STA 215: Final Exam

Time: 3 hours

Name: Solution Key

Qn	1(a)	1(b)	1(c)	1(d)	2	3(a)	3(b)	4	5(a)	5(b)	Total
Points											
Max	3	3	3	5	4	4	5	6	4	3	40

1. Let $X = (X_1, \cdots, X_n)$ with $X_i \overset{\text{IID}}{\sim} \mathcal{P}(\mu)$ (pmf: $e^{-\mu}\mu^{x_i}/x_i!$ for $x_i = 0, 1, \cdots$) with $\mu > 0$ unknown.

   (a) Show that the Fisher information is given by $I^F(\mu) = n/\mu$. [3 points]

   We have

   $$\log p(x|\theta) = \text{const.} - n\mu + n\bar{x}\log\mu$$

   and so $\frac{\partial^2}{\partial\theta^2}\log p(x|\theta) = -n\bar{x}/\mu^2$. Therefore, $I^F(\mu) = -\mathrm{E}_\mu[\frac{\partial^2}{\partial\theta^2}\log p(X|\theta)] = n/\mu$.

   (b) Is $\hat{\mu}_{\text{MLE}} = \bar{X}$ a uniformly minimum variance unbiased estimator (UMVUE) of μ? Explain. [3 points]

   Yes. $\hat{\mu}_{\text{MLE}} = \bar{X}$ is unbiased for μ, so the information bound on variance is given by $I^F(\mu)^{-1} = \mu/n$. But, $\mathrm{Var}_\mu\bar{X} = \mathrm{Var}_\mu X_1/n = \mu/n$. So $\bar{X}$ is UMVUE for this model.

   (c) Identify the posterior distribution of μ under the Jeffreys prior $\pi_J(\mu) \propto \mu^{-1/2}$ (give name of the distribution and exact expressions for parameters). [3 points]

   The posterior pdf of μ under the prior $\pi_J(\mu) \propto \mu^{-1/2}$ is given by

   $$\pi_J(\mu|x) \propto p(x|\theta)\pi_J(\theta) \propto e^{-n\mu}\mu^{n\bar{x}}\mu^{-1/2} \propto \mu^{n\bar{x}+\frac{1}{2}-1}e^{-n\mu}$$

   which matches the expression of the $\mathcal{G}a(n\bar{x} + 1/2, n)$ pdf in μ.

   (d) Derive the Wald and Rao test statistics for testing $H_0 : \mu = \mu_0$ vs. $\mu \neq \mu_0$. For this problem, is one practically more useful than the other? Explain. [5 points]

   The Wald test statistics is

   $$W = (\hat{\mu}_{\text{MLE}} - \mu_0)'\{I^F(\hat{\mu}_{\text{MLE}})\}(\hat{\mu}_{\text{MLE}} - \mu_0) = n\frac{(\bar{X} - \mu_0)^2}{\bar{X}}$$

   and the Rao test statistics is

   $$S = \dot{\ell}_X(\mu_0)'\{I^F(\mu_0)\}^{-1}\dot{\ell}_X(\mu_0) = n\frac{(\bar{X} - \mu_0)^2}{\mu_0}$$

   The Rao test statistic appears more useful because it is defined for all observable data X, whereas the Wald statistic is undefined if all X_i's are observed to be zero.

2. Consider $X = (X_1, \cdots, X_n)$ with X_i's independently distributed as $X_i \sim \mathcal{N}(\mu_i, 1)$. For testing $H_0 : \mu_1 = \mu_2 = \cdots = \mu_n = 0$ vs. H_1 : at least one $\mu_i \neq 0$, let δ be a test that rejects H_0 if any $|X_i|$ exceeds $q_{\mathcal{N}}(1 - \alpha/2)$. Is δ a level-α test? Justify your answer. [4 points]

   No. The size of δ is:

$$P_0(\text{reject } H_0) = 1 - \prod_{i=1}^{n} P_0(|X_i| \leq q_{\mathcal{N}}(1 - \alpha/2)) = 1 - (1 - \alpha)^n$$

   which equals α only for $n = 1$ and exceeds α for $n > 1$.

3. Consider a statistical model $X \sim p(x|\theta)$, $\theta \in \Theta$ and suppose we want to test $H_0 : \theta \in \Theta_0$ vs. $\theta \in \Theta_1$ where Θ_0 and Θ_1 form a partition of Θ. Let $T = T(X)$ be a statistic such that large values of $|T|$ provide evidence against H_0.

   (a) Suppose $S = S(X)$ is another statistic such that the distribution of T given S is the same for all $\theta \in \Theta_0$. Denote the common cumulative distribution function of T given $S = s$, under every $\theta \in \Theta_0$, by $F(t|s)$, $-\infty < t < \infty$. As usual, let $q_F(u, s)$ give the quantiles of $F(t|s)$ for $u \in (0, 1)$. Show that for any $\alpha \in (0, 1)$, the test that rejects H_0 for $T > q_F(1 - \alpha/2, S)$ or $T < q_F(\alpha/2, S)$ is level-α. [4 points]
   For any $\theta \in \Theta_0$ and any s in the range of S,

$$\begin{aligned} P_\theta(\text{reject } H_0|S = s) &= P_\theta(\{T < q_F(\alpha/2, s)\} \cup \{T > q_F(1 - \alpha/2, s)\} \mid S = s) \\ &= F(q_F(\alpha/2, s)|s) + 1 - F(q_F(1 - \alpha/2, s)|s) \\ &= \alpha/2 + 1 - (1 - \alpha/2) = \alpha \end{aligned}$$

   Since this conditional probability does not depend on the conditioning event $S = s$, we have, for any $\theta \in \Theta_0$,

$$P_\theta(\text{reject } H_0) = \alpha.$$

   Since this probability equals α for all $\theta \in \Theta_0$, the size of the associated test is precisely α.

   (b) Consider $X = (X_1, X_2)$ where X_1 and X_2 are modeled as independent exponential random variables with parameters $\lambda_1 > 0$ and $\lambda_2 > 0$ (i.e., for $i = 1, 2$, the pdf of X_i is $\lambda_i \exp(-\lambda_i x_i)$, $x_i > 0$). A simple calculation shows that the joint pdf of $S = X_1 + X_2$ and $T = X_1 - X_2$ is given by,

$$f(t, s|\lambda_1, \lambda_2) = \frac{\lambda_1 \lambda_2}{2} \exp\left\{-\frac{(\lambda_1 + \lambda_2)s + (\lambda_1 - \lambda_2)t}{2}\right\}, \quad s > 0, -s < t < s,$$

   and $f(t, s|\lambda_1, \lambda_2) = 0$ otherwise. Argue that for testing $H_0 : \lambda_1 = \lambda_2$ against $H_1 : \lambda_1 \neq \lambda_2$, a level-$\alpha$ test is given by "Reject H_0 if $|T| > (1 - \alpha)S$". [5 points]
   Clearly, for any $\lambda_1 = \lambda_2$, the conditional distribution of T given $S = s$ is the uniform distribution over $(-s, s)$, which does not depend on the particular values of λ_1, λ_2. Therefore part (a) applies with $q_F(\alpha/2, s) = -(1 - \alpha)s$ and $q_F(1-\alpha/2, s) = (1-\alpha)s$. So a level-$\alpha$ test is given by "reject H_0 if $|T| > (1-\alpha)S$".

4. Consider again $X = (X_1, X_2)$ where X_1 and X_2 are modeled as independent exponential random variables with parameters $\lambda_1 > 0$ and $\lambda_2 > 0$. Describe the level-α maximum likelihood test (a.k.a., the likelihood ratio test) for testing $H_0 : \lambda_1 = \lambda_2$ against $H_1 : \lambda_1 \neq \lambda_2$ (write down your test as "reject H_0 if $T > c$", give a simple expression for T and identify c in terms of α and possibly a quantile of a named distribution with exact expressions for its parameters). [Hint: $u(1-u)$ is small if $|u - 1/2|$ is large.] [6 points]

The log-likelihood function is:

$$\ell_X(\lambda_1, \lambda_2) = \log \lambda_1 - \lambda_1 X_1 + \log \lambda_2 - \lambda_2 X_2.$$

The unrestricted maxima is attained at $\hat{\lambda}_{\text{MLE},1} = 1/X_1$, $\hat{\lambda}_{\text{MLE},2} = 1/X_2$. Under H_0, the maxima is attained at $\hat{\lambda}_{H_0,1} = \hat{\lambda}_{H_0,2} = 2/(X_1 + X_2)$. So

$$\begin{aligned}
2 \log \Lambda(X) &= \ell_X(\hat{\lambda}_{\text{MLE},1}, \hat{\lambda}_{\text{MLE},2}) - \ell_x(\hat{\lambda}_{H_0,1}, \hat{\lambda}_{H_0,2}) \\
&= -\log X_1 - 1 - \log X_2 - 1 + 2\log(X_1 + X_2) - 2\log 2 + 2 \\
&= -\log \left\{ \frac{X_1}{X_1 + X_2} \left(1 - \frac{X_1}{X_1 + X_2}\right) \right\} - 2 \log 2.
\end{aligned}$$

The ML test rejects H_0 if $2 \log \Lambda(X) > k$ for some constant k which is same as rejecting H_0 if $\frac{X_1}{X_1+X_2}(1 - \frac{X_1}{X_1+X_2}) < k'$ for some constant k', which is same as rejecting H_0 if $|\frac{X_1}{X_1+X_2} - \frac{1}{2}|$ is larger than some constant c. Now, under H_0, $\frac{X_1}{X_1+X_2} \sim \mathcal{B}e(1,1) = \mathcal{U}(0,1)$ and so $\frac{X_1}{X_1+X_2} - \frac{1}{2} \sim \mathcal{U}(-1/2, 1/2)$. So to make the test level-α we must take $c = 1/2 - \alpha/2$. [You may notice that this is the same test as in 3(b)]

5. A researcher has recruited n subjects for a study with two treatment conditions C1 and C2. She models the outcome as $Y_i = z_i'\beta + \epsilon_i$, $\epsilon_i \overset{\text{IID}}{\sim} \mathcal{N}(0, \sigma^2)$ with $z_i = (A_i, C_{1i}, C_{2i})$ where A_i is the age of subject i, $C_{1i} = 1, C_{2i} = 0$ if subject i is assigned to C1, $C_{1i} = 0, C_{2i} = 1$ otherwise. Letting Z denote the design matrix, it can be shown that

$$Z'Z = \begin{pmatrix} \sum_i A_i^2 & n_1 \bar{A}_1 & n_2 \bar{A}_2 \\ n_1 \bar{A}_1 & n_1 & 0 \\ n_2 \bar{A}_2 & 0 & n_2 \end{pmatrix},$$

$$(Z'Z)^{-1} = \frac{1}{n_1 n_2 (n_1 s_1^2 + n_2 s_2^2)} \begin{pmatrix} n_1 n_2 & -n_1 n_2 \bar{A}_1 & -n_1 n_2 \bar{A}_2 \\ -n_1 n_2 \bar{A}_1 & n_2 \sum_i A_i^2 - n_2^2 \bar{A}_2^2 & n_1 n_2 \bar{A}_1 \bar{A}_2 \\ -n_1 n_2 \bar{A}_2 & n_1 n_2 \bar{A}_1 \bar{A}_2 & n_1 \sum_i A_i^2 - n_1^2 \bar{A}_1^2 \end{pmatrix}$$

where n_i, $\bar{A}_i$ and s_i^2 denote the number, average age and variance of age of subjects assigned to Ci, $i = 1, 2$. For this problem, variance is defined as follows: the variance of $x_1, \cdots, x_k$ is $\frac{1}{k}\sum_{i=1}^k (x_i - \bar{x})^2 = \frac{1}{k}\sum_i x_i^2 - \bar{x}^2$.

(a) Assume σ^2 is known and consider the flat prior $\pi(\beta) \propto 1$. What is the posterior variance of the treatment contrast $\beta_3 - \beta_2$? Simplify. [4 points]

The posterior variance of $\beta_3 - \beta_2$ is $\sigma^2 a'(Z'Z)^{-1}a$ where $a = (0, -1, 1)'$. This equals

$$\begin{aligned}
\frac{\sigma^2}{n_1 n_2 (n_1 s_1^2 + n_2 s_2^2)} &\{n_2 \sum_i A_i^2 - n_2^2 \bar{A}_2^2 + n_1 \sum_i A_i^2 - n_1^2 \bar{A}_1^2 - 2 n_1 n_2 \bar{A}_1 \bar{A}_2\} \\
&= \frac{\sigma^2}{n_1 n_2 (n_1 s_1^2 + n_2 s_2^2)} \{n \sum_i A_i^2 - (n_1 \bar{A}_1 + n_2 \bar{A}_2)^2\} \\
&= \frac{\sigma^2}{n_1 n_2 (n_1 s_1^2 + n_2 s_2^2)} \{n \sum_i A_i^2 - (n \bar{A})^2\} \\
&= n^2 \frac{\sigma^2 s^2}{n_1 n_2 (n_1 s_1^2 + n_2 s_2^2)}
\end{aligned}$$

where n, $\bar{A}$ and s^2 are the number, average age and variance of age of all subjects.

(b) Suppose $n = 2m$ and the researcher decides to assign m subjects to each group. Further suppose she has the age records available to her before she makes this assignment. Can she do any better job than just randomly splitting them into two equal halves? Explain (I don't need a mathematical proof. A careful reasoning would do!). [3 points]

When $n_1 = n_2 = m$, the variance of $\beta_3 - \beta_2$ equals $\frac{4\sigma^2 s^2}{s_1^2 + s_2^2}$. Because $\beta_3 - \beta_2$ is the quantity of importance, it makes sense to allocate subjects so that this variance is small. The numerator does not depend on the allocation, but the denominator can be large by carefully dividing the subjects so that the age variance in each group is large.