Statistics 581, Final Exam Solutions

Wellner; 12/15/97

1. (30 points) **Define** the following terms. In each case, provide an appropriate context for your definition.
(a) The *Kullback-Leibler* information between a probability measure P and another (sub-)probability measure Q on the same measurable space $(\mathcal{X}, \mathcal{A})$.
(b) The (vector) of *score function(s)* for a sample of size one in a regular parametric model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ with $\Theta \subset R^k$.
(c) The *tangent space* $\dot{\mathcal{P}}$ of a regular parametric model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ with $\Theta \subset R^k$.
(d) The information matrix for a sample of size one in a regular parametric model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ with $\Theta \subset R^k$.
(e) An *asymptotically linear estimator* with influence function ψ.
Solution: see Chapters 3 and 4.

2. (24 points) **State** *any three* of the following results:
(a) The Lindeberg - Feller Central limit theorem.
(b) The Liapunov central limit theorem.
(c) Donsker's theorem for the uniform empirical process $\mathbb{U}_n \equiv \sqrt{n}(\mathbb{G}_n - I)$.
(d) The Cramér - Rao inequality for an unbiased estimator T_n of a parameter $\nu(P_\theta) = q(\theta)$.
(e) The g' theorem or delta-method for a random vector Z_n with $X_n \equiv a_n(Z_n - b) \rightarrow_d X$ in R^m and where $g : R^m \rightarrow R^k$.
(f) A limit central limit theorem for a vector of sample quantiles $(\mathbb{F}_n^{-1}(t_1), \ldots, \mathbb{F}_n^{-1}(t_k))$ for fixed numbers $0 < t_1 < \ldots < t_k < 1$.
Solution: See Chapters 2 and 3.

Do **either** problem 3 **or** problem 4.

3. (30 points) (a) **Prove** the *Glivenko-Cantelli theorem* for the empirical distribution function $\mathbb{G}_n$ of n Uniform$(0,1)$ random variables $\xi_1, \ldots, \xi_n$: if $\mathbb{G}_n(t) = n^{-1} \sum_{i=1}^n 1_{[0,t]}(\xi_i)$, then

$$\|\mathbb{G}_n - I\|_\infty \equiv \sup_{0 \le t \le 1} |\mathbb{G}_n(t) - t| \to_{a.s} 0.$$

(b) State the Glivenko-Cantelli theorem for the empirical distribution function $\mathbb{F}_n$ of a sample $X_1, \ldots, X_n$ i.i.d. with distribution function F.
(c) Use (a) to **prove** the statement in (b).

Solution: See chapter 2.

4. (30 points) (a) State the Elementary Skorokhod theorem.
(b) State Fatou's lemma.
(c) Use the Skorokhod theorem combined with Fatou's lemma to prove the following: if $X_n \to_d X_0$, then $E|X_0| \le \liminf_{n\to\infty} E|X_n|$.
(d) Use (b) to prove that if $X_n \to_d X_0$ then $Var(X_0) \le \liminf_{n\to\infty} Var(X_n)$.

Solution: See Chapter 2 for (a) - (c). For (d), use the identity $Var(X) = \frac{1}{2}E(X-Y)^2$, where $Y =_d X$ and X and Y are independent, twice in combination with (a)-(c).

5. (42 points) Suppose that $X \sim$ Beta(α, β); i.e. as in HW problem 8.3, X has density

$$p_\theta(x) \equiv p_{\alpha,\beta}(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} 1_{(0,1)}(x).$$

Let $\mathcal{P} = \{P_\theta : \ \theta = (\alpha, \beta) \in R^{+2}$.
(a) Compute the scores for α and β based on one X.
(b) Compute the information matrix for $\theta = (\alpha, \beta)$ based on one observation X.
(c) Is this an exponential family? If so, identify the various components of the family?
(d) Compute $E_\theta \log X$ and $E_\theta \log(1-X)$. Also compute $Var_\theta(\log X)$ and $Var_\theta(\log(1-X)$. [Hint: use (a) and (b).]
(e) What is the information for α if β is known? What is the information for α if β is unknown? Draw a picture of the scores to illustrate this geometrically.
(f) Consider estimation of the parameter $q(\theta) = \psi(\alpha) - \psi(\beta)$ based on $X_1, \ldots, X_n$ i.i.d. with distribution $P_{\theta_0} \in \mathcal{P}$. What is the efficient influence function for estimation of $q(\theta)$? [You only need to do this generically; no explicit detailed computation required!]
(g) Propose a nonparametric method of moments estimator of the parameter $q(\theta)$ in (f) and find its influence function ψ. Does your estimator have an influence

function in the tangent space $\dot{\mathcal{P}}$? [Hint: recall (d).]

Solution: (a) If $X \sim$Beta(α, β), then

$$\log p_\theta(x) = \log\Gamma(\alpha+\beta) - \log\Gamma(\alpha) - \log\Gamma(\beta) + (\alpha-1)\log x + (\beta-1)\log(1-x),$$

and

$$\dot{l}_\alpha(x) = \log x - (\psi(\alpha) - \psi(\alpha+\beta)),$$
$$\dot{l}_\beta(x) = \log(1-x) - (\psi(\beta) - \psi(\alpha+\beta)).$$

(b) Hence it follows that

$$\ddot{l}_{\alpha\alpha}(x) = -\psi'(\alpha) + \psi'(\alpha+\beta),$$
$$\ddot{l}_{\beta\beta}(x) = -\psi'(\beta) + \psi'(\alpha+\beta),$$
$$\ddot{l}_{\alpha\beta}(x) = \psi'(\alpha+\beta),$$

which are all constant in x. Hence the information matrix is

$$I(\theta) = I(\alpha, \beta) = \begin{pmatrix} \psi'(\alpha) - \psi'(\alpha+\beta) & -\psi'(\alpha+\beta) \\ -\psi'(\alpha+\beta) & \psi'(\beta) - \psi'(\alpha+\beta) \end{pmatrix}.$$

(c) This is an exponential family because it can be written as

$$\begin{aligned} p_\theta(x) \equiv p_{\alpha,\beta}(x) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{-1}(1-x)^{-1}\exp(\alpha\log x + \beta\log(1-x))\, 1_{(0,1)}(x) \\ &= c(\theta)h(x)\exp(\sum_{j=1}^{2}\theta_j T_j(x)) \end{aligned}$$

where $\theta_1 \equiv \alpha$, $\theta_2 = \beta$, $T_1(x) = \log x$, $T_2(x) = \log(1-x)$.
(d) Since the scores have mean zero, it follows from (a) that

$$E_\theta \log X = (\psi(\alpha) - \psi(\alpha+\beta))$$

and

$$E_\theta \log(1-X) = (\psi(\beta) - \psi(\alpha+\beta)).$$

Then it follows from (b) that

$$Var_\theta(\log X) = E_\theta(\dot{l}_\alpha(X)^2) = I_{\alpha\alpha} = \psi'(\alpha) - \psi'(\alpha+\beta)$$

and

$$Var_\theta(\log(1-X)) = E_\theta(\dot{l}_\beta(X)^2) = I_{\beta\beta} = \psi'(\beta) - \psi'(\alpha+\beta).$$

(e) If β is known, the information for α is $I_{\alpha\alpha} = \psi'(\alpha) - \psi'(\alpha+\beta)$. If β is unknown, the information for α is

$$\begin{aligned} I_{\alpha\alpha\cdot\beta} &= I_{\alpha\alpha} - I_{\alpha\beta} I_{\beta\beta} I_{\beta\alpha} \\ &= I_{\alpha\alpha}\left(1 - \frac{I_{\alpha\beta}^2}{I_{\alpha\alpha} I_{\beta\beta}}\right) < I_{\alpha\alpha}. \end{aligned}$$

The geometry of this is shown below.

(f) The efficient influence function for estimation of $q(\theta)$ is $\dot{q}(\theta_0)^T I(\theta_0)^{-1} \dot{l}_\theta(x)$ where the score functions are as given in (a), the information matrix is as given in (b), and $\dot{q}(\theta_0) = (\psi'(\alpha), -\psi'(\beta))^T$. (g) Since $q(\theta) = \psi(\alpha) - \psi(\beta) = E_\theta \log X - E_\theta \log(1-X) = E_\theta \log \frac{X}{1-X}$, a simple methods of moments estimator of $q(\theta)$ is given by

$$\overline{\log \frac{X}{1-X}} = \frac{1}{n} \sum_{i=1}^{n} \log \frac{X_i}{1-X_i}.$$

This estimator is linear with influence function

$$\psi(x) = \log \frac{x}{1-x} - E_\theta \log \frac{X}{1-X}.$$

This is clearly in the tangent space $\dot{\mathcal{P}}$ of the model since $\psi = \dot{l}_\alpha - \dot{l}_\beta$.

6. (40 points) Suppose that $X, X_1, \ldots, X_n$ are i.i.d. with distribution function F_θ given by $F_\theta(x) = G(x-\theta)$ for $\theta \in R$ where $G(x) = 1-1/(1+x)^4$, $x \ge 0$, $G(x) = 0$ for $x \le 0$. [This is the same distribution as on problem 5 of the midterm exam, but shifted left by 1.]
(a) For what values of $r > 0$ is $E|X|^r < \infty$?
(b) Compute the density function $g(x)$ of the distribution function G. Sketch pictures of G and its density g.
(c) Suppose that $X_{n:1} \equiv X_{(n)} \equiv \min_{1 \le i \le n} X_i$. Show that $X_{n:1} \rightarrow_p \theta$.
(d) Show that $n(X_{n:1} - \theta) \rightarrow_d$ "something" and find the limiting disribution. [Hint: Start by computing $P(n(X_{n:1} - \theta) > t)$.]
(d) Is the score for θ defined in this problem? If you were to try to define the information for θ in this problem what would be a sensible definition? [Hint: it follows from part (d) that the estimator $X_{n:1}$ of θ satisfies $\sqrt{n}(X_{n:1} - \theta) \rightarrow_p 0 = N(0,0)$.]
(e) Show that $F_\theta^{-1}(t) = \theta + G^{-1}(t)$ for $t \in (0,1)$, and use this to find an estimator $\hat{\theta}_n(t)$ of θ based on the $t-$th sample quantile $\mathbb{F}_n^{-1}(t)$.
(f) Show that the estimators $\hat{\theta}_n(t)$ of θ which you derived in (e) satisfy

$$\sqrt{n}(\hat{\theta}_n(t) - \theta) \rightarrow_d N(0, V_t^2)$$

and compute V_t^2 explicitly.
(g) Find $t \in (0,1)$ which minimizes the variance V_t^2 of the estimator which you found in (f).

Solution: (a)

$$E|X|^r = \int_\theta^\infty |x|^r dF_\theta(x) = 4\int_\theta^\infty |x|^r \frac{1}{(1+(x-\theta))^5}\, dx < \infty$$

if $r - 5 < -1$; i.e. if $r < 4$, just as on the midterm exam.
(b) The density function corresponding to G is $g(x) = 4(1+x)^{-5} 1_{(0,\infty)}(x)$. A picture follows:

(c) Let $\epsilon > 0$. Then

$$\begin{aligned} P(|X_{n:1} - \theta| > \epsilon) &= P(X_{n:1} > \theta + \epsilon) + P(X_{n:1} < \theta - \epsilon) \\ &= P(X_i > \theta + \epsilon \text{ for all } i = 1, \ldots, n) + 0 \\ &= (1 - F_\theta(\theta + \epsilon))^n = (1 - G(\epsilon))^n \to 0 \end{aligned}$$

since $G(\epsilon) > 0$.
(d) Fix $t > 0$. Then

$$\begin{aligned} P(n(X_{n:1} - \theta) > t) &= P(X_{n:1} > \theta + t/n) \\ &= P(X_i > \theta + t/n \text{ for all } i = 1, \ldots, n) \\ &= (1 - F_\theta(\theta + t/n))^n = (1 - G(t/n))^n \\ &= (1 + \frac{t}{n})^{-4n} \to (e^{-t})^4 = \exp(-4t). \end{aligned}$$

In other words $n(X_{n:1} - \theta) \to_d Y$ where $Y \sim \exp(4)$.
(e) Solving $t = F_\theta(x) = G(x - \theta)$ for x yields $x = F_\theta^{-1}(t) = \theta + G^{-1}(t)$. Since $\mathbb{F}_n^{-1}(t) \to_{a.s.} F_\theta^{-1}(t) = \theta + G^{-1}(t)$, it follows that

$$\hat{\theta}_n(t) \equiv \mathbb{F}_n^{-1}(t) - G^{-1}(t) \to_{a.s.} \theta\,.$$

(f) Now

$$\sqrt{n}(\hat{\theta}_n(t) - \theta) = \sqrt{n}(\mathbb{F}_n^{-1}(t) - F_\theta^{-1}(t)) \to_d N(0, \frac{t(1-t)}{f_\theta^2(F_\theta^{-1}(t))}) \equiv N(0, V_t^2)$$

where

$$\begin{aligned} V_t^2 &= \frac{t(1-t)}{f_\theta^2(F_\theta^{-1}(t))} = \frac{t(1-t)}{g^2(G^{-1}(t))} \\ &= \frac{t(1-t)}{[4(1 + G^{-1}(t))^{-5}]^2} = \frac{t(1-t)}{16(1-t)^{5/2}} \\ &= \frac{1}{16} t(1-t)^{-3/2}. \end{aligned}$$

An easy calculation of derivatives shows that this is increasing in $t \in (0,1)$ and the minimum is 0 at $t = 0$. This is in rough qualititative agreement with the result in (d) if we think of $X_{n:1} = \mathbb{F}_n^{-1}(t_n)$ with $t_n = 1/n \to 0$.

7. (40 points) Suppose that $X, X_1, \ldots, X_n$ are i.i.d. exponential(θ_0); i.e. $X \sim P_{\theta_0} \in \mathcal{P} \equiv \{P_\theta : \theta > 0\}$ where P_θ has density

$$p_\theta(x) = \theta \exp(-\theta x)\, 1_{(0,\infty)}(x)$$

with respect to Lebesgue measure on R.
(a) Compute the score for θ based on one observation.
(b) Compute the information for θ based on one observation.
(c) Compute the Kullback-Leibler information $K(P_{\theta_0}, P_\theta)$ between P_{θ_0} and P_θ.
(d) Find the MLE $\hat{\theta}$ of θ based on the sample $X_1, \ldots, X_n$.
(e) Prove directly (using an appropriate central limit theorem) that $\hat{\theta}$ satisfies $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d N(0, 1/I(\theta_0))$.
(f) Consider estimation of the parameter $q(\theta) = \nu(P_\theta)$ given by $q(\theta) = \theta \exp(-\theta)$. Show that $q(\hat{\theta}_n)$ satisfies

$$\sqrt{n}(q(\hat{\theta}_n) - q(\theta_0)) \rightarrow_d N(0, V(\theta))$$

and find $V(\theta)$ explicitly.
(g) When $\theta_0 = 1$, show that

$$\sqrt{n}(q(\hat{\theta}_n) - q(\theta_0)) \rightarrow_p 0$$

and that

$$n(q(\hat{\theta}_n) - q(\theta_0)) \rightarrow_d \text{“}something\text{''};$$

identify the "something" explicitly. [Hint: draw a picture of $q(\theta)$ and note the special feature at the point $\theta = 1$.]

Solution: (a) $l(\theta|x) = \log\theta - \theta x$, so $\dot{l}_\theta(x) = 1/\theta - x$, $\ddot{l}_{\theta\theta}(x) = -1/\theta^2$,
(b) Computing the information via the "information identity" gives

$$I(\theta) = E_\theta(-\ddot{l}_{\theta\theta}(X)) = 1/\theta^2.$$

In this case the identity is easily checked since

$$I(\theta) = E_\theta(\dot{l}^2) = \theta^{-2}E_\theta(1 - \theta X)^2 = \theta^{-2}.$$

(c) First,

$$\log\frac{p_{\theta_0}}{p_\theta}(X) = \log\frac{\theta_0}{\theta} - (\theta_0 - \theta)X,$$

so

$$K(P_{\theta_0}, P_\theta) = E_{\theta_0}\log\frac{p_{\theta_0}}{p_\theta}(X) = \log\frac{\theta_0}{\theta} - \frac{\theta_0 - \theta}{\theta_0} = h(\frac{\theta}{\theta_0})$$

where $h(x) \equiv -\log x - 1 + x$ has $h(1) = 0$, $h'(1) = 0$, $h''(x) = x^{-2} > 0$, and $h(x) \rightarrow \infty$ as $x \downarrow 0$ or $x \uparrow \infty$.
(d) The score equation is

$$0 = \sum_{i=1}^n \dot{l}_\theta(X_i) = \sum_{i=1}^n(\frac{1}{\theta} - X_i) = \frac{n}{\theta} - n\overline{X},$$

so $\hat{\theta}_n = 1/\overline{X}$.
(e) Now $\overline{X}_n \to_p 1/\theta \equiv \mu$, and

$$\sqrt{n}(\overline{X}_n - \theta_0^{-1}) \to_d N(0, \theta_0^{-2}),$$

so with $g(x) = 1/x$, so that $g'(x) = -1/x^2$, the delta - methods yields

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \theta_0) = \sqrt{n}(g(\overline{X}_n) - g(1/\theta_0)) &\to_d g'(1/\theta_0^2) N(0, \theta_0^{-2}) \\ &= \theta_0^2 N(0, \theta_0^{-2}) = N(0, \theta_0^2) \\ &= N(0, 1/I(\theta_0)). \end{aligned}$$

(f) If $q(\theta) = \theta \exp(-\theta)$, then since $\dot{q}(\theta) = (1-\theta)\exp(-\theta)$,

$$\begin{aligned} \sqrt{n}(q(\hat{\theta}_n) - q(\theta_0)) &\to_d \dot{q}(\theta_0) N(0, \theta_0^2) \\ &= N(0, \dot{q}(\theta_0)^2 \theta_0^2) \end{aligned}$$

so that

$$V(\theta_0) = \theta_0^2 (1-\theta_0)^2 e^{-2\theta_0}.$$

(g) When $\theta_0 = 1$, $\dot{q}(1) = 0$ and $V(1) = 0$. Thus we have

$$\sqrt{n}(q(\hat{\theta}_n) - q(\theta_0)) \to_d 0,$$

and since convergence in distribution to a constant implies convergence in probability to the same constant, this yields

$$\sqrt{n}(q(\hat{\theta}_n) - q(\theta_0)) \to_p 0.$$

Now $\ddot{q}(\theta) = (\theta - 2)e^{-\theta}$ is continuous in a neighborhood of $\theta = 1$ and $\ddot{q}(1) = -e^{-1} < 0$. Hence $q(\theta)$ is maximized at $\theta = 1$ and

$$\begin{aligned} n(q(\hat{\theta}_n) - q(1)) &= \dot{q}(1) n(\hat{\theta}_n - 1) + \frac{1}{2}\ddot{q}(\theta^*) n(\hat{\theta}_n - 1)^2 \\ &= 0 + \frac{1}{2}\ddot{q}(\theta^*) n(\hat{\theta}_n - 1)^2 \\ &\to_d -\frac{e^{-1}}{2} N(0,1)^2 = -\frac{e^{-1}}{2}\chi_1^2 < 0 \end{aligned}$$

by using Slutsky's theorem and the continuous mapping theorem.