Statistics 581, Problem Set 3 Solutions, corrected

Wellner; 10/25/2006

1. Ferguson, ACILST, page 34, problem 1(b), modified slightly.
Suppose that $X_1, \ldots, X_n$ is a sample from the Poisson distribution with parameter $\lambda > 0$: $P(X_1 = k) = \exp(-\lambda)\lambda^k/k!$, $k = 0, 1, \ldots$. Let $Z_n = (1/n)\sum_{i=1}^n 1_{[X_i=1]}$.
(a) What is the joint asymptotic distribution of

$$\sqrt{n}((\overline{X}_n, Z_n)' - (\lambda, \lambda e^{-\lambda})')?$$

(b) Let $p_1(\lambda) \equiv P_\lambda(X_1 = 1)$. What is the asymptotic distribution of $\hat{p}_1 \equiv p_1(\hat{\lambda}_n)$ where $\hat{\lambda}_n = \overline{X}_n$?
(c) What is the joint asymptotic distribution of $(Z_n, \hat{p}_1)$ (after centering and rescaling)?
(d) Compute the ratio of the asymptotic variances of the two estimators Z_n and $\hat{p}_1$ of $p_1(\lambda)$. Which estimator would you prefer if the Poisson model (assumption) holds? Which estimator would you prefer if the Poisson model (assumption) fails?

Solution: (a). Let $W_i \equiv (X_i, Y_i) \equiv (X_i, 1_{[X_i=1]})$. Then the W_i's are i.i.d. with mean $E(W_1) = (\lambda, \lambda e^{-\lambda})'$ and covariance matrix

$$\Sigma = \begin{pmatrix} \lambda & \lambda e^{-\lambda} - \lambda^2 e^{-\lambda} \\ \lambda e^{-\lambda} - \lambda^2 e^{-\lambda} & \lambda e^{-\lambda}(1 - \lambda e^{-\lambda}) \end{pmatrix} = \begin{pmatrix} \lambda & \lambda(1-\lambda)e^{-\lambda} \\ \lambda(1-\lambda)e^{-\lambda} & \lambda e^{-\lambda}(1 - \lambda e^{-\lambda}) \end{pmatrix}. \quad (1)$$

Hence the multivariate CLT implies that

$$\sqrt{n}(\overline{W} - E(W_1)) = \sqrt{n}((\overline{X}_n, Z_n)' - (\lambda, \lambda e^{-\lambda})) \to_d T \sim N_2(0, \Sigma) \quad (2)$$

where Σ is given in (1).

(b). Now $\hat{p}_1 = g(\overline{X}_n)$ where $g(v) = ve^{-v}$. Hence $g'(v) = (1-v)e^{-v}$, $g'(\lambda) = (1-\lambda)e^{-\lambda}$, and $\sqrt{n}(\overline{X}_n - \lambda) \to_d N(0, \lambda)$ by the CLT (or the first component of the convergence in distribution in part (a)). Hence it follows from the delta-method that

$$\sqrt{n}(\hat{p}_1 - p_1(\lambda)) = \sqrt{n}(g(\overline{X}_n) - g(\lambda)) \to_d g'(\lambda)N(0, \lambda) = N(0, \lambda(1-\lambda)^2 e^{-2\lambda}).$$

(c). At this point it is a bit easier to study $(\hat{p}_1, Z_n) = g(\overline{X}_n, Z_n)$ where $g(u, v) \equiv (ue^{-u}, v)$. Then in view of (2) and

$$\nabla g(\lambda, e^{-\lambda}) = \begin{pmatrix} (1-\lambda)e^{-\lambda} & 0 \\ 0 & 1 \end{pmatrix},$$

it follows from the delta-method that

$$\sqrt{n}((\hat{p}_1, Z_n)' - \lambda e^{-\lambda}(1, 1)') \to_d \nabla g(\lambda, e^{-\lambda})T \sim N_2(0, \nabla g \Sigma (\nabla g)')$$

where

$$\nabla g \Sigma (\nabla g)' = \begin{pmatrix} \lambda(1-\lambda)^2 e^{-2\lambda} & \lambda(1-\lambda)^2 e^{-2\lambda} \\ \lambda(1-\lambda)^2 e^{-2\lambda} & \lambda e^{-\lambda}(1 - \lambda e^{-\lambda}) \end{pmatrix}.$$

This is a situation in which we have two estimators of $P_\lambda(X_1 = 1) = p_1(\lambda)$, namely the MLE $\hat{p}_1 = p_1(\hat{\lambda})$ and the empirical (or "plug-in" estimator $Z_n = \#\{i \le n : X_i = 0\}/n$. Note that the ratio of the asymptotic variance of $\hat{p}_1$ to the asymptotic variance of Z_n is

$$ARE(\hat{p}_1, Z_n) \equiv \frac{\lambda(1-\lambda)^2 e^{-2\lambda}}{\lambda e^{-\lambda}(1-\lambda e^{-\lambda})} = \frac{(1-\lambda)^2 e^{-\lambda}}{(1-\lambda e^{-\lambda})} < 1$$

for all $\lambda > 0$. See the figure below. If the observations really have the assumed Poisson distribution, then the MLE based on the Poisson assumption is preferable because of its smaller asymptotic variance. If the Poisson assumption fails, then the empirical estimator Z_n might be preferable since it always estimates the probability correctly.

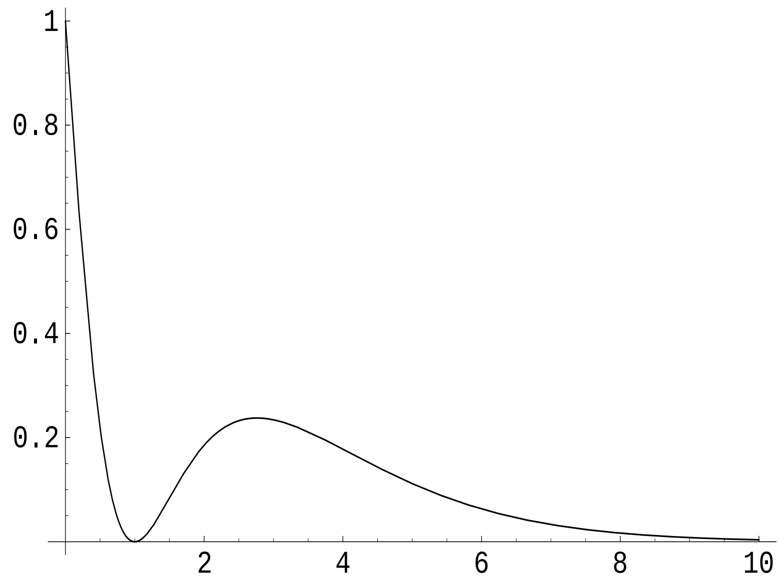

Figure 1: ARE of MLE relative to Plug-In

2. (From Ferguson, *A Course in Large Sample Theory*, page 65, modified.) In a multinomial experiment with sample size $n = 100$ and 3 cells with null hypothesis $H_0 : \underline{p}_0 = (.2, .6, .2)$, what is the approximate power at the alternative $\underline{p} = (.25, .5, .25)$ when the level of significance is $\alpha = .05$? $\alpha = .01$? How large a sample size is need to achieve power 0.8 at this alternative when $\alpha = .05$? $\alpha = .01$?

Solution: Now

$$\begin{aligned} n^{1/2}(\underline{p} - \underline{p}_0) &= 10((1/4, 1/2, 1/4) - (1/5, 3/5, 1/5)) \\ &= 10(1/20, -2/20, -1/20) = (1/2, -1, 1/2), \end{aligned}$$

so the non-centrality parameter is

$$\delta = \frac{(1/2)^2}{1/5} + \frac{(1)^2}{3/5} + \frac{(1/2)^2}{1/5} = 5 \cdot \left\{ \frac{1}{4} + \frac{1}{3} + \frac{1}{4} \right\} = \frac{25}{6}.$$

Thus the approximate power via $\chi_2^2(\delta)$ is

$$P(\chi_2^2(25/6) \ge \chi_{2,.05}) = P(\chi_2^2(25/6) \ge 5.991) = .431\,, \qquad \text{when} \quad \alpha = .05\,,$$

and

$$P(\chi_2^2(25/6) \ge \chi_{2,.01}) = P(\chi_2^2(25/6) \ge 9.210) = .215 \qquad \text{when} \quad \alpha = .01\,,$$

(b) Now we want to find n so that

$$P(\chi_2^2(\delta_n) \ge 5.991) = .80$$

where

$$\delta_n = n\left\{\frac{(1/20)^2}{1/5} + \frac{(2/20)^2}{3/5} + \frac{(1/20)^2}{1/5}\right\} = n/24\,.$$

In this case we find that $\delta_n = n/24 = 9.635$, so that $n = 24 \cdot 9.635 \approx 231$. When $\alpha = .01$ we find that $\delta_n = n/24 = 13.881$ so that $n = 24 \cdot 13.881 \approx 333$.

The alternative approximation to power that we derived in class is

$$\begin{aligned} P_p(Q_n \ge \chi^2_{k-1,\alpha}) &= P_p\left(\sqrt{n}(n^{-1}Q_n - q) \ge \sqrt{n}(n^{-1}\chi^2_{k-1,\alpha} - q)\right) \\ &\doteq P\left(N(0, d^T A d) \ge \sqrt{n}(n^{-1}\chi^2_{k-1,\alpha} - q)\right) \\ &= 1 - \Phi(\sqrt{n}(n^{-1}\chi^2_{k-1,\alpha} - q)/\sqrt{d^T A d}) \end{aligned}$$

where $d \equiv 2\text{diag}(1/p_0)(p - p_0)$, $A = \text{diag}(p) - pp^T$, and $q = \sum_{j=1}^k (p_j - p_{j0})^2/p_{j0}$. In the present case I calculate $q = 1/24$, $d = (1, -6, 1)^T/10$, and

$$A = \text{diag}(p) - pp^T = \frac{1}{16}\begin{pmatrix} 3 & -2 & -1 \\ -2 & 4 & -2 \\ -1 & -2 & 3 \end{pmatrix}$$

so that $d^T A d = 1/8$. Thus the approximation becomes

$$P_p(Q_n \ge \chi^2_{2,\alpha}) \doteq 1 - \Phi(\sqrt{n}(n^{-1}\chi^2_{2,\alpha} - 1/24)/\sqrt{1/8}).$$

When I calculate I get

$$\begin{aligned} P_p(Q_n \ge \chi^2_{2,.05}) &\doteq 1 - \Phi(\sqrt{n}(n^{-1}\chi^2_{2,.05} - 1/24)/\sqrt{1/8}) \doteq 0.303 \\ P_p(Q_n \ge \chi^2_{2,.05}) &\doteq 1 - \Phi(\sqrt{n}(n^{-1}\chi^2_{2,.01} - 1/24)/\sqrt{1/8}) \doteq 0.077, \end{aligned}$$

which are both somewhat lower than suggested by the non-central chi-square approximation. A Monte-Carlo study not shown here shows that the non-central chi-square approximation is quite accurate in this case. I suspect that the fixed alternative limit theorem and resulting normal approximation to power will do better for more extreme alternatives with a larger number of cells, but I have not carried out a thorough study.

3. Suppose that $X_1, X_2, \ldots$ are i.i.d. (μ, σ^2) with $\mu_4 < \infty$. Let $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$ and $S_n^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ be the sample mean and sample variance

respectively.
(a) Show that

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - \mu \\ S_n^2 - \sigma^2 \end{pmatrix} \to_d \underline{Z} \sim N_2(0, \Sigma)$$

where

$$\begin{pmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{pmatrix}.$$

(b) Suppose $\mu \neq 0$. Use (a) to find the limiting distribution of the sample *coefficient of variation* $C_n \equiv S_n/\overline{X}_n$; i.e. show that $\sqrt{n}(C_n - c) \to_d N(0, V^2)$ with $c \equiv \sigma/\mu$ and find V^2.

Solution: (a) Since $S_n^2 = n^{-1}\sum_{i=1}^n (X_i - \mu)^2 + o_p(1/\sqrt{n})$, we have

$$\begin{aligned} \sqrt{n}\begin{pmatrix} \overline{X}_n - \mu \\ S_n^2 - \sigma^2 \end{pmatrix} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{pmatrix} X_i - \mu \\ (X_i - \mu)^2 - \sigma^2 \end{pmatrix} + o_p(1) \\ &\to_d \underline{Z} \sim N_2(0, \Sigma) \end{aligned}$$

by the multivariate CLT where Σ is as given above.
(b) Here we first study $D_n \equiv 1/C_n = \overline{X}_n/S_n$, and then use this to obtain the result for C_n itself.

The function $g(u, v) = u/\sqrt{v}$ is differentiable at points (u, v) with $v \neq 0$, and the derivative is $\nabla g(u, v) = (1/\sqrt{v}, u(-1/2)v^{-3/2})$ so that $\nabla g(\mu, \sigma^2) = (1/\sigma, (-1/2)\mu/\sigma^{-3}) = (1/\sigma)(1, -(1/2)\mu/\sigma^2))$. Hence it follows from the delta method (g' theorem) that

$$\begin{aligned} \sqrt{n}(D_n - d) &= \sqrt{n}(g(\overline{X}_n, S_n^2) - g(\mu, \sigma^2)) \\ &\to_d \nabla g \cdot \underline{Z} \sim N(0, \nabla g^T \Sigma \nabla g) \end{aligned}$$

and it is easy to calculate that

$$\begin{aligned} \nabla g^T \Sigma \nabla g &= \frac{1}{\sigma^4}\left\{\sigma^4 - \mu\mu_3 + \frac{1}{4}d^2(\mu_4 - \sigma^4)\right\} \\ &= 1 - d\gamma_1 + \frac{1}{4}d^2(2 + \gamma_2) \end{aligned}$$

where $\gamma_1 \equiv \mu_3/\sigma^3$ and $\gamma_2 \equiv \mu_4/\sigma^4 - 3$. Note that when the X_i's are normal (so $\gamma_1 = \gamma_2 = 0$), this reduces to $1 + d^2/2$. Thus under normality we have

$$\sqrt{n}(g(D_n) - g(d)) \to_d N(0, 1)$$

if $g(x) \equiv \sqrt{2}\text{arcsinh}(x/\sqrt{2})$.

By taking $h(x) = x^{-1}$ we also get $C_n = h(D_n)$, $c = h(d)$, and hence, for $d \neq 0$ we find that

$$\begin{aligned} \sqrt{n}(C_n - c) &= \sqrt{n}(h(D_n) - h(d)) \to_d h'(d)N(0, \nabla g^T \Sigma \nabla g) \\ &= -d^{-2}N(0, 1 - d\gamma_1 + 2^{-1}d^2(1 + \gamma_2/2)) \\ &= N(0, c^4(1 - c^{-1}\gamma_1 + 2^{-1}c^{-2}(1 + \gamma_2/2) \\ &= N(0, c^4 - c^3\gamma_1 + 2^{-1}c^2(1 + \gamma_2/2)). \end{aligned}$$

When the X_i's are normal,

$$\sqrt{n}(C_n - c) \to N(0, c^4 + 2^{-1}c^2),$$

and the variance stabilizing transformation is given by

$$g(x) = \sqrt{2}\log\left(\frac{x}{1+\sqrt{1+2x^2}}\right).$$

That is, with g as given in the last display, under normality we have

$$\sqrt{n}(g(C_n) - g(c)) \to_d N(0,1).$$

4. Suppose that X is a random variable with finite fourth moment; $E|X|^4 < \infty$. Then $\mu_4 = E(X-\mu)^4$ is the fourth central moment of X. The ratio $\mu_4/\sigma^4 \equiv \kappa$ is the *kurtosis* of X (or of the distribution function F of X), and $\gamma_2 \equiv \mu_4/\sigma^4 - 3$ is called the *excess of kurtosis*; note that for any $N(\mu, \sigma^2)$ random variable, $\gamma_2 = 0$. Investigate the value of γ_2 for various classical distributions (t_r, uniform, bernoulli, Poission(λ), ...). How big can γ_2 be? How small can γ_2 be?

   Solution: Note that $\mu_4^{1/4} = \{E(X-\mu)^4\}^{1/4} \ge \{E(X-\mu)^2\}^{1/2} = \sigma$ by Liapunov's inequality. Thus $\mu_4/\sigma^4 \ge 1$ always, or $\gamma_2 \equiv \mu_4/\sigma^4 - 3 \ge -2$ with equality if $X = \pm 1$ with probability 1/2 each: then $\mu = 0$, $\sigma^2 = 1$, $\mu_4 = 1$, and $\gamma_2 = -2$.
   For $X \sim N(0,1)$, $\gamma_2 = 0$ since $EX^4 = 3$.
   For $X \sim t_r$, $r > 4$, $\gamma_2 = 6/(r-4) \nearrow \infty$ as $r \searrow 4$; $\gamma_2 \searrow 0$ as $r \nearrow \infty$.
   For $X \sim$ Gamma(α, β), $\gamma_2 = 6/\alpha \nearrow \infty$ as $\alpha \searrow 0$.
   For $X \sim$ Poisson(λ), $\gamma_2 = 1/\lambda \nearrow \infty$ as $\lambda \searrow 0$.
   For $X \sim$ Bernoulli(p), $\gamma_2 = (1-p)^2/p + p^2/(1-p) - 3$ which $= -2$ when $p = 1/2$, and $\nearrow \infty$ when $p \to 0, 1$.

5. Ferguson, ACILST, problem 7, page 34.

   Solution: Now $T_n = \sum_{k=1}^n X_k$ where $X_1, X_2, \ldots, X_n$ are independent and $X_k \sim$ Uniform on $\{0, 1, \ldots, k-1\}$ for $k = 1, \ldots, n$. Thus we compute

$$E(X_k) = \sum_{j=0}^{k-1} j\frac{1}{k} = \frac{1}{k}\frac{(k-1)k}{2} = \frac{k-1}{2},$$

$$E(X_k^2) = \sum_{j=0}^{k-1} j^2\frac{1}{k} = \frac{1}{k}\frac{(k-1)k(2k-1)}{6} = \frac{(k-1)(2k-1)}{6},$$

$$Var(X_k) = E(X_k^2) - [E(X_k)]^2 = \frac{(k-1)(2k-1)}{6} - \left(\frac{k-1}{2}\right)^2$$
$$= \frac{(k-1)(k+1)}{12}.$$

   This yields

$$\mu_n \equiv E(T_n) = \sum_{k=1}^n E(X_k) = \sum_{k=1}^n \frac{k-1}{2} = \frac{(n-1)n}{4},$$

$$
\begin{aligned}
\sigma_n^2 \equiv Var(T_n) &= \sum_{k=1}^{n} Var(X_k) = \frac{1}{12}\left\{\sum_{1}^{n} k^2 - n\right\} \\
&= \frac{n(n-1)(2n+5)}{72}.
\end{aligned}
$$

To verify the Lindeberg condition, note that $X_{nj} \equiv X_j - (j-1)/2$ satisfies $|X_{nj}| \le (j-1)/2 \le (n-1)/2$ for $1 \le j \le n$. Therefore, for any fixed $\epsilon > 0$,

$$
\begin{aligned}
\frac{1}{\sigma_n^2}\sum_{j=1}^{n} E\{|X_{nj}|^2 1_{[|X_{nj}|>\epsilon\sigma_n]}\} &\le \frac{1}{\sigma_n^2}\sum_{j=1}^{n} E\{|X_{nj}|^2 1_{[(n-1)/2>\epsilon\sigma_n]}\} \\
&= 1_{[(n-1)/2>\epsilon\sigma_n]} = 0
\end{aligned}
$$

for all $n \ge$ some N_ϵ since $\sigma_n^2 \sim n^3/36$ and hence $\sigma_n \sim n^{3/2}/6$. Thus the Lindeberg condition is satisfied and we conclude that $(T_n - \mu_n)/\sigma_n \to_d N(0,1)$ by the Lindeberg-Feller CLT.

6. Suppose the same set-up as in the chi-square testing situation considered in lecture in class on 10/8/93 and in the above problem, but now, for 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

$$
H_n^2 \equiv 4n \sum_{i=1}^{k} (\sqrt{\hat{p}_j} - \sqrt{p_{0,j}})^2 .
$$

The statistic H_n^2 is $8n$ times the square of the *Hellinger distance* between $\underline{\hat{p}}$ and $\underline{p}$.
(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 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 $\tilde{Z}_n \equiv \sqrt{n}(\hat{p}_n - p_0)$. Then $\tilde{Z}_n \to_d \tilde{Z} \sim N_k(0, \Sigma)$ where $\Sigma = \text{diag}(p_0) - p_0 p_0'$. Thus, by the delta-method (or g' theorem)

$$
\begin{aligned}
Z_n &\equiv 2\sqrt{n}(\sqrt{\hat{p}_n} - \sqrt{p_0}) \\
&\to_d \text{diag}(1/\sqrt{p_0})\tilde{Z} \equiv Z \sim N_k(0, I - \sqrt{p_0}\sqrt{p_0'}).
\end{aligned}
$$

Hence, by the continuous mapping theorem,

$$
H_n^2 = Z_n' Z_n \to_d Z'Z.
$$

It remains to answer the question: what is the distribution of $Z'Z$? This goes just exactly as in the case of the limit for the χ^2- statistic Q_n. Let Γ be an orthogonal matrix with first row $\sqrt{p_0}$. Then

$$
\Gamma Z \sim N_k\left(0, \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}\right)
$$

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

$$Z'Z = Z'\Gamma'\Gamma Z = (\Gamma Z)'(\Gamma Z) \sim \chi^2_{k-1}.$$

Thus $H_n^2 \to_d Z'Z \sim \chi^2_k$.
(b) Under fixed $p \neq p_0$, $\hat{p}_n \to_{a.s.} p$. Hence by continuous mapping

$$\begin{aligned} n^{-1}H_n^2 &= 4\sum_{j=1}^{k}\{\sqrt{\hat{p}_j} - \sqrt{p_{0j}}\}^2 \\ &\to_{a.s.} 4\sum_{j=1}^{k}\{\sqrt{p_j} - \sqrt{p_{0j}}\}^2 \\ &= 8H^2(P, P_0) > 0. \end{aligned}$$

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

$$P_p(H_n^2 \geq \chi_{k-1,\alpha}) \to 1;$$

i.e. the test based on H_n^2 is consistent.
(c) Under local alternatives $p_n = p_0 + cn^{-1/2}$ with $1'c = 0$, using problem 1 and the delta method yields

$$\begin{aligned} Z_n &= 2\sqrt{n}(\sqrt{\hat{p}_n} - \sqrt{p_n}) + 2\sqrt{n}(\sqrt{p_n} - \sqrt{p_0}) \\ &\to_d Z + \mathrm{diag}(1/\sqrt{p_0})c \\ &\equiv Z + \mu \sim N_k(\mu, I - \sqrt{p_0}\sqrt{p_0'}). \end{aligned}$$

Note that the first row of $b \equiv \Gamma\mathrm{diag}(1/\sqrt{p_0})$ is the (row) vector $1'$ of all 1's, and hence the first entry in the vector μ is 0 (since $1'c = 0$). Thus $\Gamma(Z+\mu)$ has first coordinate 0, and the remaining $k-1$ coordinates are independent $N(b_i, 1)$. Hence

$$\begin{aligned} (Z+\mu)'(Z+\mu) &= (\Gamma Z + b)'(\Gamma Z + b) \\ &\sim \chi^2_{k-1}(b'b) = \chi^2_{k-1}\left(\sum_{i=1}^{k}\frac{c_i^2}{p_{0j}}\right) \end{aligned}$$

since $b'b = \mu'\mu = \sum_{j=1}^k c_j^2/p_{0j}^2$. Thus the local asymptotic power of the test based on the Hellinger statistics H_n^2 is the same as that of the χ^2statistic Q_n.