Statistics 583, Midterm Exam Solutions

Wellner; 5/4/2016

1. (30 points) **Define** *each* of the following terms.
   (a) The bracketing number $N_{[\,]}(\epsilon, \mathcal{F}, L_1(P))$ of a class of real-valued functions $\mathcal{F}$ with respect to $L_1(P)$ on a general sample space $(\Omega, \mathcal{A}, P)$.
   (b) A $P-$Glivenko-Cantelli class of functions $\mathcal{F}$ on a general sample space $(\mathcal{X}, \mathcal{A})$.
   (c) Give a sufficient condition for a class of functions $\mathcal{F}$ to satisfy the $P-$Glivenko -Cantelli property described in (b).

   Solution: (a) Given $\epsilon > 0$ the bracketing number $N_{[\,]}(\epsilon, \mathcal{F}, L_1(P))$ is the minimal number of $\epsilon-$brackets needed to cover $\mathcal{F}$. Here a bracket is $[l, u] \equiv \{f \in \mathcal{F} : \ l(x) \le f(x) \le u(x) \text{ for all } \ x \in \mathcal{X}\}$ where l and u are two real-valued (measurable) functions on $(\mathcal{X}, \mathcal{A})$ such that $l \le u$ and it is of size ϵ if $P(u - l) = \int_{\mathcal{X}} (u - l) dP < \epsilon$.
   (b) $\mathcal{F}-$ is a $P-$Glivenko - Cantelli class of functions if

$$\|\mathbb{P}_n - P\|_{\mathcal{F}}^* = \left(\sup_{f \in \mathcal{F}} |\mathbb{P}_n(f) - P(f)| \right)^* \to_{a.s.} 0$$

   as $n \to \infty$ where $\mathbb{P}_n$ is the empirical measure of $X_1, \ldots, X_n$ i.i.d. P on $(\mathcal{X}, \mathcal{A})$.
   (c) A sufficient condition for $\mathcal{F}$ to be a $P-$Glivenko-Cantelli class of functions is $N_{[\,]}(\epsilon, \mathcal{F}, L_1(P)) < \infty$ for every $\epsilon > 0$.

2. (30 points) Give a complete **statement** of *any three* of the following results:
   (a) Wald's consistency theorem for maximum likelihood estimators.
   (b) Le Cam's (two-sided) Glivenko-Cantelli theorem related to Wald's consistency theorem.
   (c) The Cramér-Rao bound for the variance under P_{θ_0} of an unbiased estimator of $\nu(P_\theta) = q(\theta)$ for a regular parametric model under the assumption that q is differentiable at θ_0. Relate this bound to the efficient influence function $\tilde{l}_\nu$ for $\nu(P_\theta)$ (at θ_0).
   (d) A result about the joint asymptotic normality of the sample quantiles $(\mathbb{F}_n^{-1}(t_1), \ldots, \mathbb{F}_n^{-1}(t_k))$ where $0 < t_1 < \cdots < t_k < 1$.
   (e) An inequality relating Hellinger distance $H(P, Q)$ (or Hellinger affinity $\rho(P, Q)$) to the total variation distance $V(P, Q)$ (or the total variation affinity $\eta(P, Q)$) and an equality relating $\rho(P^n, Q^n)$ to $\rho(P, Q)$.
   (f) The Lindeberg-Feller central limit theorem (including context).

   Solution: See course notes.

3. (40 points) Let $\widehat{p}_n$ be the maximum likelihood estimator of $p_0 \in \mathcal{P}$ where $\mathcal{P}$ is some class of densities on $(\mathcal{X}, \mathcal{A})$ with respect to a fixed dominating measure μ.
(a) State any "basic inequality" giving a bound for $H^2(\hat{p}_n, p_0)$ in terms of $(\mathbb{P}_n - P_0)(g_{\widehat{p}_n})$ for some function $x \mapsto g_p(x)$ for $p \in \mathcal{P}$.
(b) Use the bound you state in (a) to give a proof that $H^2(\widehat{p}_n, p_0) \to_{a.s.} 0$ assuming that some appropriate class of functions is a P_0-Glivenko-Cantelli class.

Solution: (a) If $\widehat{p}_n$ is the MLE of p_0 over a class $\mathcal{P}$ of densities, then a very simple basic inequality is

$$H^2(\widehat{p}_n, p_0) \le (\mathbb{P}_n - P_0)(g_{\widehat{p}_n})$$

where

$$g_p(x) \equiv \left(\sqrt{\frac{p(x)}{p_0(x)}} - 1\right) 1_{[p_0(x)>0]}.$$

(b) If the class of functions $\mathcal{F} = \{g_p : \ p \in \mathcal{P}\}$ is $P-$Glivenko - Cantelli, then it follows that

$$H^2(\widehat{p}_n, p_0) \le \sup_{p \in \mathcal{P}} (\mathbb{P}_n - P_0)(g_p) \to_{a.s.} 0.$$

4. (40 points) Let $X \sim N(0, \sigma^2)$ under P and suppose that $X \sim N(\mu, \sigma^2)$ with $\mu > 0$ under Q.
(a) Compute $(q/p)(X)$ and $\log((q/p)(X))$ and use this to compute $K(Q, P)$ and $K(P, Q)$. Do your computations yield $K(P, Q) \ge 0$ and $K(Q, P) \ge 0$?
(b) Compute $\rho(P, Q) = \int \sqrt{pq} d\mu$ where μ denotes Lebesgue measure on $\mathbb{R}$. Use this to compute $H^2(P, Q) = (1/2) \int (\sqrt{p} - \sqrt{q})^2 d\mu$.
(c) How does $H^2(P, Q)$ relate to $K(P, Q)$ or $K(Q, P)$? State a simple inequality relating $H^2(P, Q)$ to $K(P, Q)$ or $K(Q, P)$ and verify that it holds in the special case in (a) and (b).

Solution: (a) Now Q has density

$$\begin{aligned} q(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2} + \frac{\mu x - (1/2)\sigma^2}{\sigma^2}\right) \\ &= p(x) \cdot \exp\left(\frac{\mu x - (1/2)\sigma^2}{\sigma^2}\right). \end{aligned}$$

Thus

$$\frac{q(x)}{p(x)} = \exp\left(\frac{\mu x - (1/2)\sigma^2}{\sigma^2}\right), \quad \text{and} \quad \log \frac{q(x)}{p(x)} = \frac{\mu x - (1/2)\mu^2}{\sigma^2}.$$

Thus we compute

$$\begin{aligned} K(Q, P) &= E_Q \log\left(\frac{q(X)}{p(X)}\right) = E_Q \left(\frac{\mu X - (1/2)\mu^2}{\sigma^2}\right) \\ &= \frac{\mu E_Q(X) - (1/2)\mu^2}{\sigma^2} = \frac{\mu^2}{2\sigma^2}, \end{aligned}$$

and

$$
\begin{aligned}
K(P,Q) &= E_P \log\left(\frac{p(X)}{q(X)}\right) = E_P\left(-\frac{\mu X + (1/2)\mu^2}{\sigma^2}\right) \\
&= \frac{-\mu E_P(X) + (1/2)\mu^2}{\sigma^2} = \frac{\mu^2}{2\sigma^2}.
\end{aligned}
$$

Thus in this (special!) case we have $K(Q,P) = K(P,Q) = \mu^2/(2\sigma^2) > 0$ if $\mu \neq 0$.
(b) We have

$$
\begin{aligned}
\rho(P,Q) &= \int_{-\infty}^{\infty} \sqrt{p(x)q(x)}dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{4\sigma^2}\right) \exp\left(-\frac{(x-\mu)^2}{4\sigma^2}\right) dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(x^2 - \mu x + 2^{-1}\mu^2\right)\right) dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left((x-\mu/2)^2 - 4^{-1}\mu^2 + 2^{-1}\mu^2\right)\right) \\
&= \exp\left(-\frac{\mu^2}{8\sigma^2}\right).
\end{aligned}
$$

Alternatively,

$$
\begin{aligned}
\rho(P,Q) &= \int \sqrt{p(x)q(x)}dx = \int \sqrt{(q/p)(x)}p(x)dx = E_P(q/p)^{1/2}(X) \\
&= E_P \exp\left(\frac{\mu X - (1/2)\mu^2}{2\sigma^2}\right) = \exp\left(-\frac{\mu^2}{4\sigma^2}\right) E_P \exp(rX) \quad \text{with} \quad r \equiv \mu/(2\sigma^2) \\
&= \exp\left(-\frac{\mu^2}{4\sigma^2}\right) \exp(r^2\sigma^2/2) = \exp\left(-\frac{\mu^2}{4\sigma^2}\right) \exp(\mu^2/(8\sigma^2)) \\
&= \exp\left(-\frac{\mu^2}{8\sigma^2}\right)
\end{aligned}
$$

Thus

$$
H^2(P,Q) = 1 - \rho(P,Q) = 1 - \exp\left(-\frac{\mu^2}{8\sigma^2}\right) \le \frac{\mu^2}{8\sigma^2} < \frac{1}{2}\frac{\mu^2}{2\sigma^2} = \frac{1}{2}K(P,Q)
$$

where we used $1 - e^{-x} \le x$ to get the first inequality. This verifies the general inequality $K(P,Q) \ge 2H^2(P,Q)$ proved in class.

5. (40 points) Suppose that $X \sim$ Weibull (α, β); that is, $P_\theta(X > x) = \exp(-(x/\alpha)^\beta)$ for $x \ge 0$.
(a) We showed in class that $V = (X/\alpha)^\beta \sim$ Exponential(1). Show that $\epsilon \equiv \log V$ has the double exponential (or Gumbel or extreme value) distribution function $G(x) = 1 - \exp(-e^x)$ for $x \in \mathbb{R}$.
(b) Show that $Y \equiv \log X \stackrel{d}{=} \log(\alpha) + \beta^{-1}\epsilon$ where $\epsilon \equiv \log V$ is as in (a). Thus, on

the log scale, $\log\alpha$ is a location parameter and $1/\beta$ is a scale parameter.
(c) The information matrix for θ that we computed in class is

$$I(\theta) = \begin{pmatrix} \beta^2/\alpha^2 & a/\alpha \\ a/\alpha & b^2/\beta^2 \end{pmatrix}$$

where $a = -(1-\gamma)$ and $b^2 = \pi^2/6 + (1-\gamma)^2$ and $\gamma \equiv .577216...$ is Euler's constant. Compute the Cramér-Rao lower bound for unbiased estimators of $q(\theta) = \log\alpha$ and $q(\theta) = 1/\beta$.
(d) How would you use $\overline{Y}_n = \overline{\log X}$ and $S_Y^2 \equiv n^{-1}\sum_{i=1}^n (Y_i - \overline{Y}_n)^2$ to estimate $(\log\alpha, 1/\beta)$? [Hint: $E(\log V) = -\gamma$ and $Var(\log V) = \pi^2/6$.]

Solution: (a) $V \equiv (X/\alpha)^\beta \sim$ Exponential(1). Thus the survival function $1-G$ of $\epsilon \equiv \log V$ is given by

$$1 - G(y) = P(\log V \geq y) = P(V \geq e^y) = \exp(-e^y).$$

Thus $G(y) = 1 - \exp(-e^y)$ for $y \in \mathbb{R}$.
(b) Since $V \equiv (X/\alpha)^\beta \sim$ Exponential(1), we have

$$\epsilon = \log V = \log\{(X/\alpha)^\beta\} = \beta(\log X - \log\alpha) = \beta\log X - \beta\log\alpha.$$

Thus we find that

$$\beta\log X = \beta\log\alpha + \epsilon, \quad \text{or} \quad \log X = \log\alpha + \beta^{-1}\epsilon.$$

Thus we see that for $Y \equiv \log X$, $\log\alpha$ is a location parameter and β^{-1} is a scale parameter.
(c) With $q(\theta) = \log\alpha$ we have $\dot{q}(\theta) = (1/\alpha, 0)^T$ and hence the C-R lower bound for unbiased estimates of $\log\alpha$ is $n^{-1}\dot{q}(\theta)^T I(\theta)^{-1}\dot{q}(\theta) = n^{-1}\alpha^{-2}I^{11}$ where $I^{11} = I_{11\cdot 2}^{-1}$ Now

$$I_{11\cdot 2} = I_{11} - I_{12}I_{22}^{-1}I_{21} = (\beta^2/\alpha^2) - (a/\alpha)^2(\beta^2/b^2) = (\beta^2/\alpha^2)(1 - b^2/(b^2-a^2))$$

and hence

$$n^{-1}\alpha^{-2}I^{11} = \alpha^{-2}(\alpha^2/\beta^2)\frac{b^2}{b^2-a^2} = \beta^{-2}\frac{b^2}{b^2-a^2}$$

where $b^2/(b^2-a^2) = (\pi^2/6 + (1-\gamma)^2)/(\pi^2/6)$. Similarly, for $q(\theta) = \beta^{-1}$, $\dot{q}(\theta) = (0, -\beta^{-2})^T$, and hence the information bound for unbiased estimators of $q(\theta) = 1/\beta$ is

$$n^{-1}\dot{q}(\theta)^T I(\theta)^{-1}\dot{q}(\theta) = n^{-1}\beta^{-4}I^{22}$$

where $I^{22} = I_{22\cdot 1}^{-1}$. Now

$$I_{22\cdot 1} = I_{22} - I_{21}I_{11}^{-1}I_{12} = \frac{b^2}{\beta^2} - \frac{a^2}{\alpha^2}\cdot\frac{\alpha^2}{\beta^2} = \beta^{-2}(b^2-a^2) = \beta^{-2}\frac{\pi^2}{6}.$$

Thus

$$n^{-1}\beta^{-4}I^{22} = n^{-1}\beta^{-4}\beta^2\frac{6}{\pi^2} = n^{-1}\beta^{-2}\frac{6}{\pi^2}.$$

(d) Now if the Weibull model holds, then

$$\overline{Y}_n \to_p E(Y) = \log\alpha + \beta^{-1}E(\epsilon) = \log\alpha + \beta^{-1}c$$

where $c \equiv -\gamma$, and

$$S_Y^2 \to_p \mathrm{Var}(Y) = \beta^{-2}\mathrm{Var}(\epsilon) = \beta^{-2}d^2 = \beta^{-2}\pi^2/6.$$

We conclude that $\widehat{\beta}_n^2 \equiv d^2/S_Y^2$ satisfies $\widehat{\beta}_n \to_p d/(d/\beta) = \beta$, and then

$$\widehat{\log\alpha} \equiv \overline{Y}_n - c/\widehat{\beta}_n \to \log\alpha + \frac{c}{\beta} - \frac{c}{\beta} = \log\alpha.$$