Midterm 2 Solution Key

Feng-Chang Lin

Department of Biostatistics
University of North Carolina at Chapel Hill

flin@bios.unc.edu

Problem 1

Suppose $X_1, \ldots, X_n$ are iid Poisson(θ) with a probability density function

$$f(x|\theta) = \frac{\theta^x e^{-\theta}}{x!}, \quad \theta > 0, \quad x = 0, 1, 2, \ldots .$$

(a) Show that $I(X_1 = 0)$ is an unbiased estimator of $e^{-\theta}$, where

$$I(X_1 = 0) = \begin{cases} 1 & \text{if } X_1 = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Solution: Since $E\{I(X_1 = 0)\} = P(X_1 = 0) = e^{-\theta}$, $I(X_1 = 0)$ is an unbiased estimator of $e^{-\theta}$.

Problem 1 (cont'd)

(b) Show that $\sum_{i=1}^{n} X_i$ is a complete sufficient statistic.
Solution: Since

$$L(\theta|\boldsymbol{x}) = \exp\left\{(\log\theta)\sum_{i=1}^{n} x_i - n\theta + \sum_{i=1}^{n}\log(x_i!)\right\},$$

one can claim the this distribution belongs to an exponential family and conclude that $\sum_{i=1}^{n} X_i$ is a complete sufficient statistic.

Problem 1 (cont'd)

(c) Using Lehmann-Scheffe Theorem, show that $\phi(\sum_{i=1}^{n} X_i)$ is the best unbiased estimator (UMVUE) of $e^{-\theta}$, where

$$\phi\left(\sum_{i=1}^{n} X_i\right) = \left(1 - \frac{1}{n}\right)^{\sum_{i=1}^{n} X_i}$$

Solution: By Rao-Blackwell-Lehmann-Scheffe Theorem, one just need to show $\phi\left(\sum_{i=1}^{n} X_i\right)$ is an unbiased estimator of $e^{-\theta}$ since $\sum_{i=1}^{n} X_i$ is a complete sufficient statistic. We show that in the homework 5.

Problem 1 (cont'd)

In fact, one has

$$
\begin{aligned}
E\{I(X_1 = 0) | \sum_{i=1}^{n} X_i = t\} &= P(X_1 = 0 | \sum_{i=1}^{n} X_i = t) \\
&= \frac{P(X_1 = 0, \sum_{i=1}^{n} X_i = t)}{P(\sum_{i=1}^{n} X_i = t)} = \frac{P(X_1 = 0, \sum_{i=2}^{n} X_i = t)}{P(\sum_{i=1}^{n} X_i = t)} \\
&= \frac{P(X_1 = 0) P(\sum_{i=2}^{n} X_i = t)}{P(\sum_{i=1}^{n} X_i = t)} = \frac{e^{-\theta}\{(n-1)\theta\}^t e^{-(n-1)\theta}/t!}{\{n\theta\}^t e^{-n\theta}/t!} \\
&= (1 - 1/n)^t .
\end{aligned}
$$

By Lehmann-Scheffe Theorem, $\phi\left(\sum_{i=1}^{n} X_i\right) = \left(1 - \frac{1}{n}\right)^{\sum_{i=1}^{n} X_i}$ is the UMVUE.

Problem 1 (cont'd)

(d) Compute the Cramér-Rao Lower Bound for unbiased estimators of $e^{-\theta}$.
Solution: One has $d\tau(\theta)/d\theta = -e^{-\theta}$. One also has

$$E\left\{\frac{\partial^2}{\partial\theta^2}\ell(\theta|x)\right\} = E(-x\theta^{-2}) = -\theta^{-1}.$$

Therefore, the CRLB is $\theta e^{-2\theta}/n$.

(e) Find the MLE of $e^{-\theta}$.
Solution: The MLE of θ is $\bar{X}$. By the invariance property of MLE, the MLE of $e^{-\theta}$ is $e^{-\bar{X}}$.

Problem 2

Suppose that the random variables $Y_1, \ldots, Y_n$, $n > 2$ are independent and normally distributed with $EY_i = \theta x_i$, where $x_1, \ldots, x_n$ are known constants and none of which is zero. Let $\mathrm{Var}\, Y_i = \sigma^2 > 0$ and $\theta \in (-\infty, \infty)$. Assume that σ^2 is a known constant and θ is an unknown parameter.

(a) Find the method of moments estimator $\tilde{\theta}$ of θ, matching $M_1 = n^{-1} \sum_{i=1}^n Y_i$ and $E(M_1)$.

$$n^{-1} \sum_{i=1}^n Y_i \doteq E(M_1) = n^{-1} \sum_{i=1}^n E(Y_i) = n^{-1} \sum_{i=1}^n \theta x_i.$$

This implies $\tilde{\theta} = \bar{Y}/\bar{x}$.

Problem 2 (cont'd)

(b) Find the MLE $\hat{\theta}$ of θ and show that it is an unbiased estimator of θ.
Solution: The log-likelihood equals

$$\ell(\theta|\boldsymbol{y}) = -n\log(\sqrt{2\pi\sigma^2}) - (2\sigma^2)^{-1}\sum_{i=1}^{n}(y_i - \theta x_i)^2.$$

Set $\partial\ell(\theta|\boldsymbol{y})/\partial\theta = 0$. One can have $\hat{\theta} = \sum_{i=1}^{n} Y_i x_i / \sum_{i=1}^{n} x_i^2$. Check $\partial^2\ell(\theta|\boldsymbol{y})/\partial\theta^2 = -\sum_{i=1}^{n} x_i^2/\sigma^2 < 0$. Therefore, $\hat{\theta}$ is the MLE.

Problem 2 (cont'd)

(c) Find the distribution of $\hat{\theta}$.

Solution: Since $\hat{\theta}$ is a linear combination of Y_i, one can conclude $\hat{\theta}$ is also normally distributed. One can have

$$E(\hat{\theta}) = \sum_{i=1}^{n} E(Y_i)x_i / \sum_{i=1}^{n} x_i^2 = \theta,$$

and

$$\begin{aligned} Var(\hat{\theta}) &= \sum_{i=1}^{n} Var(Y_i)x_i^2 / (\sum_{i=1}^{n} x_i^2)^2 \\ &= \sum_{i=1}^{n} \sigma^2 x_i^2 / (\sum_{i=1}^{n} x_i^2)^2 = \sigma^2 / \sum_{i=1}^{n} x_i^2. \end{aligned}$$

One can conclude $\hat{\theta} \sim N(\theta, \sigma^2 / \sum_{i=1}^{n} x_i^2)$.

Problem 2 (cont'd)

(d) Let $T_1 = \sum_{i=1}^n Y_i / \sum_{i=1}^n x_i$ and $T_2 = \sum_{i=1}^n (Y_i/x_i)/n$. Show that both T_1 and T_2 are unbiased estimators of θ.
Solution:

$$E(T_1) = \sum_{i=1}^n E(Y_i) / \sum_{i=1}^n x_i = \sum_{i=1}^n \theta x_i / \sum_{i=1}^n x_i == \theta,$$

and

$$E(T_2) = \sum_{i=1}^n \{E(Y_i)/x_i\}/n = \sum_{i=1}^n (\theta x_i / x_i)/n = \theta.$$

Both T_1 and T_2 are unbiased estimator.

Problem 2 (cont'd)

(e) Show that the variance of $\hat{\theta}$ is smaller than the variance of both T_1 and T_2, i.e., Var($\hat{\theta}$) $\leq$ Var(T_1) and Var($\hat{\theta}$) $\leq$ Var(T_2).

Solution:

$$Var(T_1) = \sum_{i=1}^{n} Var(Y_i)/(\sum_{i=1}^{n} x_i)^2 = \sigma^2 n/(\sum_{i=1}^{n} x_i)^2.$$

One can claim that $\sum_{i=1}^{n} x_i^2 \geq (\sum_{i=1}^{n} x_i)^2/n = n\bar{x}^2$ because

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \geq 0.$$

That means,

$$Var(\hat{\theta}) = \sigma^2/\sum_{i=1}^{n} x_i^2 \leq \sigma^2 n/(\sum_{i=1}^{n} x_i)^2 = Var(T_1).$$

Problem 2 (cont'd)

(e) Then,

$$Var(T_2) = \sum_{i=1}^{n} \{Var(Y_i)/x_i^2\}/n^2 = \sigma^2 \sum_{i=1}^{n} x_i^{-2}/n^2.$$

One also can claim that $\sum_{i=1}^{n} x_i^2 \geq n^2 / \sum_{i=1}^{n} x_i^{-2}$ because of the fact that $n^{-1} \sum_{i=1}^{n} x_i^2 \geq (n^{-1} \sum_{i=1}^{n} x_i^{-2})^{-1}$. Hence,

$$Var(\hat{\theta}) = \sigma^2 / \sum_{i=1}^{n} x_i^2 \leq \sigma^2 \sum_{i=1}^{n} x_i^{-2}/n^2 = Var(T_2).$$

We can claim the variance of $\hat{\theta}$ is smaller than the variance of both T_1 and T_2.

Problem 3

Let $X_1, \ldots, X_n$ be a sample from the distribution with probability density function

$$f(x|\theta) = e^{-(x-\theta)}, \ \ \theta \leq x < \infty, \ \ -\infty < \theta < \infty.$$

If one tries to test $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$.

(a) Find the probability density function of $X_{(1)} = \min\{X_1, \ldots, X_n\}$.
Solution: One has

$$F(x|\theta) = \int_\theta^x f(t|\theta)dt = 1 - e^{-(x-\theta)}.$$

$$\begin{aligned} f_{X_{(1)}}(x) &= \frac{n!}{(n-1)!} e^{-(x-\theta)} \{e^{-(x-\theta)}\}^{n-1} \\ &= ne^{-n(x-\theta)}, \ \ x > \theta. \end{aligned}$$

Problem 3 (cont'd)

(b) Find the likelihood ratio test statistic $\lambda(x)$, as a function of $X_{(1)}$. If your test statistic depends on the range of $X_{(1)}$, please indicate it.
Solution: The numerator of $\lambda(x)$ is, under H_0,

$$L(0|\boldsymbol{x}) = \sum_{i=1}^{n} e^{-x_i} I(x_{(1)} > 0).$$

The denominator of $\lambda(x)$ is, under unrestricted θ,

$$L(x_{(1)}|\boldsymbol{x}) = \sum_{i=1}^{n} e^{-(x_i - x_{(1)})}.$$

Hence, the test statistic $\lambda(x) = e^{-nX_{(1)}} I(X_{(1)} > 0)$.

Problem 3 (cont'd)

(c) Draw a figure of your test statistic $\lambda(x)$ as a function of $x_{(1)}$.

Problem 3 (cont'd)

(d) By the likelihood ratio test, one rejects H_0 if $\delta(x) = 1$, where

$$\delta(x) = \begin{cases} 1 & \text{if } \lambda(x) < c, \\ 0 & \text{if } \lambda(x) > c. \end{cases}$$

Show that, equivalently, one can use the following rejection region:

$$\delta(x) = \begin{cases} 1 & \text{if } X_{(1)} > c^*, \\ 0 & \text{if } X_{(1)} < c^*. \end{cases}$$

Solution: The rejection region for $\lambda(x) < c$ is equivalent to $X_{(1)} > -\log(c)/n$ and $X_{(1)} < 0$. However, $P(X_{(1)} < 0) = 0$. Therefore, the only possible rejection region is $X_{(1)} > c^*$ where $c^* = -\log(c)/n$.

Problem 3 (cont'd)

(e) Following (d), find c^* such that the type I error probability of the test equals 0.05.
Solution:

$$0.05 = P(X_{(1)} > c^* | \theta = 0) = \int_{c^*}^{\infty} f_{X_{(1)}}(t|\theta = 0) dt = e^{-c^* n}.$$

That shows $c^* = -\log(0.05)/n$.

Homework 6 Problem 4

Let $X_1, X_2, \ldots, X_n$ be iid (continuous) random variables from the pdf.

$$f(x|\mu) = (2\pi x^3)^{-1/2} \exp\left\{-\frac{x-2\mu}{2\mu^2} - \frac{1}{2x}\right\}, \quad x > 0,$$

where $\mu > 0$ is an unknown parameter.

(a) Derive the uniformly most powerful size α test ($0 < \alpha < 1$) for $H_0 : \mu = 1$ against $H_1 : \mu > 1$. Specify the form of the rejection (critical) region as concisely as possible.

$$f(\boldsymbol{x}|\mu) = \prod_{i=1}^{n} (2\pi x_i^3)^{-1/2} \exp\left\{-\sum_{i=1}^{n} \frac{x_i - 2\mu}{2\mu^2} - \sum_{i=1}^{n} \frac{1}{2x_i}\right\},$$

$$f(\boldsymbol{x}|\mu = 1) = \prod_{i=1}^{n} (2\pi x_i^3)^{-1/2} \exp\left\{-\sum_{i=1}^{n} \frac{x_i - 2}{2} - \sum_{i=1}^{n} \frac{1}{2x_i}\right\}.$$

Homework 6 Problem 4

(a) By Neyman-Pearson Lemma, the rejection region is

$$R = \left\{ \boldsymbol{x} | \frac{f(\boldsymbol{x}|\mu)}{f(\boldsymbol{x}|\mu = 1)} > c^* \right\}.$$

One can have

$$\frac{f(\boldsymbol{x}|\mu)}{f(\boldsymbol{x}|\mu = 1)} = \exp\left\{ -\sum_{i=1}^{n} \frac{x_i - 2\mu}{2\mu^2} + \sum_{i=1}^{n} \frac{x_i - 2}{2} \right\} > c,$$

or, equivalently,

$$n^{-1} \sum_{i=1}^{n} x_i > (\log c^*/n - 2\mu + 2\mu^2)/(\mu^2 - 1).$$

Homework 6 Problem 4 (cont'd)

(b) An investigator wants to design a study for testing the hypothesis stated above. Suppose that we wish to use the test (which may or may not be the UMP test) that rejects H_0 if $n^{-1}\sum_{i=1}^{n} X_i > c$ for some constant c. The investigator desires a type I error probability of 0.01 and a maximum type II error probability of 0.1 at $\mu = 1.1$. Find values (approximate or exact) of n and c that will achieve this.

$$\alpha = P_{\mu=1}\left(n^{-1}\sum_{i=1}^{n} X_i > c\right),$$

and

$$1-\beta = P_{\mu=1.1}\left(n^{-1}\sum_{i=1}^{n} X_i > c\right).$$