Statistics 581, Midterm Exam Solutions

Wellner; 11/13/2002

1. (24 points) **Define** *any three* of the following terms. In each case, provide an appropriate context for your definition.
(a) A normal random vector $Y = (Y_1, \ldots, Y_n)$.
(b) Convergence in rth mean (of a sequence of random variables).
(c) Convergence in probability (of a sequence of random variables).
(d) The inverse or quantile function F^{-1} of a distribution function F.
(e) A uniformly integrable sequence of random variables X_n.

   Solution: See Chapters 1 and 2 of the course notes.

2. (24 points) **State** *any three* of the following results:
(a) The dominated convergence theorem.
(b) The delta-method or g'-theorem for a differentiable function $g : R^k \to R^m$.
(c) The Cramér-Wold device.
(d) Vitale's theorem.
(e) The basic inequality and Markov's inequality.
(f) The elementary Skorokhod theorem.
(g) The Glivenko-Cantelli theorem.

   Solution: See Chapters 1 and 2 of the course notes.

   Do **either** problem 3 **or** problem 4.

3. (30 points).
A. Suppose that $X \sim N_n(\mu, I)$ where $\mu = (\mu_1, \ldots, \mu_n)' \in R^n$ and I is the $n \times n$ identity matrix. Describe the distribution of $Y \equiv X'X = |X|^2$ in terms of ordinary chi-square distributions and a Poisson random variable K, and give the distribution's name.
B. Use the description in A to compute the mean and variance of Y.
C. What is the role of the distribution of Y in a statistical problem we have discussed in class?

   Solution: A. If $X \sim N_n(\mu, I)$, then $Y \equiv |X|^2 \sim \chi^2_n(\delta)$ where $\delta = \mu'\mu$; moreover $(Y|K = k) \sim \chi^2_{2k+n}$ where $K \sim \text{Poisson}(\delta/2)$.
B. By the conditional characterization of the distribution of Y we have

$$E(Y) = E(E(Y|K)) = E(2K + n) = 2(\delta/2) + n = n + \delta\,,$$

and

$$\begin{aligned} Var(Y) &= E\{Var(Y|K)\} + Var[E(Y|K)] = E\{2(2K+n)\} + Var[2K+n] \\ &= 2(2(\delta/2)+n)) + 4\delta/2 = 2n + 4\delta \end{aligned}$$

just as in problem 1 of problem set # 4.
C. In testing $H_0 : p = p_0$ versus $K_0 : p \neq p_0$ the limiting distribution of Q_n under local alternatives of the form $p_n = p_0 + cn^{-1/2}$ is $\chi^2_{k-1}(\delta)$ with $\delta = \sum_1^k c_j^2/p_{j0}$.

4. (30 points).
A. Let $c \in (0, \infty]$. Give examples of sequences of random variables X_n with $X_n \to_p 0$ but $E(X_n) \to c$.
B. Give an example of a sequence of random variables which is not uniformly integrable.
C. Show that if $|X_n| \le Y$ where Y is integrable then $\{X_n\}$ is uniformly integrable.
D. Show that if $\{X_n\}$ is a sequence with $\limsup_{n\to\infty} E|X_n|^r < \infty$ for some $r > 1$, then $\{X_n\}$ is uniformly integrable.

Solution: A. For $c \in (0, \infty)$, and $U \sim \text{Uniform}(0,1)$, let $X_n = cn1_{[0,1/n]}(U)$. Then for $\epsilon > 0$, $P(|X_n| > \epsilon) = n^{-1}1_{[cn>\epsilon]} \to 0$, so $X_n \to_p 0$, but $E(X_n) = c \to c$. For $c = \infty$, take $X_n = n \log n 1_{[0,1/n]}(U)$. Then again $X_n \to_p 0$, but $E(X_n) = \log n \to \infty$.
B. Neither of the sequences $\{X_n\}$ in A is uniformly integrable. If they were uniformly integrable, then by Vitali's theorem we would have $E(X_n) \to E(0) = 0$, and this clearly fails.
C. If $|X_n| \le Y$, then

$$E\{|X_n|1_{[|X_n|\ge\lambda]}\} \le E\{Y1_{[Y\ge\lambda]}\},$$

so

$$\limsup_{n\to\infty} E\{|X_n|1_{[|X_n|\ge\lambda]}\} \le E\{Y1_{[Y\ge\lambda]}\} \to 0 \qquad \text{as } \lambda \to \infty,$$

and hence $\{X_n\}$ is uniformly integrable.
D. If $\limsup_{n\to\infty} E|X_n|^r < \infty$ where $r > 1$, then

$$E\{|X_n|1_{[|X_n|\ge\lambda]}\} \le E\{|X_n|\frac{|X_n|^{r-1}}{\lambda^{r-1}}1_{[|X_n|\ge\lambda]}\} \le \frac{E\{|X_n|^r\}}{\lambda^{r-1}},$$

so

$$\limsup_{n\to\infty} E\{|X_n|1_{[|X_n|\ge\lambda]}\} \le \frac{\limsup_{n\to\infty} E|X_n|^r}{\lambda^{r-1}} \to 0 \qquad \text{as } \lambda \to \infty.$$

5. (30 points)
Suppose that $X, X_1, \ldots, X_n$ are i.i.d. Exponential(θ) random variables so that $P_\theta(X > x) = \exp(-\theta x) = 1 - F_\theta(x)$ for $x > 0$.
A. Fix $x_0 > 0$ and let $\mathbb{F}_n(x) = n^{-1}\sum_{i=1}^n 1_{[X_i \le x]} = n^{-1}\sum_{i=1}^n 1_{(\infty,x]}(X_i)$ denote the empirical distribution function. Show that

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - 1/\theta \\ \mathbb{F}_n(x_0) - F_\theta(x_0) \end{pmatrix} \to_d Y \sim N_2(0, \Sigma)$$

and find Σ.
B. Let $g(\theta) \equiv F_\theta(x_0) = 1 - \exp(-\theta x_0)$, and consider the two estimators of $F = F_\theta$ given by $T_{n,1} \equiv g(\widehat{\theta}_n)$ and $T_{n,2} \equiv \mathbb{F}_n(x_0)$ where $\widehat{\theta}_n \equiv 1/\overline{X}_n$. Show that

$$\sqrt{n}\begin{pmatrix} T_{n,1} - F_\theta(x_0) \\ T_{n,2} - F_\theta(x_0) \end{pmatrix} \to_d \tilde{Y}$$

and find the distribution of $\tilde{Y}$.
C. What is the advantage of $T_{n,2} = \mathbb{F}_n(x_0)$ as an estimator even though it is inefficient when the exponential model holds?

Solution: A. First note that

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - 1/\theta \\ \mathbb{F}_n(x_0) - F_\theta(x_0) \end{pmatrix} = \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i$$

where $Y_i = (X_i - 1/\theta, 1_{[0,x_0]}(X_i) - F_\theta(x_0))'$, are i.i.d. with $E(Y_1) = 0$ and

$$\begin{aligned} E(Y_1 Y_1^T) &= \begin{pmatrix} Var_\theta(X_1) & E(X_1 - 1/\theta)1_{[0,x_0]}(X_1) \\ E(X_1 - 1/\theta)1_{[0,x_0]}(X_1) & F_\theta(x_0)(1 - F_\theta(x_0)) \end{pmatrix} \\ &= \begin{pmatrix} 1/\theta^2 & -x_0 e^{-\theta x_0} \\ -x_0 e^{-\theta x_0} & e^{-\theta x_0}(1 - e^{-\theta x_0}) \end{pmatrix} \equiv \Sigma\,, \end{aligned}$$

since

$$\begin{aligned} E(X_1 1_{[0,x_0]}(X_1)) - (1/\theta)(1 - e^{-\theta x_0}) &= \theta^{-1}\int_0^{\theta x_0} y e^{-y}\,dy - (1/\theta)(1 - e^{-\theta x_0}) \\ &= \theta^{-1} P(\text{Gamma}(2,1) \le \theta x_0) - (1/\theta)(1 - e^{-\theta x_0}) \\ &= \theta^{-1} P(\text{Poisson}(\theta x_0) \ge 2) - (1/\theta)(1 - e^{-\theta x_0}) \\ &= \theta^{-1}\left\{1 - e^{-\theta x_0} - \theta x_0 e^{-\theta x_0}\right\} - (1/\theta)(1 - e^{-\theta x_0}) \\ &= -x_0 e^{-\theta x_0}\,. \end{aligned}$$

It follows from the central limit theorem that

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - 1/\theta \\ \mathbb{F}_n(x_0) - F_\theta(x_0) \end{pmatrix} = \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i \to_d N_2(0, \Sigma)\,.$$

B. Note that

$$\sqrt{n}\begin{pmatrix} T_{n,1}-F_\theta(x_0) \\ T_{n,2}-F_\theta(x_0)\end{pmatrix} = \sqrt{n}\begin{pmatrix} g(\widehat{\theta})-g(\theta) \\ \mathbb{F}_n(x_0)-F_\theta(x_0)\end{pmatrix} = \sqrt{n}\begin{pmatrix} h(\overline{X}_n)-h(1/\theta) \\ \mathbb{F}_n(x_0)-F_\theta(x_0)\end{pmatrix}$$

where $h(y) \equiv 1-\exp(-x_0/y)$. Thus we have a map $\tilde{g} : R^2 \to R^2$ given by $\tilde{g}(y,z) = (h(y), z)$ with gradient (matrix)

$$\nabla\tilde{g}(y,z) = \begin{pmatrix} h'(y) & 0 \\ 0 & 1\end{pmatrix} = \begin{pmatrix} -e^{-x_0/y}(x_0/y^2) & 0 \\ 0 & 1\end{pmatrix},$$

and hence

$$\nabla\tilde{g}(1/\theta, F_\theta(x_0)) = \begin{pmatrix} -\theta^2 x_0 e^{-\theta x_0} & 0 \\ 0 & 1\end{pmatrix},$$

Thus by the g'-theorem it follows that

$$\begin{aligned}\sqrt{n}\begin{pmatrix} T_{n,1}-F_\theta(x_0) \\ T_{n,2}-F_\theta(x_0)\end{pmatrix} = \sqrt{n}\begin{pmatrix} h(\overline{X}_n)-h(1/\theta) \\ \mathbb{F}_n(x_0)-F_\theta(x_0)\end{pmatrix} &\to_d \nabla\tilde{g}(1/\theta, F_\theta(x_0))Y \\ &\sim N_2(0\nabla\tilde{g}\Sigma\nabla\tilde{g}^T) = N_2(0,\tilde{\Sigma})\end{aligned}$$

where

$$\tilde{\Sigma} = \nabla\tilde{g}\Sigma\nabla\tilde{g}^T = \begin{pmatrix} \theta^2 x_0^2 e^{-2\theta x_0} & \theta^2 x_0^2 e^{-2\theta x_0} \\ \theta^2 x_0^2 e^{-2\theta x_0} & e^{-\theta x_0}(1-e^{-\theta x_0})\end{pmatrix}.$$

Thus the asymptotic relative efficiency of $T_{n,2} = \mathbb{F}_n(x_0)$ to $T_{n,1} = g(\widehat{\theta}_n)$ is given by

$$ARE(T_{n,2}, T_{n,1}) = \frac{\theta^2 x_0^2 e^{-2\theta x_0}}{e^{-\theta x_0}(1-e^{-\theta x_0})} = \frac{(\theta x_0)^2 e^{-\theta x_0}}{1-e^{-\theta x_0}}.$$

A plot of this function of θx_0 is given below.

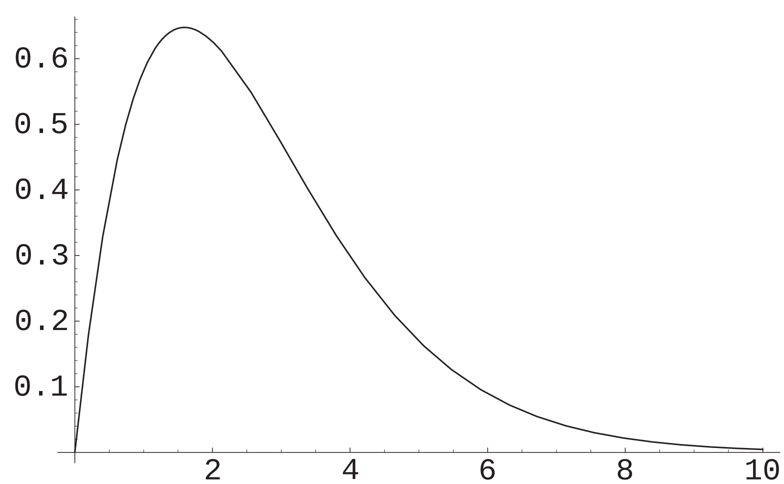

Figure 1: Efficiency of Empirical Relative to MLE under Exponential df

C. An advantage of $T_{n,2} = \mathbb{F}_n(x_0)$ as an estimator, even though it is inefficient when the exponential model holds, is that it is a consistent estimator of $F(x_0)$ even when the exponential model fails.

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

6. (32 points).
A. Suppose that X_{ni}, $i = 1, \ldots, n$ are independent and identically distributed Bernoulli(p_{ni}) random variables. Show that if $\sum_{i=1}^n p_{ni}(1 - p_{ni}) \to \infty$ then

$$\frac{\sum_{i=1}^n (X_{ni} - p_{ni})}{\sqrt{\sum_1^n p_{ni}(1 - p_{ni})}} \to_d N(0, 1) \,.$$

B. Use A to show that if $p_{ni} = p_n$ for all $i = 1, \ldots, n$ where $p_n \to p_0 > 0$, then

$$\sqrt{n}(n^{-1} \sum_1^n X_{ni} - p_n) \to_d N(0, p_0(1 - p_0)) \,.$$

C. Use the result of B to show that if $x \in R$, $t \in (0, 1)$ is fixed, and F is continuous at $F^{-1}(t)$, then

$$\sqrt{n}(\mathbb{F}_n(F^{-1}(t) + xn^{-1/2}) - F(F^{-1}(t) + xn^{-1/2})) \to_d N(0, t(1 - t)) \,.$$

D. Suppose that Δ_{ni}, $i = 1, \ldots, n$ are independent and identically distributed $\text{Mult}_k(1, p_n)$ random vectors where $p_n = (p_{n1}, \ldots, p_{nk}) \to (p_{0,1}, \ldots, p_{0,k})$. Show that

$$\sqrt{n} \left(n^{-1} \sum_1^n \Delta_{ni} - p_n \right) \to_d N_k(0, \text{diag }(p_0) - p_0 p_0^T) \,.$$

Solution: A. (From problem set 5): With $X_{ni} \equiv X_i - p_{n,i}$, $i = 1, \ldots, n, \ldots$ we have $E(X_{ni}) = 0$, $\sigma_{ni}^2 = Var(X_{ni}) = p_{n,i}(1 - p_{n,i})$, and

$$\begin{aligned}
\gamma_{ni} &= E|X_{ni}|^3 = E|X_i - p_{n,i}|^3 = |1 - p_{n,i}|^3 p_{n,i} + |0 - p_{n,i}|^3 (1 - p_{n,i}) \\
&\le p_i(1 - p_{n,i})\{(1 - p_{n,i})^2 + p_{n,i}^2\} \le 2p_{n,i}(1 - p_{n,i}) \,,
\end{aligned}$$

so that $\sigma_n^2 = \sum_{i=1}^n p_{n,i}(1 - p_{n,i})$ and $\gamma_n \le 2 \sum_{i=1}^n p_{n,i}(1 - p_{n,i})$. hence

$$\frac{\gamma_n}{\sigma_n^3} \le \frac{2}{\{\sum_{i=1}^n p_{n,i}(1 - p_{n,i})\}^{1/2}} \to 0$$

if $\sum_1^n p_{n,i}(1-p_{n,i}) \to \infty$. Hence it follows from the Liapunov CLT that

$$\frac{\sum_{i=1}^n (X_i - p_{n,i})}{\sqrt{\sum_1^n p_{n,i}(1-p_{n,i})}} \to_d N(0,1)\,,$$

and this is equivalent to the stated conclusion.
B. In the case $p_{ni} = p_n$ for $i = 1, \ldots, n$ with $p_n \to p_0 \in (0,1)$, it follows that

$$n^{-1}\sum_1^n p_{n,i}(1-p_{n,i}) = p_n(1-p_n) \to p_0(1-p_0) > 0,$$

and hence

$$\sum_1^n p_{n,i}(1-p_{n,i}) = np_n(1-p_n) \to \infty\,.$$

Thus by A we conclude that

$$\frac{\sum_{i=1}^n (X_{ni} - p_n)}{\sqrt{np_n(1-p_n)}} \to_d N(0,1)\,,$$

and since $p_n \to p_0$, this implies that

$$\sqrt{n}(n^{-1}\sum_1^n X_{ni} - p_n) \to_d N(0, p_0, (1-p_0))\,.$$

C. Note that

$$\mathbb{F}_n(F^{-1}(t) + xn^{-1/2}) = n^{-1}\sum_{i=1}^n X_{ni}$$

where $X_{ni} \equiv 1_{(-\infty, F^{-1}(t)+xn^{-1/2}]}(X_i)$, $i = 1, \ldots, n$ are independent and identically distributed Bernoulli(p_n) random variables with $p_n = F(F^{-1}(t) + xn^{-1/2}) \to F(F^{-1}(t)) = t \in (0,1)$. Thus the desired conclusion follows from part B.
D. By the Cramér - Wold device it suffices to show that for any $a \in R^k$

$$a'\sqrt{n}\left(n^{-1}\sum_1^n \Delta_{ni} - p_n\right) \to_d N_1(0, a'\left\{\text{diag }(p_0) - p_0 p_0^T\right\} a)\,.$$

But

$$a'\sqrt{n}\left(n^{-1}\sum_1^n \Delta_{ni} - p_n\right) = \sum_{i=1}^n n^{-1/2} a'(\Delta_{ni} - p_n) \equiv \sum_1^n X_{ni}$$

where

$$E(X_{ni}) = 0, \qquad \sigma_{ni} \equiv Var(X_{ni}) = n^{-1} a'\left\{\text{diag }(p_n) - p_n p_n^T\right\} a\,,$$

and

$$\begin{aligned}\gamma_{ni} &= E|X_{ni}|^3 = n^{-3/2} E\Big|\sum_{j=1}^{k} a_j(\Delta_{n,i,j} - p_{n,j})\Big|^3 \\ &\le n^{-3/2}\left(\sum_{j=1}^{k} |a_j|\right)^3 .\end{aligned}$$

It follows that

$$\sigma_n^2 = \sum_1^n \sigma_{ni}^2 = a'\left\{\text{diag }(p_n) - p_n p_n^T\right\} a \to a'\left\{\text{diag }(p_0) - p_0 p_0^T\right\} a\,, \tag{1}$$

and

$$\gamma_n = \sum_1^n \gamma_{ni} \le n^{-1/2}\left(\sum_{j=1}^{k} |a_j|\right)^3 ,$$

so that

$$\frac{\gamma_n}{\sigma_n^3} \le \frac{n^{-1/2}\left(\sum_{j=1}^k |a_j|\right)^3}{[a'\{\text{diag }(p_n) - p_n p_n^T\}\, a]^{3/2}} \to 0\,.$$

It follows from the Liapunov central limit theorem that

$$\frac{a'\sqrt{n}\,(n^{-1}\sum_1^n \Delta_{ni} - p_n)}{\sqrt{a'\{\text{diag }(p_n) - p_n p_n^T\}\, a}} \to_d N(0,1)\,,$$

and this together with (1) yields the conclucsion.

7. (32 points). Suppose that $N = (N_1, \ldots, N_k) \sim \; Mult_k(n, p)$ and consider testing $H_0 : \underline{p} = \underline{p}_0$ versus $K_0 : \underline{p} \ne \underline{p}_0$. Instead of the chi-square statistic Q_n, consider the test statistic given by

$$H_n^2 \equiv 4n \sum_{i=1}^{k} (\sqrt{\hat{p}_i} - \sqrt{p_{i0}})^2 .$$

The statistic H_n^2 is $8n$ times the square of the *Hellinger distance* between $\underline{\hat{p}}$ and $\underline{p}_0$.
A. Find the limiting distribution of H_n^2 under the null hypothesis H_0.
B. Find the limit of $n^{-1}H_n^2$ under fixed alternatives $\underline{p} \ne \underline{p}_0$ in K_0, and use this to show that the test based on H_n^2 is consistent against K_0.
C. Find the limiting distribution of H_n^2 under local alternatives $\underline{p}_n = \underline{p}_0 + \underline{c}/\sqrt{n}$

(with $\underline{c}'\underline{1} = 0$), and use this to approximate the power of this test. Compare the (local asymptotic) power of this test to the chi-square test.

Solution: A. Let $\underline{Z}_n \equiv \sqrt{n}(\hat{\underline{p}}_n - \underline{p}_0)$ Then $\underline{Z}_n \to_d \underline{Z} \sim N_k(0, \Sigma)$ with $\Sigma = \text{diag}(\underline{p}_0) - \underline{p}_0\underline{p}_0^T$. Thus, by the delta - method,

$$\begin{aligned} \underline{Y}_n &\equiv 2\sqrt{n}(\sqrt{\hat{\underline{p}}_n} - \sqrt{\underline{p}_0}) \\ &\to_d \text{diag}(1/\sqrt{\underline{p}_0})\underline{Z} \equiv \underline{Y} \sim N_k(0, I - \sqrt{\underline{p}_0}\sqrt{\underline{p}_0^T}) \end{aligned}$$

Hence, by the continuous mapping theorem,

$$H_n^2 = \underline{Y}_n^T\underline{Y}_n \to_d \underline{Y}^T\underline{Y}.$$

It remains to answer the question: what is the distribution of $\underline{Y}^T\underline{Y}$? This goes just exactly as in the case of the limit for the chi-square statistic Q_n. Let Γ be an orthogonal matrix with first row $\sqrt{\underline{p}_0}$. Then

$$\Gamma\underline{Y} \sim N_k(0, \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}),$$

which has first coordinate 0, and the remaining $k-1$ coordinates are iid $N(0,1)$. Further, $\Gamma^T\Gamma = I$ and hence

$$\underline{Y}^T\underline{Y} = \underline{Y}^T\Gamma^T\Gamma\underline{Y} = (\Gamma\underline{Y})^T(\Gamma\underline{Y}) \sim \chi^2_{k-1}.$$

Thus $H_n^2 \to_d \underline{Y}^T\underline{Y} \sim \chi^2_{k-1}$.
B. Under fixed $\underline{p} \neq \underline{p}_0$, $\hat{p}_n \to_{a.s.} p$. Hence by the continuous mapping theorem

$$\begin{aligned} n^{-1}H_n^2 &= 4\sum_{j=1}^k \left\{\sqrt{\hat{p}_j} - \sqrt{p_{j0}}\right\}^2 \\ &\to_{a.s.} 4\sum_{j=1}^k (\sqrt{p_j} - \sqrt{p_{j0}})^2 \\ &= 4d_H^2(p, p_0) > 0. \end{aligned}$$

Therefore, under $p \neq p_0$, $H_n^2 \to_{a.s.} \infty$, and hence

$$P_p(H_n^2 \geq \chi^2_{k-1,\alpha}) \to 1.$$

C. Under local alternatives, Liapunov's CLT, the Cramér - Wold device, and the delta method, yield

$$\begin{aligned} \underline{Y}_n &= 2\sqrt{n}(\sqrt{\hat{\underline{p}}_n} - \sqrt{\underline{p}_n}) + 2\sqrt{n}(\sqrt{\underline{p}_n} - \sqrt{\underline{p}_0}) \\ &\to_d \underline{Y} + \text{diag}(1/\sqrt{\underline{p}})\underline{c} \\ &\equiv \underline{Y} + \underline{\mu} \\ &\sim N_k(\underline{\mu}, I - \sqrt{p_0}\sqrt{p_0}^T). \end{aligned}$$

Now with Γ as in part (a)

$$\Gamma(\underline{Y}+\underline{\mu}) = \Gamma\underline{Y} + \Gamma\underline{\mu}) = \Gamma\underline{Y} + \underline{b}$$

where the first coordinate of $\underline{b}$ is 0. Thus $\Gamma\underline{Y}+\underline{b}$ has first coordinate 0, and the remaining $k-1$ coordinates are independent $N(b_i, 1)$. Hence

$$\begin{aligned}(\underline{Y}+\underline{\mu})^T(\underline{Y}+\underline{\mu}) &= (\Gamma\underline{Y}+\underline{b})^T(\Gamma\underline{Y}+\underline{b}) \\ &\sim \chi^2_{k-1}(\underline{b}^T\underline{b}) = \chi^2_{k-1}(\sum_{j=1}^{k} c_i^2/p_{i0})\end{aligned}$$

Thus the local asymptotic power of the test based on the Hellinger statistics H_n^2 is the same as that of the chi-square statistic Q_n.