MEASURE THEORETICAL PROBABILITY I HOMEWORK 5

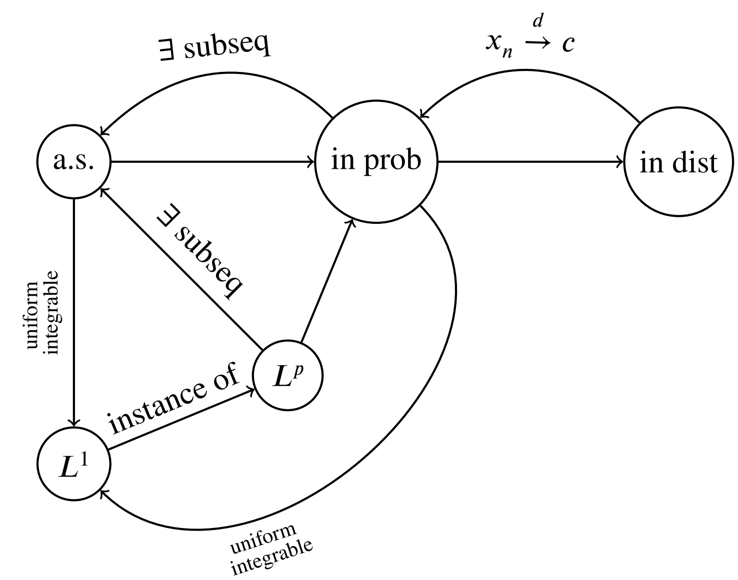
YUFEI FAN; CHENG PENG; TOMMENIX YU
STAT 38100
DUE FEB 16

Discussed with classmates.

Exercise 0.

Proof.

Prove the relations below:



conv a.s. ⇒ conv in prob:

$X_n \xrightarrow{as} X$ means that

$$\mathbb{P}\left(\forall \varepsilon > 0, \exists N s.t. \forall n > N, |X_n - X| < \varepsilon\right) = 1$$

and we can move the universal quantifier outside to get

$$\forall \varepsilon > 0, \mathbb{P}\left(\exists N s.t. \forall n > N, |X_n - X| < \varepsilon\right) = 1$$

which then implies

$$\forall \varepsilon > 0, \lim_{n\to\infty} \mathbb{P}\left(|X_n - X| < \varepsilon\right) = 1$$

which is what we want.

conv in prob $\Rightarrow \exists$ subsequence conv a.s.:

To prove convergence almost surely for some subsequence is to prove

$$\mathbb{P}\left(\limsup_{n\to\infty} |X_{\phi(n)} - X| > 0\right) = 0$$

where we can rewrite

$$\left\{\limsup_{n\to\infty} |X_{\phi(n)} - X| > 0\right\} \subset \left\{\omega \,\middle|\, |X_{\phi(n)}(\omega) - X(\omega)| \geq \frac{1}{n} i.o.\right\}$$

since for any ω in the left side it has $|X_{\phi(n)}(\omega) - X(\omega)| = c > 0$ and so for any $n > \dfrac{1}{c}$ the inequality in right side holds, so it holds i.o..

Thus, if we define

$$A_n := \left\{|X_{\phi(n)} - X| \geq \frac{1}{n}\right\}$$

then

$$\left\{\omega \,\middle|\, |X_{\phi(n)}(\omega) - X(\omega)| \geq \frac{1}{n} i.o.\right\} = \limsup_{n\to\infty} A_n$$

by definition. Now we just find suitable subsequence $\phi(n)$ to satisfy the Borel-Cantelli condition.

So we use convergence in probability to find $\phi(n) > N = N(n)$ such that

$$\mathbb{P}\left(|X_{\phi(n)} - X| \geq \frac{1}{n}\right) \leq 2^{-n}$$

and so

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = 2 < \infty$$

which implies that $\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0$. Now by monotonicity of measure we get

$$0 \leq \mathbb{P}\left(\limsup_{n\to\infty} |X_{\phi(n)} - X| > 0\right) \leq \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0$$

which means

$$\mathbb{P}\left(\limsup_{n\to\infty} |X_{\phi(n)} - X| > 0\right) = 0$$

and we are done.

conv in prob $\Rightarrow$ conv in dist:

$\forall \varepsilon > 0$,

$$F_n(x) = \mathbb{P}(X_n \leq x) \leq \mathbb{P}(X \leq x + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)$$

because in case of $\mathbb{P}(X_n \leq x)$, either $\mathbb{P}(X \leq x + \varepsilon)$ is true or $\mathbb{P}(|X_n - X| > \varepsilon)$ is true. And by plugging back

$$F_n(x) \leq \mathbb{P}(X \leq x + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon) = F(x + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)$$

and thus

$$\limsup_{n\to\infty} F_n(x) \le F(x+\varepsilon).$$

Now we need to prove the other direction, but the idea is similar. Note that for the same reason as the above, we have

$$\mathbb{P}(X \le x-\varepsilon) \le \mathbb{P}(X_n \le x) + \mathbb{P}(|X - X_n| > 0)$$

which implies

$$F(x-\varepsilon) \le F_n(x) + \mathbb{P}(|X - X_n| > 0).$$

Taking liminf this time on both sides we have

$$F(x-\varepsilon) \le \liminf_{n\to\infty} F_n(x).$$

Then, we have

$$F(x-\varepsilon) \le \liminf_{n\to\infty} F_n(x) \le \limsup_{n\to\infty} F_n(x) \le F(x+\varepsilon)$$

which exactly at the continuity points of F, we have the good definition of convergence in distribution.

conv in dist to constant $\Rightarrow$ conv in prob:

Let's say that $X_n \xrightarrow{d} c$. Since the cdf of c is

$$F_c(x) = \begin{cases} 0 & x < c \\ 1 & x \ge c \end{cases}$$

we get that every where except c is a continuity point of F_c.

Now for contradiction we assume that $X_n \xrightarrow{p} c$ doesn't hold, that is, $\exists S \subset \Omega$ with $\mathbb{P}(S) = a > 0$ with $|X_n - c| > \delta$ for some $\delta > 0$ on S.

So on S either $X_n > c + \delta$ or $X_n < c - \delta$, which then means one of the two events has probability larger or equal than $\frac{a}{2}$.

If $\mathbb{P}(S \cap \{X_n > c+\delta\}) \ge \frac{a}{2}$ this means for cdf of X_n, denoted F_n, we have

$$F_n(c+\delta) \le 1 - \frac{a}{2}$$

uniform in n, which means $F_n(c+\delta) \to F_c(c+\delta) = 1$ is not true. Contradiction!

If the other case hold, i.e. $\mathbb{P}(S \cap \{X_n < c-\delta\}) \ge \frac{a}{2}$ we know

$$F_n(c-\delta) \ge \frac{a}{2}$$

and hence $F_n(c-\delta) \to F_c(c-\delta) = 0$ does not hold. Contradiction!

Thus we must have convergence in probability.

conv in Lp $\Rightarrow$ conv in prob:

Convergent in L^p means that

$$\lim_{n\to\infty}\left(\int |X_n - X|^p\right)^{\frac{1}{p}} = 0$$

and for contradiction let's assume that $\mathbb{P}(|X_n - X| > \varepsilon) > c > 0$ for some ε and c. Then integrating we get

$$\int |X_n - X|^p d\mathbb{P} \geq \int_{|X_n - X|>\varepsilon} |X_n - X|^p d\mathbb{P} > c \cdot \varepsilon$$

which, by taking $\frac{1}{p}$ degree we see that the limit as $n \to \infty$ is larger than $(c\cdot\varepsilon)^{\frac{1}{p}}$, so it cannot go to 0. Contradiction! Thus, we have convergence in probability.

conv in prob $\Rightarrow$ conv in L^1: (Theorem 4.6.3 in Durret book)

We assume uniform integrable here. That is

$$\mathbb{E}\left[|X_n| \cdot \mathbb{1}_{|X_n|>k}\right] < \varepsilon$$

which means

$$\int_{|X_n|>k} |X_n| d\mathbb{P} < \varepsilon.$$

We do truncation of the random variable and define

$$f(x) = \begin{cases} M & x \geq M \\ x & |x| \leq M \\ -M & x \leq -M \end{cases}$$

then we have

$$|X_n - X| \leq |X_n - f(X_n)| + |f(X_n) - f(X)| + |f(X) - X|$$

and taking the expectation (integral) we have

$$\int |X_n - X| \leq \int_{|X_n - X|<\delta} \delta d\mathbb{P} + \int_{|X_n - X|>\delta} |X_n - f(X_n)| + |f(X_n) - f(X)| + |f(X) - X| d\mathbb{P}$$

where the first term is nothing but $\delta \cdot \mathbb{P}(|X_n - X| \leq \delta) \to 0$ as $\delta \to 0$.

For the first and last term we bound by uniform integrability since

$$\int_{|X_n - X|>\delta} |X_n - f(X_n)| d\mathbb{P} \leq \int |X_n - f(X_n)| d\mathbb{P} \leq \mathbb{E}\left[|X_n| \cdot \mathbb{1}_{|X_n|>M}\right] \to 0$$

since we can find fixed M for any ε we need. Similar for the third term.

Then, for the middle term we simply bound

$$\int_{|X_n - X|>\delta} |f(X_n) - f(X)| d\mathbb{P} \leq 2M\mathbb{P}(|X_n - X| > \delta) \to 0$$

as $n \to \infty$ and hence

$$\int |X_n - X| d\mathbb{P} \leq \varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon_4 \to 0.$$

conv in $L^p \Rightarrow \exists$ subsequence conv a.s.:

conv in $L^p \Rightarrow$ conv in prob $\Rightarrow \exists$ subsequence conv a.s..

conv a.s. + uniform integrable $\Rightarrow$ conv in L^1:

conv a.s.+ uniform integrable $\Rightarrow$ conv in prob + uniform integrable $\Rightarrow$ conv in L^p.

L^1 is instance of L^p:

By definition.

$\square$

Exercise 1. *Prob 1.*

Proof.

Theorem in class says: Suppose $X_1, \dots, X_n$ are independent random variables and X_i has laws μ_i, then $(X_1, \dots, X_n)$ has the law $\mu := \mu_1 \times \mu_2 \times \cdots \times \mu_n$, and

$$\mu(A_1 \times A_2 \times \cdots \times A_n) = \prod \mu_i(A_i).$$

That is, we can very naturally use the joint law to compute the expectation of $f(X,Y) = X + Y$. Moreover, the joint measure is a product measure $\mu_{X,Y} = \mu_X \times \mu_Y$. For $Z = X + Y$ we compute

$$\begin{aligned}
\mathbb{E}[Z] = \mathbb{E}[X+Y] &= \int_{\Omega^2} (x+y) d\mu_{X,Y} \\
\text{(Fubini)} \quad &= \int_\Omega \int_\Omega (x+y) d\mu_X d\mu_Y = \int_\Omega \left(\int_\Omega (x+y) g(y) dy \right) d\mu_X \\
&= \int_\Omega \left(\int_\Omega z g(z-x) dz \right) d\mu_X = \int_\Omega \left(\int_\Omega z g(z-x) d\mu_X \right) dz \\
&= \int_\Omega \mathbb{E}_X[z g(z-X)] dz = \int_\Omega z \mathbb{E}_X[g(z-X)] dz
\end{aligned}$$

where $\mathbb{E}_X[g(z-X)]$ is thus a density since that's how we did the change of variable $z = x + y$ in the middle.

□

Exercise 2. *Prob 2.*

Proof.

(Fubini approach) Let $f(X,Y) = XY$. Then $f(X,Y)$ is measurable because it is composition of measurable and continuous function. It is integrable since we have both $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ exists and by Fubini.

For the exact same theorem as in problem 1 we can use Fubini to decompose into double integral and we thus have:

$$\begin{aligned}
\mathbb{E}[XY] &= \int_{\Omega^2} xy d\mu_{X,Y} \\
\text{(Fubini)} \quad &= \int_{\Omega}\int_{\Omega} xy d\mu_X d\mu_Y = \left(\int_{\Omega} x d\mu_X\right)\cdot\left(\int_{\Omega} y d\mu_Y\right) \\
&= \mathbb{E}[X]\mathbb{E}[Y].
\end{aligned}$$

(simple function approach) Consider two non-negative independent random variables X,Y. Then X and Y can be approximated by $\sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$ and $\sum_{j=1}^{n} b_j \mathbb{1}_{B_j}$ separately. And note that the product of simple functions can be written as

$$\sum_{i=1}^{n} a_i \mathbb{1}_{A_i} \sum_{j=1}^{n} b_j \mathbb{1}_{B_j} = \sum_{i,j} a_i b_j \mathbb{1}_{A_i \cap B_j}$$

we have

$$\mathbb{E}\left[\sum_{i,j} a_i b_j \mathbb{1}_{A_i \cap B_j}\right] = \mathbb{E}\left[\sum_{i=1}^{n} a_i \mathbb{1}_{A_i}\right]\mathbb{E}\left[\sum_{j=1}^{n} b_j \mathbb{1}_{B_j}\right].$$

Then by Monotone Convergence Theorem, we have

$$\lim_{n\to\infty} \mathbb{E}\left[\sum_{i,j}^{n} a_i b_j \mathbb{1}_{A_i \cap B_j}\right] = \mathbb{E}[XY]$$

and

$$\lim_{n\to\infty} \mathbb{E}\left[\sum_{i=1}^{n} a_i \mathbb{1}_{A_i}\right]\mathbb{E}\left[\sum_{j=1}^{n} b_j \mathbb{1}_{B_j}\right] = \mathbb{E}[X]\mathbb{E}[Y].$$

When X, Y are two independent integrable random variables, we can write

$$X = X^+ - X^-$$

and

$$Y = Y^+ - Y^-$$

where $X^+ = \max\{X, 0\}$, $X^- = \max\{X, 0\}$, $Y^+ = \max\{X, 0\}$ and $Y^- = \max\{X, 0\}$ are two non-negative random variables. Then by the deduction above, we have

$$\begin{gathered}\mathbb{E}[X^+Y^+] = \mathbb{E}[X^+]\mathbb{E}[Y^+] \\ \vdots \\ \mathbb{E}[X^-Y^-] = \mathbb{E}[X^-]\mathbb{E}[Y^-]\end{gathered}$$

And this implies that

$$\begin{aligned}\mathbb{E}[XY] &= \mathbb{E}[(X^+ - X^-)(Y^+ - Y^-)] \\ &= \mathbb{E}[X^+Y^+ + X^-Y^- - X^+Y^-X^-Y^+] \\ &= \mathbb{E}[X^+]\mathbb{E}[Y^+] + \mathbb{E}[X^-]\mathbb{E}[Y^-] - \mathbb{E}[X^+]\mathbb{E}[Y^-] - \mathbb{E}[X^-]\mathbb{E}[Y^+] \\ &= \mathbb{E}[X^+ - X^-]\mathbb{E}[Y^+ - Y^-] \\ &= \mathbb{E}[X]\mathbb{E}[Y].\end{aligned}$$

$\square$

Exercise 3. *Prob 3.*

Proof.

(a) Since $F_X(t) \to 1$ as $t \to \infty$, then when $\varepsilon > 0$ is fixed, we can always find K such that $F_X(K) \geq 1 - \varepsilon$. Without loss of generality, we can assume that K is a continuity point. Otherwise since continuity points are dense and $F_X(t)$ is increasing, we can find a continuity point K' such that $K' > K$ and $F_X(K') > 1 - \varepsilon$.

Then

$$P(|X_n| > K) = 1 - F_{X_n}(K) + F_X(K) - F_X(K) \leq |1 - F_X(K)| + |F_{X_n}(K) - F_X(K)|.$$

Note that the first term is less than or equal to ε by our assumption. And for the second term, since $F_{X_n}(K) \to F_X(K)$ as $n \to \infty$, then $\exists N \in \mathbb{N}$ such that $\forall n > N$, we have $F_{X_n}(K) - F_X(K) \leq \varepsilon$.

For others, we can choose $K_0, ..., K_{N-1}$ such that $P(|X_i| > K_i) < \varepsilon$, and this implies that $\sup_n P(|X_i| > \max\{K_0, \dots, K_{N-1}, K\}) < \varepsilon$.

(b) $\forall m$ we have

$$P(|c_n X_n| > m) \leq P\left(|X_n| < K, |c_n| > \frac{m}{K}\right) + P(|X_n| > K)$$

Choose K such that $P(|X_n| > K) < \varepsilon$ for some X_n tight. And then the first term is less than $P\left(|c_n| > \frac{m}{K}\right)$.

By MCT, we have

$$\lim_{n\to\infty} P\left(|c_n| > \frac{m}{K}\right) \leq \lim_{n\to\infty} \sum_{n=N}^{\infty} P\left(|c_n| > \frac{m}{K}\right) = P\left(\limsup_{n\to\infty} \left\{|c_n| > \frac{m}{K}\right\}\right)$$

Since $c_n \to 0$, then $P\left(\limsup_{n\to\infty} \left\{|c_n| > \frac{m}{K}\right\}\right) = 0$.

□

Exercise 4. *Prob 4*

Proof.

Since $W_n \in S^n$ is symmetric we can diagonalize with respect to unitary matrices and get

$$W_n^k = (PDP^{-1})^k = PD^kP^{-1}$$

which gives that

$$\operatorname{tr}(W_n^k) = \operatorname{tr}(D^k) = \sum_{i=1}^n \lambda_i^k.$$

Thus, it suffices to show

$$\frac{1}{n}\operatorname{tr}(W_n^k) - \mathbb{E}\left[\frac{1}{n}\operatorname{tr}(W_n^k)\right] \xrightarrow{p} 0$$

where we have

$$\mathbb{P}\left(\left|\frac{1}{n}\operatorname{tr}(W_n^k) - \mathbb{E}\left[\frac{1}{n}\operatorname{tr}(W_n^k)\right]\right| > \varepsilon\right) \overset{Chebyshev}{\leq} \frac{1}{\varepsilon^2}\left(\mathbb{E}\left[\left(\frac{1}{n}\operatorname{tr}(W_n^k)\right)^2\right] - \mathbb{E}\left[\frac{1}{n}\operatorname{tr}(W_n^k)\right]^2\right)$$

so we only need to show that for any fixed α the term

$$A := \mathbb{E}\left[\left(\frac{1}{n}\operatorname{tr}(W_n^k)\right)^2\right] - \mathbb{E}\left[\frac{1}{n}\operatorname{tr}(W_n^k)\right]^2$$

is such that $A \to 0$.

By brutal computation we get that

$$\frac{1}{n}\operatorname{tr}(W_n^k) = \frac{1}{n^{k/2+1}} \sum_{1\leq i1,\ldots,ik\leq n} X_{i1,i2}\cdots X_{ik,i1}.$$

And so we have

$$A = \frac{1}{n^{2+k}}\left(\mathbb{E}\left[S_{i,i'}\right] - \mathbb{E}[S_i]\mathbb{E}[S_{i'}]\right)$$

where

$$S_{i,i'} := \sum_{\substack{1\leq i1,\ldots,ik\leq n\\ 1\leq i1',\ldots,ik'\leq n}} X_{i1,i2}\cdots X_{ik,i1}X_{i1',i2'}\cdots X_{ik',i1'}$$

and

$$S_i := \sum_{1\leq i1,\ldots,ik\leq n} X_{i1,i2}\cdots X_{ik,i1}$$

$$S_{i'} := \sum_{1\leq i1',\ldots,ik'\leq n} X_{i1',i2'}\cdots X_{ik',i1'}.$$

To further the question, we use graph theory:

(☆)

Similarly define graph V with vertices
$V_{i,i'} = \{i_1 \cdots i_k\} \cup \{i_1', \cdots i_k'\}$

and Edges $E_{i,i'} = \{i_1 i_2, \cdots i_k i_1\} \cup \{i_1' i_2', \cdots i_k' i_1'\}$

Let $w(V)$, ~~$w(V)$~~ be the weight

Similarly, it's obvious that ① each distinctive (undirectioned) edge should be repeated twice.

with some observation

② $\{i_1 i_2, \cdots i_k, i_1\} \cap \{i_1' i_2', \cdots i_k' i_1'\} \neq \phi$.

treat $i_1 i_2 = i_2 i_1$.

therefore $E_{i,i'}$ is a walk.

If $w(V) \leq k$, there are $n(n-1) \cdots (n-t+1) \leq n^k$

in each of the equivalent class.

~~By the assignment 2.34~~

~~# standard walk satisfies ① ②~~

~~$\leq$ # standard walk satisfies ①~~

standard walks satisfies ① with $w(V)$ as we

$\leq$ ~~$w(V)$~~ $w(V)^k$ is independent of n.

therefore. when $w(v) \le k$

$(\star) \le \frac{1}{n^{2+k}} \cdot \sum_{w(v) \le k} n^k \cdot C$. $C \to$ free of n.

$= O\left(\frac{1}{n^2}\right)$

when $w(v) \ge k+2$. since condition ① is required there're at most $k+1$ unique vertices encountered by the walk.

No random walk satisfies condition ①.

if $w(v) = k+1$.

every edge is repeated exactly twice.

but to satisfy condition ②

V doesn't exist if $w(v) = k+1$
because from condition ② we know $\{i\}$ has odd steps of walk. but simultaneously it starts from i_1 and returns to i_1. each edge can be repeated at most 2 times. Therefore it requires even steps to return to i_1. $\Rightarrow$ contradiction

Therefore $(\star) = O\left(\frac{1}{n^2}\right)$

$\square$