**Statistics 581, Final Exam Solutions**

Wellner; 12/13/2002

1. (40) points) **Define** *each* of the following terms. In each case, briefly provide an appropriate context for your definition.
   (a) The information matrix for $\theta$ in a regular parametric model
   $\mathcal{P} = \{P_\theta : \ \theta \in \Theta \subset R^k\}$.
   (b) The efficient score function for a parameter $\theta_1$ when $\theta = (\theta_1, \theta_2)$.
   (c) The efficient influence function $\tilde{l}_1$ for a parameter $\theta_1$ 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 $\psi$.

   **Solution:** See chapter 3 of the course notes.

2. (32 points) **State** *four* of the following five results, providing the appropriate (brief) context for your statement:
   (a) The (elementary) Skorokhod theorem.
   (b) The Cramér - Wold device.
   (c) A result about the finite-dimensional limiting distributions of the sample quantile process $\{\sqrt{n}(\mathbb{F}_n^{-1}(t) - F^{-1}(t)) : \ 0 < t < 1\}$ specifying the assumption(s) carefully.
   (d) The Mann-Wald (or continuous mapping) theorem.
   (e) The Glivenko-Cantelli theorem.

   **Solution:** Se Chapter 2 of the course notes.

3. (50 points).
   Suppose that $P = P_0 = N(0,1)$, $Q = P_\theta = N(\theta, 1)$ on $(\mathbb{X}, \mathcal{A}) = (\mathbb{R}, \mathcal{B})$.
   (a) Compute $K(P,Q) = K(P_0, P_\theta)$.
   (b) Compute $H^2(P,Q) = 1 - \rho(P,Q)$ and $\rho(P,Q) = \int \sqrt{p(x)q(x)}dx$. [It might be easiest to compute $\rho(P,Q)$ first recalling that if $Z \sim N(0,1)$ then $E\exp(tZ) = \exp(t^2/2)$.]
   (c) Compute $d_{TV}(P,Q) = 1 - \eta(P,Q)$ and $\eta(P,Q) = \int p(x) \wedge q(x)dx$. [It might be easiest to compute $\eta(P,Q)$ first.]
   (d) Show in general that $K(P,Q) \geq 2H^2(P,Q)$, thereby strengthening the fact $K(P,Q) \geq 0$ that we proved in class. [Hint: write both $K(P,Q)$ and $H^2(P,Q)$ in terms of $Y = (p(X)/q(X))^{1/2}$ and use the inequality $\log(1+x) \geq x/(1+x)$ for $x \geq 0$. You will need to relate $E_Q Y$ and $E_Q Y^2$ to $H^2(P,Q)$.]
   (e) Use the results of (a) and (d) to find a lower bound for $K(P^n, Q^n)$ in terms of $H^2(P,Q)$ or $\rho(P,Q)$; here $P^n$ and $Q^n$ are the probability distributions of $X_1, \ldots, X_n$ i.i.d. as $P$ and $Q^n$ respectively.

   **Solution:** (a) Now $p(x) = \phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ and $q(x) = (2\pi)^{-1/2}\exp(-(x-\theta)^2/2)$, so

$$\frac{p(x)}{q(x)} = \exp(-x^2/2 + (x-\theta)^2/2) = \exp(-\theta x + \theta^2/2),$$

$$\log \frac{p}{q}(x) = -\theta x + \theta^2/2,$$

and it follows that

$$K(P,Q) = E_P \left( \log \frac{p}{q} \right) = -E_P(X) + \theta^2/2 = \theta^2/2\,.$$

(b) We compute $\rho(P,Q)$ first:

$$\begin{aligned}\rho(P,Q) = \int \sqrt{p(x)q(x)}dx &= \int \frac{1}{\sqrt{2\pi}} \exp(-x^2/4)\exp(-(x-\theta)^2/4)dx \\ &= \int \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)\exp((2\theta x - \theta^2)/4)\,dx \\ &= \exp(-\theta^2/4) E_P \exp(\theta X/2) = \exp(-\theta^2/4)\exp(\theta^2/8) = \exp(-\theta^2/8)\,.\end{aligned}$$

Hence

$$H^2(P,Q) = 1 - \rho(P,Q) = 1 - \exp(-\theta^2/8)\,.$$

(c) We compute $\eta(P,Q)$ first. Since $\phi(x) \ge \phi(x-\theta)$ if and only if

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \ge \frac{1}{\sqrt{2\pi}} \exp(-(x-\theta)^2/2)$$

or equivalently, if and only if

$$1 \ge \exp(\theta x - \theta^2/2) \qquad \text{iff} \qquad \theta x - \theta^2/2 \le 0$$

and, for $\theta > 0$, this holds if and only if $x \le \theta/2$. Hence it follows that

$$\begin{aligned}\int p(x) \wedge q(x) dx &= \int \phi(x) \wedge \phi(x-\theta) dx \\ &= \int_{-\infty}^{\theta/2} \phi(x-\theta)dx + \int_{\theta/2}^{\infty} \phi(x)dx \\ &= \int_{-\infty}^{-\theta/2} \phi(y)dy + 1 - \Phi(\theta/2) \\ &= \Phi(-\theta/2) + 1 - \Phi(\theta/2)\,, \qquad \text{if } \ \theta > 0\,.\end{aligned}$$

When $\theta < 0$, $\theta x - \theta^2/2 < 0$ if and only if $x \ge \theta/2$, and this yields

$$\begin{aligned}\int p(x) \wedge q(x) dx &= \int_{\theta/2}^{\infty} \phi(x-\theta)dx + \int_{-\infty}^{\theta/2} \phi(x)dx \\ &= \Phi(\theta/2) + 1 - \Phi(-\theta/2)\,, \qquad \text{if } \ \theta > 0\,.\end{aligned}$$

It follows that

$$d_{TV}(P,Q) = 1 - \eta(P,Q) = \begin{cases} \Phi(\theta/2) - \Phi(-\theta/2), & \theta > 0 \\ \Phi(-\theta/2) - \Phi(\theta/2), & \theta < 0\,. \end{cases}$$

The following Figure shows $K(P_0, P_\theta)$, $H^2(P_0, P_\theta)$, and $d_{TV}(P_0, P_\theta)$ as functions of $\theta$.

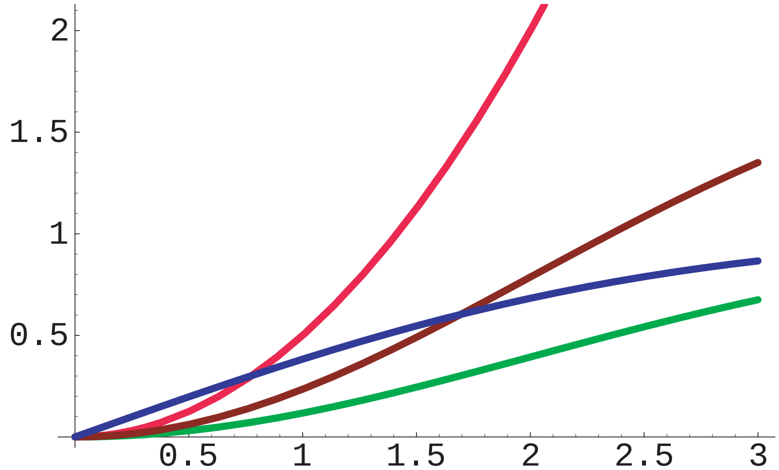


Figure 1: $K(P_0, P_\theta)$ (red), $H^2(P_0, P_\theta)$ (green), $d_{TV}(P_0, P_\theta)$ (blue), and $2H^2(P_0, P_\theta)$ (burgundy) as functions of $\theta$

.

(d) The problem is slightly easier with a different choice of $Y$, as follows. (My apologies.) Let $Y \equiv \sqrt{p/q} - 1$. Then $p/q = (1+Y)^2$ and it follows that

$$\begin{aligned} K(P,Q) &= \int p \log(p/q) d\mu = \int (p/q) \log(p/q) q d\mu \\ &= 2 \int (p/q) \log(p/q)^{1/2} dQ = 2 \int (1+Y)^2 \log(1+Y) dQ \\ &\geq 2 \int (1+Y)^2 \frac{Y}{1+Y} dQ \\ &= 2 \int Y(1+Y) dQ = 2 \left\{ \int Y dQ + \int Y^2 dQ \right\}. \end{aligned}$$

But we also have

$$\begin{aligned} \int Y dQ &= \int \sqrt{pq} d\mu - 1 = -H^2(P,Q), \\ \int Y^2 dQ &= \int [\sqrt{p/q} - 1]^2 q d\mu = \int [\sqrt{p} - \sqrt{q}]^2 d\mu = 2H^2(P,Q). \end{aligned}$$

By combining the results of these last two displays it follows that

$$K(P,Q) \geq 2\left\{2H^2(P,Q) - H^2(P,Q)\right\} = 2H^2(P,Q).$$

(e) By (d), and (c) of the next problem,

$$K(P^n, Q^n) \geq 2H^2(P^n, Q^n) = 2\{1 - \rho(P^n, Q^n)\} = 2\{1 - \rho(P,Q)^n\}$$

in general. For the particular case we began with $\rho(P,Q) = \exp(-\theta^2/8)$, and thus we conclude that in this case

$$K(P^n, Q^n) \geq 2\{1 - \exp(-n\theta^2/8)\}.$$

4. (33 points). Prove that:
(a) $d_{TV}(P,Q) \equiv \sup_{A \in \mathcal{A}} |P(A) - Q(A)| = (1/2) \int |p-q| d\mu$.
(b) $H^2(P,Q) \le d_{TV}(P,Q)$.
(c) $\rho(P^n, Q^n) = \rho(P,Q)^n$ where $\rho(P,Q) = \int \sqrt{pq} d\mu$ and $P^n$ and $Q^n$ are the distributions of $X_1, \ldots, X_n$ i.i.d. as $P$ or $Q$ respectively.

**Solution:** (a) Let $\delta \equiv p - q$ so that $\delta = \delta^+ - \delta^-$ and $|\delta| = \delta^+ + \delta^-$. Noting that

$$0 = \int (p-q) d\mu = \int \delta d\mu = \int (\delta^+ - \delta^-) d\mu = \int \delta^+ d\mu - \int \delta^- d\mu \,,$$

it follows that $\int \delta^+ d\mu = \int \delta^- d\mu$ and, moreover, that

$$\int |p-q| d\mu = \int (\delta^+ + \delta_-) d\mu = 2 \int \delta^+ d\mu \,.$$

Hence for any set $A \in \mathcal{A}$,

$$\begin{aligned} |P(A) - Q(A)| &= |\int_A \delta d\mu| = |\int_A \delta^+ d\mu - \int_A \delta^- d\mu| \\ &\le \int_A \delta^+ d\mu \le \int \delta^+ d\mu = \frac{1}{2} \int |p-q| d\mu \,. \end{aligned}$$

Noting that equality holds when $A = \{x : \ p(x) \ge q(x)\}$, the result follows.
(b) Note that

$$\begin{aligned} H^2(P,Q) &= \frac{1}{2} \int [\sqrt{p} - \sqrt{q}]^2 d\mu = \frac{1}{2} \int |\sqrt{p} - \sqrt{q}| |\sqrt{p} - \sqrt{q}| d\mu \\ &\le \frac{1}{2} \int |\sqrt{p} - \sqrt{q}| |\sqrt{p} + \sqrt{q}| d\mu = \frac{1}{2} \int |p-q| d\mu = d_{TV}(P,Q) \,. \end{aligned}$$

(c) Now by direct computation via the Tonelli part of the Fubini-Tonelli theorem,

$$\begin{aligned} \rho(P^n, Q^n) &= \int \cdots \int \sqrt{p(x_1) \cdots p(x_n) q(x_1) \cdots q(x_n)} d\mu(x_1) \cdots d\mu(x_n) \\ &= \int \cdots \int \sqrt{p(x_1) q(x_1)} \cdots \sqrt{p(x_n) q(x_n)} d\mu(x_1) \cdots d\mu(x_n) \\ &= \int \sqrt{p(x_1) q(x_1)} d\mu(x_1) \cdots \int \sqrt{p(x_n) q(x_n)} d\mu(x_n) \\ &= \rho(P,Q)^n \,. \end{aligned}$$

5. (54 points)
Suppose that $X, X_1, \ldots, X_n$ are i.i.d. with distribution function $F_\theta$ given by $F_\theta(x) = G(x - \theta)$ for $\theta \in R$ where $G(x) = 1 - 1/(1+x)^{7/2}$, $x \ge 0$, $G(x) = 0$ for $x \le 0$.
(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 $X_{n:1} \equiv X_{(1)} \equiv \min_{1 \le i \le n} X_i$. Show that $X_{n:1} \to_p \theta$.
(d) Show that $n(X_{n:1} - \theta) \to_d$ "something" and find the limiting disribution. [Hint: Start by computing $P(n(X_{n:1} - \theta) > t)$.]
(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 $\theta$ based on the $t-$th sample quantile $\mathbb{F}_n^{-1}(t)$.
(f) Show that the estimators $\hat{\theta}_n(t)$ of $\theta$ which you derived in (e) satisfy

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

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) Now

$$\begin{aligned} E|X|^r &= \int |x|^r dF_\theta(x) = \int |x|^r dG(x-\theta) = \int |y+\theta|^r dG(y) \\ &= \int_0^\infty |y+\theta|^r \frac{7/2}{[1+y]^{9/2}} dy \end{aligned}$$

and this integral converges if $9/2 - r > 1$, or equivalently if $r < 7/2$.
(b) Differentiating $G$ yields

$$g(x) = \frac{7}{2(1+x)^{9/2}}, \qquad x \ge 0\,.$$

Both $G$ and $g$ are shown in the following figure:
(c) Let $\epsilon > 0$, and note that $X_i - \theta \equiv Y_i$ are i.i.d. with distribution function $G$. Now $P(X_{n:1} < \theta - \epsilon) = 0$ for all $n$, and, on the other hand,

$$\begin{aligned} P_\theta(X_{n:1} > \theta + \epsilon) &= P(X_{n:1} - \theta > \epsilon) \\ &= P(Y_{n:1} > \epsilon) = P(Y_1 > \epsilon, \ldots, Y_n > \epsilon) \\ &= P(Y_1 > \epsilon)^n = \frac{1}{(1+\epsilon)^{7n/2}} \to 0\,, \end{aligned}$$

and hence $X_{n:1} \to_p \theta$.
(d) By a calculation just like that in (c),

$$\begin{aligned} P_\theta(n(X_{n:1} - \theta) > t) &= P_\theta(X_{n:1} > \theta + t/n) \\ &= P(Y_{n:1} > t/n) = \frac{1}{1 + t/n)^{7n/2}} \\ &\to (e^{-t})^{7/2} = \exp(-7t/2)\,. \end{aligned}$$

It follows that $n(X_{n:1} - \theta) \to_d$ Exponential(7/2).
(e) Since $F_\theta(x) = G(x - \theta) = t$, solving for $x$ we find that $x = F_\theta^{-1}(t) = \theta + G^{-1}(t)$ where

$$G^{-1}(t) = (1-t)^{-2/7} - 1\,.$$

Since $\mathbb{F}_n^{-1}(t)$ consistently estimates $F_\theta^{-1}(t) = \theta + G^{-1}(t)$, we see that

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

(f) Now

$$\begin{aligned}\sqrt{n}(\mathbb{F}_n^{-1}(t) - G^{-1}(t) - \theta) &= \sqrt{n}(\mathbb{F}_n^{-1}(t) - F_\theta^{-1}(t)) \\ &\to_d Q_\theta'(t)\mathbb{V}(t) \sim N(0, t(1-t)[Q_\theta'(t)]^2),\end{aligned}$$

where

$$Q_\theta(t) = F_\theta^{-1}(t) = (1-t)^{-2/7} - 1 + \theta .$$

Hence

$$Q_\theta'(t) = \frac{2}{7}(1-t)^{-9/7}, \qquad \text{and} \qquad [Q_\theta'(t)]^2 = \frac{4}{49}(1-t)^{-18/7},$$

and it follows that

$$V_t^2 = t(1-t)[Q_\theta'(t)]^2 = \frac{4}{49}t(1-t)^{-11/7} .$$

This is a monotone increasing function of $t$ (as can be shown by simple calculation), and hence $V_t^2$ is minimized by taking $t = 0$ (which yields a nonsensical estimator). This is, however, not surprising in view of the fact that $X_{n:1}$ estimates $\theta$ at rate $n$ as we showed in part (d).

**Remark:** The sub text to this problem is that this is a parametric model for which the hypotheses of our theorem 4.2 fail: in the present case the set $A = \{x : f_\theta(x) > 0\} = [\theta, \infty)$ clearly depends on $\theta$. Moreover, the information for $\theta$ is either undefined or $\infty$ depending on ones perspective, and the theory developed in chapters 3 and 4 does not apply. On the other hand, a direct analysis is not difficult and it is clear that it is fairly easy to estimate $\theta$ in this problem (since the density is 0 to the left of $\theta$, but strictly positive at $\theta$), and as we showed in (d), the rate of convergence of the first order statistic as an estimator of $\theta$ is $n^{-1}$.

6. (50 points) (A parametric version of the Cox model; a simpler version of problem 2, problem set 8).
Suppose that $(Y|Z) \sim \text{Exponential}(\lambda e^{\gamma Z})$ where $Z \sim \text{Bernoulli}(\eta)$. 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 $\theta$.
(c) Compute the information for $\gamma$ when $\lambda$ 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 $\theta$ 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:

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

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 $\theta$ 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 $\gamma$ when $\lambda$ is known is $I_{11} = E(Z^2)$. The information for $\gamma$ when $\lambda$ 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 $\gamma$ 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 $\gamma$ 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 $\theta$.

(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 $\psi$ explicitly.

(e) Find the efficient influence function $\tilde{l}_\nu$ for estimation of $\nu(P_\theta)$ assuming the Poisson model.

(f) Describe the relationship between $\psi$ 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 $\chi_1^2$. 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 $\lambda$ 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

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

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

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

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

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

where

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

(d) 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

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

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