Probability Theory I
Assignment 4

Claire He

October 4, 2022

Notations

Exercise 1

($\Rightarrow$) Let X, Y be two independent random variables. Let f, g be Borel and bounded.

- Step 1. First see that $f(X), g(Y)$ are independent if X, Y are independent. Assume the latter. For any $C_1, C_2 \in \mathcal{B}(\mathbb{R})$;

$$\begin{aligned}
P(\{f(X) \in C_1\}, \{g(Y) \in C_2\}) &= P((f \circ X)^{-1}(C_1), (g \circ Y)^{-1}(C_2)) \\
&= P(X^{-1}(f^{-1}(C_1)), Y^{-1}(g^{-1}(C_2))) \text{ set } C_1' = f^{-1}(C_1), C_2' = g^{-1}(C_2) \in \mathcal{B}(\mathbb{R}) \\
&= P(X^{-1}(C_1'), Y^{-1}(C_2')) \text{ by independence of } X, Y \\
&= P(X \in C_1')P(Y \in C_2') \\
&= P(\{f(X) \in C_1\})P(\{g(Y) \in C_2\})
\end{aligned}$$

So $f(X), g(Y)$ are independent. Since f and g are bounded, $f(X),\ g(X) \in L^1$, follows that $f(X)g(Y)$ is bounded as well and in L^1 by Hölder inequality. Now left to show the equality holds.

- Step 2. Suppose $f(X), g(Y)$ are indicators. $f(X) = \mathbb{1}_A$ and $f(Y) = \mathbb{1}_B$.

$$\begin{aligned}
E(f(X)g(Y)) &= \int_\Omega \mathbb{1}_A(\omega)\mathbb{1}_B(\omega)dP(\omega) \\
&= \int_\Omega \mathbb{1}_{A\cap B}(\omega)dP(\omega) \\
&= P(A \cap B) \text{ using independence} \\
&= P(A)P(B) \\
&= \int_\Omega \mathbb{1}_A(\omega)dP \int_\Omega \mathbb{1}_B(\omega)dP(\omega) \\
&= E(f(X))E(g(Y))
\end{aligned}$$

- Step 3. Suppose $f(X)$ and $g(Y)$ are simple: $f(X) = \sum_{i=1}^n a_i \mathbb{1}_{A_i}$ and $g(Y) = \sum_{i=1}^n b_i \mathbb{1}_{B_i}$. By linearity on the previous step, we get the same result:

$$\begin{aligned}
E(f(X)g(Y)) &= \int_\Omega \sum_{i,j} a_i b_j \mathbb{1}_{A_i}(\omega)\mathbb{1}_{B_j}(\omega)dP(\omega) \\
&= \int_\Omega \mathbb{1}_{A_i\cap B_j}(\omega)dP(\omega) \overset{\text{linearity}}{=} \sum_{i,j} a_i b_j P(A_i \cap B_j) \text{ using independence of } (A_i)_i, (B_j)_j \\
&= \sum_{i,j} a_i b_j P(A_i)P(B_j) = (\sum_i a_i P(A_i))(\sum_j b_j P(B_j)) \\
&= \int_\Omega \sum_i a_i \mathbb{1}_{A_i}(\omega)dP \int_\Omega \sum_j b_j \mathbb{1}_{B_j}(\omega)dP(\omega) \\
&= E(f(X))E(g(Y))
\end{aligned}$$

- Step 3. Now suppose that $f(X)$ and $g(Y)$ are non negative, we can find a sequence of simple functions $f_n(X)$ and $g_n(Y)$ such that $f_n(X) \uparrow f(X)$ and $g_n(Y) \uparrow g(Y)$. Note that $E(f_n(X)) \to_n E(f(X))$, and $E(g_n(Y)) \to_n E(g(Y))$ using MCT.

  From the previous statement, $E(f_n(X)g_n(Y)) = E(f_n(X))E(g_n(Y))$ for all n.

  So we have $E(f_n(X))E(g_n(Y)) \to_n E(f(X))E(g(Y))$ on the RHS.

  Since f, g bounded, we can take our simple functions bounded as well, and exists $0 < C < \infty$ such that for all n, $0 \leq f_n(X)g_n(Y) < C$. By DCT, $f(X)g(Y) \in L^1$ and $E(f_n(X)g_n(Y)) \to_n E(f(X)g(Y))$.

  So we have $E(f_n(X)g_n(Y)) \to_n E(f(X)g(Y))$ for the LHS. Piecing both sides, as required,

$$E(f(X)g(Y)) = \lim_n E(f_n(X)g_n(Y)) = \lim_n E(f_n(X))E(g_n(Y)) = E(f(X))E(g(Y))$$

- Step 4. For any $f(X), g(Y) \in L^1$, we write $f(X) = f^+(X) - f^-(X)$ and $g(Y) = g^+(Y) - g^-(Y)$ where $g^+(Y), g^-(Y), f^+(X), f^-(X)$ are all non negative. For each, we use step 3, with $f_n^+(X) \uparrow f^+(X)$, $g_n^+(Y) \uparrow g^+(Y)$, $f_n^-(X) \uparrow f^-(X)$ and $g_n^-(Y) \uparrow g^-(Y)$ sequences of simple functions that we can take bounded. Using monotone convergence:

$$\begin{aligned}
E(f_n(X)g_n(Y)) &= E((f_n^+(X) - f_n^-(X))(g_n^+(Y) - g_n^-(Y))) \\
&= E(f_n^+(X)g_n^+(Y) + f_n^-(X)g_n^-(Y) - f_n^-(X)g_n^+(Y) - f_n^+(X)g_n^-(Y)) \\
&= E(f_n^+(X)g_n^+(Y)) + E(f_n^-(X)g_n^-(Y)) - E(f_n^-(X)g_n^+(Y)) - E(f_n^+(X)g_n^-(Y))) \\
&\underset{n}{\longrightarrow} E(f^+(X)g^+(Y)) + E(f^-(X)g^-(Y)) - E(f^-(X)g^+(Y)) - E(f^+(X)g^-(Y))) \\
&\underset{n}{\longrightarrow} E(f(X)(g^+(Y) - g^-(Y))) \\
&\underset{n}{\longrightarrow} E(f(X)g(Y))
\end{aligned}$$

  On the other hand, see that using DCT like in step 3 on each $f_n^+(X)g_n^+(Y), f_n^-(X)g_n^-(Y), f_n^-(X)g_n^+(Y)$ and $f_n^+(X)g_n^-(Y))$, we have :

$$\begin{aligned}
E(f_n(X)g_n(Y)) &= E(f_n^+(X)g_n^+(Y)) + E(f_n^-(X)g_n^-(Y)) - E(f_n^-(X)g_n^+(Y)) - E(f_n^+(X)g_n^-(Y))) \\
&= E(f_n^+(X))E(g_n^+(Y)) + E(f_n^-(X))E(g_n^-(Y)) - E(f_n^-(X))E(g_n^+(Y)) - E(f_n^+(X))E(g_n^-(Y))) \\
&\underset{n}{\longrightarrow} E(f^+(X))E(g^+(Y)) + E(f^-(X))E(g^-(Y)) - E(f^-(X))E(g^+(Y)) - E(f^+(X))E(g^-(Y))) \\
&\underset{n}{\longrightarrow} E(f(X)E(g(Y))
\end{aligned}$$

  Now piecing both equality together, we get that $E(f(X)g(Y)) = E(f(X))E(g(Y))$ as required.

($\Rightarrow$) Have any $C_1, C_2 \in \mathcal{B}(\mathbb{R})$. Let's take $f = \mathbb{1}_{C_1}, g = \mathbb{1}_{C_2}$ who are bounded and Borel, then

$$\begin{aligned} & E(f(X)g(Y)) = E(\mathbb{1}_{C_1}(X)\mathbb{1}_{C_2}(Y)) = P(X \in C_1, Y \in C_2) \\ & E(f(X))E(g(Y)) = E(\mathbb{1}_{C_1}(X))E(\mathbb{1}_{C_2}(Y)) = P(X \in C_1)P(Y \in C_2) \\ \Rightarrow\quad & P(X \in C_1, Y \in C_2) = P(X \in C_1)P(Y \in C_2) \end{aligned}$$

So X, Y are independent as required.

Exercise 2

Let (M, d) be a metric space and $(x_n) \subset M$ a Cauchy sequence.

- Let's first show that (x_n) has at most one accumulation point. By contradiction, let x_0 and x_1 be two distinct accumulation points of (x_n). So we can extract a subsequence $(x_{\psi(n)})$ of (x_n) such that $x_{\psi(n)} \underset{\psi(n)\to\infty}{\longrightarrow} x_0$. Similarly, we can extract a subsequence $(x_{\phi(n)})$ of (x_n) such that $x_{\phi(n)} \underset{\phi(n)\to\infty}{\longrightarrow} x_1$.

  Let $\varepsilon > 0$, for any n, m large enough, since (x_n) is Cauchy, $d(x_{\phi(n)}, x_{\psi(m)}) < \varepsilon/2$. We can take n, m large enough so that $d(x_0, x_{\psi(n)}) < \varepsilon/4$ and $d(x_1, x_{\phi(n)}) < \varepsilon/4$.

  Note that by triangle inequality,

$$\begin{aligned} 0 \le d(x_0, x_1) &\le d(x_0, x_{\psi(n)}) + d(x_1, x_{\phi(m)}) + d(x_{\psi(n)}, x_{\phi(m)}) \\ &< \varepsilon/4 + \varepsilon/4 + \varepsilon/2 \\ 0 \le d(x_0, x_1) &< \varepsilon \end{aligned}$$

  Therefore, in limit, $d(x_0, x_1) = 0$, and we have a contradiction. So (x_n) has at most one accumulation point.

- ($\Leftarrow$) Suppose x_n doesn't converge to x by contradiction but for every subsequence of x_n, it has a further subsequence that converges to x. So we can pick N such that for any $n \ge N$, $d(x_n, x) > \epsilon$.

$$\begin{aligned} \epsilon < d(x_{\phi(n)}, x) &\le d(x_{\phi(n)}, x_n) + d(x_n, x) \\ &\le \epsilon/4 + \epsilon/4 = \epsilon/2 \end{aligned}$$

  Contradiction.

- ($\Rightarrow$) Suppose let $\epsilon > 0$ and have for N large enough, $d(x_N, x) < \epsilon/4$. Since (x_n) is Cauchy, can pick $N_1 > N$, $d(x_n, x_{\phi(n)}) < \epsilon/4$. By contradiction, assume there $(x_{\phi(n)})$ is a subsequence that doesn't converge to x. Then we can pick $n > N_1$ such that $d(x_{\phi(n)}, x) > \epsilon$. However, by triangle inequality,

Exercise 3

Statement. *Dominated Convergence Theorem for d-convergence*

Let $(X_n), X \in \mathcal{L}^0$ and $X_n \xrightarrow{P} X$ (ie converges in measure). Suppose there exists $Y \in \mathcal{L}^1$ such that $|X_n| \le Y$ a.e. for $n \ge 1$. Then $X_n, X \in \mathcal{L}^1$ and $E(X_n) \to E(X)$, $E(|X_n - X|) \to 0$.

Proof. Assuming convergence in measure, we know that since $X_n \xrightarrow{P} X$, X_n is Cauchy in measure, and by Riesz-Fischer, we can find a further subsequence $(X_{n_{k_l}})$ such that $X_{n_{k_l}} \to X$ *a.s.*.

Therefore, we can apply the normal DCT on $(X_{n_{k_l}})$ which yields: $(X_{n_{k_l}}), X \in \mathcal{L}^1$ and $E(|X_{n_{k_l}} - X|) \underset{n_{k_l}}{\to} 0$.

Note that $(L^1, ||\cdot||_1)$ is a Banach space and the norm induces a distance hence a metric space. We also know by exercise 2. that for a metric space, it is equivalent to show for every subsequence of X_n has a further subsequence $X_{n_{k_l}}$ which converges to X and that $X_n \to X \in L^1$.

Exercise 4

Let $X \in \mathcal{L}^\infty$ on $(\Omega, \mathcal{F}, P)$ probability space. Let's show that $||X||_\infty = \lim_{p\to\infty} ||X||_p$. First notice that $X \in L^p$ by corollary 3.20.

Note that $||X||_\infty = \inf\{c \in \mathbb{R} : |X| < c \ a.s.\}$. Can find $||X||_\infty > \epsilon > 0$, such that $|X| > ||X||_\infty - \epsilon$. By definition of $||X||_\infty$, we also know that $a := P(|X| > ||X||_\infty - \epsilon) \in (0, 1]$. In particular:

$$\begin{aligned}
\int_\Omega |X|^p dP &\geq \int_\Omega (||X||_\infty - \epsilon)^p dP \geq \int_{\{|X|>||X||_\infty - \epsilon\}} (||X||_\infty - \epsilon)^p dP \\
\Rightarrow ||X||_p^p &\geq (||X||_\infty - \epsilon)^p P(|X| > ||X||_\infty - \epsilon) \\
\Rightarrow ||X||_p &\geq (||X||_\infty - \epsilon) a^{\frac{1}{p}} \underset{p\to\infty}{\longrightarrow} ||X||_\infty - \epsilon
\end{aligned}$$

Now, for the reverse inequality, use Jensen like in corollary 3.20 and see that $||X||_\infty \geq ||X||_p$ for any $p > 0$. So $\limsup_p ||X||_p \leq ||X||_\infty$. Piecing this together as the following holds for any $\epsilon \in (0, ||X||_\infty)$,

$$||X||_\infty - \epsilon \leq \liminf_p ||X||_p \leq \limsup_p ||X||_p \leq ||X||_\infty$$

So $\lim_p ||X||_p = \liminf_p ||X||_p = \limsup_p ||X||_p = ||X||_\infty$ as required.

Exercise 5

Let's define $d(X, Y) := E(|X - Y| \wedge 1)$ for $X, Y \in L^0$.

- Let's first show that d defines a metric on L^0.

  Indeed, for $X, Y, Z \in L^0$, it is a symmetric function: $d(X, Y) = E(|X - Y| \wedge 1) = E(|Y - X| \wedge 1) = d(Y, X)$, definite: $d(X, Y) = 0 \Rightarrow E(|X - Y| \wedge 1) = 0 \Rightarrow |X - Y| \wedge 1 = 0 \ a.e. \Rightarrow X = Y \ a.e.$ and we have a triangle inequality property: $d(X, Z) = E(|X - Z| \wedge 1) \leq E(|X - Y| \wedge 1) + E(|Z - Y| \wedge 1) = d(X, Y) + d(Y, Z)$.

  So (L^0, d) defines a metric space.

- Note that the Riesz-Fischer theorem for d-convergence ensures that (L^0, d) is a complete space as for any $(X_n) \in L^0$ d-Cauchy, there exists a random variable X such that $X_n \xrightarrow{P} X$ ie (X_n) d-converges.

- Now let's show that d-convergence is equivalent to convergence in measure. We say that $X_n \xrightarrow{P} X$ ie converges in measure if $\forall \varepsilon > 0$, $P(|X_n - X| > \varepsilon) \underset{n}{\to} 0$.

  $(\Rightarrow)$ Assume $d(X_n, X) \underset{n}{\to} 0$, by Markov we have

  $$\begin{aligned}
  P(|X_n - X| \wedge 1 > \varepsilon) &\leq \frac{E(|X_n - X| \wedge 1)}{\varepsilon} \\
  &\leq \frac{d(X_n, X)}{\varepsilon} \underset{n}{\to} 0
  \end{aligned}$$

  $(\Leftarrow)$ Now for the converse, assume $X_n \xrightarrow{P} X$. First we claim it is Cauchy in measure. For any $\varepsilon > 0$.

  $$P(|X_n - X_m| > \varepsilon) \leq P(|X_n - X| > \varepsilon/2) + P(|X_m - X| > \varepsilon/2) \underset{n,m}{\to} 0$$

  Since it is Cauchy in measure, by Riesz Fischer theorem, there exists a subsequence (X_{n_k}) that converges almost surely to X. Note that $\forall n_k > 1$, $|X_{n_k} - X| \wedge 1 \leq 1 \ a.e.$ therefore by DCT, $(X_{n_k} - X) \wedge 1 \in L^1$ and $E(|X_{n_k} - X| \wedge 1) \to 0$ ie $d(X_{n_k}, X) \to 0$ and (L^0, d) is a metric space with (X_n) Cauchy so as we have shown in Exercise 2, X_n can only have one accumulation point and for X_{n_k} our subsequence, $d(X_{n_k}, X) \to 0$ implies that $d(X_n, X) \to 0$.

  Note that this last argument shows that (L^0, d) is complete as we have shown that a sequence which is Cauchy for d is d-convergent.