Statistics 581, Problem Set 4 Solutions

Wellner; 10/25/2018

1. Suppose that $\underline{N}_n = (N_{11}, N_{12}, N_{21}, N_{22}) \sim \text{Mult}_4(n, \underline{p})$ where $\underline{p} = (p_{11}, p_{12}, p_{21}, p_{22})$ where $\sum_{i=1}^2 \sum_{j=1}^2 p_{ij} = 1$. (Thus $\underline{N}_n$ is the sum of n independent $\text{Mult}_4(1, \underline{p})$ random vectors $\{\underline{Y}_i\}_{i=1}^n$.) Since there are really just three independently varying parameters for this problem, it is often useful to re-express the cell probabilities in terms of two marginal probabilities, say $p_{1\cdot} = p_{11} + p_{12}$ and $p_{\cdot 1} = p_{11} + p_{21}$, and ψ, the log of the odds-ratio, defined by

$$\psi \equiv \log \frac{p_{21}/p_{22}}{p_{11}/p_{12}} = \log \frac{p_{12}p_{21}}{p_{11}p_{22}} \,. \tag{1}$$

You may use the fact that $\psi = 0$ if and only if independence holds for the 2×2 table (i.e. $p_{ij} = p_{i\cdot}p_{\cdot j}$ for $i, j = 1, 2$).
(a) Suggest an estimator of ψ, say $\widehat{\psi}$.
(b) Show that the estimator you proposed in (a) is asymptotically normal and compute the asymptotic variance of your estimator.

Solution: (a) An obvious estimator of ψ is

$$\widehat{\psi} = \log \frac{\widehat{p}_{12}\widehat{p}_{21}}{\widehat{p}_{11}\widehat{p}_{22}}$$

where $\widehat{\underline{p}} = \underline{N}/n$.
(b) Now $\widehat{\psi} = g(\widehat{\underline{p}})$ where $g(\underline{p})$ is given in (1) and is differentiable with derivative

$$\nabla g(\underline{p}) = (-1/p_{11}, 1/p_{12}, 1/p_{21}, -1/p_{22})$$

and, by the multivariate CLT,

$$\sqrt{n}(\widehat{\underline{p}} - \underline{p}) \to_d Z \sim N_4(0, \Sigma)$$

where $\Sigma = \text{diag}(\underline{p}) - \underline{p}\underline{p}^T$. Thus the delta method (or g'-theorem) yields

$$\begin{aligned} \sqrt{n}(\widehat{\psi} - \psi) &= \sqrt{n}(g(\widehat{\underline{p}}) - g(\underline{p})) \\ &\to_d \nabla g(\underline{p}) Z \sim N(0, \nabla g^T \Sigma \nabla g) = N(0, V^2(\underline{p})) \end{aligned}$$

where

$$V^2(\underline{p}) = \frac{1}{p_{11}} + \frac{1}{p_{12}} + \frac{1}{p_{21}} + \frac{1}{p_{22}} \,.$$

2. This is a continuation of problem 1. One standard test of independence in the 2×2 table is the test based on a Pearson-type chi-square statistic.
(a) Write down the chi-square statistic Q_n for this problem, state its asymptotic distribution under the null hypothesis, and explain briefly why the claimed result holds.
(b) Suppose that the alternative hypothesis holds. Show that under the alternative

hypothesis $n^{-1}Q_n \to_p$ some constant q and compute q as explicitly as possible.
(c) Find the asymptotic distribution of Q_n under local alternatives of the form $\psi_n = tn^{-1/2}$; i.e. $\underline{p}_n \equiv (p_{11,n}, p_{12,n}, p_{21,n}, p_{22,n}) = \underline{p}_0 + \underline{c}n^{-1/2}$ where

$$\psi_0 \equiv \log\left(\frac{p_{21,0}p_{12,0}}{p_{11,0}p_{22,0}}\right) = 0$$

and $\underline{1}'\underline{c} = 0$.
(d) Suppose that $n = 40$, $\alpha = .05$, and the true $\underline{p}$ is $\underline{p} = (.3, .2, .1, .4)$. Give an approximation to the power of the chi-square test at this particular alternative.

Solution: (a) The chi-square statistic for testing independence in a 2×2 table is

$$Q_n = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(N_{ij} - n\hat{p}_{i\cdot}\hat{p}_{\cdot j})^2}{n\hat{p}_{i\cdot}\hat{p}_{\cdot j}}$$

where

$$\begin{aligned}
N_{11} - n\hat{p}_{1\cdot}\hat{p}_{\cdot 1} &= N_{11} - n\frac{N_{11}+N_{12}}{n}\cdot\frac{N_{11}+N_{21}}{n} \\
&= N_{11}\cdot n - (N_{11}+N_{12})(N_{11}+N_{21}) \\
&= N_{11}\cdot(N_{11}+N_{12}+N_{21}+N_{22}) - (N_{11}+N_{12})(N_{11}+N_{21}) \\
&= N_{11}N_{22} - N_{12}N_{21},
\end{aligned}$$

$$\begin{aligned}
N_{12} - n\hat{p}_{1\cdot}\hat{p}_{\cdot 2} &= N_{12}\cdot n - (N_{11}+N_{12})(N_{12}+N_{22}) \\
&= N_{12}\cdot(N_{11}+N_{12}+N_{21}+N_{22}) - (N_{11}+N_{12})(N_{12}+N_{22}) \\
&= N_{12}N_{21} - N_{11}N_{22},
\end{aligned}$$

and similarly for $N_{21} - n\hat{p}_{2\cdot}\hat{p}_{\cdot 1}$ and $N_{22} - n\hat{p}_{2\cdot}\hat{p}_{\cdot 2}$. Therefore

$$\begin{aligned}
Q_n &= \frac{(N_{11}N_{22} - N_{12}N_{21})^2}{n^3}\sum_{i,j}\left\{\frac{1}{\hat{p}_{i\cdot}\hat{p}_{\cdot j}}\right\} \\
&= \frac{(N_{12}N_{21} - N_{11}N_{22})^2}{n^3}\frac{1}{\hat{p}_{1\cdot}(1-\hat{p}_{1\cdot})\hat{p}_{\cdot 1}(1-\hat{p}_{\cdot 1})} \\
&= \frac{n\{\exp(\hat{\psi}_n) - 1\}^2\,(\hat{p}_{11}\hat{p}_{22})^2}{\hat{p}_{1\cdot}(1-\hat{p}_{1\cdot})\hat{p}_{\cdot 1}(1-\hat{p}_{\cdot 1})} \\
&= \frac{\{\sqrt{n}[\exp(\hat{\psi}_n) - 1]\}^2\,(\hat{p}_{11}\hat{p}_{22})^2}{\hat{p}_{1\cdot}(1-\hat{p}_{1\cdot})\hat{p}_{\cdot 1}(1-\hat{p}_{\cdot 1})} \\
&\to_d [N(0,V^2)]^2\frac{[p_{1\cdot}(1-p_{1\cdot})p_{\cdot 1}(1-p_{\cdot 1})]^2}{p_{1\cdot}(1-p_{1\cdot})p_{\cdot 1}(1-p_{\cdot 1})} \\
&= [N(0,V^2)]^2 p_{1\cdot}(1-p_{1\cdot})p_{\cdot 1}(1-p_{\cdot 1}) = [N(0,1)]^2 \stackrel{d}{=} \chi_1^2
\end{aligned}$$

by the delta method or g' theorem and result of problem 1 where we have repeatedly used the fact that $p_{ij} = p_{i\cdot}p_{\cdot j}$ under the null hypothesis of independence.

(b) When the alternative hypothesis holds, then the above argument shows that

$$\begin{aligned} n^{-1}Q_n &= \frac{(N_{12}N_{21} - N_{11}N_{22})^2}{n^4} \frac{1}{\hat{p}_{1\cdot}(1-\hat{p}_{1\cdot})\hat{p}_{\cdot 1}(1-\hat{p}_{\cdot 1})} \\ &= \frac{(\hat{p}_{12}\hat{p}_{21} - \hat{p}_{11}\hat{p}_{22})^2}{\hat{p}_{1\cdot}(1-\hat{p}_{1\cdot})\hat{p}_{\cdot 1}(1-\hat{p}_{\cdot 1})} \\ &\to_p \frac{(p_{12}p_{21} - p_{11}p_{22})^2}{p_{1\cdot}(1-p_{1\cdot})p_{\cdot 1}(1-p_{\cdot 1})} > 0 \end{aligned}$$

where $p_{1\cdot} = p_{11} + p_{12}$ and $p_{\cdot 1} = p_{11} + p_{21}$. It follows that $P_p(Q_n \geq \chi^2_{1,\alpha}) \to 1$ as $n \to \infty$; i.e. the test is consistent.

(c) Under local alternatives with $\psi_n = tn^{-1/2}$ for $t \neq 0$, the argument in (a) repeated (but using the Liapunov CLT) yields

$$\begin{aligned} \sqrt{n}(\widehat{\psi}_n - 0) &= \sqrt{n}(\widehat{\psi}_n - \psi_n) + \sqrt{n}(\psi_n - 0) \\ &= \sqrt{n}(g(\widehat{\underline{p}}) - g(\underline{p}_n)) + t \\ &\to_d \nabla g(\underline{p}_0)Z + t \sim N(t, \nabla g^T \Sigma \nabla g) = N(t, V^2(\underline{p}_0)) \end{aligned}$$

where

$$V^2(\underline{p}_0) = \frac{1}{p_{11,0}} + \frac{1}{p_{12,0}} + \frac{1}{p_{21,0}} + \frac{1}{p_{22,0}} = \frac{1}{p_{1\cdot,0}(1-p_{1\cdot,0})p_{\cdot 1,0}(1-p_{\cdot 1,0})},$$

and hence, by the delta-method again,

$$\sqrt{n}(\exp(\widehat{\psi}_n) - 1) \quad \to_d \quad \nabla g(\underline{p}_0)Z + t \sim N(t, \nabla g^T \Sigma \nabla g) = N(t, V^2(\underline{p}_0)).$$

This implies, via the same development as in (a), that under $\underline{p}_n$ we have

$$\begin{aligned} Q_n &= \frac{n\{\exp(\hat{\psi}_n) - 1\}^2 (\hat{p}_{11}\hat{p}_{22})^2}{\hat{p}_{1\cdot}(1-\hat{p}_{1\cdot})\hat{p}_{\cdot 1}(1-\hat{p}_{\cdot 1})} \\ &= \frac{\{\sqrt{n}[\exp(\hat{\psi}_n) - 1]\}^2 (\hat{p}_{11}\hat{p}_{22})^2}{\hat{p}_{1\cdot}(1-\hat{p}_{1\cdot})\hat{p}_{\cdot 1}(1-\hat{p}_{\cdot 1})} \\ &\to_d [N(t, V^2(\underline{p}_0)]^2 p_{1\cdot,0}(1-p_{1\cdot,0})p_{\cdot 1,0}(1-p_{\cdot 1,0}) \\ &= [N(t\sqrt{c}, 1)]^2 \stackrel{d}{=} \chi^2_1(\delta) \end{aligned}$$

where $\delta = ct^2$ and $c \equiv p_{1\cdot,0}(1-p_{1\cdot,0})p_{\cdot 1,0}(1-p_{\cdot 1,0})$.

(d) When $n = 40$, $\alpha = .05$, and the true $\underline{p}$ is $\underline{p} = (.3, .2, .1, .4)$, we calculate $p_{1\cdot} = 1 - p_{\cdot 1} = .5$, $p_{\cdot 1} = .4$ (so that $c = p_{1\cdot}(1-p_{1\cdot})p_{\cdot 1}(1-p_{\cdot 1}) = (.5)^2(.4)(.6) = .06$),

$$t_n \equiv \sqrt{n}\log\frac{p_{12}p_{21}}{p_{11}p_{22}} = -11.33... .$$

Thus $\delta = (11.33\ldots)^2(.06) = 7.705....$, and an approximation to the power of the chi-square test is given by

$$P(\chi^2_1(7.705...) > \chi^2_{1,.05}) = P(\chi^2_1(7.705...) > 3.841...) = .793... .$$

3. Suppose that $X_1, X_2, \ldots$ are i.i.d. positive random variables, and define $\overline{X}_n \equiv n^{-1}\sum_{i=1}^n X_i$, $H_n \equiv 1/(n^{-1}\sum_{i=1}^n (1/X_i))$, and $G_n \equiv \{\prod_{i=1}^n X_i\}^{1/n}$ to be the *arithmetic, harmonic, and geometric* means respectively. We know that $\overline{X}_n \rightarrow_{a.s.} E(X_1) = \mu$ if and only if $E|X_1| < \infty$.
(a) Use the SLLN together with appropriate additional hypotheses to show that $H_n \rightarrow_{a.s.} 1/\{E(1/X_1)\} \equiv h$, and $G_n \rightarrow_{a.s.} \exp(E\{\log X_1\}) \equiv g$.
(c) Use the multivariate CLT and the delta method to find the joint limiting distribution of $\sqrt{n}(\overline{X}_n - \mu, H_n - h, G_n - g)$. You will need to impose or assume additional moment conditions to be able to prove this. Specify these additional assumptions carefully.
(d) Suppose that $X_i \sim \text{Gamma}(r, \lambda)$ with $r > 0$. For what values of r are the hypotheses you imposed in (c) satisfied? Compute the covariance of the limiting distribution in (c) as explicitly as you can in this case.
(e) Use the result in (c) to show that $\sqrt{n}(G_n/\overline{X}_n - g/\mu) \rightarrow_d N(0, V^2)$ and compute V^2 explicitly when $X_i \sim \text{Gamma}(r, \lambda)$ with r satisfying the conditions you found in (d).

Solution: (a) If $0 < E(1/X_1) < \infty$, then

$$\frac{1}{n}\sum_{i=1}^n (1/X_i) \rightarrow_{a.s.} E(1/X_1) > 0.$$

If $E|\log(X_1)| < \infty$, then

$$\log G_n = \frac{1}{n}\sum_{i=1}^n \log(X_i) \rightarrow_{a.s.} E\log X_1.$$

Thus by the continuous mapping theorem if both $E(1/X_1) < \infty$ and $E|\log X_1| < \infty$, it follows that

$$(H_n, G_n) \rightarrow_{a.s.} (1/E(1/X_1), \exp(E\log X_1)) \equiv (h, g).$$

(c) By the multivariate CLT, if $EX_1^2 < \infty$, $E(1/X_1)^2 < \infty$, and $E(\log X_1)^2 < \infty$, then

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - \mu \\ \overline{X^{-1}}_n - E(1/X_1) \\ \overline{\log X}_n - E\log X_1 \end{pmatrix} \rightarrow_d \underline{Z} \sim N_3(0, \Sigma)$$

where

$$\Sigma = \begin{pmatrix} Var(X_1) & Cov(X_1, 1/X_1) & Cov(X_1, \log(X_1)) \\ Cov(X_1, 1/X_1) & Var(1/X_1) & Cov(1/X_1, \log X_1) \\ Cov(X_1, \log(X_1)) & Cov(1/X_1, \log X_1) & Var(\log(X_1)) \end{pmatrix}.$$

Hence by the delta method with $g(x, y, z) = (x, 1/y, \exp(z))$ so that $\nabla g(x, y, z) = \text{diag}(1, -y^{-2}, \exp(z))$ and $\nabla g(\mu, E(1/X_1), E(\log X_1)) = \text{diag}(1, -h^2, g)$, it follows that

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - \mu \\ H_n - h \\ G_n - g \end{pmatrix} \rightarrow_d \nabla g \cdot \underline{Z} \sim N_3(0, \nabla g \Sigma \nabla g^T) \equiv N_3(0, \tilde{\Sigma}).$$

(d) When $X \equiv X_1 \sim \text{Gamma}(r, \lambda)$, then $Y \equiv \lambda X \sim \text{Gamma}(r, 1)$, and it is easily seen that

$$
\begin{aligned}
&E(X^2) < \infty \qquad \text{if } \ r > 0 \\
&E((\lambda X)^{-2}) = \int_0^\infty y^{-2} \frac{y^{r-1}}{\Gamma(r)} e^{-y} dy = \frac{1}{\Gamma(r)} \int_0^\infty y^{r-3} e^{-y} dy < \infty
\end{aligned}
$$

if $r > 2$, and

$$
E((\log \lambda X)^2) = \int_0^\infty (\log y)^2 y^{r-1} e^{-y} dy / \Gamma(r) < \infty
$$

for all $r > 0$. Thus the covariance matrix Σ exists for $r > 2$; we now calculate it explicitly in this case. First,

$$
\begin{aligned}
E(X) &= \frac{r}{\lambda} \\
E(1/X) &= E(\lambda/(\lambda X) = \lambda E(1/Y) \\
&= \lambda \int_0^\infty y^{-1} y^{r-1} e^{-y} dy / \Gamma(r) \\
&= \lambda \int_0^\infty y^{r-2} e^{-y} dy / \Gamma(r) \\
&= \lambda \Gamma(r-1)/\Gamma(r) = \lambda/(r-1)\,, \\
E(\log X) &= E(\log(\lambda X/\lambda))) = E(\log Y) - \log \lambda \\
&= \int_0^\infty (\log y) y^{r-1} e^{-y} dy / \Gamma(r) - \log \lambda \\
&= \frac{\Gamma'(r)}{\Gamma(r)} - \log \lambda \equiv \psi(r) - \log \lambda\,.
\end{aligned}
$$

Next we calculate Σ:

$$
\begin{aligned}
Var(X) &= Var(\lambda X/\lambda) = Var(Y)/\lambda^2 = \frac{r}{\lambda}\,, \\
Var(1/X) &= \lambda^2 Var(1/Y) = \lambda^2 \{E(1/Y^2) - [E(1/Y)]^2\} \\
&= \lambda^2 \{\frac{1}{(r-1)(r-2)} - \frac{1}{(r-1)^2}\} \\
&= \lambda^2 \frac{1}{(r-1)^2(r-2)} \\
Var(\log X) &= Var(\log(\lambda X/\lambda)) = Var(\log(Y)) = \frac{\Gamma''(r)}{\Gamma(r)} - \left(\frac{\Gamma'(r)}{\Gamma(r)}\right)^2 \equiv C_r \\
Cov(X, 1/X) &= Cov(Y, 1/Y) = 1 - E(Y)E(1/Y) = 1 - \frac{r}{r-1} = \frac{-1}{r-1} \\
Cov(X, \log X) &= \lambda^{-1} Cov(Y, \log Y) = \lambda^{-1}\{E(Y \log Y) - E(Y)E(\log Y)\} \\
&= \lambda^{-1}\{r\psi(r+1) - r\psi(r)\} = \frac{1}{\lambda}\,, \\
Cov(1/X, \log X) &= \lambda Cov(1/Y, \log Y) = \lambda\{E((1/Y)\log Y) - E(1/Y)E(\log Y)\} \\
&= \lambda\{\frac{\psi(r-1)}{r-1} - \frac{\psi(r)}{r-1}\} \\
&= \frac{\lambda}{r-1}\{\psi(r-1) - \psi(r)\} \equiv \frac{-\lambda}{(r-1)^2}\,,
\end{aligned}
$$

where we have used

$$\begin{aligned} A_r &\equiv \psi(r+1)-\psi(r) = (\log\Gamma(r+1)-\log\Gamma(r))' = (\log r)' = 1/r\,, \\ B_r &\equiv \psi(r-1)-\psi(r) = (\log\Gamma(r-1)-\log\Gamma(r))' = (-\log(r-1))' = \frac{-1}{r-1}\,. \end{aligned}$$

Hence

$$\Sigma = \begin{pmatrix} r/\lambda^2 & -1/(r-1) & 1/\lambda \\ -1/(r-1) & \lambda^2/(r-1)^2(r-2) & -\lambda/(r-1)^2 \\ 1/\lambda & -\lambda/(r-1)^2 & C_r \end{pmatrix}.$$

For the gradient ∇g we get $diag(1, -h^2, g) = diag(1, -(r-1)^2/\lambda^2, e^{\psi(r)}/\lambda)$, so it follows by direct calculation that

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

(e) Let $g_2(x,y,z) \equiv z/x$. Then $\nabla g_2(x,y,z) = (-z/x^2, 0, 1/x) = (-z/x, 0, 1)/x$. when we evaluate this at $(\mu, h, g) = (r/\lambda, (r-1)/\lambda, e^{\psi(r)}/\lambda)$, we find that

$$\nabla g_2 = (\lambda/r)(-r^{-1}e^{\psi(r)}, 0, 1)\,.$$

Hence by the delta - method again we find that

$$\sqrt{n}\Big(\frac{G_n}{\overline{X}_n} - g/\mu\Big) \to_d N(0, \nabla g_2 \tilde{\Sigma}(\nabla g_2)') = N(0, e^{2\psi(r)}\{C_r - (1/r)\})\,.$$

Note that the scale factor λ washes out completely here, as we expect.

4. Suppose that $Y_i = \alpha + \theta'(x_i - \overline{x}) + \epsilon_i$, $i = 1, \ldots, n$, where $\epsilon_i \sim (0, \sigma^2)$ are i.i.d. and the x_i's are known vectors in R^k. Equivalently, $\underline{Y} = X\underline{\beta} + \underline{\epsilon}$ where

$$X^T = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 - \overline{x} & x_2 - \overline{x} & \cdots & x_n - \overline{x} \end{pmatrix}$$

so that X is an $n \times (k+1)$ matrix. Let $\hat{\underline{\beta}}$ be the least squares estimator of $\underline{\beta} = (\alpha, \theta')'$; i.e. $\hat{\underline{\beta}} = (X^TX)^{-1}X^T\underline{Y}$. Suppose that $n^{-1}(X^TX) \to D$ where D is positive definite.

(a) What additional condition(s) do you need to impose to prove that

$$\sqrt{n}(\hat{\beta}_n - \beta) \to_d N_{k+1}(0, \text{“}something\text{”})\,?$$

(b) Find “something” in part (a).

Solution: (a) Let $\underline{a} \in R^{k+1}$. Now

$$\begin{aligned} \hat{\beta} &= (X^TX)^{-1}X^TY \\ &= (X^TX)^{-1}X^T(X\beta + \epsilon) \\ &= \beta + (X^TX)^{-1}X^T\epsilon\,, \end{aligned}$$

so

$$\sqrt{n}(\hat{\beta}-\beta) = \sqrt{n}(X^TX)^{-1}X^T\epsilon \equiv B_n\epsilon$$

where $B_n \equiv \sqrt{n}(X^TX)^{-1}X^T$ is a $(k+1)\times n$ matrix. Thus

$$\begin{aligned} a^T(\sqrt{n}(\hat{\beta}-\beta)) &= a^TB_n\epsilon \equiv b_n^T\epsilon \\ &= \sum_{i=1}^n b_{ni}\epsilon_i \equiv \sum_{i=1}^n X_{ni} \end{aligned}$$

where $b_n^T \equiv a^TB_n$ is an $1\times n$ vector. Now we compute

$$\mu_{ni} = E(X_{ni}) = b_{ni}\cdot 0, \qquad \sigma_{ni}^2 = Var(X_{ni}) = b_{ni}^2\sigma^2\,,$$

and hence, using the hypothesized convergence of $n^{-1}X^TX \to D$ in the last line,

$$\begin{aligned} \sigma_n^2 &= \sigma^2\sum_{i=1}^n b_{ni}^2 = \sigma^2 b_n^Tb_n \\ &= \sigma^2a^TB_nB_n^Ta = n\sigma^2a^T(X^TX)^{-1}(X^TX)(X^TX)^{-1}a \\ &= \sigma^2a^T(n^{-1}X^TX)^{-1}a \to \sigma^2a^TD^{-1}a \equiv V^2(a) > 0 \end{aligned}$$

since D is nonsingular. To establish asymptotic normality of $a^T(\sqrt{n}(\hat{\beta}-\beta))/\sigma_n$, it remains to verify the Lindeberg condition: namely

$$\frac{1}{\sigma_n^2}\sum_{i=1}^n E\left\{|X_{ni}|^2 1_{[|X_{ni}|>\delta\sigma_n]}\right\} \to 0 \tag{2}$$

for every $\delta > 0$. But, as we have seen before, this holds if

$$\max_{1\le i\le n}|b_{ni}| \to 0 \qquad \text{as} \qquad n\to\infty\ : \tag{3}$$

the left side of (2) is bounded as follows:

$$\begin{aligned} &\frac{1}{\sigma_n^2}\sum_{i=1}^n b_{ni}^2 E\left\{\epsilon_1^2 1_{[|\epsilon_1|>\delta\sigma_n/|b_{ni}|]}\right\} \\ &\le \frac{1}{\sigma^2}E\left\{\epsilon_1^2 1_{[|\epsilon_1|>\delta\sigma_n/\max_{1\le i\le n}|b_{ni}|]}\right\} \\ &\to \frac{1}{\sigma^2}\cdot 0 = 0 \end{aligned}$$

by (3), $E(\epsilon_1^2) < \infty$, and the dominated convergence theorem. Thus it follows from the Lindeberg-Feller CLT that

$$a^T(\sqrt{n}(\hat{\beta}-\beta))/\sigma_n \to_d N(0,1),$$

and since $\sigma_n^2 \to \sigma^2a^TD^{-1}a$, this implies that

$$a^T(\sqrt{n}(\hat{\beta}-\beta)) \to_d N(0, a^T(\sigma^2D^{-1})a)\,,$$

which in turn, via the Cramér-Wold device, implies

$$\sqrt{n}(\hat{\beta}-\beta) \to_d N_{k+1}(0,\sigma^2D^{-1})\,.$$

5. Suppose the same set-up as in the chi-square testing situation considered in lecture in class but now, for testing $H_0 : \underline{p} = \underline{p}_0$ versus $K_0 : \underline{p} \neq \underline{p}_0$, 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 $4n$ 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 $\underline{p} \neq \underline{p}_0$ 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 $\underline{p}_n = \underline{p}_0 + \underline{c}/\sqrt{n}$, 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 $\underline{Z}_n \equiv \sqrt{n}(\underline{\hat{p}}_n - \underline{p}_0)$ 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,

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

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

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

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

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

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

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

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.

6. **Optional bonus problem 1:** Ferguson, ACILST, problem 5, page 50. Ferguson, ACILST, problem 5, page 50: (The Poisson dispersion test). A standard test of the hypothesis H_0 that a distribution is Poisson(λ) for some λ is to reject H_0 if the ratio of the sample variance to the sample mean, $S_n^2/\overline{X}_n$, is too large. This test is good against alternatives whose variance is greater than the mean, such as the negative binomial distribution or any other mixture of Poisson distributions.
(a) Find the asymptotic distribution of $S_n^2/\overline{X}_n$ for general distributions.
(b) Find the asymptotic distribution of $S_n^2/\overline{X}_n$ under H_0 and show that it is independent of λ.

7. **Optional bonus problem 2:** Ferguson, ACILST, problem 4, page 55. Suppose that $(X_i - \mu)/\sigma$, $i = 1, \ldots, m$ and $(Y_j - \nu)/\tau$, $j = 1, \ldots, n$ are i.i.d. $(0, 1, \mu_4) < \infty$; thus γ_2 is the same for the two populations. Let S_X^2 and S_Y^2 denote the sample variances of the X's and Y's respectively. The classical $F-$ test based on the assumption that all the standardized X's and Y's are $N(0,1)$ rejects $H_0: \ \tau \leq \sigma$ in favor of $H_1: \ \tau > \sigma$ if $F \equiv S_Y^2/S_X^2 > F_{n-1,m-1,\alpha}$. Assuming that $m/N \to \lambda \in [0,1]$ as $m \wedge n \to \infty$ where $N \equiv m + n$, find the true asymptotic size of this test for non-normal X's and Y's as above.