Statistics 582, Final Exam, Solutions

Wellner; 3/13/2007

1. (30 points) **Define** *any three* of the following terms. In each case, briefly provide an appropriate context for your definition.

   (a) A family of distributions $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset \mathbb{R}\}$ (with densities $p_\theta = dP_\theta/d\mu$ with respect to a dominating measure μ) with *monotone likelihood ratio.*

   (b) An *unbiased* level α test of $H : \theta \in \Theta_0$ versus $K : \theta \in \Theta_1$.

   (c) A similar test of $H : \theta \in \Theta_0$ versus $K : \theta \in \Theta_1$.

   (d) A *minimax decision rule* in a general decision problem with loss function $L(\theta, d)$.

   Solution: See the Chapter 5 and 6 notes.

2. (30 points) **State** *any three* of the following results:

   (a) The Neyman - Pearson lemma.

   (b) The Karlin-Rubin theorem.

   (c) A theorem relating similar tests to tests with Neyman structure.

   (d) A theorem about admissibility properties of the sample mean $\overline{X}$ when sampling from a normal distribution on $\mathbb{R}$ and a contrasting theorem for sampling from a normal distribution on $\mathbb{R}^d$.

   (e) A theorem relating Bayes rules to minimax rules and least favorable prior distributions.

   (f) A conditional limit theorem about the large sample behavior of posterior distributions

   Solution: See the Chapter 5 and 6 notes.

   Everyone should do problems 3.

3. (36 points) A random variable X takes on the values $1,\ 2,\ 3,\ 4$ with probability distribution $p_0(x)$ or $p_1(x)$ as follows:

x	1	2	3	4
$p_0(x)$	.05	.07	.40	.48
$p_1(x)$	.15	.35	.35	.15

(a) For the usual $0 - 1$ loss, find a most powerful test of size .10 for testing $H : p = p_0$ versus $K : p = p_1$ and determine its power.

(b) Find a test ϕ which minimizes the sum of risks $E_0\phi + E_1(1-\phi)$. What is the relationship between the minimized sum of risks and the total variation distance between P_1 and P_2 and why does this make intuitive sense?

(c) If the losses are $L(1,1) = L(0,0) = 0$, $L(0,1) = 3$, $L(1,0) = 2$, and the prior is $\lambda = (\lambda_0, \lambda_1) = (.4, .6)$, find the Bayes rule d_B and the minimax rule d_M.

Solution:

Solution:

(a). The ratios of the probabilities under the two hypotheses are given by the following table

x	1	2	3	4
$p_0(x)$	.05	.07	.40	.48
$p_1(x)$	.15	.35	.35	.15
$p_1(x)/p_0(x)$	3	5	7/8	15/48

Thus by the Neyman - Pearson lemma the most powerful test of size .10 is $\phi(x) = 1\{x = 2\} + (3/5)1\{x = 1\}$. The power of this test is $\beta_\phi = E_1\phi(X) = P_1(X = 2) + (3/5)P_1(X = 1) = .35 + (3/5)(.15) = .35 + .09 = .44$.

(b). The sum of risks $E_0\phi + E_1(1-\phi) = 1 + \sum_x \phi(x)\{p_0(x) - p_1(x)\}$ is minimized by any rule $\phi(x) = 1\{x : \; p_1(x) > p_0(x)\} = 1_{\{1,2\}}(x)$. The minimum sum of risks is $E_0\phi(X) = 1 + (p_0(1) - p_1(1)) + (p_0(2) - p_1(2)) = 1 - .1 - .28 = .62$. [Note that the total risk of the Neyman-Pearson test of size .10 is $.1 + .56 = .66$.] The total variation distance between P_0 and P_1 is $d_{TV}(P_0, P_1) = 1 - \eta(P_0, P_1)$ where $\eta(P_0, P_1) = \int p_0 \wedge p_1 d\mu = .05 + .07 + .35 + .15 = .62$. Thus

$$\min_\phi \{E_0\phi + E_1(1-\phi)\} = \eta(P_0, P_1) = 1 - d_{TV}(P_0, P_1)$$

This is intuitively reasonable: if $d_{TV}(P_0, P_1)$ is large, then we can make the minimum sum of errors small, but if $d_{TV}(P_0, P_1)$ is small, then the minimum sum of errors will be large.

(c). For a rule $d = (d_1, d_2, d_3, d_4)$ which chooses 1 with probability d_x when x is observed, the (ordinary) risks are

$$\begin{aligned} R(0,d) &= E_0L(0,d) = 3\{d_1(.05) + d_2(.07) + d_3(.40) + d_4(.48) \\ &= \{.15d_1 + .21d_2 + 1.2d_3 + 1.44d_4\}, \\ R(1,d) &= E_1L(1,d) \\ &= 2\{(1-d_1)(.15) + (1-d_2)(.35) + (1-d_3)(.35) + (1-d_4)(.15)\} \\ &= 2 - .3d_1 - .7d_2 - .7d_3 - .3d_4. \end{aligned}$$

Thus the Bayes risk for the prior $\lambda = (.4, .6)$ is

$$\begin{aligned}
\mathcal{R}(\lambda, d) &= .4R(0,d) + .6R(1,d) \\
&= \frac{1}{10}\{.6d_1 + .84d_2 + 4.8d_3 + 4 \times 1.44d_4\} \\
&\qquad + \frac{1}{10}\{12 - (1.8d_1 + 4.2d_2 + 2.8d_3 + 1.2d_4)\} \\
&= \frac{1}{10}\{12 + (.6 - 1.8)d_1 + (.84 - 4.2)d_2 + (4.8 - 2.8)d_3 + (4 \times 1.44 - 1.2)d_4\}.
\end{aligned}$$

Since the coefficients of d_1 and d_2 are negative, and the coefficients of d_3 and d_4 are positive, the Bayes rule is given by $d_B = (1, 1, 0, 0)$ with corresponding Bayes risk $\mathcal{R}(\lambda, d_B) = (12 - 1.2 - 4.2 + .84)/10 = 7.44/10 = .744 = 93/125$. The ordinary risks of this rule are $R(0, d_B) = .36$ and $R(1, d_B) = 1$. [Crosscheck: $.4R(0, d_B) + .6R(1, d_B) = .4 \times .36 + .6 = .744$.]

To find the minimax rule d_M, we first consider the non-random rule $d = (1, 1, 1, 0)$. This rule has ordinary risks $R(0, d) = 1.56$, $R(1, d) = .3$. Thus we suspect that a minimax rule is of the form $d' = (1, 1, d'_3, 0)$. Any rule of this form has risks $R(0, d') = .36 + 1.2d'_3$, $R(1, d') = 1 - .7d'_3$. Equating these and solving for d'_3 yields $1.9d'_3 = .64$, or $d'_3 = .64/1.9 = 32/95$. Then $R(0, d') = .36 + 1.2(32/95) = .7642$, while $R(1, d) = 1 - .7(32/95) = .7642$. Thus the particular rule $d_M = (1, 1, 32/95, 0)$ is minimax.

Do any two of problems 4 - 6.

4. (40 points) State and prove the "short form" of the generalized Neyman-Pearson lemma.

   Solution: See the Chapter 6 notes.

5. (40 points) Prove the following theorem (in the context of finitely many states of nature): Suppose that d_0 is Bayes with respect to $\lambda = (\lambda_1, \ldots, \lambda_l)$ and $\lambda_i > 0$ for $i = 1, \ldots, l$. Then d_0 is admissible.

   Solution: See the Chapter 5 notes.

6. (40 points) In the context of $X_1, \ldots, X_n$ i.i.d. either P or Q with densities p and q with respect to some dominating measure μ, show that Neyman - Pearson tests are both size and power consistent under a simple restriction on the constants specifying the tests.

   Solution: See the Chapter 6 notes.

Do either problem 7 or problem 8.

7. (36 points) Suppose that $X_1, \ldots, X_n$ are i.i.d. Poisson(θ) random variables so that $P(X_1 = x) = \exp(-\theta)\theta^x/x!$ for $x \in \{0, 1, \ldots\}$ and $\theta > 0$.

(a) Find the MLE of θ.

(b) If $\theta \sim \text{Gamma}(\alpha, \beta)$ so that

$$\lambda(\theta) = \frac{\beta^\alpha \theta^{\alpha-1}}{\Gamma(\alpha)} \exp(-\beta\theta) 1_{(0,\infty)}(\theta)$$

and $E(\theta) = \alpha/\beta$, find the posterior distribution of θ.

(c) Find the Bayes estimator of θ for squared error loss. Is it consistent?

(d) What is the asymptotic behavior of the posterior distributions you found in (b) when appropriately centered and normalized?

Solution: (a) The score function for a sample of size 1 is $\dot{\mathbf{l}}_\theta(x) = x/\theta - 1 = (x-\theta)/\theta$, and hence the likelihood equation is $\dot{\mathbf{l}}_n(\theta) = (\sum_1^n X_i - n\theta)/\theta = 0$, and hence the MLE is $\widehat{\theta}_n = \overline{X}_n$.

(b) The posterior density of θ is proportional to

$$e^{-n\theta}\theta^{\sum_1^n x_i} \cdot \theta^{\alpha-1}e^{-\beta\theta} = \theta^{\sum_1^n x_i + \alpha - 1}e^{-(\beta+n)\theta}.$$

It follows that the posterior distribution of θ, i.e. the distribution of θ given $\underline{X}$ is Gamma $(\sum_{i=1}^n X_i + \alpha, n+\beta)$.

(c) It follows from (b) that the Bayes estimator of θ is

$$\begin{aligned} E(\theta|\underline{X}) &= \frac{\sum_{i=1}^n X_i + \alpha}{n+\beta} \\ &= \frac{n}{n+\beta}\overline{X}_n + \frac{\beta}{n+\beta}\frac{\alpha}{\beta} \\ &\rightarrow_{a.s.} 1 \cdot \theta + 0 \cdot \frac{\alpha}{\beta} = \theta, \end{aligned}$$

so $d_B(\underline{X}) = E(\theta|\underline{X})$ is a consistent estimator of θ.

(d) Now T_n of the Bernstein-von Mises theorem is given by

$$T_n = \theta_0 + \frac{1}{I(\theta_0)} n^{-1}\dot{\mathbf{l}}_n(\theta_0) = \theta_0 + \theta_0\left(\frac{\overline{X}_n}{\theta_0} - 1\right) = \overline{X}_n,$$

since $I(\theta_0) = 1/\theta_0$, so the Bernstein - von Mises theorem says that

$$d_{TV}(P(\sqrt{n}(\theta - \overline{X}_n) \le \cdot|\underline{X}), \sqrt{I(\theta_0)}\Phi(\sqrt{I(\theta_0)}\cdot)) \rightarrow_p 0$$

as $n \rightarrow 0$.

8. (36 points) Suppose that $X_1, \ldots, X_m$ are i.i.d. Geometric(μ) and that $Y_1, \ldots, Y_n$ are i.i.d. Geometric(ν) and independent of the X_i's. (Thus $P(X_i = x) = \mu(1-\mu)^{x-1}$, $x \in \{1, 2, \ldots\}$ and $P(Y_i = y) = \nu(1-\nu)^{y-1}$ for $y \in \{1, 2, \ldots\}$ where $0 \le \mu, \nu \le 1$. (Reminder: $E(X_1) = 1/\mu$, $Var(X_1) = (1-\mu)/\mu^2$.)

(a) Consider testing $H: \ \nu \le \mu$ versus $K: \ \nu > \mu$. Find the joint density function of $(\underline{X}, \underline{Y})$ and show that it can be written in the form of an exponential family with parameter of interest θ which takes a fixed value θ_0 on the boundary Θ_B (with an associated statistic $U(\underline{X}, \underline{Y})$), and a nuisance parameter ξ (with an associated sufficient statistic $T(\underline{X}, \underline{Y})$).

(b) Find the conditional distribution of $U(\underline{X}, \underline{Y})$ given the sufficient statistic $T(\underline{X}, \underline{Y})$ for Θ_B.

(c) Find the UMPU test of H versus K, specifying the constants involved as precisely as possible.

(d) Describe the Bayes test of H versus K for $0-1$ loss assuming a prior distribution Λ of (μ, ν). Can you describe the rejection region of this test when Λ is the product of two independent Beta priors (i.e. $\mu \sim$ Beta(α, β) and $\nu \sim$ Beta(γ, δ))?

Solution: (a) The joint density of the observations is

$$
\begin{aligned}
p_{\mu,\nu}(\underline{x}, \underline{y}) &= \mu^m (1-\mu)^{\sum_1^m x_i - m} \nu^n (1-\nu)^{\sum_1^n y_j - n} \\
&= \left(\frac{\mu}{1-\mu}\right)^m \left(\frac{\nu}{1-\nu}\right)^n (1-\mu)^{\sum_1^m x_i} (1-\nu)^{\sum_1^n y_j} \\
&\equiv c_{m,n}(\mu,\nu) \exp\left(\sum_1^m x_i \log(1-\mu) + \sum_1^n y_j \log(1-\nu)\right) \\
&= c_{m,n}(\mu,\nu) \exp\left(\sum_1^n y_j \log(1-\nu) - \sum_1^n y_j \log(1-\mu) + (\sum_1^m x_i + \sum_1^n y_j)\log(1-\mu)\right) \\
&= C(\theta, \xi) \exp(\theta U(\underline{x}, \underline{y}) + \xi T(\underline{x}, \underline{y}))
\end{aligned}
$$

where

$$
\theta \equiv \log\left(\frac{1-\nu}{1-\mu}\right), \qquad U(\underline{x}, \underline{y}) = \sum_1^n y_j,
$$

$$
\xi \equiv \log(1-\mu), \qquad T(\underline{x}, \underline{y}) = \sum_1^m x_i + \sum_1^n y_j.
$$

Thus $\mu = \nu$ if and only if $\theta = 0 (\equiv \theta_0)$, and $\nu > \mu$ if and only if $\theta < 0 = \theta_0$. Thus $T \equiv T(\underline{X}, \underline{Y}) = \sum_1^m X_i + \sum_1^n Y_j$ is sufficient for Θ_B.
(b) With $R \equiv \sum_1^m X_i$ and $S \equiv \sum_1^n Y_j$, we know that

$$R \sim \text{Negative Binomial}(m, \mu), \quad P_\mu(R = r) = \binom{r-1}{m-1} \mu^m (1-\mu)^{r-m}, \quad r = m, m+1, \ldots$$
$$S \sim \text{Negative Binomial}(n, \nu), \quad P_\nu(S = s) = \binom{s-1}{n-1} \nu^n (1-\nu)^{s-n}, \quad s = n, n+1, \ldots.$$

Thus the conditional distribution of $U = S$ given $T = R + S$ is given by

$$\begin{aligned}
P_{\mu,\nu}(S = s | T = t) &= P(R = t - s, S = s)/P(T = t) \\
&= \frac{\binom{t-s-1}{m-1} \mu^m (1-\mu)^{t-s-m} \binom{s-1}{n-1} \nu^n (1-\nu)^{s-n}}{\sum_{s'} \binom{t-s'-1}{m-1} \mu^m (1-\mu)^{t-s'-m} \binom{s'-1}{n-1} \nu^n (1-\nu)^{s'-n}} \\
&= \frac{\binom{t-s-1}{m-1}\binom{s-1}{n-1} \left(\frac{1-\nu}{1-\mu}\right)^s}{\sum_{s'} \binom{t-s'-1}{m-1}\binom{s'-1}{n-1} \left(\frac{1-\nu}{1-\mu}\right)^{s'}} \\
&\overset{\mu=\nu}{=} \frac{\binom{t-s-1}{m-1}\binom{s-1}{n-1}}{\sum_{s'=n}^{t-m} \binom{t-s'-1}{m-1}\binom{s'-1}{n-1}}.
\end{aligned}$$

When $\mu = \nu$, $T = R + S \sim$ Negative binomial$(m + n, \mu)$ with

$$P_\mu(T = t) = \binom{t-1}{m+n-1} \mu^{m+n} (1-\mu)^{t-(m+n)}, \qquad t = m+n, m+n+1, \ldots,$$

and it therefore follows that the normalizing constant in the denominator of the last display is simply

$$\binom{t-1}{m+n-1}.$$

Thus we have

$$P_{\mu,\mu}(S = s | T = t) = \frac{\binom{t-s-1}{m-1}\binom{s-1}{n-1}}{\binom{t-1}{m+n-1}}, \qquad s \in \{n, \ldots, t-m\}.$$

I do not know a name for this distribution.

(c) The UMPU test of H versus K is given (conditionally on $T = t$) by

$$\varphi(\underline{X}, \underline{Y}) = \begin{cases} 1 & \text{if } \sum_1^n Y_j < c(t), \\ \gamma(t) & \text{if } \sum_1^n Y_j = c(t), \\ 0 & \text{if } \sum_1^n Y_j > c(t) \end{cases}$$

where $c(t)$ and $\gamma(t)$ satisfy $E\{\varphi(\underline{X},\underline{Y})|T=t)=\alpha$.

(d) The Bayes test for $0-1$ loss and a prior Λ for (μ,ν) is "reject H if

$$P\left((\mu,\nu)\in K\middle|\underline{X},\underline{Y}\right) > P\left((\mu,\nu)\in H\middle|\underline{X},\underline{Y}\right)\text{";}$$

i.e. reject H if

$$P\left(\nu>\mu\middle|\underline{X},\underline{Y}\right) > P\left(\nu\le\mu\middle|\underline{X},\underline{Y}\right).$$

Since

$$P\left(\nu\le\mu\middle|\underline{X},\underline{Y}\right) = 1-P\left(\nu>\mu\middle|\underline{X},\underline{Y}\right),$$

this can be rewritten as "reject H if

$$P\left(\nu>\mu\middle|\underline{X},\underline{Y}\right) > 1/2\text{".} \tag{1}$$

When the prior distribution Λ of (μ,ν) is Beta$(\alpha,\beta)\times$ Beta(γ,δ), then

$$\begin{aligned} p(\mu,\nu|\underline{X},\underline{Y}) &\propto \mu^m(1-\mu)^{\sum_1^m X_i-m}\nu^n(1-\nu)^{\sum_1^n Y_j-n}\cdot\mu^{\alpha-1}(1-\mu)^{\beta-1}\cdot\nu^{\gamma-1}(1-\nu)^{\delta-1} \\ &= \mu^{m+\alpha-1}(1-\mu)^{\sum_1^m X_i+\beta-m-1}\cdot\nu^{n+\gamma-1}(1-\nu)^{\sum_1^n Y_j+\delta-n-1}, \end{aligned}$$

and hence the joint posterior density of (μ,ν) is

$$\text{Beta}(m+\alpha,\sum_1^m X_i+\beta-m)\times\text{Beta}(n+\gamma,\sum_1^n Y_j+\delta-n).$$

Thus the rejection region of (1) can be written as

$$\begin{aligned} &c(m+\alpha,\sum_1^m X_i+\beta-m)c(n+\gamma,\sum_1^n Y_j+\delta-n) \\ &\quad\cdot\int_0^1\int_0^v v^{m+\alpha-1}(1-v)^{\sum_1^m X_i+\beta-m-1}u^{n+\gamma}(1-u)^{\sum_1^n Y_j+\delta-n}du dv \\ &\quad > 1/2 \end{aligned}$$

where

$$c(\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}.$$