ORFE 524: Statistical Theory and Methods
Homework 5

Zachary Hervieux-Moore

Friday 2$^{\text{nd}}$ December, 2016

Exercise 1: Let $X_n \in L^p$ and $X \in L^p$. Show that $X_n \xrightarrow{L_p} X$ implies $\|X_n\|_p \to \|X\|_p$. Conversely, does $\|X_n\|_p \to \|X\|_p$ imply $X_n \xrightarrow{L_p} X$? Why?

Note: Here L^p is defined as the set of all random variables on probability space (Ω, Σ, P) such that $E[|X_n|^p] < \infty$.

Answer: Since $X_n \xrightarrow{L_p} X$, then

$$E[|X_n - X|^p] \to 0$$

By the the reverse triangle inequality,

$$\Big| \|X_n\|_p - \|X\|_p \Big| \leq \|X_n - X\|_p = E[|X_n - X|^p]^{1/p}$$

However, since we have $E[|X_n - X|^p] \to 0$, then the right hand side goes to 0. Thus the left side also goes to 0 by the squeeze theorem. Therefore

$$\|X_n\|_p \to \|X\|_p$$

The converse does not hold true. Consider the constant random variable of $X_n = 1$ and $X \sim 2^{1/p} \cdot Bernoulli(1/2)$. Then clearly

$$\|X_n\|_p = 1$$
$$\|X\|_p = (0^p \cdot 1/2 + (2^{1/p})^p \cdot 1/2))^{1/p} = 1$$

and so $\|X_n\|_p \to \|X\|_p$. However,

$$E[|X_n - X|^p] = E[|X_n - X|^p] = |-1|^p \cdot 1/2 + |2^{1/p} - 1|^p \cdot 1/2$$

Which holds for all n and so does not converge to 0. Thus, we do not have convergence in L^p.

Exercise 2: Let $X_n, Y_n \in \mathbb{R}^d$ with $X_n \xrightarrow{\bullet} X$ and $Y_n \xrightarrow{\bullet} Y$, where $\bullet$ is either "a.s.", "in probability", or "in L^p" (for L^p, assuming $d = 1$).

a) Show that $X_n + Y_n \xrightarrow{\bullet} X + Y$ (in the cases of "a.s.", "in probability", and "in L^p").

b) Show that $X_n^T Y_n \xrightarrow{\bullet} X^T Y$ (in the cases of "a.s." and "in probability").

c) Let $d = 1$, and $P(Y = 0) = 0$, show that

$$\frac{X_n}{Y_n} \xrightarrow{\bullet} \frac{X}{Y}$$

in the cases of "a.s." and "in probability".

d) Explain why convergence in distribution fails for a), b), and c).

Note: You are expected to answer these questions without using continuity theorems.

Answer:

a) For "a.s." convergence, we have

$$P(X_n + Y_n \to X + Y) \geq P(X_n \to X \text{ and } Y_n \to Y)$$

Since the intersection of two sets of probability 1 (by a.s.) occurs with probability 1, then

$$P(X_n + Y_n \to X + Y) = 1$$

Thus, $X_n + Y_n \to X + Y$ "a.s."

Now for "in probability", we first note the following inequality for random variables U and V

$$1_{\{U+V<\epsilon\}} \geq 1_{\{U<\epsilon/2\}} \cdot 1_{\{V<\epsilon/2\}}$$

Taking expectations

$$P(U + V < \epsilon) \geq P(U < \epsilon/2, V < \epsilon/2)$$

Now using a bound on intersection of events,

$$P(U < \epsilon/2, V < \epsilon/2) \geq \max(0, P(U < \epsilon/2) + P(V < \epsilon/2) - 1)$$

Picking $U = |X_n - X|$ and $V = |Y_n - Y|$

$$P(|X_n - X| + |Y_n - Y| < \epsilon)$$
$$\geq \max(0, P(|X_n - X| < \epsilon/2) + P(|Y_n - Y| < \epsilon/2) - 1)$$

We now upper bound using the triangle inequality,

$$P(|X_n - X + Y_n - Y| < \epsilon) \geq P(|X_n - X| + |Y_n - Y| < \epsilon)$$

Thus,

$$P(|X_n + Y_n - (X + Y)| < \epsilon)$$
$$\geq \max(0, P(|X_n - X| < \epsilon/2) + P(|Y_n - Y| < \epsilon/2) - 1)$$

For "in probability", both of the lower bound terms converge to 1 due to convergence "in probability". Thus,

$$P(|X_n + Y_n - (X + Y)| < \epsilon) \to 1$$

which gives convergence "in probability"

For L^p we have that,

$$\|X_n - X\|_p \to 0 \text{ and } \|Y_n - Y\|_p \to 0$$

Minkowski inequality gives

$$\|X_n + Y_n - (X + Y)\|_p \leq \|X_n - X\|_p + \|Y_n - Y\|_p$$

Thus, the right hand side converges to zero and so does the left hand side. This gives us convergence in L^p. Note that all the p-norms are well defined since L^p is closed under addition and we know all the individual terms are in L^p by assumption.

b) For "a.s.", we note

$$P(X_n^T Y_n \to X^T Y) \geq P(X_n \to X \text{ and } Y_n \to Y)$$

Since the intersection of two sets of probability 1 (by a.s.) occurs with probability 1, then

$$P(X_n^T Y_n \to X^T Y) = 1$$

Thus, $X_n^T Y_n \to X^T Y$ "a.s."

Now for "in probability". Since $X_n \to X$ and $Y_n \to Y$ in probability. For a given subsequence $X_n'^T Y_n'$, we can find a suitable subsubsequence such that X_n'' and Y_n'' converge a.s. By the previous part, since they converge a.s. then $X_n''^T Y_n'' \to X^T Y$ a.s. Thus, for every subsequence, there is a subsubsequence that converges a.s. and so $X_n^T Y_n \to X^T Y$ "in probability".

c) Since all the random variables are well defined ($P(Y = 0) = P(Y_n = 0) = 0$) then we can simply define $Y_n' = \frac{1}{Y_n}$ and $Y' = \frac{1}{Y}$.

$$\frac{X_n}{Y_n} = X_n Y_n' \text{ and } \frac{X}{Y} = XY'$$

This changes the problem to $X_n Y_n' \xrightarrow{\bullet} XY'$. Using the previous part b) with $d = 1$ completes the proof.

d) Consider symmetric variables. That is, $X \sim N(0, 1)$. Then $X \to -X$ in distribution. However, $X + (-X) = 0$ while $N(0, 1) + N(0, 1) \neq 0$ necessarily. Thus we have that a) fails. The issue is that there is no information on the joint distributions for convergence in distributions. Thus, $N(0, 1) + N(0, 1)$ can equal many things depending on how the distributions are dependent. Independent means $N(0, 1) + N(0, 1) = N(0, 2)$. In my case, $N(0, 1) + N(0, 1) = 0$. Thus adding distributions is only meaningful when the joint distribution is known.

Exercise 3: Suppose that X_n and X are random variables.

1) Prove that $X_n \xrightarrow{L^p} X$ implies $X_n \xrightarrow{p} X$.

2) Let $\bullet$ represent convergence a.s., in probability, in distribution, or L^p. Show that $X_n \xrightarrow{\bullet} X$ if and only if every subsequence $X_{n_k} \xrightarrow{\bullet} X$.

3) If $X_n \xrightarrow{d} c$ where $c \in \mathbb{R}$, then $X_n \xrightarrow{p} c$.

Answer:

1) Since we have convergence in L^p we have

$$E[|X_n - X|^p] \to 0$$

By Markov's inequality

$$P(|X_n - X| > \epsilon^{1/p}) \leq \frac{E[|X_n - X|^p]}{\epsilon} \to 0$$

Thus convergence in L^p implies convergence in probability. Since the left hand side is 0, then $P(|X_n - X| < \epsilon^{1/p}) \to 1$

2) If every subsequence converges $X_{n_k} \xrightarrow{\bullet} X$, then it is immediate that $X_n \xrightarrow{\bullet} X$ since we just pick n_k to be n. Now, suppose that $X_n \xrightarrow{\bullet} X$. By contradiction, suppose such a subsequence X_{n_k} existed that did not converge. Then, for every ϵ in the definitions of the $\bullet$ convergence, we could not find a N_ϵ sufficiently large to satisfy the conditions. Because if you pick N_ϵ, I simply have to choose $n_k > N_\epsilon$ which is always possible because the subsequence is infinite. Then we cannot say statements of the sort "$\forall n \geq N_\epsilon$" and hence no convergence is possible. Thus, no such X_{n_k} can exist if we have $\bullet$ convergence.

3) If $X_n \xrightarrow{d} X = c$ then the pdf becomes $P(X = c) = 1$. Thus, we have

$$P(|X_n - X| < \epsilon) = P(|X_n - c| < \epsilon)$$

Picking N_ϵ sufficiently large

$$P(|X_{N_\epsilon} - c| < \epsilon) = 1$$

Which achieves convergence in probability.

Exercise 4: Suppose X is a random variable and $F(x)$ is its distribution function.

1) Assume for simplicity that F is invertible, and let $X^* = F^{-1}(U)$, where $U \sim U[0,1]$. Prove that $X \stackrel{\text{d}}{=} X^*$.

2) In general case where F is not necessarily invertible, let

$$X^* = F^{-1}(U) := \inf\{x : F(x) \geq U\}, \text{ where } U \sim U[0,1]$$

Prove that the claim in part 1) is still true.

Answer:

1) We begin with $X^* = F^{-1}(U)$ exists and well defined. This implies $F^{-1}(\cdot)$ is strictly increasing and thus strictly monotone.

$$\begin{gathered} P(X^* \leq x^*) \\ = P(F^{-1}(U) \leq x^*) \\ = P(U \leq F(x^*)) \end{gathered}$$

Where $P(U \leq F(x^*)) = F(x^*) = P(X \leq x^*)$. Thus, $P(X^* \leq x^*) = P(X \leq x^*)$ and we conclude $X \stackrel{\text{d}}{=} X^*$.

2) In this case, we have the following

$$F(F^{-1}(u)) \geq u$$

By right-continuity. We wish to show $u \leq F(x^*)$ is equivalent to $F^{-1}(u) \leq x^*$. If $u \leq F(x^*)$, then by definition of $F^{-1}(U)$, $F^{-1}(u) \leq x^*$. If $F^{-1}(u) \leq x^*$ then we use the fact $F(F^{-1}(u)) \geq u$ to get $u \leq F(x^*)$. Thus we showed

$$u \leq F(x^*) \iff F^{-1}(u) \leq x^*$$

Thus we can use the previous reasoning now

$$\begin{gathered} P(X^* \leq x^*) \\ = P(F^{-1}(U) \leq x^*) \\ \iff P(U \leq F(x^*)) \end{gathered}$$

Thus, $P(X^* \leq x^*) = P(X \leq x^*)$ and we conclude $X \stackrel{\text{d}}{=} X^*$.

Exercise 5: Suppose that $\{X_n\}_{n\geq 1}$ and X are d-dimensional random vectors. Prove the following claims by reducing them to the scalar case.

1) Show that $X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{p} X \implies X_n \xrightarrow{d} X$

2) Show that $X_n \xrightarrow{p} X$ if and only if for every subsequence $\{X_{n_k}\}_{k\geq 1}$, there is a further subsequence $\{X_{n_{k_j}}\}_{j\geq 1}$ such that $X_{n_{k_j}} \xrightarrow{a.s.} X$.

Note: You may use (without proof) the fact that these two arguments hold in the scalar case.

Answer:

1) First let's show $X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{p} X$. If $X_n \xrightarrow{a.s.} X$ then

$$P(\forall \epsilon, \exists N_\epsilon, \forall n \geq N_\epsilon, \|X_n - X\|_d < \epsilon) = 1$$

Expanding the norm,

$$P(\forall \epsilon, \exists N_\epsilon, \forall n \geq N_\epsilon, \sum_{i=1}^{d} |X_n^i - X^i|^d < \epsilon^d) = 1$$

Now define new scalar random variables $Y_n = \sum_{i=1}^{d} |X_n^i - X^i|^d$ and $Y = 0$. Since $Y_n \xrightarrow{a.s.} Y$ then we have $Y_n \xrightarrow{p} Y$. That is,

$$\begin{aligned}
&P(|Y_n - Y| < \epsilon^d) \to 1 \\
&= P(|\sum_{i=1}^{d} |X_n^i - X^i|^d - 0| < \epsilon^d) \to 1 \\
&= P(\sum_{i=1}^{d} |X_n^i - X^i|^d < \epsilon^d) \to 1 \\
&= P(\|X_n - X\|_d < \epsilon) \to 1
\end{aligned}$$

Now to show $X_n \xrightarrow{p} X \implies X_n \xrightarrow{d} X$. If $X_n \xrightarrow{p} X$ then

$$P(\|X_n - X\|_d < \epsilon) \to 1$$

In particular, for $\|v\|_d = 1$,

$$P(|v^T X_n - v^T X| < \epsilon) \to 1$$
$$= P(|\sum_{i=1}^{d} v^i (X_n^i - X^i)|_d < \epsilon) \to 1$$

Thus, picking $Y_n = \sum_{i=1}^d v^i X_n^i$ and $Y = \sum_{i=1}^d v^i X^i$. Then $Y_n \xrightarrow{p} Y$ in the scalar case and so applying the scalar fact, $Y_n \xrightarrow{d} Y$. But this implies, $\sum_{i=1}^d v^i X_n^i \xrightarrow{d} \sum_{i=1}^d v^i X^i$ which is exactly $v^T X_n \xrightarrow{d} v^T X$. Since v was arbitrary, by the Cramer-Wold device, this is an equivalent definition of convergence in distribution in higher dimensions.

2) We are in the random vector case. Thus $X_n \xrightarrow{p} X$ mean that

$$P(\|X_n - X\|_d < \epsilon) \to 1$$

Expanding the norm,

$$P(\sum_{i=1}^{d} |X_n^i - X^i|^d < \epsilon^d) \to 1$$

Again, define $Y_n = \sum_{i=1}^d |X_n^i - X^i|^d$ and $Y = 0$. Then $Y_n \xrightarrow{p} Y$. Then applying the scalar case yields the result. That is, there exists $Y_{n_{k_j}} \xrightarrow{a.s.} Y$ for every Y_{n_k}. Thus, there exists $\sum_{i=1}^d |X_{n_{k_j}}^i - X_{n_{k_j}}^i|^d$ which converges to 0 a.s. which implies that there exists $X_{n_{k_j}} \xrightarrow{a.s.} X$ for any X_{n_k}.

Exercise 6: Prove the following statements. Suppose that X_n and Y_n are random vectors.

1) Suppose $X_n \xrightarrow{d} X$, then $\|X_n\| = O_P(1)$

2) Suppose that $\|X_n\| = O_P(1)$ and $\|Y_n\| = o_P(1)$. Then, $Y_n^T X_n = o_P(1)$.

Answer:

1) Since $X_n \xrightarrow{d} X$ we also have $\|X_n\| \xrightarrow{d} \|X\|$ by continuity mapping results. Now, pick c_ϵ from the continuity set of $F_{\|X\|}$ such that

$$P(\|X\| \leq c_\epsilon) \geq 1 - \epsilon/2$$

By convergence in distributions

$$P(\|X_n\| \leq c_\epsilon) \to P(\|X\| \leq c_\epsilon)$$

So for large enough n, we have

$$P(\|X_n\| \leq c_\epsilon) \geq 1 - \epsilon$$

Thus, $\|X_n\|$ is $O_P(1)$.

2) Let $\epsilon, \delta > 0$, and $\|X_n\| \leq C_\delta$ with

$$P(\|X_n\| \leq C_\delta) \geq 1 - \delta/2 \ \forall n$$

Now, for $n \geq N_{\epsilon,\delta}$ let $\|Y_n\| \leq \epsilon/C_\delta$ with

$$P(\|Y_n\| \leq \epsilon/C_\delta) \geq 1 - \delta/2$$

Thus,

$$\|Y_n^T X_n\| \leq \frac{\epsilon}{C_\delta} \cdot C_\delta = \epsilon$$

With probability greater than $1 - \delta$. Thus, $Y_n^T X_n$ is $o_P(1)$.

Exercise 7: Let g be a bounded function on $\mathbb{R}$, and suppose that it is continuous a.e. $\mathbb{P}_X$. In other words, $\mathbb{P}(\omega : g \text{ continuous at } X(\omega)) = 1$. Show that $X_n \xrightarrow{\bullet} X \implies \mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$, where $\bullet$ means convergence "a.s.", "in probability", or "in distribution".

Answer: We showed in class that $X_n \xrightarrow{\bullet} X \implies g(X_n) \xrightarrow{\bullet} g(X)$. Now taking expectation,

$$\lim_{n\to\infty} \mathbb{E}[g(X_n)] = \lim_{n\to\infty} \int_{\mathbb{R}} g(X_n) dP$$

By the Dominated Convergence Theorem, $g(X_n)$ is dominated (since it is bounded) and it is integrable since

$$\int_{\mathbb{R}} |g(X_n)| dP \leq \int_{\mathbb{R}} M dP = M$$

So we can switch the limit and the integral

$$= \int_{\mathbb{R}} \lim_{n\to\infty} g(X_n) dP$$

Since we have $g(X_n) \xrightarrow{\bullet} g(X)$

$$= \int_{\mathbb{R}} g(X) dP = \mathbb{E}[g(X)]$$

Therefore, as $n \to \infty$ then $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$.

Exercise 8: Let $\{X_i\} \sim P^n$, with $\mathbb{E}[X_1^4] < \infty$. Let $var(X_1) = \sigma^2$. Recall that $S_n^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$ and $S_{n-1}^2 = \frac{n}{n-1} S_n^2$.

1) Derive the asymptotic distribution of $\sqrt{n}(S_{n-1}^2 - \sigma^2)$.

2) Consider $\{(x_i, y_i)\}_1^n$ i.i.d. and $\widehat{C}_n = \frac{1}{n}\sum_{i=1}^n x_i y_i - \bar{x}\bar{y}$. Suppose that $\mathbb{E}[x_1^{l_1} y_1^{l_2}] < \infty$ for any integers $0 \le l_1 \le 2$ and $0 \le l_2 \le 2$. Use Delta method to derive the asymptotic distribution of $\sqrt{n}(\widehat{C}_n - Cov(x_1, y_1))$. (You may give an expression without simplificaiton).

Answer:

1) We notice that

$$\sqrt{n}(\frac{n}{n-1} S_{n-1}^2 - \sigma^2) = \sqrt{n}(S_n^2 - \sigma^2) + \frac{\sqrt{n}}{n-1} S_n^2$$

Asymptotically, the second term vanishes and we are left with the same part we did in class which is distributed $N(0, \mathbb{E}[(X_1 - \mathbb{E}[X_1])^4] - \sigma^4)$

2) Let $Z_i = \begin{bmatrix} X_i Y_i \\ X_i \\ Y_i \end{bmatrix}$. Then, $\bar{Z} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n X_i Y_i \\ \bar{X} \\ \bar{Y} \end{bmatrix}$ and $\mathbb{E}[Z] = \begin{bmatrix} \mathbb{E}[XY] \\ \mathbb{E}[X] \\ \mathbb{E}[Y] \end{bmatrix}$.
Invoking CLT,

$$\sqrt{n}(\bar{Z} - \mathbb{E}[Z]) \xrightarrow{d} N(0, \Sigma)$$

Where $\Sigma = \begin{bmatrix} Var(XY) & Cov(XY, X) & Cov(XY, Y) \\ Cov(XY, X) & Var(X) & Cov(X, Y) \\ Cov(XY, Y) & Cov(X, Y) & Var(Y) \end{bmatrix}$.
Now pick $h(x, y, z) = x - yz$. Then

$$\nabla h = \begin{bmatrix} 1 \\ -z \\ -y \end{bmatrix} \implies \nabla h(\mathbb{E}[Z]) = \begin{bmatrix} 1 \\ -\mathbb{E}[Y] \\ -\mathbb{E}[X] \end{bmatrix}$$

Then using Delta Method

$$\sqrt{n}(h(Z) - h(\mathbb{E}[\mathbb{Z}]))$$
$$= \sqrt{n}(\widehat{C}_n - Cov(X, Y)) \xrightarrow{d} N(0, \nabla h^T(\mathbb{E}[Z])\Sigma\nabla h(\mathbb{E}[Z]))$$