STAC58S: 2014 Final Exam

Name:

Student Number:

Time: 3 hours
Instructions: Write your answers on the exam paper. You do not have to fully evaluate numerical results for questions that ask for these as full marks will be given when it is clear that you have the correct expression.
Aids: The exam is open book. Students may use any notes, books and calculators in writing this exam. You are free to refer to any results derived in class or on the assignments.

1. (a) A researcher is interested in grade point averages for students enrolled in a Specialist program in the Department of Computer and Mathematical Sciences and in particular, assessing whether or not this is dependent on whether or not a student works part-time.

(a) (5 marks) For a statistical study of this problem identify the population, the response measurement, the predictor measurement and the relevant conditional distributions.

Π = set of all students enrolled in a Specialist program in this department

$Y(\pi)$ = response = grade point average of $\pi \in \Pi$

$X(\pi)$ = predictor $= \begin{cases} 1 & \pi \text{ works part-time} \\ 0 & \text{otherwise} \end{cases}$

$f_Y(\cdot \mid X=1)$ = conditional distribution of GPA given works part-time

$f_Y(\cdot \mid X=0)$ = " " " " given doesn't

(b) (5 marks) Is it possible to definitively answer the question as to whether or not there is a relationship between the variables? If so describe how you would do this.

Yes by conducting a census and computing $f_Y(\cdot \mid X=1)$, $f_Y(\cdot \mid X=2)$ and comparing them. If they are different there is a relationship.

(c) (5 marks) An apparent paradox arises in probability and statistics when we use continuous probability distributions to model a response as the probability of any specific point occurring is 0 but we do in fact observe data. Explain what the role is for continuous probability distributions when modeling responses.

We use continuous probability distributions to approximate distributions of quantities that are (finitely) discrete. This approximation is used to simplify an analysis.

2. Suppose we have a sample $x = (x_1, \ldots, x_n)$ from a $N(\mu, \sigma^2)$ distribution where $\mu \in R^1$ and $\sigma > 0$. In the following you can use any results and calculations established in class.

(a) (5 marks) What is the MLE of $\psi = (\mu + \sigma z_{0.25}, \mu + \sigma z_{0.75})$ where z_p is the p-th quantile of the $N(\mu, \sigma^2)$ distribution? Justify your answer.

The MLE is $\left(\bar{x} + \sqrt{\frac{(n-1)s_x^2}{n}}\, z_{0.25},\ \bar{x} + \sqrt{\frac{(n-1)s_x^2}{n}}\, z_{0.75}\right)$

since $\left(\bar{x}, \frac{(n-1)s_x^2}{n}\right)$ is the MLE of (μ, σ^2) and ψ is a 1-1 fn of (μ, σ^2).

(b) (5 marks) What is the profile MLE of the interquartile range of the $N(\mu, \sigma^2)$ distribution? Justify your answer.

We know that the profile MLE is determined by "plugging in" the MLE into the function $\psi = \Psi(\mu, \sigma^2) = \sigma(z_{0.75} - z_{0.25})$ and so the profile MLE is $s(z_{0.75} - z_{0.25})$.

(c) (10 marks) Determine the form of the .95-profile likelihood region for μ. Explain how this region is to be used.

The joint likelihood is $(\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2\} \exp\{-\frac{(n-1)}{2\sigma^2}s^2\}$

$= (\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{1}{2\sigma^2}[n(\bar{x}-\mu)^2 + (n-1)s^2]\}$. For fixed μ we max over σ^2 to obtain the profile likelihood for μ. So taking logs and differentiating wrt σ^2

$$\left(-\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}[\ ]\right)' = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}[\ ] = 0$$

or $\hat{\sigma}^2 = \frac{1}{n}[\ ] = (\bar{x}-\mu)^2 + \frac{1}{n}\frac{(n-1)}{n}s^2$ which implies that $L_{\mathbb{I}}(\mu \mid x) = (\hat{\sigma}^2)^{-\frac{n}{2}} \exp\{-\frac{1}{2\hat{\sigma}^2}[\ ]\}$

$= e^{-\frac{n}{2}}\left((\bar{x}-\mu)^2 + \frac{1}{n}\left(\frac{n-1}{n}\right)s^2\right)^{-\frac{n}{2}}$ which is maximized at $\mu = \bar{x}$.

So 0.95-profile likelihood region for μ is

$$C_{0.95}(x) = \{\mu : \cancel{e^{-\frac{n}{2}}}\left((\bar{x}-\mu)^2 + \frac{1}{n}\left(\frac{n-1}{n}\right)s^2\right)^{-\frac{n}{2}} \geq (1-\gamma)e^{-\frac{n}{2}}\left(\frac{1}{n}\left(\frac{n-1}{n}\right)s^2\right)^{-\frac{n}{2}}$$

$$= \{\mu : (\bar{x}-\mu)^2 + \frac{1}{n}\left(\frac{n-1}{n}\right)s^2 \leq (1-\gamma)^{2/n}\left(\frac{1}{n}\left(\frac{n-1}{n}\right)s^2\right)^{-\frac{n}{2}}\}$$

$$= \{\mu : (\bar{x}-\mu)^2 \leq [(1-\gamma)^{2/n} - 1]\left(\frac{1}{n}\left(\frac{n-1}{n}\right)s^2\right)^{-\frac{n}{2}}\}$$

$$= \bar{x} \pm [(1-\gamma)^{2/n} - 1]^{1/2}\left(\frac{1}{n}\left(\frac{n-1}{n}\right)s^2\right)^{-\frac{n}{4}}$$

3. (a) (5 marks) Suppose that $x = (x_1, \ldots, x_n)$ is a sample from f_θ where $\theta \in \Theta$. Prove that the order statistic $(x_{(1)}, \ldots, x_{(n)})$ is a sufficient statistic for this model.

The joint density is $f_\theta(x_1) \cdots f_\theta(x_n)$ $= f_\theta(x_{(1)}) \cdots f_\theta(x_{(n)})$. Therefore by the Factorization Thm $(x_{(1)}, \ldots, x_{(n)})$ is sufficient.

(b) (5 marks) Suppose that $x = (x_1, \ldots, x_n)$ is a sample from a Geometric(θ) distribution where $\theta \in (0, 1]$ is unknown. Determine a minimal sufficient statistic for this model.

The likelihood is $L(\theta \mid x) = \prod_{i=1}^{n} (1-\theta)^{x_i} \theta$

$= (1-\theta)^{n\bar{x}} \theta^n$ and so $\bar{x}$ is sufficient by Factorization. We have

$(\log L(\theta \mid x))' = (n\bar{x} \log(1-\theta) + n \log \theta)'$

$= -\frac{n\bar{x}}{1-\theta} + \frac{n}{\theta} = 0$ or $\theta_{MLE}(x) = \frac{1}{1+\bar{x}}$

and we can compute $\bar{x}$ from the likelihood so it's a MSS.

4. Suppose we have two unbiased estimators T_1 and T_2 of $\psi(\theta) \in R^1$.

(a) (5 marks) Show that $\alpha T_1 + (1-\alpha)T_2$ is also an unbiased estimator of $\psi(\theta)$ whenever $\alpha \in [0,1]$.

$$E_\theta(\alpha T_1 + (1-\alpha)T_2) = \alpha E_\theta(T_1) + (1-\alpha) E_\theta(T_2)$$

$$= \alpha\psi(\theta) + (1-\alpha)\psi(\theta) = \psi(\theta) \quad \forall\, \theta \in \Theta,$$

So $\alpha T_1 + (1-\alpha)T_2$ is unbiased.

(b) (5 marks) If T_1 and T_2 are also independent, e.g., determined from independent samples, then calculate $\mathrm{Var}_\theta(\alpha T_1 + (1-\alpha)T_2)$ in terms of $\mathrm{Var}_\theta(T_1)$ and $\mathrm{Var}_\theta(T_2)$.

$$\mathrm{Var}_\theta(\alpha T_1 + (1-\alpha)T_2) = \alpha^2 \mathrm{Var}_\theta(T_1) + (1-\alpha)^2 \mathrm{Var}_\theta(T_2)$$

(c) (5 marks) For the situation in part (b), determine the best choice of α in the sense that for this choice $\mathrm{Var}_\theta(\alpha T_1 + (1-\alpha)T_2)$ is smallest. What is the effect on this combined estimator of T_1 having a very large variance relative to T_2?

$$\frac{d}{d\alpha}\left(\alpha^2 \mathrm{Var}_\theta(T_1) + (1-\alpha)^2 \mathrm{Var}_\theta(T_2)\right)$$

$$= 2\alpha \mathrm{Var}_\theta(T_1) - 2(1-\alpha)\mathrm{Var}_\theta(T_2)$$

$$= 2\alpha\left(\mathrm{Var}_\theta(T_1) + \mathrm{Var}_\theta(T_2)\right) - 2\mathrm{Var}_\theta(T_2) = 0$$

$$\hat{\alpha} = \mathrm{Var}_\theta(T_2) / \left(\mathrm{Var}_\theta(T_1) + \mathrm{Var}_\theta(T_2)\right)$$

when $\mathrm{Var}_\theta(T_1) >> \mathrm{Var}_\theta(T_2)$ the estimator is approximately equal to T_2.

(d) (5 marks) Repeat parts (b) and (c), but now do not assume that T_1 and T_2 are independent, so $\mathrm{Var}_\theta(\alpha T_1 + (1-\alpha)T_2)$ will also involve $\mathrm{Cov}_\theta(T_1, T_2)$.

$$\mathrm{Var}_\theta(\alpha T_1 + (1-\alpha)T_2) = \alpha^2 \mathrm{Var}_\theta(T_1) + 2\alpha(1-\alpha)\mathrm{Cov}_\theta(T_1,T_2) + (1-\alpha)^2 \mathrm{Var}_\theta(T_2)$$

$$\frac{d}{d\alpha}(\ ) = 2\alpha \mathrm{Var}_\theta(T_1) + 2(1-2\alpha)\mathrm{Cov}_\theta(T_1,T_2) - 2(1-\alpha)\mathrm{Var}_\theta(T_2)$$

$$= 2\alpha\left[\mathrm{Var}_\theta(T_1) - 2\mathrm{Cov}_\theta(T_1,T_2) + \mathrm{Var}_\theta(T_2)\right] + 2\left[\mathrm{Cov}_\theta(T_1,T_2) - \mathrm{Var}_\theta(T_2)\right]$$

$$\hat{\alpha} = \frac{\mathrm{Var}_\theta(T_2) - \mathrm{Cov}_\theta(T_1,T_2)}{\mathrm{Var}_\theta(T_1) - 2\mathrm{Cov}_\theta(T_1,T_2) + \mathrm{Var}_\theta(T_2)}$$ which again

goes to 0 as $\mathrm{Var}_\theta(T_1) \to \infty$ provided $\mathrm{Cov}_\theta(T_1,T_2)/\mathrm{Var}_\theta(T_1) \to 0$ as well

(e) (5 marks) Discuss the motivation for considering the concept of unbiased estimators and their variances.

We are looking for an estimator of $\psi = \Psi(\theta)$ that minimizes $MSE_\theta(T) = E_\theta(T - \Psi(\theta))^2$ for each $\theta \in \Theta$. We proved in class that no such estimator exists. So we need to restrict the class of estimators we consider and one such restriction is to require that the estimator be unbiased, namely,

$E_\theta(T) = \Psi(\theta)$ for every $\theta \in \Theta$.

When we do this $MSE_\theta(T) = Var_\theta(T)$, and so we are minimizing variance.

5. Suppose that a statistical model is given by the two distributions in the following table.

	$s=1$	$s=2$	$s=3$	$s=4$
$f_a(s)$	$1/3$	$1/6$	$1/12$	$5/12$
$f_b(s)$	$1/2$	$1/4$	$1/6$	$1/12$

(a) (5 marks) Determine the UMP size 0.10 test for testing $H_0 : \theta = a$ versus $H_a : \theta = b$. What is your conclusion when you observe $s = 3$?

$f_b(s)/f_a(s)$ = $\underline{s=1}$ $3/2$, $\underline{s=2}$ $3/2$, $\underline{s=3}$ 2, $\underline{s=4}$ $1/5$

$P_a(f_b(s)/f_a(s) > k) = \frac{1}{12}, \frac{1}{12}, 0, 7/12$

$\therefore$ take $k = 3/2$ and $\gamma = (1/10 - 1/12)/1/2 = 1/30$

when $s = 3$ we reject H_0.

(b) (5 marks) What is the power of the test determined in (a)?

$P_b(f_b(s)/f_a(s) > \frac{3}{2}) + \frac{1}{30} P_b(f_b(s)/f_a(s) = \frac{3}{2})$

$= \frac{1}{6} + \frac{1}{30} \cdot \frac{3}{4}$

6. Suppose that the statistical model is given by

	$f_\theta(1)$	$f_\theta(2)$	$f_\theta(3)$	$f_\theta(4)$
$\theta = a$	1/3	1/6	1/3	1/6
$\theta = b$	1/2	1/4	1/8	1/8

and that we consider the family of priors given by

	$\pi_\tau(a)$	$\pi_\tau(b)$
$\tau = 1$	1/2	1/2
$\tau = 2$	1/3	2/3

and we observe the sample $x_1 = 1, x_2 = 1, x_2 = 3$.

(a) (5 marks) If we use the maximum value of the prior predictive for the data to determine the value of τ, and hence the prior, which prior is selected here?

$m_1(1,1,3) = \frac{1}{2}\frac{1}{3}\frac{1}{3}\frac{1}{3} + \frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{8} = \frac{1}{2}\left(\frac{1}{27}+\frac{1}{32}\right) = \frac{59}{2\cdot 27\cdot 32}$

$m_2(1,1,3) = \frac{1}{3}\frac{1}{3}\frac{1}{3}\frac{1}{3} + \frac{2}{3}\frac{1}{2}\frac{1}{2}\frac{1}{8} = \frac{1}{3}\left(\frac{1}{27}+\frac{2}{32}\right) = \frac{86}{3\cdot 27\cdot 32}$

$\therefore m_2(1,1,3) < m_1(1,1,3)$ we use prior given by $\tau = 1$

(b) (5 marks) Determine the posterior of θ based on the selected prior.

$\pi(a \mid 1,1,3) \propto \frac{1}{2}\cdot\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}$, $\pi(b \mid 1,1,3) \propto \frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{8}$

$= \frac{1}{54}$ $\qquad = \frac{1}{64}$

(c) (5 marks) The procedure described in parts (a) and (b) is known as *empirical Bayes* because the prior is selected based on the data. Explain how empirical Bayes inferences violate a basic principle of inference.

When we select the prior based on the data we no longer have that $\pi_x(\theta) f_\theta(x)$ gives the joint distribution of (θ, x). As such we cannot apply the principle of conditional probability to this to obtain the posterior.

(d) (5 marks) Using the prior given by $\tau = 1$ determine the LRSE of θ.

Since the prior is uniform the LRSE is the same as the MAP estimator and from (b) part we see that this is given by $\theta_{MAP}(1,1,3) = \phi$.