Statistics 582, Problem Set 6 Solutions

Wellner; 2/16/2011

1. Consider Example 5.5.4 on pages 16 and 17 of the Chapter 5 notes.
(a) Show that the variance of $\hat{\psi}$ is given by

$$Var(\hat{\psi}_n) = \frac{1}{n}\left\{ \frac{1}{B}\sum_{j=1}^{B} \frac{\theta_j}{\xi_j} - \psi(\theta)^2 \right\}.$$

[Hint: use the formula $Var(Y) = EVar(Y|X) + Var[E(Y|X)]$ twice.]
(b) Use the result of (a) to show that

$$Var(\hat{\psi}_n) \le \frac{1}{n\delta}$$

under the assumption that $\xi_j \ge \delta > 0$ for all $1 \le j \le B$.
(c) What if the sampling probabilities $\xi_i = 1$ for all i: do the conclusions of Example 4.5.4 and (a) and (b) above still hold?

Solution: (a) Since the (X_i, R_i, Y_i)'s are i.i.d.,

$$\begin{aligned}
Var(\hat{\psi}_n) &= n^{-1} Var\left(\frac{R_1 Y_1}{\xi_{X_1}}\right) \\
&= n^{-1}\left\{ EVar\left(\frac{R_1 Y_1}{\xi_{X_1}}|R_1, X_1\right) + Var\left(E\left(\frac{R_1 Y_1}{\xi_{X_1}}|R_1, X_1\right)\right)\right\} \\
&= n^{-1}\left\{ E\left(\frac{R_1^2}{\xi_{X_1}^2}\theta_{X_1}(1-\theta_{X_1})\right) + Var\left(\frac{R_1}{\xi_{X_1}}\theta_{X_1}\right)\right\} \\
&= n^{-1}\left\{ EE\left(\frac{R_1^2}{\xi_{X_1}^2}\theta_{X_1}(1-\theta_{X_1})|X_1\right)\right. \\
&\qquad \left. + EVar\left(\frac{R_1}{\xi_{X_1}}\theta_{X_1}|X_1\right) + Var\left(E\left(\frac{R_1}{\xi_{X_1}}\theta_{X_1}|X_1\right)\right)\right\} \\
&= n^{-1}\left\{ E\left(\frac{\theta_{X_1}(1-\theta_{X_1})}{\xi_{X_1}}\right)\right. \\
&\qquad \left. + E\left(\frac{\theta_{X_1}^2}{\xi_{X_1}^2}\xi_{X_1}(1-\xi_{X_1})\right) + Var(\theta_{X_1})\right\} \\
&= n^{-1}\left\{ \frac{1}{B}\sum_{j=1}^{B}\frac{\theta_j(1-\theta_j)}{\xi_j} + \frac{1}{B}\sum_{j=1}^{B}\theta_j^2\frac{1-\xi_j}{\xi_j} + \frac{1}{B}\sum_{j=1}^{B}(\theta_j - \overline{\theta})^2\right\} \\
&= n^{-1}\left\{ \frac{1}{B}\sum_{j=1}^{B}\frac{\theta_j}{\xi_j} - \psi(\theta)^2\right\}.
\end{aligned}$$

(b) Since $\xi_j \geq \delta$ and $\theta_j \leq 1$ for all j, it follows that

$$Var(\hat{\psi}_n) \leq n^{-1} \frac{1}{B} \sum_{j=1}^{B} \frac{1}{\delta} = \frac{1}{n\delta}.$$

(c) When $\xi_i = 1$ for all $i \in \{1, \ldots, B\}$ the calculations of Example 4.5.4 and (a) and (b) above remain valid. The estimator becomes simply $\hat{\psi}_n = \overline{Y}_n$, and the variance computed in (a) reduces to

$$Var(\hat{\psi}_n) = n^{-1} \left\{ \psi(\theta) - \psi^2(\theta) \right\} = n^{-1}\psi(\theta)(1 - \psi(\theta)) \leq \frac{1/4}{n}.$$

Thus the missing data aspect of the problem (and the introduction of the Horvitz-Thompson estimator) is not crucial to the essence of the problem which is the high-dimensionality of the parameter space.

2. A random variable X takes on values in the set $\{1, 2, 3, 4\}$ with probability distributions $p_0(x)$ or $p_1(x)$ given in the following table.

x	1	2	3	4
$p_0(x)$	.1	.3	.4	.2
$p_1(x)$	.2	.2	.2	.4

(a) Find a most powerful test of size $\alpha = .2$ for testing p_0 versus p_1 and determine its power.
(b) Find a test ϕ which minimizes the sum of risks $a + b$ where $a \equiv E_0\phi$ and $b = E_1(1 - \phi)$.
(c) Compute $d_{TV}(P_0, P_1)$, $H(P_0, P_1)$, and the affinity $\rho(P_0, P_1)$. For the product laws P_0^n and P_1^n (corresponding to observation of $X_1, \ldots, X_n$ i.i.d. P_0 or P_1 respectively), compute $\rho(P_0^n, P_1^n)$ and $H(P_0^n, P_1^n)$ for $n = 10, 25, 100$. What does this imply about the test ϕ_n based on $X_1, \ldots, X_n$ which minimizes the sum of risks?

Solution: (a) Now $p_1(x)/p_0(x) = 2,\ 2/3,\ 1/2,\ 2$ according as $x = 1,\ 2,\ 3,\ 4$, so a most powerful test of size $\alpha = .2$ is given by

$$\phi(x) = \begin{cases} 2/3 & \text{if } x = 1 \text{ or } 4 \\ 0 & \text{if } x = 2 \text{ or } 3. \end{cases}$$

Then

$$E_0\phi(X) = \frac{2}{3}\{P_0(X = 1) + P_0(X = 4)\} = \frac{2}{3}\{.1 + .2\} = .2,$$

while

$$\text{Power} \;=\; E_1\phi(X) = \frac{2}{3}\{P_1X = 1) + P_1(X = 4)\} = \frac{2}{3}\{.2 + .4\} = .4.$$

In fact, the whole family of MP tests is given by

$$\phi_\gamma(x) \;=\; \begin{cases} \gamma & \text{if } \; x = 1, \\ 1 - \gamma/2 & \text{if } \; x = 4, \\ 0 & \text{if } \; x = 2, 3. \end{cases}$$

for some $0 \le \gamma \le 1$. The first test above is just $\phi_{2/3}$. Then

$$E_0\phi_\gamma(X) = \gamma P_0(X = 1) + (1 - \gamma/2)P_0(X = 4) = .1\gamma + .2 - .2\gamma/2 = .2$$

while

$$\begin{aligned} Power \;&=\; E_1\phi_\gamma(X) = \gamma P_1(X = 1) + (1 - \gamma/2)P_1(X = 4) \\ &=\; \gamma(.2) + (1 - \gamma/2)(.4) = .4. \end{aligned}$$

(b) The test which minimizes $a + b$ is given by

$$\phi(x) = \left\{ \begin{array}{ll} 1, & \text{if } \; p_1(x) > p_0(x) \\ 0, & \text{if } \; p_1(x) < p_0(x) \end{array} \right\} = \left\{ \begin{array}{ll} 1, & \text{if } \; x = 1 \text{ or } 4 \\ 0, & \text{if } \; x = 2 \text{ or } 3. \end{array} \right.$$

Then $a = E_0\phi(X) = .3$ and $b = E_1(1 - \phi(X)) = .4$ and for the test minimizing $a + b$ we have $a + b = .7$. Note that $1 - d_{TV}(P_0, P_1) = \int p_0 \wedge p_1 d\mu = .7$.
(c) We compute

$$d_{TV}(P_0, P_1) = \frac{1}{2}\{.1 + .1 + .2 + .2\} = \frac{1}{2}(.6) = .3.$$

Furthermore,

$$\rho(P_0, P_1) = \sqrt{.02} + \sqrt{.06} + \sqrt{.08} + \sqrt{.08} = .9520558...$$

so that

$$H^2(P_0, P_1) = 1 - \rho(P_0, P_1) = .0479442,$$

and $H(P_0, P_1) = .2189616$. Note that the inequalities of proposition 2.1.15 (Chapter 2, notes) are indeed satisfied:

$$\begin{aligned} H^2(P_0, P_1) = .0479 < .3 \;&=\; d_{TV}(P_0, P_1) \\ &<\; H(P_0, P_1)(2 - H^2(P_0, P_1))^{1/2} = .2189616(2 - .0479442)^{1/2} = .3059244 < \end{aligned}$$

For $n = 10,\ 25,\ 100$ we have:

n	$\rho(P_0^n, P_1^n)$	$H(P_0^n, P_1^n)$
10	.61182	.6230
25	.29279	.8410
100	.00735	.9963

Since the test $\phi = \phi(X)$ which minimizes the sum of risks has

$$\begin{aligned} E_0\phi(\underline{X}) + E_1(1 - \phi(\underline{X})) &= \int p_0(\underline{x}) \wedge p_1(\underline{x}) d\mu(\underline{x}) \\ &\leq \rho(P_0^n, P_1^n) \leq \rho^n(P_0, P_1) \to 0. \end{aligned}$$

From the table above we see that this happens quite rapidly.

3. Suppose that $X_1, \ldots, X_n$ are i.i.d. $N_k(\theta, \Sigma)$ with Σ known. Suppose that $\theta \sim N_k(\mu, \tau^2 I)$.
(a) Find the Bayes estimator for estimating θ with squared error loss $L(\theta, a) = \|\theta - a\|^2 \equiv \sum_{j=1}^k (\theta_j - a_j)^2$.
(b) Use the result of (a) to show that $\overline{X}_n$ is a minimax estimator of θ.

Solution: If $X_1, \ldots, X_n$ are $N_k(\theta, \Sigma)$, and $\theta \sim N_k(\mu, \tau^2 I)$, then the posterior distribution of θ is

$$(\theta | X_1, \ldots, X_n) \sim N_k(V(n\Sigma^{-1}\overline{X}_n + \tau^{-2} I \mu), V)$$

where $V \equiv (n\Sigma^{-1} + \tau^{-2} I)^{-1}$; this follows from straightforward calculation, assuming that Σ^{-1} exists. Hence the Bayes estimator of θ is

$$d_\Lambda(X) = V(n\Sigma^{-1}\overline{X}_n + \tau^{-2}\mu).$$

This follows from Theorem 5.5.1 since for any rule $d = d(\underline{X})$

$$\begin{aligned} \int |\theta - d(\underline{X})|^2 d\Lambda(\theta|\underline{X}) &= \int |\theta - E(\theta|\underline{X})|^2 d\Lambda(\theta|\underline{X}) + |E(\theta|\underline{X}) - d(\underline{X})|^2 \\ &\leq \int |\theta - E(\theta|\underline{X})|^2 d\Lambda(\theta|\underline{X}) \end{aligned}$$

with equality if and only if $d(\underline{X}) = E(\theta|\underline{X})$.
(b) Now

$$\begin{aligned} \mathcal{R}(\Lambda, d_\Lambda) &= E\{(\theta - d_\Lambda(\underline{X}))^T(\theta - d_\Lambda(\underline{X}))\} \\ &= EE\{(\theta - d_\Lambda(\underline{X}))^T(\theta - d_\Lambda(\underline{X}))\Big|\underline{X}\} \\ &= E\{\text{trace}(n\Sigma^{-1} + \tau^{-2}I)^{-1}\} \\ &= \text{trace}(n\Sigma^{-1} + \tau^{-2}I)^{-1} \\ &\to \text{trace}(n^{-1}\Sigma) \quad \text{as} \quad \tau^2 \to \infty \\ &= R(\theta, \overline{X}) \quad \text{for all} \quad \theta. \end{aligned}$$

Hence by Theorem 5.6.7, $\overline{X}_n$ is minimax.