MIDTERM EXAM - MATH 563, SPRING 2016

NAME: SOLUTION ______________________________

Exam rules:

There are **5** problems on this exam.
You must show **all** work to receive credit, state any theorems and definitions clearly.
The instructor will NOT answer any questions during the exam.

Problem	Points
1	5+10+10/ 25
2	5+5+5+5/ 20
3	7+7+6 / 20
4	5+15 / 20
5	15 / 15

TOTAL GRADE: **(out of 90)**

Problem 1. *[This problem is about statistics and data reduction techniques.]*

a) What does *data reduction* mean? In other words, how can a statistic $T(X)$, where X is a random sample in the sample space $\mathcal{X}$, be used for data reduction?

SOLUTION:

A statistic $T(X)$ is used for data reduction by partitioning the sample space $\mathcal{X}$ into equivalence classes $\{X : T(X) = t\}$. Any different samples in the same equivalence class take the same value of $T(X)$, and are viewed as containing the same amount of information with regards to T.

b) Define a *sufficient* and *ancillary* statistic.

SOLUTION:

Let X be a random sample from a population, whose distribution depends on a parameter θ,

- A statistic $T(X)$ is called sufficient if the conditional distribution of X given $T(X)$ does not depend on θ;
- A statistic $S(X)$ is called ancillary if its own distribution does not depend on θ.

(continued on next page)

Problem 1, continued.

Solve only ONE of the following two problems, c) OR d):

c) Let N be a random variable taking values $1, 2, \ldots$ with known probabilities $p_1, p_2, \ldots$, where $\sum_i p_i = 1$. Having observed $N = n$, perform n Bernoulli trials with success probability θ, obtaining X successes. Prove that the pair (X, N) is minimal sufficient and N is ancillary for θ.
[Hint: use a criterion, not the definition of minimal sufficient.]

d) Prove Basu's Theorem in the DISCRETE case: If $T(X)$ is a complete and minimal sufficient statistic, then $T(X)$ is independent of every ancillary statistic.
[Hint: Let $S(X)$ be any ancillary statistic, and show that $P(S(X) = s|T(X) = t) - P(S(X) = s) = 0$.]

SOLUTION of c):

This is Exercise 6.12(a) from textbook. See Solution to HW#3 for its answer.

SOLUTION of d):

Let $S(X)$ be any ancillary statistic. As is stated in the hint, to prove that $T(X)$ and $S(X)$ are independent, it suffices to show that

$$P(S(X) = s|T(X) = t) - P(S(X) = s) = 0 \tag{1}$$

Since $S(X)$ is ancillary, $P(S(X) = s)$ does not depend on θ.
Also, since $T(X)$ is sufficient, $P(S(X) = s|T(X) = t)$ does not depend on θ.
Note that

$$P(S(X) = s) = \sum_t P(S(X) = s|T(X) = t)P_\theta(T(X) = t) \tag{2}$$

Also, since $\sum_t P_\theta(T(X) = t) = 1$, we have

$$P(S(X) = s) = \sum_t P(S(X) = s)P_\theta(T(X) = t) \tag{3}$$

If we substract (3) from (2), we obtain

$$\sum_t [P(S(X) = s|T(X) = t) - P(S(X) = s)] \cdot P_\theta(T(X) = t) = 0 \tag{4}$$

Since $T(X)$ is a complete statistic, we can obtain (1) from (4). This completes the proof.

Problem 2. *[This problem is about computing point estimators.]*

1) Let $Y_1, \ldots, Y_n$ denote a random sample from the probability density function

$$f(y|\theta) = \begin{cases} (\theta+1)y^\theta, & \text{for } 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Find the method of moments estimator for θ.

SOLUTION:

We have $E(Y|\theta) = \int_{-\infty}^{\infty} y f(y|\theta) dy = \int_0^1 (\theta+1) y^{\theta+1} dy = \frac{\theta+1}{\theta+2}$.

Solving θ from $E(Y|\theta) = \bar{Y}$, we obtain $\hat{\theta}_{MoM} = \frac{1-2\bar{Y}}{\bar{Y}-1}$.

2) Let $X_1, \ldots, X_n$ be a random sample from the pdf $f_\theta(x) = \theta x^{\theta-1}$, for $x \in (0,1)$ (a beta distribution). Compute the MLE of θ.

SOLUTION:

The likelihood is

$$L(\theta|X) = \prod_{i=1}^n f(x_i|\theta) = \theta^n (x_1 \cdots x_n)^{\theta-1}$$

Thus the log likelihood is

$$l(\theta|X) = \log L(\theta|X) = n \log\theta + (\theta-1)\sum_{i=1}^n \log x_i$$

.

Thus

$$\frac{\partial l(\theta|X)}{\partial\theta} = 0 \Rightarrow \frac{n}{\theta} + \sum_{i=1}^n \log x_i = 0 \Rightarrow \hat{\theta}_{MLE} = -\frac{n}{\sum_{i=1}^n \log x_i}$$

.

It is also easy to verify $\frac{\partial^2 l(\theta|X)}{\partial\theta^2} < 0$, thus $l(\theta|X)$ attains its maximum at $\hat{\theta}_{MLE}$.

(continued on next page)

Problem 2, continued.

3) Let $X_n, \ldots, X_n$ be a random sample from $Poisson(\lambda)$. Since Gamma is a conjugate family for Poisson, let λ have a Gamma distribution. As written on the board during the exam, the X_i are meant to just be Y_i here.

a) Write down the formula you would use for computing the posterior distribution of λ, and explain the meaning of each term in your formula. *[Please do NOT write down the formulas for the densities!]*

b) **Assume** that the calculation in part a) would provide that the posterior distribution of λ is $Gamma(y+\alpha, \frac{\beta}{n\beta+1})$, with $E[\lambda|y] = (y+\alpha)\frac{\beta}{n\beta+1}$ and $Var(\lambda|y) = (y+a)\frac{\beta^2}{(n\beta+1)^2}$.
What would you use as the Bayes estimator for λ?

SOLUTION:

(a) The formula used here is

$$\pi(\lambda|y) = \frac{f(y|\lambda)\pi(\lambda)}{m(y)}$$

.

- $\pi(\lambda|y)$ is the posterior density of λ, which is the density estimated after the data is observed.
- $\pi(\lambda)$ is the prior density of λ, which is the density set before the data is collected.
- $f(y|\lambda)$ is the sampling density of data given λ.
- $m(y) = \int f(y|\lambda)\pi(\lambda)d\lambda$ is the marginal distribution of data.

(b) We can use $E(\lambda|y) = \frac{(y+\alpha)\beta}{n\beta+1}$ as the natural Bayesian estimator.

Problem 3. *[This problem is about evaluating point estimators.]*

Let $X_1, \ldots, X_n$ be a random sample from $Bernoulli(\theta)$.[1] Then, $T = \sum_{i=1}^n X_i$ is a sufficient statistic. The goal of this problem is to estimate $\eta = \theta(1-\theta)$.

a) Show that the 'naive' estimator $\tilde{\eta} = X_1(1 - X_2)$ is unbiased.

SOLUTION:

Since X_1 and X_2 are independent, so are X_1 and $1 - X_2$. We then have

$$E(\tilde{\eta}) = E[X_1(1 - X_2)] = (EX_1)(1 - EX_2) = \theta(1-\theta) = \eta$$

. Thus $\tilde{\eta}$ is unbiased.

b) Find a better estimator $\hat{\eta} = E[\tilde{\eta}|T = t]$ using the Rao-Blackwell Theorem.

(Hint: The random variable $X_1(1 - X_2)$ is either 0 or 1; it's 1 if and only if $X_1 = 1$ and $X_2 = 0$.)

SOLUTION:

According to the hint, we have

$$\begin{aligned}
\hat{\eta} =& E(\tilde{\eta}|T = t) \\
=& P(\tilde{\eta} = 0|T = t) \cdot 0 + P(\tilde{\eta} = 1|T = t) \cdot 1 \\
=& P(\tilde{\eta} = 1|T = t) \\
=& P(X_1 = 1, X_2 = 0| \sum_{i=1}^n X_i = t) \\
=& \frac{P(X_1 = 1, X_2 = 0, \sum_{i=1}^n X_i = t)}{P(\sum_{i=1}^n X_i = t)} \\
=& \frac{P(X_1 = 1)P(X_2 = 0)P(\sum_{i=3}^n X_i = t - 1)}{P(\sum_{i=1}^n X_i = t)} \\
=& \frac{\theta(1-\theta)\binom{n-2}{t-1}\theta^{t-1}(1-\theta)^{(n-2)-(t-1)}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \\
=& \frac{\binom{n-2}{t-1}}{\binom{n}{t}} \\
=& \frac{t(n-t)}{n(n-1)}
\end{aligned}$$

c) What is a minimum variance unbiased estimator (MVUE) of η? Could it be that $\hat{\eta}$ is an MVUE?

SOLUTION:

According to Rao-Blackwell Theorem, if $\tilde{\eta}$ is an unbiased estimator of η (which was proven in part a), and if $T(X)$ is a sufficient statistic (which is given in the problem), then $E(\tilde{\eta}|T)$ is a MVUE. Therefore, yes, $\hat{\eta}$ computed in (b) is a MVUE.

[1] Recall that $\theta = E[X_i]$ in this case.

Problem 4. *[This problem is about convergence.]*

Let $X_1, \ldots, X_n$ denote a random sample from a distribution with mean μ and variance σ^2.

a) State the Weak Law of Large Numbers. *(Reminder: this is a statement about the sample mean $\overline{X_n}$.)*

SOLUTION:

WLLN states that if $\sigma^2 < \infty$, then $\bar{X}_n \xrightarrow{P} \mu$ as $n \to \infty$. This means

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 1$$

. Also, see P232, Theorem 5.5.2 in the textbook.

b) Show that the sample variance S_n^2 is a *consistent* estimator of σ^2.

SOLUTION:

We need to assume $\text{Var}(X_i^2) < \infty$ for $i = 1, \cdots, n$.

Then by WLLN, as $n \to \infty$, we have

$$\bar{X}_n \xrightarrow{P} \mu \tag{5}$$

$$\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{P} E(X_i^2) = \mu^2 + \sigma^2. \tag{6}$$

Thus

$$\begin{aligned}
S_n^2 =& \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2 \\
=& \frac{1}{n-1}(\sum_{i=1}^n X_i^2 - n\bar{X}_n^2) \\
=& \frac{n}{n-1}(\frac{1}{n}\sum_{i=1}^n X_i^2) - \frac{n}{n-1}\bar{X}_n^2
\end{aligned}$$

From (5) and (6), and the fact that $\lim_{n\to\infty} \frac{n}{n-1} = 1$, we can conclude from Slutsky's Theorem that as $n \to \infty$,

$$S_n^2 \xrightarrow{P} (\mu^2 + \sigma^2) - \mu^2 = \sigma^2$$

This completes the proof.

Problem 5. *[This problem is about evaluating point estimators asymptotically.]*

Let $X_1, \ldots, X_n$ be a random sample from $Bernoulli(p)$.[2] Recall that the MLE of p is $\hat{p} = \overline{X}$ and that it is an unbiased estimator.

Show that the MLE attains the Cramer-Rao lower bound.

SOLUTION:

Note that this example is exactly from the lectures (when the Fisher information was introduced and when the example of the Cramer-Rao lower bound was computed). Here is the complete answer:

From Corollary 7.3.10 (P337), we need to prove

$$\mathrm{Var}(\bar{\mathrm{X}}) = \frac{(\frac{d}{dp}\mathrm{E}\bar{X})^2}{n\mathrm{E}[(\frac{\partial}{\partial p}\log f(X|p))^2]}.$$

Since Bernoulli distribution belongs to the exponential family, from Lemma 7.3.11(P338), it is equivalent to show

$$\mathrm{Var}(\bar{\mathrm{X}}) = \frac{(\frac{d}{dp}\mathrm{E}\bar{X})^2}{-n\mathrm{E}[\frac{\partial^2}{\partial p^2}\log f(X|p)]} \tag{7}$$

Since $X_i \sim Bernoulli(p)$, we have $\mathrm{E}X_i = p, \mathrm{Var}(X_i) = p(1-p)$, for $i = 1, \cdots, n$.

Thus $\mathrm{E}\bar{X} = \mathrm{E}X_i = p$, $\mathrm{Var}(\bar{X}) = \frac{1}{n}\mathrm{Var}(X_i) = \frac{p(1-p)}{n}$. We then have

$$\frac{d}{dp}\mathrm{E}\bar{X} = 1 \tag{8}$$

$$\mathrm{Var}(\bar{X}) = \frac{p(1-p)}{n} \tag{9}$$

Also,

$$\begin{aligned} \log f(X|p) = \log[p^X(1-p^{1-X})] &= X\log p + (1-X)\log(1-p), \\ \Rightarrow \frac{\partial}{\partial p}\log f(X|p) &= \frac{X}{p} - \frac{1-X}{1-p} \\ \Rightarrow \frac{\partial^2}{\partial p^2}\log f(X|p) &= -\frac{X}{p^2} - \frac{1-X}{(1-p)^2} \end{aligned}$$

Bearing in mind that $\mathrm{E}X = p$, we have

$$\mathrm{E}[\frac{\partial^2}{\partial p^2}\log f(X|p)] = -\frac{1}{p} - \frac{1}{1-p} = -\frac{1}{p(1-p)} \tag{10}$$

Taking (8) and (10) into (7), we find the right side of it to be $\frac{1}{-\frac{n}{-p(1-p)}} = \frac{p(1-p)}{n}$, which, according to (9), is the left side of (7). This completes our proof of (7).

[2]Recall that $E[X_i] = p$ and $Var(X_i) = p(1-p)$.