Statistics 521, Problem Set 6 Solutions

Wellner; 11/8/2019

1. (a) Give an example of a sequence of random variables X_n, X (all defined on a common probability space $(\Omega, \mathcal{A}, P)$) satisfying $X_n \to_{a.s.} X$, but $E(X_n) \nrightarrow E(X)$.
(b) Give an example of a sequence of non-negative random variables X_n, X on a common probability space satisfying $E(X_n) \to E(X)$ but $X_n \nrightarrow_{a.s.} X$.
(c) Give an example of a sequence of random variables X_n, X satisfying $X_n \to_d X$, but $X_n \nrightarrow_{p,a.s.,1} X$.

Solution: (a) Let $U \sim \text{Uniform}[0,1]$. For $\alpha \ge 0$, let $X_n \equiv n^\alpha 1_{(1/(n+1),1/n]}(U)$. Then $X_n \to_{a.s.} 0 \equiv X$ since $X_n = 0$ for $n > 1/U$ and $P(U \in (0,1]) = 1$. But

$$E(X_n) = n^\alpha (n^{-1} - (n+1)^{-1}) = n^{\alpha-1}/(n+1) \to 1_{\{2\}}(\alpha) + \infty \cdot 1_{(2,\infty)}(\alpha).$$

Thus $E(X_n) \nrightarrow E(X)$ for $\alpha \ge 2$.
(b) Let $U \sim \text{Uniform}[0,1]$. For $0 < \alpha < 1$, $m \ge 1$, and $1 \le k \le 2^m$ define

$$Y_{m,k} \equiv (2^{m\alpha}) 1_{((k-1)/2^m, k/2^m]}(U);$$

Then, note that $\sum_{j=1}^m 2^{j-1} = 2(2^{m-1} - 1)$ and let $X_n \equiv X_{2(2^{m-1}-1)+k} \equiv Y_{m,k}$ for $m \ge 1$ and $1 \le k \le 2^m$. Then $E(X_n) = E(Y_{m,k}) = 2^{m\alpha} \cdot 2^{-m} = 2^{(\alpha-1)m} \to 0$ as $n = 2(2^{m-1} - 1) + k \to \infty$, but $X_n = Y_{m,k} > 0$ i.o. with probability 1 since $P(U \in (0,1]) = 1$. (The indicator functions $\{1_{((k-1)/2^m, k/2^m]}(U) : m \ge 1,\ 1 \le k \le 2^m\}$ are sometimes called the "dancing functions".)
(c) Suppose that $X_1, X_2, \ldots$ are independent and identically distributed random variables with common distribution function F all defined on a common probability space. (We will make this completely rigorous in chapter 5.) Then $F_n(x) = P(X_n \le x) = F(x)$ for all $n \ge 1$, so F_n certainly converges to F for all $x \in \mathbb{R}$ and in particular at all $x \in C_F$. Thus $X_n \to_d X$, but $X_n \nrightarrow_{a.s.,p,1}$. Alternatively, the X_n's could be taken to be defined on separate probability spaces $(\Omega_n, \mathcal{A}_n, P_n)$ with induced

distributions P_{X_n} on $\mathbb{R}$ with distribution functions $F_n(x) \equiv P(X_n \le x)$ satisfying $F_n \to_d F$. Now we cannot even talk about the random variables $X_n - X$, so $X_n \nrightarrow_{a.s.,p,1} X$. Here is yet a further simple example: Suppose that $X_n \equiv U \sim \text{Uniform}[0,1]$ for all $n \ge 1$. Suppose that $X \equiv 1 - U$. Then $X \sim \text{Uniform}[0,1]$ so $X_n \stackrel{d}{=} X$ for all n and hence $X_n \to_d X$, while $X_n \equiv U \ne_{a.s.} 1 - U \equiv_X$, so $X_n \nrightarrow_{p,a.s.} X$. Note that

$$\begin{aligned}\{|X_n - X| \ge \epsilon\} &= \{|U - (1-U)| \ge \epsilon\} = \{|2U - 1| \ge \epsilon\} \\ &= \{(1-\epsilon)/2 \le U \le (1+\epsilon)/2\}^c\end{aligned}$$

has $P(|X_n - X| \ge \epsilon) = 1 - \epsilon$ for every $n \ge 1$ and hence $X_n \nrightarrow_p X$. This also implies that $X_n \nrightarrow_{a.s.} X$.

2. PfS, Exercise 3.5.7, page 61, modified as follows: Suppose that $f_0, f_1, \ldots$ are ≥ 0, defined on a sigma-finite measure space $(\Omega, \mathcal{A}, \mu)$. (a) Suppose that $\int_\Omega f_n d\mu = 1$ for $n = 0, 1, \ldots$, and $f_n \to_{a.e.} f_0$ with respect to μ. Show that

$$\sup_{A \in \mathcal{A}} \Big| \int_A f_n d\mu - \int_A f_0 d\mu \Big| \to 0 \qquad \text{as} \ \ n \to \infty\,.$$

(b) Show that the conclusion of (a) holds if just $f_n \to_\mu f_0$ and $\int_\Omega f_n d\mu \to \int_\Omega f_0 d\mu$.

Solution: (a) As we saw in class,

$$\sup_{A \in \mathcal{A}} \left| \int_A f_n d\mu - \int_A f_0 d\mu \right| = \int (f_0 - f_n)^+ d\mu$$

where $(f_0 - f_n)^+ \to_{a.e.} 0$ and is dominated by the integrable function f_0. Hence the right side converges to 0 by the dominated convergence theorem. For the record, here are the details of the argument again: letting $\delta_n \equiv f_0 - f_n$ we have $\delta_n \to_{a.e.} 0$ and $\delta_n^{\pm} \to_{a.e.} 0$.

$$\begin{aligned}d_{TV}(P_n, P_0) &= \sup_{A \in \mathcal{A}} |P_n(A) - P_0(A)| \\ &= \sup_{A \in \mathcal{A}} \left| \int_A \delta_n d\mu \right|\end{aligned}$$

where

$$\int_A \delta_n d\mu + \int_{A^c} \delta_n d\mu = \int_\Omega \delta_n d\mu = 0.$$

so that $\int_{A^c} \delta_n d\mu = -\int_A \delta_n d\mu$. It follows that

$$\begin{aligned} 2\left|\int_A \delta_n d\mu\right| &= \left|\int_A \delta_n d\mu\right| + \left|\int_{A^c} \delta_n d\mu\right| \leq \int |\delta_n| d\mu \\ &= \int (\delta_n^+ + \delta_n^-) d\mu = 2\int \delta_n^+ d\mu \end{aligned}$$

and equality holds if $A = \{\delta_n \geq 0\}$. Note that $0 = \int \delta_n d\mu = \int (\delta_n^+ - \delta_n^-) d\mu$ implies that $\int \delta_n^- d\mu = \int \delta_n^+ d\mu$ was used in the last equality. It follows that

$$d_{TV}(P_n, P_0) = \sup_{A \in \mathcal{A}} \left|\int_A \delta_n d\mu\right| \leq \frac{1}{2}\int |\delta_n| d\mu = \int \delta_n^+ d\mu.$$

(b) If we have $f_n \to_\mu f_0$ and $\int f_n d\mu \to \int f_0 d\mu$, then we still have

$$\begin{aligned} &\sup_{A \in \mathcal{A}} \left|\int_A f_n d\mu - \int_A f_0 d\mu\right| && (1) \\ &\leq \sup_A \int_A |f_n - f_0| d\mu \\ &\leq \int_\Omega |f_n - f_0| d\mu = \int_\Omega |f_0 - f_n| d\mu \\ &= \int (f_0 - f_n)^+ d\mu + \int (f_0 - f_n)^- d\mu \\ &= \int (f_0 - f_n)^+ d\mu + \int (f_0 - f_n)^+ d\mu - D_n \\ &= 2\int (f_0 - f_n)^+ d\mu - D_n && (2) \end{aligned}$$

where

$$D_n \equiv \int_\Omega (f_0 - f_n) d\mu = \int_\Omega \{(f_0 - f_n)^+ - (f_0 - f_n)^-\} d\mu \to 0\,.$$

But the right side of (2) converges to 0 by the dominated convergence theorem together with $D_n \to 0$ by the hypothesis $\int f_n d\mu \to \int f_0 d\mu$.

3. Let $X_{n1}, \ldots, X_{nn}$ be independent, $X_{nk} \sim \text{Bernoulli}(p_{nk})$, and let $Y_n \sim \text{Poisson}(\sum_{k=1}^n p_{nk})$. Let P_n be the distribution of $\sum_{k=1}^n X_{nk}$ and let Q_n be the distribution of Y_n. Show that

$$d_{TV}(P_n, Q_n) \equiv \sup_{A \in \mathcal{B}} |P(S_n \in A) - P(Y_n \in A)| \le \sum_{k=1}^n p_{nk}^2.$$

Note that when $p_{nk} = p_n \to 0$ for all k and $np_n \to \lambda$, then $\sum_{k=1}^n p_{nk}^2 = np_n^2 = (np_n)^2/n = O(n^{-1})$.

Hint: Construct S_n and Y_n on a common probability space as follows: let $T_{nk} \sim \text{Poisson}(p_{nk})$, $k = 1, \ldots, n$ be independent, and let $Z_{nk} \sim \text{Bernoulli}(1 - (1 - p_{nk})e^{p_{nk}})$, $k = 1, \ldots, n$ be independent and independent of the T_{nk}'s. Define $X_{nk} = 1_{[T_{nk} \ge 1]} + 1_{[T_{nk}=0]} 1_{[Z_{nk}=1]}$. Set $S_n = \sum_{k=1}^n X_{nk}$, $Y_n = \sum_{k=1}^n T_{nk}$. Check that $X_{nk} \sim \text{Bernoulli}(p_{nk})$, $Y_n \sim \text{Poisson}(\sum_1^n p_{nk})$, and

$$\begin{aligned} P(T_{nk} = 0, X_{nk} = 1) &= e^{-p_{nk}} - (1 - p_{nk}) \\ P(T_{nk} \ge 1, X_{nk} = 0) &= 0, \quad P(T_{nk} \ge 2) = 1 - e^{-p_{nk}} - p_{nk} e^{-p_{nk}}. \end{aligned}$$

Show that

$$d_{TV}(P_n, Q_n) \le P(S_n \ne Y_n) \le \sum_{k=1}^n P(X_{nk} \ne T_{nk}) \le \sum_{k=1}^n p_{nk}^2.$$

Solution: Using the notation in the hint we first show that $X_{nk} \sim \text{Bern}(p_{nk})$: this follows since, using $P(Y_\lambda \ge 1) = 1 - P(Y_\lambda = 0) = 1 - e^{-\lambda}$ if $Y_\lambda \sim \text{Poisson}(\lambda)$,

$$\begin{aligned} P(X_{nk} = 1) &= P(T_{nk} \ge 1) + P(T_{nk} = 0) P(Z_{nk} = 1) \\ &= 1 - e^{-p_{nk}} + e^{-p_{nk}}(1 - (1 - p_{nk})e^{p_{nk}}) \\ &= p_{nk}. \end{aligned}$$

Next,

$$P(T_{nk} = 0, X_{nk} = 1) = e^{-p_{nk}}(1 - (1 - p_{nk})e^{p_{nk}} = e^{-p_{nk}} - (1 - p_{nk}),$$

while

$$P(T_{nk} \ge 1, X_{nk} = 0) = P(T_{nk} \ge 1,\ T_{nk} = 0) = 0,$$

and

$$P(T_{nk} \geq 2) = 1 - P(T_{nk} = 0 \text{ or } 1) = 1 - e^{-p_{nk}} - p_{nk} e^{-p_{nk}}.$$

Thus

$$\begin{aligned} P(X_{nk} \neq T_{nk}) &= P(X_{nk} = 0, T_{nk} = 1) + P(X_{nk} = 1, T_{nk} = 0) + P(T_{nk} \geq 2) \\ &= 0 + e^{-p_{nk}} - (1 - p_{nk}) + 1 - e^{-p_{nk}} - p_{nk} e^{-p_{nk}} \\ &= p_{nk}(1 - e^{-p_{nk}}) \leq p_{nk}^2. \end{aligned}$$

Thus for any $A \in 2^{\mathbb{N}}$,

$$\begin{aligned} &P(S_n \in A) - P(Y_n \in A) \\ &\quad = P([S_n \in A] \cap [S_s = Y_n]) + P([S_n \in A] \cap [S_n \neq Y_n]) \\ &\qquad - P([Y_n \in A] \cap [S_n = Y_n]) + P([Y_n \in A] \cap [S_n \neq Y_n]) \\ &\quad = P([S_n \in A] \cap [S_n \neq Y_n]) - P([Y_n \in A] \cap [S_n \neq Y_n]) \\ &\quad \leq P([S_n \in A] \cap [S_n \neq Y_n]) \leq P(S_n \neq Y_n). \end{aligned}$$

Similarly, by a symmetric argument,

$$P(S_n \in A) - P(Y_n \in A) \geq -P(S_n \neq Y_n),$$

and this yields

$$\begin{aligned} d_{TV}(P_n, Q_n) &\equiv \sup_{A \in 2^{\Omega}} |P(S_n \in A) - P(Y_n \in A))| \\ &\leq P(S_n \neq Y_n) \leq \sum_{k=1}^{n} P(X_{nk} \neq T_{nk}) \leq \sum_{k=1}^{n} p_{nk}^2. \end{aligned}$$

If $p_{nk} = \lambda/n$ for $1 \leq k \leq n$ for some $\lambda > 0$, then this bound yields

$$d_{TV}(P_n, Q_n) \leq n(\lambda/n)^2 = \frac{\lambda^2}{n}.$$

For still stronger results, see Barbour, Holst, and Janson (1992), *Poisson Approximation*.

4. Let $Z \sim N(0,1)$ and let $Y_r \sim \chi_r^2$ for $r > 0$. Thus $X_r \equiv Z/\sqrt{Y_r/r} \sim t_r$, the Student $t-$ distribution with r "degrees of freedom" with density

$$f_r(t) \equiv C_r \left(1 + \frac{t^2}{r}\right)^{-(r+1)/2}$$

where $C_r \equiv \Gamma((r+1)/2)/(\sqrt{\pi r}\Gamma(r/2)$. Let P_r denote the corresponding probability measure on $\mathbb{R}$.
(a) Show that $Y_r/r \to_p 1$ as $r \to \infty$.
(b) Show that $X_r \to_d Z \sim N(0,1)$ as $r \to \infty$.
(c) Show that $f_r(t) \to \phi(t) \equiv (2\pi)^{-1/2} e^{-t^2/2}$.
(d) Use the result of (c) to show that $d_{TV}(P_r, P_Z) \to 0$ as $r \to \infty$.

Solution: (a) Now $E(Y_r/r) = r/r = 1$ for all $r > 0$. Since the density of Y_r is given by

$$f_r(x) = \frac{1}{\Gamma(r/2)2^{r/2}} x^{r/2-1} \exp(-x/2) \quad \text{for} \quad x > 0,$$

it follows that

$$\begin{aligned} E(Y_r^2) &= \int_0^\infty x^2 f_r(x) dx = \frac{1}{\Gamma(r/2)2^{r/2}} \int_0^\infty x^{r/2+2-1} e^{-x/2} dx \\ &= \frac{\Gamma((r+4)/2) 2^{(r+4)/2}}{\Gamma(r/2)2^{r/2}} = 2^2 \left(\frac{r+4}{2} - 1\right)\left(\frac{r+4}{2} - 2\right) \\ &= r^2 + 2r \end{aligned}$$

where we used the duplication formula for the Γ function,

$$\Gamma(x) = (x-1)\Gamma(x-1).$$

Hence $Var(Y_r) = EY_r^2 - (E(Y_r))^2 = 2r$. Thus

$$P(|Y_r/r - 1| > \epsilon) \le \frac{2r}{r^2 \epsilon^2} \to 0$$

as $r \to \infty$. Thus $Y_r/r \to_p 1$.
(b) Now $X_r = Z/\sqrt{Y_r/r} \to_d Z \sim N(0,1)$ by (a) and Slutsky's theorem.
(c) The non-constant term in the t_r-density is

$$\left(1 + \frac{t^2}{r}\right)^{-(r+1)/2} \to \exp(-t^2/2) \quad \text{as} \quad r \to \infty$$

since $(1 + z/r)^r \to e^z$ for $z \in \mathbb{R}$. Since

$$1 = \int_{\mathbb{R}} f_r(t) dt = C_r \int_{\mathbb{R}} \left(1 + \frac{t^2}{r}\right)^{-(r+1)/2} dt$$

we have

$$\begin{aligned} C_r &= \left(\int_{\mathbb{R}} \left(1 + \frac{t^2}{r}\right)^{-(r+1)/2} dt \right)^{-1} \\ &= \Gamma((r+1)/2)/(\sqrt{\pi r}\Gamma(r/2) \\ &\to (2\pi)^{-1/2} \end{aligned}$$

by a direct argument using Stirling's formula for the Gamma function twice. It then follows that $f_r(t) \to (2\pi)^{-1/2} e^{-t^2/2} \equiv \phi(t)$ for all $t \in \mathbb{R}$.
(d) Thus by (c) and Scheffé's lemma we conclude that $d_{TV}(P_r, P_Z) \to 0$. (How fast does this convergence happen? Pinelis (2015) showed that $r d_{TV}(P_r, P_Z) \to C$ where

$$C = \frac{1}{2}\sqrt{\frac{7 + 5\sqrt{2}}{\pi e^{1+\sqrt{2}}}} \approx 0.3165 \ldots .)$$