Statistics 581, Final Exam Solutions

Wellner; 12/16/2009

1. (40) points) **Define** *each* of the following terms. In each case, briefly provide an appropriate context for your definition.
(a) The information matrix for θ in a regular parametric model
$\mathcal{P} = \{P_\theta : \ \theta \in \Theta \subset R^d\}$.
(b) The efficient score function for a parameter θ_1 when $\theta = (\theta_1, \theta_2)$.
(c) The efficient influence function $\tilde{l}_1$ for a parameter θ_1 when $\theta = (\theta_1, \theta_2)$.
(d) The efficient influence function for $\tilde{l}_\nu$ for a differentiable parameter $q(\theta) = \nu(P_\theta)$ in a regular parametric model $\mathcal{P}$.
(e) An asymptotically linear estimator of a parameter $\nu(P)$ with influence function ψ.

Solution: See 581 Course Notes, Chapters 3 and 4.

2. (32 points) **State** *four* of the following five results, providing the appropriate (brief) context for your statement:
(a) The (elementary) Skorokhod theorem.
(b) The Cramér - Wold device.
(c) A result about the finite-dimensional limiting distributions of the sample quantile process $\{\sqrt{n}(\mathbb{F}_n^{-1}(t) - F^{-1}(t)) : \ 0 < t < 1\}$ specifying the assumption(s) carefully.
(d) Le Cam's third lemma.
(e) The Glivenko-Cantelli theorem.

Solution: See 581 Course Notes, Chapters 2 and 4.

Do either problem 3 or problem 4:

3. (50 points). (A two-sample model.) Suppose that $Z \sim \text{Bernoulli}(\eta)$, and conditional on Z, a random variable X has conditional distributions given Z described by $(X|Z) \sim \text{Exponential}(\mu Z + \nu(1-Z))$ where $\mu, \nu > 0$. Thus

$$P(X > x|Z = 1) = \exp(-\mu x), \quad x \geq 0,$$
$$P(X > x|Z = 0) = \exp(-\nu x), \quad x \geq 0.$$

(a) Let $\theta = (\mu, \nu, \eta) \equiv (\theta_1, \theta_2, \theta_3) \in (0, \infty) \times (0, \infty) \times (0, 1)$, and write $(X, Z) \sim P_\theta$. Show that the joint density p_θ of (X, Z) (with respect to the product of Lebesgue measure λ on $(0, \infty)$ and counting measure $\#$ on $\{0, 1\}$ is given by

$$p_\theta(x, z) = (\eta \mu e^{-\mu x})^z ((1-\eta)\nu e^{-\nu x})^{1-z} 1_{(0,\infty)}(x) 1_{\{0,1\}}(z).$$

(b) Find the score(s) for $\theta = (\mu, \nu, \eta)$ and the information matrix for θ (when $n = 1$).
(c) If $(X_1, Z_1), \ldots, (X_n, Z_n)$ are i.i.d. as $(X, Z) \sim P_{\theta_0}$, what is the Cramér - Rao bound for estimation of $q(\theta) = \nu(P_\theta) \equiv \theta_1 - \theta_2 = \mu - \nu$?

(d) For a sample of (X, Z) pairs as in (c), find the maximum likelihood estimator $\widehat{\theta}_n$ of θ.
(e) Find the limiting distribution of $\sqrt{n}(q(\widehat{\theta}_n) - q(\theta_0))$ when $\theta = \theta_0$ is true.
(f) Propose a test statistic for testing $H : \mu = \nu$ versus $K : \mu \neq \nu$ based on the result of (e). What is the limiting distribution of your test statistic under H? What is the limiting distribution of your test statistic under local alternatives of the form $\mu = \nu + tn^{-1/2}$?

Solution: (a) The conditional densities given $Z = 1$ and $Z = 0$ are $p_\theta(x|1) = \mu \exp(-\mu x) 1_{(0,\infty)}(x)$ and $p_\theta(x|0) = \nu \exp(-\nu x) 1_{(0,\infty)}(x)$ respectively (with respect to Lebesgue measure); equivalently

$$p_\theta(x|z) = (\mu \exp(-\mu x))^z \cdot (\nu \exp(-\nu x))^{1-z} 1_{(0,\infty)} 1_{\{0,1\}}(z).$$

Since $Z \sim$ Bernoulli(η), its density (or mass function) with respect to counting measure is $p_\eta(z) = \eta^z(1-\eta)^{1-z} 1_{\{0,1\}}(z)$. The resulting joint density of (X, Z) is therefore

$$p_\theta(x, z) = p_\theta(x|zp_\eta(z) = (\eta\mu \exp(-\mu x))^z \cdot ((1-\eta)\nu \exp(-\nu x))^{1-z} 1_{(0,\infty)} 1_{\{0,1\}(z).}$$

(b) First,

$$\log p_\theta(x, z) = z \log \eta + (1-z)\log(1-\eta) + z(\log \mu - \mu x) + (1-z)(\log \nu - \nu x).$$

Thus the scores $\dot{l}_\mu$, $\dot{l}_\nu$, and $\dot{l}_\eta$ are given by

$$\begin{aligned}
\dot{l}_\mu(x, z) &= z\left(\frac{1}{\mu} - x\right), \\
\dot{l}_\nu(x, z) &= (1-z)\left(\frac{1}{\nu} - y\right), \\
\dot{l}_{eta}(x, z) &= \frac{z}{\eta} - \frac{1-z}{1-\eta},
\end{aligned}$$

and the matrix of second derivatives $\ddot{l}_{ij}$ is given by

$$\left(\begin{array}{ccc}
\ddot{l}_{\mu\mu}(x, z) & \ddot{l}_{\mu\nu}(x, z) & \ddot{l}_{\mu\eta}(x, z) \\
\ddot{l}_{\nu\mu}(x, z) & \ddot{l}_{\nu\nu}(x, z) & \ddot{l}_{\nu\eta}(x, z) \\
\ddot{l}_{\eta\mu}(x, z) & \ddot{l}_{\eta\nu}(x, z) & \ddot{l}_{\eta\eta}(x, z)
\end{array}\right) = \left(\begin{array}{ccc}
-\frac{z}{\mu^2} & 0 & 0 \\
0 & -\frac{1-z}{\nu^2} & 0 \\
0 & 0 & -\frac{z}{\eta^2} - \frac{1-z}{(1-\eta)^2}
\end{array}\right).$$

Thus the information matrix is

$$I(\theta) = -E_\theta\left(\ddot{l}_{ij}(\theta|X_1)\right) = \left(\begin{array}{ccc}
\frac{\eta}{\mu^2} & 0 & 0 \\
0 & \frac{1-\eta}{\nu^2} & 0 \\
0 & 0 & \frac{1}{\eta(1-\eta)}
\end{array}\right).$$

(c) For estimation of $q(\theta) = \theta_1 - \theta_2 = \mu - \nu$, we have $\dot{q}(\theta) = (1, -1)^T$. Thus the Cramér Rao bound for estimation of $q(\theta)$ is given by

$$Var_\theta(T_n) \geq \frac{\alpha^T I^{-1}(\theta)\alpha}{n}$$

where $\alpha = \nabla E_\theta(T(\underline{X})) = \dot{q}(\theta) + \dot{b}(\theta)$ and $b(\theta) \equiv E_\theta(T(\underline{X})) - q(\theta)$ is the bias of T. If T is unbiased, then

$$Var_\theta(T_n) \geq \frac{\dot{q}^T(\theta)I^{-1}(\theta)\dot{q}(\theta)}{n}.$$

(d) The score equations are given by

$$\begin{aligned}
0 &= \sum_{i=1}^n \dot{l}_\mu(X_i, Z_i) = \sum_{i=1}^n Z_i(\frac{1}{\mu} - X_i), \\
0 &= \sum_{i=1}^n \dot{l}_\nu(X_i, Z_i) = \sum_{i=1}^n (1 - Z_i)(\frac{1}{\nu} - X_i), \\
0 &= \sum_{i=1}^n \dot{l}_\eta(Z_i, Z_i) = \sum_{i=1}^n \frac{Z_i - \eta}{\eta(1-\eta)},
\end{aligned}$$

and it is easily seen that the solutions are given by

$$\begin{aligned}
\hat{\mu}_n &= \frac{\sum_{i=1}^n Z_i}{\sum_1^n Z_i X_i}, \\
\hat{\nu}_n &= \frac{\sum_{i=1}^n (1 - Z_i)}{\sum_1^n (1 - Z_i) X_i}, \\
\hat{\eta}_n &= \overline{Z}_n.
\end{aligned}$$

In this case the Hessian is negative definite, and the log-likelihood is strictly concave, the solution of the likelihood equations is unique and yields a maximizer of the log-likelihood.

(e) Now under P_{θ_0}, with $\theta_0 = (\mu_0, \nu_0, \eta_0) \in (0, \infty)^2 \times (0, 1)$, our general theory applies and it follows that $\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d N_3(0, I^{-1}(\theta_0))$ where $I(\theta)$ is as calculated in (b). Since $q(\theta)$ is differentiable with derivative $\dot{q}(\theta_0) = (1, -1, 0)^T$, it follows by the delta method that

$$\sqrt{n}(q(\widehat{\theta}_n) - q(\theta_0)) \to_d N(0, \dot{q}(\theta_0)I^{-1}(\theta_0)\dot{q}(\theta_0)) = N\left(0, \frac{\mu_0^2}{\eta_0} + \frac{\nu_0^2}{1 - \eta_0}\right).$$

(f) One natural (Wald-type) test statistic is

$$\begin{aligned}
W_n &\equiv \{\sqrt{n}(q(\widehat{\theta}_n) - q(\theta_0))\}^2 \dot{q}^T(\widehat{\theta}_n)I^{-1}(\widehat{\theta}_n)\dot{q}(\widehat{\theta}_n) \\
&= nq(\widehat{\theta}_n)^2 \dot{q}^T(\widehat{\theta}_n)I^{-1}(\widehat{\theta}_n)\dot{q}(\widehat{\theta}_n) \\
&= \frac{n(\hat{\mu}_n - \hat{\nu}_n)^2}{\frac{\hat{\mu}_n^2}{\hat{\eta}_n} + \frac{\hat{\nu}_n^2}{1 - \hat{\eta}_n}} \\
&\to_d \frac{(D_1 - D_2)^2}{Var(D_1 - D_2)} \sim \chi_1^2
\end{aligned}$$

under the null hypothesis $\mu = \nu$. If $\theta_n = \theta_0 + \underline{t}n^{-1/2}$ with with $\mu_0 = \nu_0$ (and, for simplicity $t_2 = t_3 = 0$, $t \equiv t_1$), then it follows easily from regularity of the MLE's that

$$\sqrt{n}q(\widehat{\theta}_n) \to_d D_1 - D_2 + t \sim N(t, \dot{q}(\theta_0)^T I^{-1}(\theta_0)\dot{q}(\theta_0)) = N(t, \frac{\mu_0^2}{\eta_0(1 - \eta_0)}).$$

Then it follows that the statistic W_n in (f) satisfies

$$W_n \to_d N(t/\sqrt{\mu_0^2/(\eta_0(1-\eta_0))}, 1)^2 \sim \chi_1^2(\delta)$$

where $\delta = t^2\eta_0(1-\eta_0)/\mu_0^2$.

4. (48 points).
Suppose that $X_1, \ldots, X_n$ are independent and identically distributed real-valued random variables with distribution function F and density f.
(a) Consider the sample median $\mathbb{F}_n^{-1}(1/2)$ and the sample mean $\overline{X}_n$. Give conditions under which the sample median $\mathbb{F}_n^{-1}(1/2)$ is an asymptotically linear estimator of the population median $F^{-1}(1/2)$. Identify the influence function $\psi(x)$ and the limiting distribution of $\sqrt{n}(\mathbb{F}_n^{-1}(1/2) - F^{-1}(1/2))$.
(b) Give conditions under which the sample median $\mathbb{F}_n^{-1}(1/2)$ and the sample mean $\overline{X}_n$ have a joint limiting distribution; i.e. conditions which imply that the random vector

$$\begin{pmatrix} \sqrt{n}(\mathbb{F}_n^{-1}(1/2) - F^{-1}(1/2)) \\ \sqrt{n}(\overline{X}_n - \mu) \end{pmatrix}$$

converges in distribution where $\mu = \mu_F = E_F(X_1)$. Find the limiting distribution explicitly.
(c) A simple test for asymmetry of a distribution function is based on the difference of the mean and median: $\gamma(F) \equiv \mu_F - F^{-1}(1/2)$. Note that $\gamma_F = 0$ if F is symmetric about some point, while $\gamma(F)$ is positive for F skewed to the right, and negative for F skewed to the left. Use the results of (b) to find the limiting distribution of

$$\sqrt{n}(\gamma(\mathbb{F}_n) - \gamma(F)) = \sqrt{n}(\overline{X}_n - \mathbb{F}_n^{-1}(1/2) - (\mu_F - F^{-1}(1/2))).$$

Compute the limiting variance in terms of expectations of functions of F. Is $\gamma(\mathbb{F}_n)$ asymptotically linear?

Solution: (a) Assuming that F has a positive density f at $F^{-1}(1/2)$, we know that with $Q(t) \equiv F^{-1}(t)$, so $Q'(t) = 1/f(F^{-1}(t))$,

$$\begin{aligned}
\sqrt{n}(\mathbb{F}_n^{-1}(1/2) - F^{-1}(1/2)) &\to_d -Q'(1/2)\mathbb{U}(1/2) \\
&= -\frac{1}{f(F^{-1}(1/2)}\mathbb{U}(1/2) \sim N(0, \frac{1/4}{f^2(F^{-1}(1/2)}).
\end{aligned}$$

Moreover,

$$\begin{aligned}
\sqrt{n}(\mathbb{F}_n^{-1}(1/2) - F^{-1}(1/2)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{-1}{f(F^{-1}(1/2))}(1_{(-\infty, F^{-1}(1/2)]}(X_i) - 1/2) + o_p(1) \\
&\equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(X_i) + o_p(1)
\end{aligned}$$

where $\psi(x) = -(1_{(-\infty, F^{-1}(1/2)]}(x) - 1/2)/f(F^{-1}(1/2))$ has $E\psi(X) = 0$ and $E\psi^2(X) = (1/4)/f^2(F^{-1}(1/2))$.

(b) If $E(X_1^2) < \infty$, then from the asymptotically linear representation of $\mathbb{F}_n^{-1}(1/2)$ in (a) together with the multivariate central limit theorem it follows that

$$\begin{aligned}
\sqrt{n}\begin{pmatrix} \mathbb{F}_n^{-1}(1/2) - F^{-1}(1/2) \\ \overline{X}_n - \mu_F \end{pmatrix} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{pmatrix} \psi(X_i) \\ X_i - \mu_F \end{pmatrix} + o_p(1) \\
&\to_d \underline{V} \sim N_2(0, \Sigma)
\end{aligned}$$

where $\Sigma_{11} = E\psi^2(X_1) = (1/4)/f^2(F^{-1}(1/2))$, $\Sigma_{22} = \text{Var}(X_1)$, and

$$\Sigma_{12} = \Sigma_{21} = E\psi(X_1)(X_1 - \mu) = -\frac{E\{1_{(-\infty, F^{-1}(1/2)]}(X_1)(X_1 - \mu)\}}{f(F^{-1}(1/2))}.$$

(c) From the joint convergence result in (b) it follows by the continuous mapping theorem that

$$\begin{aligned}
\sqrt{n}(\gamma(\mathbb{F}_n) - \gamma(F)) &= \sqrt{n}(\overline{X}_n - \mathbb{F}_n^{-1}(1/2) - (\mu_F - F^{-1}(1/2))) \\
&= \sqrt{n}(\overline{X}_n - \mu_F) - \sqrt{n}(\mathbb{F}_n^{-1}(1/2) - F^{-1}(1/2)) \\
&\to_d V_2 - V_1 \sim N(0, \tau^2)
\end{aligned}$$

where

$$\begin{aligned}
\tau^2 &\equiv \tau^2(F) \qquad (0.1) \\
&= \frac{1/4}{f^2(F^{-1}(1/2))} + Var_F(X) \\
&\qquad + 2E_F\left\{(X - \mu_F)(1_{(-\infty, F^{-1}(1/2)]}(X) - 1/2)\right\}/f(F^{-1}(1/2)) \\
&= Var_F(X) + \frac{1/4}{f^2(F^{-1}(1/2))} \\
&\qquad + 2E_F\left\{(X - \mu_F)1_{(-\infty, F^{-1}(1/2)]}(X)\right\}/f(F^{-1}(1/2)) \\
&= Var_F(X) + \frac{1/4}{f^2(F^{-1}(1/2))} \\
&\qquad + 2(E_F\left\{X1_{(-\infty, F^{-1}(1/2)]}(X)\right\} - \mu_F)/f(F^{-1}(1/2))\,. \qquad (0.2)
\end{aligned}$$

Indeed, $\gamma(\mathbb{F}_n)$ is asymptotically linear as well: from the asymptotic linearity of $\mathbb{F}_n^{-1}(1/2)$ given in (a) and (b)

$$\sqrt{n}(\gamma(\mathbb{F}_n) - \gamma(F)) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \{X_i - \mu_F - \psi(X_i)\} + o_p(1)$$

where $\psi = -(1_{(-\infty, F^{-1}(1/2)]}(x) - 1/2)/f(F^{-1}(1/2))$ as in (a)

Remarks: This statistic for testing symmetry was suggested by Edgeworth (1887). See the discussion on pages 105 and 106 of Stigler (1999), which also indicates that the joint asymptotic distribution of the mean and median was known to Laplace in the early 1800's.]

Do either problem 5 or problem 6.

5. (48 points).
Suppose that $\underline{N}_n \equiv \sum_{i=1}^n \underline{\Delta}_i \sim \text{Mult}_k(n, \underline{p})$ where $\underline{\Delta}_i \sim \text{Mult}_k(1, \underline{p})$ are independent. The usual MLE $\hat{\underline{p}}_n$ of $\underline{p}$ is $\hat{\underline{p}}_n = n^{-1}\underline{N}_n$. For $a \in [0, 1)$ and $\underline{p}_0 = (1/k, \ldots, 1/k)$, consider the alternative estimator $\widetilde{\underline{p}}_n$ of $\underline{p}$ defined by

$$\widetilde{\underline{p}}_n = \begin{cases} \widehat{\underline{p}}_n & \text{if } \|\widehat{\underline{p}}_n - \underline{p}_0\|_2 > n^{-1/4}, \\ a\widehat{\underline{p}}_n + (1-a)\underline{p}_0 & \text{if } \|\widehat{\underline{p}}_n - \underline{p}_0\|_2 \le n^{-1/4}. \end{cases}$$

Let $\underline{Z} \sim N_k(\underline{0}, \text{diag}(\underline{p}) - \underline{p}\underline{p}^T)$, $\underline{Z}_0 \sim N_k(\underline{0}, \text{diag}(\underline{p}_0) - \underline{p}_0\underline{p}_0^T)$.
(a) Find the limiting distribution of $\sqrt{n}(\widetilde{\underline{p}}_n - \underline{p}_0)$ when $\underline{p} \ne \underline{p}_0$ and when $\underline{p} = \underline{p}_0$. (Express these in terms of $\underline{Z}$ and $\underline{Z}_0$.)
(b) Let $\underline{p}_n = \underline{p}_0 + n^{-1/2}\underline{c}$ where $\underline{1}^T\underline{c} = 0$. Find the limiting distribution of $\sqrt{n}(\widetilde{\underline{p}}_n - \underline{p}_n)$ under $\underline{p}_n$.
(c) Is $\widetilde{\underline{p}}_n$ a locally regular estimator of $\underline{p}$ at $\underline{p} = \underline{p}_0$? Explain why or why not.
(d) Find the limit of $E_{\underline{p}_n} n\|\widetilde{\underline{p}}_n - \underline{p}_n\|^2$ where $\|\underline{v}\|^2 \equiv \sum_{j=1}^k v_j^2$ for $\underline{v} \in \mathbb{R}^k$. How does this compare to the limit of $E_{\underline{p}_n} n\|\widehat{\underline{p}}_n - \underline{p}_n\|^2$?

Solution: (a) Now

$$\begin{aligned}
\sqrt{n}(\widetilde{\underline{p}}_n - \underline{p}) &= \sqrt{n}(\widehat{\underline{p}}_n - \underline{p})1\{\|\widehat{\underline{p}}_n - \underline{p}_0\| > n^{-1/4}\} \\
&\quad + \sqrt{n}(a\widehat{\underline{p}}_n + (1-a)\underline{p}_0 - \underline{p})1\{\|\widehat{\underline{p}}_n - \underline{p}_0\| \le n^{-1/4}\} \\
&= \sqrt{n}(\widehat{\underline{p}}_n - \underline{p})1\{\|\sqrt{n}(\widehat{\underline{p}}_n - \underline{p}) + \sqrt{n}(\underline{p} - \underline{p}_0)\| > n^{+1/4}\} \\
&\quad + \{a\sqrt{n}(\widehat{\underline{p}}_n - \underline{p}) + (1-a)\sqrt{n}(\underline{p}_0 - \underline{p})\} \\
&\qquad \cdot 1\{\|\sqrt{n}(\widehat{\underline{p}}_n - \underline{p}) + \sqrt{n}(\underline{p} - \underline{p}_0)\| \le n^{+1/4}\}.
\end{aligned}$$

If $\underline{p} \ne \underline{p}_0$, then $\|\sqrt{n}(\widehat{\underline{p}}_n - \underline{p}) + \sqrt{n}(\underline{p} - \underline{p}_0)\|n^{-1/4} \to_p \infty$, and hence on a set with probability converging to 1,

$$\sqrt{n}(\widetilde{\underline{p}}_n - \underline{p}) = \sqrt{n}(\widehat{\underline{p}}_n - \underline{p}) \to_d \underline{Z}.$$

On the other hand, if $\underline{p} = \underline{p}_0$, then

$$\|\sqrt{n}(\widehat{\underline{p}}_n - \underline{p}) + \sqrt{n}(\underline{p} - \underline{p}_0)\|n^{-1/4} = \|\sqrt{n}(\widehat{\underline{p}}_n - \underline{p})\|n^{-1/4} \to_p 0,$$

and hence on a set with probability converging to 1,

$$\sqrt{n}(\widetilde{\underline{p}}_n - \underline{p}_0) = a\sqrt{n}(\widehat{\underline{p}}_n - \underline{p}_0) \to_d a\underline{Z}_0.$$

(b) If $\underline{p}_n = \underline{p}_0 + cn^{-1/2}$ with $\underline{1}^T\underline{c} = 0$, then

$$\begin{aligned}
\sqrt{n}(\widetilde{\underline{p}}_n - \underline{p}_n) &= \sqrt{n}(\widehat{\underline{p}}_n - \underline{p}_n)1\{\|\sqrt{n}(\widehat{\underline{p}}_n - \underline{p}_n) + \sqrt{n}(\underline{p}_n - \underline{p}_0)\| > n^{+1/4}\} \\
&\quad + \{a\sqrt{n}(\widehat{\underline{p}}_n - \underline{p}_n) + (1-a)\sqrt{n}(\underline{p}_0 - \underline{p}_n)\} \\
&\qquad \cdot 1\{\|\sqrt{n}(\widehat{\underline{p}}_n - \underline{p}_n) + \sqrt{n}(\underline{p}_n - \underline{p}_0)\| \le n^{+1/4}\} \\
&= \sqrt{n}(\widehat{\underline{p}}_n - \underline{p}_n)1\{\|\sqrt{n}(\widehat{\underline{p}}_n - \underline{p}_n) + \underline{c}\| > n^{+1/4}\} \\
&\quad + \{a\sqrt{n}(\widehat{\underline{p}}_n - \underline{p}_n) - (1-a)\underline{c}\}1\{\|\sqrt{n}(\widehat{\underline{p}}_n - \underline{p}_n) + \underline{c}\| \le n^{+1/4}\} \\
&\to_d a\underline{Z}_0 - (1-a)\underline{c}.
\end{aligned}$$

(c) Since the limiting distribution in (b) depends on $\underline{c}$, $\underline{\widetilde{p}}_n$ is not a locally regular estimator of $\underline{p}$ at $\underline{p}_0$.
(d) Now

$$E_{\underline{p}_n}\{n\|\underline{\widetilde{p}}_n - \underline{p}_n\|^2 \quad \rightarrow \quad E\|a\underline{Z}_0 - (1-a)\underline{c}\|^2 = a^2 E\|\underline{Z}_0\|^2 + (1-a)^2\|\underline{c}\|^2$$
$$\begin{cases} = E\|\underline{Z}_0\|^2 & \text{if } \ a = 1 \\ > E\|\underline{Z}_0\|^2 & \text{if } \ \|\underline{c}\|^2 > \frac{1+a}{1-a} \\ = a^2 E\|\underline{Z}_0\|^2 < E\|\underline{Z}_0\|^2 & \text{if } |a| < 1 \text{ and } \underline{c} = 0. \end{cases}$$

Thus even though $\underline{\widetilde{p}}_n$ has smaller mean-squared error than $\widehat{\underline{p}})n$ at $\underline{p}_0$, it has larger mean squared error as soon as $\|\underline{c}\| > \sqrt{(1+a)/(1-a)}$.

6. (48 points).
Consider a parametric model $\mathcal{P} = \{P_\theta : \ \theta \in \Theta \subset \mathbb{R}^d\}$ satisfying the hypotheses A0-A4 of section 4.1 of the Chapter 4 notes. Suppose that we are using the Rao (or score statistic) $R_n = \underline{Z}_n(\theta_0)^T I(\theta_0)^{-1}\underline{Z}_n(\theta_0)$ for testing the null hypothesis $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$ where $\underline{Z}_n(\theta_0) \equiv n^{-1/2}\sum_{i=1}^n \dot{\underline{l}}_\theta(\theta_0|X_i)$.
(a) Suppose that $\theta_n = \theta_0 + tn^{-1/2}$ and consider $l_n(\theta_n) - l_n(\theta_0) = \sum_{i=1}^n \log\{p_{\theta_n}(X_i)/p_{\theta_0}(X_i)\}$. Use the expansion developed in HW (and stated in part (vi) of Theorem 4.1.2) which implies LAN at θ_0, together with the linearity of $\underline{Z}_n(\theta_0)$, to find the joint limit distribution of

$$\begin{pmatrix} \underline{c}^T\underline{Z}_n(\theta_0) \\ l_n(\theta_n) - l_n(\theta_0) \end{pmatrix} = \begin{pmatrix} \underline{c}^T\underline{Z}_n(\theta_0) \\ \log\left(\frac{dP^n_{\theta_n}}{dP^n_{\theta_0}}\right) \end{pmatrix}$$

under P_{θ_0}.
(b) Use the result of (a) together with Le Cam's 3rd lemma to find the joint limiting distribution of

$$\begin{pmatrix} \underline{c}^T\underline{Z}_n(\theta_0) \\ l_n(\theta_n) - l_n(\theta_0) \end{pmatrix} = \begin{pmatrix} \underline{c}^T\underline{Z}_n(\theta_0) \\ \log\left(\frac{dP^n_{\theta_n}}{dP^n_{\theta_0}}\right) \end{pmatrix}$$

under P_{θ_n}.
(c) Use the result of (b) to find the limiting distribution of $\underline{Z}_n(\theta_0)$ under P_{θ_n}.
(d) Use the result of (c) to find the limiting distribution of the Rao statistic R_n under P_{θ_n}. What does this imply about the power of the Rao statistic?

Solution: (a) From HW 9 and Theorem 4.1.2 (vi),

$$l_n(\theta_n) - l_n(\theta_0) = \underline{t}^T\underline{Z}_n(\theta_0) - \frac{1}{2}\underline{t}^T I(\theta_0)\underline{t} + o_p(1)$$

under P_{θ_0}. Thus for any $\underline{c} \in \mathbb{R}^d$, with $\underline{Z} \sim N_d(0, I(\theta_0))$,

$$\begin{aligned} \begin{pmatrix} \underline{c}^T\underline{Z}_n(\theta_0) \\ l_n(\theta_n) - l_n(\theta_0) \end{pmatrix} &= \begin{pmatrix} \underline{c}^T\underline{Z}_n(\theta_0) \\ \underline{t}^T\underline{Z}_n(\theta_0) - \frac{1}{2}\underline{t}^T I(\theta_0)\underline{t} \end{pmatrix} + o_p(1) \\ &\rightarrow_d \begin{pmatrix} \underline{c}^T\underline{Z} \\ \underline{t}^T\underline{Z} - \frac{1}{2}\underline{t}^T I(\theta_0)\underline{t} \end{pmatrix} \\ &\sim N_2\left(\begin{pmatrix} 0 \\ -\frac{1}{2}\underline{t}^T I(\theta_0)\underline{t} \end{pmatrix}, \begin{pmatrix} \underline{c}^T I(\theta_0)\underline{c} & \underline{c}^T I(\theta_0)\underline{t} \\ \underline{c}^T I(\theta_0)\underline{t} & \underline{t}^T I(\theta_0)\underline{t} \end{pmatrix}\right). \end{aligned}$$

(b) By Le Cam's third lemma it follows that under P_{θ_n}

$$\begin{pmatrix} \underline{c}^T \underline{Z}_n(\theta_0) \\ l_n(\theta_n) - l_n(\theta_0) \end{pmatrix} \to_d \begin{pmatrix} \underline{c}^T \underline{Z} + \underline{c}^T I(\theta_0)\underline{t} \\ \underline{t}^T \underline{Z} + \frac{1}{2}\underline{t}^T I(\theta_0)\underline{t} \end{pmatrix}$$
$$\sim N_2\left(\begin{pmatrix} 0 \\ \frac{1}{2}\underline{t}^T I(\theta_0)\underline{t} \end{pmatrix}, \begin{pmatrix} \underline{c}^T I(\theta_0)\underline{c} & \underline{c}^T I(\theta_0)\underline{t} \\ \underline{c}^T I(\theta_0)\underline{t} & \underline{t}^T I(\theta_0)\underline{t} \end{pmatrix} \right).$$

(c) In particular, it follows from (b) that under P_{θ_n},

$$\underline{c}^T \underline{Z}_n(\theta_0) \to_d \underline{c}^T \underline{Z} + \underline{c}^T I(\theta_0)\underline{t} = \underline{c}^T(\underline{Z} + I(\theta_0)\underline{t}).$$

Since this holds for every $\underline{c} \in \mathbb{R}^d$ it follows from the Cramér - Wold device that

$$\underline{Z}_n(\theta_0) \to_d \underline{Z} + I(\theta_0)\underline{t}.$$

(d) By (c), when P_{θ_n} is true, the Rao statistic R_n satisfies

$$R_n = \underline{Z}_n(\theta_0)^T I(\theta_0)^{-1} \underline{Z}_n \to_d (\underline{Z} + I(\theta_0)\underline{t}) I^{-1}(\theta_0)(\underline{Z} + I(\theta_0)\underline{t}) \sim \chi_d^2(\delta)$$

where $\delta = \underline{t}^T I(\theta_0) I^{-1}(\theta_0) I(\theta_0)\underline{t} = \underline{t}^T I(\theta_0)\underline{t}$. Thus for these alternatives, the power of the Rao test satisfies

$$P_{\theta_n}(R_n \geq \chi^2_{d,\alpha}) \to P(\chi_d^2(\delta) \geq \chi^2_{d,\alpha}) > \alpha$$

if $\delta > 0$.

7. (50 points).
Suppose that $P = P_0 = N(0,1)$, $Q = P_\theta = N(\theta, 1)$ on $(\mathbb{X}, \mathcal{A}) = (\mathbb{R}, \mathcal{B})$.
(a) Compute $K(P,Q) = K(P_0, P_\theta)$.
(b) Compute $H^2(P,Q) = 1 - \rho(P,Q)$ and $\rho(P,Q) = \int \sqrt{p(x)q(x)}dx$. [It might be easiest to compute $\rho(P,Q)$ first recalling that if $Z \sim N(0,1)$ then $E\exp(tZ) = \exp(t^2/2)$.]
(c) Compute $d_{TV}(P,Q) = 1 - \eta(P,Q)$ and $\eta(P,Q) = \int p(x) \wedge q(x)dx$. [It might be easiest to compute $\eta(P,Q)$ first.]
(d) Show in general that $K(P,Q) \geq 2H^2(P,Q)$, thereby strengthening the fact $K(P,Q) \geq 0$ that we proved in class. [Hint: write both $K(P,Q)$ and $H^2(P,Q)$ in terms of $Y = (p(X)/q(X))^{1/2}$ and use the inequality $\log(1+x) \geq x/(1+x)$ for $x \geq 0$. You will need to relate $E_Q Y$ and $E_Q Y^2$ to $H^2(P,Q)$.]
(e) Use the results of (a) and (d) to find a lower bound for $K(P^n, Q^n)$ in terms of $H^2(P,Q)$ or $\rho(P,Q)$; here P^n and Q^n are the probability distributions of $X_1, \ldots, X_n$ i.i.d. as P and Q^n respectively.

Solution: (a) Now $p(x) = \phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ and $q(x) = (2\pi)^{-1/2}\exp(-(x-\theta)^2/2)$, so

$$\frac{p(x)}{q(x)} = \exp(-x^2/2 + (x-\theta)^2/2) = \exp(-\theta x + \theta^2/2),$$
$$\log\frac{p}{q}(x) = -\theta x + \theta^2/2,$$

and it follows that

$$K(P,Q) = E_P\left(\log \frac{p}{q}\right) = -\theta E_P(X) + \theta^2/2 = \theta^2/2\,.$$

(b) We compute $\rho(P,Q)$ first:

$$\begin{aligned}
\rho(P,Q) = \int \sqrt{p(x)q(x)}dx &= \int \frac{1}{\sqrt{2\pi}} \exp(-x^2/4)\exp(-(x-\theta)^2/4)dx \\
&= \int \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)\exp((2\theta x - \theta^2)/4)\,dx \\
&= \exp(-\theta^2/4)E_P \exp(\theta X/2) = \exp(-\theta^2/4)\exp(\theta^2/8) = \exp(-\theta^2/8)\,.
\end{aligned}$$

Hence

$$H^2(P,Q) = 1 - \rho(P,Q) = 1 - \exp(-\theta^2/8)\,.$$

(c) We compute $\eta(P,Q)$ first. Since $\phi(x) \ge \phi(x-\theta)$ if and only if

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \ge \frac{1}{\sqrt{2\pi}} \exp(-(x-\theta)^2/2)$$

or equivalently, if and only if

$$1 \ge \exp(\theta x - \theta^2/2) \qquad \text{iff} \qquad \theta x - \theta^2/2 \le 0$$

and, for $\theta > 0$, this holds if and only if $x \le \theta/2$. Hence it follows that

$$\begin{aligned}
\int p(x) \wedge q(x)dx &= \int \phi(x) \wedge \phi(x-\theta)dx \\
&= \int_{-\infty}^{\theta/2} \phi(x-\theta)dx + \int_{\theta/2}^{\infty} \phi(x)dx \\
&= \int_{-\infty}^{-\theta/2} \phi(y)dy + 1 - \Phi(\theta/2) \\
&= \Phi(-\theta/2) + 1 - \Phi(\theta/2)\,, \qquad \text{if } \ \theta > 0 \\
&= 2\Phi(-|\theta|/2), \quad \text{if } \theta > 0.
\end{aligned}$$

When $\theta < 0$, $\theta x - \theta^2/2 < 0$ if and only if $x \ge \theta/2$, and this yields

$$\begin{aligned}
\int p(x) \wedge q(x)dx &= \int_{\theta/2}^{\infty} \phi(x-\theta)dx + \int_{-\infty}^{\theta/2} \phi(x)dx \\
&= \Phi(\theta/2) + 1 - \Phi(-\theta/2)\,, \qquad \text{if } \ \theta > 0 \\
&= 2\Phi(-|\theta|/2), \quad \theta < 0.
\end{aligned}$$

It follows that $\eta(P,Q) = 2\Phi(-|\theta|/2)$ for all θ, and hence

$$d_{TV}(P,Q) \;=\; 1 - \eta(P,Q) = 1 - 2\Phi(-|\theta|/2).$$

The following Figure shows $K(P_0,P_\theta)$, $H^2(P_0,P_\theta)$, and $d_{TV}(P_0,P_\theta)$ as functions of θ.

(d) Let $Y \equiv \sqrt{p/q}$. Then

$$\begin{aligned} K(P,Q) &= 2\int p\log(\sqrt{p/q})d\mu = 2E_P \log Y = 2E_P\log(1+(Y-1)) \\ &\ge 2E_P \frac{Y-1}{1+Y-1} \qquad \text{using } \log(1+x) \ge \frac{x}{1+x}, \\ &= 2E_P\frac{Y-1}{Y} = 2\int p\left\{\sqrt{\frac{p}{q}}-1\right\}\sqrt{\frac{q}{p}}d\mu \\ &= 2\int p\left\{1-\sqrt{\frac{q}{p}}\right\}d\mu = 2\left\{1-\int\sqrt{pq}d\mu\right\} \\ &= 2H^2(P,Q). \end{aligned}$$

Here is a another way of organizing the argument with a slightly different choice of Y, as follows. Let $Y \equiv \sqrt{p/q}-1$. Then $p/q = (1+Y)^2$ and it follows that

$$\begin{aligned} K(P,Q) &= \int p\log(p/q)d\mu = \int (p/q)\log(p/q)qd\mu \\ &= 2\int (p/q)\log(p/q)^{1/2}dQ = 2\int (1+Y)^2\log(1+Y)dQ \\ &\ge 2\int (1+Y)^2\frac{Y}{1+Y}dQ \\ &= 2\int Y(1+Y)dQ = 2\left\{\int YdQ + \int Y^2dQ\right\}. \end{aligned}$$

But we also have

$$\begin{aligned} \int YdQ &= \int \sqrt{pq}d\mu - 1 = -H^2(P,Q)\,, \\ \int Y^2dQ &= \int[\sqrt{p/q}-1]^2qd\mu = \int[\sqrt{p}-\sqrt{q}]^2d\mu = 2H^2(P,Q)\,. \end{aligned}$$

By combining the results of these last two displays it follows that

$$K(P,Q) \ge 2\left\{2H^2(P,Q) - H^2(P,Q)\right\} = 2H^2(P,Q)\,.$$

(e) By (d), and (c) of the next problem,

$$K(P^n,Q^n) \ge 2H^2(P^n,Q^n) = 2\{1-\rho(P^n,Q^n)\} = 2\{1-\rho(P,Q)^n\}$$

in general. For the particular case we began with $\rho(P,Q) = \exp(-\theta^2/8)$, and thus we conclude that in this case

$$K(P^n,Q^n) \ge 2\{1-\exp(-n\theta^2/8)\}\,.$$

Another lower bound follows from the identity $K(P^n,Q^n) = nK(P,Q)$: we conclude from this together with the inequality from (d) that

$$K(P^n,Q^n) = nK(P,Q) \ge n2H^2(P,Q) = 2n(1-\exp(-\theta^2/8)).$$

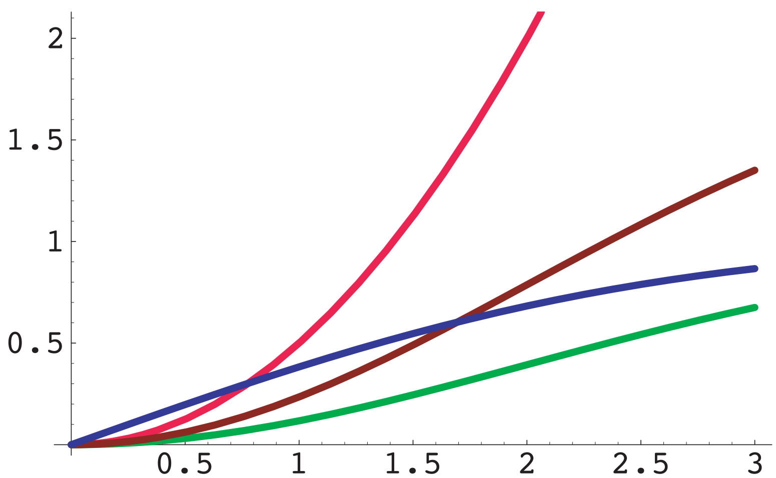

Figure 1: $K(P_0, P_\theta)$ (red), $H^2(P_0, P_\theta)$ (green), $d_{TV}(P_0, P_\theta)$ (blue), and $2H^2(P_0, P_\theta)$ (burgundy) as functions of θ
.