Statistics 581, Final Exam, Solution
Wellner; 12/11/2008

1. (48) points) **Define** *each* of the following terms. In each case, provide an appropriate context for your definition.
(a) An asymptotically linear estimator T_n of a parameter $\nu(P)$ with influence function ψ.
(b) A locally regular estimator T_n of a parameter $\nu(P_\theta)$.
(c) An $n^{1/4}-$consistent *preliminary estimator* of a (vector) parameter θ in a parametric model.
(d) A *one-step estimator* of a (vector) parameter θ in a parametric model.

Solution: (a) T_n is an asymptotically linear estimator (based on $X_1, \ldots, X_n$ i.i.d. P) of a parameter $\nu(P)$ with influence function ψ if

$$\sqrt{n}(T_n - \nu(P)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i) + o_p(1)$$

where $E\psi(X_1) = 0$ and $E\psi^2(X_1) < \infty$.
(b) An estimator T_n of a parameter $\nu(P_\theta)$ is locally regular at θ_0 if for every sequence $\{\theta_n\}$ with $\sqrt{n}(\theta_n - \theta_0) \to t$ for some $t \in \mathbb{R}^d$ it follows that

$$\sqrt{n}(T_n - \nu(P_{\theta_n})) \to_d \mathbb{Z}$$

where the distribution of $\mathbb{Z}$ depends on θ_0 but not on t.
(c) An estimator T_n is $n^{1/4}-$consistent if $n^{1/4}|T_n - \theta| = o_p(1)$ under P_θ.
(d) Suppose that $\bar{\theta}_n$ is a a preliminary estimator of θ. Then

$$\check{\theta}_n \equiv \bar{\theta}_n + I(\bar{\theta}_n)^{-1} n^{-1} l_n(\bar{\theta}_n)$$

is a one-step estimator of θ. If $\bar{\theta}_n$ is $n^{1/4}-$ consistent (as defined in (c)) and conditions A0-A4 hold, then $\check{\theta}_n$ is asymptotically efficient:

$$\sqrt{n}(\check{\theta}_n - \theta_0) \to_d N_d(0, I(\theta_0)^{-1}).$$

2. (32 points) **State** *four* of the following six results, providing the appropriate (brief) context for your statement:
(a) The Mann-Wald (or continuous mapping) theorem.
(b) The basic event identity connected with the inverse transformation.
(c) A result concerning the Kullback-Leibler discrepancy $K(P, Q)$ for two probability measures P and Q.

(d) An identity which gives two ways of calculating the information matrix; specify the regularity conditions which are needed for the identity to hold.
(e) The Lindeberg-Feller central limit theorem.
(f) LAN (Local Asymptotic Normality) of the local - log likelihood ratios for a regular parametric model satisfying the Cramér hypotheses.

Solution: See Chapters 2 - 4.

Do either problem 3 or problem 4.

3. (40 points) Let $X_1, \ldots, X_n$ be i.i.d. $P_\theta = \text{Normal}(\theta, 1)$.
(a) Give the Hodges superefficient estimator T_n of θ (with superefficiency at $\theta = 0$).
(b) What is the limiting distribution of $\sqrt{n}(T_n - \theta)$ as a function of θ?
(c) What is the limiting distribution of $\sqrt{n}(T_n - \theta_n)$ when sampling from $\theta = \theta_n$ when $\theta_n = cn^{-1/2}$?
(d) Does the limit distribution in (c) depend on c? Is T_n a locally regular estimator of θ at $\theta = 0$?
(e) What is the limit of $E_{\theta_n}\{[\sqrt{n}(T_n - \theta_n)]^2\}$ when $\theta_n = cn^{-1/2}$ as in (c)? For what values of c does the limiting risk of T_n exceed the (limiting) risk of $\overline{X}_n$?

Solution: See course notes, chapter 3, pages 25 - 26.

4. (40 points)
(a) Give an example of a parametric model for which the likelihood equations may have multiple roots, but for which our general theorem proving existence of consistent and asymptotically efficient roots applies.
(b) Give an example of a parametric model for which the maximum likelihood estimator is inconsistent.
(c) Give an example of a parametric model for which the maximum likelihood estimator does not exist.
(d) Give an example of a (non-Gaussian, non-normal) regular parametric model with dimension d of the parameter space at least 2 for which the maximum likelihood estimator exists almost surely and is asymptotically efficient. (e) Give an example of a one-dimensional parametric model for which the second derivative identity for information fails to hold.

Solution: See course notes, chapter 4, pages 16 - 24.

5. (30 points)
(a) **State** the Glivenko-Cantelli theorem. Then **prove** that it holds if it holds for the case of i.i.d. Uniform$(0, 1)$ random variables.

(b) **Prove** the Glivenko-Cantelli theorem for i.i.d. Uniform$(0,1)$ random variables: if $\xi_1,\ldots,\xi_n,\ldots$ are i.i.d. Uniform$(0,1)$ with empirical distribution function

$$\mathbb{G}_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{[0,t]}(\xi_i)\,, \qquad \text{then} \qquad \sup_{0\le t\le 1} |\mathbb{G}_n(t) - t| \rightarrow_{a.s.} 0\,.$$

Solution: See course notes, chapter 2, pages 17 - 18.

6. (50 points) (A parametric version of the Cox model.)
Suppose that $(Y|Z) \sim$ Exponential$(\lambda e^{\gamma Z})$ where $Z \sim$ Bernoulli(η). Thus the density of $X = (Y,Z)$ is given by

$$p_\theta(y,z) = \lambda e^{\gamma z}\exp(-\lambda e^{\gamma z} y) 1_{[0,\infty)}(y)\eta^z(1-\eta)^{1-z}1_{\{0,1\}}(z)$$

where $\theta = (\gamma,\lambda,\eta)$. Suppose that $X_1 = (Y_1,Z_1),\ldots,X_n = (Y_n,Z_n)$ are i.i.d. as X.
(a) Find the scores for $\theta = (\gamma,\lambda,\eta)$ based on one observation.
(b) Find the information matrix for θ.
(c) Compute the information for γ when λ is known (I_{11}) and unknown ($I_{11\cdot 2}$), and explain the difference based on the geometry of the scores.
(d) Write down the score equations for θ and briefly discuss the existence and uniqueness of solutions of these equations.
(e) What does our theory from chapter 4 say about the limiting distribution of $\sqrt{n}(\widehat{\theta} - \theta_0)$ and of $\sqrt{n}(\widehat{\gamma} - \gamma_0)$?

Solution:
(a) Now

$$\log p_\theta(y,z) = \log\lambda + \gamma z - \lambda e^{\gamma z} y + z\log\eta + (1-z)\log(1-\eta)\,,$$

so it follows that the scores are given by

$$\begin{aligned}
\dot{l}_\gamma(y,z) &= z - z\lambda e^{\gamma z} y = z(1-\lambda e^{\gamma z} y)\,,\\
\dot{l}_\lambda(y,z) &= \lambda^{-1} - e^{\gamma z} y = \lambda^{-1}(1-\lambda e^{\gamma z} y)\,,\\
\dot{l}_\eta(y,z) &= \frac{z}{\eta} - \frac{1-z}{1-\eta}\,.
\end{aligned}$$

(b) Calculating second derivatives, we find:

$$\begin{aligned}
\ddot{l}_{\gamma\gamma}(y,z) &= -\lambda z^2 e^{\gamma z} y\,,\\
\ddot{l}_{\gamma\lambda}(y,z) &= -z e^{\gamma z} y = \ddot{l}_{\lambda\gamma}(y,z)\,,\\
\ddot{l}_{\gamma\eta}(y,z) &= \ddot{l}_{\lambda\eta}(y,z) = \ddot{l}_{\eta\gamma}(y,z) = \ddot{l}_{\eta\lambda}(y,z) = 0\,,\\
\ddot{l}_{\lambda\lambda}(y,z) &= -\lambda^{-2}\,,\\
\ddot{l}_{\eta\eta}(y,z) &= -\frac{z}{\eta^2} - \frac{1-z}{(1-\eta)^2}\,.
\end{aligned}$$

Thus, computing the expectations of these second derivatives, and using the fact that $(\lambda e^{\gamma Z} Y|Z) \sim$ Exponential(1), we find that the information matrix for θ is given by

$$I(\theta) = \left(\begin{array}{ccc} E(Z^2) & E(Z)/\lambda & 0 \\ E(Z)/\lambda & 1/\lambda^2 & 0 \\ 0 & 0 & 1/(\eta(1-\eta)) \end{array} \right) .$$

(c) Thus the information for γ when λ is known is $I_{11} = E(Z^2) = \eta$. The information for γ when λ is unknown is

$$\begin{aligned} I_{11\cdot 2} &= I_{11} - I_{12} I_{22}^{-1} I_{21} \\ &= E(Z^2) - (\lambda^{-1}E(Z), 0) \left(\begin{array}{cc} \lambda^2 & 0 \\ 0 & \eta(1-\eta) \end{array} \right) (\lambda^{-1}E(Z), 0)^T \\ &= E(Z^2) - (E(Z))^2 = Var(Z) = \eta(1-\eta) . \end{aligned}$$

Geometrically, I_{11} is the squared length of the (raw) score $\dot{l}_\gamma$ for $\gamma = \theta_1$ in $L_2(P_\theta)$, while $I_{11\cdot 2}$ is the squared length of the efficient score for γ given by

$$l_\gamma^* = \dot{l}_\gamma - I_{12} I_{22}^{-1} \dot{l}_2 \;= \Pi(\dot{l}_\gamma | \dot{\mathcal{P}}_2^\perp) .$$

(d) The score equations for $\theta = (\gamma, \lambda, \eta)$ are given by

$$\begin{aligned} 0 &= \sum_{i=1}^n \dot{l}_\gamma(Y_i, Z_i) = \sum_{i=1}^n Z_i (1 - \lambda e^{\gamma Z_i} Y_i) , \\ 0 &= \sum_{i=1}^n \dot{l}_\lambda(Y_i, Z_i) = \sum_{i=1}^n \lambda^{-1} (1 - \lambda e^{\gamma Z_i} Y_i) , \\ 0 &= \sum_{i=1}^n \dot{l}_\eta(Y_i, Z_i) = \sum_{i=1}^n \left\{ \frac{Z_i}{\eta} - \frac{1 - Z_i}{1-\eta} \right\} = \frac{1}{\eta(1-\eta} n(\overline{Z} - \eta) . \end{aligned}$$

The third equation has the unique solution $\widehat{\eta} = \overline{Z}$. The second equation can be solved for $\widehat{\lambda}(\gamma)$ for each fixed value of γ to obtain

$$\widehat{\lambda}(\gamma) = \frac{n}{\sum_{i=1}^n Y_i e^{\gamma Z_i}} .$$

Substituting this into the first equation, we see that $\widehat{\gamma}$ satisfies

$$\overline{Z} = \frac{\sum_1^n Z_i Y_i e^{\gamma Z_i}}{\sum_1^n Y_i e^{\gamma Z_i}} ,$$

which has a unique solution if not all Z's are either 0 or 1.
(e) From our theory in Chapter 4 we know that

$$\sqrt{n}(\widehat{\theta} - \theta) \to_d N_3(0, I(\theta)^{-1})$$

where $I(\theta)$ was computed in (b), and

$$\sqrt{n}(\widehat{\gamma}-\gamma) \to_d N_1(0, 1/I_{11\cdot 2}(\theta)) = N_1(0, 1/Var(Z))$$

as was computed in (c).

7. (60 points). (Parametric Cox model, continued).
(a) Suggest three tests of the (composite!) null hypothesis $H : \gamma = 0$ versus $K : \gamma \neq 0$. What is the asymptotic distribution of each of these three statistics under the null hypothesis and under local alternatives of the form $\gamma_n = tn^{-1/2}$?
(b) If in (a) we were testing the simple null hypothesis $H_s : \theta = \theta_0$ versus $K_2 : \theta \neq \theta_0$, briefly describe the limiting behavior of the three statistics corresponding to those in (a) for a fixed alternative.
(c) Consider estimation of the function

$$q(\theta) = \nu(P_\theta) = P_\theta(Y > 1|Z = 1) = \frac{P_\theta(Y > 1, Z = 1)}{P_\theta(Z = 1)}.$$

Compute $q(\theta)$ explicitly as a function of θ.
(d) Suggest a natural empirical estimator of this (conditional) probability which does not rely on the exponential model. If this estimator is called $\tilde{\nu}_n$, show that $\tilde{\nu}_n$ is asymptotically linear and find its influence function ψ explicitly.
(e) Find the efficient influence function $\tilde{l}_\nu$ for estimation of $\nu(P_\theta)$ assuming the exponential model.
(f) Describe the relationship between ψ and $\tilde{l}_\nu$ geometrically.

Solution:
(a) Three possible statistics for testing $H : \gamma = 0$ versus $K : \gamma \neq 0$ are the LR, Wald, and Rao statistics given by

$$\begin{aligned}
2\log\lambda_n &= 2\log\left(\frac{\sup_{\theta\in\Theta} L_n(\theta)}{\sup_{\theta\in\Theta_0} L_n(\theta)}\right) \\
W_n &= [n^{1/2}(\widehat{\gamma}-0)]\hat{I}_{11\cdot 2}[n^{1/2}(\widehat{\gamma}-0)] = n\widehat{\gamma}^2\hat{I}_{11\cdot 2}\,, \\
R_n &= [Z_n(\widehat{\theta}_n^0)]^T\hat{I}(\widehat{\theta}_n^0)^{-1}[Z_n(\widehat{\theta}_n^0)]\,.
\end{aligned}$$

All three of these test statistics converge in distribution under the null hypothesis to χ_1^2. Under local alternatives of the form $\theta_n = \theta_0 + tn^{-1/2}$ where $\theta_0 = (1, \lambda_0, \eta_0)$ they all converge in distribution to $\chi_1^2(\delta)$ with $\delta = t_1 I_{11\cdot 2}(\theta_0)t_1$. The Rao statistic is the easiest to calculate, it simply entails calculation of $\widehat{\theta}_n^0 = (0, \widehat{\lambda}^0, \widehat{\eta}^0)$, and this is easy because the score equation for λ has the explicit solution $\widehat{\lambda}_n^0 = 1/\overline{Y}_n$ while $\widehat{\eta}^0 = \widehat{\eta} = \overline{Z}$.

(b) If we were testing a simple null hypothesis $H_s : \theta = \theta_0$ versus $K_s : \theta \neq \theta_0$ and $\theta \neq \theta_0$ is true, then

$$\begin{aligned} &n^{-1} 2 \log \lambda_n \to_p 2K(P_\theta, P_{\theta_0}) \,, \\ &W_n \to_p (\theta - \theta_0)^T I(\theta)(\theta - \theta_0) \,, \\ &R_n \to_p E_\theta \dot{\mathbf{l}}_\theta(\theta_0 | X_1)^T I^{-1}(\theta_0) E_\theta \dot{\mathbf{l}}_\theta(\theta_0 | X_1) \,. \end{aligned}$$

(c) First write (with, for example, $y_0 = 1$)

$$\begin{aligned} \nu(P_\theta) &= P_\theta(Y \geq y_0 | Z = 1) = \frac{P_\theta(Y \geq y_0, Z = 1)}{P_\theta(Z = 1)} \\ &= \frac{\eta \exp(-\lambda e^\gamma y_0)}{\eta} = \exp(-\lambda e^\gamma y_0) \,. \end{aligned}$$

Thus a natural nonparametric estimator of $\nu(P_\theta)$ is

$$\widehat{\nu} \equiv \nu(\mathbb{P}_n) = \frac{\mathbb{P}_n(Y \geq y_0, Z = 1)}{\mathbb{P}_n(Z = 1)} = \frac{n^{-1} \sum_1^n 1\{Y_i \geq y_0, Z_i = 1\}}{n^{-1} \sum_1^n 1\{Z_i = 1\}} \,.$$

(d) To show that $\widehat{\nu}$ is asymptotically linear, we write

$$\begin{aligned} \sqrt{n}(\widehat{\nu} - \nu(P_\theta)) &= \sqrt{n} \left\{ \frac{\mathbb{P}_n(Y \geq y_0, Z = 1) - P(Y \geq y_0, Z = 1)}{\mathbb{P}_n(Z = 1)} \right. \\ &\qquad \left. + P(Y \geq y_0, Z = 1) \left(\frac{1}{\mathbb{P}_n(Z = 1)} - \frac{1}{P(Z = 1)} \right) \right\} \\ &= \frac{1}{\mathbb{P}_n(Z = 1)} \sqrt{n} \left\{ \mathbb{P}_n(Y \geq y_0, Z = 1) - P(Y \geq y_0, Z = 1) \right. \\ &\qquad \left. - \frac{P(Y \geq y_0, Z = 1)}{P(Z = 1)} (\mathbb{P}_n(Z = 1) - P(Z = 1)) \right\} \\ &= \frac{1}{P(Z = 1)} \sqrt{n} \left\{ \mathbb{P}_n(Y \geq y_0, Z = 1) - P(Y \geq y_0, Z = 1) \right. \\ &\qquad \left. - \frac{P(Y \geq y_0, Z = 1)}{P(Z = 1)} (\mathbb{P}_n(Z = 1) - P(Z = 1)) \right\} + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \frac{1}{E(Z)} (1\{Y_i \geq y_0, Z_i = 1\} - \nu(P) 1\{Z_i = 1\}) \right\} + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(Y_i, Z_i) + o_p(1) \end{aligned}$$

where

$$\psi(y,z) = \frac{1}{E(Z)}\left(1\{y \geq y_0, z=1\} - \nu(P)1\{z=1\}\right) .$$

(e) The efficient influence function for $\nu(P_\theta) = q(\theta)$ is

$$\tilde{\mathbf{l}}_\nu(y,z) = \dot{q}(\theta)^T I(\theta)^{-1} \dot{\mathbf{l}}_\theta(y,z) \,,$$

and the information bound for $q(\theta)$ is given by

$$\dot{q}(\theta)^T I(\theta)^{-1} \dot{q}(\theta) \,.$$

(f) Geometrically, $\tilde{l}_\nu \in \dot{\mathcal{P}}$ is the projection of ψ onto $\dot{\mathcal{P}}$: $\tilde{l}_\nu = \Pi(\psi|\dot{\mathcal{P}})$. Thus $\psi - \tilde{l}_\nu \perp \dot{\mathcal{P}}$.