Statistics 581, Final Exam Solutions

Wellner; 12/17/2001

1. (40) 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 $\mathcal{P} = \{P_\theta : \ \theta \in \Theta \subset R^k\}$.
   (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 with influence function ψ.

2. (30 points) **State** *each* of the following results, providing the appropriate (brief) context for your statement:
   (a) The Glivenko-Cantelli theorem.
   (b) Donsker's theorem for the empirical process.
   (c) A result about the finite-dimensional limiting distributions of the sample quantile process.

3. (45 points). Suppose that $X, X_1, \ldots, X_n$ are i.i.d. positive random variables. Let

$$A_n = \overline{X}_n \ = \ n^{-1} \sum_{i=1}^n X_i \, ,$$

$$G_n = \left\{ \prod_{i=1}^n X_i \right\}^{1/n}, \qquad \text{and} \qquad H_n = \frac{1}{\frac{1}{n}\sum_{i=1}^n X_i^{-1}}$$

be the arithmetic, geometric, and harmonic means, respectively.

A. Use the weak law of large numbers and the continuous mapping theorem to show that $A_n \rightarrow_p a$, $G_n \rightarrow_p g$, and $H_n \rightarrow_p h$ for some constants a, g, and h depending on the distribution of X under appropriate integrability assumptions. Specify these assumptions precisely and identify the constants a, g, and h.

B. Compute the constants a, g, and h when $X \sim (1/2)$ Uniform$(1,2) + (1/2)\delta_1$; i.e. when the distribution function F of the X's is given by

$$F(x) = \left\{ \begin{array}{rl} 0 \, , & x < 1 \\ x/2 \, , & 1 \le x \le 2 \\ 1 \, , & 2 \le x < \infty \end{array} \right\} .$$

C. Use the multivariate central limit theorem and the delta-method to show that under appropriate assumptions on the distribution of X

$$\sqrt{n} \left(\begin{array}{c} A_n - a \\ G_n - g \\ H_n - h \end{array} \right) \rightarrow_d N_3(0, V)$$

for some covariance matrix V. Identify the assumptions you need explicitly. Compute V as explicitly as possible.

Solution: A. Now $A_n = \overline{X}_n \to_p E(X) \equiv a$. Also,

$$\log G_n \;=\; \frac{1}{n}\sum i = 1^n \log X_i \to_p E(\log X)$$

if $E|\log X| < \infty$, and

$$1/H_n = \frac{1}{n}\sum_{i=1}^{n} X_i^{-1} \to_p E(1/X)$$

if $E(1/X) < \infty$. Thus by the continuous mapping theorem

$$G_n \to_p \exp(E(\log(X))) \equiv g, \qquad H_n \to_p \frac{1}{E(1/X)} \equiv h\,.$$

B. When $X \sim F = (1/2)Uniform(1,2) + \delta_1$,

$$a = E(X) = \frac{1}{2} + \frac{1}{2}\frac{3}{2} = \frac{1}{2} + \frac{3}{4} = \frac{5}{4}\,;$$

$$E(\log X) = \frac{1}{2}\log 1 + \frac{1}{2}\int_1^2 \log x\, dx = \frac{1}{2}\int_1^2 \log x dx = \log 2 - \frac{1}{2}\,,$$

and hence $g = \exp(E(\log X)) = 2e^{-1/2}$; and

$$E(1/X) = \frac{1}{2} + \frac{1}{2}\int_1^2 x^{-1} dx = (1 + \log 2)/2\,.$$

Thus we have $(a, g, h) = (5/4, 2e^{-1/2}, 2/(1 + \log 2)) = (1.25, 1.21306..., 1.18123...)$.
C. By the Multivariate CLT, if $E(X^2) < \infty$, $E((\log X)^2) < \infty$, and $E(1/X^2) < \infty$, it follows that

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - a \\ \log G_n - \log g \\ 1/H_n - 1/h \end{pmatrix} \to_d Z \sim N_3(0, \Sigma)$$

where

$$\Sigma = \begin{pmatrix} Var(X) & d & e \\ d & Var(\log X) & f \\ e & f & Var(1/X) \end{pmatrix}$$

where

$$\begin{aligned} d &= E(X\log X) - E(X)E(\log X), \\ e &= 1 - E(X)E(1/X), \\ f &= E((1/X)\log X) - E(1/X)E(\log X)\,. \end{aligned}$$

Let $g(x, y, z) = (x, e^y, 1/z)$. Then

$$\nabla g(x, y, z) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^y & 0 \\ 0 & 0 & -z^{-2} \end{pmatrix}\,,$$

and hence

$$\nabla g(a, \log g, 1/h) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & g & 0 \\ 0 & 0 & -h^2 \end{pmatrix}.$$

Thus by the delta-method it follows that

$$\sqrt{n} \begin{pmatrix} \overline{X}_n - a \\ G_n - g \\ H_n - h \end{pmatrix} \to_d \nabla g(a, \log g, 1/h) Z \sim N_3(0, V)$$

where $V = \nabla g(a, \log g, 1/h) \Sigma \nabla g(a, \log g, 1/h)$.

Do **either** problem 4 **or** problem 5.

4. (40 points).
For a regular parametric model satisfying the Cramér hypotheses of our Theorem 4.1.5 of Chapter 4, the local likelihood ratios satisfy the LAN condition: under P_{θ_0}

$$\log \frac{L_n(\theta_0 + tn^{-1/2})}{L_n(\theta_0)} = tZ_n - \frac{1}{2} t^T I(\theta_0) t + o_p(1) \to_d t^T Z - \frac{1}{2} t^T I(\theta_0) t \qquad (1)$$

where $Z \sim N_k(0, I(\theta_0))$.
A. Show that the right side of (1) can be written as

$$-\frac{1}{2}(t - W)^T I(\theta_0)(t - W) + \frac{1}{2} W^T I(\theta_0) W$$

where W is some linear transformation of Z. Identify W explicitly in terms of $I(\theta_0)$ and Z.
B. The displayed equation in part A shows that the limit is maximized as a function of t by $t = W$ and the maxmimum value is $(1/2) W^T I(\theta_0) W$. Rewrite both of these relations in terms of $I(\theta_0)$ and Z and interpret the result.
Hint: If we interpret the maximum likelihood estimator $\hat{\theta}_n$ of θ (assuming it exists) as $\text{argmax}_\theta \log L_n(\theta)$, then

$$\text{argmax}_t \left(\log L_n(\theta_0 + tn^{-1/2}) / L_n(\theta_0) \right) = \sqrt{n}(\hat{\theta}_n - \theta_0)$$

while the argmax of the right side of (1) is just W; when W is rewritten in terms of $I(\theta_0)$ and Z this becomes quite natural!

Solution: A. Evidently $W = I(\theta_0)^{-1} Z$. Thus

$$t^T Z - \frac{1}{2} t^T I(\theta_0) t = -\frac{1}{2}(t - I(\theta_0)^{-1} Z)^T I(\theta_0)(t - I(\theta_0)^{-1} Z) + \frac{1}{2} Z^T I(\theta_0)^{-1} Z.$$

B. $W = I(\theta_0)^{-1} Z \sim N_k(0, I(\theta_0)^{-1})$, while $(1/2) W^T I(\theta_0) W = (1/2) Z^T I(\theta_0)^{-1} Z \sim (1/2)\chi_k^2$. since

$$\text{argmax}_t \frac{L_n(\theta_0 + tn^{-1/2})}{L_n(\theta_0)} = \sqrt{n}(\hat{\theta}_n - \theta_0)$$

and

$$\text{argmax}_t (t^T Z - \frac{1}{2} t^T I(\theta_0) t) = I(\theta_0)^{-1} Z,$$

it is natural to suspect that when LAN (i.e. (1)) holds

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d I(\theta_0)^{-1} Z \sim N_k(0, I(\theta_0)^{-1}),$$

and this is what we proved in Theorem 4.1.5.

5. (40 points). Suppose that $Z \sim P_{\mu,\Sigma} = N_k(\mu, \Sigma)$ where Σ is positive definite.
A. What is the density of Z with respect to Lebesgue measure on R^k?
B. Is $P_{\mu,\Sigma} << P_{0,\Sigma}$ as measures? If your answer is yes, explain why and compute the Radon-Nikodym derivative

$$\frac{dP_{\mu,\Sigma}}{dP_{0,\Sigma}}(z).$$

Hint: if $P << \mu$, $Q << \mu$, and $Q << P$, then

$$\frac{dQ}{dP} = \frac{dQ/d\mu}{dP/d\mu}.$$

C. Suppose $\mu = I(\theta_0)t$, $\Sigma = I(\theta_0)$, so that $Z \sim P_{I(\theta_0)t, I(\theta_0)}$. Use the result of B to compute

$$\log\left\{\frac{dP_{I(\theta_0)t, I(\theta_0)}}{dP_{0,I(\theta_0)}}\right\}(Z).$$

Explain the meaning of this computation for the Local Asymptotic Normality (LAN) part of Theorem 4.1.5.

Solution: A. Since Σ is positive definite the density with respect to Lebesgue measure μ is given by

$$p(x; \mu, \Sigma) = \frac{1}{(2\pi|\Sigma|)^{k/2}} \exp(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)).$$

B. Yes; $P_{\mu,\sigma} \prec\prec P_{0,\Sigma}$: if $P_{0,\Sigma}(A) = 0$, then since $p(x; 0, \Sigma) > 0$ for all x, it follows that $\mu(A) = 0$, and hence

$$P_{\mu,\Sigma}(A) = \int_A p(x; \mu, \Sigma) d\mu(x) = 0.$$

By the hint it follows that

$$\begin{aligned}
\frac{dP_{\mu,\Sigma}}{dP_{0,\Sigma}}(z) &= \frac{p(z; \mu, \Sigma)}{p(z; 0, \Sigma)} \\
&= \frac{\exp(-(1/2)(z-\mu)^T \Sigma^{-1}(z-\mu))}{\exp(-(1/2)z^T \Sigma^{-1} z)} \\
&= \exp(\mu^T \Sigma^{-1} z - (1/2)\mu^T \Sigma^{-1} \mu).
\end{aligned}$$

C. When $\mu = I(\theta_0)t$ and $\Sigma = I(\theta_0)$, the result of B yields

$$\log\left\{\frac{dP_{\mu,\Sigma}}{dP_{0,\Sigma}}(Z)\right\} = t^T Z - \frac{1}{2} t^T I(\theta_0) t,$$

which is exactly what appear on the right side in the LAN condition. Thus the right side in the LAN condition is exactly

$$\log\left\{\frac{dP_{\mu,\Sigma}}{dP_{0,\Sigma}}(Z)\right\}.$$

Do **either** problem 6 **or** problem 7.

6. (40 points). Suppose that $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset R^k\}$ is a regular parametric model with densities p_θ with respect to a dominating measure μ. Suppose that $\theta_0 \in \Theta$ is fixed.
A. Express $K(P_{\theta_0}, P_\theta)$ in terms of the densities p_{θ_0} and p_θ.
B. If $g(\theta) \equiv K(P_{\theta_0}, P_\theta)$, compute (assuming that the interchange of integration and differentiation is permissible)

$$\dot{g}(\theta) = (\frac{\partial}{\partial\theta_1} g(\theta), \ldots, \frac{\partial}{\partial\theta_k} g(\theta))^T$$

and the $k \times k$ matrix

$$\ddot{g}(\theta) = \left(\frac{\partial^2}{\partial\theta_i \partial\theta_j} g(\theta)\right).$$

Interpret $\dot{g}(\theta)$ in terms of a particular constant involved in the limit behavior of the Rao (or score) statistic under fixed alternatives. Evaluate both $\dot{g}(\theta)$ and $\ddot{g}(\theta)$ at θ_0, and identify the special values involved.
C. Carry out the computations in A and B explicitly when $p_\theta(x) = \theta^x e^{-\theta}/x!$ for $x = 0, 1, 2, \ldots$, $\theta > 0$.

Solution: A. By definition

$$\begin{aligned} K(P_{\theta_0}, P_\theta) &= E_{\theta_0} \log \frac{p_{\theta_0}(X)}{p_\theta(X)} \\ &= \int p_{\theta_0}(x) \log \frac{p_{\theta_0}(x)}{p_\theta(x)} d\mu(x) \\ &= \int p_{\theta_0}(x) \log p_{\theta_0}(x) d\mu(x) - \int p_{\theta_0}(x) \log p_\theta(x) d\mu(x). \end{aligned}$$

B. From the last equation in A we easily compute

$$\dot{g}(\theta) = -\int p_{\theta_0}(x) \dot{l}_\theta(x;\theta) d\mu(x) = -E_{\theta_0} \dot{l})\theta(X).$$

This is minus the expected value under the "true" θ_0 of the scores evaluated at the "hypothesized" point θ, and hence it is exactly minus one times the vector involved in describing the behavior of the Rao statistic under fixed alternatives. Furthermore,

$$\ddot{g}(\theta) = -E_{\theta_0} \ddot{l}_{\theta\theta}(X;\theta).$$

Evaluating these at θ_0 we find

$$\dot{g}(\theta_0) = -E_{\theta_0} \dot{l}_\theta(X;\theta_0) = 0,$$

while

$$\ddot{g}(\theta_0) = -E_{\theta_0} \ddot{l}_{\theta\theta}(X;\theta_0) = I(\theta_0).$$

C. When $p_\theta(x) = \theta^x e^{-\theta}/x!$ it follows that

$$\frac{p_{\theta_0}(x)}{p_\theta(x)} = \left(\frac{\theta_0}{\theta}\right)^x e^{-(\theta_0 - \theta)}$$

so that

$$\log\left(\frac{p_{\theta_0}(x)}{p_\theta(x)}\right) = x\log\left(\frac{\theta_0}{\theta}\right) - (\theta_0 - \theta)$$

and hence

$$K(P_{\theta_0}, P_\theta) = \theta_0 \log\left(\frac{\theta_0}{\theta}\right) - (\theta_0 - \theta) = \theta h(\theta_0/\theta)$$

where $h(x) = x(\log x - 1) + 1$. [The function h arises frequently in connection with the Binomial and Poisson distributions; see e.g. Shorack and Wellner (1986), pages 416 and 445.] Thus we easily compute

$$\dot{g}(\theta) = 1 - \frac{\theta_0}{\theta}$$

which equals zero at θ_0, and

$$\ddot{g}(\theta) = \frac{\theta_0}{\theta^2}$$

which reduces to $1/\theta_0 = I(\theta_0)$ at θ_0. We can easily check these by computing

$$l(\theta|X) = X\log\theta - \theta - \log(X!)\,,$$

$$\dot{l}_\theta(X;\theta) = \frac{X}{\theta} - 1\,,$$

and hence

$$-E_{\theta_0}\dot{l}_\theta(X,\theta) = 1 - \frac{\theta_0}{\theta}\,.$$

Moreover,

$$I(\theta_0) = E_{\theta_0}\dot{l}_\theta(X,\theta_0)^2 = \frac{1}{\theta_0^2}Var_{\theta_0}(X) = \frac{1}{\theta_0}\,.$$

7. (40 points)
Suppose that $\underline{X}, \underline{X}_1, \ldots, \underline{X}_n$ are i.i.d. $\text{Mult}_k(1, \underline{p})$, so that $\underline{N}_n \equiv \sum_{i=1}^n \underline{X}_i \sim \text{Mult}_k(n, \underline{p})$. Thus

$$\begin{aligned} P_{\underline{p}}(\underline{X} = \underline{x}) &= \prod_{j=1}^k p_j^{x_j} \quad \text{for} \;\; x_i \in \{0,1\}, \;\; \sum_1^k x_i = 1\,, \\ P_{\underline{p},n}(\underline{N}_n = \underline{m}) &= \frac{n!}{\prod_{j=1}^k m_j!}\prod_{j=1}^k p_j^{m_j} \quad \text{for} \;\; m_i \ge 0, \;\text{integers}\;\; \sum_{j=1}^k m_j = n\,. \end{aligned}$$

A. Compute $K(P_{\underline{q}}, P_{\underline{p}}) \equiv K(\underline{q}, \underline{p})$ for vectors $\underline{q}$, $\underline{p}$ with $\sum p_j = \sum q_j = 1$.
B. Evaluate $K(\widehat{\underline{p}}, \underline{p})$ where $\widehat{\underline{p}} = n^{-1}\underline{N}_n$. Relate this to the log-likelihood $\log L_n(\underline{p}|\underline{N}_n)$.
C. Use the result of B to show, without using any calculus, that the MLE of $\underline{p}$ is $\widehat{\underline{p}} = \underline{N}/n$.

Solution: A. First,

$$\log\frac{p_{\underline{q}}(\underline{x})}{p_{\underline{p}}(\underline{x})} = \log\prod_{j=1}^k \frac{q_j^{x_j}}{p_j^{x_j}} = \sum_{j=1}^k x_j \log(\frac{q_j}{p_j})\,.$$

Thus

$$K(\underline{q}, \underline{p}) = \sum_{j=1}^{k} q_j \log \frac{q_j}{p_j}\,.$$

B. From A it follows that

$$K(\widehat{\underline{p}}, \underline{p}) = \sum_{j=1}^{k} \widehat{p}_j \log \frac{\widehat{p}_j}{p_j} = -\sum_{j=1}^{k} \widehat{p}_j \log \frac{p_j}{\widehat{p}_j}\,.$$

Now

$$\begin{aligned}
\log L_n(\underline{p}|\underline{N}_n) &= \sum_{j=1}^{k} N_j \log p_j + \log\left(\frac{n!}{\prod N_j!}\right) \\
&= n\sum_{j=1}^{k} \widehat{p}_j \log p_j + \log\left(\frac{n!}{\prod N_j!}\right) \\
&= n\sum_{j=1}^{k} \widehat{p}_j \log\left(\frac{p_j}{\widehat{p}_j}\right) + n\sum_{j=1}^{k} \widehat{p}_j \log \widehat{p}_j + \log\left(\frac{n!}{\prod N_j!}\right) \\
&= -K(\widehat{\underline{p}}, \underline{p}) + \text{terms constant in } \underline{p}\,.
\end{aligned}$$

Even more neatly, as several of you noted,

$$\log \frac{L_n(\widehat{\underline{p}}|\underline{N}_n)}{L_n(\underline{p}|\underline{N}_n)} = n\sum_{j=1}^{k} \{\widehat{p}_j \log \widehat{p}_j - \widehat{p}_j \log p_j\} = nK(\widehat{\underline{p}}, \underline{p})\,.$$

C. Since $K(\widehat{\underline{p}}, \underline{p}) \ge 0$ with equality if and only if $\underline{p} = \widehat{\underline{p}}$, we see from the identity in B that $L_n(\underline{p}|\underline{N}_n)$ is maximized by $\underline{p} = \widehat{\underline{p}}$.

Do **either** problem 8 **or** problem 9.

8. (48 points). (Poisson regression). Much as in problem set 8, Suppose that $(Y|Z) \sim \text{Poisson}(\lambda e^{\gamma Z})$, and $Z \sim \text{Bernoulli}(\eta)$ on $\{0, 1\}$ You may assume that η is known. Thus Z is a "covariate" or "predictor variable", γ is a "regression parameter" which affects the intensity of the (conditionally) Poisson variable Y, and $\theta = (\lambda, \gamma)$.
(a) Find the information matrix for θ.
(b) Find the information and information bound for estimation γ if the parameter λ is unknown.
(c) Find the efficient score function and the efficient influence function for estimation of γ when λ is unknown. Interpret these in terms of the scores for γ and λ
(d) If we observe $X_i = (Y_i, Z_i)$, $i = 1, \ldots, n$, i.i.d. P_θ, write down the likelihood equations for the maximum likelihood estimator $\hat{\theta}_n = (\hat{\lambda}_n, \hat{\gamma}_n)$. What do our theorems tell us about the asymptotic normality of $\hat{\theta}_n$?

Solution: (a) Now

$$P(Y = y|Z = z) = (\lambda e^{\gamma z})^y \frac{\exp(-\lambda e^{\gamma z})}{y!}\,,$$

so

$$\log p(y,z;\lambda,\gamma) = y\log(\lambda e^{\gamma z}) - \lambda e^{\gamma z} - \text{const.}$$

and we calculate the scores and second derivatives as follows:

$$\dot{l}_\lambda(y,z) = \frac{y}{\lambda} - e^{\gamma z}, \quad \dot{l}_\gamma(y,z) = z(y-\lambda e^{\gamma z})\,;$$

$$\ddot{l}_{\lambda,\lambda}(y,z) = -\frac{y}{\lambda^2}\,, \quad \ddot{l}_{\gamma,\gamma}(y,z) = -z^2\lambda e^{\gamma z}\,, \quad \ddot{l}_{\lambda,\gamma}(y,z) = -ze^{\gamma z}\,.$$

Since $I(\theta) = -E_\theta(\ddot{l}(Y,Z))$ we find that

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

B. The information for γ when λ is unknown is

$$\begin{aligned} I_{\gamma\gamma\cdot\lambda} &= I_{\gamma\gamma} - I_{\gamma\lambda}I_{\lambda\lambda}^{-1}I_{\lambda\gamma} \\ &= \lambda\eta e^\gamma - (\eta e^\gamma)^2\frac{\lambda}{\eta e^\gamma + (1-\eta)} \\ &= \lambda\eta e^\gamma\frac{1-\eta}{\eta e^\gamma + (1-\eta)} = \lambda\eta(1-\eta) \quad \text{when} \quad \gamma = 0\,. \end{aligned}$$

The information bound for estimation of γ when λ is unknonwn is

$$1/I_{\gamma\gamma\cdot\lambda} = \frac{\eta e^\gamma + (1-\eta)}{1-\eta}\frac{1}{\lambda\eta e^\gamma}\,.$$

(c) The efficient score function for γ when λ is unknown is

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

Note that this gives, by computing conditionally on Z,

$$I_{\gamma,\gamma\cdot\lambda} = E(l_\gamma^{*2}(Y,Z)) = \lambda E\left\{e^{\gamma Z}\left(Z - \frac{\eta e^\gamma}{\eta e^\gamma + (1-\eta)}\right)^2\right\}$$

which is $\lambda E(e^{\gamma Z})$ times the variance of Z in the $\gamma-$tilted distribution corresponding to $Z \sim$ Bernoulli(η). The efficient influence function for γ when λ is unknown is

$$\tilde{l}_\gamma(y,z) = I_{\gamma,\gamma\cdot\lambda}^{-1}l_\gamma^*(y,z)\,.$$

Both l_γ^* and $\tilde{l}_\gamma$ are orthogonal to $\dot{l}_\lambda$ in $L_2(P_\theta)$.
(d) The likelihood equations are given by

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

The solution $\widehat{\theta}_n = (\widehat{\lambda}_n, \widehat{\gamma}_n)$, which exists with probability converging to 1 as $n \to \infty$, is the MLE of $\theta = (\lambda, \gamma)$. Theorem 4.1.5 tells us that when P_{θ_0} is true,

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d I(\theta_0)^{-1} Z \sim N_2(0, I(\theta_0)^{-1})$$

where $I(\theta_0)$ is the information matrix (at θ_0) calcualted in (a).

9. (48 points). (Poisson regression, continued).
(a) Suggest three tests of the (composite!) null hypothesis $H : \gamma = 0$ versus $K : \gamma \neq 0$. What is the distribution of each of these three statistics under the null hypothesis and under local alternatives of the form $\gamma_n = tn^{-1/2}$?
(b) Consider estimation of the function

$$q(\theta) = \nu(P_\theta) = P_\theta(Y = 0)\,.$$

Compute $q(\theta)$ explicitly as a function of θ.
(c) Suggest a natural empirical estimator of this probability (which does not rely on the Poisson model). If this estimator is called $\tilde{\nu}_n$, show that $\tilde{\nu}_n$ is asymptotically linear and find its influence function ψ explicitly.
(d) Find the efficient influence function $\tilde{l}_\nu$ for estimation of $\nu(P_\theta)$ assuming the Poisson model.
(e) 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 Wald, likelihood ratio, and score statistics:

$$\begin{aligned} W_n &= n^{1/2}\widehat{\gamma}_n \widehat{I}_{\gamma\gamma\cdot\lambda}(\widehat{\lambda}_n, \widehat{\gamma}_n)(n^{1/2}\widehat{\gamma}_n)\,, \\ 2\log\lambda_n &= 2\log\frac{\prod_{i=1}^n p(Y_i, Z_i; \widehat{\theta}_n)}{\prod_{i=1}^n p(Y_i, Z_i; \widehat{\theta}_n^0)}\,, \\ R_n &= Z_n(\widehat{\theta}_n^0)^T I^{-1}(\widehat{\theta}_n^0) Z_n(\widehat{\theta}_n^0)\,; \end{aligned}$$

here $\widehat{\theta}_n = (\widehat{\lambda}_n, \widehat{\gamma}_n)$ solve the likelihood equations in problem 8(d) above while $\widehat{\theta}_n^0 = (\overline{Y}_n, 0)$, and

$$\underline{Z}_n(\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{pmatrix} (Y_i - \lambda e^{\gamma Z_i})/\lambda \\ Z_i(Y_i - \lambda e^{\gamma Z_i}) \end{pmatrix}\,,$$

so that

$$\underline{Z}_n(\widehat{\theta}_n^0) = \begin{pmatrix} 0 \\ n^{-1/2}\sum_{i=1}^n Z_i(Y_i - \overline{Y}) \end{pmatrix}\,.$$

Hence the Rao statistic reduces to

$$R_n = \frac{\left\{n^{-1/2}\sum_{i=1}^n Z_i(Y_i - \overline{Y})\right\}^2}{\widehat{I}_{\gamma\gamma\cdot\lambda}(\widehat{\theta}_n^0)}\,.$$

One easy estimator for the denominator is $\overline{Y}\eta(1-\eta)$; another is $\overline{YZ}(1-\overline{Z})$. Under H all three statistics converge in distribution to χ_1^2. Under local alternatives of the form $\gamma_n = tn^{-1/2}$

$$W_n,\ 2\log\lambda_n,\ R_n \ \to_d \ \chi_1^2(\delta)$$

where $\delta = t^2 I_{\gamma\gamma\cdot\lambda)=t^2\lambda\eta(1-\eta)}$ (in views of problem 8(b)).
(b) The parameter $q(\theta)$ is easily computed by conditioning on Z:

$$\begin{aligned} q(\theta) &= P_\theta(Y=0) = E\{P(Y=0|Z)\} \\ &= E\{\exp(-\lambda e^{\gamma Z})\} \\ &= \eta \exp(-\lambda e^\gamma) + (1-\eta)\exp(-\lambda)\,. \end{aligned}$$

(c) A natural empirical estimator of $\nu(P_\theta) = P_\theta(Y=0)$ is $\nu(\mathbb{P}_n) = \mathbb{P}_n(Y=0) = n^{-1}\sum_{i=1}^n 1_{[Y_i=0]}$. Now

$$\sqrt{n}(\nu(\mathbb{P}_n) - \nu(P_\theta)) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (1_{[Y_i=0]} - P(Y=0)) \rightarrow_d N(0, P(Y=0)(1-P(Y=0))\,,$$

so $\tilde{\nu}_n = \nu(\mathbb{P}_n)$ is asymptotically linear with influence function

$$\psi(y,z) = 1_{[y=0]} - P(Y=0)\,.$$

The efficient influence function $\tilde{l}_\nu$ for estimation of $q(\theta) = \nu(P_\theta)$ is given by

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

where

$$\dot{q}(\theta) = \begin{pmatrix} -e^\gamma \eta \exp(-\lambda e^\gamma) - (1-\eta)e^{-\lambda} \\ -\lambda e^\gamma \eta \exp(-\lambda e^\gamma) \end{pmatrix}$$

and the information matrix is as given in the solution for problem 8(a).
(d) The efficient influence function $\tilde{l}_\nu$ is the projection of ψ onto $\dot{\mathcal{P}} = [\dot{l}_\lambda, \dot{l}_\gamma]$ in $L_2(P_\theta)$ as shown in the following figure.