Stat 8111 # Final Exam December '14

Twenty two people took the exam. The maximum possible score was 90. The highest score was 70. There were 7 in the 60's, 9 in the 50's, 3 in the 40's and 2 in the 30's.

You may use any fact or theorem proved in class. Please put the answer to each question on a separate page with your name on it.

1. Let $X_1, \ldots, X_n$ be independent and identically distributed each Bernoulli $(\exp -\theta)$ where $\theta \in \Theta = (0, \infty)$ is unknown.

i) Find the form of the UMP level α test of H: $\theta \le \theta_0$ against K: $\theta > \theta_0$ where θ_0 is some known positive number.

ii) Assuming the loss function is convex identify the class of functions of θ for which a best unbiased estimator exists. For such a function what is its best unbiased estimator.

2. Let X be Normal(θ, σ^2) where σ^2 is known and $\theta \in \Theta = \{-1, 0, 1\}$. Let the decision space $D = \Theta$ and the loss function be given by

$$L(\theta, d) = \begin{cases} 0 & \text{if } \theta = d \\ 1 & \text{if } \theta \neq d \end{cases}$$

Find a Bayes rule against the prior which puts equal mass on the three points of Θ.

3. Let $\Theta = \{0, 1\}$, $D = [0, 1]$ and $L(\theta, d) = |\theta - d|$. For this no data problem draw the risk set. Suppose now that given θ X_1, X_2 and X_3 are independent and identically distributed each Bernoulli $((1+\theta)/3)$. Draw the lower boundary of the risk set for this second problem.

4. Let $X_1, \ldots, X_n$ be independent and identically distributed with common density function given by

$$f(x; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp \frac{-\lambda(x-\mu)^2}{2\mu^2 x} \quad \text{for} \quad x > 0$$

where $0 < \mu < \infty$ and $0 < \lambda < \infty$ are unknown parameters.

(Note: this density was misspecified on the exam.)

i) Find the complete sufficient statistic for $\theta = (\mu, \lambda)$.

ii) Suppose μ is known and the prior density for λ is given by

$$g(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp -\beta\lambda \quad \text{for} \quad \lambda > 0$$

where $\alpha > 0$ and $\beta > 0$ are known. Assuming squared error loss find the Bayes estimator of λ.

5. Let $X_1, \ldots, X_n$ be independent and identically distributed where each is uniform$(\theta, \theta+1)$. For testing $H: \theta <= \theta_0$ against $K: \theta > \theta_0$ at level α where $0 < \alpha < 1$ show that there exists a UMP level α test and find its form.

6. Let X be a vector valued random variable with a family of possible distributions $\{P_\theta : \theta \in \Theta\}$ indexed by the parameter θ. Consider the problem of estimating some real valued function $\gamma(\theta)$ with squared error loss. Let Δ, assumed to be nonempty, be the class of all unbiased estimators of $\gamma(\theta)$ with finite variance. In what follows all integrals discussed are assumed to be finite.

Suppose now θ is no longer considered to be a parameter but an unobserved realization of a probability distribution G on Θ. Suppose G is not known and assumed to belong to $\mathcal{G}$ some family of distributions over Θ. Then $\{P_G : G \in \mathcal{G}\}$ is the family of unknown distributions for X where

$$P_G(A) = \int_A d\left\{\int_\Theta P_\theta(x)\, dG(\theta)\right\}$$

for a measurable set A in the sample space of X. The problem now is to estimate

$$\overline{\gamma}(G) = \int_\Theta \gamma(\theta)\, dG(\theta)$$

with squared error loss. Let $\overline{\Delta}$ be the class of all unbiased estimators of $\overline{\gamma}(G)$.

i) Show that if $\delta \in \Delta$ then $\delta \in \overline{\Delta}$.

ii) Show that if δ is a minimum variance unbiased estimator (MVU) of $\gamma(\theta)$ and it has finite variance for $G \in \mathcal{G}$ it is also a MVU estimator of $\overline{\gamma}(G)$ within the class of estimators Δ.

iii) Give a condition on $\mathcal{G}$ which implies that $\Delta = \overline{\Delta}$.

iv) Give an example where Δ is a proper subset of $\overline{\Delta}$ and a MVU estimator of $\overline{\gamma}(G)$ for $\overline{\Delta}$ does not belong to Δ.

Solutions

1. i) For $\theta_1 < \theta_2$ we have $f_{\theta_2}(x_1, \ldots, x_n)/f_{\theta_1}(x_1, \ldots, x_n)$ is decreasing in $\sum x_i$ or it has monotone likelihood property in $-\sum x_i$. So we reject when $-\sum x_i > c$ or when $\sum x_i < c'$.

ii) $\sum X_i$ is a complete and sufficient statistic. $E_\theta \delta(X_1, \ldots, X_n) = \gamma(\theta)$ implies that $\gamma(\theta)$ is a polynomial of degree $\leq n$ in $\exp^- \theta$. Just as in class all such functions will have a best unbiased estimator.

2. For the no data problem the Bayes rule takes d to be the member of the parameter space with the largest prior probability. Hence for the data problem a Bayes rule selects the d with the largest posterior probability. Now $P\{\theta = 0 \,|\, x\} > P\{\theta = -1 \,|\, x\}$ if and only if

$$\exp\{-x^2/2\sigma^2\} > \exp\{-(x+1)^2/2\sigma^2\}$$

which occurs if and only if $x > -0.5$. In the same way $P\{\theta = 0 \,|\, x\} > P\{\theta = 1 \,|\, x\}$ in and only if $x < 0.5$. So the Bayes rule will select -1 when $x < 0.5$, 0 when $-0.5 < x < 0.5$ and 1 when $x > 0.5$. When $x = -0.5$ or $x = 0.5$ it can be either of the two appropriate values.

3. For the no data problem for a given d its risk vector is just $(d, 1-d)$. So the risk set is just the line segment joining (0,1) and (1,0).

Let π be the prior probability that $\theta = 0$. Then the Bayes risk of d against π is just $\pi d + (1-\pi)(1-d) = 1 - \pi + d(2\pi - 1)$. So a Bayes rule against π is to take $d = 0$ when $\pi \geq 0.5$ and $d = 1$ when $\pi < 0.5$.

For the data problem reduce to a sufficient statistic $Y = \sum X_i$. A non-randomized rule δ can be characterized by the vector $(\delta(0), \delta(1), \delta(2), \delta(3))$. By the previous paragraph we only need to consider cases where $\delta(y)$ is either 0 or 1 for each value of y. There are five obvious sensible rules which are (0,0,0,0), (0,0,0,1), (0,0,1,1), (0,1,1,1) and (1,1,1,1). Their respective risk vectors are easily found to be (0,1), (1/27,19/27), (7/27,7/27), (19/27,1/27) and (1,0). The piece wise linear path you get by joining these five points gives you the lower boundary of the risk set because it is easy to check that the other non-randomized rules give you risk vectors which are above this path.

4. i) Since we are in the exponential family and

$$\sum_i \frac{-\lambda(x_i - \mu)^2}{2\mu^2} = -\frac{\lambda}{2\mu^2} \sum_i x_i + \frac{\lambda}{\mu} - \frac{\lambda}{2} \sum_i \frac{1}{x_i}$$

we see that $(\sum_i X_i, \sum_i 1/x_i)$ is complete sufficient,

ii) It is easy to check that the posterior is also a gamma distribution with parameters $\alpha' = \alpha + n/2$ and

$$\beta' = \beta + \sum_i \frac{(x_i - \mu)^2}{2\mu^2 x_i}$$

So the Bayes estimate of λ is α'/β'.

5. Consider testing $\theta = \theta_0$ against $\theta = \theta_1$ where $\theta_0 < \theta_1$. We need to consider two cases: where $\theta_1 \geq \theta_0$ and $\theta_0 < \theta_1 < \theta_0 + 1$.

For the first case any test which rejects H when $\max x_i > \theta_0 + 1$ has power 1 and size 0 so it is MP level α.

For the second case let $r(x) = f_{\theta_1}(x_1, \ldots, x_n)/f_{\theta_0}(x_1, \ldots, x_n)$. We see that

$$\begin{aligned} r(x) &= \infty \quad \text{when } \theta_0 + 1 < \max x_i < \theta_1 + 1 \\ &= 1 \quad \text{when } \theta_1 < \min x_i \leq \max x_i < \theta_0 + 1 \\ &= 0 \quad \text{when } \theta_0 < \min x_i < \theta_1 \text{ and } \max x_i < \theta_0 + 1 \end{aligned}$$

Let $c(\alpha)$ be chosen to satisfy $P_{\theta_0}(\min X_i > \theta_0 + c(\alpha)) = \alpha$

It follows from the NP lemma that the test which rejects $\theta = \theta_0$ when $\max x_i > \theta_0 + 1$ or when $\min x_i > \theta_0 + c(\alpha)$ is most powerful against θ_1. In the NP lemma we use the cut off point $k = 0$ when $\theta_1 > \theta_0 + c(\alpha)$ and use $k = 1$ when $\theta_1 \le \theta_0 + c(\alpha)$.

6. i)

$$\begin{aligned} E_G\delta &= \int \delta(x)\, dP_G(x) = \int \delta(x)\, d\left\{ \int_\Theta P_\theta(x)\, dG(\theta) \right\} \\ &= \int_\Theta \left\{ \int \delta(x)\, dP_\theta(x) \right\} dG(\theta) = \int_\Theta \gamma(\theta)\, dG(\theta) = \overline{\gamma}(G) \end{aligned}$$

ii) For $\delta \in \Delta$ since $E(\delta(X)|\theta) = \theta$ we have

$$\begin{aligned} E_G(\delta(X) - \overline{\gamma}(G))^2 &= E\big(\delta(X) - E(\delta(X)|\theta) + E(\delta(X)|\theta) - \overline{\gamma}(G)\big)^2 \\ &= E(\delta(X) - \gamma(\theta))^2 + E(\gamma(\theta) - \overline{\gamma}(G))^2 \\ &= \int_\Theta Var_\theta\delta\, dG(\theta) + E(\gamma(\theta) - \overline{\gamma}(G))^2 \end{aligned}$$

Note the second term does not depend on δ. So if δ^* is MVU for estimating $\gamma(\theta)$ it minimizes the first term over all $\delta \in \Delta$.

Note that the above calculation is just the familiar formula

$$Var_G\delta = E_G\{Var(\delta|\theta)\} + Var_G\{E(\delta|\theta)\}$$

since δ is unbiased for $\gamma(\theta)$.

iii) One easy condition is that $\mathcal{G}$ contains all distributions which put probability one on a single point of Θ.

iv) Let $\mathcal{G} = \{G : \int_\Theta \gamma(\theta)\, dG(\theta) = c\}$ for some constant c. Then $\delta^*(X) \equiv c$ is MVU for estimating $\overline{\gamma}(G)$ for $G \in \mathcal{G}$ but typically it will not belong to Δ.