Ma 494 — Theoretical Statistics

Solutions for Final — May 10, 2010

Prof. Sawyer — Washington University

Seven problem solutions on four pages. Closed book and closed notes. Two $8\frac{1}{2} \times 11$ sheets of paper with notes on both sides and a calculator are allowed. Parts of a problem may be weighted according to difficulty.

1. The likelihood is

$$L(\theta, X) \;=\; \prod_{i=1}^{n} \Big(\theta e^{-\theta e^{X_i}} e^{X_i}\Big) \;=\; \theta^n \exp\left(-\theta \sum_{i=1}^{n} e^{X_i}\right) \exp\left(-\sum_{i=1}^{n} X_i\right)$$

The condition for a sufficient statistic is $L(\theta, X) = g\big(S(X), \theta\big) h(X)$ where $h(X)$ depends only on X. The function $S(X) = \sum_{i=1}^{n} e^{X_i}$ satisfies this condition with $g(S, \theta) = \theta^n e^{-\theta S}$.

2. Note $P(Y_i = j) = (1-p)^{j-1} p$ and

$$\begin{aligned} P(Y_i \ge j) \;&=\; \sum_{i=j}^{\infty} (1-p)^{i-1} p \;=\; (1-p)^{j-1} \sum_{i=0}^{\infty} (1-p)^i p \\ &=\; (1-p)^{j-1} \frac{1}{1-(1-p)}\, p \;=\; (1-p)^{j-1} \end{aligned}$$

so that $P(Y_i \ge 6) = (1-p)^5$. Thus the likelihood in terms of $K_1, \ldots, K_5, R_6$ given p is

$$\begin{aligned} L(p, K, R) \;&=\; \left(\prod_{j=1}^{5} \big((1-p)^{j-1} p\big)^{K_j}\right) \big((1-p)^5\big)^{R_6} \\ &=\; p^{\sum_{j=1}^{5} K_j} (1-p)^{\sum_{j=1}^{5} (j-1) K_j + 5R_6} \end{aligned}$$

Since the maximum of $p^A (1-p)^B$ for $A, B > 0$ and $0 < p < 1$ is attained at $p = A/(A+B)$, the MLE is

$$\begin{aligned} \widehat{p} \;&=\; \frac{\sum_{j=1}^{5} K_j}{\sum_{j=1}^{5} K_j + \sum_{j=1}^{5} (j-1) K_j + 5R_6} \;=\; \frac{\sum_{j=1}^{5} K_j}{\sum_{j=1}^{5} j K_j + 5R_6} \\ &=\; \frac{8+7+4+6+5}{8+14+12+24+25+50} \;=\; \frac{30}{133} \;=\; 0.2256 \end{aligned}$$

3. The likelihood is

$$\begin{aligned} L(\theta, X) &= \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(X_i - \mu)^2 \right) \right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i^2 - 2\mu X_i + \mu^2) \right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^{n} X_i^2 - 2\mu \sum_{i=1}^{n} X_i + n\mu^2 \right) \right) \\ &= g\left(\sum_{i=1}^{n} X_i^2,\ \sum_{i=1}^{n} X_i,\ \mu, \sigma^2 \right) \end{aligned}$$

for an appropriate function $g(T_1, T_2, \mu, \sigma^2)$. Thus $T(X) = \big(T_1(X), T_2(X)\big)$ is a sufficient statistic for $\theta = (\mu, \sigma^2)$.

4. (ia) The likelihood is

$$L(\theta, Y) = \prod_{j=1}^{n} \Big(\theta/(Y_j + 1)^{\theta+1} \Big) = \frac{\theta^n}{\Big(\prod_{j=1}^{n} (Y_j + 1) \Big)^{\theta+1}}$$

Thus

$$\begin{aligned} \log L(\theta, Y) &= n \log \theta - (\theta + 1) \log \left(\prod_{j=1}^{n} (Y_j + 1) \right) \\ \frac{\partial}{\partial \theta} \log L(\theta, Y) &= \frac{n}{\theta} - \log \left(\prod_{j=1}^{n} (Y_j + 1) \right) = \frac{n}{\theta} - \sum_{j=1}^{n} \log(Y_j + 1) \end{aligned}$$

Setting the last expression equal to zero yields

$$\widehat{\theta} = \frac{n}{\sum_{i=1}^{n} \log(Y_j + 1)}$$

(ib) For $f(y, \theta) = \theta/(Y_j + 1)^{\theta+1}$, we have as above

$$\begin{aligned} \log f(Y_j, \theta) &= \log \theta - (\theta + 1) \log(Y_j + 1) \\ \frac{\partial}{\partial \theta} \log f(Y_j, \theta) &= \frac{1}{\theta} - \log(Y_j + 1) \qquad \text{and} \\ \frac{\partial^2}{\partial \theta^2} \log f(Y_j, \theta) &= -\frac{1}{\theta^2} \end{aligned}$$

Since the Fisher information $I(f,\theta)$ is minus the expected value of the last expression, $I(f,\theta) = 1/\theta^2$.

(ii) The asymptotic (central) 95% confidence interval in general is

$$\left(\widehat{\theta} - \frac{1.96}{\sqrt{nI(f,\widehat{\theta})}},\ \widehat{\theta} + \frac{1.96}{\sqrt{nI(f,\widehat{\theta})}} \right)$$

which, since $I(f,\widehat{\theta}) = 1/\widehat{\theta}^2$, equals

$$\left(\widehat{\theta} - \frac{1.96\,\widehat{\theta}}{\sqrt{n}},\ \widehat{\theta} + \frac{1.96\,\widehat{\theta}}{\sqrt{n}} \right)$$

5. (i) The power is the probability of accepting H_1 when H_1 is true, which is

$$P(X \in \mathcal{C} \mid H_1) \ = \ \int_{\mathcal{C}} g(\widetilde{x})\, d\widetilde{x}$$

where $\widetilde{x} = (x_1, \ldots, x_n) \in R^n$, $g(\widetilde{x})$ is the n-dimensional density $g(x_1)g(x_2)\ldots g(x_n)$, and the integral is n-dimensional. Similarly, the level of significance is the probability of accepting H_1 (or rejecting H_0) when H_0 is true, which is

$$P(X \in \mathcal{C} \mid H_0) \ = \ \int_{\mathcal{C}} f(\widetilde{x})\, d\widetilde{x}$$

(ii) The Neyman-Pearson Lemma says that, for any pair of densities $f(x)$ and $g(x)$, among set $\mathcal{C} \subseteq R^n$ with $P(X \in \mathcal{C} \mid H_0) = \alpha$ fixed, the set $\mathcal{C}$ with the largest power is

$$\mathcal{C}(\lambda) = \left\{ \widetilde{x} : \frac{g(\widetilde{x})}{f(\widetilde{x})} \ge \lambda \right\}$$

where λ is chosen so that $P(X \in \mathcal{C}(\lambda) \mid H_0) = \int_{\mathcal{C}} f(\widetilde{x})\, d\widetilde{x} = \alpha$.

6. Since $\mathrm{Var}(X) = E(X^2) - E(X)^2$ for any random variable X, we have

$$E\big((X-\theta)^2\big) \ = \ \mathrm{Var}(X) \ + \ (E(X) - \theta)^2$$

Thus $R(T_B,\theta) = E\big((T_B - \theta)^2\big) = \mathrm{Var}(T_B) + Q^2$ where $Q = |E_\theta(T_B) - \theta|$. Thus $R(T_B,\theta) = \mathrm{Var}(T_B) + Q^2 = 0.80\,\mathrm{Var}(T_U) + Q^2 \le \mathrm{Var}(T_U)$ if and only if $Q^2 \le (1 - 0.80)\,\mathrm{Var}(T_U) = 0.20\,\mathrm{Var}(T_U)$, or if and only if

$$Q = |E_\theta(T_B) - \theta| \ \le \ \sqrt{0.20}\sqrt{\mathrm{Var}(T_U)} \ = \ 0.4472\sqrt{\mathrm{Var}(T_U)}$$

Thus T_B will have smaller (quadratic-loss) risk as long as its bias Q is less than 0.4472 times the standard deviation of T_U.

7. (i) By the Neyman-Pearson lemma, the most powerful test is

$$\begin{aligned}\mathcal{C}(\lambda) &= \left\{ \widetilde{x} : \frac{g(\widetilde{x}, r)}{g(\widetilde{x}, 1)} \geq \lambda \right\} = \left\{ \widetilde{x} : C(r) \frac{\prod_{j=1}^n x_j^{r-1} e^{-x_j}}{\prod_{j=1}^n e^{-x_j}} \geq \lambda \right\} \\ &= \left\{ \widetilde{x} : C(r) \Big(\prod_{j=1}^n x_j \Big)^{r-1} \geq \lambda \right\} = \left\{ \widetilde{x} : \prod_{j=1}^n x_j \geq \lambda_r \right\}\end{aligned}$$

where $\lambda_r = (\lambda / C(r))^{1/(r-1)}$. The constant λ_r is determined by $P(X \in \mathcal{C}(\lambda) \mid H_0) = \alpha$. For part (i), $r = 3$.

(ii) The set $\mathcal{C}$ is determined by the last set $\mathcal{C}$ above, where λ_r is determined by α. In particular, since λ_r depends only on α and H_0, the constant λ_r is the same for all $r > 1$. Thus the most powerful test does not depend on r as long as $r > 1$. Hence the test is UMP for all $r > 1$.