Statistics 581, Final Exam Solutions
Wellner; 12/13/2006

1. (48) 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.
   (f) A locally regular estimator T_n of a parameter $\nu(P_\theta)$.

   Solution: see Chapter 3 and 4 of the notes.

2. (32 points) **State** *four* of the following six results, providing the appropriate (brief) context for your statement:
   (a) The (elementary) Skorokhod theorem.
   (b) The multiparameter Cramér - Rao inequality (for an unbiased estimator) $T = T(\underline{X})$ of a real-valued parameter $q(\theta) = \nu(P_\theta)$.
   (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 Glivenko-Cantelli theorem.
   (e) LAN (Local Asymptotic Normality) of the local - log likelihood ratios for a regular parametric model satisfying the Cramér hypotheses).
   (f) The asymptotic behavior of the likelihood ratio statistic $2\log\lambda_n$ for testing a simple null hypothesis $\theta = \theta_0$ versus $\theta \ne \theta_0$ under a fixed alternative P_θ with $\theta \ne \theta_0$.

   Solution: see Chapters 2-4 of the course notes.

3. (40 points) Suppose that $\mathcal{P} = \{P_\theta : \ \theta \in (0,1)\}$ where P_θ has density with respect to Lebesgue measure λ on $[0,1]$ given by

$$p_\theta(x) = 2\left\{\frac{x}{\theta}1_{[0,\theta]}(x) + \frac{1-x}{1-\theta}1_{(\theta,1]}(x)\right\}.$$

   We showed in problem 5 of problem set # 9 that

$$\begin{aligned}\dot{\mathbf{l}}_\theta(x) &= -\frac{1}{\theta}1_{[0,\theta]}(x) + \frac{1}{1-\theta}1_{(\theta,1]}(x),\\ I(\theta) &= \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta(1-\theta)},\end{aligned}$$

and that if $X_1, \ldots, X_n$ are i.i.d. P_θ with $\theta \in (0,1)$, then a $\sqrt{n}-$consistent preliminary estimator of θ is given by $\overline{\theta}_n = 3\overline{X}_n - 1$ which satisfies

$$\sqrt{n}(\overline{\theta}_n - \theta) \to_d N(0, (1-\theta+\theta^2)/2).$$

(a) Use the above facts to suggest a one-step estimator $\check{\theta}_n$ of θ.
(b) Show that the estimator $\check{\theta}_n$ you proposed in (a) can be written as

$$\check{\theta}_n = \overline{\theta}_n - (\mathbb{F}_n(\overline{\theta}_n) - \overline{\theta}_n)$$

where $\mathbb{F}_n(x) = n^{-1}\sum_{i=1}^n 1_{[X_i \le x]}$ is the empirical distribution of the X_i's.
(c) Show by a direct argument using the result of (b) that $\check{\theta}_n \to_p \theta$. Hints: Calculate the distribution function F_θ corresponding to the density p_θ, show that $F_\theta(\theta) = \theta$, and use the Glivenko-Cantelli theorem.
(d) Show by a direct argument using the result of (b) that $\sqrt{n}(\check{\theta}_n - \theta) \to_d N(0, I(\theta)^{-1})) = N(0, \theta(1-\theta))$. Hint: Use Donsker's theorem and the delta-method applied to $g(x) = F_\theta(x) - x$.
(e) Compare the asymptotic variance of the preliminary estimator $\overline{\theta}_n$ to the asymptotic variance of the one- step estimator $\check{\theta}_n$.

Solution: (a) Given the preliminary estimator $\overline{\theta}_n = 3\overline{X}_n - 1$, a natural one-step estimator of θ is given by

$$\begin{aligned}
\check{\theta}_n &= \overline{\theta}_n + I(\overline{\theta}_n)^{-1} n^{-1} \dot{\mathbf{l}}_n(\overline{\theta}_n) \\
&= \overline{\theta}_n + \overline{\theta}_n(1-\overline{\theta}_n)\left\{ -\frac{1}{\overline{\theta}_n}\mathbb{F}_n(\overline{\theta}_n) + \frac{1}{1-\overline{\theta}_n}(1-\mathbb{F}_n(\overline{\theta}_n)) \right\}.
\end{aligned}$$

(b) From the last display in (a) we see that

$$\begin{aligned}
\check{\theta}_n &= \overline{\theta}_n + \overline{\theta}_n(1-\overline{\theta}_n)\left\{ -\frac{1}{\overline{\theta}_n}\mathbb{F}_n(\overline{\theta}_n) + \frac{1}{1-\overline{\theta}_n}(1-\mathbb{F}_n(\overline{\theta}_n)) \right\} \\
&= \overline{\theta}_n - (1-\overline{\theta}_n)\mathbb{F}_n(\overline{\theta}_n) + \overline{\theta}_n(1-\mathbb{F}_n(\overline{\theta}_n)) \\
&= \overline{\theta}_n - (\mathbb{F}_n(\overline{\theta}_n) - \overline{\theta}_n).
\end{aligned}$$

(c) Since $\overline{\theta}_n \to_{p,a.s.} \theta$, and

$$F_\theta(x) = \frac{x^2}{\theta}1_{[x\le\theta]} + (1 - \frac{(1-x)^2}{1-\theta})1_{[\theta<x<1]}$$

satisfies $F_\theta(\theta) = \theta$, it follows from the last display in (b) that

$$\begin{aligned}
\check{\theta}_n &= \overline{\theta}_n - (\mathbb{F}_n(\overline{\theta}_n) - \overline{\theta}_n) \\
&= \overline{\theta}_n - (\mathbb{F}_n(\overline{\theta}_n) - F_\theta(\overline{\theta}_n)) - (F_\theta(\overline{\theta}_n) - \overline{\theta}_n) \\
&\to_p \theta - 0 - 0;
\end{aligned}$$

here the second 0 follows from the Glivenko-Cantelli theorem since

$$|\mathbb{F}_n(\overline{\theta}_n) - F_\theta(\overline{\theta}_n)| \le \sup_x |\mathbb{F}_n(x) - F_\theta(x)| \to_{a.s.} 0.$$

(d) Letting $g(x) \equiv F_\theta(x) - x$, so that $g(\theta) = 0$ and $g'(x) = p_\theta(x) - 1$ has $g'(\theta) = p_\theta(\theta) - 1 = 2 - 1 = 1$, we can write

$$\begin{aligned}
\sqrt{n}(\check{\theta}_n - \theta) &= \sqrt{n}(\overline{\theta}_n - \theta) - \sqrt{n}(\mathbb{F}_n(\overline{\theta}_n) - \overline{\theta}_n) \\
&= \sqrt{n}(\overline{\theta}_n - \theta) - \sqrt{n}(\mathbb{F}_n(\overline{\theta}_n) - F_\theta(\overline{\theta}_n)) \\
&\qquad -\sqrt{n}(F_\theta(\overline{\theta}_n)) - \overline{\theta}_n) \\
&= \sqrt{n}(\overline{\theta}_n - \theta) - \sqrt{n}(\mathbb{F}_n(\overline{\theta}_n) - F_\theta(\overline{\theta}_n)) \\
&\qquad -\sqrt{n}(g(\overline{\theta}_n)) - g(\theta)) \\
&= \sqrt{n}(\overline{\theta}_n - \theta)\,\{1 - g'(\theta_n^*)\} - \sqrt{n}(\mathbb{F}_n(\overline{\theta}_n) - F_\theta(\overline{\theta}_n)) \\
&= O_p(1)o_p(1) - \sqrt{n}(\mathbb{F}_n(\overline{\theta}_n) - F_\theta(\overline{\theta}_n)) \\
&\to_d -\mathbb{U}(F_\theta(\theta)) = -\mathbb{U}(\theta) \sim N(0, \theta(1-\theta));
\end{aligned}$$

here $|\theta_n^* - \theta| \le |\overline{\theta}_n - \theta| \to_p 0$, and the convergence in the last line of the display follows from Donsker's theorem:

$$\sqrt{n}(\mathbb{F}_n - F) = \sqrt{n}(\mathbb{F}_n - F_\theta) \stackrel{d}{=} \mathbb{U}_n(F_\theta) \Rightarrow \mathbb{U}(F_\theta)$$

where $\mathbb{U}$ is a standard Brownian bridge process on $(0,1)$.
(e) The asymptotic variance function $V^2(\theta) = (1 - \theta + \theta^2)/2$ of the preliminary estimator $\overline{\theta}_n$ of θ has $V^2(0) = V^2(1) = 1/2$, while $V^2(\theta)$ is a minimum at $\theta = 1/2$ when it equals $3/8$. On the other hand, the asymptotic variance $I(\theta)^{-1} = \theta(1-\theta)$ of the one-step estimator $\check{\theta}_n$ of θ is 0 at $\theta = 0$ and 1, and is a maximum at $\theta = 1/2$ where it equals $1/4$; note that this maximum $1/4$ is less than the minimum variance of the one-step estimator, namely $3/8$. The gains achieved near $\theta = 0$ and $\theta = 1$ are substantial. here is a plot comparing the two variance functions:

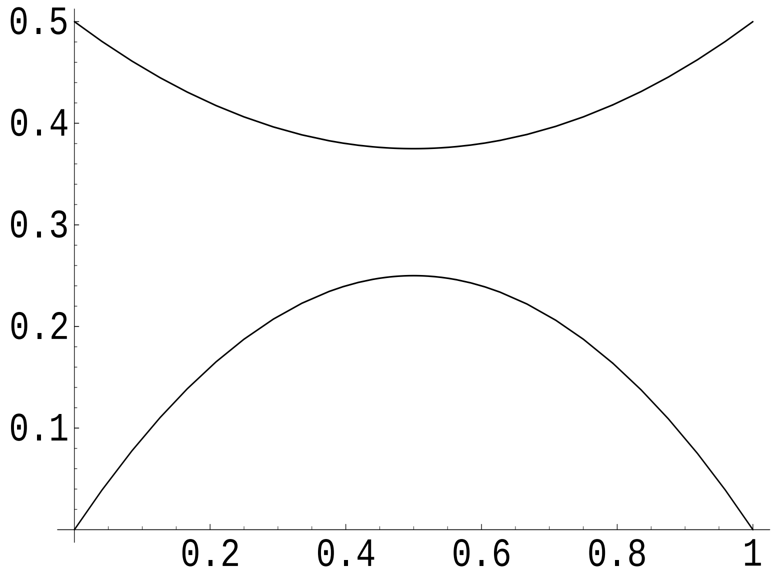

Figure 1: Plot of the two variance functions $(1-\theta+\theta^2)/2$ and $\theta(1-\theta)$

4. (40 points) Suppose that $X_1,\ldots,X_n$ are i.i.d. $N(\mu,\sigma^2)$, and let $\theta=(\mu,\sigma^2)\in\Theta\equiv\mathbb{R}\times(0,\infty)$. Consider testing $H:\ \sigma^2=\sigma_0^2$ where σ_0^2 is a known constant, versus $K:\ \sigma^2\neq\sigma_0^2$.
(a) What are the score functions for μ and σ^2, and information matrix for θ in this model?
(b) Identify the null hypothesis H in terms of a subset Θ_0 of Θ.
(c) What is the maximum likelihood estimator $\widehat{\theta}_n$ of $\theta=(\mu,\sigma^2)$ under $\theta\in\Theta$? What is the maximum likelihood estimator $\widehat{\theta}_n^0$ of $\theta=(\mu,\sigma^2)$ under $\theta\in\Theta_0$?
(d) What are the likelihood ratio, Wald, and score (or Rao) statistics for testing H versus K? What is the limiting distribution of these three test statistics under the null hypothesis H?
(e) Suppose that $X_1,\ldots,X_n$ are i.i.d. P_θ with $\theta\notin\Theta_0$. Find the limits of $n^{-1}2\log\lambda_n$, $n^{-1}W_n$, $n^{-1}R_n$ (almost surely or in probability).

Solution: (a) As in example 4.3.1 of the course notes,

$$\log p_\theta(x) = -\frac{1}{2}\log\sigma^2-\frac{1}{2\sigma^2}(x-\mu)^2-\frac{1}{2}\log(2\pi),$$

so the score functions $\dot{\mathbf{l}}_\mu$ and $\dot{\mathbf{l}}_{\sigma^2}$ are given by

$$\begin{aligned}\dot{\mathbf{l}}_\mu(x) &= \frac{1}{\sigma^2}(x-\mu),\\ \dot{\mathbf{l}}_{\sigma^2}(x) &= \frac{(x-\mu)^2}{2\sigma^4}-\frac{1}{2\sigma^2},\end{aligned}$$

and the information matrix for $\theta = (\mu, \sigma^2)$ is

$$I(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}.$$

(b) Under Θ, the MLE of θ is $\widehat{\theta}_n = (\overline{X}_n, S_n^2)$ where $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$, $S_n^2 = n^{-1}\sum_{i=1}^n (X_i - \overline{X})^2$. Under Θ_0, the MLE of θ is $\widehat{\theta}_n^0 = (\overline{X}_n, \sigma_0^2)$.
(c) The parameter space $\Theta = \mathbb{R} \times (0, \infty)$, the upper half-plane in $\mathbb{R}^2$, and the null hypothesis consists of the horizontal line (μ, σ_0^2) at constant height σ_0^2 above the x axis.
(d) The likelihood ratio statistic for testing H versus K is given by

$$\begin{aligned} 2\log\lambda_n &= 2\log\frac{L_n(\overline{X}_n, S_n^2)}{L_n(\overline{X}_n, \sigma_0^2)} \\ &= 2\{l_n(\overline{X}_n, S_n^2) - l_n(\overline{X}_n, \sigma_0^2)\} \\ &= 2\left\{-\frac{1}{S_n^2}\sum_{i=1}^n (X_i - \overline{X}_n)^2 - \frac{n}{2}\log S_n^2 \right. \\ &\qquad \left. - \left(-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (X_i - \overline{X}_n)^2 - \frac{n}{2}\log\sigma_0^2\right)\right\} \\ &= n\left\{\left(\frac{S_n^2}{\sigma_0^2} - 1\right) - \log\left(\frac{S_n^2}{\sigma_0^2}\right)\right\} \\ &= ng(S_n^2/\sigma_0^2) \end{aligned}$$

where $g(x) \equiv x - 1 - \log x$ has $g(1) = 0$, $g'(1) = 0$, and $g''(1) = 1$.
The Wald statistic W_n for testing H versus K is given by

$$\begin{aligned} W_n &= \sqrt{n}(S^2 - \sigma_0^2)\hat{I}_{22\cdot 1}\sqrt{n}(S^2 - \sigma_0^2) \\ &= \sqrt{n}(S^2 - \sigma_0^2)\frac{1}{2S_n^4}\sqrt{n}(S^2 - \sigma_0^2) \\ &= \frac{n}{2}\left(\frac{S_n^2 - \sigma_0^2}{S_n^2}\right)^2 = \frac{n}{2}\left(1 - \frac{\sigma_0^2}{S_n^2}\right)^2, \end{aligned}$$

while the Rao (or score) statistic is

$$\begin{aligned} R_n &= Z_n(\hat{\theta}_n^0)^T I(\hat{\theta}_n^0)^{-1} Z_n(\hat{\theta}_n^0) \\ &= \left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\frac{(X_i - \overline{X}_n)^2}{2\sigma_0^4} - \frac{1}{2\sigma_0^2}\right)\right\}^2 \hat{I}_{22\cdot 1}^{-1} \\ &= \left\{\sqrt{n}\left(\frac{S_n^2}{\sigma_0^2} - 1\right)\right\}^2 \frac{1}{(2\sigma_0^2)^2} 2\sigma_0^4 \end{aligned}$$

$$= \left\{ \sqrt{\frac{n}{2}} \left(\frac{S_n^2}{\sigma_0^2} - 1 \right) \right\}^2 .$$

Under the null hypothesis H we have $2 \log \lambda_n$, W_n, and R_n all $\to_d \chi_1^2$.
(e) Under a fixed alternative $\theta \notin \Theta_0$ we have $S_n^2 \to_p \sigma^2 \neq \sigma_0^2$, and hence by the Mann-Wald or continuous mapping theorem

$$n^{-1} 2 \log \lambda_n \to_p g(\sigma^2/\sigma_0^2) > 0,$$
$$n^{-1} W_n \to_p \frac{1}{2} \left(1 - \frac{\sigma_0^2}{\sigma^2} \right)^2 > 0,$$
$$n^{-1} R_n \to_p \frac{1}{2} \left(\frac{\sigma^2}{\sigma_0^2} - 1 \right)^2 > 0.$$

Comment: Many of you answered this as if Θ_0 was a *simple* null hypthesis, but here it is *composite*, so the general result in Theorem 4.2.1, page 11 of the chapter 4 notes, does not apply. A comparable general result for Θ_0 composite is not included in Section 4.2. Can you formulate the correct general result for Θ_0 composite?

5. (48 points). (Poisson regression). 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 and that $0 < \eta < 1$. 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 geometrically 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:

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

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

Since $I(\theta) = -E_\theta(\ddot{\mathbf{l}}(Y,Z))$ 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 unknown 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

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

Note that this gives, by computing conditionally on Z,

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

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

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

Both $\mathbf{l}^*_\gamma$ and $\tilde{\mathbf{l}}_\gamma$ are orthogonal to $\dot{\mathbf{l}}_\lambda$ in $L_2(P_\theta)$.

(d) The likelihood equations are given by

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

These can be rewritten as

$$\mathbb{P}_n Y = \hat{\lambda}_n \mathbb{P}_n(\exp(\hat{\gamma Z})), \qquad \mathbb{P}_n(YZ) = \hat{\lambda}_n = \mathbb{P}_n(Z\exp(\hat{\gamma Z})),$$

and dividing the second by the first to eliminate $\hat{\lambda}$ shows that $\hat{\gamma}$ is the solution of

$$\frac{\mathbb{P}_n(YZ)}{\mathbb{P}_n(Z)} = \frac{\mathbb{P}_n(Z\exp(\hat{\gamma}Z))}{\mathbb{P}_n(\exp(\hat{\gamma}Z))}.$$

The function $h(\gamma) \equiv \mathbb{P}_n(Z\exp(\gamma Z))/\mathbb{P}_n(\exp(\gamma Z))$ is a strictly increasing function of γ if not all the Z_i's are 0 or 1, with $h(-\infty) = 0$, $h(\infty) = 1$, so there is a unique solution of the equation in the last display if not all the Z_i's are equal. 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 4.1.2 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) calculated in (a).

6. (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

$$q(\theta) = \nu(P_\theta) = E_\theta(Y|Z = 1)\,.$$

Compute $q(\theta)$ explicitly as a function of θ.
(c) Suggest a natural empirical estimator of this conditional expectation 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 $\tilde{\mathbf{l}}_\nu$ for estimation of $\nu(P_\theta)$ assuming the Poisson model.
(e) Describe the relationship between ψ and $\tilde{\mathbf{l}}_\nu$ 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 5(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 $\overline{Y}\,\overline{Z}(1-\overline{Z})$. 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 5(b)).
(b) The parameter $q(\theta)$ is easily computed by conditioning on Z:

$$q(\theta) = E_\theta(Y|Z=1) = \lambda e^\gamma.$$

(c) A natural empirical estimator of $\nu(P_\theta) = E_\theta(Y=0|Z=1)$ is

$$\nu(\mathbb{P}_n) = \mathbb{P}_n(Y 1_{[Z=1]})/\mathbb{P}_n 1_{[Z_i=1]} = n^{-1}\sum_{i=1}^n Y_i Z_i / n^{-1} \sum_{i=1}^n Z_i.$$

since $Z_i \in \{0,1\}$. Now

$$\begin{aligned}
\sqrt{n}(\nu(\mathbb{P}_n) - \nu(P)) &= \sqrt{n}\left(\frac{\mathbb{P}_n(YZ)}{\mathbb{P}_n Z} - \frac{P(YZ)}{P(Z)}\right) \\
&= \frac{\sqrt{n}(\mathbb{P}_n(YZ) - P(YZ))}{\mathbb{P}_n(Z)} + \sqrt{n}\left(\frac{1}{\mathbb{P}_n Z} - \frac{1}{PZ}\right) P(YZ) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ (Y_i Z_i) - E(Y|Z=1) - \frac{1}{P(Z)}(Z_i - E(Z))P(YZ) \right\} + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \frac{Z_i}{\eta}(Y_i - E(Y|Z=1)) \right\} + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(Y_i, Z_i) + o_p(1),
\end{aligned}$$

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

$$\psi(y,z) = \frac{z}{\eta}(y - E(Y|Z=1))\,.$$

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

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

where

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

and the information matrix and scores are as given in the solution for problem 5(a).

(e) In general the efficient influence function $\tilde{\mathbf{l}}_\nu$ is the projection of ψ onto $\dot{\mathcal{P}} = [\dot{\mathbf{l}}_\lambda, \dot{\mathbf{l}}_\gamma]$ in $L_2(P_\theta)$ as shown in the following figure.

In this particular case we note that

$$\psi(y,z) = \frac{1}{\eta} z(y - \lambda e^\gamma) = \frac{1}{\eta} \dot{\mathbf{l}}_\gamma(y,z),$$

so ψ is already in $\dot{\mathcal{P}}$ since it is simply η^{-1} times the score for γ. Hence our preliminary estimator is already asymptotically efficient.