MATH 507A, GRADUATE PROBABILITY
SELECTED HOMEWORK SOLUTIONS

STEVEN HEILMAN

Contents

1. Homework 1

Exercise 1.3. Let $n \geq 1$. Show that the Borel σ-algebra on $\mathbb{R}^n$ is generated by sets of the form $A_1 \times \cdots \times A_n$ where $A_i \subseteq \mathbb{R}$ is a Borel set for every $1 \leq i \leq n$.

Solution. For any $1 \leq i \leq n+1$, let $\mathcal{B}_i$ denote the σ-algebra generated by sets of the form

$$(a_1, b_1) \times \cdots \times (a_{i-1}, b_{i-1}) \times A_i \times \cdots \times A_n,$$

where for any $1 \leq j \leq i-1$, $a_j, b_j \in \mathbb{R}$, $a_j < b_j$, and for any $i \leq j \leq n$, $A_j \in \mathcal{B}(\mathbb{R})$.

Claim 1. $\mathcal{B}_1 = \mathcal{B}_{n+1}$.

We will prove this claim by showing that $\mathcal{B}_i = \mathcal{B}_{i+1}$ for every $1 \leq i \leq n$. Since open intervals are Borel sets in $\mathbb{R}$, we immediately have $\mathcal{B}_i \supseteq \mathcal{B}_{i+1}$ for every $1 \leq i \leq n$. So, it remains to show the reverse containment.

Fix $1 \leq i \leq n$. Fix $a_1, b_1, \ldots, a_{i-1}, b_{i-1} \in \mathbb{R}$ with $a_j < b_j$ for all $1 \leq j \leq i-1$, and fix $A_j \in \mathcal{B}(\mathbb{R})$ for all $i+1 \leq j \leq n$. Let $I \subseteq \mathbb{R}$ be an open interval. By the definition of $\mathcal{B}_{i+1}$, note that

$$(a_1, b_1) \times \cdots \times (a_{i-1}, b_{i-1}) \times I \times A_{i+1} \times \cdots \times A_n \in \mathcal{B}_{i+1}. \qquad (*)$$

Let $U \subseteq \mathbb{R}$ be any open set. From the Structure Theorem for open sets of $\mathbb{R}$, there exists a (disjoint) countable collection of open intervals $I_1, I_2, \ldots$ such that $U = \cup_{m=1}^{\infty} I_m$. Since $\mathcal{B}_{i+1}$ is a σ-algebra, it is closed under countable unions. We therefore conclude from $(*)$ that

$$(a_1, b_1) \times \cdots \times (a_{i-1}, b_{i-1}) \times U \times A_{i+1} \times \cdots \times A_n \in \mathcal{B}_{i+1}. \qquad (**)$$

Date: November 13, 2020

With all other parameters here fixed other than U, and considering $U \subseteq \mathbb{R}$ to be any open set, we see that the smallest σ-algebra containing all sets of the form given in $(**)$ must contain all sets of the form

$$(a_1, b_1) \times \cdots \times (a_{i-1}, b_{i-1}) \times A_i \times A_{i+1} \times \cdots \times A_n, \qquad (***)$$

where A_i is any Borel subset of $\mathbb{R}$ (since $\mathcal{B}(\mathbb{R})$ is the σ-algebra generated by open subsets of $\mathbb{R}$). Since $\mathcal{B}_{i+1}$ is a σ-algebra containing all sets of the form $(***)$, we conclude that $\mathcal{B}_{i+1} \supseteq \mathcal{B}_i$, by the definition of $\mathcal{B}_i$. We deduce that Claim 1 holds.

Combining Claim 1 with the following Claim 2 concludes the proof. So it remains to prove Claim 2.

Claim 2. $\mathcal{B}_{n+1} = \mathcal{B}(\mathbb{R}^n)$. We will prove Claim 2 using Claim 3 below. Assuming Claim 3, let us prove Claim 2.

From Claim 3, any open $U \subseteq \mathbb{R}^n$ is a countable union of sets that generate $\mathcal{B}_{n+1}$. Since $\mathcal{B}_{n+1}$ is a σ-algebra, it is closed under countable unions, so $\mathcal{B}_{n+1}$ must contain all open sets in $\mathbb{R}^n$. Since $\mathcal{B}(\mathbb{R}^n)$ is the smallest σ-algebra containing all open sets in $\mathbb{R}^n$, we conclude that $\mathcal{B}(\mathbb{R}^n) \subseteq \mathcal{B}_{n+1}$. On the other hand, since open boxes are open, and since $\mathcal{B}_{n+1}$ is generated by open boxes, we immediately have $\mathcal{B}_{n+1} \subseteq \mathcal{B}(\mathbb{R}^n)$, since the latter is generated by all open sets (including open boxes in particular). Claim 2 is therefore proven, assuming Claim 3.

Claim 3. Any open $U \subseteq \mathbb{R}^n$ can be written as a countable union of open boxes of the form

$$(a_1, b_1) \times \cdots \times (a_n, b_n),$$

where for any $1 \leq j \leq n$, $a_j, b_j \in \mathbb{R}$ and $a_j < b_j$.

Below, we denote $\|x\| := (x_1^2 + \cdots + x_n^2)^{1/2}$ for any $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, and for any $r > 0$, we denote the open ball centered at x of radius r as $B(x, r) := \{(y \in \mathbb{R}^n \colon \|x - y\| < r\}$ and $B_\infty(x, r) := \{(y \in \mathbb{R}^n \colon \max_{1 \leq i \leq n} |x_i - y_i| < r\}$ as the open box centered at x with ℓ_∞ radius r.

Let $U \subseteq \mathbb{R}^n$ be any open set. By the definition of the standard topology of $\mathbb{R}^n$, U can be written as a (possibly uncountable) union of open balls. By the Structure Theorem for open sets, U can be written as a finite or countable union of open balls. That is, $U = \cup_{j=1}^\infty B_j$ for some open balls $B_j \subseteq \mathbb{R}^n$ of the form $B_j = B(x^{(j)}, r_j)$ for some $0 < r_j < \infty$ and for some $x^{(j)} \in \mathbb{R}^n$ where $j \geq 1$ (and we can take some B_j to be empty). Fix $j \geq 1$. We claim that B_j can be written as a countable union of open boxes. Since $U = \cup_{j=1}^\infty B_j$, this statement implies the claim.

Let $C_j \subseteq B_j$ be a countable dense set of points of B_j. For any $c \in C_j$, since $S_j := \{(y \in \mathbb{R}^n \colon \left\|y - x^{(j)}\right\| = r_j\}$ is a compact set (being a closed and bounded subset of Euclidean space), there exists some $\varepsilon_c > 0$ such that $\varepsilon_c := \inf_{y \in S_j} \|c - y\| > 0$. Similarly, if $b \in B_j$, then $\delta_b := \inf_{y \in S_j} \|b - y\| > 0$. If $b \in B_j$, since C_j is dense in B_j, there exists $c \in C_j$ such that $\|b - c\| < \delta_b/(4n)$. We claim now that

$$b \in B_\infty(c, \varepsilon_c/(2n)) \subseteq B_j. \qquad (\ddagger)$$

Since $b \in B_j$ is arbitrary, this containment implies that $B_j = \cup_{c \in C_j} B_\infty(c, \varepsilon_c/(2n))$, thereby completing the claim, since then U can be written as a countable union of open boxes. So, it remains to prove $(\ddagger)$. Since $\|b - c\| < \delta_b/(4n)$, we have $|b_i - c_i| \leq \|b - c\| < \delta_b/(4n)$ for all $1 \leq i \leq n$, implying that $b \in B_\infty(c, \delta_b/(4n))$. We now claim that $\delta_b/2 \leq \varepsilon_c$, thereby proving the first containment of $(\ddagger)$. This claim follows since, if $d \in S_j$, then $\|b - d\| \leq$

$\|d-c\| + \|c-b\|$ from the triangle inequality. So, taking the infimum of both sides over $d \in S_j$, we get $\delta_b \leq \varepsilon_c + \delta_b/4$, so that $\delta_b(3/4) \leq \varepsilon_c$. The first containment of (‡) is proven, since $\delta_b/2 \leq \varepsilon_c$ implies that

$$b \in B_\infty(c, \delta_b/(4n)) \subseteq B_\infty(c, \varepsilon_c/(2n)).$$

To prove the last containment of (‡), note that if $y \in B_\infty(c, \varepsilon_c/(2n))$, then $|y_i - c_i| \leq \varepsilon_c/(2n)$ for all $1 \leq i \leq n$, so that $\|y-c\| \leq \varepsilon_c/2$, so by the triangle inequality,

$$\|x^{(j)} - y\| \leq \|x^{(j)} - c\| + \|c-y\| = r_j - \varepsilon_c + \|c-y\| < r_j.$$

That is, $y \in B_j$, as desired.

□

Exercise 1.5. Let μ be a measure on a measurable space $(\Omega, \mathcal{F})$. Using the axioms for a measure, show:

- (Monotonicity) If $A \subseteq B$ are measurable, then $\mu(A) \leq \mu(B)$.
- (Subadditivity) If $A_1, A_2, \ldots$ are measurable (but not necessarily disjoint), then

$$\mu(\cup_{n=1}^\infty A_n) \leq \sum_{n=1}^\infty \mu(A_n).$$

- (Continuity from below) If $A_1 \subseteq A_2 \subseteq \cdots$ are measurable, then $\mu(\cup_{n=1}^\infty A_n) = \lim_{n\to\infty} \mu(A_n)$.
- (Continuity from above) If $A_1 \supseteq A_2 \supseteq \cdots$ are measurable and if $\mu(A_1) < \infty$, then $\mu(\cap_{n=1}^\infty A_n) = \lim_{n\to\infty} \mu(A_n)$. Then, find a measurable space $(\Omega, \mathcal{F})$ and measurable subsets $B_1 \supseteq B_2 \supseteq \cdots$ such that $\mu(\cap_{n=1}^\infty B_n) \neq \lim_{n\to\infty} \mu(B_n)$.

Exercise 1.8. Let μ be a probability measure on $\mathbb{R}^n$, where $\mathbb{R}^n$ has the Borel σ-algebra. Define the **distribution function** $F\colon \mathbb{R}^n \to [0,1]$ associated to μ by

$$\begin{aligned} F(t_1, \ldots, t_n) &:= \mu((-\infty, t_1] \times \cdots \times (-\infty, t_n]) \\ &= \mu(\{(x_1, \ldots, x_n) \in \mathbb{R}^n\colon\ -\infty < x_i \leq t_i, \forall\, 1 \leq i \leq n\}), \qquad \forall\, t_1, \ldots, t_n \in \mathbb{R}. \end{aligned}$$

Show the following properties of F:

- F is nondecreasing. ($F(t_1, \ldots, t_n) \leq F(t_1', \ldots, t_n')$ whenever $t_i \leq t_i'\ \forall\ 1 \leq i \leq n$.)
- $\lim_{t_1,\ldots,t_n \to -\infty} F(t_1, \ldots, t_n) = 0$ and $\lim_{t_1,\ldots,t_n\to\infty} F(t_1, \ldots, t_n) = 1$.
- F is right continuous, i.e. $F(t_1, \ldots, t_n) = \lim_{(s_1,\ldots,s_n)\to(t_1,\ldots,t_n)^+} F(s_1, \ldots, t_n)$ for all $t_1, \ldots, t_n \in \mathbb{R}$, where the limit restricts that $s_i \geq t_i\ \forall\ 1 \leq i \leq n$.
- If $t_{i,0} \leq t_{i,1}\ \forall\ 1 \leq i \leq n$, then

$$\sum_{(\omega_1,\ldots,\omega_n)\in\{0,1\}^n} (-1)^{\omega_1+\cdots+\omega_n+n} F(t_{1,\omega_1}, \ldots, t_{n,\omega_n}) \geq 0.$$

Solution. We prove the last property only. Fix real numbers $t_{i,0} \leq t_{i,1}\ \forall\ 1 \leq i \leq n$. Define

$$A := \{(x_1, \ldots, x_n) \in \mathbb{R}^n\colon x_i \leq t_{i,1}, \qquad \forall\, 1 \leq i \leq n\}.$$

And for any fixed $1 \leq i \leq n$, define

$$A_i := A \cap \{(x_1, \ldots, x_n) \in \mathbb{R}^n\colon x_i \leq t_{i,0}\}.$$

From the Inclusion-Exclusion Formula, we have

$$\mu(\cup_{i=1}^n A_i) = \sum_{k=1}^n \sum_{S\subseteq\{1,\ldots,n\}:\ |S|=k} (-1)^{k+1}\mu(\cap_{i\in S} A_i). \qquad (*)$$

For any fixed $S \subseteq \{1,\ldots,n\}$, if we define $(\omega_1,\ldots,\omega_n) \in \{0,1\}^n$ so that $\omega_i = 0$ if and only if $i \in S$, then by definition of F and $A_1,\ldots,A_n$, we have

$$\mu(\cap_{i\in S} A_i) = F(t_{1,\omega_1},\ldots,t_{n,\omega_n}).$$

In fact, we have just described a bijection between subsets S of $\{1,\ldots,n\}$ and vectors in $\{0,1\}^n$, where if S has k elements, then $\omega_1+\cdots+\omega_n = n-k$. Combining these with $(*)$,

$$\begin{aligned}\mu(\cup_{i=1}^n A_i) &= \sum_{k=1}^n \sum_{\substack{(\omega_1,\ldots,\omega_n)\in\{0,1\}^n:\\ \omega_1+\cdots+\omega_n=n-k}} (-1)^{k+1}F(t_{1,\omega_1},\ldots,t_{n,\omega_n}) \\ &= F(t_{1,1},\ldots,t_{n,1}) + \sum_{(\omega_1,\ldots,\omega_n)\in\{0,1\}^n} (-1)^{n-\omega_1-\cdots-\omega_n+1}F(t_{1,\omega_1},\ldots,t_{n,\omega_n}).\end{aligned}$$

Finally, we note that

$$F(t_{1,1},\ldots,t_{n,1}) - \mu(\cup_{i=1}^n A_i) = \mu(A) - \mu(\cup_{i=1}^n A_i) = \mu(A\smallsetminus(\cup_{i=1}^n A_i)) = \mu(\prod_{i=1}^n (t_{i,0},t_{i,1})) \geq 0.$$

□

2. Homework 2

Exercise 2.1. Let $(\Omega,\mathcal{F},\mathbf{P})$ be a finite or countable probability space. If $X\colon \Omega\to[0,\infty]$ is a random variable with $\mathbf{E}\,|X|<\infty$, show that

$$\mathbf{E}X = \sum_{\omega\in\Omega} X(\omega)\mathbf{P}(\omega).$$

Solution. The case that Ω is finite follows from the definition of $\mathbf{E}X$, since every function on Ω is then a simple function. Consider the case that Ω is countable. Without loss of generality, we assume $\Omega = \mathbb{N} = \{1,2,3,\ldots\}$. For any $n \geq 1$, define $X_n(t) := X(t)1_{t\leq n}$, for any $t\in\Omega$. Then X_n increases monotonically to X as $n\to\infty$, so by the Monotone Convergence Theorem and the case of finite Ω,

$$\sum_{t=1}^\infty X(t)\mathbf{P}(t) = \lim_{n\to\infty}\sum_{t=1}^n X(t)\mathbf{P}(t) = \lim_{n\to\infty}\mathbf{E}X_n = \mathbf{E}\lim_{n\to\infty}X_n = \mathbf{E}X.$$

Alternatively, the exercise follows by applying Fubini's Theorem to the random variable $Y\colon \Omega\times\mathbb{N}\to[0,\infty]$ defined by $Y(\omega,n) := X(\omega)1_{\omega=n}$ with the measure $\mathbf{P}\times\mu$ where $\mu(n) := 1$ for all $n\geq 1$. □

Exercise 2.2. Let X,Y be random variables such that $X,Y\geq 0$ or $\mathbf{E}\,|X|\,,\mathbf{E}\,|Y|<\infty$. For any $a\in[-\infty,\infty]$, define $0\cdot a := 0$. Show:

- $\mathbf{E}(X+Y) = \mathbf{E}X+\mathbf{E}Y$ and if $c\in\mathbb{R}$, then $\mathbf{E}(cX) = c\mathbf{E}X$.
- If $\mathbf{P}(X=Y)=1$, then $\mathbf{E}X=\mathbf{E}Y$.
- $\mathbf{E}\,|X|\geq 0$ with equality only when $X=0$ almost surely.

- If $X \leq Y$ almost surely, then $\mathbf{E}X \leq \mathbf{E}Y$.

Solution. We prove the third and fourth items only. We consider the third item. Since $|X| \geq 0$, we conclude that $\mathbf{E}\,|X| \geq 0$, since $\mathbf{E}\,|X|$ is the supremum of expected values of simple functions that have nonnegative expected value.

Now, suppose $\mathbf{E}\,|X| = 0$. Suppose for the sake of contradiction that $\mathbf{P}(|X| > 0) > 0$. Note that $\cup_{n=1}^{\infty}\{|X| > 1/n\} = \{|X| > 0\}$, and the events $\{|X| > 1/n\}$ are increasing as n increases, so $\mathbf{P}(|X| > 0) = \lim_{n\to\infty}\mathbf{P}(|X| > 1/n)$, by Continuity of the probability law $\mathbf{P}$. Since $\mathbf{P}(|X| > 0) > 0$, there exists $n \geq 1$ such that $\mathbf{P}(|X| > 1/n) > 0$. But then $\frac{1}{n}1_{|X|>1/n}$ is a simple function such that $\frac{1}{n}1_{|X|>1/n} \leq |X|$. So, by the definition of $\mathbf{E}\,|X|$, we have

$$\mathbf{E}\,|X| \geq \frac{1}{n}\mathbf{E}1_{|X|>1/n} = \frac{1}{n}\mathbf{P}(|X| > 1/n) > 0.$$

Since we have found a contradiction, we conclude that $\mathbf{P}(|X| = 0) = \mathbf{P}(X = 0) = 1$.

For the fourth item in the case $X, Y \geq 0$, note that if W is an unsigned simple function such that $W \leq X$, then $W1_{X\leq Y}$ is also a simple function such that $W \leq Y$. And $\mathbf{E}W1_{X\leq Y} = \mathbf{E}W$ by the second item. So, by the definition of expected value, we must have $\mathbf{E}X \leq \mathbf{E}Y$.

For the fourth item in the remaining case $\mathbf{E}\,|X|\,, \mathbf{E}\,|Y| < \infty$, we have $\mathbf{E}\max(X, 0) \leq \mathbf{E}\max(Y, 0)$ and $-\mathbf{E}\max(-X, 0) \leq -\mathbf{E}\max(-Y, 0)$, from the case already proven, so

$$\mathbf{E}X = \mathbf{E}\max(X, 0) - \mathbf{E}\max(-X, 0) \leq \mathbf{E}\max(Y, 0) - \mathbf{E}\max(-Y, 0) \leq \mathbf{E}Y.$$

□

Exercise 2.4. Let $\phi\colon \mathbb{R} \to \mathbb{R}$. We say that ϕ is **convex** if, for any $x, y \in \mathbb{R}$ and for any $t \in [0, 1]$, we have

$$\phi(tx + (1 - t)y) \leq t\phi(x) + (1 - t)\phi(y).$$

Let $\phi\colon \mathbb{R} \to \mathbb{R}$. Show that ϕ is convex if and only if: for any $y \in \mathbb{R}$, there exists a constant a and there exists a function $L\colon \mathbb{R} \to \mathbb{R}$ defined by $L(x) = a(x - y) + \phi(y)$, $x \in \mathbb{R}$, such that $L(y) = \phi(y)$ and such that $L(x) \leq \phi(x)$ for all $x \in \mathbb{R}$. (In the case that ϕ is differentiable, the latter condition says that ϕ lies above all of its tangent lines.)

(Hint: Suppose ϕ is convex. If x is fixed and y varies, show that $\frac{\phi(y)-\phi(x)}{y-x}$ increases as y increases. Draw a picture. What slope a should L have at x?)

Solution. Assume ϕ is convex.

Now, fix $x \in \mathbb{R}$ and let $b < x < c$. Let $M_R := \{\frac{\phi(c)-\phi(x)}{c-x}\colon c > x\}$, $M_L := \{\frac{\phi(x)-\phi(b)}{x-b}\colon b < x\}$ be the slopes of the secant lines through ϕ using points to the right and left of x, respectively. We claim that for any $m \in M_R, p \in M_L$, we have $m \geq p$. Let $m \in M_R, p \in M_L$. By definition, there exist $b < x < c$ such that $m = \frac{\phi(c)-\phi(x)}{c-x}$ and $p = \frac{\phi(x)-\phi(b)}{x-b}$. Let $t \in (0, 1)$ such that $tb+(1-t)c = x$. Since ϕ is convex, $t\phi(b)+(1-t)\phi(c) \geq \phi(x)$. The following list of statements

are all equivalent:

$$\begin{aligned} m \geq p &\iff \frac{\phi(c)-\phi(x)}{c-x} \geq \frac{\phi(x)-\phi(b)}{x-b} \\ &\iff (x-b)(\phi(c)-\phi(x)) \geq (c-x)(\phi(x)-\phi(b)) \\ &\iff (x-b)\phi(c)+b\phi(x) \geq c\phi(x)-(c-x)\phi(b) \\ &\iff (1-t)(c-b)\phi(c)+b\phi(x) \geq c\phi(x)-t(c-b)\phi(b) \\ &\qquad\qquad \text{, using } (x-b)=(1-t)(c-b), \quad c-x=t(c-b) \\ &\iff (1-t)(c-b)\phi(c) \geq (c-b)\phi(x)-t(c-b)\phi(b) \\ &\iff (1-t)\phi(c)+t\phi(b) \geq \phi(x) \end{aligned}$$

We verified that the last line holds. We conclude that $m \geq p$. So, if $x \in \mathbb{R}$ is fixed, we can choose some $a \in \mathbb{R}$ such that $p \leq a \leq m$ for all $p \in M_L, m \in M_R$. Consider the function $L(z) = a(z-x)+\phi(x)$, $z \in \mathbb{R}$. Then $L(x) = \phi(x)$. We claim that $L(z) \leq \phi(z)$ for all $z \in \mathbb{R}$. We argue by contradiction. Suppose there is some $z \in \mathbb{R}$ with $L(z) > \phi(z)$. Without loss of generality, $z > x$. By definition of L, $a(z-x)+\phi(x) > \phi(z)$, so $a > \frac{\phi(z)-\phi(x)}{z-x}$. But $z > x$, so $\frac{\phi(z)-\phi(x)}{z-x} \in M_R$, and by choice of a, we must have $a \leq \frac{\phi(z)-\phi(x)}{z-x}$, a contradiction. Having found a contradiction, we conclude that $L(z) \leq \phi(z)$, as desired.

Now, assume: for any $y \in \mathbb{R}$, there exists a constant a and there exists a function $L\colon \mathbb{R} \to \mathbb{R}$ defined by $L(x) = a(x-y)+\phi(y)$, $x \in \mathbb{R}$, such that $L(y) = \phi(y)$ and such that $L(x) \leq \phi(x)$ for all $x \in \mathbb{R}$. We now show that ϕ is convex.

Fix $b, c \in \mathbb{R}$. Let $t \in (0,1)$. Set $y := tb+(1-t)c$. By assumption, there is a function $L(x) = a(x-y)+\phi(y)$ such that $L(x) \leq \phi(x)$ for all $x \in \mathbb{R}$. In particular,

$$a(b-y)+\phi(y) \leq \phi(b), \qquad a(c-y)+\phi(y) \leq \phi(c).$$

Multiplying by $t > 0$ and $(1-t) > 0$ respectively,

$$ta(b-y)+t\phi(y) \leq t\phi(b), \qquad (1-t)a(c-y)+(1-t)\phi(y) \leq (1-t)\phi(c)$$

Note that $t(b-y)+(1-t)(c-y) = tb+(1-t)c-y = 0$ by definition of y. So, adding the inequalities,

$$t\phi(b)+(1-t)\phi(c) \geq \phi(y)+a[t(b-y)+(1-t)(c-y)] = \phi(y) = \phi(tb+(1-t)c).$$

□

3. Homework 3

Exercise 3.1 (**Stein Identity**)**.** Let X be a standard Gaussian random variable, so that X has density $x \mapsto e^{-x^2/2}/\sqrt{2\pi}$, $\forall\ x \in \mathbb{R}$. Let $g\colon \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function such that g and g' have polynomial volume growth. That is, $\exists\ a, b > 0$ such that $|g(x)|, |g'(x)| \leq a(1+|x|)^b$, $\forall\ x \in \mathbb{R}$. Prove the **Stein identity**

$$\mathbf{E}Xg(X) = \mathbf{E}g'(X).$$

Using this identity, recursively compute $\mathbf{E}X^k$ for any positive integer k.

Alternatively, for any $t > 0$, show that $\mathbf{E}e^{tX} = e^{t^2/2}$, i.e. compute the **moment generating function** of X. Then, using $\frac{d^k}{dt^k}|_{t=0}\mathbf{E}e^{tX} = \mathbf{E}X^k$ and using the power series expansion of the exponential, compute $\mathbf{E}X^k$ directly from the identity $\mathbf{E}e^{tX} = e^{t^2/2}$.

Solution. From the "Change of Variables" formula and integration by parts

$$\begin{aligned}\mathbf{E}Xg(X) &= \int_{-\infty}^{\infty} xg(x)e^{-x^2/2}\frac{dx}{\sqrt{2\pi}} = -\int_{-\infty}^{\infty} g(x)\frac{d}{dx}[e^{-x^2/2}]\frac{dx}{\sqrt{2\pi}} \\ &= \int_{-\infty}^{\infty} \frac{d}{dx}g(x)e^{-x^2/2}\frac{dx}{\sqrt{2\pi}} = \mathbf{E}g'(X).\end{aligned}$$

The fact that no boundary terms arise at $\pm\infty$ follows since g, g' have polynomial volume growth.

For any odd positive integer k, we have $\mathbf{E}X^k = \int_{-\infty}^{\infty} x^k e^{-x^2/2}\frac{dx}{\sqrt{2\pi}} = 0$ since the integrand is odd. When k is an even positive integer, we apply the Stein identity to get

$$\mathbf{E}X^k = \mathbf{E}XX^{k-1} = (k-1)\mathbf{E}X^{k-2} = (k-1)(k-3)\mathbf{E}X^{k-4} = \cdots = (k-1)!!.$$

Alternatively, for any $t \in \mathbb{R}$, we complete the square to get

$$\begin{aligned}\mathbf{E}e^{tX} &= \int_{-\infty}^{\infty} e^{tx}e^{-x^2/2}\frac{dx}{\sqrt{2\pi}} \\ &= \int_{-\infty}^{\infty} e^{t^2/2}e^{-(x-t)^2/2}\frac{dx}{\sqrt{2\pi}} = e^{t^2/2}\int_{-\infty}^{\infty} e^{-x^2/2}\frac{dx}{\sqrt{2\pi}} = e^{t^2/2}.\end{aligned}$$

We now claim that the derivative of the function on the left exists in $t \in \mathbb{R}$. Fix $t \in \mathbb{R}$ and $h > 0$ and observe

$$h^{-1}[\mathbf{E}e^{(t+h)X} - e^{tX}] = h^{-1}\int_{-\infty}^{\infty}[e^{(t+h)x} - e^{tx}]e^{-x^2/2}\frac{dx}{\sqrt{2\pi}} = \int_{-\infty}^{\infty} e^{tx}\Big(\frac{e^{hx}-1}{h}\Big)e^{-x^2/2}\frac{dx}{\sqrt{2\pi}}.$$

By the Dominated Convergence Theorem, the limit exists as $h \to 0$ and

$$\frac{d}{dt}\mathbf{E}e^{tX} = \int_{-\infty}^{\infty} e^{tx}xe^{-x^2/2}\frac{dx}{\sqrt{2\pi}}.$$

(By the Mean Value Theorem, $|\frac{e^{hx}-1}{h}| = \left|x_0e^{hx_0}\right| \le |x|\max(e^{hx}, 1)$ for some $x_0 \in [0, x]$, so the Dominated Convergence Theorem applies.) In particular, setting $t = 0$ we get

$$\frac{d}{dt}|_{t=0}\mathbf{E}e^{tX} = \int_{-\infty}^{\infty} xe^{-x^2/2}\frac{dx}{\sqrt{2\pi}} = \mathbf{E}X.$$

Repeating this argument, we see that $\mathbf{E}e^{tX}$ is infinitely differentiable in t and for any $k > 0$,

$$\frac{d^k}{dt^k}|_{t=0}\mathbf{E}e^{tX} = \int_{-\infty}^{\infty} x^k e^{-x^2/2}\frac{dx}{\sqrt{2\pi}} = \mathbf{E}X^k.$$

Meanwhile, using the power series expansion of $e^{t^2/2}$, we get

$$\mathbf{E}e^{tX} = e^{t^2/2} = \sum_{k=0}^{\infty}\frac{(t^2/2)^k}{k!}.$$

Equating the k^{th} derivatives at zero of both sides gives

$$\mathbf{E}X^{2k} = \frac{d^{2k}}{dt^{2k}}|_{t=0}\frac{(t^2/2)^k}{k!} = \frac{(2k)!}{2^k k!}.$$

□

Exercise 3.2 (**Finite Product Measure**). Let $(\Omega_i, \mathcal{F}_i, \mu_i)$ be probability spaces for any $1 \leq i \leq n$. Show that there exists a unique probability measure, denoted $\prod_{i=1}^n \mu_i$ on $(\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \mathcal{F}_i)$ (where the latter measurable space is defined in the notes) such that

$$\Big(\prod_{i=1}^n \mu_i\Big)\Big(\prod_{i=1}^n A_i\Big) = \prod_{i=1}^n \mu_i(A_i), \qquad \forall\, A_1 \in \mathcal{F}_1, \ldots, A_n \in \mathcal{F}_n.$$

Solution. [Solution 1] By induction, it suffices to prove the case $n = 2$. Let $\mathcal{A}$ be the set of measurable subsets E of $\mathcal{F}_1 \times \mathcal{F}_2$ such that the function $f_E(x) := \mu_2\{y \in \Omega_2 \colon (x, y) \in E\}$ is measurable, where $x \in \Omega_1$. Then $\mathcal{A}$ is closed under complements and finite disjoint unions. Moreover $\mathcal{A}$ is closed under countable disjoint unions since if $E = \cup_{i=1}^\infty E_i$ disjointly, then $f_E = \sum_{i=1}^\infty f_{E_i}$, and the latter sum is an infinite sum of measurable functions, so f_E is measurable. Therefore, $\mathcal{A}$ is closed under countable increasing unions. By the Monotone Class Lemma, $\mathcal{A}$ is therefore equal to $\mathcal{F}_1 \times \mathcal{F}_2$. For any $E \in \mathcal{F}_1 \times \mathcal{F}_2$, define

$$\mu(E) := \int_{\Omega_1} \mu_2\{y \in \Omega_2 \colon (x, y) \in E\} d\mu_1(x).$$

It follows from the Monotone Convergence Theorem that μ is a measure on $\mathcal{F}_1 \times \mathcal{F}_2$. Uniqueness follows by the Carathéodory Extension Theorem. (We call a *rectangle* any set of the form $\prod_{i=1}^2 A_i$ where $A_i \in \mathcal{F}_i$ for all $1 \leq i \leq 2$. Let $\mathcal{B}$ be the set of finite disjoint unions of rectangles in $\prod_{i=1}^2 \Omega_i$. It can be shown that $\mathcal{B}$ is an algebra. By assumption, the values of μ are known on rectangles, so the values of μ on the whole σ-algebra must be unique by the uniqueness part of the Extension Theorem.)

□

Solution. [Solution 2] We sketch an argument using rectangles and the Extension Theorem.

By induction, it suffices to prove the case $n = 2$. We call a *rectangle* any set of the form $\prod_{i=1}^2 A_i$ where $A_i \in \mathcal{F}_i$ for all $1 \leq i \leq 2$. Let $\mathcal{A}$ be the set of finite disjoint unions of rectangles in $\prod_{i=1}^2 \Omega_i$. It can be shown that $\mathcal{A}$ is an algebra. We will not prove this. For any finite disjoint union of rectangles $\cup_{i=1}^n A_i \times B_i$, define

$$\mu\Big(\cup_{i=1}^n A_i \times B_i\Big) := \sum_{i=1}^n \mu_1(A_i)\mu_2(B_i).$$

In order to apply the Carathéodory Extension Theorem, we need to show that μ evaluated on a rectangle R agrees with μ when R is split up into a countable disjoint union of rectangles.

Claim 1. Suppose $R = A \times B$ is a rectangle such that R is a countable disjoint union of rectangles, $R = \cup_{i=1}^\infty A_i \times B_i$. Then

$$\mu(R) = \sum_{i=1}^\infty \mu(A_i \times B_i).$$

Proof of Claim 1. By assumption, $1_A(x)1_B(y) = 1_R(x, y) = \sum_{i=1}^\infty 1_{A_i}(x)1_{B_i}(y)$ for any $x \in \Omega_1, y \in \Omega_2$. So, applying the Monotone Convergence Theorem in y,

$$1_A(x)\mu_2(B) = \int_{\Omega_2} 1_A(x)1_B(y)d\mu_2(y) = \sum_{i=1}^\infty \int_{\Omega_2} 1_{A_i}(x)1_{B_i}(y)d\mu_2(y) = \sum_{i=1}^\infty 1_{A_i}(x)\mu_2(B_i).$$

Applying the Monotone Convergence Theorem again in x,

$$\mu(R) = \mu_1(A)\mu_2(B) = \int_{\Omega_1} 1_A(x)\mu_2(B)d\mu_1(x) = \int_{\Omega_1} \sum_{i=1}^{\infty} 1_{A_i}(x)\mu_2(B_i)d\mu_1(x)$$
$$= \sum_{i=1}^{\infty} \mu_2(B_i) \int_{\Omega_1} 1_{A_i}(x)d\mu_1(x) = \sum_{i=1}^{\infty} \mu_1(A_i)\mu_2(B_i).$$

□

It follows from Claim 1 that μ is a completely additive set function on the algebra $\mathcal{F}$. So, the Carathéodory Extension Theorem implies that μ can be extended to a measure on the product σ-algebra. (Technically we also need to verify Claim 1 not just for rectangles, but also when R is a finite disjoint union of rectangles.) □

Exercise 3.4. Let $X_1, \ldots, X_n$ be discrete random variables (i.e. they take values in finite or countable spaces $S_1, \ldots, S_n$ with their discrete σ-algebras). Show that $X_1, \ldots, X_n$ are independent if and only if:

$$\mathbf{P}\big(\bigcap_{i=1}^{n}\{X_i = x_i\}\big) = \prod_{i=1}^{n} \mathbf{P}(X_i = x_i), \qquad \forall\, x_1 \in S_1, \ldots, x_n \in S_n.$$

Solution. It suffices to show that the second condition implies the first. Let $A_1 \subseteq S_1, \ldots, A_n \subseteq S_n$. Then using disjointness,

$$\mathbf{P}\big(\bigcap_{i=1}^{n}\{X_i \in A_i\}\big) = \mathbf{P}\Big(\bigcup_{x_1\in A_1,\ldots,x_n\in A_n} \bigcap_{i=1}^{n}\{X_i = x_i\}\Big) = \sum_{x_1\in A_1,\ldots,x_n\in A_n} \mathbf{P}\big(\bigcap_{i=1}^{n}\{X_i = x_i\}\big)$$
$$= \sum_{x_1\in A_1,\ldots,x_n\in A_n} \prod_{i=1}^{n} \mathbf{P}(X_i = x_i) = \prod_{i=1}^{n} \sum_{x_i\in A_i} \mathbf{P}(X_i = x_i) = \prod_{i=1}^{n} \mathbf{P}(X_i \in A_i).$$

□

Exercise 3.5. Show that $X_1, \ldots, X_n \colon \Omega \to \mathbb{R}$ are independent if and only if:

$$\mathbf{P}\big(\bigcap_{i=1}^{n}\{X_i \le x_i\}\big) = \prod_{i=1}^{n} \mathbf{P}(X_i \le x_i), \qquad \forall\, x_1, \ldots, x_n \in \mathbb{R}.$$

Solution. It suffices to show that the second condition implies the first. Fix $x_2, \ldots, x_n \in \mathbb{R}$. Let $\mathcal{A}$ be the set of inverse images $X_1^{-1}(D)$ of measurable subsets D of $\mathbb{R}$ such that

$$\mathbf{P}(X_1 \in D, X_2 \le x_2, \ldots, X_n \le x_n) = \mathbf{P}(X_1 \in D) \prod_{i=2}^{n} \mathbf{P}(X_i \le x_i).$$

We will show that $\mathcal{A}$ is a monotone class. Denote $C := \{X_2 \le x_2, \ldots, X_n \le x_n\}$. (Note that $\prod_{i=2}^{n} \mathbf{P}(X_i \le x_i) = \mathbf{P}(C)$ by assumption.)

By choosing $D = \emptyset$ or $D = \mathbb{R}$, note that $\emptyset, \mathbb{R} \in \mathcal{A}$. Let $A, B \in \mathcal{A}$ with $A \subseteq B$. Then $(B \smallsetminus A) \cap C = (B \cap C) \smallsetminus (A \cap C)$, so

$$\mathbf{P}((B \smallsetminus A) \cap C) = \mathbf{P}(B \cap C) - \mathbf{P}(A \cap C) = \mathbf{P}(B)\mathbf{P}(C) - \mathbf{P}(A)\mathbf{P}(C)$$
$$= (\mathbf{P}(B) - \mathbf{P}(A))\mathbf{P}(C) = \mathbf{P}(B \smallsetminus A)\mathbf{P}(C).$$

Therefore, $B \setminus A \in \mathcal{A}$. In particular, $\Omega \setminus A = A^c \in \mathcal{A}$. Since $\mathcal{A}$ is closed under complements, it remains to show that $\mathcal{A}$ is closed under increasing unions. Let $A_1 \subseteq A_2 \subseteq \cdots$ be sets in $\mathcal{A}$. Using Exercise 1.5 twice,

$$\begin{aligned}\mathbf{P}((\cup_{m=1}^{\infty} A_m) \cap C) &= \mathbf{P}(\cup_{m=1}^{\infty}(A_m \cap C)) = \lim_{m\to\infty} \mathbf{P}(A_m \cap C) \\ &= \lim_{m\to\infty} \mathbf{P}(A_m)\mathbf{P}(C) = \mathbf{P}(\cup_{m=1}^{\infty} A_m)\mathbf{P}(C).\end{aligned}$$

So, $(\cup_{m=1}^{\infty} A_m) \in \mathcal{A}$. That is, $\mathcal{A}$ is a monotone class, as desired.

Since $\mathcal{A}$ is a monotone class, by the Monotone Class Lemma, $\mathcal{A}$ contains the σ-algebra generated by $\{X_1^{-1}(-\infty, x_1]\}_{x_1\in\mathbb{R}}$. (Here we are applying the Monotone Class Lemma to $\mathcal{F}$ which is the algebra generated by $\{X_1^{-1}(-\infty, x_1]\}_{x_1\in\mathbb{R}}$.) That is, $\mathcal{A}$ contains the set of all $X_1^{-1}(D)$ where $D \subseteq \mathbb{R}$ is measurable. So, by definition of $\mathcal{A}$, for any measurable $D \subseteq \mathbb{R}$, we have

$$\mathbf{P}(X_1 \in D, X_2 \leq x_2, \ldots, X_n \leq x_n) = \mathbf{P}(X_1 \in D) \prod_{i=2}^{n} \mathbf{P}(X_i \leq x_i).$$

We can now iterate this argument. Fix measurable $D \subseteq \mathbb{R}$ and $x_3, \ldots, x_n \in \mathbb{R}$. Let $\mathcal{A}'$ be the set of inverse images $X_2^{-1}(D')$ of measurable subsets D' of $\mathbb{R}$ such that

$$\mathbf{P}(X_1 \in D, X_2 \in D', X_3 \leq x_3, \ldots, X_n \leq x_n) = \mathbf{P}(X_1 \in D)\mathbf{P}(X_2 \in D') \prod_{i=3}^{n} \mathbf{P}(X_i \leq x_i).$$

The above argument implies that $\mathcal{A}'$ is a monotone class, etc. Running this argument for each of the indices $2, \ldots, n$ concludes the proof. □

Exercise 3.6 (Optional)**.** Let V be a finite-dimensional vector space over a finite field $\mathbb{F}$. Let X be a random variable uniformly distributed in V. Let $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ be a non-degenerate bilinear form on V (if $v \in V$ satisfies $\langle v, w \rangle = 0$ for all $w \in V$, then $v = 0$.). Let $v_1, \ldots, v_n$ be non-zero vectors in V. Show that the random variables $\langle X, v_1 \rangle, \ldots, \langle X, v_n \rangle$ are independent if and only if the vectors $v_1, \ldots, v_n$ are linearly independent.

Solution.

Suppose V has dimension m over the field $\mathbb{F}$, denoted $\dim_{\mathbb{F}}(V) = m$. Let $\alpha_1 \in \mathbb{F}$ and let $V_1 := \{v \in V : \langle v, v_1 \rangle = \alpha_1\}$. We first claim that V_1 is an affine subspace of V of dimension $n-1$, denoted $\dim_{\mathbb{F}}(V_1) = m - 1$. To prove the claim, note that since $v_1 \neq 0$, there exists $w \in V$ such that $\langle v_1, w \rangle \neq 0$. So, $\langle v_1, w \rangle \cdot (\alpha_1 \langle v_1, w \rangle^{-1}) = \alpha_1$, i.e. the vector $u := w\alpha\langle v_1, w \rangle^{-1}$ satisfies $\langle v_1, u \rangle = \alpha_1$, so that $u \in V_1$ (in case $\alpha_1 = 0$, we just define $u := 0 \in V$). Meanwhile, note that $\{v \in V : \langle v, v_1 \rangle = 0\}$ is an $(m-1)$-dimensional subspace of V since it is the kernel of the linear map $v \mapsto \langle v, v_1 \rangle$ that maps V to $\mathbb{F}$, so $\{v \in V : \langle v, v_1 \rangle = 0\}$ has dimension at least $m - 1$ by the rank-nullity theorem, and it does not have dimension n since the bilinear form is nondegenerate. In conclusion, $V_1 = u + \{v \in V : \langle v, v_1 \rangle = 0\}$, so that V_1 is indeed an affine subspace of dimension $m - 1$.

Now, for any $1 \leq i \leq n$, let $\alpha_i \in \mathbb{F}$ and define

$$V_i := \{v \in V : \langle v, v_i \rangle = \alpha_i\}.$$

From the above claim, for all $1 \leq i \leq n$,

$$\dim_{\mathbb{F}}(V_i) = m - 1. \qquad (*)$$

More generally, suppose $\cap_{i=1}^n V_i$ is nonempty. Then $\cap_{i=1}^n V_i$ is equal to some vector $u \in V$ plus kernel of the map $v \mapsto (\langle v, v_1\rangle, \ldots, \langle v, v_n\rangle)$ from V to $\mathbb{F}^n$. The latter map can be represented by a matrix with n rows $v_1, \ldots, v_n$. By the rank-nullity theorem, the kernel of this map is equal to m minus the dimension of the span of $v_1, \ldots, v_n$. That is, if $\cap_{i=1}^n V_i$ is nonempty,

$$\dim_{\mathbb{F}}(\cap_{i=1}^n V_i) = m - \dim_{\mathbb{F}}\text{span}(v_1, \ldots, v_n).$$

So, letting $|\cdot|$ denote cardinality,

$$|\cap_{i=1}^n \{v \in V : \langle v, v_i\rangle = \alpha_i\}| = |\mathbb{F}|^{m-n} \text{ if and only if } v_1, \ldots, v_n \text{ are linearly independent.}$$

Since V has dimension m, $|V| = |\mathbb{F}|^m$, so that by definition of X,

$$\mathbf{P}(\cap_{i=1}^n \{\langle X, v_i\rangle = \alpha_i\}) = \frac{|\mathbb{F}|^{m-n}}{|\mathbb{F}|^m} = |\mathbb{F}|^{-n} = \left(\frac{|\mathbb{F}|^{m-1}}{|\mathbb{F}|^m}\right)^n \overset{(*)}{=} \prod_{i=1}^n \mathbf{P}(\langle X, v_i\rangle = \alpha_i)$$

if and only if $v_1, \ldots, v_n$ are linearly independent.

So, by Exercise 3.4, $\langle X, v_1\rangle, \ldots, \langle X, v_n\rangle$ are independent if and only if $v_1, \ldots, v_n$ are linearly independent.

The above argument works in the case $m \geq n$. In the case $m < n$, the above argument shows that $\langle X, v_1\rangle, \ldots, \langle X, v_n\rangle$ are not independent, and since $m < n$ the vectors $v_1, \ldots, v_n$ are automatically linearly dependent. □

Exercise 3.7. Give an example of three random variables $X, Y, Z : \Omega \to [-\infty, \infty]$ that are pairwise independent (any two of the random variables X, Y, Z are independent of each other), but such that X, Y, Z are not independent. (Hint: Exercise 3.6 might be helpful.)

Solution. Let $p \geq 3$ be a prime and let $n \geq 2$. Let $\mathbb{F} := \mathbb{Z}/p\mathbb{Z}$ and let $V := (\mathbb{Z}/p\mathbb{Z})^n$. Note that V is an n-dimensional vector space over $\mathbb{F}$. Equip V with the standard bilinear form $\langle u, v\rangle := \sum_{i=1}^n u_i v_i$ where $u = (u_1, \ldots, u_n), v = (v_1, \ldots, v_n) \in V$. Then there exist three vectors $x, y, z \in V$ such that any two of x, y, z are linearly independent, but such that x, y, z are linearly dependent. Let U be uniformly distributed in V. Define $X := \langle U, x\rangle$, $Y := \langle U, y\rangle$, $Z := \langle U, z\rangle$. From Exercise 3.6, any two of X, Y, Z are independent, but X, Y, Z are not independent. Also, X, Y, Z take values in the integers. □

Exercise 3.8 (Optional)**.** Let $X : \Omega \to \mathbb{R}^n$ be a random variable with the **standard Gaussian distribution**:

$$\mathbf{P}(X \in A) := \int_A e^{-(x_1^2 + \cdots + x_n^2)/2} dx (2\pi)^{-n/2}, \qquad \forall\, A \subseteq \mathbb{R}^n \text{ measurable.}$$

Let $v_1, \ldots, v_m$ be vectors in $\mathbb{R}^n$. Let $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be the standard inner product on $\mathbb{R}^n$, so that $\langle x, y\rangle := \sum_{i=1}^n x_i y_i$ for any $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \mathbb{R}^n$. Show that the random variables $\langle X, v_1\rangle, \ldots, \langle X, v_m\rangle$ are independent if and only if the vectors $v_1, \ldots, v_m$ are pairwise orthogonal.

Solution. Without loss of generality, all of the vectors are nonzero. Suppose for now that $m \leq n$. Suppose the vectors $v_1, \ldots, v_m$ are pairwise orthogonal. For any $1 \leq i \leq n$, let $e_i \in \mathbb{R}^n$ be the vector with a 1 in the i^{th} entry and zeros in all other entries. Let Q be any $n \times n$ real orthogonal matrix such that $Qe_i = v_i / \|v_i\|$ for all $1 \leq i \leq m$ (this is possible since $m \leq n$). Specifically, we let the first m columns of Q be $v_1/\|v_1\|, \ldots, v_m/\|v_m\|$. Suppose the remaining columns of Q are $Q_{m+1}, \ldots, Q_n$. Recall that $Q^{-1} = Q^T$ so that

$QQ^T = Q^TQ = I_n$, where I_n denotes the $n \times n$ identity matrix. Note the rows of Q^T are $v_1/\left\|v_1\right\|, \ldots, v_m/\left\|v_m\right\|, Q_{m+1}, \ldots, Q_n$, so

$$Q \begin{pmatrix} \langle X, \frac{v_1}{\|v_1\|} \rangle \\ \vdots \\ \langle X, \frac{v_m}{\|v_m\|} \rangle \\ \langle X, Q_{m+1} \rangle \\ \vdots \\ \langle X, Q_n \rangle \end{pmatrix} = QQ^T X = X. \qquad (*)$$

So, if $A_1, \ldots, A_m \subseteq \mathbb{R}$ are measurable, we have

$$\begin{aligned}
&\mathbf{P}(\langle X, v_1 \rangle \in A_1, \ldots, \langle X, v_m \rangle \in A_m) = \mathbf{P}(\left(\langle X, v_1 \rangle, \ldots, \langle X, v_m \rangle\right) \in A_1 \times \cdots \times A_m) \\
&= \mathbf{P}(\left(\langle X, v_1 \rangle, \ldots, \langle X, v_m \rangle, \langle X, Q_{m+1} \rangle, \ldots, \langle X, Q_n \rangle\right) \in A_1 \times \cdots \times A_m \times \mathbb{R}^{n-m}) \\
&= \mathbf{P}(\left(\langle X, \frac{v_1}{\|v_1\|} \rangle, \ldots, \langle X, \frac{v_m}{\|v_m\|} \rangle, \langle X, Q_{m+1} \rangle, \ldots, \langle X, Q_n \rangle\right) \in \frac{A_1}{\|v_1\|} \times \cdots \times \frac{A_m}{\|v_m\|} \times \mathbb{R}^{n-m}) \\
&= \mathbf{P}(Q^TQ(\langle X, \frac{v_1}{\|v_1\|} \rangle, \ldots, \langle X, \frac{v_m}{\|v_m\|} \rangle, \langle X, Q_{m+1} \rangle, \ldots, \langle X, Q_n \rangle)^T \in \frac{A_1}{\|v_1\|} \times \cdots \times \frac{A_m}{\|v_m\|} \times \mathbb{R}^{n-m}) \\
&= \mathbf{P}(Q(\langle X, \frac{v_1}{\|v_1\|} \rangle, \ldots, \langle X, \frac{v_m}{\|v_m\|} \rangle, \langle X, Q_{m+1} \rangle, \ldots, \langle X, Q_n \rangle)^T \in Q(\frac{A_1}{\|v_1\|} \times \cdots \times \frac{A_m}{\|v_m\|} \times \mathbb{R}^{n-m})) \\
&= \mathbf{P}(X \in Q(\frac{A_1}{\|v_1\|} \times \cdots \times \frac{A_m}{\|v_m\|} \times \mathbb{R}^{n-m})) \qquad \text{, by } (*) \\
&= \int_{Q(\frac{A_1}{\|v_1\|} \times \cdots \times \frac{A_m}{\|v_m\|} \times \mathbb{R}^{n-m})} e^{-(x_1^2 + \cdots + x_n^2)/2} dx (2\pi)^{-n/2} \quad \text{, by definition of } X \\
&= \int_{\frac{A_1}{\|v_1\|} \times \cdots \times \frac{A_m}{\|v_m\|} \times \mathbb{R}^{n-m}} e^{-x^T Q^T Q x/2} dx (2\pi)^{-n/2} \quad \text{, changing variables} \\
&= \int_{\frac{A_1}{\|v_1\|} \times \cdots \times \frac{A_m}{\|v_m\|} \times \mathbb{R}^{n-m}} e^{-\|x\|^2/2} dx (2\pi)^{-n/2} = \prod_{i=1}^m \int_{A_i/\|v_i\|} e^{-x_i^2/2} dx_i/\sqrt{2\pi} \quad \text{, by Fubini's Theorem} \\
&= \prod_{i=1}^m \mathbf{P}(\langle X, \frac{v_i}{\|v_i\|} \rangle \in \frac{A_i}{\|v_i\|}) = \prod_{i=1}^m \mathbf{P}(\langle X, v_i \rangle \rangle \in A_i).
\end{aligned}$$

In the penultimate equality, we used that $\langle X, \frac{v_i}{\|v_i\|} \rangle$ is a standard one-dimensional Gaussian random variable. This follows from the $m = 1$ case of the above.

More generally, for any linearly independent vectors $v_1, \ldots, v_m$, we construct Q as above, where now Q may not be orthogonal, and

$$(QQ^T)^{-1}Q\begin{pmatrix} \langle X, \frac{v_1}{\|v_1\|}\rangle \\ \vdots \\ \langle X, \frac{v_m}{\|v_m\|}\rangle \\ \langle X, Q_{m+1}\rangle \\ \vdots \\ \langle X, Q_n\rangle \end{pmatrix} = (QQ^T)^{-1}QQ^TX = X. \qquad (*)$$

So we then get

$$\begin{aligned}
&\mathbf{P}(\langle X, v_1\rangle \in A_1, \ldots, \langle X, v_m\rangle \in A_m) = \mathbf{P}(\,(\langle X, v_1\rangle, \ldots, \langle X, v_m\rangle) \in A_1 \times \cdots \times A_m) \\
&= \mathbf{P}(\,(\langle X, v_1\rangle, \ldots, \langle X, v_m\rangle, \langle X, Q_{m+1}\rangle, \ldots, \langle X, Q_n\rangle) \in A_1 \times \cdots \times A_m \times \mathbb{R}^{n-m}) \\
&= \mathbf{P}(\,(\langle X, \frac{v_1}{\|v_1\|}\rangle, \ldots, \langle X, \frac{v_m}{\|v_m\|}\rangle, \langle X, Q_{m+1}\rangle, \ldots, \langle X, Q_n\rangle) \in \frac{A_1}{\|v_1\|} \times \cdots \times \frac{A_m}{\|v_m\|} \times \mathbb{R}^{n-m}) \\
&= \mathbf{P}((QQ^T)^{-1}Q(\langle X, \frac{v_1}{\|v_1\|}\rangle, \ldots, \langle X, \frac{v_m}{\|v_m\|}\rangle, \langle X, Q_{m+1}\rangle, \ldots, \langle X, Q_n\rangle)^T \\
&\qquad\qquad \in (QQ^T)^{-1}Q(\frac{A_1}{\|v_1\|} \times \cdots \times \frac{A_m}{\|v_m\|} \times \mathbb{R}^{n-m})) \\
&= \mathbf{P}(X \in (QQ^T)^{-1}Q(\frac{A_1}{\|v_1\|} \times \cdots \times \frac{A_m}{\|v_m\|} \times \mathbb{R}^{n-m})) \qquad \text{, by } (*) \\
&= \int_{(QQ^T)^{-1}Q(\frac{A_1}{\|v_1\|}\times\cdots\times\frac{A_m}{\|v_m\|}\times\mathbb{R}^{n-m})} e^{-(x_1^2+\cdots+x_n^2)/2}dx(2\pi)^{-n/2} \quad \text{, by definition of } X \\
&= |\det(Q)|^{-1} \int_{\frac{A_1}{\|v_1\|}\times\cdots\times\frac{A_m}{\|v_m\|}\times\mathbb{R}^{n-m}} e^{-x^TQ^T(QQ^T)^{-2}Qx/2}dx(2\pi)^{-n/2} \quad \text{, changing variables}
\end{aligned}$$

Letting D be the diagonal matrix with diagonal entries $\|v_1\|^{-1}, \ldots, \|v_m\|^{-1}, 1, \ldots, 1$,

$$\begin{aligned}
&\mathbf{P}(\langle X, v_1\rangle \in A_1, \ldots, \langle X, v_m\rangle \in A_m) \\
&= |\det(Q)|^{-1} \int_{D(A_1\times\cdots\times A_m\times\mathbb{R}^{n-m})} e^{-x^TQ^T(QQ^T)^{-2}Qx/2}dx(2\pi)^{-n/2} \\
&= |\det(Q)|^{-1}\,|\det(D)| \int_{A_1\times\cdots\times A_m\times\mathbb{R}^{n-m}} e^{-x^TDQ^T(QQ^T)^{-2}QDx/2}dx(2\pi)^{-n/2}
\end{aligned}$$

So, the joint density of $\langle X, v_1\rangle, \ldots, \langle X, v_m\rangle$ is

$$|\det(Q)|^{-1}\,|\det(D)| \int_{\mathbb{R}^{n-m}} e^{-x^TDQ^T(QQ^T)^{-2}QDx/2}dx_{n-m+1}\cdots dx_n(2\pi)^{-n/2}.$$

In the special case $n = m = 2$ with $\|v_1\| = \|v_2\| = 1$, the joint density becomes

$$|\det(Q)|^{-1}\, e^{-x^TQ^T(QQ^T)^{-2}Qx/2}(2\pi)^{-n/2}.$$

In this case Q is the matrix whose columns are v_1, v_2. Let $B := Q^T(QQ^T)^{-2}Q$. Suppose we write Q in its singular value decomposition as $Q = UDV$ where U, V are orthogonal and D is diagonal. If $v_1 = \pm v_2$, then $\langle X, v_1\rangle, \langle X, v_2\rangle$ are dependent, so we may assume that $v_1 \neq v_2$ and $v_1 \neq -v_2$. Then Q has rank 2, so D has nonzero diagonal entries. Since $Q^T = V^T D U^T$, using $UU^T = I$ and $VV^T = I$, we get

$$Q^T(QQ^T)^{-2}Q = V^T D^{-2} V.$$

If v_1, v_2 are not orthogonal, $D \neq I$, so $Q^T(QQ^T)^{-2}Q$ is not a diagonal matrix. Therefore, $|\det(Q)|^{-1} e^{-x^T Q^T (QQ^T)^{-2} Qx/2} (2\pi)^{-n/2}$ is not a product function, so $\langle X, v_1\rangle, \langle X, v_2\rangle$ are not independent. The general case then follows from the $m = n = 2$ case.

□

4. Homework 4

Exercise 4.1. Let $\varepsilon_1, \varepsilon_2, \cdots \in \{0, 1\}$ be random variables that are independent and identically distributed copies of the Bernoulli random variable with expectation $1/2$, so that $\mathbf{P}(\varepsilon_n = 1) = \mathbf{P}(\varepsilon_n = 0) = 1/2$ for all $n \geq 1$.

- Show that the random variable $\sum_{n=1}^{\infty} 2^{-n}\varepsilon_n$ is uniformly distributed on the unit interval $[0, 1]$.
- Show that the random variable $\sum_{n=1}^{\infty} 2 \cdot 3^{-n}\varepsilon_n$ is uniformly distributed on the standard middle third Cantor set (where the Cantor set's center is $1/2$.)
- Let μ be a probability measure on $\mathbb{R}$. The Fourier Transform of μ at $\xi \in \mathbb{R}$ is defined by $\widehat{\mu}(\xi) := \int_{\mathbb{R}} e^{i x \xi} d\mu(x)$. where $i = \sqrt{-1}$. For example, if μ is uniform on $[-1/2, 1/2]$, then

$$\widehat{\mu}(\xi) = \int_{-1/2}^{1/2} e^{ix\xi} dx = \frac{e^{i\xi/2} - e^{-i\xi/2}}{i\xi} = \frac{2\sin(\xi/2)}{\xi}, \qquad \forall\, \xi \neq 0.$$

  Using the first item, find an expression for $\sin(\xi)/\xi$ in terms of an infinite product of cosines. (Hint: if a random variable X has distribution μ_X, then $\widehat{\mu_X}(\xi) = \mathbf{E}e^{iX\xi}$ for any $\xi \in \mathbb{R}$. So the Fourier transform of the sum of independent random variables is the product of the Fourier transforms.) Similarly, find an expression for the Fourier transform of the uniform measure on the middle third Cantor set (when the Cantor set's center is $0 \in \mathbb{R}$) in terms of an infinite product of cosines.

Solution. Let $I \subseteq [0, 1]$ be an interval of the form $[k/2^n, (k+1)/2^n]$ where $0 \leq k \leq 2^n - 1$ (such an interval is called a dyadic interval). We claim that $X := \sum_{n=1}^{\infty} 2^{-n}\varepsilon_n$ satisfies $\mathbf{P}(X \in I) = 2^{-n}$. To see this, note that the event $X \in I$ corresponds to $\varepsilon_1, \cdots, \varepsilon_n$ being equal to the binary expansion of k. And the latter event has probability 2^{-n}. It then follows that $\mathbf{P}(X \leq k/2^n) = k/2^n$ for any $n \geq 1$ and for any $0 \leq k \leq 2^n$, since we can write $\mathbf{P}(X \leq k/2^n)$ as the sum of probabilities of X lying in disjoint dyadic intervals of width 2^{-n}. So, the cumulative distribution function F of X is known at any number in $[0, 1]$ of the form $k/2^n$, namely, $F(k/2^n) = k/2^n$. Since the numbers $\{k/2^n\}_{n\geq 1, 0\leq k\leq 2^n}$ form a dense set in $[0, 1]$, it follows by right-continuity of F that $F(t) = t$ for all $t \in [0, 1]$. Therefore, X is uniform on $[0, 1]$.

The proof for the middle third Cantor set is analogous. Let $f_0(x) := x/3$ and let $f_1(x) := 2/3 + x/3$ for any $x \in \mathbb{R}$. Let I be any interval formed by n compositions $f_{a_1}(f_{a_2}(\cdots(f_{a_n}[0, 1])))$ where $a_1, \ldots, a_n \in \{0, 1\}$. Then $X := \sum_{n=1}^{\infty} 2 \cdot 3^{-n}\varepsilon_n$ satisfies $\mathbf{P}(X \in$

$I) = 2^{-n}$. We take this as a definition that X is uniformly distributed on the middle third Cantor set.

By shifting the first result by $1/2$, the following random variable is uniformly distributed on $[-1/2, 1/2]$

$$-\frac{1}{2} + \sum_{n=1}^{\infty} 2^{-n}\varepsilon_n = \sum_{n=1}^{\infty} 2^{-n}(\varepsilon_n - 1/2) =: X.$$

Using that the expected value of a product of independent random variables is the product of expected values,

$$\frac{\sin(\xi/2)}{\xi} = \frac{e^{-i\xi/2} - e^{i\xi/2}}{-i\xi} = \int_{-1/2}^{1/2} e^{-ix\xi} dx = \mathbf{E}e^{-iX\xi} = \widehat{\mu_X}(\xi) = \prod_{n=1}^{\infty} \mathbf{E}e^{2^{-n}i[\varepsilon_n - 1/2]\xi}$$

$$= \prod_{n=1}^{\infty} \frac{1}{2}[e^{i2^{-n-1}\xi} + e^{-i2^{-n-1}\xi}] = \prod_{n=1}^{\infty} \cos(\xi 2^{-n-1}).$$

Therefore, for any $\xi \in \mathbb{R}$,

$$\frac{\sin \xi}{\xi} = \prod_{n=1}^{\infty} \cos(\xi 2^{-n}).$$

By shifting the second result by $1/2$, the following random variable is uniformly distributed on the middle third Cantor set, centered at 0

$$-\frac{1}{2} + \sum_{n=1}^{\infty} 2 \cdot 3^{-n}\varepsilon_n = \sum_{n=1}^{\infty} 2 \cdot 3^{-n}(\varepsilon_n - 1/2) =: X.$$

Using that the expected value of a product of independent random variables is the product of expected values,

$$\widehat{\mu_X}(\xi) = \mathbf{E}e^{-iX\xi} = \prod_{n=1}^{\infty} \mathbf{E}e^{2\cdot 3^{-n}i[\varepsilon_n - 1/2]\xi}$$

$$= \prod_{n=1}^{\infty} \frac{1}{2}[e^{i3^{-n}\xi} + e^{-i3^{-n}\xi}] = \prod_{n=1}^{\infty} \cos(\xi 3^{-n}).$$

□

Exercise 4.3 (**MAX-CUT**)**.** The probabilistic method is a very useful way to prove the existence of something satisfying some properties. This method is based upon the following elementary statement: If $\alpha \in \mathbb{R}$ and if a random variable $X \colon \Omega \to \mathbb{R}$ satisfies $\mathbf{E}X \geq \alpha$, then there exists some $\omega \in \Omega$ such that $X(\omega) \geq \alpha$. We will demonstrate this principle in this exercise.

Let $G = (V, E)$ be an undirected graph on the vertices $V = \{1, \ldots, n\}$ so that the edge set E is a subset of unordered pairs $\{i, j\}$ such that $i, j \in V$ and $i \neq j$. Let $S \subseteq V$ and denote $S^c := V \setminus S$. We refer to (S, S^c) as a cut of the graph G. The goal of the MAX-CUT problem is to maximize the number of edges going between S and S^c over all cuts of the graph G.

Prove that there exists a cut (S, S^c) of the graph such that the number of edges going between S and S^c is at least $|E|/2$. (Hint: define a random $S \subseteq V$ such that, for every $i \in V$, $\mathbf{P}(i \in S) = 1/2$, and the events $1 \in S, 2 \in S, \ldots, n \in S$ are all independent. If

$\{i,j\} \in E$, show that $\mathbf{P}(i \in S, j \notin S) = 1/4$. So, what is the expected number of edges $\{i,j\} \in E$ such that $i \in S$ and $j \notin S$?)

Solution. Define S as noted above. (Recall that these random variables can exist by a Corollary of the Kolmogorov Extension Theorem.) If $\{i,j\} \in E$, then by independence we have $\mathbf{P}(i \in S, j \notin S) = \mathbf{P}(i \in S)\mathbf{P}(j \notin S) = (1/2)^2 = 1/4$. Similarly, $\mathbf{P}(i \notin S, j \in S) = 1/4$. And the event that one of i,j is in S and the other is not in S is the disjoint union $\{i \in S, j \notin S\} \cup \{i \notin S, j \in S\}$. So, the probability that one of i,j is in S and the other is not in S is $1/4 + 1/4 = 1/2$. By linearity of expected value, the expected number of cut edges is

$$\mathbf{E} \sum_{\{i,j\} \in E} 1_{\{i \in S, j \notin S\} \cup \{i \notin S, j \in S\}} = \sum_{\{i,j\} \in E} \mathbf{E} 1_{\{i \in S, j \notin S\} \cup \{i \notin S, j \in S\}} = |E| \cdot (1/2).$$

□

Exercise 4.4. Let $X_1, X_2, \ldots : \Omega \to S$ be random variables. Show that

$$\sigma(X_1, X_2, \ldots) = \sigma(\cup_{i=1}^{\infty} \sigma(X_1, \ldots, X_i)).$$

Exercise 4.5. Let $(X_i)_{i \in I}$ be a collection of independent random variables. Show that $(X_i)_{i \in I}$ are independent if and only if $(\sigma(X_i))_{i \in I}$ are independent σ-algebras. (Hint: Let $i \in I$ and let $J \subseteq I \setminus \{i\}$ be finite. Are the sets in $\sigma(X_i)$ that are independent of $(\sigma(X_j))_{j \in J}$ a monotone class?)

Solution. It suffices to show: if $(X_i)_{i \in I}$ are independent, then $(\sigma(X_i))_{i \in I}$ are independent. Suppose $X_i \colon \Omega \to S_i$ and S_i has the σ-algebra $\mathcal{B}_i$, for any $i \in I$. Let $i \in I$ and let $J \subseteq I \setminus \{i\}$ be finite. For any $j \in J$, let $B_j \in \mathcal{B}_j$ and let $B_i \in \mathcal{B}_i$. We know that the sets $X_i^{-1}(B_i)$ and $(X_j^{-1}(B_j))_{j \in J}$ are independent, by assumption. Consider the set $\mathcal{A} := \{A \in \sigma(X_i) \colon A \text{ is independent of } (X_j^{-1}(B_j))_{j \in J}\}$. We know that $\mathcal{A}$ contains $\{X_i^{-1}(B_i) \colon B_i \in \mathcal{B}_i\}$. Also $\mathcal{A}$ is closed under complements and countable increasing unions (repeating the proof of the Kolmogorov's Zero-One Law). Therefore $\mathcal{A}$ contains $\sigma(X_i)$. Therefore, $\sigma(X_i)$ is independent of $(X_j^{-1}(B_j))_{j \in J}$. Repeating this argument once for every index in J, we conclude that $(\sigma(X_j))_{j \in J}$ are independent. □

Exercise 4.6. Let $X_1, X_2, \ldots$ be random variables. Show that $X_1, X_2, \ldots$ are independent if and only if: for every $i \geq 1$, $\sigma(X_{i+1})$ is independent of $\sigma(X_1, \ldots, X_i)$. And the previous cases occur if and only if: for every $i \geq 1$, $\sigma(X_{i+1}, X_{i+2}, \ldots)$ is independent of $\sigma(X_1, \ldots, X_i)$.

Solution. From Exercise 4.5, $X_1, X_2, \ldots$ are independent if and only if $\sigma(X_1), \sigma(X_2), \ldots$ are independent. If $\sigma(X_1), \sigma(X_2), \ldots$ are independent, then for every $i \geq 1$, $\sigma(X_{i+1})$ is independent of $\sigma(\sigma(X_1) \cup \cdots \cup \sigma(X_i)) = \sigma(X_1, \ldots, X_i)$. Conversely, assume: for every $i \geq 1$, $\sigma(X_{i+1})$ is independent of $\sigma(X_1, \ldots, X_i)$. Suppose $X_i \colon \Omega \to S_i$ and S_i has the σ-algebra $\mathcal{B}_i$, for any $i \in \mathbb{N}$. Let $J = (j_1, \ldots, j_n) \subseteq \mathbb{N}$ be finite and let $B_j \in S_j$ for all $j \in J$. Then $X_{j_n}^{-1}(B_{j_n})$ is independent of $\cap_{k=1}^{n-1} \{X_{j_k}^{-1}(B_{j_k})\}$. So,

$$\mathbf{P}\Big(\cap_{k=1}^{n} \{X_{j_k}^{-1}(B_{j_k})\} \Big) = \mathbf{P}(X_{j_n}^{-1}(B_{j_n})) \mathbf{P}\Big(\cap_{k=1}^{n-1} \{X_{j_k}^{-1}(B_{j_k})\} \Big).$$

Iterating this argument, we conclude that

$$\mathbf{P}\Big(\cap_{k=1}^{n} \{X_{j_k}^{-1}(B_{j_k})\} \Big) = \prod_{k=1}^{n} \mathbf{P}(X_{j_k}^{-1}(B_{j_k})).$$

That is, $X_1, X_2, \ldots$ are independent so that $\sigma(X_1), \sigma(X_2), \ldots$ are independent. The first equivalence is proven.

We now show the second equivalence. The last condition implies the condition: for every $i \geq 1$, $\sigma(X_{i+1})$ is independent of $\sigma(X_1, \ldots, X_i)$. Conversely, if for every $i \geq 1$, $\sigma(X_{i+1})$ is independent of $\sigma(X_1, \ldots, X_i)$, then from the case already proven, $X_1, X_2, \ldots$ are independent and $\sigma(X_1), \sigma(X_2), \ldots$ are independent. Fix $i \geq 1$. Let $\mathcal{A} = \{A \in \sigma(X_{i+1}, X_{i+2}, \ldots) \colon A \text{is independent of} \sigma(X_1, \ldots, X_i)\}$. Since $\sigma(\sigma(X_1), \ldots \sigma(X_i)) = \sigma(X_1, \ldots, X_i)$, we know that $\sigma(X_{i+1}), \sigma(X_{i+1}), \ldots \subseteq \mathcal{A}$. As usual, $\mathcal{A}$ is a monotone class, so $\mathcal{A}$ therefore contains $\sigma(X_{i+1}, X_{i+2}, \ldots)$, as desired. $\square$

Exercise 4.7. Let $X_1, X_2, \ldots : \Omega \to \mathbb{R}$ be a sequence of independent random variables. For any $n \geq 1$, let $S_n := X_1 + \cdots + X_n$. Show the following:

- $\{\lim_{n\to\infty} S_n \text{ exists}\} \in \mathcal{T}$.
- If $t \in [-\infty, \infty]$, then it can occur that $\{\limsup_{n\to\infty} S_n > t\} \notin \mathcal{T}$.
- If $t \in [-\infty, \infty]$ and if $c_1 \leq c_2 \leq \cdots$ is a sequence of real numbers such that $\lim_{n\to\infty} c_n = \infty$, then
$$\{\limsup_{n\to\infty} \frac{S_n}{c_n} > t\} \in \mathcal{T}.$$

Solution. For any $k > 1$, note that

$$\begin{aligned}
&\limsup_{n\to\infty}(X_1 + \cdots + X_n) - \liminf_{n\to\infty}(X_1 + \cdots + X_n) \\
&= \lim_{n\to\infty} \Big(\sup_{m\geq n}(X_1 + \cdots + X_m) - \inf_{m\geq n}(X_1 + \cdots + X_m) \Big) \\
&= \lim_{n\to\infty} \Big(\sup_{m\geq n+k}(X_1 + \cdots + X_m) - \inf_{m\geq n+k}(X_1 + \cdots + X_m) \Big) \\
&= \lim_{n\to\infty} \Big(X_1 + \cdots + X_k - (X_1 + \cdots + X_k) + \sup_{m\geq n}(X_{k+1} + \cdots + X_m) - \inf_{m\geq n}(X_{k+1} + \cdots + X_m) \Big) \\
&= \lim_{n\to\infty} \Big(\sup_{m\geq n}(X_{k+1} + \cdots + X_m) - \inf_{m\geq n}(X_{k+1} + \cdots + X_m) \Big).
\end{aligned}$$

The latter random variable is measurable with respect to $\sigma(X_{k+1}, X_{k+2}, \ldots)$ for any $k > 1$. Therefore, this random variable is measurable with respect to $\mathcal{T} := \cap_{k\geq 1}\sigma(X_{k+1}, X_{k+2}, \ldots)$. In particular, the event that $\limsup_{n\to\infty} S_n - \liminf_{n\to\infty} S_n = 0$ is in $\mathcal{T}$.

For the second item, note that if $\mathbf{P}(X_1 = 1) = \mathbf{P}(X_1 = -1) = 1/2$ and $X_n = 0$ for all $n > 1$, then all of these random variables are independent, $\limsup_{n\to\infty} S_n = X_1$, $\mathcal{T} = \{\emptyset, \Omega\}$, so $\{X_1 > 0\} \notin \mathcal{T}$.

Lastly, for any $k > 1$, note that

$$\begin{aligned}
\limsup_{n\to\infty} \frac{X_1 + \cdots + X_n}{c_n} &= \lim_{n\to\infty} \sup_{m\geq n} \frac{X_1 + \cdots + X_m}{c_m} \\
&= \lim_{n\to\infty} \Big(\frac{X_1 + \cdots + X_k}{c_n} + \sup_{m\geq n+k} \frac{X_{k+1} + \cdots + X_m}{c_m} \Big) \\
&= \lim_{n\to\infty} \sup_{m\geq n+k} \frac{X_{k+1} + \cdots + X_m}{c_m}.
\end{aligned}$$

The latter random variable is measurable with respect to $\sigma(X_{k+1}, X_{k+2}, \ldots)$ for any $k > 1$. Therefore, this random variable is measurable with respect to $\mathcal{T} := \cap_{k\geq 1}\sigma(X_{k+1}, X_{k+2}, \ldots)$. In particular, for any $t > 0$, the event that $\limsup_{n\to\infty} S_n/c_n > t$ is in $\mathcal{T}$.

□

5. Homework 5

Exercise 5.2. Let $0 < p \leq \infty$. Show that, if $Y_1, Y_2, \ldots : \Omega \to \mathbb{R}$ converge to $Y : \Omega \to \mathbb{R}$ in L_p, then $Y_1, Y_2, \ldots$ converges to Y in probability.

Then, show that the converse is false.

Solution. The first part follows from Markov's inequality since $\mathbf{P}(|Y_n - Y| > \varepsilon) \leq \varepsilon^{-p}\mathbf{E}\,|Y_n - Y|^p$ for all $\varepsilon > 0$, $\forall\ n \geq 1$. For the second part, fix $0 < p < \infty$. Consider $\Omega = [0, 1]$, $\mathbf{P}$ uniform on Ω and $Y_n(t) := n^{1+1/p}1_{(0,1/n]}(t)$ for all $n \geq 1\ \forall\ t \in [0, 1]$. Then $Y_1, Y_2, \ldots$ converges in probability to 0, since $\mathbf{P}(Y_n \neq 0) = 1/n$ for all $n \geq 1$, but Y_1, Y_2 does not converge in L_p since $\mathbf{E}\,|Y_n|^p = n^p \to \infty$ as $n \to \infty$. So, if $Y \in L_p$ then $\|Y_n - Y\|_p \geq \|Y_n\|_p - \|Y\|_p \to \infty$ as $n \to \infty$. □

Exercise 5.3. Suppose random variables $Y_1, Y_2, \ldots : \Omega \to \mathbb{R}$ converge in probability to a random variable $Y : \Omega \to \mathbb{R}$. Prove that $Y_1, Y_2, \ldots$ converge in distribution to Y.

Then, show that the converse is false.

Solution. For the converse, see e.g. Proposition 3.1 in the notes. □

Exercise 5.4. Prove the following statement. Almost sure convergence does not imply convergence in L_2, and convergence in L_2 does not imply almost sure convergence. That is, find random variables that converge in L_2 but not almost surely. Then, find random variables that converge almost surely but not in L_2.

Solution. We first find random variables that converge in L_2 but not almost surely. We can do this by a "traveling bump" construction with "decreasing width." Let $\Omega = [0, 1]$ with $\mathbf{P}$ uniform on Ω. For any integer $n \geq 1$, write $n = 2^j + b$ where b is a positive integer with $b < 2^j$ and $j \geq 0$ is a nonnegative integer. Let $X_n : [0, 1] \to \{0, 1\}$ be the function

$$X_n := 1_{[2^{-j}b, 2^{-j}(b+1)]}.$$

As $n \to \infty$, $j \to \infty$, so $\mathbf{E}X_n^2 \to 0$ as $n \to \infty$, so that $X_1, X_2, \ldots$ converges to 0 in L_2. However, $X_1, X_2, \ldots$ does not converge almost surely since, for any $t \in [0, 1]$, the sequence $X_1(t), X_2(t), \ldots$ has infinitely many zero and one values.

To find random variables that converge almost surely but not in L_2, we re-use a previous construction. Consider $\Omega = [0, 1]$, $\mathbf{P}$ uniform on Ω and $Y_n(t) := n1_{(0,1/n]}(t)$ for all $n \geq 1\ \forall\ t \in [0, 1]$. Then $Y_1, Y_2, \ldots$ converges almost surely to 0, $\lim_{n\to\infty} Y_n(t) = 0$ for all $t \in [0, 1]$, but Y_1, Y_2 does not converge in L_2 since $\mathbf{E}\,|Y_n|^2 = n \to \infty$ as $n \to \infty$. So, if $Y \in L_2$ then $\|Y_n - Y\|_2 \geq \|Y_n\|_2 - \|Y\|_2 \to \infty$ as $n \to \infty$. □

Exercise 5.5 (Optional)**.** Let $X, X_1, X_2, \ldots : \Omega \to \mathbb{R}$.

(i) Suppose that $\sum_{i=1}^{\infty} \mathbf{P}(|X_i - X| > \varepsilon) < \infty$ for all $\varepsilon > 0$. Show that $X_1, X_2, \ldots$ converges to X almost surely. Show that the converse does not hold in general.

(ii) Suppose $X_1, X_2, \ldots$ converges to X in probability. Show there is a subsequence $X_{i_1}, X_{i_2}, \ldots$ of $X_1, X_2, \ldots$ such that $X_{i_1}, X_{i_2}, \ldots$ converges to X almost surely. (Here $i_1 < i_2 < \cdots$)

(iii) (Urysohn subsequence principle) Suppose that every subsequence $X_{i_1}, X_{i_2}, \dots$ of $X_1, X_2, \dots$ has a further subsequence $X_{i_{j_1}}, X_{i_{j_2}}, \dots$ that converges to X in probability. Show that $X_1, X_2, \dots$ also converges to X in probability.

(iv) Suppose $X_1, X_2, \dots$ converges in probability. Let $F\colon \mathbb{R} \to \mathbb{R}$ be continuous. Show that $F(X_1), F(X_2), \dots$ converges in probability to $F(X)$. More generally, suppose $\forall\ 1 \le j \le k$, $X_1^{(j)}, X_2^{(j)}, \dots \colon \Omega \to \mathbb{R}$ is a sequence of random variables that converge in probability to $X^{(j)}$. Let $F\colon \mathbb{R}^k \to \mathbb{R}$ be continuous. Show that $F(X_i^{(1)}, \dots, X_i^{(k)})$ converges in probability to $F(X^{(1)}, \dots, X^{(k)})$. For example, if $k = 2$, then $X_1^{(1)} + X_1^{(2)}, X_1^{(2)} + X_2^{(2)}, \dots$ converges in probability to $X^{(1)} + X^{(2)}$, and $X_1^{(1)} \cdot X_1^{(2)}, X_1^{(2)} \cdot X_2^{(2)}, \dots$ converges in probability to $X^{(1)} \cdot X^{(2)}$.

(v) (Fatou's lemma for convergence in probability) If $X_1, X_2, \dots \colon \Omega \to [0, \infty)$ converges in probability to X, show that $\mathbf{E}X \le \liminf_{n\to\infty} \mathbf{E}X_n$.

(vi) (Dominated convergence in probability) If $X_1, X_2, \dots$ converge in probability to X, and there exists a random variable $Y\colon \Omega \to [0, \infty)$ such that, for any $n \ge 1$, $|X_n| \le Y$ and $\mathbf{E}Y < \infty$, then $\lim_{n\to\infty} \mathbf{E}X_n = \mathbf{E}X$.

Solution. Part (i) follows from Borel-Cantelli. For the converse, re-use the example from Exercise 5.4.

Part (ii) follows by taking, for any $\varepsilon > 0$, a subsequence $\{X_{i_{n,\varepsilon}}\}$ such that $\sum_{i=1}^{\infty} \mathbf{P}(\left|X_{i_{n,\varepsilon}} - X\right| > \varepsilon) < \infty$, and then applying a Cantor diagonalization argument.

Part (iii) follows since any sequence of real numbers in $[0, 1]$ has a convergent subsequence.

The remaining parts follow by applying the previous parts. □

Exercise 5.8. A random variable $X\colon \Omega \to \mathbb{R}$ is said to be in weak L_1 if

$$\sup_{t>0} t\,\mathbf{P}(|X| > t) < \infty.$$

For example, a Cauchy distributed random variable X has density $f(x) = \frac{1}{\pi(1+x^2)}$ for any $x \in \mathbb{R}$, and X is in weak L_1 while $\mathbf{E}\,|X| = \infty$.

Show that, if $X_1, X_2, \dots \colon \Omega \to (0, \infty)$ are i.i.d. such that X_1 is in weak L_1, then there exist real numbers $a_1, a_2, \dots$ such that $\lim_{n\to\infty} a_n = \infty$ such that $\frac{1}{a_n}(X_1 + \cdots + X_n)$ converges in probability to 1.

(Hint: If you want to build up your intuition, assume $\mathbf{P}(X_1 > t) = 1/t$ for all $t > 2$, and use $b_n := n \log n$ in the Weak Law for Triangular Arrays.)

(Hint: Let $f(s) := \mathbf{E}X_1 1_{X_1 \le s}$ for any $s > 0$. Note that $f(s)/s = \int_0^s (1/s)\mathbf{P}(X > t)dt = \int_0^1 \mathbf{P}(X > sx)dx \to 0$ as $s \to \infty$ by the Bounded Convergence Theorem. Choose $b_1 \le b_2 \le \cdots$ going to infinity such that $nf(b_n) \le b_n$ for all large $n \ge 1$ as follows. When n is fixed and large, $nf(s)/s$ is larger than 1, and it converges to 0 as $s \to \infty$. Also, $nf(s)/s$ is right continuous in s, so let $b_n := \inf\{s > 0\colon nf(s)/s \le 1\}$. Assume $\mathbf{E}X_1 = \infty$. Note that $\lim_{s\to\infty} \frac{f(s)}{s\mathbf{P}(X_1 > s)} = \infty$, so $\infty = \lim_{n\to\infty} \frac{f(b_n)}{b_n \mathbf{P}(X_1 > b_n)} = \lim_{n\to\infty} \frac{1}{n\mathbf{P}(X_1 > b_n)}$, i.e. $\lim_{n\to\infty} n\mathbf{P}(X_1 > b_n) = 0$. Now, use the Weak Law for Triangular arrays. Note that $\lim_{n\to\infty} \frac{b_n}{n} = \lim_{n\to\infty} f(b_n) = \lim_{s\to\infty} f(s) = \infty$, using $\mathbf{E}X_1 = \infty$.)

Solution. First, note that if $\mathbf{E}X_1 < \infty$, or more generally if $\lim_{x\to\infty} x\mathbf{P}(X_1 > x) = 0$, then we choose $a_n := n\mu_n$ where $\mu_n := \mathbf{E}X_1 1_{X_1 \le n}$, and we are done by the Weak Law of Large Numbers. So, we may assume e.g. that $\mathbf{E}X_1 = \infty$, in which case $\lim_{n\to\infty} \frac{b_n}{n} = \infty$ and $\lim_{n\to\infty} n\mathbf{P}(X_1 > b_n) = 0$.

Repeating the proof of the Weak Law, with $X_{n,k} := X_k$ and $\overline{X}_{n,k} := X_k 1_{X_k < b_n}$,

$$\mathbf{E}\overline{X}_{n,k}^2 = \int_0^\infty 2t\mathbf{P}(|\overline{X}_{n,k}| > t)dt = \int_0^{b_n} 2t\mathbf{P}(|X_k| > t)dt = b_n^2 \int_0^1 2s\mathbf{P}(|X_1| > sb_n)ds.$$

$$b_n^{-2} \sum_{k=1}^n \mathbf{E}\overline{X}_{n,k}^2 = \int_0^1 2sn\mathbf{P}(|X_1| > sb_n)ds = \frac{n}{b_n} \int_0^1 2sb_n\mathbf{P}(|X_1| > sb_n)ds.$$

From the Bounded Convergence Theorem (since X_1 is in Weak L_1), and using $\lim_{n\to\infty} \frac{b_n}{n} = \infty$, we have

$$\lim_{n\to\infty} b_n^{-2} \sum_{k=1}^n \mathbf{E}\overline{X}_{n,k}^2 = 0.$$

So, we have verified the second assumption of the Weak Law for Triangular Arrays. To verify the first assumption, we similarly have

$$\sum_{k=1}^n \mathbf{P}(|X_{n,k}| > b_n) = \sum_{k=1}^n \mathbf{P}(|X_k| > b_n) = n\mathbf{P}(|X_1| > b_n).$$

So, as we mentioned in the hint, $\lim_{n\to\infty} \sum_{k=1}^n \mathbf{P}(|X_{n,k}| > b_n) = 0$. We can therefore conclude from the Weak Law for Triangular Arrays that

$$b_n^{-1} \sum_{k=1}^n (X_{n,k} - \mathbf{E}\overline{X}_{n,k})$$

converges to 0 in probability as $n \to \infty$. That is, $b_n^{-1} \sum_{k=1}^n (X_k - f(b_n))$ converges to 0 in probability as $n \to \infty$. That is, $\frac{X_1+\cdots+X_n}{b_n}$ converges to 1 in probability as $n \to \infty$, since $f(b_n)/b_n = 1/n$ for all $n \geq 1$. □

6. Homework 6

Exercise 6.3. For each $n \geq 1$, let $A_n = (a_{ij,n})_{1\leq i,j\leq n}$ be a random $n \times n$ matrix (i.e. a random variable taking values in the space $\mathbb{R}^{n\times n}$ or $\mathbb{C}^{n\times n}$ of $n \times n$ matrices) such that the entries $a_{ij,n}$ of A_n are independent in i, j and take values in $\{-1, 1\}$ with a probability of $1/2$ each. We do not assume any independence for the sequence $A_1, A_2, \ldots$.

(i) Show that the random variables $\mathrm{Tr}A_nA_n^*/n^2$ are equal to the constant 1, where A_n^* denotes the matrix adjoint (which, in this case, is also the transpose) of A_n and Tr denotes the trace (or sum of the diagonal entries) of a matrix.

(ii) Show that for any natural number $k \geq 1$, the quantities $\mathbf{E}\mathrm{Tr}(A_nA_n^*)^k/n^{k+1}$ are bounded uniformly in $n \geq 1$ (i.e. they are bounded by a quantity C_k that can depend on k but not on n). (It may be helpful to first try $k = 2$ and $k = 3$.)

(iii) Let $\|A_n\|$ denote the operator norm of A_n, and let $\varepsilon > 0$. Show that $\|A_n\|/n^{1/2+\varepsilon}$ converges almost surely to zero, and that $\|A_n\|/n^{1/2-\varepsilon}$ diverges almost surely to infinity. (Hint: use the spectral theorem to relate $\|A_n\|$ with the quantities $\mathrm{Tr}(A_nA_n^*)^k$.

Solution.

$$\mathrm{Tr}A_nA_n^* = \sum_{i,j=1}^n a_{ij}^2 = n^2.$$

Below, we consider addition in sub-indices as addition mod $2k$. Note that

$$\mathbf{E}\mathrm{Tr}(A_nA_n^*)^k = \sum_{1\le i_1,\dots,i_{2k}\le n} \mathbf{E}\prod_{j=1}^{k} a_{i_{2j-1}i_{2j}}a_{i_{2j+1}i_{2j}}.$$

Since $\mathbf{E}a_{ij} = 0$ for all $1 \le i, j \le n$, if any a_{ij} term appears exactly once in the expected value, independence implies that that term has expected value zero. So, if any a_{ij} term appears once in the product, it must appear at least twice. So, although n^{2k} terms appear in the above sum, the number of nonzero terms in the sum is at most $O(n^{k+1})$. (In order to have repeated terms, it follows by induction that there have to be at least $k-1$ equality constraints on the indices being summed, and $2k-(k-1) = k+1$.) Meanwhile, since $|a_{ij}| = 1$ for all $1 \le i, j \le n$, we have $|\mathbf{E}\prod_{j=1}^{k} a_{i_{2j-1}i_{2j}}a_{i_{2j+1}i_{2j}}| \le 1$ for any $1 \le i_1, \dots, i_{2k} \le n$. So, $\mathbf{E}\mathrm{Tr}(A_nA_n^*)^k \le O(n^{k+1})$, where the implied constant can depend on k but not on n.

Let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ let $k \ge 1$ and denote $\|x\|_k := (\sum_{i=1}^n |x_i|^k)^{1/k}$, $\|x\|_\infty := \max_{1\le i\le n} |x_i|$.

Claim 1. For any $x \in \mathbb{R}^n$, $\|x\|_\infty \le \|x\|_{2k} \le n^{1/(2k)} \|x\|_\infty$

Proof of Claim 1. $\|x\|_\infty \le \|x\|_{2k} \le (n\|x\|_\infty^{2k})^{1/(2k)} = n^{1/(2k)}\|x\|_\infty$. The Claim is proven.

From the Spectral Theorem, if $\lambda_{(n)}$ denotes the vector of square roots of eigenvalues of $A_nA_n^*$, then

$$\mathrm{Tr}(A_nA_n^*)^k = \left\|\lambda_{(n)}\right\|_{2k}^{2k}. \qquad (*)$$

Note that we can write $A_nA_n^* = Q_nD_nQ_n^{-1}$ where Q_n is an orthogonal $n\times n$ (random) matrix and D_n is a diagonal real $n \times n$ (random) matrix containing the eigenvalues of $A_nA_n^*$, so $(A_nA_n^*)^k = Q_nD_n^kQ_n^{-1}$, and then

$$\mathrm{Tr}(A_nA_n^*)^k = \mathrm{Tr}Q_nD_n^kQ_n^{-1} = \mathrm{Tr}Q_n^{-1}Q_nD_n^k = \mathrm{Tr}D_n^k = \left\|\lambda_{(n)}\right\|_{2k}^{2k}.$$

Combining $(*)$ and Claim 1, and noting that $\|A_n\|$ is the maximum element of $\lambda_{(n)}$,

$$\|A_n\| \le (\mathrm{Tr}(A_nA_n^*)^k)^{1/(2k)} \le n^{1/(2k)}\|A_n\|. \qquad (**)$$

Dividing by $n^{1/2+\varepsilon}$, we get

$$\frac{\|A_n\|}{n^{1/2+\varepsilon}} \le \Big(\frac{\mathrm{Tr}(A_nA_n^*)^k}{n^{k(1+2\varepsilon)}}\Big)^{1/(2k)} = \Big(\frac{\mathrm{Tr}(A_nA_n^*)^k}{n^{k+1}}n^{1-2k\varepsilon}\Big)^{1/(2k)}.$$

Fix $\varepsilon > 0$. Choose $k > 0$ such that $1 - 2k\varepsilon < 0$. Then, by the first part of the problem and Jensen's inequality,

$$\mathbf{E}\frac{\|A_n\|}{n^{1/2+\varepsilon}} \le \Big(\mathbf{E}\frac{\mathrm{Tr}(A_nA_n^*)^k}{n^{k+1}}n^{1-2k\varepsilon}\Big)^{1/(2k)} \to 0 \text{ as } n\to\infty.$$

Since $\|A_n\|$ is a nonnegative random variable, we conclude that $\|A_n\|/n^{1/2+\varepsilon}$ converges to zero almost surely as $n \to \infty$. Meanwhile, dividing the right inequality of $(**)$ by $n^{1/2-\varepsilon}$,

$$\frac{\|A_n\|}{n^{1/2-\varepsilon}} \ge \Big(\frac{\mathrm{Tr}(A_nA_n^*)^k}{n^{k+1}}n^{2k\varepsilon}\Big)^{1/(2k)}.$$

Using e.g. $k = 1$, we get

$$\frac{\|A_n\|}{n^{1/2-\varepsilon}} \ge \Big(\frac{\mathrm{Tr}(A_nA_n^*)}{n^2}n^{2\varepsilon}\Big)^{1/2} = n^{\varepsilon}.$$

Letting $n \to \infty$ shows that $\|A_n\|/n^{1/2-\varepsilon}$ converges to ∞ almost surely to ∞. □

Exercise 6.4. The Cramér random model for the primes is a random subset $\mathcal{P}$ of the natural numbers such that $1 \notin \mathcal{P}$, $2 \in \mathcal{P}$, and the events $n \in \mathcal{P}$ for $n = 3, 4, \ldots$ are independent with $\mathbf{P}(n \in \mathcal{P}) := \frac{1}{\log n}$. Here we used the restriction $n \geq 3$ so that $\frac{1}{\log n} < 1$. This random set of integers $\mathcal{P}$ gives a reasonable way to model the primes $2, 3, 5, 7, \ldots$, since by the Prime Number Theorem, the number of primes less than n is approximately $n/\log n$, so the probability of n being a prime should be about $1/\log n$. The Cramér random model can provide heuristic confirmations for many conjectures in analytic number theory:

- (Probabilistic prime number theorem) Show that $\frac{1}{x/\log x}|\{n \leq x : n \in \mathcal{P}\}|$ converges almost surely to one as $x \to \infty$.
- (Probabilistic Riemann hypothesis) Let $\varepsilon > 0$. Show that
$$\frac{1}{x^{1/2+\varepsilon}}\Big(|\{n \leq x : n \in \mathcal{P}\}| - \int_2^x \frac{dt}{\log t}\Big)$$
converges almost surely to zero as $x \to \infty$.
- (Probabilistic twin prime conjecture) Show that almost surely, there are an infinite number of elements p of $\mathcal{P}$ such that $p + 2$ also lies in $\mathcal{P}$.
- (Probabilistic Goldbach conjecture) Show that almost surely, all but finitely many natural numbers n are expressible as the sum of two elements of $\mathcal{P}$.

Proof. **Part 1**. For any $n \geq 3$, let $X_n := 1_{n\in\mathcal{P}} - \frac{1}{\log n}$. Fix $x > 0$ and consider the random variable
$$U_x := \sum_{n\leq x} \frac{\log n}{n} X_n.$$
Note that $\sum_{n=3}^{\infty} \operatorname{var}\Big(\frac{\log n}{n} X_n\Big) \leq 10 \sum_{n=3}^{\infty} \frac{(\log n)^2}{n^2} < \infty$. So, from Theorem 2.25 (Convergence of Random Series), U_x converges almost surely as $x \to \infty$. From Kronecker's Lemma, Exercise 6.6, with $b_n := n/\log n$,
$$\frac{1}{x/\log x} \sum_{n\leq x} X_n \to 0 \text{ as } x \to \infty. \qquad (*)$$
It remains to consider the constant terms in the definition of X_n. We have
$$\sum_{n=3}^{x} \frac{1}{\log n} = O(1) + \int_3^x \frac{1}{\log y} dy.$$
If $g(y) = y/\log y$ for any $y > 0$, then $g'(y) = 1/\log y - 1/(\log y)^2$, so
$$\int_3^x \frac{1}{\log y} dy = O(1) + x/\log x + \int_3^x \frac{1}{(\log y)^2} dy = O(1) + x/\log x + o(x/\log x).$$
It then follows from $(*)$ that, almost surely,
$$\frac{1}{x/\log x} \sum_{n\leq x} 1_{n\in\mathcal{P}} \to 1 \text{ as } x \to \infty.$$

Part 2. We repeat the above argument for T_x, where for any $x > 0$ we have
$$T_x := \sum_{n\leq x} \frac{\log n}{n^{1/2+\varepsilon}} X_n.$$

(Alternatively, for parts 1 and 2, we could use a variant of the Chernoff inequality.)

Part 3. For any $n \geq 3$, let $A_n := \{n \in \mathcal{P}, n+2 \in \mathcal{P}\}$. By the definition of $\mathcal{P}$,

$$\mathbf{P}(A_n) = \mathbf{P}(n \in \mathcal{P})\mathbf{P}(n+2 \in \mathcal{P}) = \frac{1}{\log n}\frac{1}{\log(n+2)}.$$

So, $\sum_{n\geq 3} \mathbf{P}(A_{4n}) = \infty$. Also, the events $A_4, A_8, A_{12}, \ldots$ are independent. So, by the second Borel-Cantelli Lemma, with probability one, infinitely many of the events $A_4, A_8, A_{12}, \ldots$ occur.

Part 4. Note that any integer $n \geq 6$ can be written as $n = (n-m)+m$ where $2 \leq m \leq n-2$. So, fix $n \geq 6$ and for any $2 \leq m \leq n-2$, let $X_m := 1_{m\in\mathcal{P},(n-m)\in\mathcal{P}}$. Assume for now that n is odd. Define $Y_n := \max(X_2, \ldots, X_{(n-1)/2})$. Then the event $Y_n = 1$ is equal to the event that n can be expressed as a sum of two prime numbers and $\{Y_n = 0\} = \{Y_n = 1\}^c$. Using independence and the definition of Y_n,

$$\mathbf{P}(Y_n = 0) = \mathbf{P}(X_2 = 0, \ldots, X_{(n-1)/2} = 0) = \mathbf{P}(X_2 = 0)\cdots\mathbf{P}(X_{(n-1)/2} = 0).$$

So, by the definition of $\mathcal{P}$, when n is odd,

$$\mathbf{P}(Y_n = 0) = \prod_{m=2}^{(n-1)/2} \frac{1}{\log m}\frac{1}{\log(n-m)} = \frac{1}{\prod_{m=3}^{n-2}\log m}.$$

Similarly, when n is even

$$\mathbf{P}(Y_n = 0) = \frac{1}{\prod_{m=3}^{n-2}\log m}.$$

By the Borel-Cantelli Lemma, we will be done once we show that the following sum is finite

$$\sum_{n=6}^{\infty} \mathbf{P}(Y_n = 0) = \sum_{n=6}^{\infty} \frac{1}{\prod_{m=3}^{n-2}\log m}.$$

Note that once $m > 10$, $\log m > 2$, so $\frac{1}{\prod_{m=3}^{n-2}\log m} \leq 100 \cdot 2^{-(n-10)}$, so that

$$\sum_{n=6}^{\infty} \mathbf{P}(Y_n = 0) \leq 10^3 \sum_{n=6}^{\infty} 2^{-(n-10)} < \infty.$$

□

Exercise 6.5. This exercise proves the Hardy-Ramanujan Theorem. This theorem, with probabilistic proof due to Turán, says that a typical large $n \in \mathbb{N}$ has about $\log\log n$ distinct prime factors. Unlike the previous exercise, the probabilistic proof here proves a rigorous result about primes.

Let $\mathcal{P} \subseteq \mathbb{N}$ denote the set of prime numbers (in this exercise $\mathcal{P}$ is deterministic, not random). When $p \in \mathcal{P}$ and $n \in \mathbb{N}$, we use the notation $p|n$ to denote "p divides n," i.e. n/p is a positive integer. Let $x \geq 100$ with $x \in \mathbb{N}$ (so that $\log\log x \geq 1$), and let N be a natural number that is uniformly distributed in $\{1, 2, \ldots, x\}$. Assume Mertens' theorem

$$\sum_{p\in\mathcal{P}:\, p\leq x} \frac{1}{p} = \log\log x + O(1).$$

- Show that the random variable $\sum_{p\in\mathcal{P}:\ p\le x^{1/10}} 1_{p|N}$ has mean $\log\log x + O(1)$ and variance $O(\log\log x)$. (Hint: up to reasonable errors, compute the means, variances and covariances of the random variables $1_{p|N}$.)
- For any $n \in \mathbb{N}$, let $f(n)$ denote the number of distinct prime factors of n. Show that $\frac{f(N)}{\log\log N}$ converges to 1 in probability as $x \to \infty$. (Hint: first show that $f(N) = \sum_{p\in\mathcal{P}:\ p\le x^{1/10}} 1_{p|N} + O(1)$.) More precisely, show that

$$\frac{f(N) - \log\log N}{g(N)\sqrt{\log\log N}}$$

converges in probability to zero as $x \to \infty$, whenever $g\colon \mathbb{N} \to \mathbb{R}$ is any function satisfying $\lim_{n\to\infty} g(n) = \infty$.

Proof. Note that $\mathbf{E}1_{p|N} = \mathbf{P}(p|N) = \frac{|\{p,2p,3p,\ldots,\lfloor x/p\rfloor\}|}{x} = \frac{\lfloor x/p\rfloor}{x}$. Writing $x = ap + r$ where $a, r > 0$ are integers with $r < p$, we have $\mathbf{E}1_{p|N} = \frac{a}{ap+r} = \frac{a}{ap+r} = \frac{1}{p+r/a}$. Since $p + r/a \le p(1 + 1/a)$ and $a \ge x/p - 1$, $1 + 1/a \le 1 + 1/[x/p - 1]$, so $p + r/a \le p(1 + 1/[x/p - 1]) = p(1 + p/[x - p]) = px/(x - p) = p/(1 - p/x)$, and

$$(1 - p/x)\frac{1}{p} \le \mathbf{E}1_{p|N} \le \frac{1}{p}.$$

That is,

$$\mathbf{E}1_{p|N} = \frac{1}{p} + O(1/x),$$

where the O term does not depend on p. Consequently,

$$\mathrm{var}1_{p|N} = \mathbf{E}1_{p|N} - (\mathbf{E}1_{p|N})^2 = \frac{1}{p} - \frac{1}{p^2} + \frac{1}{p}O(1/x) + O(1/x^2).$$

Also,

$$\begin{aligned}\mathrm{cov}(1_{p|N}, 1_{q|N}) &= \mathbf{E}1_{pq|N} - \mathbf{E}1_{p|N}\mathbf{E}1_{q|N} = \frac{1}{pq} + O(1/x) - (\frac{1}{p} + O(1/x))(\frac{1}{q} + O(1/x)) \\ &= \Big(\frac{1}{p} + \frac{1}{q}\Big)O(1/x) + O(1/x^2).\end{aligned}$$

So, adding these up and using Mertens' Theorem in the form

$$\sum_{p\in\mathcal{P}:\ p\le x^{1/10}} \frac{1}{p} = \log\log(x^{1/10}) + O(1) = \log\log x + \log(1/10) + O(1) = \log\log x + O(1),$$

we get

$$\begin{aligned}\mathrm{var}\Big(\sum_{p\in\mathcal{P}:\ p\le x^{1/10}} 1_{p|N}\Big) &= \sum_{p\in\mathcal{P}:\ p\le x^{1/10}} \Big(\frac{1}{p} - \frac{1}{p^2} + \frac{1}{p}O(1/x) + O(1/x^2)\Big) \\ &\qquad + \sum_{p,q\in\mathcal{P}:\ p,q\le x^{1/10}} \Big(\Big(\frac{1}{p} + \frac{1}{q}\Big)O(1/x) + O(1/x^2)\Big) \\ &\le [\log\log x + O(1)][1 + O(1/x)] + x^{1/10}O(1/x^2) \\ &\qquad + 2x^{1/10}[\log\log x + O(1)]O(1/x) + x^{1/10}O(1/x^2) \\ &= \log\log x + O(1).\end{aligned}$$

Lastly, by Mertens' Theorem

$$\mathbf{E}\sum_{p\in\mathcal{P}:\ p\le x^{1/10}} 1_{p|N} = \sum_{p\in\mathcal{P}:\ p\le x^{1/10}} \Big(\frac{1}{p}+O(1/x)\Big) = \log\log x+O(1)+x^{1/10}O(1/x) = \log\log x+O(1).$$

Now, let $p_1, p_2, \ldots, p_{11}$ be distinct prime factors larger than $x^{1/10}$. Since $\prod_{i=1}^{11} p_i > x$, it cannot be the case that an integer n with $n \le x$ has 11 or more prime factors larger than $x^{1/10}$. So, by definition of N, it cannot be the case that N has 11 or more prime factors larger than $x^{1/10}$. That is, with probability one,

$$\sum_{p\in\mathcal{P}:\ p\ge x^{1/10}} 1_{p|N} \le 11.$$

So, by definition of f,

$$f(N) = \sum_{p\in\mathcal{P}:\ p\le x} 1_{p|N} = O(1) + \sum_{p\in\mathcal{P}:\ p\le x^{1/10}} 1_{p|N}.$$

From the first part of the exercise, $f(N)/\log\log x$ has mean $1+O(1/\log\log x)$ and variance $O(1/\log\log x)$, so Chebyshev's inequality implies that

$$\mathbf{P}\Big(\Big|\frac{f(N)}{\log\log x} - 1 + O(1/\log\log x)\Big| > \varepsilon\Big) \le \varepsilon^{-2}O(1/\log\log x), \qquad \forall\,\varepsilon > 0.$$

Letting $x\to\infty$ shows that $f(N)/\log\log x$ converges in probability to 1 as $x\to\infty$. Also, $1_{N\ge 100}\log\log N/\log\log x$ converges in probability to 1 as $x\to\infty$, so

$$f(N)/\log\log N = \frac{f(N)}{\log\log x}\frac{\log\log x}{\log\log N}$$

converges in probability to 1 as $x\to\infty$. [For any positive integer m, we have $\log\log x^{1/m}/\log\log x = 1+\log(1/m)/\log\log x$ which goes to 1 as $x\to\infty$, implying that $1_{N\ge 100}\log\log N/\log\log x$ converges in probability to 1 as $x\to\infty$.]

More precisely, from the first part of the exercise,

$$\frac{f(N)-\log\log x}{g(x)\sqrt{\log\log x}}$$

has mean $O(1/g(x)\sqrt{\log\log x})$ and variance $O(1/g(x))$, so Chebyshev's inequality implies

$$\mathbf{P}\Big(\Big|\frac{f(N)-\log\log x}{g(x)\sqrt{\log\log x}} + O(1/g(x)\sqrt{\log\log x})\Big| > \varepsilon\Big) \le \varepsilon^{-2}O(1/g(x)), \qquad \forall\,\varepsilon > 0.$$

That is,

$$\frac{f(N)-\log\log x}{g(x)\sqrt{\log\log x}}$$

converges in probability to 0 as $x\to\infty$. As before, we can then extend this conclusion to

$$\frac{f(N)-\log\log N}{g(N)\sqrt{\log\log N}}$$

□

Exercise 6.6 (**Kronecker's Lemma**). Let $y_1, y_2, \ldots$ be a sequence of real numbers. Let $0 < b_1 \leq b_2 \leq \cdots$ be a sequence of real numbers that goes to infinity. Assume that $\lim_{n\to\infty} \sum_{m=1}^{n} y_m$ exists. Then $\lim_{n\to\infty} \frac{1}{b_n} \sum_{m=1}^{n} b_m y_m = 0$. (Hint: if $s_n := \sum_{m=1}^{n} y_m$, then the summation by parts formula implies that $\frac{1}{b_n} \sum_{m=1}^{n} b_m y_m = s_n - \frac{1}{b_n} \sum_{m=1}^{n-1} (b_{m+1} - b_m) s_m$.)

Solution. Let $\varepsilon > 0$. Let $s := \sum_{m=1}^{\infty} y_m \in \mathbb{R}$. Choose $p > 0$ such that $|s - s_n| < \varepsilon$ for all $n > p$. Then, splitting the sum in the summation by parts formula into two parts, then adding and subtracting s, we get

$$\begin{aligned}
\frac{1}{b_n} \sum_{m=1}^{n} b_m y_m &= s_n - \frac{1}{b_n} \sum_{m=1}^{p-1} (b_{m+1} - b_m) s_m - \frac{1}{b_n} \sum_{m=p}^{n-1} (b_{m+1} - b_m) s_m \\
&= s_n - \frac{1}{b_n} \sum_{m=1}^{p-1} (b_{m+1} - b_m) s_m - \frac{1}{b_n} \sum_{m=p}^{n-1} (b_{m+1} - b_m)(s_m - s) - \frac{1}{b_n} \sum_{m=p}^{n-1} (b_{m+1} - b_m) s \\
&= s_n - \frac{1}{b_n} \sum_{m=1}^{p-1} (b_{m+1} - b_m) s_m - \frac{1}{b_n} \sum_{m=p}^{n-1} (b_{m+1} - b_m)(s_m - s) - \frac{1}{b_n} (b_n - b_p) s \\
&= s_n - \frac{1}{b_n} \sum_{m=1}^{p-1} (b_{m+1} - b_m) s_m - \frac{1}{b_n} \sum_{m=p}^{n-1} (b_{m+1} - b_m)(s_m - s) - (1 - b_p/b_n) s
\end{aligned}$$

As $n \to \infty$, $b_n \to \infty$, so the last term converges to s. By definition, $\lim_{n\to\infty} s_n = s$, so the first and last terms cancel. The third term is bounded in absolute value by

$$\frac{1}{b_n} \sum_{m=p}^{n-1} (b_{m+1} - b_m) \varepsilon = \varepsilon \frac{b_n - b_p}{b_n} \leq \varepsilon.$$

The second term goes to zero as $n \to \infty$ since there are only a finite number of terms divided by b_n. In conclusion, $\frac{1}{b_n} \sum_{m=1}^{n} b_m y_m \to 0$ as $n \to \infty$. □

Exercise 6.10 (**Cheap Law of the Iterated Logarithm**). Let $X_1, X_2, \ldots : \Omega \to \mathbb{R}$ be random variables with mean zero and variance one. The Strong Law of Large Numbers says that $\frac{1}{n}(X_1 + \cdots + X_n)$ converges almost surely to zero (if the random variables are also identically distributed). The Central Limit Theorem says that $\frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)$ converges in distribution to a standard Gaussian random variable. But what happens if we divide by some function of n in between $n^{1/2}$ and n ? This Exercise gives a partial answer to this question.

Let $\varepsilon > 0$. Show that

$$\frac{X_1 + \cdots + X_n}{n^{1/2} (\log n)^{(1/2)+\varepsilon}}$$

converges to zero almost surely as $n \to \infty$. (Hint: Re-do the proof of the Strong Law of Large Numbers, but divide by $n^{1/2}(\log n)^{(1/2)+\varepsilon}$ instead of n. You don't need to do any truncation.)

Solution. We are required to show that $\frac{X_1+\cdots+X_n}{n}$ converges to 0 almost surely. By integral comparison, and the substitution $u = \log x$, $du = (1/x)dx$, note that

$$\sum_{m=1}^{\infty} m^{-1}(\log m)^{-1-2\varepsilon} = O(1) + \int_3^{\infty} \frac{1}{x(\log x)^{1+2\varepsilon}} dx = O(1) + \int_{\log 3}^{\infty} u^{-1-2\varepsilon} du < \infty.$$

So,

$$\sum_{m=1}^{\infty} \operatorname{var}\Big(\frac{X_m}{m^{1/2}(\log m)^{1/2+\varepsilon}}\Big) = \sum_{m=1}^{\infty} \frac{\operatorname{var}(X_m)}{m(\log m)^{1+2\varepsilon}} = \sum_{m=1}^{\infty} m^{-1}(\log m)^{-1-2\varepsilon} < \infty.$$

So, by Theorem 2.25 (Convergence of Random Series), $\sum_{m=1}^{\infty} \frac{X_m}{m^{1/2}(\log m)^{1/2+\varepsilon}}$ converges almost surely. Let $b_m := m^{1/2}(\log m)^{1/2+\varepsilon}$ for any $m \geq 3$. From Kronecker's Lemma, Exercise 6.6, we conclude that $\frac{1}{\sqrt{n}(\log n)^{1/2+\varepsilon}} \sum_{m=1}^{n} X_m$ converge to 0 almost surely as $n \to \infty$. □

7. Homework 7

Exercise 7.2 (**Chernoff Inequality**)**.** Let $0 < p < 1$. Let $X_1, X_2, \ldots$ be independent identically distributed random variables with $\mathbf{P}(X_1 = 1) = p$ and $\mathbf{P}(X_1 = 0) = 1 - p$ for any $i \geq 1$. Then for any $n \geq 1$

$$\mathbf{P}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i \geq t\Big) \leq e^{-np}\Big(\frac{ep}{t}\Big)^{tn}, \qquad \forall\, t \geq p.$$

Prove the same estimate for $\mathbf{P}(\frac{1}{n}\sum_{i=1}^{n} X_i \leq t)$ for any $t \leq p$. (Hint: $1 + x \leq e^x$ for any $x \in \mathbb{R}$, so $1 + (e^{\alpha} - 1)p \leq e^{(e^{\alpha}-1)p}$.)

Exercise 7.3 (Optional)**.** We return to the Erdös-Renyi random graph $G = (V, E)$ on n vertices with parameter $0 < p < 1$ from an earlier homework. Define $d := p(n-1)$.

- Show that d is the expected degree of each vertex in G. (The degree of a vertex $v \in V$ is the number of vertices connected to v by an edge in E.)
- Show that there exists a constant $c > 0$ such that the following holds. Assume $p \geq \frac{c \log n}{n}$. Then with probability larger than .9, all vertices of G have degrees in the range $(.9d, 1.1d)$. (Hint: first consider a single vertex, then use the union bound over all vertices.)

Solution. The degree Y_i of a single vertex $1 \leq i \leq n$ is a binomial random variable with parameters $n - 1$ and p. That is, Chernoff's inequality implies that

$$\mathbf{P}\Big(\frac{1}{n-1}Y_i \geq t\Big) \leq e^{-(n-1)p}\Big(\frac{ep}{t}\Big)^{t(n-1)}, \qquad \forall\, t \geq p.$$

Taking the union bound over all $1 \leq i \leq n$,

$$\mathbf{P}\Big(\forall\, 1 \leq i \leq n,\ \frac{1}{n-1}Y_i \geq t\Big) \leq ne^{-(n-1)p}\Big(\frac{ep}{t}\Big)^{t(n-1)}, \qquad \forall\, t \geq p.$$

Choosing $t = 1.1d/(n-1) = 1.1p$, so that $t/p = 1.1$, we get

$$\begin{aligned}\mathbf{P}\Big(\forall\, 1 \le i \le n,\ \frac{1}{n-1}Y_i \ge t\Big) &\le ne^{-(n-1)p}\Big(\frac{e}{1.1}\Big)^{1.1p(n-1)} = n(1.1)^{-p(n-1)}e^{(.1)p(n-1)} \\ &= ne^{-p(n-1)(1.1)\log(1.1)}e^{(.1)p(n-1)} = ne^{-p(n-1)[(1.1)\log(1.1)-.1]} \\ &\le ne^{-p(n-1)(.001)}.\end{aligned}$$

Assuming $p \ge c\log n/n$, we then get

$$\mathbf{P}\Big(\forall\, 1 \le i \le n,\ \frac{1}{n-1}Y_i \ge t\Big) \le ne^{-c\log n(.001)} = n^{1-c(.001)}.$$

Choosing $c = 2000$ shows that this probability is exponentially small. We can get a similar inequality for the lower probability. □

Exercise 7.5. Let $X_1, X_2, \ldots : \Omega \to \mathbb{R}$ be i.i.d. with $\mathbf{E}\,|X_1| = \infty$. Then $\mathbf{P}(|X_n| > n$ for infinitely many $n \ge 1) = 1$. And $\mathbf{P}(\lim_{n\to\infty} \frac{X_1+\cdots+X_n}{n} \in (-\infty, \infty)) = 0$. (Hint: show $\sum_{n=1}^{\infty} \mathbf{P}(|X_n| > n) = \infty$, then apply the second Borel-Cantelli Lemma. Write $\frac{S_n}{n} - \frac{S_{n+1}}{n+1} = \frac{S_n}{n(n+1)} - \frac{X_{n+1}}{n+1}$, and consider what happens to both sides on the set where $\lim_{n\to\infty} \frac{S_n}{n} \in \mathbb{R}$.)

Solution.

$$\begin{aligned}\infty = \mathbf{E}\,|X_1| &= \int_0^\infty \mathbf{P}(|X_1| > t)dt = \sum_{n=1}^{\infty}\int_{n-1}^{n} \mathbf{P}(|X_1| > t)dt \\ &\le \sum_{n=1}^{\infty}\int_{n-1}^{n} \mathbf{P}(|X_1| > n)dt = \sum_{n=1}^{\infty}\mathbf{P}(|X_1| > n) = \sum_{n=1}^{\infty}\mathbf{P}(|X_n| > n).\end{aligned}$$

Since the events $\{X_1 > 1\}, \{X_2 > 2\}, \ldots$ are independent, we can apply the second Borel-Cantelli Lemma, Exercise 7.7, to conclude that $\mathbf{P}(|X_n| > n$ for infinitely many $n \ge 1) = 1$. Now, note that, for any $n \ge 1$,

$$\frac{S_n}{n} - \frac{S_{n+1}}{n+1} = \frac{S_n}{n(n+1)} - \frac{X_{n+1}}{n+1}$$

So, if $\lim_{n\to\infty} \frac{S_n}{n} \in \mathbb{R}$, then the left side is zero, $\lim_{n\to\infty} \frac{S_n}{n(n+1)} = 0$, and the remaining term is larger than one in absolute value, for infinitely many $n \ge 1$, with probability one. Therefore, $\lim_{n\to\infty} \frac{S_n}{n} \in \mathbb{R}$ with probability zero. □

Also, unfortunately the strong law cannot hold for triangular arrays.

Exercise 7.6. Let X be a random variable taking values in the natural numbers with $\mathbf{P}(X = n) = \frac{1}{\zeta(3)}\frac{1}{n^3}$, where $\zeta(3) := \sum_{m=1}^{\infty} \frac{1}{m^3}$.

- Show that X is absolutely integrable.
- For any $n \ge 1$, let $X_{n,1}, \ldots, X_{n,n} : \Omega \to \mathbb{R}$ be independent copies of X. Show that the random variables $\frac{X_{n,1}+\cdots+X_{n,n}}{n}$ are almost surely unbounded. (Hint: for any constant c, show that $\frac{X_{n,1}+\cdots+X_{n,n}}{n} > c$ occurs with probability at least ε/n for some $\varepsilon > 0$. Then use the second Borel-Cantelli lemma.)

Proof. Note that

$$\mathbf{E}\,|X| = \mathbf{E}X = \sum_{n=1}^{\infty} n\mathbf{P}(X = n) = \frac{1}{\zeta(3)}\sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{1}{\zeta(3)}\frac{\pi^2}{6} \approx 1.368 < \infty.$$

Let $S_n := \frac{X_{n,1}+\cdots+X_{n,n}}{n}$ for any $n \geq 1$. Then for any integer $c \geq 1$,

$$\begin{aligned}\mathbf{P}(S_n \geq c) &\geq \mathbf{P}(\max_{1\leq j\leq n} X_{n,j} \geq cn) = 1 - \mathbf{P}(\max_{1\leq j\leq n} X_{n,j} < cn) = 1 - [\mathbf{P}(X < cn)]^n \\ &= 1 - [1 - \mathbf{P}(X \geq cn)]^n = 1 - \Big[1 - o(1/(cn)^2) - \sum_{m=cn}^{\infty} \frac{1}{m^3}\frac{1}{\zeta(3)}\Big]^n \\ &= 1 - \Big[1 - o(1/(cn)^2) - \frac{1}{(cn)^2}\frac{1}{\zeta(3)}\Big]^n \geq \frac{1}{10cn} + o(1/cn).\end{aligned}$$

The second Borel Cantelli Lemma then implies that $\mathbf{P}(S_n > c \text{ for infinitely many } n \geq 1) = 1$, since the events $S_1 > c, S_2 > c, \ldots$ are independent. □

Exercise 7.7 (**Second Borel-Cantelli Lemma**)**.** Let $A_1, A_2, \ldots$ be independent events with $\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \infty$. Then $\mathbf{P}(A_n \text{ occurs for infinitely many } n \geq 1) = 1$. (Hint: using $1 - x \leq e^{-x}$ for any $x \in \mathbb{R}$, show $\mathbf{P}(\cap_{n=s}^{t} A_n^c) \leq \exp(-\sum_{n=s}^{t} \mathbf{P}(A_n))$, let $t \to \infty$ to conclude $\mathbf{P}(\cup_{n=s}^{\infty} A_n) = 1$ for all $s \geq 1$, then let $s \to \infty$.)

Exercise 7.8 (Optional)**.** Let $X, X_1, X_2, \ldots$ and let $Y, Y_1, Y_2, \ldots$ be random variables with values in $\mathbb{R}$.

(i) Assume that X is constant almost surely. Show that $X_1, X_2, \ldots$ converges to X in distribution if and only if $X_1, X_2, \ldots$ converges to X in probability.

(ii) Prove this Lemma from the notes: Let $\mu_1, \mu_2, \ldots$ be a sequence of probability measures on $\mathbb{R}$. Then any subsequential limit of the sequence (with respect to vague convergence) is a probability measure if and only if $\mu_1, \mu_2, \ldots$ is **tight**: $\forall\ \varepsilon > 0$, $\exists$ $m = m(\varepsilon) > 0$ such that

$$\limsup_{n\to\infty}(1 - \mu_n([-m, m])) \leq \varepsilon.$$

(iii) Suppose that $X_1, X_2, \ldots$ converges in distribution to X. Show there exist random variables $Z, Z_1, Z_2, \ldots : \Omega \to \mathbb{R}$ such that $\mu_Z = \mu_X$, $\mu_{Z_n} = \mu_{X_n}$ for any $n \geq 1$, and such that $Z_1, Z_2, \ldots$ converges almost surely to Z. (Hint: use the sample space $\Omega = [0, 1]$ and using an exercise from a previous homework, represent each random variable on Ω as the "inverse" of its cumulative distribution function.)

(iv) (Slutsky's Theorem) Suppose $X_1, X_2, \ldots$ converges in distribution to X and $Y_1, Y_2, \ldots$ converges in probability to Y. Assume Y is constant almost surely. Show that $X_1+Y_1, X_2+Y_2, \ldots$ converges in distribution to $X+Y$. Show also that $X_1Y_1, X_2Y_2, \ldots$ converges in distribution to XY. (Hint: either use (iii) or use (ii) to control error terms.) What happens if Y is not constant almost surely?

(v) (Fatou's lemma) If $g\colon \mathbb{R} \to [0, \infty)$ is continuous, and if $X_1, X_2, \ldots$ converges in distribution to X, show that $\liminf_{n\to\infty} \mathbf{E}g(X_n) \geq \mathbf{E}g(X)$.

(vi) (Bounded convergence) If $g\colon \mathbb{R} \to \mathbb{C}$ is continuous and bounded, and if $X_1, X_2, \ldots$ converges in distribution to X, show that $\lim_{n\to\infty} \mathbf{E}g(X_n) = \mathbf{E}g(X)$.

(vii) (Dominated convergence) If $X_1, X_2, \ldots : \Omega \to \mathbb{R}$ converges in distribution to X, and if there exists a random variable $Y\colon \Omega \to [0, \infty)$ with $|X_n| \leq Y$ for all $n \geq 1$ and $\mathbf{E}Y < \infty$, show that $\lim_{n\to\infty} \mathbf{E}X_n = \mathbf{E}X$.

Solution. Part (ii) is Theorem 3.2.13 in Durrett's book (5th Edition)

Part (iii) is Theorem 3.2.8 in Durrett's book.

Parts (v) (vi) and (vii) follow by applying part (iii) to the corresponding statements where almost sure convergence is assumed.

□

Exercise 7.9 (Optional; **Portmanteau Theorem**)**.** Let $X, X_1, X_2, \ldots$ be random variables with values in $\mathbb{R}$. Show that the condition ($X_1, X_2, \ldots$ converges in distribution to X) is equivalent to the following three statements:

- For any closed $K \subseteq \mathbb{R}$, $\limsup_{n\to\infty} \mathbf{P}(X_n \in K) \leq \mathbf{P}(X \in K)$.
- For any open $U \subseteq \mathbb{R}$, $\liminf_{n\to\infty} \mathbf{P}(X_n \in U) \leq \mathbf{P}(X \in U)$.
- For any Borel set $E \subseteq \mathbb{R}$ whose topological boundary ∂E satisfies $\mathbf{P}(X \in \partial E) = 0$, $\lim_{n\to\infty} \mathbf{P}(X_n \in E) = \mathbf{P}(X \in E)$.

(Hint: Urysohn's Lemma might be helpful.)

Solution. See Theorem 3.2.11 in Durrett's book (5th edition). □

8. Homework 8

Exercise 8.1. Let $k \geq 1$ be an integer. Let $X : \Omega \to \mathbb{R}$ be a random variable with finite k^{th} moment: $\mathbf{E}|X|^k < \infty$. Show that $\phi_X(t)$ is k-times continuously differentiable in t, and

$$\frac{d^k}{dt^k}|_{t=0}\phi_X(t) = i^k \mathbf{E} X^k.$$

In particular, we get the Taylor expansion

$$\phi_X(t) = \sum_{n=0}^{k} \frac{(it)^n}{n!}\mathbf{E}X^n + o(|t|^k), \qquad \forall\, t \in \mathbb{R}.$$

Solution. Let $t > 0$ and consider the difference quotient

$$\frac{\phi_X(t) - \phi_X(0)}{t} = \frac{\mathbf{E}e^{itX} - \mathbf{E}e^0}{t} = \mathbf{E}\frac{e^{itX} - 1}{t}.$$

From the Fundamental Theorem of Calculus, we have

$$e^{is} = e^0 + \int_0^s \frac{d}{dt}e^{it}dt = 1 + \int_0^s ie^{it}dt.$$

So,

$$\left|e^{is} - 1\right| \leq |s|.$$

Substituting $s = tX$, we get $\left|e^{itX} - 1\right| \leq |tX|.$, so that

$$\left|\frac{e^{itX} - 1}{t}\right| \leq |X|, \qquad \forall\, t \in \mathbb{R} \setminus \{0\}.$$

So, from the Dominated Convergence Theorem,

$$\lim_{t\to 0} \frac{\phi_X(t) - \phi_X(0)}{t} = \mathbf{E}\lim_{t\to 0}\frac{e^{itX} - 1}{t} = \mathbf{E}\frac{d}{dt}|_{t=0}e^{itX} = i\mathbf{E}X.$$

Similarly, for all $s \in \mathbb{R}$, $\phi'(s)$ exists, since

$$\lim_{t\to 0} \frac{\phi_X(s+t) - \phi_X(s)}{t} = \lim_{t\to 0} \mathbf{E}e^{isX}\frac{e^{itX} - \mathbf{E}e^0}{t} = \mathbf{E}iXe^{isX}.$$

The case of general k follows by induction.

$$\frac{\phi_X^{(k-1)}(t) - \phi_X^{(k-1)}(0)}{t} = \frac{\mathbf{E}(iX)^{k-1}e^{itX} - \mathbf{E}(iX)^{k-1}e^0}{t} = \mathbf{E}(iX)^{k-1}\frac{e^{itX}-1}{t}.$$

From the Fundamental Theorem of Calculus, we again use the inequality $|e^{is} - 1| \leq |s|$ with $s = tX$ to get $\left|(iX)^{k-1}\right| \left|e^{itX} - 1\right| \leq \left|tX^k\right|.$, so that

$$\left|(iX)^{k-1}\frac{e^{itX}-1}{t}\right| \leq |X|^k, \qquad \forall\, t \in \mathbb{R} \setminus \{0\}.$$

So, from the Dominated Convergence Theorem,

$$\lim_{t\to 0}\frac{\phi_X^{(k-1)}(t) - \phi_X^{(k-1)}(0)}{t} = \mathbf{E}(iX)^{k-1}\lim_{t\to 0}\frac{e^{itX}-1}{t} = \mathbf{E}(iX)^{k-1}\frac{d}{dt}|_{t=0}e^{itX} = (i\mathbf{E}X)^k.$$

Similarly, for all $s \in \mathbb{R}$, $\phi^{(k)}(s)$ exists. □

Exercise 8.4 (Lévy Continuity Theorem). Let $X, X_1, X_2, \ldots$ be real-valued random variables (possibly on different sample spaces). Assume that, $\forall\, t \in \mathbb{R}$, $\phi(t) := \lim_{n\to\infty}\phi_{X_n}(t)$ exists. Then the following are equivalent.

(i) ϕ is continuous at 0.
(ii) $\mu_{X_1}, \mu_{X_2}, \ldots$ is tight. ($\forall\ \varepsilon > 0$, $\exists\ m = m(\varepsilon) > 0$ such that $\limsup_{n\to\infty}(1 - \mu_{X_n}([-m, m])) \leq \varepsilon$.)
(iii) There exists a random variable X such that $\phi_X = \phi$.
(iv) $X_1, X_2, \ldots$ converges in distribution to X.

(Hint: Use the Exercise from the previous homework that characterizes tightness for vague convergence to get from (ii) to other conditions.)

Proof. We already know that (iii) and (iv) are equivalent by the Lévy Continuity Theorem, Special Case, from the notes.

We now show that (i) implies (ii). Let $a > 0, t \in \mathbb{R}$ and note that

$$\frac{1}{2a}\int_{-a}^{a}(1 - e^{itx})dt = 1 - \frac{\sin(ax)}{ax} \geq 0.$$

Taking expected values with $x = X_n$, we have

$$\begin{aligned}\frac{1}{2a}\int_{-a}^{a}(1 - \phi_{X_n}(t))dt &= \mathbf{E}\Big(1 - \frac{\sin(aX_n)}{aX_n}\Big) \geq \mathbf{E}1_{|X_n|>2/a}\Big(1 - \frac{\sin(aX_n)}{aX_n}\Big)\\ &\geq \mathbf{E}1_{|aX_n|>2}\Big(1 - \frac{1}{aX_n}\Big) \geq \frac{1}{2}\mathbf{E}1_{|X_n|>2/a} = \frac{1}{2}\mathbf{P}(|X_n| > 2/a).\end{aligned}$$

Taking the lim sup of both sides,

$$\frac{1}{2a}\int_{-a}^{a}(1 - \phi_X(t))dt \geq \frac{1}{2}\limsup_{n\to\infty}\mathbf{P}(|X_n| > 2/a).$$

Since $\phi_X(t)$ is continuous at zero, the left side goes to zero as $a \to 0^+$. We therefore conclude that (ii) holds.

We now show that (ii) implies (iv). From Exercise 7.8(ii), any subsequential limit of the sequence $\mu_{X_1}, \mu_{X_2}, \ldots$ (with respect to vague convergence) is a probability measure. Let $\mu_{X_{j_1}}, \mu_{X_{j_2}}, \ldots$ be a subsequence that converges vaguely to some probability measure μ. Let X be a random variable such that $\mu_X = \mu$. Then by Proposition 3.3 in the notes, $X_{j_1}, X_{j_2}, \ldots$

converges in distribution to X. From Theorem 3.20 in the notes, $\lim_{n\to\infty} \phi_{X_{j_n}}(t) = \phi_X(t)$ for all $t \in \mathbb{R}$. Since it is assumed that $\lim_{n\to\infty} \phi_{X_n}(t) = \phi(t)$ for all $t \in \mathbb{R}$, any other subsequence must converge vaguely to μ_X (also by Theorem 3.20 in the notes). That is, (iv) holds.

It remains to show that (iv) implies (i). But this follows since $\phi(t) = \phi_X(t)$, and ϕ_X must be continuous at 0, since it is the characteristic function of a random variable

□

Exercise 8.6. Give an alternate proof of the fact $\mathcal{F}[e^{-x^2/2}](\xi) = \sqrt{2\pi}e^{-\xi^2/2}$ using the following strategy:

- Let $g(\xi) := (2\pi)^{-1/2}\mathcal{F}[e^{-x^2/2}](\xi)$. Show that $g'(\xi) = -\xi g(\xi)$ for all $\xi \in \mathbb{R}$.
- Deduce that $(d/d\xi)(g(\xi)e^{\xi^2/2}) = 0$.
- Finally, conclude that $g(\xi) = e^{-\xi^2/2}$.

Exercise 8.7 (Weak Berry-Esséen theorem)**.** Let $X, X_1, X_2, \ldots$ be i.i.d. real-valued random variables with mean zero, variance 1 and with $\mathbf{E}\,|X|^3 < \infty$. Let Z be a standard Gaussian random variable.

(i) Show that for any compactly supported $g\colon \mathbb{R} \to \mathbb{R}$ with three continuous derivatives, and for any $n \geq 1$,

$$\mathbf{E}g\Big(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\Big) = \mathbf{E}g(Z) + O(n^{-1/2} \sup_{x\in\mathbb{R}} |g'''(x)|\, \mathbf{E}|X|^3),$$

where the implied constant does not depend on g, n or on any of the random variables $X, X_1, X_2, \ldots$.

(ii) Show that, for any $n \geq 1$, and for any $t \in \mathbb{R}$

$$\mathbf{P}\Big(\frac{X_1 + \cdots + X_n}{\sqrt{n}} \leq t\Big) = \mathbf{P}(Z \leq t) + O(n^{-1/2}\mathbf{E}|X|^3)^{1/4},$$

where the implied constant does not depend on n, t or on any of the random variables $X, X_1, X_2, \ldots$.

(Hint: for the first item, try to modify the argument we did in class.) (Hint: for the second item, consider $g = 1_{[-\infty,t]} * \phi_\varepsilon$, where $\phi(x) = e^{-x^2/2}/\sqrt{2\pi}$ and $\phi_\varepsilon(x) = \varepsilon^{-1}\phi(x/\varepsilon)$ for any $x \in \mathbb{R}$ and for appropriately chosen $\varepsilon > 0$.)

Solution. The first part follows by the argument we gave in class. We also note that if ϕ is bounded and all of its derivatives are bounded, then we can apply part (i) to ϕ, which is the case in the second part of the problem. As suggested, we use

$$g(y) = 1_{[-\infty,t]} * \phi_\varepsilon(y) = \int_{\mathbb{R}} 1_{x\leq t}\varepsilon^{-1}\phi((y-x)/\varepsilon)dx.$$

We then have

$$|g'''(y)| \leq \varepsilon^{-4}\int_{\mathbb{R}} 1_{x\leq t}\phi'''((y-x)/\varepsilon)dx \leq 100\varepsilon^{-3}, \qquad \forall\, y \in \mathbb{R}$$

Let $f := 1_{[-\infty,t]}$. Then

$$|\phi_\varepsilon * f(x) - f(x)| = \left|\int_{\mathbb{R}} \phi_\varepsilon(y)(f(x-y) - f(x))dy\right| = \left|\int_{\mathbb{R}} \phi(y)(f(x-\varepsilon y) - f(x))dy\right|$$

By definition of f, we have

$$|f(x-\varepsilon y)-f(x)| = |1_{x-\varepsilon y\le t} - 1_{x\le t}| = \begin{cases} 1_{0\le x-t\le \varepsilon y} & \text{, if } y>0 \\ 1_{\varepsilon y\le x-t\le 0} & \text{, if } y\le 0. \end{cases}$$

So

$$\begin{aligned}\sup_{x\in\mathbb{R}:\ |x-t|>\delta} |\phi_\varepsilon * f(x) - f(x)| &\le \sup_{x\in\mathbb{R}:\ |x-t|>\delta} \left|\int_{\mathbb{R}} \phi(y) 1_{0\le |x-t|\le \varepsilon y} dy\right| \\ &\le \left|\int_{\mathbb{R}} \phi(y) 1_{y\ge \delta/\varepsilon} dy\right| = \int_{|y|>\delta/\varepsilon} \phi(y)dy.\end{aligned}$$

Choosing $\varepsilon=\delta$ and then choosing ϕ to be a C^∞ nonnegative function supported in $[-1,1]$, the last quantity is zero, so that

$$\sup_{x\in\mathbb{R}:\ |x-t|>\delta} |\phi_\varepsilon * f(x) - f(x)| = 0.$$

Using this fact and then the first part of the exercise twice,

$$\begin{aligned}\mathbf{P}\Big(\frac{X_1+\cdots+X_n}{\sqrt{n}}\le t\Big) &\le \mathbf{E}(1_{[-\infty,t+2\delta]} * \phi_\varepsilon)\Big(\frac{X_1+\cdots+X_n}{\sqrt{n}}\Big) \\ &\le \mathbf{E}(1_{[-\infty,t+2\delta]} * \phi_\varepsilon)(Z) + O(n^{-1/2}\varepsilon^{-3}\mathbf{E}|X|^3) \\ &\le \mathbf{P}(Z\le t+4\delta) + O(n^{-1/2}\varepsilon^{-3}\mathbf{E}|X|^3) \\ &\le \mathbf{P}(Z\le t) + \mathbf{P}(0<Z-t<4\delta) + O(n^{-1/2}\varepsilon^{-3}\mathbf{E}|X|^3).\end{aligned}$$

Similarly,

$$\mathbf{P}\Big(\frac{X_1+\cdots+X_n}{\sqrt{n}}\le t\Big) \ge \mathbf{P}(Z\le t) - \mathbf{P}(0<Z-t<4\delta) - O(n^{-1/2}\varepsilon^{-3}\mathbf{E}|X|^3).$$

Since Z is a standard Gaussian, $\mathbf{P}(0<Z-t<4\delta) = O(\delta)$. In summary,

$$\left|\mathbf{P}\Big(\frac{X_1+\cdots+X_n}{\sqrt{n}}\le t\Big) - \mathbf{P}(Z\le t)\right| = O(n^{-1/2}\varepsilon^{-3}\mathbf{E}|X|^3) + O(\delta).$$

Choosing $\delta=\varepsilon=(n^{-1/2}\mathbf{E}|X|^3)^{1/4}$ concludes the proof. □

Exercise 8.10. Show that there exists a nonzero random variable X such that, if $X_1, X_2, \dots$ are i.i.d. copies of X, then $\frac{X_1+\cdots+X_n}{n}$ is equal in distribution to X, for any $n\ge 1$. (Optional: can you write out an explicit formula for the density of X?) (Hint: take the Fourier transform.)

Show that there exists a nonzero random variable X such that, if $X_1, X_2, \dots$ are i.i.d. copies of X, then $\frac{X_1+\cdots+X_n}{n^2}$ is equal in distribution to X, for any $n\ge 1$.

Solution. If X satisfies the above with $\phi_X(t) := \mathbf{E}e^{itX}$ for any $t\in\mathbb{R}$, then

$$\phi_X(t) := \phi_{\frac{X_1+\cdots+X_n}{n}}(t) = \prod_{j=1}^n \phi_{X_j}(t/n) = (\phi_X(t/n))^n.$$

Solving the functional equation, or just guessing (recalling also that $|\phi_X(t)|\le 1$ for all $t\in\mathbb{R}$), we anticipate that there is a $c>0$ such that

$$\phi_X(t) = e^{-c|t|}, \qquad \forall\, t\in\mathbb{R}.$$

Similarly, if X satisfies the last condition, then $\phi_X(t) := \phi_{\frac{X_1+\cdots+X_n}{n^2}}(t) = (\phi_X(t/n^2))^n$, and we anticipate that there is a $c > 0$ such that $\phi_X(t) = e^{-c|t|^{1/2}}$ $\forall\, t \in \mathbb{R}$.

As shown in Durrett in Section 2.7, if Y satisfies $\mathbf{P}(Y > t) = \mathbf{P}(Y < -t) = t^{-\alpha}/2$ for all $t > 1$, then $\lim_{t\to 0} \frac{1-\phi_Y(t)}{|t|^\alpha} = 1/c$ for some $c > 0$, if also $0 < \alpha < 2$. Then for any $t > 0$, if $Y_1, Y_2, \ldots$ are independent copies of Y,

$$\mathbf{E}e^{it(Y_1+\cdots+Y_n)/n^{1/\alpha}} = [\phi_Y(t/n^{1/\alpha})]^n = [1 - (1 - \phi_Y(t/n^{1/\alpha}))]^n.$$

Taking the logarithm and letting $n \to \infty$, we have $\lim_{n\to\infty} n(1 - \phi(t/n^{1/\alpha})) = c\,|t|^\alpha$, so that, for any $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbf{E}e^{it(Y_1+\cdots+Y_n)/n^{1/\alpha}} = e^{-c|t|^\alpha}.$$

It then follows from Exercise 8.4 that there exists a random variable X such that $\phi_X(t) = e^{-c|t|^\alpha}$, as long as this function is continuous at 0. And this function is continuous at zero when $\alpha = 1$, $\alpha = 1/2$, or more generally when $0 < \alpha < \infty$, though to find the behavior of ϕ_Y near 0 we required $0 < \alpha < 2$.

In the case $\alpha = 1$, we can let X have density $x \mapsto \frac{1}{\pi(1+x^2)}$. We can see this via contour integration, since the function $z \mapsto e^{itz}/(1 + z^2) = \frac{e^{itz}}{(z+i)(z-i)}$, $z \in \mathbb{C}$ has a simple pole at $z = i$. By integrating over a half circle in the upper half plane, and then letting the radius go to infinity, we get for any $t > 0$,

$$\mathbf{E}e^{itX} = \int_{\mathbb{R}} \frac{e^{itx}}{\pi(1 + x^2)} dx = 2\pi i \lim_{z\to i} \frac{e^{itz}}{\pi(z + i)} = e^{-t}.$$

Similarly, when $t < 0$, we integrate over a half circle in the lower half plane (with a counter-clockwise contour integral), and then letting the radius go to infinity to get

$$\mathbf{E}e^{itX} = \int_{\mathbb{R}} \frac{e^{itx}}{\pi(1 + x^2)} dx = -2\pi i \lim_{z\to -i} \frac{e^{itz}}{\pi(z - i)} = e^{t}.$$

So, for any $t \in \mathbb{R}$, we have $\mathbf{E}e^{itX} = e^{-|t|}$.

When $\alpha = 1/2$, we can let X have density $x \mapsto (2\pi x^3)^{-1/2} e^{-1/2x}$ when $x > 0$, though this is less obvious.

The case $\alpha = 2$ corresponds to the Gaussian density; we verified in Proposition 8.7 or Exercise 8.6 that a standard Gaussian X satisfies $\mathbf{E}e^{itX} = e^{-t^2}$ for any $t \in \mathbb{R}$.

However, if $\alpha > 2$, it is not possible that X can satisfy $\phi_X(t) = e^{-c|t|^\alpha}$, since by the moment formula for derivatives of ϕ_X, we would have $\mathbf{E}X^2 = -\frac{d^2}{dt^2}|_{t=0}\phi_X(t) = 0$, implying that $X = 0$ so that $c = 0$ as well. □

9. Homework 9

Exercise 9.1. Let $X_1, X_2, \ldots : \Omega \to \mathbb{R}$ be i.i.d. In each of the cases below, show that with probability one, $-\infty = \liminf_{n\to\infty} S_n$ and $\limsup_{n\to\infty} S_n = \infty$.

- The distribution μ_{X_1} is symmetric about 0 (i.e. $\mu_{-X_1} = \mu_{X_1}$) and $\mathbf{P}(X_1 = 0) < 1$.
- $\mathbf{E}X_1 = 0$ and $\mathbf{E}X_1^2 \in (0, \infty)$. (Hint: use the Central Limit Theorem.)

For example, when $\mathbf{P}(X_1 = 1) = \mathbf{P}(X_1 = -1) = 1/2$ and $S_0 = 0$, show that with probability one, $S_0, S_1, \ldots$ takes every integer value infinitely many times.

Solution. The distribution of $X_1 + \cdots + X_n$ is symmetric about 0, implying that cases (ii) and (iii) of Theorem 4.6 cannot occur. Also case (i) does not occur by definition of X_1. So, by Theorem 4.6, $-\infty = \liminf_{n\to\infty} S_n$ and $\limsup_{n\to\infty} S_n = \infty$ with probability one.

In the second case, the central limit theorem says

$$\lim_{n\to\infty} \mathbf{P}(\frac{X_1 + \cdots + X_n}{\sqrt{n}\sqrt{\mathbf{E}X_1^2}} \geq 1) = \int_1^\infty e^{-t^2/2} dt/\sqrt{2\pi}.$$

So, there exists $m > 0$ such that, for all $n \geq m$, we have

$$\mathbf{P}(X_1 + \cdots + X_n \geq \sqrt{\mathbf{E}X_1^2}\sqrt{n}) \geq \frac{1}{10}.$$

Similarly, there exists $m' > 0$ such that, for all $n \geq m'$, we have

$$\mathbf{P}(X_1 + \cdots + X_n \leq \sqrt{\mathbf{E}X_1^2} - \sqrt{n}) \geq \frac{1}{10}.$$

This condition implies that parts (i), (ii) and (iii) of Theorem 4.6 do not occur. So, by Theorem 4.6, $-\infty = \liminf_{n\to\infty} S_n$ and $\limsup_{n\to\infty} S_n = \infty$ with probability one.

□

Exercise 9.2. Let M, N be stopping times for a random walk $S_0, S_1, \ldots$. Show that $\max(M, N)$ and $\min(M, N)$ are stopping times. In particular, if $n \geq 1$ is fixed, then $\max(M, n)$ and $\min(M, n)$ are stopping times

Solution. Let $p \geq 1$ be an integer. Then

$$\begin{aligned}
&\{\max(M,N) = p\} \\
&= \cup_{j=0}^{p} \{\max(M,N) = p, \min(M,N) = j\} \\
&= \cup_{j=0}^{p} \{\max(M,N) = p, \min(M,N) = j, M \leq N\} \cup \{\max(M,N) = p, \min(M,N) = j, M > N\} \\
&= \cup_{j=0}^{p} \{N = p, M = j, M \leq N\} \cup \{M = p, N = j, M > N\} \\
&= \Big(\cup_{j=0}^{p} \{N = p, M = j\} \Big) \cup \Big(\cup_{j=0}^{p-1} \{M = p, N = j\} \Big).
\end{aligned}$$

All of the events on the right are in $\mathcal{F}_p$ since N, M are stopping times and $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots$. Since $\mathcal{F}_p$ is a σ-algebra, we conclude that $\{\max(M, N) = p\} \in \mathcal{F}_p$, as desired. Similarly,

$$\begin{aligned}
&\{\min(M,N) = p\} \\
&= \cup_{j=p}^{\infty} \{\min(M,N) = p, \max(M,N) = j\} \\
&= \cup_{j=p}^{\infty} \{\min(M,N) = p, \max(M,N) = j, M \leq N\} \cup \{\min(M,N) = p, \max(M,N) = j, M > N\} \\
&= \cup_{j=p}^{\infty} \{M = p, N = j, M \leq N\} \cup \{N = p, M = j, M > N\} \\
&= \Big(\cup_{j=p}^{\infty} \{M = p, N = j\} \Big) \cup \Big(\cup_{j=p+1}^{\infty} \{N = p, M = j\} \Big) \\
&= \{M = p, N \geq p\} \cup \{N = p, M > p\} \\
&= \big(\{M = p\} \cap \{N < p\}^c \big) \cup \big(\{N = p\} \cap \{M \leq p\}^c \big) \\
&= \big(\{M = p\} \cap (\cup_{j=0}^{p-1} \{N = j\})^c \big) \cup \big(\{N = p\} \cap (\cup_{j=0}^{p} \{M = j\})^c \big).
\end{aligned}$$

All of the events on the right are in $\mathcal{F}_p$, and since $\mathcal{F}_p$ is a σ-algebra, we conclude that $\{\max(M, N) = p\} \in \mathcal{F}_p$, as desired. □

Exercise 9.6. Let $0 < p < 1$. Consider the random walk on $\mathbb{Z}$ such that $\mathbf{P}(X_1 = 1) = p$ and $\mathbf{P}(X_1 = -1) = 1 - p$. Show that the corresponding random walk $S_0, S_1, \ldots$ is transient when $p \neq 1/2$.

Solution. Since $(S_n + n)/2$ has binomial distribution with parameters n and p, $(S_{2n}/2) + n$ has binomial distribution with parameters $2n$ and p, and $\{S_{2n} = 0\} = \{S_{2n}/2 + n = n\}$, so

$$\mathbf{P}(S_{2n} = 0) = \binom{2n}{n} p^n (1-p)^n.$$

Using Stirling's formula in the form $n! \sim \sqrt{2\pi n}(n/e)^n$, we have

$$\mathbf{P}(S_{2n} = 0) = \frac{(2n)!}{(n!)^2} p^n (1-p)^n \sim \frac{1}{\sqrt{2\pi}} \sqrt{\frac{2}{n}} \frac{(2n)^{2n}}{n^{2n}} p^n (1-p)^n = \frac{1}{\sqrt{2\pi n}} (4p(1-p))^n.$$

Since $p \neq 1/2$, $4p(1-p) < 1$, so this quantity decays exponentially in n, i.e.

$$\sum_{n=0}^{\infty} \mathbf{P}(S_{2n} = 0) < \infty.$$

So, the random walk is transient by Theorem 4.14.

□

Exercise 9.8. Let $1/2 < p < 1$. Consider the random walk on $\mathbb{Z}$ such that $\mathbf{P}(X_1 = 1) = p$ and $\mathbf{P}(X_1 = -1) = 1-p$. Let $S_0, S_1, \ldots$ be the corresponding random walk with $S_0 := 0$. Let $N := \min\{n \geq 1 \colon S_n > 0\}$. Using Wald's equation for $\min(N, n)$ and then letting $n \to \infty$, show that $\mathbf{E}N = 1/\mathbf{E}X_1 = 1/(2p-1)$.

Solution. Since N is a stopping time, if $n > 1$ is fixed, $\min(N, n)$ is a stopping time with $\mathbf{E}\min(N, n) \leq n < \infty$. Also, $\mathbf{E}|X_1| \leq 1 < \infty$. Note that $\mathbf{E}X_1 = p - (1-p) = 2p - 1$. So, from Wald's equation,

$$\mathbf{E}S_{\min(N,n)} = \mathbf{E}X_1 \mathbf{E}\min(N, n) = (2p-1)\mathbf{E}\min(N, n). \qquad (*)$$

Recall that $\mathbf{E}X_1 = p - (1-p) = 2p - 1 > 0$ since $p > 1/2$. Using Markov's inequality for the fourth moment as in Proposition 2.14, $\exists\, c = c(p) > 0$ such that, for all $n \geq 1$,

$$\mathbf{P}(S_n \leq 0) \leq \mathbf{P}\left(\left|\frac{S_n - n\mathbf{E}X_1}{n}\right| > \frac{\mathbf{E}X_1}{2}\right) \leq \frac{c}{n^2}. \qquad (**)$$

In particular, $\mathbf{P}(N = \infty) = 0$, so the right side of $(*)$ converges monotonically to $(2p-1)\mathbf{E}N$ as $n \to \infty$, by the Monotone Convergence Theorem. Meanwhile, by the definition of N, the left side of $(*)$ can be estimated as

$$\begin{aligned} 1 &\geq \mathbf{E}S_{\min(N,n)} = \mathbf{E}S_{\min(N,n)}(1_{N \leq n} + 1_{N > n}) = \mathbf{E}S_N 1_{N \leq n} + \mathbf{E}S_n 1_{N > n} \\ &= \mathbf{E}1_{N \leq n} + \mathbf{E}S_n 1_{N > n} = \mathbf{P}(N \leq n) + \mathbf{E}S_n 1_{N > n} \geq \mathbf{P}(N \leq n) - n\mathbf{P}(N > n) \\ &= 1 - (n+1)\mathbf{P}(N > n). \end{aligned}$$

(We always have $S_n \geq -n$.) By definition of N, we have $\mathbf{P}(N > n) \leq \mathbf{P}(S_n \leq 0)$. So, from $(**)$ and the Squeeze Theorem, we conclude that $\lim_{n\to\infty} \mathbf{E}S_{\min(N,n)} = 1$, so that $(*)$ says $1 = (2p-1)\mathbf{E}N$, as desired. □

10. Homework 10

Exercise 10.4. Let $\Omega = [0,1]$. Let $\mathbf{P}$ be the uniform probability law on Ω. Let $X\colon [0,1] \to \mathbb{R}$ be a random variable such that $X(t) = t^2$ for all $t \in [0,1]$. For every integer $k > 1$, let $s = 2^{-k}$, let $\mathcal{G}_k := \sigma\{[0,s), [s,2s), [2s,3s), \dots, [1-2s, 1-s), [1-s, 1)\}$, and let $M_k := \mathbf{E}(X|\mathcal{G}_k)$. Show that the increments $M_2 - M_1, M_3 - M_2, \dots$ are orthogonal in the following sense. For any $i, j \geq 1$ with $i \neq j$,

$$\mathbf{E}(M_{i+1} - M_i)(M_{j+1} - M_j) = 0.$$

This property is sometimes called **orthogonality of martingale increments**.

Solution. Let $i, j \geq 1$ with $i \neq j$. Without loss of generality, $i < j$. Consider an interval $I \in \mathcal{G}_{i+1}$. By definition, M_i and M_{i+1} are constant on I. So, $M_{i+1} - M_i$ is constant on I. Since $i < j$, we can write I in the form $I = [ps, (p+1)s) \cup [(p+1)s, (p+2)s) \cup \dots \cup [(p+n-1)s, (p+n)s)$ where $n, p \in \mathbb{Z}$ and $s = 2^{-j-1}$. That is, I is a union of an even number of smaller intervals $J_1, \dots, J_n$. Since $M_{i+1} - M_i$ is constant on I, $M_{i+1} - M_i$ is constant on each subinterval $J_1, \dots, J_n$. By definition, M_{j+1} is constant on J_1 and M_j is constant on $J_1 \cup J_2$. Moreover, by their definition,

$$\begin{aligned}\mathbf{E}(M_{j+1} - M_j)1_{J_1 \cup J_2} &= \mathbf{E}M_{j+1}1_{J_1 \cup J_2} - \mathbf{E}M_j 1_{J_1 \cup J_2} \\ &= \mathbf{E}M_{j+1}1_{J_1} + \mathbf{E}M_{j+1}1_{J_2} - \mathbf{E}M_j 1_{J_1 \cup J_2} = \int_{J_1} X(t)dt + \int_{J_2} X(t)dt - \int_{J_1 \cup J_2} X(t)dt = 0.\end{aligned}$$

Similarly, for any $1 \leq \ell \leq n/2$,

$$\mathbf{E}(M_{j+1} - M_j)1_{J_{2\ell-1} \cup J_{2\ell}} = 0.$$

Then

$$\mathbf{E}(M_{j+1} - M_j)1_I = \mathbf{E}(M_{j+1} - M_j)(\sum_{p=1}^{n} 1_{J_p}) = (\sum_{\ell=1}^{n/2} \mathbf{E}(M_{j+1} - M_j)1_{J_{2\ell-1} \cup J_{2\ell}}) = 0.$$

And since $M_{i+1} - M_i$ is constant on I,

$$\mathbf{E}(M_{j+1} - M_j)(M_{i+1} - M_i)1_I = (M_{i+1} - M_i)\mathbf{E}(M_{j+1} - M_j)1_I = 0.$$

Finally, we write $[0,1]$ as a union of intervals I that generate $\mathcal{G}_{i+1}$ to get

$$\begin{aligned}\mathbf{E}(M_{i+1} - M_i)(M_{j+1} - M_j) &= \mathbf{E}(M_{i+1} - M_i)(M_{j+1} - M_j)(\sum_{I \text{ generating } \mathcal{G}_{i+1}} 1_I) \\ &= \sum_{I \text{ generating } \mathcal{G}_{i+1}} \mathbf{E}(M_{i+1} - M_i)(M_{j+1} - M_j)1_I = 0.\end{aligned}$$

□

Exercise 10.5. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, and let $X\colon \Omega \to \mathbb{R}$ be a random variable with $\mathbf{E}|X| < \infty$. Let $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ be σ-algebras. Let $\mathcal{H}$ be a σ-algebra that is independent of $\sigma(\sigma(X), \mathcal{G})$. Show that

$$\mathbf{E}(X|\sigma(\mathcal{G}, \mathcal{H})) = \mathbf{E}(X|\mathcal{G}).$$

In particular, if we choose $\mathcal{G} = \{\emptyset, \Omega\}$, we get: if $\mathcal{H}$ is independent of $\sigma(X)$, then $\mathbf{E}(X|\mathcal{H}) = \mathbf{E}X$.

(Hint: Let $G \in \mathcal{G}, H \in \mathcal{H}$, let $Y := \mathbf{E}(X|\mathcal{G})$. Compare $\mathbf{E}(X1_{G\cap H})$ and $\mathbf{E}(Y1_{G\cap H})$. Is the set of $A \in \sigma(\mathcal{G},\mathcal{H})$ such that $\mathbf{E}(X1_A) = \mathbf{E}(Y1_A)$ a monotone class?)

Solution. Let $Y := \mathbf{E}(X|\mathcal{G})$. Let $G \in \mathcal{G}, H \in \mathcal{H}$. By assumption, $\mathcal{H}$ is independent of $\sigma(\sigma(X),\mathcal{G})$. So, 1_H is independent of $X1_G$ and $Y1_G$. Therefore,

$$\mathbf{E}X1_{G\cap H} = \mathbf{E}X1_G1_H = \mathbf{E}X1_G\mathbf{E}1_H.$$

$$\mathbf{E}Y1_{G\cap H} = \mathbf{E}Y1_G1_H = \mathbf{E}Y1_G\mathbf{E}1_H.$$

By definition of Y, $\mathbf{E}Y1_G = \mathbf{E}X1_G$. Combining these facts, $\mathbf{E}X1_{G\cap H} = \mathbf{E}Y1_{G\cap H}$. Let $\mathcal{A} := \{G \cap H \colon G \in \mathcal{G}, H \in \mathcal{H}\}$. We have shown that $\mathbf{E}X1_A = \mathbf{E}Y1_A$ for all $A \in \mathcal{A}$. Note that $\mathcal{A}$ is closed under finite intersection, i.e. $\mathcal{A}$ is an algebra. Let $\mathcal{A}' := \{A \in \sigma(\mathcal{G},\mathcal{H}) \colon \mathbf{E}X1_A = \mathbf{E}Y1_A\}$. Note that $\mathcal{A}'$ is closed under complements.

Assume for now that $X \geq 0$. Then $Y \geq 0$. So a monotone convergence argument implies that $\mathcal{A}'$ is closed under increasing unions. Therefore, $\mathcal{A}'$ is a monotone class (since it is also closed under complements). Also, $\mathcal{A}' \supseteq \mathcal{A}$. So, the Monotone Class Lemma implies that $\mathcal{A}' = \sigma(\mathcal{G},\mathcal{H})$. That is, for every $A \in \sigma(\mathcal{G},\mathcal{H})$, we have $\mathbf{E}X1_A = \mathbf{E}Y1_A$. That is, $Y = \mathbf{E}(X|\sigma(\mathcal{G},\mathcal{H}))$. By definition of Y, $Y = \mathbf{E}(X|\mathcal{G})$ as well, so we are done when $X \geq 0$.

To prove the general case, we write $X = \min(X,0) - \min(-X,0)$ and apply the above to each part of X separately. □

Exercise 10.8 (Conditional Markov Inequality)**.** Let $p > 0$. Let $(\Omega,\mathcal{F},\mathbf{P})$ be a probability space, and let $X \colon \Omega \to \mathbb{R}$ be a random variable with $\mathbf{E}\,|X|^p < \infty$. Let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra. For any $A \in \mathcal{F}$, we denote $\mathbf{P}(A|\mathcal{G}) := \mathbf{E}(1_A|\mathcal{G})$.

- Show that, almost surely,

$$\mathbf{E}(|X|^p\,|\mathcal{G}) = \int_0^\infty pt^{p-1}\mathbf{P}(|X| > t|\mathcal{G})dt.$$

- Deduce a conditional version of Markov's inequality: for any $t > 0$, almost surely,

$$\mathbf{P}(|X| > t|\mathcal{G}) \leq \frac{\mathbf{E}(|X|^p\,|\mathcal{G})}{t^p}.$$

Solution. Let $A \in \mathcal{G}$. By the definition of conditional expectation,

$$\begin{aligned}\mathbf{E}\int_0^\infty pt^{p-1}\mathbf{P}(|X| > t|\mathcal{G})dt1_A &= \int_0^\infty pt^{p-1}\mathbf{E}[\mathbf{E}(1_{|X|>t}|\mathcal{G})1_A]dt = \int_0^\infty pt^{p-1}\mathbf{E}[1_{|X|>t}1_A]dt \\ &= \int_0^\infty pt^{p-1}\mathbf{P}(|X|\,1_A > t)dt = \mathbf{E}\,|X1_A|^p = \mathbf{E}\,|X|^p\,1_A.\end{aligned}$$

The penultimate equality used Example 1.87. So, by the definition of conditional expectation, we conclude that

$$\mathbf{E}(|X|^p\,|\mathcal{G}) = \int_0^\infty pt^{p-1}\mathbf{P}(|X| > t|\mathcal{G})dt$$

Now, fix $s > 0$. If $0 \leq t \leq s$, then $1_{|X|>t} \geq 1_{|X|>s}$ so $\mathbf{P}(|X| > t|\mathcal{G}) \geq \mathbf{P}(|X| > s|\mathcal{G})$ by monotonicity of conditional expectation (Corollary 5.14). So, beginning with our above

equality, we get

$$\begin{aligned}\mathbf{E}(|X|^p\,|\mathcal{G}) &= \int_0^\infty pt^{p-1}\mathbf{P}(|X|>t|\mathcal{G})dt \geq \int_0^s pt^{p-1}\mathbf{P}(|X|>t|\mathcal{G})dt \\ &\geq \int_0^s pt^{p-1}\mathbf{P}(|X|>s|\mathcal{G})dt = \mathbf{P}(|X|>s|\mathcal{G})\int_0^s pt^{p-1}dt = \mathbf{P}(|X|>s|\mathcal{G})s^p.\end{aligned}$$

□

Exercise 10.9 (Conditional Hölder Inequality)**.** Let $p,q>1$ with $\frac{1}{p}+\frac{1}{q}=1$. Let $(\Omega,\mathcal{F},\mathbf{P})$ be a probability space, and let $X,Y\colon\Omega\to\mathbb{R}$ be random variables with $\mathbf{E}\,|X|^p\,,\mathbf{E}\,|Y|^q<\infty$. Let $\mathcal{G}\subseteq\mathcal{F}$ be a σ-algebra. Show that, almost surely,

$$\mathbf{E}(|XY|\,|\mathcal{G}) \leq [\mathbf{E}(|X|^p\,|\mathcal{G})]^{1/p}[\mathbf{E}(|Y|^q\,|\mathcal{G})]^{1/q}.$$

Solution. Let $W := [\mathbf{E}(|X|^p\,|\mathcal{G})]^{1/p}$ and let $Z := [\mathbf{E}(|Y|^q\,|\mathcal{G})]^{1/q}$. By the usual Hölder inequality, $\mathbf{E}\,|XY|<\infty$, so that $\mathbf{E}(|XY|\,|\mathcal{G})$ is well-defined. Also, by the definition of conditional expectation,

$$\mathbf{E}\,|X|^p\,1_{W=0} = \mathbf{E}\mathbf{E}(|X|^p\,|\mathcal{G})1_{W=0} = \mathbf{E}W^p1_{W=0} = 0.$$

So, $|X|^p\,1_{W=0} = 0 = |X|\,1_{W=0}$ almost surely. Then, by Proposition 5.16, since $1_{W=0}$ is $\mathcal{G}$-measurable,

$$\mathbf{E}(|XY|\,|\mathcal{G})1_{W=0} = \mathbf{E}(|XY|\,1_{W=0}|\mathcal{G}) = 0.$$

Similarly, $\mathbf{E}(|XY|\,|\mathcal{G})1_{Z=0}=0$ almost surely. It then suffices to show that, almost surely,

$$\frac{\mathbf{E}(|XY|\,|\mathcal{G})}{WZ}1_{Z\neq0,W\neq0}\leq 1.$$

Let $A\in\mathcal{G}$ with $A\cap\{W=0\}=\emptyset$ and $A\cap\{Z=0\}=\emptyset$. Then, using Proposition 5.16, since W,Z are $\mathcal{G}$-measurable, and then using the definition of conditional expectation, and the usual Hölder inequality,

$$\begin{aligned}\mathbf{E}\frac{\mathbf{E}(|XY|\,|\mathcal{G})}{WZ}1_A &= \mathbf{E}\mathbf{E}\Big(\frac{|XY|}{WZ}\Big|\mathcal{G}\Big)1_A = \mathbf{E}\frac{|XY|}{WZ}1_A \leq \Big(\mathbf{E}\frac{|X|^p}{W^p}1_A\Big)^{1/p}\Big(\mathbf{E}\frac{|Y|^q}{Z^q}1_A\Big)^{1/q} \\ &= \Big(\mathbf{E}\mathbf{E}\Big(\frac{|X|^p}{W^p}\Big|\mathcal{G}\Big)1_A\Big)^{1/p}\Big(\mathbf{E}\mathbf{E}\Big(\frac{|Y|^q}{Z^q}\Big|\mathcal{G}\Big)1_A\Big)^{1/q} = \Big(\mathbf{E}\frac{\mathbf{E}(|X|^p\,|\mathcal{G})}{W^p}1_A\Big)^{1/p}\Big(\mathbf{E}\frac{\mathbf{E}(|Y|^q\,|\mathcal{G})}{Z^q}1_A\Big)^{1/q} \\ &= (\mathbf{E}1_A)^{1/p}(\mathbf{E}1_A)^{1/q} = (\mathbf{P}(A))^{1/p+1/q} = \mathbf{P}(A). \qquad (*)\end{aligned}$$

In the middle equality, we used again the definition of conditional expectation and Proposition 5.16. Finally, let

$$A := \Big\{\frac{\mathbf{E}(|XY|\,|\mathcal{G})}{WZ}>1\Big\}, \qquad U := \Big(\frac{\mathbf{E}(|XY|\,|\mathcal{G})}{WZ}-1\Big)1_A.$$

Then $U\geq 0$ almost surely, and $\mathbf{E}U\leq 0$ by $(*)$, so that $U=0$ almost surely, i.e. $\mathbf{P}(A)=0$, thereby completing the proof. □

11. Appendix: Notation

Let n, m be a positive integers. Let A, B be sets contained in a universal set Ω.

$\mathbb{N} = \{1, 2, \ldots\}$ denotes the set of natural numbers

$\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ denotes the set of integers

$\mathbb{Q} = \{a/b \colon a, b, \in \mathbb{Z}, b \neq 0\}$ denotes the set of rational numbers

$\mathbb{R}$ denotes the set of real numbers

$\mathbb{C} = \{a + b\sqrt{-1} \colon a, b \in \mathbb{R}\}$ denotes the set of complex numbers

$\in$ means "is an element of." For example, $2 \in \mathbb{R}$ is read as "2 is an element of $\mathbb{R}$."

$\forall$ means "for all"

$\exists$ means "there exists"

$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) \colon x_i \in \mathbb{R} \, \forall \, 1 \leq i \leq n\}$

$f \colon A \to B$ means f is a function with domain A and range B. For example,

$f \colon \mathbb{R}^2 \to \mathbb{R}$ means that f is a function with domain $\mathbb{R}^2$ and range $\mathbb{R}$

$\emptyset$ denotes the empty set

$A \subseteq B$ means $\forall \, a \in A$, we have $a \in B$, so A is contained in B

$A \setminus B := \{a \in A \colon a \notin B\}$

$A^c := \Omega \setminus A$, the complement of A in Ω

$A \cap B$ denotes the intersection of A and B

$A \cup B$ denotes the union of A and B

$A \Delta B := (A \setminus B) \cup (B \setminus A)$

$\mathbf{P}$ denotes a probability law on Ω

Let $a_1, \ldots, a_n$ be real numbers. Let n be a positive integer.

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + \cdots + a_{n-1} + a_n.$$

$$\prod_{i=1}^{n} a_i = a_1 \cdot a_2 \cdots a_{n-1} \cdot a_n.$$

$\min(a_1, a_2)$ denotes the minimum of a_1 and a_2.

$\max(a_1, a_2)$ denotes the maximum of a_1 and a_2.

The min of a set of nonnegative real numbers is the smallest element of that set. We also define $\min(\emptyset) := \infty$.

Let $z \in \mathbb{C}$, so that $z = a + b\sqrt{-1}$ for some $a, b \in \mathbb{R}$. Then $\Re(z) := a$ denotes the real part of z, and

$$\mathrm{Re}(z) := a \text{ denotes the real part of } z.$$
$$\mathrm{Im}(z) := b \text{ denotes the imaginary part of } z.$$

Let $X\colon \Omega \to \mathbb{R}$ be a random variable on a probability space $(\Omega, \mathcal{F}, \mu)$.

$$\mathbf{E}(X) \text{ denotes the expected value of } X$$
$$\|X\|_p := (\mathbf{E}\,|X|^p)^{1/p}, \text{ denotes the } L_p\text{-norm of } X \text{ when } 1 \le p < \infty$$
$$\|X\|_\infty := \inf\{c > 0\colon \mathbf{P}(|X| \le c) = 1\}, \text{ denotes the } L_\infty\text{-norm of } X$$
$$\mathrm{var}(X) = \mathbf{E}(X - \mathbf{E}(X))^2, \text{ the variance of } X$$
$$\sigma_X = \sqrt{\mathrm{var}(X)}, \text{ the standard deviation of } X$$

Let $A \subseteq \Omega$. Let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra. Let $Y\colon \Omega \to \mathbb{R}$. Assume $\mathbf{E}\,|X| < \infty$. Let $\sigma(Y)$ denote the σ-algebra generated by Y.

$\mathbf{E}(X|A) := \mathbf{E}(X1_A)/\mathbf{P}(A)$ denotes the expected value of X conditioned on the event A.

$\mathbf{E}(X|\mathcal{G})$ denotes the conditional expectation of X given $\mathcal{G}$.

$\mathbf{E}(X|Y) := \mathbf{E}(X|\sigma(Y))$ denotes the conditional expectation of X given Y.

$1_A\colon \Omega \to \{0, 1\}$, denotes the indicator function of A, so that

$$1_A(\omega) = \begin{cases} 1 & \text{, if } \omega \in A \\ 0 & \text{, otherwise.} \end{cases}$$

Let H be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$. Let $h \in H$.

$$\|h\| := \langle h, h \rangle^{1/2}, \text{ denotes the norm of } h$$

Let X be a random variable on a sample space Ω, so that $X\colon \Omega \to \mathbb{R}$. Let $\mathbf{P}$ be a probability law on Ω. Let $x, t \in \mathbb{R}$. Let $i := \sqrt{-1}$.

$F_X(x) = \mathbf{P}(X \le x) = \mathbf{P}(\{\omega \in \Omega\colon X(\omega) \le x\})$
the Cumulative Distibution Function of X.

$M_X(t) = \mathbf{E}e^{tX}$ denotes the Moment Generating Function of X at $t \in \mathbb{R}$

$\phi_X(t) = \mathbf{E}e^{itX}$ denotes the Characteristic Function (or Fourier Transform) of X at $t \in \mathbb{R}$

Let $g, h\colon \mathbb{R} \to \mathbb{R}$. Let $t \in \mathbb{R}$.

$$(g * h)(t) = \int_{-\infty}^{\infty} g(x)h(t - x)dx \text{ denotes the convolution of } g \text{ and } h \text{ at } t \in \mathbb{R}$$

Let $f, g\colon \mathbb{R} \to \mathbb{C}$. We use the notation $f(t) = o(g(t))$, $\forall\ t \in \mathbb{R}$ to denote $\lim_{t\to 0} \left|\frac{f(t)}{g(t)}\right| = 0$. Let $A \subseteq \mathbb{R}$ and let $f, g\colon A \to \mathbb{C}$. We use the notation $f(t) = O(g(t))$ to denote that $\exists\ c > 0$ such that $|f(t)| \leq c\,|g(t)|$ for all $t \in A$.

USC Department of Mathematics, Los Angeles, CA
E-mail address: stevenmheilman@gmail.com