Statistics 522, Final Exam Solutions

Wellner; 3/21/2013

1. (24 points). **Define** *three* of the following six terms:
   (a) An asymptotically tight sequence $\{X_n\}$ in $\mathbb{R}$ (or $\mathbb{R}^d$).
   (b) The Lévy metric (for convergence in distribution of distribution functions on $\mathbb{R}$).
   (c) The characteristic function of a real-valued random variable X and of a random vector $\underline{X}$ with values in $\mathbb{R}^d$.
   (d) A Brownian bridge process $\mathbb{U}$ on $[0,1]$, and Brownian motion process $\mathbb{S}$ on $[0,\infty)$.
   (e) The conditional expectation $E(Y|\mathcal{D})$ of Y given a sub-σ-field $\mathcal{D}$.
   (f) The tail $\sigma-$field of a sequence of random variables $X_1, X_2, \ldots$.

   Solution: See PfS, chapters 7-10.

2. (40 points). Give careful **statements** of *any four* of the following seven theorems or results:
   (a) A result connecting convergence of characteristic functions of a sequence of random variables $\{X_n\}$ to tightness of the sequence $\{X_n\}$.
   (b) The Mann-Wald or continuous mapping theorem.
   (c) Helly's selection theorem.
   (d) Any result connecting a Brownian motion process $\mathbb{S}$ on $[0,1]$ or $[0,\infty)$ to a Brownian bridge process $\mathbb{U}$ on $[0,1]$.
   (e) The Cramér - Lévy continuity theorem for characteristic functions.
   (f) A basic triangular array central limit theorem.
   (g) The Lindeberg-Feller central limit theorem.

   Solution: See PfS, chapters 7-10.

3. (40 points) Suppose that $\underline{X} = (X_1, \ldots, X_d)$ is a random vector in $\mathbb{R}^d$ with distribution function
$$F_{\underline{X}}(\underline{x}) = P(\underline{X} \le \underline{x}) = P(X_1 \le x_1, \ldots, X_d \le x_d)$$
for all $\underline{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$.
(a) Define the characteristic function $\phi_{\underline{X}}(\underline{t})$ of $\underline{X}$ at $\underline{t} \in \mathbb{R}^d$.
(b) Let $\underline{a} \in \mathbb{R}^d$ be fixed and define the random variable Y by $Y = \underline{a}^T\underline{X} = \sum_{j=1}^d a_j X_j$ Express the characteristic function ϕ_Y in terms of $\phi_{\underline{X}}$, the characteristic function of the random vector $\underline{X}$.
(c) Now suppose that $\underline{X}_1, \underline{X}_2, \ldots$ is a sequence of random vectors in $\mathbb{R}^d$. The Cramér-Wold device says that $\underline{X}_n \to_d \underline{X}$ in $\mathbb{R}^d$ if and only if $\underline{a}^T\underline{X}_n \to_d \underline{a}^T\underline{X}$ in $\mathbb{R}$ for all $\underline{a} \in \mathbb{R}^d$. Sketch a proof of this result using characteristic functions and (b).

Solution: (a) The characteristic function $\phi_{\underline{X}}(\underline{t})$ of $\underline{X}$ is defined by

$$\phi_{\underline{X}}(\underline{t}) = E\exp(i\underline{t}^T\underline{X}) = E\exp(i\sum_{j=1}^{d} t_j X_j).$$

(b) If $Y = \underline{a}^T\underline{X} = \sum_{j=1}^{d} a_j X_j$, then the characteristic function ϕ_Y of Y is given by

$$\phi_Y(t) = E\exp(itY) = E\exp(it\underline{a}^T\underline{X}) = \phi_{\underline{X}}(t\underline{a})$$

(c) Suppose that $Y_n \equiv \underline{a}^T\underline{X}_n \to_d a^T\underline{X} \equiv Y$ for all $\underline{a} \in \mathbb{R}^d$. Then using (b)

$$\begin{aligned} \phi_{\underline{X}_n}(t\underline{a}) &= \phi_{Y_n}(t) = E\exp(itY_n) \\ &\to E\exp(itY) = \phi_Y(t) = \phi_{\underline{X}}(t\underline{a}) \end{aligned}$$

for all $t \in \mathbb{R}$ and all $\underline{a} \in \mathbb{R}^d$. Thus the characteristic functions of the sequence of random vectors $\underline{X}_n$ converge to the characteristic function of the random vector $\underline{X}$. Since this is the characteristic function of a proper (multivariate) distribution, it follows that $\underline{X}_n \to_d \underline{X}$.

Conversely, if $\underline{X}_n \to_d \underline{X}$, then

$$Y_n \equiv \underline{a}^T\underline{X}_n \to_d \underline{a}^T\underline{X} = Y$$

by the Mann-Wald or continuous mapping theorem.

4. (40 points) (a) Suppose that $X_1, X_2, \ldots$ are i.i.d. random variables with $E(X_1) = 0$, $Var(X_1) = 1$, and let $X_{n,i} \equiv a_{n,i}X_i$ for $i = 1, \ldots, n$ where $\{a_{n,i} : \ 1 \le i \le n\}$ satisfy $\sum_{i=1}^n a_{n,i}^2 \to v^2 \in (0,\infty)$ and $\max_{1\le i\le n}|a_{n,i}| \to 0$. Show that $S_n \equiv \sum_{i=1}^n X_{n,i}$ satisfies $S_n \to_d vZ \sim N(0, v^2)$.
(b) Suppose that $X_1, X_2, \ldots$ are i.i.d. as in (a) and let $a_{n,i} = n^{-1/2}(i/n)$ for $i \in \{1, \ldots, n\}$. Show that the hypotheses of (a) are satisfied for some v^2 (find the particular value in this case) and hence that $S_n = n^{-1/2}\sum_{i=1}^n (i/n)X_i \to_d N(0, v^2)$.
(c) Now suppose that the X_i's are as in (a), but that $a_{n,i} = n^{-1/2}(i/n)^\alpha$ for some $\alpha \in \mathbb{R}$. For what values of α does it hold that $S_n \to_d N(0, v_\alpha^2)$ for some $v_\alpha^2 < \infty$? For the values for which this holds, compute v_α^2 as a function of α.

Solution: (a) Now $E(X_{n,i}) = 0$ and $Var(X_{n,i}) = a_{n,i}^2$, so

$$Var(S_n) \equiv \sigma_n^2 = \sum_{i=1}^n a_{n,i}^2 \to v^2$$

by hypothesis and we have

$$M_n^2 \equiv \frac{\max_{1\le i\le n} a_{n,i}^2}{\sum_{i=1}^n a_{n,i}^2} \to \frac{0}{v^2} = 0.$$

To apply the Lindeberg-Feller CLT we need to check the Lindeberg condition. Let $\epsilon > 0$. Then

$$
\begin{aligned}
L_n(\epsilon) &\equiv \frac{1}{\sigma_n^2}\sum_{i=1}^n E\{X_{n,i}^2 1_{[|X_{n,i}|\ge \epsilon\sigma_n]}\} \\
&= \frac{1}{\sigma_n^2}\sum_{i=1}^n a_{n,i}^2 E\{X_i^2 1_{[|a_i||X_i|\ge \epsilon\sigma_n]}\} \\
&\le \frac{1}{\sigma_n^2}\sum_{i=1}^n a_{n,i}^2 E\{X_1^2 1_{[|X_1|\ge \epsilon/(\max_{1\le i\le n}|a_{n,i}|/\sigma_n)]}\} \\
&= E\{X_1^2 1_{[|X_1|\ge \epsilon/M_n]}\} \to 0
\end{aligned}
$$

by the dominated convergence theorem since $M_n \to 0$. Hence the Lindeberg condition holds and we conclude that $S_n/\sigma_n \to_d Z \sim N(0,1)$. But $\sigma_n^2 \to v^2$ so we conclude that $S_n \to_d vZ \sim N(0, v^2)$.

(b) When $a_{n,i} = n^{-1/2}(i/n)$ we have

$$
\begin{aligned}
\sigma_n^2 &= \sum_{i=1}^n a_{n,i}^2 = \frac{1}{n^3}\sum_{i=1}^n i^2 \\
&= \frac{n(n+1)(2n+1)}{6n^3} \to \frac{2}{6} = \frac{1}{3} \equiv v^2
\end{aligned}
$$

(Alternatively $\sigma_n^2 = n^{-1}\sum_{i=1}^n (i/n)^2 \to \int_0^1 t^2 dt = 1/3$.) Furthermore,

$$
M_n = \frac{\max_{1\le i\le n} a_{n,i}^2}{\sigma_n^2} = \frac{n^{-1/2}}{\sigma_n^2} \to 0.
$$

Thus the conditions of (a) hold and we conclude that $S_n \to_d N(0, 1/3)$.

(c) When $a_{n,i} = n^{-1/2}(i/n)^\alpha$ we have

$$
\begin{aligned}
\sigma_n^2 &= \sum_{i=1}^n a_{n,i}^2 = \frac{1}{n}\sum_{i=1}^n (i/n)^2\alpha \\
&\to \int_0^1 t^{2\alpha} dt = \frac{1}{2\alpha+1} t^{2\alpha+1}|_0^1 = \frac{1}{2\alpha+1} \equiv v^2
\end{aligned}
$$

if $2\alpha + 1 > 0$; i.e. if $\alpha > -1/2$. Furthermore,

$$
\begin{aligned}
M_n &\equiv \frac{\max_{1\le i\le n} a_{n,i}^2}{\sigma_n^2} = \frac{n^{-1/2}(i/n)^{2\alpha}}{\sigma_n^2} \\
&= \begin{cases} \frac{n^{-1/2}}{\sigma_n^2}, & \text{if } \ \alpha \ge 0, \\ \frac{n^{-1/2} n^{2\alpha}}{\sigma_n^2}, & \text{if } \ -1/2 < \alpha < 0 \end{cases} \\
&\to_d \ 0 \ \text{ if } \ \alpha > -1/2.
\end{aligned}
$$

Thus the hypotheses of (a) are satisfied for $\alpha > -1/2$ and we conclude that $S_n \to_d vZ \sim N(0, v^2)$ with $v^2 = 1/(2\alpha + 1)$.

5. (25 points). Suppose that X and Y are random variables on the probability space $(\Omega, \mathcal{A}, P)$ with $X \in L_2(P)$ and $Y \in L_2(P)$ (so that $XY \in L_1(P)$), and suppose that $\mathcal{D}$ is a sub sigma-field of $\mathcal{A}$. Show that

$$Cov(X,Y) = E\left[Cov(X,Y|\mathcal{D})\right] + Cov(E(X|\mathcal{D}), E(Y|\mathcal{D}))$$

where

$$Cov(X,Y|\mathcal{D}) = E\left[(X - E(X|\mathcal{D})(Y - E(Y|\mathcal{D}))|\mathcal{D}\right].$$

(This generalizes our formula for the variance of a random variable X obtained in the midterm exam.)

Solution: Now by adding and subtracting $E(X|\mathcal{D})$ and $E(Y|\mathcal{D})$ we have

$$\begin{aligned} Cov(X,Y) &= E(X - EX)(Y - E(Y)) \\ &= E\{(X - E(X|\mathcal{D}) + E(X|\mathcal{D}) - E(X))((Y - E(Y|\mathcal{D}) + E(Y|\mathcal{D}) - E(Y))\} \\ &= E\{(X - E(X|\mathcal{D}))(Y - E(Y|\mathcal{D}))\} + E\{(E(X|\mathcal{D}) - E(X))(E(Y|\mathcal{D}) - E(Y))\} \\ &= E[Cov(X,Y|\mathcal{D})] + Cov[E(X|\mathcal{D}), E(Y|\mathcal{D})] \end{aligned}$$

since we have

$$\begin{aligned} E\{(X - E(X|\mathcal{D}))(E(Y|\mathcal{D}) - EY)\} &= EE\{(X - E(X|\mathcal{D}))(E(Y|\mathcal{D}) - EY)|\mathcal{D}\} \\ &= E\{(E(Y|\mathcal{D}) - EY)E[(X - E(X|\mathcal{D}))|\mathcal{D}]\} \\ &= E\{(E(Y|\mathcal{D}) - EY)(E(X|\mathcal{D})) - E(X|\mathcal{D}))\} \\ &= 0, \end{aligned}$$

and similarly for $E\{(Y - E(Y|\mathcal{D}))(E(X|\mathcal{D}) - EX)\}$.

6. (40 points).
Let $X \sim N(\mu, 1)$ with $\mu > 0$.
(a) Suppose that ϵ is a Rademacher random variable independent of X; i.e. $P(\epsilon = \pm 1) = 1/2$. Compute the distribution function G of the random variable $Y = \epsilon X$ and find its density function g. Give rough plots of g and G for $\mu = 3$.
(b) Now suppose the $\epsilon \equiv 2 \cdot 1_{[X \geq \mu]} - 1$. Find the distribution of ϵ. Is it independent of X?
(c) Find the distribution function H of $W = \epsilon X$ when ϵ is as defined in (b) with $\mu = 0$ and with $X \sim N(0,1)$. Does it have a density h with respect to Lebesgue measure? If so, compute it. Is the distribution function H the same as the distribution G you found in (a) specialized to have $\mu = 0$?
(d) What is the characteristic function of $X \sim N(\mu, 1)$? What is the characteristic function of ϵ with $X \sim N(\mu, 1)$ and ϵ as in (a)? What is the characteristic function of W with W as defined in (c). Do these all agree when $\mu = 0$?

Solution:

Solution: (a) Since X has distribution function

$$F(x) = P(X \le x) = P(X - \mu \le x - \mu) = P(Z \le x - \mu) = \Phi(x - \mu)$$

where Φ is the distribution function of $Z \sim N(0,1)$, it follows that the d.f. of $Y = \epsilon X$ is given by

$$\begin{aligned} G(y) \equiv P(Y \le y) &= P(\epsilon X \le y) = P(X \le y,\ \epsilon = 1) + P(-X \le y,\ \epsilon = -1) \\ &= \Phi(y - \mu) \cdot (1/2) + (1 - \Phi(-y - \mu)) \cdot (1/2). \end{aligned}$$

The density g of G is given by

$$g(y) = (1/2)\{\phi(y - \mu) + \phi(-y - \mu)\} = \frac{1}{2}\{\phi(y - \mu) + \phi(y + \mu)\}$$

where $\phi(z) = (2\pi)^{-1/2} e^{-z^2/2}$ is the standard normal density. Here are plots of the d.f. G and the density g when $\mu = 3$.

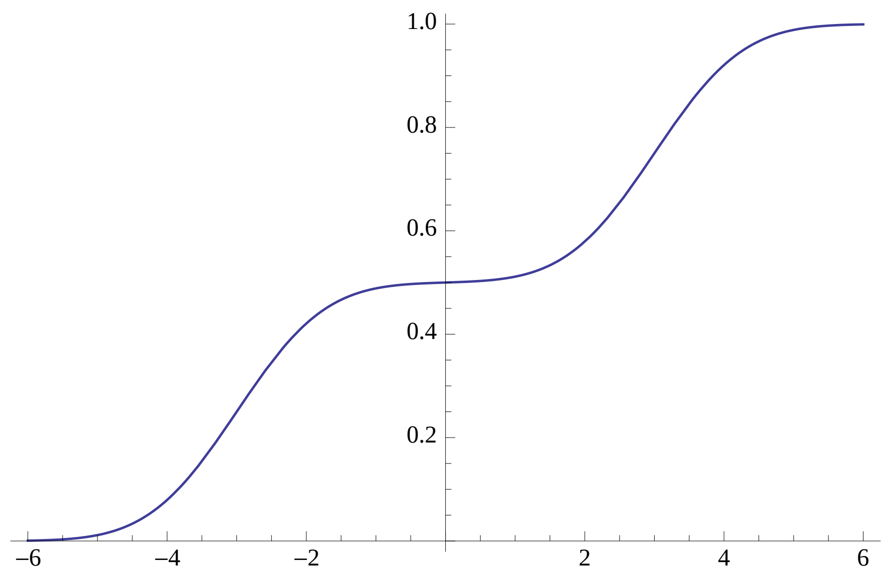

Figure 1: The distribution function G of $Y = \epsilon X$ with $X \sim N(3,1)$

(b) We compute

$$P(\epsilon = 1) = P(X \ge \mu) = P(X - \mu \ge 0) = P(Z \ge 0) = 1/2$$

and, similarly $P(\epsilon = -1) = P(X < \mu) = 1/2$. Thus $\epsilon \sim$ Rademacher. It is *not* independent of X since

$$P(X \ge \mu, \epsilon = 1) = P(X \ge \mu) = 1/2 \ne (1/2) \cdot (1/2) = P(X \ge \mu) P(\epsilon = 1).$$

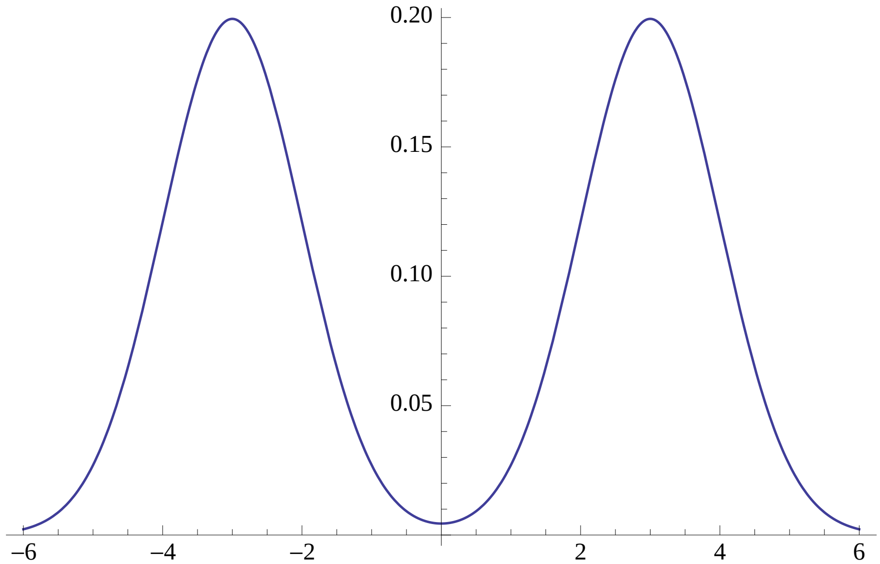

Figure 2: The density function g of $Y = \epsilon X$ with $X \sim N(3,1)$

(c) First note that when $\mu = 0$, $W = (2 \cdot 1_{[X \geq 0]} - 1)X \equiv T(X)$ where

$$\begin{aligned} T(x) &= (2 \cdot 1_{[\mu,\infty)}(x) - 1)x = \begin{cases} x, & x \geq 0, \\ -x, & x < 0. \end{cases} \\ &= |x| \end{aligned}$$

Thus we compute, for $w \geq 0$,

$$\begin{aligned} H(w) &= P(W \leq w) = P(|X| \leq w) = P(-w \leq X \leq w) \\ &= \Phi(w) - \Phi(-w), \end{aligned}$$

while $H(w) = 0$ for $w \leq 0$. This distribution function has density

$$h(w) = 2\phi(w) 1_{[0,\infty)}(w).$$

This d.f. H and density h are *not at all the same* as the d.f. G with density g found in (a) since with $\mu = 0$ the density in (a) is given by $g(y) = \phi(y)$ for all $y \in \mathbb{R}$.
(d) The characteristic function of $X \sim N(\mu, 1)$ is $\phi_X(t) = \exp(it\mu - (1/2)t^2)$.
The characteristic function of $Y = \epsilon X$ with $\epsilon \sim$ Rademacher independent of X is $\phi_Y(t) = Re(\phi_X)(t) = e^{-t^2/2} \cdot \cos(\mu t)$. When $\mu = 0$ this becomes $\phi_Y(t) = e^{-t^2/2}$. The characteristic function of W in (c) is given by

$$\phi_W(t) = \int_0^\infty e^{itw} 2\phi(w) dw = \sqrt{\frac{2}{\pi}} \int_0^\infty e^{itw - w^2/2} dw$$

which takes on complex values, so it is not the same as the previous two characteristic functions which are equal and real-valued when $\mu = 0$

7. (40 points)
Suppose that $X_n \sim N(\mu_n, \sigma_n^2)$ where $\mu_n \in \mathbb{R}$, $\sigma_n^2 > 0$.
(a) Suppose that $\{\mu_n\}$ and $\{\sigma_n^2\}$ are bounded. Show that $\{X_n\}$ (or the corresponding sequence of distribution functions $\{F_n\}$ of $\{X_n\}$) are tight.
(b) If $\{\mu_n\}$ is unbounded, show that $\{X_n\}$ is not tight.
(c) If $\{\mu_n\}$ is bounded but $\{\sigma_n^2\}$ is unbounded, show that $\{X_n\}$ is not tight.
(d) Suppose that $\mu_n = (-1)^n a$ with $a > 0$ and $\sigma_n^2 = 1 + \exp(-\pi n(-1)^n)$ for $n \geq 1$. Is there any tight subsequence $\{X_{n'}\}$ of $\{X_n\}$? Identify all the subsequences $\{X_n'\}$ of $\{X_n\}$ which converge in sub-distribution or in distribution.

Solution: (a) If $\{\mu_n\}$ and $\{\sigma_n^2\}$ are bounded, then $EX_n^2 = Var(X_n)+(EX_n)^2 = \sigma_n^2+\mu_n^2$ is bounded by some finite constant M. Thus it follows from Markov's inequality that $\{X_n\}$ is tight:

$$\begin{aligned}\limsup_{n\to\infty} P(|X_n| \geq \lambda) &\leq \limsup_{n\to\infty} \frac{E(X_n^2)}{\lambda^2} \leq \frac{M}{\lambda^2} \\ &\to 0 \text{ as } \lambda \to \infty.\end{aligned}$$

(b) If $\{\mu_n\}$ is unbounded, then either $\mu_n > b$ or $\mu_n < a$ for some a, b, and n. In the first case $P(X_n > b) \geq 1/2$, while in the second case $P(X_n < a) \geq 1/2$. Thus $\{X_n\}$ is not tight if μ_n is unbounded.
(c) If $\{\mu_n\}$ is bounded, say by B, then

$$\begin{aligned}P(X_n \leq a) &= P(X_n - \mu_n \leq a - \mu_n) \geq P(X_n - \mu_n \leq (a - B)) \\ &= P(Z \geq (a - B)/\sigma_n)\end{aligned}$$

where $Z \sim N(0, 1)$. If $\{\sigma_n^2\}$ is unbounded, then $P(Z \geq (a-B)/\sigma_n) \to 1/2$ along some subsequence, and hence $\{X_n\}$ cannot be tight.
(d) Now the sequence μ_n oscillates between $-a$ and $+a$ according as n is odd or even, while $\sigma_{2n}^2 = 1 + \exp(-\pi 2n) \to 1$ as $2n \to \infty$ (through the even integers), and $\sigma_{2n+1}^2 = 1+\exp(+\pi(2n+1)) \to +\infty$ as $2n+1 \to$ (through the odd integers). Thus the subsequence $\{X_{2n}\}$ is tight and $X_{2n} \to_d N(a, 1)$. On the other hand the subsequence $\{X_{2n+1}\}$ is not tight, and we have $X_{2n+1} \to_{s.d.} (\delta_{-\infty} + \delta_{+\infty})/2$, the sub-distribution function with mass 1/2 at each of $\pm\infty$.