# Midterm 1 Solution Key

Feng-Chang Lin

Department of Biostatistics
University of North Carolina at Chapel Hill

flin@bios.unc.edu

# Problem 1

Suppose that iid random variables $X_1, \ldots, X_n$ follow a uniform distribution on the interval $(0, 1)$ with pdf

$$f_X(x) = 1, \;\; 0 < x < 1.$$

Let random variables $U = X_{(1)}$ and $V = 1 - X_{(n)}$, where $X_{(1)} = \min_i X_i$ and $X_{(n)} = \max_i X_i$ are minimum and maximum order statistics, respectively.

(a) Find an explicit expression for the joint distribution of the random variables $U$ and $V$.

$$f_{X_{(1)},X_{(n)}}(x_{(1)}, x_{(n)}) = \frac{n!}{1!(n-2)!1!}(x_{(n)} - x_{(1)})^{n-2}.$$

$U$ and $V$ are transformations of $X_{(1)}$ and $X_{(n)}$.

# Problem 1 (cont'd)

Hence, one can have $X_{(1)} = U$, $X_{(n)} = 1 - V$, and $|J| = -1$. The joint pdf of $(U, V)$ is

$$f_{U,V}(u,v) = n(n-1)(1-v-u)^{n-2}, \;\; 0 < u < 1, \;\; 0 < u+v < 1.$$

(b) Let $R = nU$ and $S = nV$. Show that

$$P(R > r, S > s) = \left(1 - \frac{r}{n} - \frac{s}{n}\right)^n.$$

$$P(R > r, S > s) = P(U > \frac{r}{n}, V > \frac{s}{n}) = \int_{s/n}^{1-r/n} \int_{1-v}^{r/n} f_{U,V}(u,v)dudv.$$

# Problem 1 (cont'd)

(c) Following the result in (b), show that $R$ and $S$ are asymptotically independent. You may need the fact that

$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x.$$

$$\lim_{n\to\infty} P(R>r, S>s) = \lim_{n\to\infty}\left(1-\frac{r}{n}-\frac{s}{n}\right)^n = e^{-(r+s)} = e^{-r}e^{-s}$$

$$\lim_{n\to\infty} P(R>r) = \lim_{n\to\infty}\left(1-\frac{r}{n}\right)^n = e^{-r}$$

$$\lim_{n\to\infty} P(S>s) = \lim_{n\to\infty}\left(1-\frac{s}{n}\right)^n = e^{-s}$$

$R$ and $S$ are asymptotically independent because the joint survival function equals the product of two individual survival function.

# Problem 2

A certain simple biological system involves exactly two independently functioning components. If one of these two components fails, then entire systems fails. For $i = 1, 2$, let $Y_i$ be the random variable representing the time to failure of the $i$th component, with the pdf of $Y_i$ being

$$f_{Y_i}(y_i) = \theta_i e^{-\theta_i y_i}, \quad 0 < y_i < \infty, \quad \theta_i > 0.$$

Clearly, if this biological system fails, then only two random variables are observable, namely $U$ and $W$, where $U = \min(Y_1, Y_2)$ and

$$W = \begin{cases} 1, & \text{if } \; Y_1 < Y_2, \\ 0, & \text{if } \; Y_2 < Y_1. \end{cases}$$

# Problem 2 (cont'd)

(a) Show that the joint distribution $f_{U,W}(u,w)$ of random variables $U$ and $W$ is

$$f_{U,W}(u,w) = \theta_1^{(1-w)}\theta_2^w e^{-(\theta_1+\theta_2)u}, \quad 0 < u < \infty, \quad w = 0, 1,$$

by working on the derivation of $P(U \le u, W = 0)$ and $P(U \le u, W = 1)$.

$$\begin{aligned}
P(U \le u, W = 0) &= P(Y_2 \le u, Y_2 < Y_1) \\
&= \int_0^u \int_{y_2}^\infty \left(\theta_1 e^{-\theta_1 y_1}\right)\left(\theta_2 e^{-\theta_2 y_2}\right) dy_1 dy_2 \\
&= \left(\frac{\theta_1}{\theta_1+\theta_2}\right)\left\{1 - e^{-(\theta_1+\theta_2)u}\right\}
\end{aligned}$$

$$f_{U,W}(u, w = 0) = \theta_1 e^{-(\theta_1+\theta_2)u}, \quad 0 < u < \infty. \tag{1}$$

# Problem 2 (cont'd)

$$\begin{aligned} P(U \le u, W = 1) &= P(Y_1 \le u, Y_1 < Y_2) \\ &= \int_0^u \int_{y_1}^\infty \left(\theta_1 e^{-\theta_1 y_1}\right)\left(\theta_2 e^{-\theta_2 y_2}\right) dy_2 dy_1 \\ &= \left(\frac{\theta_2}{\theta_1 + \theta_2}\right)\left\{1 - e^{-(\theta_1+\theta_2)u}\right\} \end{aligned}$$

$$f_{U,W}(u, w = 1) = \theta_2 e^{-(\theta_1+\theta_2)u}, \quad 0 < u < \infty. \tag{2}$$

Combining (1) and (2), one has

$$f_{U,W}(u, w) = \theta_1^{(1-w)}\theta_2^w e^{-(\theta_1+\theta_2)u}, \quad 0 < u < \infty, \quad w = 0, 1.$$

$$P(U \le u, W = 1) = P(U \le u | W = 1)P(W = 1) \ne P(Y_1 \le u)P(Y_1 < Y_2).$$

# Problem 2 (cont'd)

(b) Find the marginal distribution $f_W(w)$ of the random variable $W$.

$$f_W(w) = \int_0^\infty f_{U,W}(u,w)du = \int_0^\infty \theta_1^{(1-w)}\theta_2^w e^{-(\theta_1+\theta_2)u} du$$
$$= \theta_1^{(1-w)}\theta_2^w(\theta_1+\theta_2)^{-1} = \left(\frac{\theta_1}{\theta_1+\theta_2}\right)^{(1-w)}\left(\frac{\theta_2}{\theta_1+\theta_2}\right)^w$$

(c) Find the marginal distribution $f_U(u)$ of the random variable $U$.

$$f_U(u) = \sum_{w=0}^{1} f_{U,W}(u,w) = (\theta_1+\theta_2)e^{-(\theta_1+\theta_2)u}$$

(d) Are $U$ and $W$ independent random variables?

$U$ are $W$ are independent because of $f_{U,W}(u,w) = f_U(u)f_W(w)$.

# Problem 3

Suppose that $X_1, X_2, \ldots, X_n$ are iid random variables distributed as Poisson with mean $\mu > 0$. Denote $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. We are interested in constructing a confidence interval for $\mu$.

(a) State the central limit theorem for $\bar{X}_n$.

$$\sqrt{n}(\bar{X}_n - \mu)/\sqrt{\mu} \to_d N(0,1).$$

(b) What is the asymptotic variance of $T_n = \sqrt{n}(\bar{X}_n - \mu)$?

$$\lim_{n\to\infty} Var(T_n/\sqrt{\mu}) = 1 \Rightarrow \lim_{n\to\infty} Var(T_n) = \mu.$$

(c) What is the appropriate function $h(\bar{X}_n)$ so that $h(\bar{X}_n)T_n \to_d N(0,1)$?

$$T_n/\sqrt{\mu} \to_d N(0,1) \Rightarrow h(\bar{X}_n) = 1/\sqrt{\bar{X}_n}.$$

# Problem 3 (cont'd)

(d) Use the last part to construct an approximate 95% confidence interval for $\mu$.

$$
\begin{aligned}
0.95 &= P(-1.96 < h(\bar{X}_n)T_n < 1.96) \\
&= P(-1.96 < h(\bar{X}_n)\sqrt{n}(\bar{X}_n - \mu) < 1.96) \\
&= P\left(\bar{X}_n - \frac{1.96}{h(\bar{X}_n)\sqrt{n}} < \mu < \bar{X}_n + \frac{1.96}{h(\bar{X}_n)\sqrt{n}}\right)
\end{aligned}
$$

# Problem 3 (cont'd)

(e) Another approach to eliminate $\mu$ from the asymptotic variance is to find a function $g$ such that $\sqrt{n}(g(\bar{X}_n) - g(\mu)) \to_d N(0,1)$. Find an explicit expression for $g(\mu)$.

According to the delta method,

$$\sqrt{n}(g(\bar{X}_n) - g(\mu)) \to_d N(0, \{g'(\mu)\}^2\mu) \equiv N(0,1).$$

That means $\{g'(\mu)\}^2\mu = 1$ and $g(\mu) = 2\mu^{1/2}$.

(f) Use the last part to construct an approximate 95% confidence interval for $\mu$.

$$\begin{aligned} 0.95 &= P(-1.96 < \sqrt{n}(g(\bar{X}_n) - g(\mu)) < 1.96) \\ &= P(g(\bar{X}_n) - 1.96/\sqrt{n} < g(\mu) < g(\bar{X}_n) + 1.96/\sqrt{n}) \\ &= P\left(\frac{1}{4}\{g(\bar{X}_n) - 1.96/\sqrt{n}\}^2 < \mu < \frac{1}{4}\{g(\bar{X}_n) + 1.96/\sqrt{n}\}^2\right) \end{aligned}$$

# Problem 4

Let $Y_1, \ldots, Y_n$ constitute a random sample from $N(0, \sigma^2)$.

(a) Show that $T = \sum_{i=1}^n Y_i^2$ is a sufficient statistic for unknown $\theta = \sigma^r$, where $r$ is a known positive integer.

$$
\begin{aligned}
f(y_1, \ldots, y_n | \sigma^2) &= \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{\sum_{i=1}^n y_i^2/\sigma^2} \\
&= \left( \frac{1}{\sqrt{2\pi(\sigma^r)^{2/r}}} \right)^n e^{\sum_{i=1}^n y_i^2/(\sigma^r)^{2/r}}
\end{aligned}
$$

By factorization theorem, $\sum_{i=1}^n Y_i^2$ is a sufficient statistic for $\sigma^r$.

(b) What is the distribution of $T/\sigma^2$? Justify your answer.

$$
\frac{T}{\sigma^2} = \sum_{i=1}^n \left( \frac{Y_i}{\sigma} \right)^2 = \sum_{i=1}^n Z_i^2 \sim \chi_n^2.
$$

# Problem 4 (cont'd)

(c) Develop an explicit expression for an unbiased estimator $\hat{\theta}$ that is a function of $T$.

$$E\left(\frac{T}{\sigma^2}\right)^{r/2} = \frac{\Gamma(n/2 + r/2)}{\Gamma(n/2)} 2^{r/2}$$

$$E(T^{r/2}) = \frac{\Gamma(n/2 + r/2)}{\Gamma(n/2)} 2^{r/2} \sigma^r$$

Therefore,

$$E\left(T^{r/2} 2^{-r/2} \frac{\Gamma(n/2)}{\Gamma(n/2 + r/2)}\right) = \sigma^r.$$

# Problem 5

(a) Let $X$ and $Y$ be random variable such that $E(X^k)$ and $E(Y^k) \neq 0$ exist for a positive integer $k$. If the ratio of $X/Y$ and its denominator $Y$ are independent, prove that $E\{(X/Y)^k\} = E(X^k)/E(Y^k)$.

$$E(X^k) = E\left(Y^k \frac{X^k}{Y^k}\right) = E(Y^k)E\left(\frac{X^k}{Y^k}\right) \Rightarrow E\left(\frac{X}{Y}\right)^k = \frac{E(X)^k}{E(Y)^k}$$

(b) Let $X_{(1)}, \ldots, X_{(n)}$ be the order statistics of a random sample $X_1, \ldots, X_n$ from a distribution that has pdf

$$f_X(x|\beta) = \frac{1}{\beta} e^{-x/\beta}, \ \ 0 < x < \infty, \quad 0 < \beta < \infty.$$

Show that $T = \sum_{i=1}^n X_{(i)}$ is complete sufficient and $nX_{(1)}/T$ is ancillary for $\beta$.

# Problem 5 (cont'd)

- Claim $f_X(x|\beta)$ belongs to the exponential family:

$$f_X(x|\beta) = h(x)c(\beta)\exp\{w_1(\beta)t_1(x)\}.$$

  One can let $h(x) = 1$, $c(\beta) = 1/\beta$, $w_1(\beta) = -1/\beta$, and $t_1(x) = x$. Hence we can conclude $\sum_{i=1}^n t_1(X_i) = \sum_{i=1}^n X_i = \sum_{i=1}^n X_{(i)}$ is a complete sufficient statistic.

- Let $Z_i = X_i/\beta$. One can show $f_Z(z) = e^{-z}$, which does not have any information of $\beta$. One can write

$$\frac{nX_1}{T} = \frac{n\beta Z_{(1)}}{\sum_{i=1}^n \beta Z_{(i)}} = \frac{nZ_{(1)}}{\sum_{i=1}^n Z_{(i)}},$$

  which does not have any information of $\beta$ since it is a function of $Z_1, \cdots, Z_n$. One hence can claim $nX_{(1)}/T$ is ancillary for $\beta$.

# Problem 5 (cont'd)

(c) Use the result in (a) and Basu's Theorem to determine $E(R)$, where $R = nX_{(1)}/T$.

By Basu's Theorem, $T$ and $R$ are independent. One can have, by the result in (a),

$$E(nX_{(1)}) = E\left(\frac{nX_{(1)}}{T}T\right) = E(R)E(T).$$

One can show $E(nX_{(1)}) = \beta$ and $E(T) = E(\sum_{i=1}^{n} X_i) = n\beta$. We have $E(R) = 1/n$.

$$f_{Z_{(1)}}(z_{(1)}) = \frac{n!}{1!(n-1)!}e^{-z_{(1)}}(e^{-z_{(1)}})^{n-1} = ne^{-nz_{(1)}}.$$

$Z_{(1)}$ follows $\text{Exp}(1/n)$ and $E(Z_{(1)}) = 1/n$.