Statistics 581, Final Exam Solution

Wellner; 12/11/2000

1. (32) points) **Define** *each* of the following terms. In each case, 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 Hellinger distance between two probability measures P and Q.
   (c) The Kullback-Leibler information number $K(P, Q)$ between a probability hfill measure P and a (sub-)probability measure Q.
   (d) The efficient influence function $\tilde{l}_\nu$ for a differentiable parameter $q(\theta) = \nu(P_\theta)$ in a regular parametric model $\mathcal{P}$.

   Solution: See the Notes, Chapters 1-4.

2. (40) **State** *each* of the following results, providing the appropriate (brief) context for your statement:
   (a) A basic result concerning the Kullback-Leibler "distance" $K(P, Q)$.
   (b) The dominated convergence theorem.
   (c) An identity which gives two ways of calculating the information matrix; specify the regularity conditions which are needed for the identity to hold.
   (d) The Lindeberg-Feller central limit theorem.
   (e) The Mann-Wald or continuous mapping theorem.

   Solution: See the Notes, Chapters 1-4.

3. (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

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

   Solution: See the proof of Theorem 4.1 in Chapter 2 of the notes.

4. Suppose that $\underline{N} \sim \text{Mult}_k(n, \underline{p})$; thus the probability distribution of $\underline{N}$ is given by

$$P_p(\underline{N} = \underline{m}) = \frac{n!}{m_1! \cdots m_k!} \prod_{j=1}^{k} p_j^{m_j} \,,$$

and the likelihood is

$$L_n(\underline{p} | \underline{N}) = \frac{n!}{N_1! \cdots N_k!} \prod_{j=1}^{k} p_j^{N_j} \,.$$

Show, without using any calculus, that the MLE of $\underline{p}$ is $\widehat{\underline{p}} = \underline{N}/n$.

Solution: Note that n^{-1} times the log-likelihood is

$$n^{-1} l_n(\underline{p}) \quad = \quad \sum_{j=1}^{k} (N_j/n) \log p_j = \sum_{j=1}^{k} \widehat{p}_j \log p_j \,.$$

Thus we have, by Jensen's inequality,

$$\sum_{j=1}^{k} \widehat{p}_j \log \left(\frac{p_j}{\widehat{p}_j} \right) \quad \leq \quad \log \left(\sum_{j=1}^{k} p_j \right) = \log 1 = 0$$

since $\sum_1^k p_j = 1$. Note that equality holds if and only if all the $p_j/\widehat{p}_j = 1$, or equivalently $p_j = \widehat{p}_j$, $j = 1, \ldots, k$. This implies that

$$n^{-1} l_n(\underline{p}) = \sum_{j=1}^{k} (N_j/n) \log p_j = \sum_{j=1}^{k} \widehat{p}_j \log p_j \leq \sum_{j=1}^{k} \widehat{p}_j \log \widehat{p}_j = n^{-1} l_n(\widehat{\underline{p}})$$

with equality if and only if $\underline{p} = \widehat{\underline{p}}$. Thus the MLE of $\underline{p}$ is $\widehat{\underline{p}} = \underline{N}/n$.

5. (54 points) Suppose that $\underline{N}_n = (N_{11}, N_{12}, N_{21}, N_{22}) \sim \text{Mult}_4(n, \underline{p})$ where $\underline{p} = (p_{11}, p_{12}, p_{21}, p_{22})$ where $\sum_{i=1}^2 \sum_{j=1}^2 p_{ij} = 1$. (Thus $\underline{N}_n$ is the sum of n independent $\text{Mult}_4(1, \underline{p})$ random vectors $\{\underline{Y}_i\}_{i=1}^n$.) Since there are really just three independently varying parameters for this problem, it is often useful to re-express the cell probabilities in terms of two marginal probabilities, say $p_{1\cdot} = p_{11} + p_{12}$ and $p_{\cdot 1} = p_{11} + p_{21}$, and ψ, the log of the odds-ratio, defined by

$$\psi \equiv \log \frac{p_{21}/p_{22}}{p_{11}/p_{12}} = \log \frac{p_{12}p_{21}}{p_{11}p_{22}} \,. \tag{0.1}$$

You may use the fact that $\psi = 0$ if and only if independence holds for the 2×2 table (i.e. $p_{ij} = p_{i\cdot}p_{\cdot j}$ for $i, j = 1, 2$).

(a) Suggest an estimator of ψ, say $\widehat{\psi}$.

(b) Show that the estimator you proposed in (b) is asymptotically normal and compute the asymptotic variance of your estimator.

(c) One standard test of independence in the 2×2 table is the test based on a Pearson-type chi-square statistic. Write down the chi-square statistic Q_n for this problem, state its asymptotic distribution under the null hypothesis, and explain briefly why the claimed result holds.

(d) Another statistic for testing independence is the likelihood ratio statistic $2 \log \lambda_n$ for testing $H : \underline{p} \in \Theta_0$ versus $K : \underline{p} \notin \Theta_0$, or, equivalently, $H : \psi = 0$ versus $K : \psi \neq 0$. Find this statistic and state its asymptotic distribution under the null hypothesis.

(e) Without doing any additional calculation, what is the asymptotic distribution of the likelihood ratio statistic under local alternatives of the form $\psi_n = tn^{-1/2}$? (Hint: use the result of (b) to find an expression for the (efficient) information for ψ in the presence of the nuisance parameters $p_{1\cdot}$, $p_{\cdot 1}$.)

(f) Suppose that $\psi \neq 0$; i.e. the alternative hypothesis holds. Show that for the statistics Q_n and $2 \log \lambda_n$ from (d) and (e) we have $n^{-1}Q_n \to_p q$ and $n^{-1} 2 \log \lambda_n \to_p J$ for some positive constants q and J respectively; you should compute q and J as explicitly as possibly in terms of $\underline{p}$ and/or $(p_{1\cdot}, p_{\cdot 1}, \psi)$.

Solution: (a) An obvious estimator of ψ is

$$\widehat{\psi} = \log \frac{\widehat{p}_{12}\widehat{p}_{21}}{\widehat{p}_{11}\widehat{p}_{22}}$$

where $\widehat{\underline{p}} = \underline{N}/n$.

(b) Now $\widehat{\psi} = g(\widehat{\underline{p}})$ where $g(\underline{p})$ is given in (0.1) and is differentiable with derivative

$$\nabla g(\underline{p}) = (-1/p_{11}, 1/p_{12}, 1/p_{21}, -1/p_{22})$$

and, by the multivariate CLT,

$$\sqrt{n}(\widehat{\underline{p}} - \underline{p}) \to_d Z \sim N_4(0, \Sigma)$$

where $\Sigma = \mathrm{diag}(\underline{p}) - \underline{p}\underline{p}^T$. Thus the delta method (or g'-theorem) yields

$$\begin{aligned} \sqrt{n}(\widehat{\psi} - \psi) &= \sqrt{n}(g(\widehat{\underline{p}}) - g(\underline{p})) \\ &\to_d \nabla g(\underline{p}) Z \sim N(0, \nabla g^T \Sigma \nabla g) = N(0, V^2(\underline{p})) \end{aligned}$$

where

$$V^2(\underline{p}) = \frac{1}{p_{11}} + \frac{1}{p_{12}} + \frac{1}{p_{21}} + \frac{1}{p_{22}}.$$

(c) The Pearson chi-square statistic for testing independence is

$$Q_n \equiv \sum_{i,j=1}^{2} \frac{(N_{ij} - n(N_{i\cdot}/n)(N_{\cdot j}/n))^2}{n(N_{i\cdot}/n)(N_{\cdot j}/n)}$$

where $N_{i\cdot} = N_{i1} + N_{i2}$, $i = 1, 2$ and $N_{\cdot j} = N_{1j} + N_{2j}$, $j = 1, 2$ are the column and row sums respectively. Under the null hypothesis of independence, $Q_n \to_d \chi_1^2$ where there is one degree of freedom because there are three $= 4 - 1$ basic degrees of freedom in a Multinomial distribution with four cells, and we have estimated two parameters (the marginal probabilities $p_{1\cdot}$ and $p_{\cdot 1}$.

(d) The log-likelihood ratio statistic for testing independence is $2\log\lambda_n$ where

$$\lambda_n = \frac{\sup_{\underline{p}} L_n(\underline{p})}{\sup_{\underline{p}\in\Theta_0} L_n(\underline{p})}.$$

Now

$$\sup_{\underline{p}} L_n(\underline{p}) = L_n(\widehat{\underline{p}}) = \frac{n!}{N_{11}!N_{12}!N_{21}!N_{22}!} \prod_{i,j=1}^{2} \widehat{p}_{ij}^{N_{ij}},$$

and

$$\sup_{\underline{p}\in\Theta_0} L_n(\underline{p}) = L_n(\widehat{\underline{p}^0}) = \frac{n!}{N_{11}!N_{12}!N_{21}!N_{22}!} \prod_{i=1}^{2} \widehat{p}_{i\cdot}^{N_{i\cdot}} \prod_{j=1}^{2} \widehat{p}_{\cdot j}^{N_{\cdot j}}$$

where $p_{1\cdot} + p_{2\cdot} = 1$, $p_{\cdot 1} + p_{\cdot 2} = 1$, and $N_{1\cdot} + N_{2\cdot} = n$, $N_{\cdot 1} + N_{\cdot 2} = n$. Thus the log-likelihood ratio statistic is

$$2\log\lambda_n = 2\left\{ \sum_{i,j=1}^{2} N_{ij}\log\widehat{p}_{ij} - \sum_{i=1}^{2} N_{i\cdot}\log\widehat{p}_{i\cdot} - \sum_{j=1}^{2} N_{\cdot j}\log\widehat{p}_{\cdot j} \right\}.$$

Under the hull hypothesis, $2\log\lambda_n \to_d \chi_1^2$.
(e) Under local alternatives of the form $\psi_n = tn^{-1/2}$ we know that $2\log\lambda_n \to_d \chi_1^2(\delta)$ where $\delta = t^2 I_{11\cdot 2}$ where $\theta = (\theta_1, \theta_2)$, $\theta_1 \equiv \psi$ and $\theta_2 = (p_{1\cdot}, p_{\cdot 1})$. Now the asymptotic variance of $\widehat{\psi}$ in (b) must be the reciprocal of the efficient information $I_{11\cdot 2}$ for ψ in the presence of the nuisance parameters $\theta_2 = (p_{1\cdot}, p_{\cdot 1})$. Hence we deduce that

$$\begin{aligned} I_{11\cdot 2} &= \left(\frac{1}{p_{11}} + \frac{1}{p_{12}} + \frac{1}{p_{21}} + \frac{1}{p_{22}}\right)^{-1} \\ &= \text{the harmonic mean of the } p_{ij}\text{'s} \end{aligned}$$

Under the null hypothesis we have

$$\begin{aligned} p_{11} &= p_{1\cdot}p_{\cdot 1}\,, \\ p_{12} &= p_{1\cdot}p_{\cdot 2} = p_{1\cdot}(1 - p_{\cdot 1})\,, \\ p_{21} &= p_{2\cdot}p_{\cdot 1} = (1 - p_{1\cdot})p_{\cdot 1}\,, \\ p_{22} &= p_{2\cdot}p_{\cdot 1} = (1 - p_{1\cdot})(1 - p_{\cdot 1})\,. \end{aligned}$$

Figure 1 shows a plot of $I_{11\cdot 2}$ as a function of $p_{1\cdot}$ and $p_{\cdot 1}$.

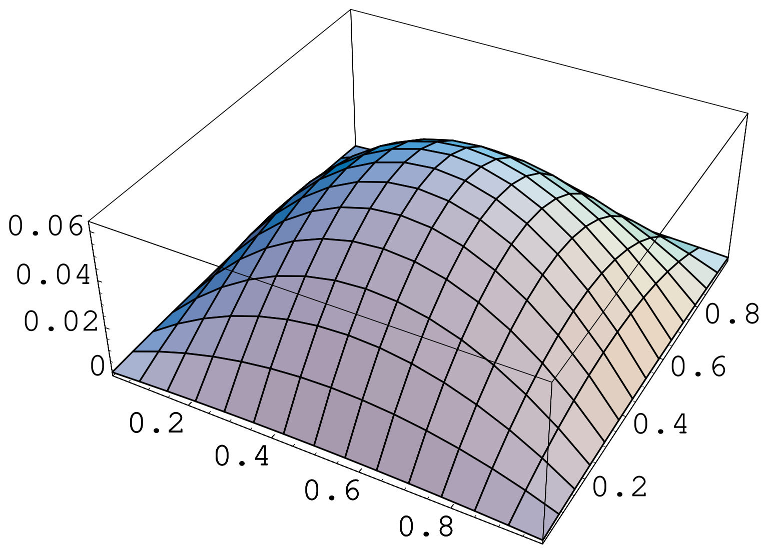

Figure 1: Information $I_{11\cdot 2}$ for ψ with $p_{1\cdot}$ and $p_{\cdot 1}$ unknown.

(f) Now under the alternative hypothesis $\psi \neq 0$, we have

$$\underline{\widehat{p}} \to_p \underline{p}$$

while

$$\widehat{p}_{i\cdot} = n^{-1}(N_{i1} + N_{i2}) \to_p p_{i1} + p_{i2} = p_{i\cdot}$$

and

$$\widehat{p}_{\cdot j} = n^{-1}(N_{1j} + N_{2j}) \to_p p_{1j} + p_{2j} = p_{\cdot j}\,.$$

Hence it follows that

$$
\begin{aligned}
n^{-1}Q_n &= \sum_{i,j=1}^{2} \frac{(N_{ij}/n - (N_{i\cdot}/n)(N_{\cdot j}/n))^2}{(N_{i\cdot}/n)(N_{\cdot j}/n)} \\
&= \sum_{i,j=1}^{2} \frac{(\widehat{p}_{ij} - \widehat{p}_{i\cdot}\widehat{p}_{\cdot j})^2}{\widehat{p}_{i\cdot}\widehat{p}_{\cdot j}} \\
&\to_p \sum_{i,j=1}^{2} \frac{(p_{ij} - p_{i\cdot}p_{\cdot j})^2}{p_{i\cdot}p_{\cdot j}} \equiv q\,.
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
2n^{-1}\log\lambda_n &= 2\left\{\sum_{i,j=1}^{2} \widehat{p}_{ij}\log\widehat{p}_{ij} - \sum_{i=1}^{2} \widehat{p}_{i\cdot}\log\widehat{p}_{i\cdot} - \sum_{j=1}^{2} \widehat{p}_{\cdot j}\log\widehat{p}_{\cdot j}\right\} \\
&\to_p 2\left\{\sum_{i,j=1}^{2} p_{ij}\log p_{ij} - \sum_{i=1}^{2} p_{i\cdot}\log p_{i\cdot} - \sum_{j=1}^{2} p_{\cdot j}\log p_{\cdot j}\right\} \equiv J\,.
\end{aligned}
$$

6. (54 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

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

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 $\sqrt{n}(\widehat{\gamma} - \gamma_0)$?
(f) Consider testing $H:\ \gamma = 0$ versus $K: \gamma \neq 0$. Suggest three test statistics, and briefly discuss the pro's and con's of each. What is the asymptotic distribution of these test statistics under the null hypothesis and local alternatives.

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

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

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

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

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

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

(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

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

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

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

as was computed in (c).
(f) 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. 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 $\widehat{\eta}^0 = \widehat{\eta} = \overline{Z}$.

7. (36 points) (A parametric version of the Cox model, continued). The following problem is in the same context as that of problem 6 above:
Consider estimation of the parameter

$$q(\theta) = \exp(-\lambda e^{\gamma} y_0) = P_\theta(Y \geq y_0 | Z = 1) \equiv \nu(P_\theta)$$

for a fixed number $y_0 > 0$.
(a) Suggest a natural empirical estimator $\widehat{\nu}_n$ of this probability (taking care to note that it is a *conditional probability*).
(b) Show that $\widehat{\nu}_n$ is asymptotically linear, and find its influence function, ψ.
(c) Find the efficient influence function $\tilde{l}_\nu$ for estimation of $q(\theta)$ and the related information bound.
(d) Describe the relationship between ψ and $\tilde{l}_\nu$ geometrically.

Solution:
(a) First write

$$\nu(P_\theta) = P_\theta(Y \geq y_0 | Z = 1) = \frac{P_\theta(Y \geq y_0, Z = 1)}{P_\theta(Z = 1)}\,.$$

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

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

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

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

where

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

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

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

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

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

(d) 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 $\psi - \tilde{l}_\nu \perp \dot{\mathcal{P}}$.