_{i=1}^{n-1} x_i\right)!} e^{-\lambda} = \frac{\lambda^t}{\left(t - \sum_{i=1}^{n-1} x_i\right)! \prod_{i=1}^{n-1} x_i!} e^{-n\lambda}\end{aligned} \tag{7}$$

Furthermore, recall that the sum of n Poisson random variables with parameter λ is a again a Poisson random variable with parameter $n\lambda$, so the density of T is

$$f(t) = \frac{(n\lambda)^t}{t!} e^{-n\lambda}. \tag{8}$$

Then to find the conditional density of $X_1, \ldots, X_n$ given T, we divide (7) by (8). Note that the terms that depend on λ (λ^t and $e^{-n\lambda}$) cancel in the numerator and denominator, so the conditional density indeed does not depend on λ. This means T is sufficient.

(b) As in part (a), we have

$$f(x_1, \ldots, x_n, t) = f(x_1 = t, \ldots, x_n) = \frac{\lambda^t}{t!} e^{-\lambda} \times \prod_{i=2}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \quad \text{and} \tag{9}$$

$$f(t) = f(x_1 = t) = \frac{\lambda^t}{t!} e^{-\lambda}, \tag{10}$$

so the conditional density of $X_1, \ldots, X_n$ given T is given by

$$f(x_1, \ldots, x_n \mid t) = \frac{f(x_1, \ldots, x_n, t)}{f(t)} = \prod_{i=2}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}, \tag{11}$$

which obviously does depend on λ, and hence T is *not* sufficient.

(c) We have the factorization

$$f(x_1, \ldots, x_n \mid \lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \underbrace{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}_{g(T(x_1,\ldots,x_n),\lambda)} \times \underbrace{\left(\prod_{i=1}^{n} x_i! \right)^{-1}}_{h(x_1,\ldots,x_n)}, \tag{12}$$

so T is a sufficient statistic by the factorization theorem.

10 points for (a) and (b), 5 points for (c).

Rice 8.70 (10 points)

Let $X_1, \ldots, X_n$ be independent and identically distributed exponential random variables with parameter τ. Note that with the choice of $T(x_1, \ldots, x_n) := \boxed{\sum_{i=1}^{n} x_i}$, we have the trivial factorization

$$f(x_1, \ldots, x_n \mid \tau) = \prod_{i=1}^{n} \frac{1}{\tau} e^{\frac{-x_i}{\tau}} = \underbrace{\frac{1}{\tau^n} e^{-\sum_{i=1}^{n} x_i \big/ \tau}}_{g(T(x_1,\ldots,x_n),\tau)} \times \underbrace{1}_{h(x_1,\ldots,x_n)}, \tag{13}$$

so $\sum_{i=1}^{n} x_i$ is a sufficient statistic.

Rice 8.71 (10 points)

With the choice of $T(x_1, \ldots, x_n) := \boxed{\prod_{i=1}^{n} (1 + x_i)}$, we have the trivial factorization

$$f(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} \frac{\theta}{(1 + x_i)^{\theta+1}} = \underbrace{\frac{\theta^n}{\left(\prod_{i=1}^{n} (1 + x_i) \right)^{\theta+1}}}_{g(T(x_1,\ldots,x_n),\theta)} \times \underbrace{1}_{h(x_1,\ldots,x_n)}. \tag{14}$$