Statistics 581, Midterm Exam Solutions

Wellner; 11/08/2010

1. (24 points) **Define** any **three** of the following six terms.
   (a) The *Hellinger distance* between two probability measures P and Q on a measurable space $(\mathcal{X}, \mathcal{A})$.
   (b) A *non-central chi-square distribution* with m degrees of freedom and non-centrality parameter δ.
   (c) *Convergence in distribution* of a sequence of random variables.
   (d) A sequence of random variables that is *bounded in probability*.
   (e) A *uniformly integrable* sequence of random variables $\{X_n\}$.
   (f) The uniform empirical distribution $\mathbb{G}_n$ and the uniform empirical process $\mathbb{U}_n$.

   Solution: See chapters 1-2, 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 Glivenko-Cantelli theorem.
   (f) Scheffé's theorem.

   Solution: See chapters 1-2, notes.

3. (30 points) Use the ordinary (univariate) central limit theorem and the Cramér-Wold device to prove the multivariate central limit theorem.

   Solution: Let $X_1, \ldots, X_n$ be i.i.d. random vectors in $\mathbb{R}^d$ with $E(X_1^T X_1) < \infty$ so that $\Sigma \equiv E(X-\mu)(X-\mu)^T$ with $\mu \equiv E(X)$ is well-defined. Without loss of generality suppose that $\mu = 0$. We want to show that $\sqrt{n}\overline{X}_n \to_d Z \sim N_d(0, \Sigma)$. To do this let $a \in \mathbb{R}^d$, and consider

$$a^T \sqrt{n}\overline{X}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n a^T X_i \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$$

   where $Y_i = a^T X_i$ are i.i.d. with $EY_1 = 0$ and $E(Y_1)^2 = E(a^T X_1 X_1^T a) = a^T \Sigma a$. Thus it follows from the univariate (Lindeberg) central limit theorem that

$$a^T \sqrt{n}\overline{X}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i \to_d N_1(0, a^T \Sigma a) \stackrel{d}{=} a^T Z.$$

   Thus it follows from the Cramér-Wold device that $\sqrt{n}\overline{X}_n \to_d Z \sim N_d(0, \Sigma)$.

4. (30 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 $Y_n = X_{(n)} - c_n \to_d Y$ 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?

Solution: (a) Note that for $F_\theta(x) = 1 - \exp(-\theta x)$ we have $F_\theta^{-1}(t) = \theta^{-1} \log(1/(1-t))$, and hence that $F_\theta^{-1}(1 - 1/n) = \theta^{-1} \log n$. Thus with $c_n \equiv \theta^{-1} \log n$, for every $y \in \mathbb{R}$,

$$\begin{aligned} P_\theta(Y_n \le y) &= P(X_{(n)} \le y + \theta^{-1} \log n) = P(X_1 \le y + \theta^{-1} \log n)^n \\ &= \left(1 - \exp(-\theta(y + \theta^{-1} \log n))\right)^n \\ &= (1 - \exp(-\theta y)/n)^n \\ &\to \exp(-\exp(-\theta y)) \equiv F_Y(y). \end{aligned}$$

(b) Note that

$$\begin{aligned} f_{Y_n}(y) &= n(1 - \exp(-\theta y)/n)^{n-1} \cdot \exp(-\theta y) \cdot \theta/n \\ &= (1 - \exp(-\theta y)/n)^{n-1} \cdot \exp(-\theta y) \cdot \theta \\ &\to \exp(-\exp(-\theta y)) \cdot \exp(-\theta y) \cdot \theta = f_Y(y). \end{aligned}$$

(c) It follows from (b) and Scheffé's theorem that with $P_n(A) \equiv \int_A f_{Y_n}(y) dy$ and $P(A) \equiv \int_A f_Y(y) dy$,

$$d_{TV}(P_n, P) = \frac{1}{2} \int_{\mathbb{R}} |f_{Y_n}(y) - f_Y(y)| dy \to 0.$$

5. (30 points). Suppose that $X_1, \ldots, X_n$ are i.i.d. with distribution function F having a continuous density function f. Let $\mathbb{F}_n$ be the empirical distribution function of the X_i's, and suppose that b_n is a sequence of positive numbers, and let

$$\hat{f}_n(x) = \frac{\mathbb{F}_n(x + b_n) - \mathbb{F}_n(x - b_n)}{2b_n}.$$

(a) Show that $E\hat{f}_n(x) \to f(x)$ if $b_n \to 0$.
(b) Show that $Var(\hat{f}_n(x)) \to 0$ if $b_n \to 0$ and $nb_n \to \infty$.
(c) Use some appropriate central limit theorem to show that (perhaps under some suitable further conditions that you might need to specify)

$$\sqrt{2nb_n}(\hat{f}_n(x) - E\hat{f}_n(x)) \to_d N(0, f(x)).$$

Hint: Write $\hat{f}_n(x)$ in terms of some Bernoulli random variables and identify $p = p_n$.

Solution: [This $\hat{f}_n$ is a *kernel density estimator* based on the uniform kernel $k(x) = 1_{[-1,1]}(x)/2$, and can be rewritten as

$$\hat{f}_n(x) = \int_{-\infty}^{\infty} \frac{1}{b_n} k((x-y)/b_n) d\mathbb{F}_n(y);$$

other kernel density estimators result when the uniform kernel is replaced by some other density function.]

(a) First note that $2nb_n\widehat{f}_n(x) = n(\mathbb{F}_n(x+b_n) - \mathbb{F}_n(x-b_n))$ is a Binomial(n, p_n) random variable with $p_n = F(x+b_n) - F(x-b_n)$. Hence if $b_n \to 0$

$$\begin{aligned}
E\widehat{f}_n(x) &= \frac{F(x+b_n) - F(x-b_n)}{2b_n} = \frac{p_n}{2b_n} \\
&= \frac{1}{2}\left\{\frac{F(x+b_n) - F(x)}{b_n} + \frac{F(x) - F(x-b_n)}{b_n}\right\} \\
&\to \frac{1}{2}\{f(x) + f(x)\} = f(x).
\end{aligned}$$

(b) Furthermore

$$\begin{aligned}
Var(\widehat{f}_n(x)) &= \frac{np_n(1-p_n)}{(2nb_n)^2} \\
&= \frac{1}{2nb_n}\frac{p_n}{2b_n}(1-p_n) \\
&\to 0 \cdot f(x) \cdot 1 = 0
\end{aligned}$$

if $nb_n \to \infty$ and $b_n \to 0$.

(c) Since $2nb_n\widehat{f}_n(x) = \sum_{i=1}^n X_{ni}$ where $X_{ni} \sim$ Bernoulli(p_n), it follows that $\sigma_{ni}^2 = p_n(1-p_n)$ so that $\sigma_n^2 = \text{Var}(\sum_{i=1}^n X_{ni}) = np_n(1-p_n)$, and

$$\begin{aligned}
\gamma_n \equiv \sum_{i=1}^n \gamma_{ni} &= \sum_{i=1}^n E|X_{ni} - \mu_{ni}|^3 \\
&= np_n(1-p_n)\{(1-p_n)^2 + p_n^2\} \\
&\le 2np_n(1-p_n)
\end{aligned}$$

so that

$$\gamma_n/\sigma^3 \le \frac{2}{\sqrt{np_n(1-p_n)}} = \frac{2}{\sqrt{nb_n(p_n/b_n)(1-p_n)}} \to 0$$

if $b_n \to 0$ and $nb_n \to \infty$. Thus, by the Liapunov CLT,

$$\frac{2nb_n(\widehat{f}_n(x) - E\widehat{f}(x))}{\sqrt{np_n(1-p_n)}} \to N(0,1)$$

if $b_n \to 0$ and $nb_n \to \infty$. Thus

$$\begin{aligned}\sqrt{2nb_n}(\widehat{f}_n(x) - E\widehat{f}_n(x)) &= \frac{2nb_n(\widehat{f}_n(x) - E\widehat{f}_n(x))}{\sqrt{np_n(1-p_n)}}\sqrt{\frac{np_n(1-p_n)}{2nb_n}} \\ &\to N(0,1)\sqrt{f(x)} = N(0, f(x)).\end{aligned}$$

6. (30 points). Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. Uniform$(0,\theta)$. Consider $\hat{\theta}_n = \max_{1\le i\le n} X_i = X_{(n)}$ as an estimator of θ.
   (a) Compute $E(\hat{\theta}_n)$.
   (b) Show that $Y_n = n(\theta - \hat{\theta}_n) \to_d Y$ and find the distribution of Y.
   (c) Show that $\hat{\theta}_n \to_p \theta$.
   (d) Consider the function $g(x) = (1-x)^{-2}$. Does $g(Y_n) \to_d$ something? If the answer is yes, what is the limit (expressed in terms of the random variable Y)?
   (e) Consider the function $g(x) = \log x$. Find the limiting distribution of $n(g(\hat{\theta}_n) - g(\theta))$.

Solution: (a) Note that $X_i \stackrel{d}{=} \theta U_i$ for $i = 1, \ldots, n$ where $U_1, \ldots, U_n$ are i.i.d. $U(0,1)$. Since $EU_{(j)} = j/(n+1)$ for $1 \le j \le n$, it follows that

$$E_\theta(\hat{\theta}_n) = E_\theta(X_{(n)}) = \theta E U_{(n)} = \frac{n}{n+1}\theta.$$

(b) Now $Y_n = n(\theta - \hat{\theta}_n) \stackrel{d}{=} \theta n(1 - U_{(n)})$. Therefore

$$\begin{aligned}P_\theta(Y_n \ge y) &= P(\theta n(1-U_{(n)}) \ge y) = P(1 - U_{(n)} \ge y/(n\theta)) = P(U_{(n)} \le 1 - y/(n\theta)) \\ &= \left(1 - \frac{y}{n\theta}\right)^n \to \exp(-y/\theta).\end{aligned}$$

Thus $Y_n \to_d Y \sim$ Exponential$(1/\theta) \stackrel{d}{=} \theta$Exp$(1)$.
(c) This follows immediately from (b):

$$\widehat{\theta}_n - \theta = -n^{-1}n(\theta - \widehat{\theta}_n) \to_d -0 \cdot Y = 0,$$

which implies convergence to zero in probability.
(d) The function $g(x) = 1/(1-x)^2$ is continuous at all $x \ne 1$. Since $P(Y \in \{1\}) = 0$, it follows by the continuous mapping (or Mann-Wald) theorem that $g(Y_n) \to_d g(Y)$.
(e) Since $g(x) = \log x$ is differentiable at θ with $g'(\theta) = 1/\theta$, it follows from the g' theorem that

$$\begin{aligned}n(g(\hat{\theta}_n) - g(\theta)) &= \frac{g(\hat{\theta}_n) - g(\theta)}{\hat{\theta}_n - \theta} n(\hat{\theta}_n - \theta) \\ &\to_d g'(\theta)(-Y) = -Y/\theta \sim -\text{Exp}(1).\end{aligned}$$

7. (30 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 and the Hellinger distance statistic $4nH_n^2$ as test statistics 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 likelihood ratio statistic $2 \log \lambda_n$ where

$$\lambda_n \equiv \frac{\sup_{\underline{p}} L_n(\underline{p})}{L_n(\underline{p}_0)} = \frac{\prod_{j=1}^k \widehat{p}_j^{N_j}}{\prod_{j=1}^k p_{0j}^{N_j}} = \prod_{j=1}^k \left\{ \frac{\widehat{p}_j}{p_{0j}} \right\}^{N_j} .$$

(a) Show that

$$2 \log \lambda_n = 2n \sum_{j=1}^k \widehat{p}_j \log \left(\frac{\widehat{p}_j}{p_{0j}} \right) .$$

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

$$n^{-1} 2 \log \lambda_n = g(\underline{\widehat{p}}) \rightarrow_p g(\underline{p}) ,$$

and identify $g(\underline{p})$ as a function of $\underline{p}$ and $\underline{p_0}$.
(c) If the alternative hypothesis K is true, so $\underline{p} \neq \underline{p}_0$, show that

$$\sqrt{n}(2n^{-1} \log \lambda_n - g(\underline{p})) = \sqrt{n}(g(\underline{\widehat{p}}) - g(\underline{p})) \rightarrow_d N(0, V^2(\underline{p})) ,$$

and compute $V^2(\underline{p})$. Could you use this to approximate the power of the likelihood-ratio test? How?

Solution: (a) This is straightforward algebra:

$$2 \log \lambda_n = 2 \sum_{j=1}^k N_j \log \left(\frac{\widehat{p}_j}{p_{0j}} \right) = 2n \sum_{j=1}^k \widehat{p}_j \log \left(\frac{\widehat{p}_j}{p_{0j}} \right) .$$

(b) Since $\widehat{p}_n \rightarrow_p p$ in $\mathbb{R}^k$ and the function $g(p) \equiv 2 \sum_{j=1}^k p_j \log \left(\frac{p_j}{p_{0j}} \right)$ is continuous at $p = (p_1, \ldots, p_k)$ with all $p_j > 0$, it follows from the Mann-Wald theorem that

$$\begin{aligned} n^{-1} 2 \log \lambda_n &= 2 \sum_{j=1}^k \widehat{p}_j \log \left(\frac{\widehat{p}_j}{p_{0j}} \right) \\ &= g(\widehat{p}_n) \rightarrow_p g(p) \\ &= 2 \sum_{j=1}^k p_j \log \left(\frac{p_j}{p_{0j}} \right) . \end{aligned}$$

In fact this is twice the Kullback-Leibler divergence between P and P_0 (which we will meet in more generality in Chapter 4.)
(c) Now $g(p) \equiv 2 \sum_{j=1}^k p_j \log \left(\frac{p_j}{p_{0j}} \right)$ is differentiable with derivative

$$\nabla g(p) = 2 \left(\underline{1} + (\log(p_1/p_{0,1}), \ldots, \log(p_k/p_{0,k}))^T \right) ,$$

and $\sqrt{n}(\widehat{p}_n - p) \to_d Z \sim N_d(0, \Sigma)$ where $\Sigma = \text{diag}(p) - pp^T$. Thus it follows from the delta-method that

$$\begin{aligned}\sqrt{n}(2n^{-1}\log\lambda_n - g(\underline{p})) &= \sqrt{n}(g(\widehat{p}_n) - g(p)) \\ &\to_d (\nabla g(p))^T Z \sim N(0, (\nabla g(p))^T \Sigma \nabla g(p)) \equiv N(0, V^2(p)).\end{aligned}$$

This can be used to approximate the power of the likelihood ratio test as follows: under the null hypothesis it is known (and easily proved) that $2\log\lambda_n \to_d \chi^2_{k-1}$, and therefore the test "reject $H: p = p_0$ if $2\log\lambda_n \geq \chi^2_{k-1,\alpha}$ has asymptotic size α. Thus the power of the test at $p \neq p_0$ is

$$\begin{aligned}P_p(2\log\lambda_n \geq \chi^2_{k-1,\alpha}) &= P_p\left(\sqrt{n}(2n^{-1}\log\lambda_n - g(p)) \geq \sqrt{n}(n^{-1}\chi^2_{k-1,\alpha} - g(p))\right) \\ &\approx P\left(N(0, V^2(p)) \geq \sqrt{n}(n^{-1}\chi^2_{k-1,\alpha} - g(p))\right) \\ &= 1 - \Phi\left(\frac{\sqrt{n}(n^{-1}\chi^2_{k-1,\alpha} - g(p))}{V(p)}\right).\end{aligned}$$

This is *not the usual approximation of the power* of the likelihood ratio statistic: along the same lines as discussed in class for the chi-square statistic Q_n and in homework for the Hellinger statistic H_n, if $p = p_n = p_0 + cn^{-1/2}$ where $1^T c = 0$, it is known (and not hard to show) that $2\log\lambda_n \to_d \chi^2_{k-1}(\delta)$ where $\delta = \sum_{j=1}^k c_j^2/p_{0,j}$. Thus the usual approximation of the power of the likelihood ratio test is given by

$$P_p(2\log\lambda_n \geq \chi^2_{k-1,\alpha}) \approx P(\chi^2_{k-1}(\delta_n) \geq \chi^2_{k-1,\alpha})$$

where $\delta_n \equiv \sum_{j=1}^k \{\sqrt{n}(p_j - p_{0,j})\}^2/p_{0,j}$.