Statistics 581, Midterm Exam Solutions

Wellner; 11/06/2017

1. (24 points) **Define** any **three** of the following six terms.
   (a) The *total variation distance* between two probability measures P and Q.
   (b) The *Hellinger distance* between two probability measures P and Q.
   (c) A normal random vector $Y = (Y_1, \ldots, Y_n)$.
   (d) A *uniformly integrable* sequence of random variables $\{X_n\}$.
   (f) A *standard Brownian bridge process* on $[0, 1]$.
   (g) The *inverse or quantile* F^{-1} of a distribution function F.

   Solution: See course notes.

2. (30 points). **State** any **three** of the following:
   (a) The Lindeberg-Feller CLT.
   (b) The inverse transformation theorem.
   (c) The multivariate delta method or g' theorem.
   (d) Skorokhod's theorem (for real-valued random variables).
   (e) The (classical) Glivenko-Cantelli theorem for the empirical d.f. $\mathbb{F}_n$.
   (f) Scheffé's theorem.

   Solution: See course notes.

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

3. (32 points) Use the ordinary (univariate) central limit theorem and the Cramér-Wold device to prove the multivariate central limit theorem.

   Solution: See course notes.

4. (32 points) Suppose that $X_1, \ldots, X_n$ are i.i.d. exponential(θ); i.e. with density $p_\theta(x) = \theta \exp(-\theta x) 1_{[0,\infty)}(x)$. Let $X_{(n)} = X_{n:n}$ be the largest order statistic of $X_1, \ldots, X_n$.
   (a) Find constants c_n so that $Y_n = X_{(n)} - c_n \rightarrow_d Y$ for some random variable Y and find the limiting distribution of F_Y.
   (b) Compute the density of Y_n and show that it converges to the density f_Y of Y.
   (c) What can you conclude from the result of (b) and Scheffé's theorem?
   (d) Why does the result in (c) imply that $Y_n \rightarrow_d Y$? Does the Hellinger distance $H(P_{Y_n}, P_Y)$ converge to zero?

   Solution:

(a) Let $c_n = \theta^{-1} \log n = F_\theta^{-1}(1 - 1/n)$. Then

$$\begin{aligned} F_n(y) = P(Y_n \le y) &= P(X_{(n)} - c_n \le y) = P(X_{(n)} \le y + c_n) \\ &= P(X_j \le y + c_n \text{ for all } \ 1 \le j \le n) \\ &= P(X_1 \le y + c_n)^n = (1 - \exp(-\theta(y + c_n)))^n \\ &= \left(1 - \frac{e^{-\theta y}}{n}\right)^n \to \exp(-e^{-\theta y}) \\ &\equiv F_Y(y); \end{aligned}$$

this is an extreme - value distribution of the the "double-exponential" or "Gumbel" type; see part (c) of Theorem 14, Ferguson, ACILST page 95.
(b) The density of Y_n is found easily by differentiating in the previous display. The result is that

$$\begin{aligned} f_n(y) &= \theta(1 - n^{-1} \exp(-\theta y))^{n-1} \exp(-\theta y) \\ &\to \theta \exp(-e^{-\theta y}) \exp(-\theta y) = f_Y(y) \equiv f(y). \end{aligned}$$

(c) Since the densities f_n converge pointwise to the limiting density, we conclude by Scheffé's theorem that with P_n begin the probability measure on $\mathbb{R}$ corresponding to F_n and P the corresponding probability measure on $\mathbb{R}$ corresponding to F_Y,

$$d_{TV}(P_n, P) = \frac{1}{2} \int |f_n(y) - f(y)| dy \to 0.$$

(d) Since $F_n(y) = P_n((-\infty, y])$ and $F(y) = P((-\infty, y])$,

$$|F_n(y) - F(y)| \le \sup_{A \in \mathcal{B}} |P_n(A) - P(A)| = d_{TV}(P_n, P) \to 0.$$

Since $H^2(P_n, P) \le d_{TV}(P_n, P) \to 0$, the Hellinger distance $H(P_n, P) \to 0$.

Do **either** problem 5 **or** problem 6.

5. (40 points) Suppose that $X, X_1, X_2, \ldots, X_n$ are independent Geometric(p) random variables: $P(X = k) = p(1-p)^{k-1}$, $k = 1, 2, 3, \ldots$. Recall that $E(X) = 1/p$ and $Var(X) = (1-p)/p^2 \equiv q/p^2$.
(a) Use the weak law of large numbers to show that the random vector

$$\underline{\overline{Y}}_n \equiv \frac{1}{n}\sum_{i=1}^{n}(X_i, 1_{[X_i=1]}, 1_{[X_i=2]})'$$

converges in probability to some vector $(a,b,c)' \equiv \underline{y}$ where (a,b,c) depends on p. Give (a,b,c) explicitly in terms of p.
(b) Use the multivariate CLT to show that

$$\sqrt{n}(\underline{\overline{Y}}_n - \underline{y}) \to_d \underline{W} \sim N_3(0, \Sigma)$$

for some covariance matrix Σ; compute Σ explicitly in terms of p.
(c) The usual estimator of p is $\widehat{p}_n = 1/\overline{X}_n$. A friend suggests the following alternative estimator of p:

$$\tilde{p}_n = 1 - \frac{\sum_{i=1}^n 1_{[X_i=2]}}{\sum_{i=1}^n 1_{[X_i=1]}} = 1 - \frac{\overline{Y}_{3,n}}{\overline{Y}_{2,n}} .$$

Is $\tilde{p}_n$ a consistent estimator of p?
(d) If the answer to (c) is yes, find the asymptotic variance of $\tilde{p}$ as an estimator of p.

Solution: (a) By the weak law of large numbers applied to each coordinate of the vector $\underline{\overline{Y}}_n \equiv \frac{1}{n}\sum_{i=1}^n (X_i, 1_{[X_i=1]}, 1_{[X_i=2]})^T$ it follows that

$$\begin{aligned}
\overline{Y}_{1,n} &= \overline{X}_n \to_p E(X_1) = 1/p,\\
\overline{Y}_{2,n} &= n^{-1}\sum_{i=1}^n 1_{[X_i=1]} \to E1_{[X_1=1]} = P(X_1 = 1) = p,\\
\overline{Y}_{3,n} &= n^{-1}\sum_{i=1}^n 1_{[X_i=2]} \to E1_{[X_1=2]} = P(X_1 = 2) = p(1-p).
\end{aligned}$$

Since a random vector converges in probability if and only each coordinate converges in probability it follows that $\underline{\overline{Y}}_n \to_p (1/p,\ p,\ p(1-p)) \equiv (a,b,c) \equiv \underline{y}$.
(b) The vectors $\underline{Y}_i - \underline{y} = (X_i, 1_{[X_i=1]}, 1_{[X_i=2]})$, $i = 1, \ldots, n$ are i.i.d. with $E(\underline{Y}_i - \underline{y}) = 0$ and covariance matrix

$$\begin{aligned}
\Sigma &= E(\underline{Y}_1 - \underline{y})(\underline{Y}_1 - \underline{y})^T = E(\underline{Y}_1\underline{Y}_1)^T - \underline{y}\underline{y}^T\\
&= \begin{pmatrix} (1-p)/p^2 & p-1, & (1-p)(2p-1)\\ p-1 & p(1-p) & -p^2(1-p)\\ (1-p)(2p-1) & -p^2(1-p) & p(1-p)(1-p(1-p)) \end{pmatrix}\\
&= (1-p)\begin{pmatrix} 1/p^2 & -1, & 2p-1\\ -1 & p & -p^2\\ 1 & -p^2 & p(1-p(1-p)) \end{pmatrix}.
\end{aligned}$$

Thus by the multivariate CLT it follows that

$$\sqrt{n}(\underline{\overline{Y}_n} - \underline{y}) \to_d \underline{W} \sim N_3(0, \Sigma).$$

with Σ as in the previous display.
(c) By the Mann - Wald (or continuous mapping) theorem

$$\widetilde{p}_n = 1 - \frac{\overline{Y}_{3,n}}{\overline{Y}_{2,n}} \to_p 1 - \frac{y_3}{y_2} = 1 - \frac{p(1-p)}{p} = p;$$

thus $\widetilde{p}_n$ is a consistent estimator of p.
(d) Now

$$\sqrt{n}(\widetilde{p}_n - p) = \sqrt{n}\left(g(\underline{\overline{Y}_n}) - g(\underline{y})\right)$$

where $g(a, b, c) = 1 - c/b$ is differentiable with derivative

$$\nabla g(a, b, c) = (0, c/b^2, -1/b)^T = b^{-1}(0, c/b, -1)^T,$$

so that

$$\nabla g(1/p, p, p(1-p)) = p^{-1}(0, 1-p, -1)^T.$$

It then follows from the delta method that

$$\begin{aligned} \sqrt{n}(\widetilde{p}_n - p) &\to_d \nabla g(1/p, p, p(1-p))\underline{W} = p^{-1}\left((1-p)W_2 - W_3\right) \\ &\sim N(0, V^2(p)) \end{aligned}$$

where

$$\begin{aligned} V^2(p) &= \frac{(1-p)^2}{p^2} p(1-p) + \frac{1}{p^2} p(1-p)\,(1 - p(1-p)) + 2\frac{1-p}{p^2}\,(p(1-p)(1-p(1-p))) \\ &= \frac{(1-p)(2-p)}{p}. \end{aligned}$$

In comparison, the asymptotic variance of $\widehat{p}_n \equiv 1/\overline{X}_n$ is $p^2(1-p)$: by the delta method with $g(v) = v^{-1}$,

$$\sqrt{n}(g(\overline{X}_n) - g(1/p)) \to_d -(1/p)^{-2} N(0, p^{-2}(1-p)) = N(0, p^2(1-p)).$$

The ratio of the asymptotic variance of $\widehat{p}_n$ to the asymptotic variance of $\widetilde{p}_n$ is

$$\text{Rel-Effiency}(p) = \frac{p^2(1-p)}{(1-p)(2-p)/p} = \frac{p^3}{2-p}$$

which increases from 0 to 1 as p goes from 0 to 1: the relative efficiency of $\widetilde{p}_n$ is very low for small values of p. Since it never exceeds 1 we would probably not recommend use of $\widetilde{p}_n$, at least if we believe that the geometric(p) model holds.

6. (40 points) Suppose that $\underline{N} = (N_1, \ldots, N_k) \sim \text{Mult}_k(n, \underline{p})$ where $\underline{p} = (p_1, \ldots, p_k)$. In class and homework problems we have discussed the chi-square statistic Q_n for testing $H : \underline{p} = \underline{p_0}$ versus $K : \underline{p} \neq \underline{p}_0$. An alternative statistic for testing H versus K is the Hellinger statistic $4nH_n^2$where

$$4nH_n^2 = 4n \sum_{j=1}^{k} \left(\sqrt{\widehat{p}_j} - \sqrt{p_{0j}} \right)^2$$

(a) Recall the following result from class: under $H : \underline{p} = \underline{p}_0$

$$\begin{pmatrix} \sqrt{n}(\widehat{p}_1 - p_{01}) \\ \vdots \\ \sqrt{n}(\widehat{p}_k - p_{0k}) \end{pmatrix} \to_d \underline{\widetilde{Z}} \sim N_k(0, \text{diag}(\underline{p}_0) - \underline{p}_0\underline{p}_0^T)$$

Use the delta-method to find the limiting distribution of

$$\begin{pmatrix} \sqrt{n}(\sqrt{\widehat{p}_1} - \sqrt{p_{01}}) \\ \vdots \\ \sqrt{n}(\sqrt{\widehat{p}_k} - \sqrt{p_{0k}}) \end{pmatrix};$$

express the result in terms of $\underline{\widetilde{Z}}$ and give the resulting distribution.

(b) Show that if $H : \underline{p} = \underline{p}_0$ is true, then

$$4nH_n^2 \to_d Q \sim \chi_{k-1}^2,$$

and hence if $P(\chi_{k-1}^2 \geq \chi_{k-1,\alpha}^2) = \alpha$, it follows that $P_{\underline{p}_0}(4nH_n^2 \geq \chi_{k-1,\alpha}^2) \to \alpha$.

(c) If the alternative hypothesis K is true, so $\underline{p} \neq \underline{p}_0$, show that

$$n^{-1} 4nH_n^2 = 4H_n^2 \to_p g(\underline{p})$$

and identify $g(\underline{p})$ as a function of $\underline{p}$ and $\underline{p_0}$.

(d) If the alternative hypothesis K is true, so $\underline{p} \neq \underline{p}_0$, show that

$$\sqrt{n} \left(4H_n^2 - g(\underline{p}) \right) \to_d N(0, V^2(\underline{p}))$$

and compute $V^2(\underline{p})$. Could you use this to approximate the power of the Hellinger test of H versus K at a particular alternative $\underline{p} \neq \underline{p}_0$? How?

Solution: (a) Let $g(\underline{v}) = (\sqrt{v_1}, \ldots, \sqrt{v_k})$. Then $\nabla g(\underline{v}) = \text{diag}(2^{-1} v_j^{-1/2})$, and $\nabla g(\underline{p}_0) = \text{diag}(2^{-1} p_j^{-1/2})$. Thus

$$\begin{aligned} \underline{Z}_n &\equiv 2\sqrt{n} \begin{pmatrix} \sqrt{\widehat{p}_1} - \sqrt{p_{0,1}} \\ \vdots \\ \sqrt{\widehat{p}_k} - \sqrt{p_{0,k}} \end{pmatrix} \\ &\to_d 2\nabla g(\underline{p_0})\underline{\widetilde{Z}}, \quad \text{where } \underline{\widetilde{Z}} \sim N_k(0, \text{diag}(\underline{p}_0) - \underline{p}_0\underline{p}_0^T) \\ &\equiv \underline{Z} \sim N_k(0, I - \sqrt{\underline{p}_0}\sqrt{\underline{p}_0}^T). \end{aligned}$$

(b) When $\underline{p} = \underline{p}_0$ is true, by (a) and the Mann-Wald theorem

$$4nH_n^2 = \underline{Z}_n^T \underline{Z}_n \to_d \underline{Z}^T \underline{Z} = \underline{Z}^T \Gamma^T \Gamma \underline{Z}$$

for any orthogonal matrix Γ. Choosing Γ to have first row $= \sqrt{p_0}^T$, we find that a

$$\underline{Y} \equiv \Gamma \underline{Z} \sim N_k \left(0, \begin{pmatrix} 0 & \underline{0}^T \\ \underline{0} & I_{k-1} \end{pmatrix} \right),$$

so that $Y_1 = 0$ a.s. and $\underline{Y}^T \underline{Y} = Y_2^2 + \cdots + Y_k^2 \sim \chi_{k-1}^2$. Thus $P_{p_0}(4nH_n^2 \ge \chi_{k-1,\alpha}^2) \to P(\chi_k^2 \ge \chi_{k-1,\alpha}^2) = \alpha$ as $n \to \infty$.
(c) When $\underline{p} \ne \underline{p}_0$ is true, $\widehat{\underline{p}}_n \to_p \ne \underline{p}_0$, and by the Mann-Wald theorem,

$$4H_n^2 = 4 \sum_{j=1}^k (\sqrt{\widehat{p}_j} - \sqrt{p_{0,j}})^2 \to_p 4 \sum_{j=1}^k (\sqrt{p_j} - \sqrt{p_{0,j}})^2 \equiv g(\underline{p}).$$

Note that $g(\underline{p}) = 8H^2(\underline{p}, \underline{p}_0)$ where H denotes the Hellinger distance.
(d) If $\underline{p} \ne \underline{p}_0$ is true, then, using $a^2 - b^2 = (a-b)(a+b)$

$$\begin{aligned}
&\sqrt{n}\left(4H_n^2 - g(\underline{p})\right) \\
&= 4\sqrt{n} \sum_{j=1}^k \left\{ \left(\sqrt{\widehat{p}_j} - \sqrt{p_{0,j}} \right)^2 - \left(\sqrt{p_j} - \sqrt{p_{0,j}} \right)^2 \right\} \\
&= 4\sqrt{n} \sum_{j=1}^k \left\{ \left(\sqrt{\widehat{p}_j} - \sqrt{p_{0,j}} \right) - \left(\sqrt{p_j} - \sqrt{p_{0,j}} \right) \right\} \cdot \left\{ \left(\sqrt{\widehat{p}_j} - \sqrt{p_{0,j}} \right) + \left(\sqrt{p_j} - \sqrt{p_{0,j}} \right) \right\} \\
&= \sum_{j=1}^k \left\{ 2\sqrt{n} \left(\sqrt{\widehat{p}_j} - \sqrt{p_j} \right) \right\} \cdot \left\{ 4\left(\sqrt{p_j} - \sqrt{p_{0,j}} \right) + 2\left(\left(\sqrt{\widehat{p}_j} - \sqrt{p_{0,j}} \right) - \left(\sqrt{p_j} - \sqrt{p_{0,j}} \right) \right) \right\} \\
&\equiv \underline{d}^T \underline{Z}_n^{alt} + o_p(1)
\end{aligned}$$

where $\underline{d} = (d_1, \ldots, d_k)^T$ with $d_j \equiv 4(\sqrt{p_j} - \sqrt{p_{0,j}})$ and where, much as in (a)

(1) $$\underline{Z}_n^{alt} \equiv 2\sqrt{n}\left(\sqrt{\widehat{p}_j} - \sqrt{p_j} \right) \to_d \underline{Z}^{alt} \sim N_k(0, I - \sqrt{\underline{p}}\sqrt{\underline{p}}^T).$$

Thus it follows from the identity in the first display above and (1) together with the Mann-Wald theorem and Slutsky's theorem that

$$\sqrt{n}(4H_n^2 - g(p)) \to_d \underline{d}^T \underline{Z}^{alt} \sim N_1(0, \underline{d}^T (I - \sqrt{\underline{p}}\sqrt{\underline{p}}^T) \underline{d}).$$