Statistics 522, Problem Set 8 Solutions

Wellner; 3/11/2020

1. Exercise 11.6.4, page 34, Wellner, Chapter 11 notes.
Give a direct proof of the equivalence of (i) and (iv) in Proposition 2.2. (Hint: consider the functions $\psi_\epsilon(x) \equiv \psi(x/\epsilon)$ with ψ as given on page 34.

Solution: Suppose that (iv) holds. Let $x \in C_F$ and $\epsilon > 0$. Suppose that ψ is defined by

$$\psi(y) \equiv \begin{cases} \frac{\int_y^1 \exp(-1/(u(1-u)))du}{\int_0^1 \exp(-1/(u(1-u)))du}, & 0 \le y \le 1, \\ 1, & y \le 0, \\ 0, & y \ge 1. \end{cases}$$

Consider the function $f_u(y) \equiv f_u(y; x, \epsilon)$ defined by

$$f_u(y) \equiv \psi_{\epsilon,x}(y) = \begin{cases} 0, & y \ge x + \epsilon, \\ 1, & y \le x, \\ \psi((y-x)/\epsilon), & x \le y \le x + \epsilon. \end{cases}$$

Then $f_u \in C^\infty(\mathbb{R})$ and it satisfies

$$1_{(-\infty,x]}(y) \le f_u(y) \le 1_{(-\infty,x+\epsilon]}(y),$$

and hence it follows that

$$F_n(x) = E1_{(-\infty,x]}(X_n) \le Ef_u(X_n) \le E1_{(-\infty,x+\epsilon]}(X_n) = F_n(x+\epsilon), \quad \text{and}$$
$$F(x) = E1_{(-\infty,x]}(X) \le Ef_u(X) \le E1_{(-\infty,x+\epsilon]}(X) = F(x+\epsilon).$$

Therefore, since (iv) holds,

$$\limsup_n F_n(x) \le \limsup_n Ef_u(X_n) = Ef_u(X) \le E1_{(-\infty,x+\epsilon]}(X) = F(x+\epsilon).$$

Letting $\epsilon \searrow 0$ and using right-continuity of F, this yields

$$\limsup_n F_n(x) \le F(x). \tag{1}$$

Similarly the function $f_l(y) \equiv f_l(y; x, \epsilon) \equiv \psi((y - (x - \epsilon))/\epsilon)$ satisfies $f_l \in C^\infty(\mathbb{R})$ and

$$1_{(-\infty, x-\epsilon]}(y) \le f_l(y) \le 1_{(-\infty, x]}(y),$$

and hence, since (iv) holds,

$$F(x - \epsilon) \le E f_l(X) = \lim_n E f_l(X_n) = \liminf_n E f_l(X_n) \le \liminf_n F_n(x).$$

Letting $\epsilon \searrow 0$ and using $x \in C_F$ (here is where $x \in C_F$ is used!), this yields

$$F(x) \le \liminf_n F_n(x). \tag{2}$$

Combining (1) and (2) yields $F_n \to_d F$; i.e. (i) holds.

For the reverse implication, note that (i) implies (ii) by the Helly-Bray Theorem 3.5.1; (i.e. $Ef(X_n) \to Ef(X)$ for all $f \in C_b(\mathbb{R})$), and since $C^\infty(\mathbb{R}) \subset C_b(\mathbb{R})$ (iv) holds. (Recall that the simplest proof of the Helly-Bray theorem is via the elementary Skorokhod theorem 6.3.2, PfS page 110.

2. Exercise 9.2.4, PfS Course Notes, page 199. (Exercise 11.8.4, page 293, PfS 2000.)
   Suppose that $\log X \sim N(0,1)$.
   (i) Show that the density of X is given by $f_X(x) = x^{-1} \exp(-(\log x)^2/2)/\sqrt{2\pi}$ for $x > 0$ (and 0 otherwise).
   (ii) For each $a \in [-1, 1]$ consider the random variable Y_a with density

$$f_a(y) = f_X(y)(1 + a \sin(2\pi \log y)) \quad \text{for} \quad y > 0.$$

Show that $EX^k = EY_a^k$ for all integers $k \ge 1$ and $a \in [-1, 1]$.

Solution: (i) Since $X \stackrel{d}{=} e^Z$ where $Z \sim N(0,1)$, it is clear that $X > 0$ with probability 1. Hence we compute, for $x > 0$,

$$F_X(x) = P(X \le x) = P(e^Z \le x) = P(Z \le \log x) = \Phi(\log x)$$

where $\Phi(z) \equiv \int_{-\infty}^z \phi(y) dy$ is the standard normal distribution function and $\phi(z) \equiv (2\pi)^{-1/2} \exp(-z^2/2)$ is the standard normal density function. Thus

$$f_X(x) = F_X'(x) = \phi(\log x) \cdot x^{-1} \quad \text{for} \quad x > 0.$$

(ii) Now by changing variables to $y = \log x$ we get

$$\begin{aligned} EX^k &= \int_0^\infty x^k x^{-1} (2\pi)^{-1/2} \exp(-(\log x)^2/2) dx \\ &= \int_{-\infty}^\infty e^{ky}\phi(y)dy \quad \text{where} \quad \phi(z) = (2\pi)^{-1/2} e^{-z^2/2} \\ &= e^{k^2/2} \int_{-\infty}^\infty \phi(z-k)dz = e^{k^2/2}. \end{aligned}$$

On the other hand, by the same change of variables,

$$\begin{aligned} EY_a^k &= \int_0^\infty x^k x^{-1} (2\pi)^{-1/2} \exp(-(\log x)^2/2) \left(1 + a\sin(2\pi \log x)\right) dx \\ &= \int_{-\infty}^\infty e^{ky}\phi(y) \left(1 + a\sin(2\pi y)\right) dy \\ &= e^{k^2/2} \int_{-\infty}^\infty \phi(y-k) \left(1 + a\sin(2\pi y)\right) dy \\ &= e^{k^2/2} \int_{-\infty}^\infty \phi(z) \left(1 + a\sin(2\pi (z+k))\right) dz \\ &= e^{k^2/2} \int_{-\infty}^\infty \phi(z) \left(1 + a\sin(2\pi z)\right) dz \\ &\qquad \text{by using} \\ &\qquad \sin(2\pi(z+k)) = \sin(2\pi z)\cos(2\pi k) + \cos(2\pi z)\sin(2\pi k) \\ &\qquad\qquad = \sin(2\pi z)\cdot 1 + \cos(2\pi z)\cdot 0 \\ &\qquad\qquad = \sin(2\pi z) \\ &= e^{k^2} \left(1 + a\int_{-\infty}^\infty \sin(2\pi z)\phi(z)dz\right) \\ &= e^{k^2/2} \quad \text{since } \sin \text{ is odd and } \phi \text{ is even} \\ &= e^{k^2/2}. \end{aligned}$$

Thus the distribution function F in (i) is not uniquely determined by its moments. The moments $\{\mu_k\}_{k\ge 1}$ *do determine* a unique distribution function if either $\limsup_{k\to\infty} |\mu_k|^{1/k}/k < \infty$ or $\sum_{k=1}^\infty \mu_{2k} t^{2k}/(2k)! < \infty$ for some interval of t's.

3. Exercise 11.6.7, page 34, Wellner, Chapter 11 notes. (This asks for a version of the Lindeberg replacement inequality in the case of random vectors X, Y, and W with values in $\mathbb{R}^k$.)

Solution: First the needed multivariate Taylor's theorem with remainder. (Thanks to Mathias Hudoba de Badyn for pointing out a nice source for this, namely G. Folland's course notes for Math 425: https://www.math.washington.edu/ folland/Math425/taylor2.pdf .) Suppose that $f : \mathbb{R}^k \to \mathbb{R}$ with $f \in C^3(\mathbb{R}^k)$. Then for $x, y \in \mathbb{R}^k$

$$f(x+y) = f(x) + y^T \nabla f(x) + \frac{1}{2} y^T H(x) y + R_{x,2}(y)$$

where

$$\begin{aligned}
\nabla f(x) &= \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_1} \right)^T \equiv \text{the gradient of } f \text{ at } x, \\
H(x) \equiv \nabla^2 f(x) &= \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{i,j=1}^k \equiv \text{the Hessian of } f \text{ at } x, \\
R_{x,2}(y) &\equiv \sum_{|\alpha|=3} \partial^\alpha f(x + \theta y) \frac{y^\alpha}{\alpha!} \quad \text{for some } 0 \le \theta \le 1,
\end{aligned}$$

and where, for $\alpha = (\alpha_1, \dots, \alpha_k) \in \mathbb{N}^k$,

$$\begin{aligned}
&|\alpha| \equiv \sum_{j=1}^k \alpha_j, \qquad \alpha! = \prod_{i=1}^k \alpha_j!, \\
&x^\alpha \equiv x_1^{\alpha_1} \cdots x_k^{\alpha_k} \quad \text{for} \quad x \in \mathbb{R}^k, \\
&\partial^\alpha f(x) = \partial_1^{\alpha_1} \partial_2^{\alpha_2} \cdots \partial_k^{\alpha_k} f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_k^{\alpha_k}}.
\end{aligned}$$

Thus if $|\partial^3 f(x)| \le M$ for all $x \in \mathbb{R}^k$ (and all α with $|\alpha| = 3$), it follows from the multinomial formula

$$(x_1 + x_2 + \cdots + x_k)^m = \sum_{|\alpha|=m} \frac{m!}{\alpha!} x^\alpha$$

with $m = 3$, that

$$|R_{x,2}(y)| \le \frac{M}{3!} \|y\|^3$$

where $\|y\|_1 = \sum_{j=1}^k |y_j|$ is the ℓ_1 norm of y.

Now back to the problem at hand. From the Taylor expansion we find, using the independence of X and Y and then the hypothesized equality of first and second moments, that

$$\begin{aligned} Ef(X+Y) &= Ef(X) + E\left[Y^T \nabla f(X)\right] + \frac{1}{2} E\left[Y^T H(X) Y\right] + E[R_{X,2}(Y)] \\ &= Ef(X) + E\left[Y^T\right] E\nabla f(X) + \frac{1}{2} E\left[YY^T\right] \cdot EH(X)\big] + E[R_{X,2}(Y)] \\ &= Ef(X) + E\left[W^T\right] E\nabla f(X) + \frac{1}{2} E\left[WW^T\right] \cdot EH(X)\big] + E[R_{X,2}(Y)] \\ &= Ef(X+W) - E[R_{X,2}(W)] + E[R_{X,2}(Y)]. \end{aligned}$$

The second equality above is based on a computation via the trace operation as follows: with $H \equiv H(X)$ and using the fact that equal means and covariance (matrices) for Y and W holds if and only if Y and W have equal means and equal second moment matrices (so $E(YY^T) = E(WW^T)$),

$$\begin{aligned} & E\left[Y^T H(X) Y\right] \\ &= E\left[tr(Y^T H Y)\right] = E\left[tr(YY^T \cdot H)\right] \\ &= tr\left(E\left[YY^T \cdot H\right]\right) \\ &= tr\left(E\left[YY^T\right] \cdot E(H)\right) \quad \text{by independence of } Y \text{ and } X \\ &= E\left[W^T H(X) W\right] \quad \text{by the same argument as for } Y \text{above.} \end{aligned}$$

This yields

$$\begin{aligned} |Ef(X+Y) - Ef(X+W)| &\le E|R_{X,2}(Y)| + E|R_{X,2}(W)| \\ &\le \frac{M}{6} \left\{ E\|Y\|^3 + E\|W\|^3 \right\} \end{aligned}$$

as claimed. Note that this can be formulated in terms of the usual Euclidean norm $|v|_2 \equiv \{\sum_{j=1}^k v_j^2\}^{1/2}$ since the ℓ_p norms on $\mathbb{R}^k$ are related by $|v|_p \le |v|_{p'} \le k^{1/p' - 1/p} |v|_p$ if $1 \le p' \le p$. In particular, $|v|_1 \le k^{1/2} |v|_2$.

4. Exercise 11.6.9, page 35, Wellner, Chapter 11 notes. Use the Cramér - Wold device to prove the multivariate CLT from the classical (one-dimensional) CLT.

Solution: Suppose that $\underline{X}_1, \ldots, \underline{X}_n$ are i.i.d with $E(\underline{X}) = \mu$ and $E((\underline{X}-\underline{\mu})(\underline{X}-\underline{\mu})^T \equiv \Sigma$. Let $a \in \mathbb{R}^k$ and consider $Y_n \equiv \underline{a}^T \sqrt{n}(\overline{\underline{X}}_n - \underline{\mu})$. Then since

$$\underline{V}_n \equiv \sqrt{n}(\overline{\underline{X}}_n - \underline{\mu}) = n^{-1/2} \sum_{i=1}^{n} (\underline{X}_i - \underline{\mu}),$$

we see that

$$\begin{aligned} Y_n &= a^T \underline{V}_n = a^T \sqrt{n}(\overline{\underline{X}}_n - \underline{\mu}) \\ &= n^{-1/2} \sum_{i=1}^{n} \underline{a}^T (\underline{X}_i - \underline{\mu}) \\ &\equiv n^{-1/2} \sum_{i=1}^{n} Z_i = n^{1/2} \overline{Z}_n \end{aligned}$$

where $Z_i \equiv \underline{a}^T (\underline{X}_i - \underline{\mu})$ are i.i.d. with mean 0 and variance

$$\begin{aligned} Var(Z_1) &= Var(\underline{a}^T (\underline{X}_1 - \underline{\mu})) \\ &= \underline{a}^T E\{(\underline{X}_i - \underline{\mu})(\underline{X}_i - \underline{\mu})^T\} \underline{a} \\ &= \underline{a}^T \Sigma \underline{a}. \end{aligned}$$

Thus the one-dimensional (Lindeberg) CLT yields

$$Y_n \to_d Y \sim N_1(0, \underline{a}^T \Sigma \underline{a}) \stackrel{d}{=} \underline{a}^T \underline{V}$$

where $\underline{V} \sim N_k(0, \Sigma)$. That is, $\underline{a}^T \underline{V}_n \to_d \underline{a}^T \underline{V}$ for every $\underline{a} \in \mathbb{R}^k$. By the Cramér-Wold device this implies that $\underline{V}_n \to_d \underline{V} \sim N_k(0, \Sigma)$.