

MATH 407A, PROBABILITY, HOMEWORK SOLUTIONS

STEVEN HEILMAN

CONTENTS

0.	Homework 0	1
1.	Homework 1	2
2.	Homework 2	7
3.	Homework 3	11
4.	Homework 4	16
5.	Homework 5	18
6.	Homework 6	22
7.	Homework 7	28
8.	Homework 8	30
9.	Homework 9	34
10.	Homework 10	39
11.	Homework 11	41
12.	Homework 12	45

0. HOMEWORK 0

Exercise 0.3.

$$\sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}$$

for all natural numbers n .

Solution. If $n = 1$, then the left-hand side is 1, while the right-hand side is $\frac{1 \cdot 2 \cdot 3}{6} = 1$, which establishes the base case.

For the induction step, assume the desired formula holds for a natural number n . We must then prove that the formula holds in the case $n + 1$. We have

$$\sum_{k=1}^{n+1} k^2 = \sum_{k=1}^n k^2 + (n+1)^2 = \frac{n(n+1)(2n+1)}{6} + (n+1)^2.$$

The last line used the inductive hypothesis. Continuing, we have

$$\sum_{k=1}^{n+1} k^2 = \frac{2n^3 + 3n^2 + n + 6(n+1)^2}{6} = \frac{2n^3 + 9n^2 + 13n + 6}{6} = \frac{(n+1)(n+2)(2n+3)}{6}$$

which is the desired formula for $n + 1$. □

Date: April 30, 2023 © 2020 Steven Heilman, All Rights Reserved.

Exercise 0.4.

$$\mathbb{R} = \bigcup_{j=1}^{\infty} [-j, j].$$

$$\{0\} = \bigcap_{j=1}^{\infty} \left[-\frac{1}{j}, \frac{1}{j}\right].$$

Solution. We first show that $\bigcup_{j=1}^{\infty} [-j, j] \subseteq \mathbb{R}$. Let $x \in \bigcup_{j=1}^{\infty} [-j, j]$. By definition of countable union, there exists $j \geq 1$ such that $x \in [-j, j]$. Since $[-j, j] \subseteq \mathbb{R}$, we have $x \in \mathbb{R}$. In conclusion, $\bigcup_{j=1}^{\infty} [-j, j] \subseteq \mathbb{R}$.

We now show the reverse inclusion $\bigcup_{j=1}^{\infty} [-j, j] \supseteq \mathbb{R}$. Let $x \in \mathbb{R}$. Since $|x|$ is a nonnegative real number, there exists a positive integer $j \geq 1$ such that $|x| \leq j$, i.e. $x \in [-j, j]$. By definition of countable union, we therefore have $x \in \bigcup_{j=1}^{\infty} [-j, j]$. In conclusion, $\bigcup_{j=1}^{\infty} [-j, j] \supseteq \mathbb{R}$.

Thus we have shown that $\bigcup_{j=1}^{\infty} [-j, j] \subseteq \mathbb{R}$ and $\mathbb{R} \subseteq \bigcup_{j=1}^{\infty} [-j, j]$, which implies that $\mathbb{R} = \bigcup_{j=1}^{\infty} [-j, j]$.

Part b. We first show that $\bigcap_{j=1}^{\infty} [-1/j, 1/j] \subseteq \{0\}$. Let $x \in \bigcap_{j=1}^{\infty} [-1/j, 1/j]$. By definition of countable intersection, $x \in [-1/j, 1/j]$ for all $j \geq 1$. That is, $|x| < 1/j$ for all $j \geq 1$. By e.g. the Archimedean property of the real numbers, the only real number x satisfying $|x| < 1/j$ for all $j \geq 1$ is the real number $x = 0$. That is, $\bigcap_{j=1}^{\infty} [-1/j, 1/j] \subseteq \{0\}$.

We now show the reverse inclusion $\bigcap_{j=1}^{\infty} [-1/j, 1/j] \supseteq \{0\}$. Let $x := 0$. Then x satisfies $|x| < 1/j$ for all $j \geq 1$. That is, $x \in [-1/j, 1/j]$ for all $j \geq 1$. By definition of countable intersection, $x \in \bigcap_{j=1}^{\infty} [-1/j, 1/j]$. In conclusion, $\bigcap_{j=1}^{\infty} [-1/j, 1/j] \supseteq \{0\}$.

Thus we have shown that $\bigcap_{j=1}^{\infty} [-1/j, 1/j] \subseteq \{0\}$ and $\bigcap_{j=1}^{\infty} [-1/j, 1/j] \supseteq \{0\}$, so that $\bigcap_{j=1}^{\infty} [-1/j, 1/j] = \{0\}$, as desired. □

Exercise 0.5. Let $\Omega = \{1, 2, \dots, 10\}$. Find subsets A_1, A_2, A_3 of Ω such that $A_1 \cap A_2 = \{2, 3\}$, $A_1 \cap A_3 = \{3, 4\}$, $A_2 \cap A_3 = \{3, 5\}$, $A_1 \cap A_2 \cap A_3 = \{3\}$, and $A_1 \cup A_2 \cup A_3 = \{2, 3, 4, 5\}$.

Solution. All of A_1, A_2, A_3 must contain 3, and must be contained in the set $\{2, 3, 4, 5\}$.

Moreover, 2 must be contained in both A_1 and A_2 , but cannot be contained in A_3 , or else it would be contained in $A_1 \cap A_3$. Similarly 4 is contained in A_1 and A_3 but not A_2 , and 5 is contained in A_2 and A_3 but not A_1 . From this we obtain

$$A_1 = \{2, 3, 4\} \quad A_2 = \{2, 3, 5\} \quad A_3 = \{3, 4, 5\}$$

□

1. HOMEWORK 1

Exercise 1.1. Let A, B, C be subsets of a set Ω . Prove that

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

Then, show that

$$(A^c)^c = A.$$

Solution. We consider all $2^3 = 8$ possibilities for the three statements “ $x \in A$ ”, “ $x \in B$ ”, “ $x \in C$ ” being true or false. These eight possibilities correspond to the eight rows of the following truth table. The final two columns of the truth table correspond to each of the sets $A \cap (B \cup C)$ and $(A \cap B) \cup (A \cap C)$. Since these two columns of the truth table are identical, we conclude that the $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

$x \in A$	$x \in B$	$x \in C$	$x \in B \cup C$	$x \in A \cap B$	$x \in A \cap C$	$x \in (A \cap B) \cup (A \cap C)$	$x \in A \cap (B \cup C)$
True	True	True	True	True	True	True	True
True	True	False	True	True	False	True	True
True	False	True	True	False	True	True	True
False	True	True	True	False	False	False	False
True	False	False	False	False	False	False	False
False	True	False	True	False	False	False	False
False	False	True	True	False	False	False	False
False	False	False	False	False	False	False	False

We now show that $(A^c)^c = A$. We first show that $A \subseteq (A^c)^c$. Let $x \in A$. By definition of the complement, $x \notin A^c$. By definition of the complement again, $x \in (A^c)^c$. That is, $A \subseteq (A^c)^c$.

We now show that $A \supseteq (A^c)^c$. Let $x \in (A^c)^c$. By definition of the complement, $x \notin A^c$. By definition of the complement again, $x \in A$. That is, $A \supseteq (A^c)^c$.

Since $A \supseteq (A^c)^c$ and $A \subseteq (A^c)^c$, we conclude that $A = (A^c)^c$. □

Exercise 1.2. Let $\{A_i\}_{i=1}^\infty$ be subsets of a set Ω . Prove that $\left(\bigcap_{i=1}^\infty A_i\right)^c = \bigcup_{i=1}^\infty A_i^c$.

Solution. Suppose $x \in (\bigcap_{i=1}^\infty A_i)^c$. That is, $x \notin \bigcap_{i=1}^\infty A_i$. Recall that $\bigcap_{i=1}^\infty A_i = \{x \in \Omega : \forall j \geq 1, x \in A_j\}$. Since x is not in the set $\bigcap_{i=1}^\infty A_i$, the negation of the definition of $\bigcap_{i=1}^\infty A_i$ applies to x . That is, x satisfies the negation of the statement: “*forall* positive integers $j \geq 1$, $x \in A_j$ ”. The negation of this statement is: “ $\exists$ a positive integer $j \geq 1$ such that $x \notin A_j$.” That is, $\exists$ a positive integers $j \geq 1$ such that $x \in A_j^c$. By the definition of countable union, we conclude that $x \in \bigcup_{i=1}^\infty A_i^c$.

So, we showed that $(\bigcap_{i=1}^\infty A_i)^c \subseteq \bigcup_{i=1}^\infty A_i^c$. To conclude, we must show that $(\bigcap_{i=1}^\infty A_i)^c \supseteq \bigcup_{i=1}^\infty A_i^c$. So, let $x \in \bigcup_{i=1}^\infty A_i^c$. By reversing the above implications, we conclude that $x \in (\bigcap_{i=1}^\infty A_i)^c$. That is, $(\bigcap_{i=1}^\infty A_i)^c \supseteq \bigcup_{i=1}^\infty A_i^c$, and the proof is complete. □

Exercise 1.3. Let $\{A_i\}_{i=1}^\infty$ and B be subsets of a set Ω . Prove that $\left(\bigcup_{i=1}^\infty A_i\right) \cap B = \bigcup_{i=1}^\infty A_i \cap B$.

Solution. Let $x \in \left(\bigcup_{i=1}^\infty A_i\right) \cap B$. By definition of intersection, the previous sentence is equivalent to: $x \in B$ and $x \in \bigcup_{i=1}^\infty A_i$. By definition of countable union, the previous sentence is equivalent to: $x \in A_j$ for some $j \geq 1$, and $x \in B$. By the definition of intersection, the previous is equivalent to: $x \in A_j \cap B$ for some $j \geq 1$. So, by definition of countable union, the previous statement is equivalent to: $x \in \bigcup_{i=1}^\infty A_i \cap B$.

We proved the equivalence of $x \in \left(\bigcup_{i=1}^\infty A_i\right) \cap B$ and $x \in \bigcup_{i=1}^\infty A_i \cap B$. We conclude that these two sets are equal. □

Exercise 1.4. Let n be a positive integer, and Ω a set with n elements. For each $A \subseteq \Omega$, define $\mathbf{P}(A) = \frac{|A|}{n}$, where $|A|$ is the number of elements in A . Show that $\mathbf{P}$ is a probability law.

Solution. We first verify that axiom (i) holds. For any $A \subseteq \Omega$, $|A| \geq 0$, so that $\mathbf{P}(A) = |A|/n \geq 0$ as well, i.e. axiom (i) holds. We now verify that axiom (iii) holds. Since Ω has n elements, by definition of $\mathbf{P}$ we have $\mathbf{P}(\Omega) = n/n = 1$.

We now verify that axiom (ii) holds. If A and B are disjoint subsets of Ω , then $|A \cup B| = |A| + |B|$. So, by definition of $\mathbf{P}$, we have

$$\mathbf{P}(A \cup B) = \frac{|A \cup B|}{n} = \frac{|A|}{n} + \frac{|B|}{n} = \mathbf{P}(A) + \mathbf{P}(B).$$

The last equality used the definition of $\mathbf{P}(A)$ and $\mathbf{P}(B)$.

More generally, if $A_1, A_2, \dots \subseteq \Omega$ are pairwise disjoint (that is, $A_i \cap A_j = \emptyset$ for all $i \neq j$, $i, j \geq 1$), then the $|\cup_{i=1}^{\infty} A_i| = \sum_{i=1}^{\infty} |A_i|$. (We note in passing that only finitely many of the A_i 's will be nonempty.) So, by definition of $\mathbf{P}$, we have

$$\mathbf{P}(\cup_{i=1}^{\infty} A_i) = \frac{|\cup_{i=1}^{\infty} A_i|}{n} = \sum_{i=1}^{\infty} \frac{|A_i|}{n} = \sum_{i=1}^{\infty} \mathbf{P}(A_i).$$

The last equality used the definition of $\mathbf{P}(A_i)$ for all $i \geq 1$. In conclusion, axiom (ii) holds. $\square$

Exercise 1.5. Let $\Omega = \mathbb{R}^2$, and define a probability law by

$$\mathbf{P}(A) = \frac{1}{2\pi} \int_A e^{-\frac{(x^2+y^2)}{2}} dx dy$$

Verify that $\mathbf{P}(\Omega) = 1$. Then compute $\mathbf{P}(A)$ when $A = \{(x, y) : x^2 + y^2 \leq 1\}$.

Solution. Using polar coordinates, we have

$$\mathbf{P}(\Omega) = \frac{1}{2\pi} \int_{\mathbb{R}^2} e^{-\frac{(x^2+y^2)}{2}} dx dy = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} r e^{-\frac{r^2}{2}} dr d\theta = \int_0^{\infty} r e^{-\frac{r^2}{2}} dr$$

Setting $u = \frac{r^2}{2}$ then yields

$$\int_0^{\infty} r e^{-\frac{r^2}{2}} dr = \int_0^{\infty} e^{-u} du = 1$$

Again using polar coordinates yields

$$\mathbf{P}(A) = \frac{1}{2\pi} \int_0^{2\pi} \int_0^1 r e^{-\frac{r^2}{2}} dr d\theta = \int_0^1 r e^{-\frac{r^2}{2}} dr$$

Setting $u = \frac{r^2}{2}$ leads to

$$\int_0^1 r e^{-\frac{r^2}{2}} dr = \int_0^{\frac{1}{2}} e^{-u} du = -e^{-u} \Big|_0^{\frac{1}{2}} = 1 - e^{-\frac{1}{2}}$$

$\square$

Exercise 1.6. Let A, B be subsets of a set Ω .

Prove that $A = (A \setminus B) \cup (A \cap B)$ and $(A \setminus B) \cap (A \cap B) = \emptyset$.

Then, prove that $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$, and that these sets are pairwise disjoint.

Solution. We first prove that $(A \setminus B) \cap (A \cap B) = \emptyset$. Recall that $A \setminus B = A \cap B^c$. So, using this definition, commuting the intersections, and using $B \cap B^c = \emptyset$,

$$(A \setminus B) \cap (A \cap B) = (A \cap B^c) \cap (A \cap B) = (B \cap B^c) \cap A \cap A = \emptyset \cap A = \emptyset.$$

We now show that $A = (A \setminus B) \cup (A \cap B)$. We first show that $A \subseteq (A \setminus B) \cup (A \cap B)$. Let $x \in A$. Then either $x \in B$ or $x \notin B$. If $x \in B$, then since $x \in A$ as well, the definition of intersection implies that $x \in A \cap B$. If $x \notin B$ then $x \in B^c$ by definition of complement, and since $x \in A$ as well, we have by definition of intersection that $x \in A \cap B^c$, i.e. $x \in A \setminus B$. So, in any case either $x \in A \cap B$ or $x \in A \cap B^c$. So, by definition of union, we have $x \in (A \cap B^c) \cup (A \cap B)$.

We now show that $A \supseteq (A \setminus B) \cup (A \cap B)$. Let $x \in (A \setminus B) \cup (A \cap B)$. By definition of union, $x \in A \cap B$ or $x \in A \cap B^c$. In either case, by definition of intersection, we have $x \in A$. We have therefore shown that $A \supseteq (A \setminus B) \cup (A \cap B)$.

Combining $A \supseteq (A \setminus B) \cup (A \cap B)$ with $A \subseteq (A \setminus B) \cup (A \cap B)$ implies that $A = (A \setminus B) \cup (A \cap B)$.

We now show that $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$, and that these sets are pairwise disjoint.

We first show disjointness. Since $A \setminus B = A \cap B^c$, commuting the intersections, and using $B \cap B^c = \emptyset$,

$$(A \setminus B) \cap (B \setminus A) = (A \cap B^c) \cap (B \cap A^c) = (B \cap B^c) \cap (A \cap A^c) = \emptyset \cap \emptyset = \emptyset.$$

Similarly,

$$(A \setminus B) \cap (A \cap B) = (A \cap B^c) \cap (A \cap B) = (B \cap B^c) \cap A \cap A = \emptyset \cap A = \emptyset.$$

$$(B \setminus A) \cap (A \cap B) = (B \cap A^c) \cap (A \cap B) = (A \cap A^c) \cap B \cap B = \emptyset \cap B = \emptyset.$$

In conclusion, the three sets $(A \setminus B)$, $(B \setminus A)$, $(A \cap B)$ are pairwise disjoint.

The three sets $A \setminus B$, $B \setminus A$, and $A \cap B$ are subsets of $A \cup B$, hence so is their union. That is, $(A \setminus B) \cup (B \setminus A) \cup (A \cap B) \subseteq A \cup B$. On the other hand, if $x \in A \cup B$, then x by definition of union, $x \in A$ or $x \in B$. So, x is either in both A and B or x is in exactly one of them. That is, $x \in A \cap B$ or $x \in A \cap B^c$ or $x \in B \cap A^c$. So, by definition of union, $x \in (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$. In conclusion, $(A \setminus B) \cup (B \setminus A) \cup (A \cap B) \supseteq A \cup B$. Combined with $(A \setminus B) \cup (B \setminus A) \cup (A \cap B) \subseteq A \cup B$, we get the claim

$$A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$$

□

Exercise 1.7. Let Ω be a sample space, and $\mathbf{P}$ a probability law on Ω . Let A, B, C be subsets of Ω .

- Prove that $\mathbf{P}(A \cup B) \leq \mathbf{P}(A) + \mathbf{P}(B)$.
- Prove that $\mathbf{P}(A \cup B \cup C) = \mathbf{P}(A) + \mathbf{P}(A^c \cap B) + \mathbf{P}(A^c \cap B^c \cap C)$.

Solution. By the previous problem, we can write

$$A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$$

with these three sets being pairwise disjoint. Therefore, by axiom (i) for $\mathbf{P}$,

$$\mathbf{P}(A \cup B) = \mathbf{P}(A \setminus B) + \mathbf{P}(B \setminus A) + \mathbf{P}(A \cap B) \quad (*)$$

By the previous problem again, $B = (A \cap B) \cup (B \setminus A)$ with these two sets disjoint, hence

$$\mathbf{P}(B) = \mathbf{P}(A \cap B) + \mathbf{P}(B \setminus A) \quad (**)$$

Similarly, interchanging the roles of A and B gives

$$\mathbf{P}(A) = \mathbf{P}(A \cap B) + \mathbf{P}(A \setminus B) \quad (***)$$

Combining $(*)$, $(**)$ and $(***)$ gives

$$\begin{aligned} \mathbf{P}(A \cup B) &\stackrel{(*)}{=} \mathbf{P}(A \setminus B) + \mathbf{P}(B \setminus A) + \mathbf{P}(A \cap B) \\ &\stackrel{(***)}{=} \mathbf{P}(A) + \mathbf{P}(B \setminus A) \stackrel{(**)}{=} \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B) \leq \mathbf{P}(A) + \mathbf{P}(B). \end{aligned}$$

The last inequality used axiom (i) to get $\mathbf{P}(A \cap B) \geq 0$, i.e. $-\mathbf{P}(A \cap B) \leq 0$.

We now show that $\mathbf{P}(A \cup B \cup C) = \mathbf{P}(A) + \mathbf{P}(A^c \cap B) + \mathbf{P}(A^c \cap B^c \cap C)$.

An element of $A \cup B \cup C$ is either in A or A^c , and in the latter case it must also be in $B \cup C$, by the definition of the union $A \cup B \cup C$. Therefore

$$A \cup B \cup C = A \cup (A^c \cap (B \cup C))$$

Similarly

$$B \cup C = B \cup (B^c \cap C)$$

so Exercise 1.1 implies that

$$A \cup (A^c \cap (B \cup C)) = A \cup (A^c \cap B) \cup (A^c \cap B^c \cap C)$$

Combining the above shows that

$$A \cup B \cup C = A \cup (A^c \cap B) \cup (A^c \cap B^c \cap C).$$

The three sets on the right are pairwise disjoint, hence by axiom (i) for $\mathbf{P}$

$$\mathbf{P}(A \cup B \cup C) = \mathbf{P}(A) + \mathbf{P}(A^c \cap B) + \mathbf{P}(A^c \cap B^c \cap C)$$

To see the pairwise disjointness, note that

$$\begin{aligned} A \cap (A^c \cap B) &= (A \cap A^c) \cap B = \emptyset \cap B = \emptyset \\ A \cap (A^c \cap B^c \cap C) &= (A \cap A^c) \cap B^c \cap C = \emptyset \cap B^c \cap C = \emptyset \\ (A^c \cap B) \cap (A^c \cap B^c \cap C) &= (B \cap B^c) \cap A^c \cap C = \emptyset \cap A^c \cap C = \emptyset. \end{aligned}$$

□

Exercise 1.8. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a function. Show that

$$\bigcup_{y \in \mathbb{R}} \{x \in \mathbb{R} : f(x) = y\} = \mathbb{R}$$

and that the sets on the left-hand side are disjoint.

Solution. Let $y, z \in \mathbb{R}$ with $y \neq z$. We first show that $\{x \in \mathbb{R} : f(x) = y\} \cap \{x \in \mathbb{R} : f(x) = z\} = \emptyset$. Since f is a function, there cannot exist $x \in \mathbb{R}$ such that $f(x) = y$ and $f(x) = z$, since $y \neq z$. We conclude that $\{x \in \mathbb{R} : f(x) = y\} \cap \{x \in \mathbb{R} : f(x) = z\} = \emptyset$.

We now show that $\cup_{y \in \mathbb{R}} \{x \in \mathbb{R} : f(x) = y\} = \mathbb{R}$. Each set $\{x \in \mathbb{R} : f(x) = y\}$ by its definition is a subset of $\mathbb{R}$, hence $\cup_{y \in \mathbb{R}} \{x \in \mathbb{R} : f(x) = y\} \subseteq \mathbb{R}$, by the definition of union. We now show that $\cup_{y \in \mathbb{R}} \{x \in \mathbb{R} : f(x) = y\} \supseteq \mathbb{R}$. Let $z \in \mathbb{R}$. Define y so that $y = f(z)$. Since $f(z) = y$, we have $z \in \{x \in \mathbb{R} : f(x) = y\}$. So, by definition of union, $\cup_{y \in \mathbb{R}} \{x \in \mathbb{R} : f(x) = y\} \supseteq \mathbb{R}$. Combined with $\cup_{y \in \mathbb{R}} \{x \in \mathbb{R} : f(x) = y\} \subseteq \mathbb{R}$, we conclude that $\cup_{y \in \mathbb{R}} \{x \in \mathbb{R} : f(x) = y\} = \mathbb{R}$. $\square$

2. HOMEWORK 2

Exercise 2.1. Two fair coins are flipped. If at least one of the coins lands heads, what is the probability that the first coin is heads?

Solution. Let H denotes a coin flip of “heads” and let T denote “tails.” The sample space is

$$\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$$

with each of the four elements of Ω having probability $\frac{1}{4}$. That is, $\mathbf{P}$ is uniform on Ω . If A is the event that the first coin is heads and B is the event that at least one coin is heads, then $A = \{(H, T), (H, H)\}$, $B = \{(H, H), (H, T), (T, H)\}$, $A \subseteq B$. So, letting $|\cdot|$ denote the number of elements of a set, we have $|A \cap B| = |A| = 2$ and $|B| = 3$. So, by definition of $\mathbf{P}$ being uniform on Ω ,

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{2}{3}$$

$\square$

Exercise 2.2. In the Monty Hall problem, show that switching wins with probability $\frac{2}{3}$ and staying put wins with probability $\frac{1}{3}$.

Solution. At the beginning of the game, there are three possibilities, each of which is equally likely. Either you select the door with the prize, or you select one of the other two doors which does not contain the prize. In the first case, you will lose by switching your door choice. In the two second cases, the host will open the empty door you did not select. So, in the two second cases, you will win if you switch your door choice. So, in exactly two out of three cases, you will win by switching your door choice. That is, you will win with probability $2/3$ by switching, and you will win with probability $1/3$ by keeping your door choice the same. $\square$

Exercise 2.3. Suppose that you roll 3 distinct, fair, four-sided dice. What is the probability that the sum of the dice is 7?

Solution. The sample space is the set of ordered triples of integers between 1 and 4:

$$\Omega = \{(x, y, z) : 1 \leq x, y, z \leq 4\}$$

which has cardinality $4^3 = 64$. Also, $\mathbf{P}$ is uniform on Ω . The possible (unordered) values for the dice to sum up to 7 are

$$1, 2, 4 \quad 1, 3, 3, \quad 2, 2, 3$$

There are $3! = 6$ different ways to order the triple $(1, 2, 4)$, while there are $\frac{3!}{2!} = 3$ different ways to order the triples $(1, 3, 3)$ and $(2, 2, 3)$, respectively. Therefore the probability is $\frac{6+3+3}{64} = \frac{12}{64} = \frac{3}{16}$. $\square$ *Solution.* The sample space is the set of ordered triples of integers between 1 and 4:

$$\Omega = \{(x, y, z) : 1 \leq x, y, z \leq 4\}$$

which has cardinality $4^3 = 64$. Also, $\mathbf{P}$ is uniform on Ω . For each $1 \leq i \leq 4$, let A_i be the event that the first roll is $1 \leq i \leq 4$. Then $A_i \cap A_j = \emptyset$ for all $i \neq j$ with $1 \leq i, j \leq 4$ (since it is not possible for the first die roll to have two different values) and $\cup_{i=1}^4 A_i = \Omega$ (since the first die roll must be something between 1 and 4, inclusive). So, by the Total Probability Theorem, if B is the event that the sum of the rolls is 7,

$$\mathbf{P}(B) = \sum_{i=1}^4 \mathbf{P}(B|A_i)\mathbf{P}(A_i). \quad (*)$$

Since $\mathbf{P}$ is uniform on Ω we have $\mathbf{P}(A_i) = 1/4$. Also, the event $B|A_i$ is the event that the last two rolls sum to $7 - i$. So, e.g. $B|A_1$ has three elements, namely when the last two rolls are $(2, 4)$, $(3, 3)$ or $(4, 2)$, so that $\mathbf{P}(B|A_1) = 3/16$. Using similar reasoning we get

$$\mathbf{P}(B|A_1) = 3/16, \quad \mathbf{P}(B|A_2) = 4/16, \quad \mathbf{P}(B|A_3) = 3/16, \quad \mathbf{P}(B|A_4) = 2/16.$$

So,

$$\mathbf{P}(B) \stackrel{(*)}{=} \frac{1}{4} \sum_{i=1}^4 \mathbf{P}(B|A_i) = \frac{1}{4 \cdot 16} (3 + 4 + 3 + 2) = \frac{12}{64} = \frac{3}{16}.$$

$\square$

Exercise 2.4. Two people take turns throwing at a dartboard. Person A goes first, and has probability $\frac{1}{4}$ of hitting the bullseye on each throw. Person B goes second, and has probability $\frac{1}{3}$ of hitting the bullseye. Then person A throws after person B, and so on. What is the probability that person A hits the bullseye before person B?

Solution. Let E be the event that person A hits the bullseye first, and E_n the event that person A hits the bullseye before the other person on person A's n th throw, where $n \geq 1$. Then

$$E = \bigcup_{n=1}^{\infty} E_n$$

and the E_n are disjoint ($E_n \cap E_m = \emptyset$ for all $n \neq m$, $n, m \geq 1$), hence by axiom (ii) for $\mathbf{P}$,

$$\mathbf{P}(E) = \sum_{n=1}^{\infty} \mathbf{P}(E_n)$$

If person A hits the bullseye on their n th throw, then they hit the bullseye on the n th throw and missed it on the previous $n - 1$ throws. Also, person B must have missed the bullseye $n - 1$ times. Therefore (by e.g. independence of each individual throw)

$$\mathbf{P}(E_n) = \frac{1}{4} \left(\frac{3}{4}\right)^{n-1} \left(\frac{2}{3}\right)^{n-1} = \frac{1}{4} \left(\frac{1}{2}\right)^{n-1}$$

and therefore

$$\mathbf{P}(E) = \sum_{n=1}^{\infty} \mathbf{P}(E_n) = \frac{1}{4} \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^{n-1} = \frac{1}{4} \sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^n = \frac{2}{4} = \frac{1}{2}$$

□

Exercise 2.5. Suppose that you roll two distinct, fair, six-sided dice two separate times. What is the probability that both rolls have the same sum?

Solution. For each $2 \leq i \leq 12$, let A_i be the event that the first two dice sum to i . Then $A_i \cap A_j = \emptyset$ for all $i \neq j$ with $2 \leq i, j \leq 12$ (since it is not possible for the first two rolls to have two different sums) and $\cup_{i=2}^{12} A_i = \Omega$ (since the first two die rolls' sum must be something between 2 and 12, inclusive). So, by the Total Probability Theorem, if B is the event that both separate rolls of two dice have the same sum,

$$\mathbf{P}(B) = \sum_{i=2}^{12} \mathbf{P}(B|A_i) \mathbf{P}(A_i). \quad (*)$$

Since $\mathbf{P}$ is uniform on Ω we have

$$\mathbf{P}(A_2) = \frac{1}{36} \quad \mathbf{P}(A_3) = \frac{2}{36} \quad \mathbf{P}(A_4) = \frac{3}{36} \quad \mathbf{P}(A_5) = \frac{4}{36} \quad \mathbf{P}(A_6) = \frac{5}{36}$$

$$\mathbf{P}(A_7) = \frac{6}{36} \quad \mathbf{P}(A_8) = \frac{5}{36} \quad \mathbf{P}(A_9) = \frac{4}{36} \quad \mathbf{P}(A_{10}) = \frac{3}{36} \quad \mathbf{P}(A_{11}) = \frac{2}{36} \quad \mathbf{P}(A_{12}) = \frac{1}{36}$$

Also, the event $B|A_i$ is the event that the last two rolls sum to i . Consequently, $\mathbf{P}(B|A_i) = \mathbf{P}(A_i)$ for all $2 \leq i \leq 12$. So,

$$\mathbf{P}(B) \stackrel{(*)}{=} \sum_{k=2}^{12} [\mathbf{P}(A_i)]^2 = \frac{73}{648} \approx 0.113$$

□

Exercise 2.6. Around 5% of men are colorblind, and around .25% of women are colorblind. Given that someone is colorblind, what is the probability that they are a man? (For the purpose of this problem, half of all people are men, and the other half are women.)

Solution. Let C be the event that a person is colorblind, M the event that the person is a man, and W the event that the person is a woman. Assume that $\mathbf{P}(M) = \mathbf{P}(W) = \frac{1}{2}$. Note that $M = W^c$ and $M \cup W = \Omega$.

Then, using Bayes' Theorem (as in the Medical Testing Example)

$$\mathbf{P}(M|C) = \frac{\mathbf{P}(C|M)\mathbf{P}(M)}{\mathbf{P}(C|M)\mathbf{P}(M) + \mathbf{P}(C|W)\mathbf{P}(W)} = \frac{\frac{1}{20} \cdot \frac{1}{2}}{\frac{1}{20} \cdot \frac{1}{2} + \frac{1}{400} \cdot \frac{1}{2}} = \frac{20}{21} \approx .95$$

□

Exercise 2.7. Two people are flipping fair coins. Let n be a positive integer. Person I flips $n + 1$ coins. Person II flips n coins. Show that the following event has probability $1/2$: Person I has more heads than Person II .

Solution 1. Let A be the event that Person I has more heads than Person II . Let S_I be the number of heads from the first n coin flips of person I . Let S_{II} be the number of heads from the first n coin flips of person II . Let B_1 be the event that the $(n+1)^{st}$ coin flip of person I is heads. Let B_2 be the event that the $(n+1)^{st}$ coin flip of person I is tails. Then $B_1 \cap B_2 = \emptyset$ since the $(n+1)^{st}$ coin flip cannot be both heads and tails. And $B_1 \cup B_2 = \Omega$, since the $(n+1)^{st}$ coin flip must be either heads or tails. So, by the total probability theorem,

$$\mathbf{P}(A) = \mathbf{P}(A|B_1)\mathbf{P}(B_1) + \mathbf{P}(A|B_2)\mathbf{P}(B_2).$$

Now, since the $(n+1)^{st}$ coin flip is a fair coin, $\mathbf{P}(B_1) = \mathbf{P}(B_2) = 1/2$. That is,

$$\mathbf{P}(A) = \frac{1}{2} (\mathbf{P}(A|B_1) + \mathbf{P}(A|B_2)).$$

Given that B_1 occurs, the event A is equal to the event that $S_I \geq S_{II}$. Given that B_2 occurs, the event A is equal to the event $S_I > S_{II}$. So,

$$\mathbf{P}(A) = \frac{1}{2} (\mathbf{P}(S_I \geq S_{II}) + \mathbf{P}(S_I > S_{II})).$$

Now, $\mathbf{P}(S_I > S_{II}) = \mathbf{P}(S_I < S_{II})$ by symmetry (with respect to interchanging the roles of person I and person II). So,

$$\mathbf{P}(A) = \frac{1}{2} (\mathbf{P}(S_I \geq S_{II}) + \mathbf{P}(S_I < S_{II})) = \frac{1}{2}.$$

In the last line, we used that the events $S_I \geq S_{II}$ and $S_I < S_{II}$ are disjoint, and their union is all of Ω , so $\mathbf{P}(S_I \geq S_{II}) + \mathbf{P}(S_I < S_{II}) = 1$. $\square$

Solution 2. Let A be the event that Person I has more heads than Person II . Let B be the event that person I has more heads than person II after they both flip n coins. Let C be the event that person I has less heads than person II after they both flip n coins. Let D be the event that person I has the same number of heads as person II after they both flip n coins. Then $B \cap C = C \cap D = B \cap D = \emptyset$, since any such intersection involves mutually exclusive events. Also, $B \cup C \cup D = \Omega$, since after the players each flip n coins, one of the three events B, C, D must occur.

So, by the total probability theorem,

$$\mathbf{P}(A) = \mathbf{P}(A|B)\mathbf{P}(B) + \mathbf{P}(A|C)\mathbf{P}(C) + \mathbf{P}(A|D)\mathbf{P}(D).$$

Given that B has occurred, we already know that A has occurred, so that $\mathbf{P}(A|B) = 1$. Given that C has occurred, it is impossible for A to occur, so that $\mathbf{P}(A|C) = 0$. And given that D has occurred, person I has only one more coin flip; if it is a heads, then A occurs, and if it is tails, then A does not occur. Since the coin is fair, we conclude that $\mathbf{P}(A|D) = 1/2$. That is,

$$\mathbf{P}(A) = \mathbf{P}(B) + \frac{1}{2}\mathbf{P}(D) = \frac{1}{2}(2\mathbf{P}(B) + \mathbf{P}(D)).$$

To conclude, it remains to show that $2\mathbf{P}(B) + \mathbf{P}(D) = 1$. As noted already, $B \cap C = C \cap D = B \cap D = \emptyset$, and $B \cup C \cup D = \Omega$, so Axiom (ii) for Probability Laws says that

$$\mathbf{P}(B) + \mathbf{P}(C) + \mathbf{P}(D) = \mathbf{P}(B \cup C \cup D) = \mathbf{P}(\Omega) = 1.$$

Now, events B and C are symmetric with respect to relabeling the players I and II . Consequently, $\mathbf{P}(B) = \mathbf{P}(C)$. That is, $2\mathbf{P}(B) + \mathbf{P}(D) = 1$, as desired. $\square$

Solution 3. Let C_1 be the number of heads of Person I . Let C_2 be the number of heads of Person II . Let $A = \{C_1 > C_2\}$. Since $A \cup A^c = \Omega$ and $A \cap A^c = \emptyset$, we have $\mathbf{P}(A) + \mathbf{P}(A^c) = 1$. Note that $A^c = \{C_1 \leq C_2\}$. Since the coins are fair, the probability $\mathbf{P}(A^c)$ can be equivalently stated by relabeling the head and tail of the coin. That is, $\mathbf{P}(A^c)$ is equal to the probability of the event that Person I has less than or equal to the number of tails of Person II . The latter event is equal to $\{C_1 > C_2\}$. That is, $\mathbf{P}(A^c) = \mathbf{P}(C_1 > C_2) = \mathbf{P}(A)$. So, $2\mathbf{P}(A) = 1$, and $\mathbf{P}(A) = 1/2$. $\square$

Exercise 2.8. Suppose that a test for a disease is 99.9% accurate, in that if you have the disease then you will test positive with probability 99.9%, and if you do not have the disease then you will test negative with probability 99.9%. Suppose also that the prevalence of the disease is $\frac{1}{20000}$. If you test positive for the disease, what is the probability that you actually have the disease?

Solution. Let D be the event that you have the disease, P the event that you test positive, and N the event that you test negative. Then using Bayes' Theorem (as in the Medical Testing Example in class)

$$\mathbf{P}(D|P) = \frac{\mathbf{P}(P|D)\mathbf{P}(D)}{\mathbf{P}(P|D)\mathbf{P}(D) + \mathbf{P}(P|D^c)\mathbf{P}(D^c)}$$

We know that $\mathbf{P}(D) = \frac{1}{20,000}$, and that $\mathbf{P}(P|D) = \frac{999}{1000}$. Moreover, $\mathbf{P}(D^c) = 1 - \mathbf{P}(D) = \frac{19999}{20000}$ and $\mathbf{P}(P|D^c) = \frac{1}{1000}$. Therefore

$$\frac{\mathbf{P}(P|D)\mathbf{P}(D)}{\mathbf{P}(P|D)\mathbf{P}(D) + \mathbf{P}(P|D^c)\mathbf{P}(D^c)} = \frac{\frac{999}{1000} \cdot \frac{1}{20,000}}{\frac{999}{1000} \cdot \frac{1}{20,000} + \frac{1}{1000} \cdot \frac{19999}{20000}} = \frac{999}{20998} \approx 0.047576$$

$\square$

3. HOMEWORK 3

Exercise 3.1. Let $\Omega = [0, 1] \times [0, 1]$. and define $\mathbf{P}(A)$ to be the area of A . Suppose that $A = \{(x, y) : a_1 \leq x \leq a_2\}$ and $B = \{(x, y) \in \Omega : b_1 \leq y \leq b_2\}$, where $0 \leq a_1 \leq a_2 \leq 1$ and $0 \leq b_1 \leq b_2 \leq 1$. Show that A and B are independent.

Solution. A is a rectangle with area $a_2 - a_1$, and B is a rectangle with area $b_2 - b_1$. And

$$A \cap B = \{(x, y) \in \Omega : a_1 \leq x \leq a_2, b_1 \leq y \leq b_2\}$$

is a rectangle of area $(a_2 - a_1)(b_2 - b_1)$. Therefore

$$\mathbf{P}(A \cap B) = (a_2 - a_1)(b_2 - b_1) = \mathbf{P}(A)\mathbf{P}(B)$$

so A and B are independent. $\square$

Exercise 3.2. Let $\Omega = \mathbb{R}^2$, and define

$$\mathbf{P}(A) = \frac{1}{2\pi} \int_A e^{-\frac{(x^2+y^2)}{2}} dx dy$$

Suppose that $A = \{(x, y) : a_1 \leq x \leq a_2\}$ and $B = \{(x, y) \in \Omega : b_1 \leq y \leq b_2\}$, where $a_1 \leq a_2$ and $b_1 \leq b_2$. Show that A and B are independent.

Solution. If $A = \{(x, y) : a_1 \leq x \leq a_2\}$, then

$$\mathbf{P}(A) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{a_1}^{a_2} e^{-\frac{x^2+y^2}{2}} dx dy$$

and similarly if $B = \{(x, y) : b_1 \leq y \leq b_2\}$, then

$$\mathbf{P}(B) = \frac{1}{2\pi} \int_{b_1}^{b_2} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dx dy$$

Since $A \cap B = \{(x, y) : a_1 \leq x \leq a_2, b_1 \leq y \leq b_2\}$, we have

$$\mathbf{P}(A \cap B) = \frac{1}{2\pi} \int_{b_1}^{b_2} \int_{a_1}^{a_2} e^{-\frac{x^2+y^2}{2}} dx dy$$

Finally,

$$\begin{aligned} \mathbf{P}(A)\mathbf{P}(B) &= \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{a_1}^{a_2} e^{-\frac{x^2+y^2}{2}} dx dy \right] \left[\frac{1}{2\pi} \int_{b_1}^{b_2} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dx dy \right] \\ &= \frac{1}{4\pi^2} \left[\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \right] \left[\int_{a_1}^{a_2} e^{-\frac{x^2}{2}} dx \right] \left[\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right] \left[\int_{b_1}^{b_2} e^{-\frac{y^2}{2}} dy \right] \\ &= \left[\frac{1}{2\pi} \int_{b_1}^{b_2} \int_{a_1}^{a_2} e^{-\frac{x^2+y^2}{2}} dx dy \right] \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dx dy \right] = \mathbf{P}(A \cap B) \end{aligned}$$

since the second term in brackets is $\mathbf{P}(\Omega) = 1$. Therefore A and B are independent. $\square$

Exercise 3.3. Let Ω be a sample space and $\mathbf{P}$ a probability law on Ω . Suppose that A and B are events with $A \subseteq B$. Is it possible for A to be independent of B ? Justify your answer.

Solution. Suppose that A and B are independent. Then since $A \cap B = A$, we must have

$$\mathbf{P}(A) = \mathbf{P}(A)\mathbf{P}(B)$$

This is possible if and only if $\mathbf{P}(A) = 0$ or $\mathbf{P}(B) = 1$. So it is possible for A and B to be independent, but only under some restrictive assumptions. $\square$

Exercise 3.4. Let Ω be a sample space, and $A_1, \dots, A_n$ events. Show that

$$\mathbf{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbf{P}(A_i) - \sum_{1 \leq i < j \leq n} \mathbf{P}(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} \mathbf{P}(A_i \cap A_j \cap A_k) + \dots + (-1)^{n-1} \mathbf{P}(A_1 \cap \dots \cap A_n)$$

Solution. For any $1 \leq m \leq n$, let B_m be the set of points $x \in \Omega$ such that x appears in exactly m of the sets $A_1, \dots, A_n$. Then the sets $B_1, \dots, B_m$ are disjoint, and $\bigcup_{m=1}^n B_m = \bigcup_{i=1}^n A_i$. So, by axiom (ii) for probability laws,

$$\mathbf{P}(\bigcup_{i=1}^n A_i) = \mathbf{P}(\bigcup_{m=1}^n B_m) = \sum_{m=1}^n \mathbf{P}(B_m).$$

We are therefore required to show that

$$\begin{aligned} \sum_{m=1}^n \mathbf{P}(B_m) &= \sum_{i=1}^n \mathbf{P}(A_i) - \sum_{1 \leq i < j \leq n} \mathbf{P}(A_i \cap A_j) \\ &\quad + \sum_{1 \leq i < j < k \leq n} \mathbf{P}(A_i \cap A_j \cap A_k) + \dots + (-1)^{n-1} \mathbf{P}(A_1 \cap \dots \cap A_n). \end{aligned} \quad (*)$$

Consider now the identity $0 = (1-1)^m = \sum_{k=0}^m (-1)^k \binom{m}{k}$, which follows from the Binomial Theorem. That is, $1 = \sum_{k=1}^m (-1)^{k+1} \binom{m}{k}$. Fix $1 \leq m \leq k$. On the left side of (*), the “number of times” that the elements of B_m are counted is once, and on the right side of (*), the “number of times” the set B_m is counted is $\sum_{k=1}^m (-1)^{k+1} \binom{m}{k}$, which is also 1. We conclude that (*) holds.

[This proof can be made a bit more precise; see the third solution below.] $\square$

Solution. We induct on n . If $n = 1$ then there is nothing to show. The case $n = 2$ was proven in class in Proposition 2.33. So, the base case is proven and we proceed to the inductive step.

For the induction step, suppose that $n \geq 2$ and that the desired result holds for all unions of n events. Then given $n + 1$ events $A_1, \dots, A_n, A_{n+1}$, by the $n = 2$ case we have

$$\mathbf{P}\left(\bigcup_{i=1}^{n+1} A_i\right) = \mathbf{P}\left(\bigcup_{i=1}^n A_i\right) + \mathbf{P}(A_{n+1}) - \mathbf{P}\left(A_{n+1} \cap \left(\bigcup_{i=1}^n A_i\right)\right) \quad (1)$$

Next, the induction hypothesis implies that

$$\mathbf{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbf{P}(A_i) - \sum_{1 \leq i < j \leq n} \mathbf{P}(A_i \cap A_j) + \dots + (-1)^{n-1} \mathbf{P}(A_1 \cap \dots \cap A_n) \quad (2)$$

and since

$$A_{n+1} \cap \bigcup_{i=1}^n A_i = \bigcup_{i=1}^n (A_i \cap A_{n+1})$$

another application of the induction hypothesis shows that

$$\begin{aligned} \mathbf{P}\left(A_{n+1} \cap \left(\bigcup_{i=1}^n A_i\right)\right) &= \mathbf{P}\left(\bigcup_{i=1}^n (A_i \cap A_{n+1})\right) \\ &= \sum_{i=1}^n \mathbf{P}(A_i \cap A_{n+1}) - \sum_{1 \leq i < j \leq n} \mathbf{P}(A_i \cap A_j \cap A_{n+1}) + \dots + (-1)^{n-1} \mathbf{P}(A_1 \cap \dots \cap A_n \cap A_{n+1}) \quad (3) \end{aligned}$$

Substituting equations (2) and (3) into equation (1) and collecting terms completes the induction step. $\square$

Solution. [This proof will only make sense later on in the class when we have covered expected value.] For any $A \subseteq \Omega$, we define $1_A: \Omega \rightarrow \mathbb{R}$ so that $1_A(\omega) = 1$ when $\omega \in A$, and $1_A(\omega) = 0$ whenever $\omega \notin A$. Also, by definition of expected value, $\mathbf{E}1_A = 1 \cdot \mathbf{P}(A) + 0 \cdot \mathbf{P}(A^c) = \mathbf{P}(A)$. It follows by the definition of union that

$$1_{\bigcup_{i=1}^n A_i}(\omega) = 1 - \prod_{i=1}^n (1 - 1_{A_i}(\omega)),$$

since each side is equal to one only when ω is in at least one of the sets $A_1, \dots, A_n$ (in which case the product on the right is zero). Multiplying out the right sides gives

$$\begin{aligned} 1_{\cup_{i=1}^n A_i}(\omega) &= \sum_{i=1}^n 1_{A_i}(\omega) - \sum_{1 \leq i < j \leq n} 1_{A_i}(\omega) 1_{A_j}(\omega) \\ &\quad + \sum_{1 \leq i < j < k \leq n} 1_{A_i}(\omega) 1_{A_j}(\omega) 1_{A_k}(\omega) - \dots + 1_{A_1} \dots 1_{A_n} \\ &= \sum_{i=1}^n 1_{A_i}(\omega) - \sum_{1 \leq i < j \leq n} 1_{A_i \cap A_j}(\omega) \\ &\quad + \sum_{1 \leq i < j < k \leq n} 1_{A_i \cap A_j \cap A_k}(\omega) - \dots + (-1)^{n-1} 1_{A_1 \cap \dots \cap A_n}(\omega). \end{aligned}$$

Taking expected values of both sides completes the proof, since, as we noted above, $\mathbf{E}1_A = \mathbf{P}(A)$, so that

$$\begin{aligned} \mathbf{P}(\cup_{i=1}^n A_i) &= \mathbf{E}1_{\cup_{i=1}^n A_i} \\ &= \sum_{i=1}^n \mathbf{E}1_{A_i} - \sum_{1 \leq i < j \leq n} \mathbf{E}1_{A_i \cap A_j} \\ &\quad + \sum_{1 \leq i < j < k \leq n} \mathbf{E}1_{A_i \cap A_j \cap A_k} - \dots + (-1)^{n-1} \mathbf{E}1_{A_1 \cap \dots \cap A_n} \\ &= \sum_{i=1}^n \mathbf{P}(A_i) - \sum_{1 \leq i < j \leq n} \mathbf{P}(A_i \cap A_j) \\ &\quad + \sum_{1 \leq i < j < k \leq n} \mathbf{P}(A_i \cap A_j \cap A_k) - \dots + (-1)^{n-1} \mathbf{P}(A_1 \cap \dots \cap A_n). \end{aligned}$$

□

Exercise 3.5.

- Suppose that your car has four tires, and your mechanic removes all four, then later puts the tires back on the car randomly. What is the probability that no tire is put on its original wheel?
- Now suppose that your car has n tires, and your mechanic removes all of them, then later puts the tires back on the car randomly. What is the probability that no tire is put on its original wheel?
- What is the limit of the probability in (b) as $n \rightarrow \infty$?

Solution. Let A_i be the event that the i th tire ends up in its original position, for $i = 1, 2, 3, 4$. The union $\cup_{i=1}^4 A_i$ is the event that at least one tire is put on the original wheel, and according to the previous problem,

$$\mathbf{P}\left(\bigcup_{i=1}^4 A_i\right) = \sum_{i=1}^4 \mathbf{P}(A_i) - \sum_{i < j} \mathbf{P}(A_i \cap A_j) + \sum_{i < j < k} \mathbf{P}(A_i \cap A_j \cap A_k) - \mathbf{P}(A_1 \cap A_2 \cap A_3 \cap A_4)$$

Next, $\mathbf{P}(A_i) = \frac{3!}{4!} = \frac{1}{4}$ since there are $4!$ ways to put the tires on the car, and $3!$ ways to put the tires on if tire i must be placed on wheel i .

Similarly, $\mathbf{P}(A_i \cap A_j) = \frac{2!}{4!} = \frac{1}{4 \cdot 3}$ for each pair $i < j$, and $\mathbf{P}(A_i \cap A_j \cap A_k) = \frac{1}{4!} = \mathbf{P}(A_1 \cap A_2 \cap A_3 \cap A_4)$.

Finally, there are $\binom{4}{2} = \frac{4 \cdot 3}{2}$ ways to choose a pair $i < j$, and $\frac{4}{3} = 4$ ways to choose a triple $i < j < k$. Therefore

$$\mathbf{P}\left(\bigcup_{i=1}^4 A_i\right) = 4 \cdot \frac{1}{4} - \frac{4 \cdot 3}{2} \cdot \frac{1}{4 \cdot 3} + 4 \frac{1}{4!} - 1 \cdot \frac{1}{4!} = 1 - \frac{1}{2} + \frac{1}{6} - \frac{1}{24} = \frac{5}{8}$$

Therefore the probability that no tire ends up on its original wheel is

$$1 - \frac{5}{8} = \frac{3}{8}$$

□

Solution. As in part (i), let A_i be the event that the i th tire is put in its original position. If $i_1 < i_2 < \dots < i_k$, then

$$\mathbf{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \frac{(n-k)!}{n!}$$

and moreover there are $\binom{n}{k}$ ways to choose the indices $i_1 < \dots < i_k$. Therefore by problem 4,

$$\mathbf{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k-1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^n \frac{(-1)^{k-1}}{k!}$$

so the probability that no tire ends up on its original wheel is

$$1 - \sum_{k=1}^n \frac{(-1)^{k-1}}{k!} = \sum_{k=0}^n \frac{(-1)^k}{k!}$$

□

Solution. The limit is

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} = \frac{1}{e}$$

□

Exercise 3.6. Let A, B, C be pairwise independent events such that

$$\mathbf{P}(A) = \frac{1}{2} \quad \mathbf{P}(B) = \frac{1}{3} \quad \mathbf{P}(C) = \frac{1}{4} \quad \mathbf{P}(A \cup B \cup C) = \frac{35}{48}$$

Are the sets A, B, C independent? Explain.

Solution. Since A, B, C are pairwise independent, we have

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B) = \frac{1}{6} \quad \mathbf{P}(A \cap C) = \mathbf{P}(A)\mathbf{P}(C) = \frac{1}{8} \quad \mathbf{P}(B \cap C) = \mathbf{P}(B)\mathbf{P}(C) = \frac{1}{12}$$

From problem 4, we know that

$$\mathbf{P}(A \cup B \cup C) = \mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C) - \mathbf{P}(A \cap B) - \mathbf{P}(A \cap C) - \mathbf{P}(B \cap C) + \mathbf{P}(A \cap B \cap C)$$

so solving for $\mathbf{P}(A \cap B \cap C)$ yields

$$\mathbf{P}(A \cap B \cap C) = \frac{35}{48} - \frac{1}{2} - \frac{1}{3} - \frac{1}{4} + \frac{1}{6} + \frac{1}{8} + \frac{1}{12} = \frac{1}{48}$$

On the other hand,

$$\mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C) = \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{24}$$

so A, B, C are not independent. $\square$

4. HOMEWORK 4

Exercise 4.1. You spin a wheel 24 times, and each time there is a 1 in 72 chance of winning a prize. Let X be the number of prizes won. Use a Poisson approximation to estimate $\mathbf{P}(X = 0)$, $\mathbf{P}(X = 1)$, and $\mathbf{P}(X = 2)$.

Solution. X is a binomial random variable with $n = 24$ and $p = \frac{1}{72}$, so X can be approximated by a Poisson random variable Y with parameter $\lambda = \frac{1}{3}$.

Therefore

$$\mathbf{P}(X = 0) \approx \mathbf{P}(Y = 0) = e^{-\frac{1}{3}} \approx 0.7165$$

$$\mathbf{P}(X = 1) \approx \mathbf{P}(Y = 1) = e^{-\frac{1}{3}} \cdot \frac{1}{3} \approx 0.2388$$

and

$$\mathbf{P}(X = 2) \approx \mathbf{P}(Y = 2) = \frac{1}{18}e^{-\frac{1}{3}} \approx 0.0896$$

$\square$

Exercise 4.2. Count the number of distinct ways which you can arrange the letters of the words CATTERPILLAR and ARUGULA.

Solution. There are 12 letters in the word catterpillar, with 2 Ts, 2 Rs, 2 As, and 2 Ls. Therefore the number of words is

$$\frac{12!}{(2!)^4} = 29,937,600$$

Similarly, the number of words which can be formed from arugula is

$$\frac{7!}{(2!)^2} = 1260$$

$\square$

Exercise 4.3. Suppose that X is a random variable with $p_X(-3) = \frac{1}{10}$, $p_X(-2) = \frac{1}{5}$, $p_X(-1) = \frac{3}{20}$, $p_X(0) = \frac{1}{5}$, $p_X(3) = \frac{1}{10}$, $p_X(5) = \frac{3}{20}$, $p_X(6) = \frac{1}{20}$, $p_X(10) = \frac{1}{20}$.

Compute the probabilities of the following events

- $X > 3$
- $4 < X < 7$ or $X > 9$
- $0 < X < 4$ or $7 < X \leq 10$.

Solution. (a) The event $\{X > 3\}$ is equal to the event $\{X = 5, 6, 10\}$, hence by axiom (ii) for probability laws,

$$\mathbf{P}(X > 3) = \mathbf{P}(X = 5) + \mathbf{P}(X = 6) + \mathbf{P}(X = 10) = \frac{5}{20} = \frac{1}{4}$$

(b) If $4 < X < 7$ or $X < 9$, then either $X = 5$, $X = 6$, or $X = 10$, so by axiom (ii) for probability laws,

$$\mathbf{P}(4 < X < 7) = \mathbf{P}(X = 5) + \mathbf{P}(X = 6) + \mathbf{P}(X = 10) = \frac{5}{20} = \frac{1}{4}$$

(c) If $0 < X < 4$ or $7 < X \leq 10$, then either $X = 3$ or $X = 10$, so by axiom (ii) for probability laws,

$$\mathbf{P}(4 < X < 7) = \mathbf{P}(X = 3) + \mathbf{P}(X = 10) = \frac{1}{10} + \frac{1}{20} = \frac{3}{20}$$

□

Exercise 4.4. Suppose that the probability that you receive a prize in the mail is $\frac{1}{7000000}$. Show that you need to receive roughly 7000000 pieces of mail in order for your probability of winning at least one prize to be about $1 - \frac{1}{e}$.

Solution. For any $1 \leq i \leq n$, let A_i be the event that you do not receive a prize in the mail. We assume that $A_1, \dots, A_n$ are independent. By assumption,

$$\mathbf{P}(A_i) = 1 - \frac{1}{7000000}, \quad \forall 1 \leq i \leq n.$$

The probability of not receiving a prize in n pieces of mail is, by independence of $A_1, \dots, A_n$,

$$\mathbf{P}(\cap_{i=1}^n A_i) = \prod_{i=1}^n \mathbf{P}(A_i) = \prod_{i=1}^n \left(1 - \frac{1}{7000000}\right) = \left(1 - \frac{1}{7000000}\right)^n$$

Since

$$\lim_{k \rightarrow \infty} \left(1 - \frac{1}{k}\right)^k = \frac{1}{e}$$

choosing $n = 7000000$, means that

$$\mathbf{P}(\cap_{i=1}^n A_i) \approx \frac{1}{e}.$$

That is, when $n = 7000000$, the probability of receiving a prize is

$$\mathbf{P}([\cap_{i=1}^n A_i]^c) = 1 - \mathbf{P}(\cap_{i=1}^n A_i) \approx 1 - \frac{1}{e}.$$

□

Exercise 4.5. Let $\Omega = \{-3, -2, -1, 0, 1, 2, 3\}$ with the uniform probability law, and suppose that $X(\omega) = \omega$ for all $\omega \in \Omega$. If $f : \mathbb{R} \rightarrow \mathbb{R}$ is defined by $f(x) = x^2$, find the probability mass function of $f(X)$.

Solution. Let $Y = f(X) = X^2$. Then by definition of Y and axiom (ii) for probability laws

$$\mathbf{P}(Y = 0) = \mathbf{P}(X = 0) = \frac{1}{7},$$

$$\mathbf{P}(Y = 1) = \mathbf{P}(X^2 = 1) = \mathbf{P}(\{X = 1\} \cup \{X = -1\}) = \mathbf{P}(X = 1) + \mathbf{P}(X = -1) = \frac{2}{7},$$

$$\mathbf{P}(Y = 4) = \mathbf{P}(X^2 = 4) = \mathbf{P}(\{X = 2\} \cup \{X = -2\}) = \mathbf{P}(X = 2) + \mathbf{P}(X = -2) = \frac{2}{7},$$

$$\mathbf{P}(Y = 9) = \mathbf{P}(X^2 = 9) = \mathbf{P}(\{X = 3\} \cup \{X = -3\}) = \mathbf{P}(X = 3) + \mathbf{P}(X = -3) = \frac{2}{7}.$$

$\mathbf{P}(Y = y) = 0$ for all real numbers $y \neq 0, 1, 4, 9$ (which is also clear from the fact that the above probabilities sum to 1). □

Exercise 4.6. Let X be geometrically distributed with parameter p , let $Y = X^4$, and let $Z = \min\{X, n\}$ for a fixed positive integer n . Find the probability mass functions of Y and Z .

Solution. By definition, the probability mass function of X is given by

$$\mathbf{P}(X = k) = (1 - p)^{k-1}p \quad \forall k = 1, 2, 3, \dots$$

Therefore the only values that Y takes with non-zero probability are 4th powers of positive integers, and

$$\mathbf{P}(Y = k^4) = (1 - p)^{k-1}p \quad \forall k = 1, 2, 3, \dots$$

The random variable Z only takes the values $k = 1, 2, \dots, n$ with non-zero probability. If $1 \leq k < n$ then the events $\{X = k\}$ and $\{\min(X, n) = k\}$ are identical, so

$$\mathbf{P}(Z = k) = \mathbf{P}(\min(X, n) = k) = \mathbf{P}(X = k) = (1 - p)^{k-1}p,$$

where the last line used the definition of X . Meanwhile, the events $\{X \geq n\}$ and $\{\min(X, n) = n\}$ are identical, so

$$\mathbf{P}(Z = n) = \mathbf{P}(X \geq n) = \sum_{k=n}^{\infty} (1 - p)^{k-1}p = \frac{p(1 - p)^{n-1}}{1 - (1 - p)} = (1 - p)^{n-1}$$

Alternately,

$$\begin{aligned} \mathbf{P}(Z = n) &= \mathbf{P}(X \geq n) = 1 - \mathbf{P}(X < n) = 1 - \sum_{k=1}^{n-1} (1 - p)^{k-1}p = 1 - p \frac{1 - (1 - p)^{n-1}}{1 - (1 - p)} \\ &= 1 - (1 - (1 - p)^{n-1}) = (1 - p)^{n-1} \end{aligned}$$

□

5. HOMEWORK 5

Exercise 5.1. Let X be a discrete random variable with finite variance, and define $f : \mathbb{R} \rightarrow \mathbb{R}$ by $f(t) = \mathbf{E}[(X - t)^2]$. Show that $f(t)$ has its unique minimum when $t = \mathbf{E}[X]$.

Solution. $(X - t)^2 = X^2 - 2Xt + t^2$, hence by the linearity of expectation

$$f(t) = \mathbf{E}[(X - t)^2] = \mathbf{E}[X^2] - 2\mathbf{E}[X]t + t^2$$

Therefore $f(t)$ is a quadratic function of t which is concave up, hence has a unique minimum at its vertex, namely

$$t = \mathbf{E}[X]$$

We know this from calculus, since the solution t of the equation $f'(t) = 0$ satisfies $2t - 2\mathbf{E}[X] = 0$, i.e. $t = \mathbf{E}[X]$, and the unique critical point of a concave up parabola is its global minimum.

□

Exercise 5.2. Let n be a positive integer and $p \in [0, 1]$. Compute the mean of a binomial random variable with parameters n and p .

Then, compute the mean of a Poisson random variable with parameter $\lambda > 0$.

Solution. Let X be a binomial random variable with parameters n and p . Recall that X is the number of heads that result from n biased coin flips. That is, for any $1 \leq i \leq n$, if we let $X_i = 1$ if the i^{th} coin flip is heads, and $X_0 = 0$ if the i^{th} coin flip is tails, then

$$X = \sum_{i=1}^n X_i.$$

More specifically, for each $1 \leq i \leq n$, X_i is a Bernoulli random variables with parameter p . Therefore, by definition of expected value,

$$\mathbf{E}[X_i] = 1 \cdot p + 0 \cdot (1 - p) = p.$$

Finally, by linearity of expected value,

$$\mathbf{E}[X] = \sum_{i=1}^n \mathbf{E}X_i = np.$$

(Once we cover independence, we will see that $X_1, \dots, X_n$ are independent random variables.)

Suppose that X is a Poisson random variable with parameter λ . Then by definition of expected value,

$$\begin{aligned} \mathbf{E}[X] &= \sum_{n=0}^{\infty} n \mathbf{P}(X = n) = e^{-\lambda} \sum_{n=1}^{\infty} \frac{n \lambda^n}{n!} = e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} \\ &= \lambda e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = \lambda e^{-\lambda} e^{\lambda} = \lambda. \end{aligned}$$

□

Exercise 5.3. Let X be a random variable on a sample space Ω taking values in the non-negative integers. Show that $\mathbf{E}[X] = \sum_{n=1}^{\infty} \mathbf{P}(X \geq n)$. Use this to compute the mean of a geometric random variable with parameter p .

Solution. The event $\{X \geq n\} = \cup_{k=n}^{\infty} \{X = k\}$, with the events on the right being disjoint. Hence, by the Total Probability Theorem,

$$\sum_{n=1}^{\infty} \mathbf{P}(X \geq n) = \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \mathbf{P}(X = k) = \sum_{n \geq 1} \sum_{k \geq 1: k \geq n} \mathbf{P}(X = k) = \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \mathbf{P}(X = k) 1_{k \geq n}$$

(Here $1_{k \geq n}$ is equal to 1 when $k \geq n$ and it is equal to 0 otherwise.) Since each term in the double sum is non-negative, as suggested in the hint, we can interchange the order of summation, yielding

$$\sum_{k=1}^{\infty} \sum_{n=1}^{\infty} \mathbf{P}(X = k) 1_{k \geq n} = \sum_{k=1}^{\infty} \sum_{n=1}^k \mathbf{P}(X = k) = \sum_{k=1}^{\infty} k \mathbf{P}(X = k) = \sum_{k=0}^{\infty} k \mathbf{P}(X = k) = \mathbf{E}[X].$$

The last line used the definition of expected value.

Now suppose that X is a geometric random variable with parameter $p \in (0, 1]$. Then

$$\mathbf{P}(X \geq n) = \sum_{k=n}^{\infty} p(1-p)^{k-1} = p \frac{(1-p)^{n-1}}{1 - (1-p)} = (1-p)^{n-1}$$

hence

$$\mathbf{E}[X] = \sum_{n=1}^{\infty} \mathbf{P}(X \geq n) = \sum_{n=1}^{\infty} (1-p)^{n-1} = \sum_{m=0}^{\infty} (1-p)^m = \frac{1}{1-(1-p)} = \frac{1}{p}$$

Alternately, $\mathbf{E}[X]$ can be computed directly from the definition:

$$\mathbf{E}[X] = \sum_{n=1}^{\infty} np(1-p)^{n-1} = p \sum_{n=1}^{\infty} n(1-p)^{n-1}$$

This sum is the result of differentiating the function

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n, \quad \forall |x| < 1$$

term-by-term and setting $x = p$. (Term-by-term differentiation is valid when $|x| < 1$, since the Taylor expansion of $1/(1-x)$ converges absolutely for all $|x| < 1$.) Since $\frac{d}{dx} \frac{1}{1-x} = \frac{1}{(1-x)^2}$, the result is

$$\mathbf{E}[X] = \frac{p}{(1-(1-p))^2} = \frac{1}{p}$$

□

Exercise 5.4. Find real numbers $a_{i,j}$ ($i, j \geq 0$) for which $\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_{ij} \neq \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} a_{ij}$.

Solution. Let $a_{ij} = 1$ if $i = j$, $a_{ij} = -1$ if $i = j + 1$, and $a_{ij} = 0$ otherwise. Then for any fixed j ,

$$\sum_{i=0}^{\infty} a_{ij} = 1 + (-1) = 0$$

hence

$$\sum_{j=0}^{\infty} \sum_{i=0}^{\infty} a_{ij} = 0$$

On the other hand, if $i \geq 1$ then

$$\sum_{j=0}^{\infty} a_{ij} = (-1) + 1 = 0$$

while

$$\sum_{j=0}^{\infty} a_{0j} = 1$$

so

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_{ij} = 1$$

If we put the entries a_{ij} into an infinite matrix, they appear as follows:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ -1 & 1 & 0 & 0 & \cdots \\ 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & -1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

□

Exercise 5.5. Let $\Omega = [0, 1]$.

Consider the probability law on Ω defined by $\mathbf{P}([a, b]) = b - a$ for $0 \leq a \leq b \leq 1$. Let n be a positive integer, and X a random variable which is constant on each interval $[\frac{i}{n}, \frac{i+1}{n})$, $0 \leq i \leq n-1$. Show that $\mathbf{E}[X] = \int_0^1 X(t) dt$.

Now, consider the probability law on Ω defined by $\mathbf{P}([a, b]) = \int_a^b \frac{1}{2\sqrt{t}} dt$. If $X(t)$ is as in part (a), show that $\mathbf{E}[X] = \int_0^1 X(t) \frac{1}{2\sqrt{t}} dt$.

Solution.

Since X is constant on each interval $[\frac{i}{n}, \frac{i+1}{n})$, we can write

$$X(t) = X(0)X_0(t) + X\left(\frac{1}{n}\right)X_1(t) + \cdots + X\left(\frac{n-1}{n}\right)X_{n-1}(t), \quad \forall t \in [0, 1]$$

where $X_i(t) = 1$ if $\frac{i}{n} \leq t < \frac{i+1}{n}$ and $X_i(t) = 0$ otherwise, for all $0 \leq i \leq n-1$.

Therefore

$$\mathbf{E}[X] = \sum_{i=0}^{n-1} X\left(\frac{i}{n}\right) \mathbf{E}[X_i] = \frac{1}{n} \sum_{i=0}^{n-1} X\left(\frac{i}{n}\right)$$

On the other hand,

$$\int_0^1 X(t) dt = \sum_{i=0}^{n-1} \int_{\frac{i}{n}}^{\frac{i+1}{n}} X(t) dt = \frac{1}{n} \sum_{i=0}^{n-1} X\left(\frac{i}{n}\right)$$

since X is constant on each interval $[\frac{i}{n}, \frac{i+1}{n})$. Therefore

$$\mathbf{E}[X] = \int_0^1 X(t) dt$$

as claimed.

Essentially the same proof as part (a) works: again write

$$X(t) = X(0)X_0(t) + X\left(\frac{1}{n}\right)X_1(t) + \cdots + X\left(\frac{n-1}{n}\right)X_{n-1}(t)$$

so that

$$\begin{aligned} \mathbf{E}[X] &= \sum_{i=0}^{n-1} X\left(\frac{i}{n}\right) \mathbf{E}[X_i] = \sum_{i=0}^{n-1} X\left(\frac{i}{n}\right) \int_{\frac{i}{n}}^{\frac{i+1}{n}} \frac{1}{2\sqrt{t}} dt \\ &= \sum_{i=0}^{n-1} \int_{\frac{i}{n}}^{\frac{i+1}{n}} X(t) \frac{1}{2\sqrt{t}} dt = \int_0^1 X(t) \frac{1}{2\sqrt{t}} dt \end{aligned}$$

□

Exercise 5.6. Let $b_1, \dots, b_n$ be distinct numbers, representing the quality of n people. Suppose n people arrive to interview for a job, one at a time, in a random order. That is, every possible arrival order of these people is equally likely. We can think of an arrival ordering of the people as an ordered list of the form $a_1, \dots, a_n$, where the list $a_1, \dots, a_n$ is a permutation of the numbers $b_1, \dots, b_n$. Moreover, we interpret a_1 as the rank of the first

person to arrive, a_2 as the rank of the second person to arrive, and so on. And all possible permutations of the numbers $b_1, \dots, b_n$ are equally likely to occur.

For each $i \in \{1, \dots, n\}$, upon interviewing the i^{th} person, if $a_i > a_j$ for all $1 \leq j < i$, then the i^{th} person is hired. That is, if the person currently being interviewed is better than the previous candidates, she will be hired. What is the expected number of hirings that will be made? (Hint: let $X_i = 1$ if the i^{th} person to arrive is hired, and let $X_i = 0$ otherwise. Consider $\sum_{i=1}^n X_i$.)

Solution. Let $X_i = 1$ if the i^{th} person to arrive is hired, and let $X_i = 0$ otherwise. Person 1 will always be hired, i.e. $\mathbf{P}(X_1 = 1) = 1$, so $\mathbf{E}X_1 = 1$. Since any arrival order is equally likely, $\mathbf{P}(X_2 = 1) = 1/2$. So, $\mathbf{E}X_2 = 1/2$. In general, if i is a positive integer, then $\mathbf{P}(X_i = 1) = 1/i$. This follows since any ordering of the people is equally likely, so there is a probability of $1/i$ of the i^{th} person having the largest number a_i among the numbers $a_1, \dots, a_i$. So, $\mathbf{E}X_i = 1/i$. (More formally, fix $i \in \{1, \dots, n\}$, and let $j \in \{1, \dots, i\}$. Let A_j be the event that $a_j > a_k$ for every $k \in \{1, \dots, i\}$ such that $k \neq j$. Then $\cup_{j=1}^i A_j = \Omega$, and $A_j \cap A_{j'} = \emptyset$ for every $j, j' \in \{1, \dots, i\}$ with $j \neq j'$. So, $1 = \mathbf{P}(\Omega) = \sum_{j=1}^i \mathbf{P}(A_j)$. We now claim that $\mathbf{P}(A_j) = \mathbf{P}(A_{j'})$ for every $j, j' \in \{1, \dots, i\}$ with $j \neq j'$. Given that this is true, it immediately follows that $\mathbf{P}(A_i) = 1/i$, as desired. To prove our claim, suppose we write any arrival order of the people as $c_1, \dots, c_n$ where $c_1, \dots, c_n$ are distinct elements of $\{1, \dots, n\}$. Then for any $k < i$, any arrival order $c_1, \dots, c_n$ where a_{c_i} exceeds $a_{c_1}, \dots, a_{c_{i-1}}$ can be uniquely associated to the arrival order $c_1, \dots, c_{k-1}, c_i, c_{k+1}, \dots, c_{i-1}, c_k, c_{i+1}, \dots, c_n$. That is, the number of orderings where the i^{th} number exceeds the previous ones is equal to the number of orderings where the k^{th} number exceeds the first i numbers. That is, $\mathbf{P}(A_i) = \mathbf{P}(A_k)$. $\square$

Exercise 5.7. Let X be a Poisson random variable with parameter $\lambda > 0$. Compute $\mathbf{E}[(1 + X)^{-1}]$.

Solution. Since X takes values in the negative integers,

$$\begin{aligned} \mathbf{E}[(1 + X)^{-1}] &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{(n+1)n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{(n+1)!} = \lambda^{-1} e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^n}{n!} \\ &= \lambda^{-1} e^{-\lambda} \left(\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} - 1 \right) = \lambda^{-1} e^{-\lambda} (e^{\lambda} - 1) = \lambda^{-1} (1 - e^{-\lambda}) \end{aligned}$$

$\square$

6. HOMEWORK 6

Exercise 6.1. There are 10 different bins, and 20 balls are placed in the bins uniformly at random. What is the expected number of empty bins?

Solution. Let X be the random variable which counts the number of empty bins. Write

$$X = X_1 + \dots + X_{10}$$

where $X_i = 1$ if bin i is empty, and $X_i = 0$ otherwise. Then

$$\mathbf{E}[X] = \mathbf{E}[X_1] + \dots + \mathbf{E}[X_{10}] = \mathbf{P}(X_1 = 1) + \dots + \mathbf{P}(X_{10} = 1)$$

If the i th bin is empty, then all 20 balls must have been placed in the other 9 bins. There are 9^{20} ways to do this, and 10^{20} ways to place the balls in the bins overall, hence

$$\mathbf{P}(X_i = 1) = \left(\frac{9}{10}\right)^{20}$$

for each i . Therefore

$$\mathbf{E}[X] = 10 \cdot \left(\frac{9}{10}\right)^{20}$$

□

Exercise 6.2. There are 100 different baseball cards in a set, and the cards are sold in packs of 10. Each card is equally likely to be included in a pack. What is the expected number of packs you must buy in order to collect the complete set?

Solution. First assume that you buy the cards one at a time, and let T_i be the time when you receive the i th new card, setting $T_0 = 0$ for convenience. We want to find T_{100} .

To do so, consider the random variable $T_i - T_{i-1}$. This random variable is geometrically distributed with parameter $p = \frac{100-(i-1)}{100} = \frac{101-i}{100}$, hence

$$\mathbf{E}[T_i - T_{i-1}] = \frac{100}{101-i}$$

by a result from the previous homework assignment. Moreover, note that

$$T_{100} = (T_{100} - T_{99}) + (T_{99} - T_{98} + \cdots + (T_1 - t_0))$$

hence

$$\mathbf{E}[T_{100}] = \sum_{i=1}^{100} \mathbf{E}[T_i - T_{i-1}] = \sum_{i=1}^{100} \frac{100}{101-i} = 100 \sum_{j=1}^{100} \frac{1}{j} \approx 518.7$$

by setting $j = 101 - i$. Finally, to account for the fact that the cards come in packs of 10, round up to the nearest multiple of 10 to obtain

$$\left\lceil \frac{\mathbf{E}[T_{100}]}{10} \right\rceil = 52$$

□

Exercise 6.3. If you draw cards from a standard 52-card deck without replacement, how many cards can you expect to draw before finding (a) a King or (b) a Heart?

Solution.

(a)

Suppose we label the non-king cards as $\{1, \dots, 48\}$. Let $i \in \{1, \dots, 48\}$. Let $X_i = 1$ if the i^{th} card is drawn before any king is drawn, and $X_i = 0$ otherwise. The number of cards drawn before the first king is

$$\sum_{i=1}^{48} X_i.$$

It remains to compute the expected value of this quantity. We claim that $\mathbf{E}X_i = 1/5$ for all $i \in \{1, \dots, 48\}$. Assuming this claim, the expected number of cards to be drawn before the second heart is

$$\mathbf{E}\left(\sum_{i=1}^{48} X_i\right) = \sum_{i=1}^{48} \mathbf{E}X_i = 48/5.$$

We now prove the claim. Suppose we label the heart at the highest point in the deck as $j = 1$, we label the next highest position heart as $j = 2$ and so on, up to $j = 4$. Then there are five possible locations for a non-king card: above the $j = 1$ king, in between the $j = 1$ and $j = 2$ kings, in between the $j = 2$ and $j = 3$ kings, etc. For any fixed $i \in \{1, \dots, 48\}$, the i^{th} card is equally likely to be in any of these 5 locations. To see this, for any of the five $k \in \{1, \dots, 5\}$ non-heart card locations, let A_k be the event that the i^{th} card is in location k . Then $\cup_{k=1}^5 A_k = \Omega$ and if $k, k' \in \{1, \dots, 5\}$ with $k \neq k'$, then $A_k \cap A_{k'} = \emptyset$. Given any arrangement of cards such that the i^{th} card is in location k , we can uniquely associate this arrangement to another arrangement where the i^{th} card occurs in location k' . We can do this, for example, by swapping all cards in location k with all cards in location k' . Since the probability law $\mathbf{P}(A_k)$ counts the number of arrangements in A_k divided by $52!$, we conclude that $\mathbf{P}(A_k) = \mathbf{P}(A_{k'})$ for all $k \neq k'$, $k, k' \in \{1, \dots, 5\}$. So, $1 = \mathbf{P}(\Omega) = \sum_{k=1}^5 \mathbf{P}(A_k) = 5\mathbf{P}(A_1)$. So, $\mathbf{P}(A_1) = \mathbf{P}(A_2) = 1/5$. That is, $\mathbf{P}(X_i = 1) = 1/5$. And since X_i only take values 1 or 0, the definition of expected value says $\mathbf{E}X_i = 1/5$ for all $i \in \{1, \dots, 48\}$, as desired.

(b) Suppose we label the non-heart cards as $\{1, \dots, 39\}$. Let $i \in \{1, \dots, 39\}$. Let $X_i = 1$ if the i^{th} card is drawn before any heart is drawn, and $X_i = 0$ otherwise. The number of cards drawn before the first heart is

$$\sum_{i=1}^{39} X_i.$$

It remains to compute the expected value of this quantity. We claim that $\mathbf{E}X_i = 1/14$ for all $i \in \{1, \dots, 39\}$. Assuming this claim, the expected number of cards to be drawn before the second heart is

$$\mathbf{E}\left(\sum_{i=1}^{39} X_i\right) = \sum_{i=1}^{39} \mathbf{E}X_i = 39/14.$$

We now prove the claim. Suppose we label the heart at the highest point in the deck as $j = 1$, we label the next highest position heart as $j = 2$ and so on, up to $j = 13$. Then there are fourteen possible locations for a non-heart card: above the $j = 1$ heart, in between the $j = 1$ and $j = 2$ hearts, in between the $j = 2$ and $j = 3$ hearts, etc. For any fixed $i \in \{1, \dots, 39\}$, the i^{th} card is equally likely to be in any of these 14 locations. To see this, for any of the fourteen $k \in \{1, \dots, 14\}$ non-heart card locations, let A_k be the event that the i^{th} card is in location k . Then $\cup_{k=1}^{14} A_k = \Omega$ and if $k, k' \in \{1, \dots, 14\}$ with $k \neq k'$, then $A_k \cap A_{k'} = \emptyset$. Given any arrangement of cards such that the i^{th} card is in location k , we can uniquely associate this arrangement to another arrangement where the i^{th} card occurs in location k' . We can do this, for example, by swapping all cards in location k with all cards in location k' . Since the probability law $\mathbf{P}(A_k)$ counts the number of arrangements in A_k divided by $52!$, we conclude that $\mathbf{P}(A_k) = \mathbf{P}(A_{k'})$ for all $k \neq k'$, $k, k' \in \{1, \dots, 14\}$. So, $1 = \mathbf{P}(\Omega) = \sum_{k=1}^{14} \mathbf{P}(A_k) = 14\mathbf{P}(A_1)$. So, $\mathbf{P}(A_1) = \mathbf{P}(A_2) = 1/14$. That is, $\mathbf{P}(X_i = 1) = 1/14$. And since X_i only take values 1 or 0, the definition of expected value says $\mathbf{E}X_i = 1/14$ for all $i \in \{1, \dots, 39\}$, as desired. $\square$

Exercise 6.4. Let f be a twice differentiable convex function, and X a discrete random variable such that $\mathbf{E}[X]$ and $\mathbf{E}[f(X)]$ exist. Prove Jensen's inequality: $\mathbf{E}[f(X)] \geq f(\mathbf{E}[X])$.

Solution. Since f is convex, it lies above all its tangent lines. Therefore

$$f(x) \geq f(t) + f'(t)(x - t)$$

for all real numbers x and t , hence

$$f(X) \geq f(t) + f'(t)(X - t)$$

for all $t \in \mathbb{R}$, and taking expectations yields

$$\mathbf{E}[f(X)] \geq f(t) + f'(t)(\mathbf{E}[X] - t)$$

If we set $t = \mathbf{E}[X]$ then the second term on the right is zero, hence

$$\mathbf{E}[f(X)] \geq f(\mathbf{E}[X])$$

as desired. $\square$

Exercise 6.5. Let n be a positive integer, $\Omega = \{0, 1\}^n$, and $p \in (0, 1)$. Define a probability law $\mathbf{P}$ on Ω by

$$\mathbf{P}(\{\omega\}) = p^{\sum_{i=1}^n \omega_i} (1 - p)^{n - \sum_{i=1}^n \omega_i}$$

where $\omega = (\omega_1, \dots, \omega_n)$. For each $1 \leq i \leq n$, define a random variable $X_i : \Omega \rightarrow \mathbb{R}$ by $X(\omega) = \omega_i$. Finally, define $X = X_1 + \dots + X_n$.

Show that $\mathbf{P}(\Omega) = 1$. Find $\mathbf{E}[X_i]$ and $\mathbf{E}[X]$. Show that X is a binomial random variable with parameters n and p .

Solution.

Below, we will repeatedly use the binomial theorem in the form

$$\sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} = 1. \quad (*)$$

By definition of $\mathbf{P}$,

$$\begin{aligned} \mathbf{P}(\Omega) &= \sum_{\omega \in \Omega} \mathbf{P}(\{\omega\}) = \sum_{\omega \in \Omega} p^{\sum_{i=1}^n \omega_i} (1 - p)^{n - \sum_{i=1}^n \omega_i} \\ &= \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} \stackrel{(*)}{=} 1 \end{aligned}$$

since there are $\binom{n}{k}$ elements ω such that exactly k of the ω_i are equal to 1.

For any $1 \leq i \leq n$, $\mathbf{E}[X_i] = \mathbf{P}(X_i = 1)$ since X_i only takes the values 0 and 1, and by definition of X_i ,

$$\begin{aligned} \mathbf{P}(X_i = 1) &= \mathbf{P}(\omega \in \Omega : X_i(\omega) = 1) = \mathbf{P}(\omega \in \Omega : \omega_i = 1) = \sum_{\omega \in \Omega : \omega_i = 1} \mathbf{P}(\{\omega\}) \\ &= \sum_{\omega \in \Omega : \omega_i = 1} p^{\sum_{j=1}^n \omega_j} (1 - p)^{n - \sum_{j=1}^n \omega_j} = p \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1 - p)^{n-1-k} \stackrel{(*)}{=} p \end{aligned}$$

Therefore, $\mathbf{E}[X] = \sum_{i=1}^n \mathbf{E}[X_i] = n\mathbf{P}(X_1 = 1) = np$.

Lastly, fix $0 \leq k \leq n$, and note that

$$\begin{aligned}
\mathbf{P}(X = k) &= \mathbf{P}(\omega \in \Omega: \sum_{i=1}^n X_i(\omega) = k) = \mathbf{P}(\omega \in \Omega: \omega_1 + \cdots + \omega_n = k) \\
&= \sum_{\omega \in \Omega: \omega_1 + \cdots + \omega_n = k} \mathbf{P}(\{\omega\}) = \sum_{\omega \in \Omega: \omega_1 + \cdots + \omega_n = k} p^{\sum_{i=1}^n \omega_i} (1-p)^{n - \sum_{i=1}^n \omega_i} \\
&= \sum_{\omega \in \Omega: \omega_1 + \cdots + \omega_n = k} p^k (1-p)^{n-k} = p^k (1-p)^{n-k} \sum_{\omega \in \Omega: \omega_1 + \cdots + \omega_n = k} 1 \\
&= p^k (1-p)^{n-k} \binom{n}{k},
\end{aligned}$$

since the number of $\omega = (\omega_1, \dots, \omega_n) \in \Omega$ such that $\omega_1 + \cdots + \omega_n = k$ is $\binom{n}{k}$.

Therefore, X is a binomial random variable. □

Exercise 6.6. Let $A_1, \dots, A_n$ be events in a probability space Ω . Define random variables X and X_i , $1 \leq i \leq n$ so that $X = 1$ on $\cup_i A_i$ and $X = 0$ otherwise, while $X_i = 1$ on A_i and $X = 0$ otherwise.

Show that $X = 1 - \prod_{i=1}^n (1 - X_i)$.

Establish the inclusion-exclusion formula by taking expected values of this identity.

Solution. See the second solution of Exercise 3.4. □

Exercise 6.7. You are trapped in a maze, starting in a room with 3 doors. Door 1 leads to a corridor which lets you exit the maze after 3 hours of walking. Door 2 leads to a corridor which returns to the starting point after 7 hours of walking. Door 3 leads to a corridor which returns to the starting point after 9 hours of walking. You aren't good at learning from your mistakes, so every time you return to the starting point, you choose a door uniformly at random.

Let X be the number of hours it takes you to exit the maze, and Y be the number of the door you initially choose.

Find $\mathbf{E}[X|Y = y]$ for $y = 1, 2, 3$ in terms of $\mathbf{E}[X]$.

Determine $\mathbf{E}[X]$.

Solution. It is given that $\mathbf{E}[X|Y = 1] = 3$, $\mathbf{E}[X|Y = 2] = 7 + \mathbf{E}[X]$, and $\mathbf{E}[X|Y = 3] = 9 + \mathbf{E}[X]$. By conditioning on the result of the first door choice and using the Total Expectation Theorem, we have

$$\begin{aligned}
\mathbf{E}[X] &= \mathbf{E}[X|Y = 1]\mathbf{P}(Y = 1) + \mathbf{E}[X|Y = 2]\mathbf{P}(Y = 2) + \mathbf{E}[X|Y = 3]\mathbf{P}(Y = 3) \\
&= \frac{1}{3}(\mathbf{E}[X|Y = 1] + \mathbf{E}[X|Y = 2] + \mathbf{E}[X|Y = 3]).
\end{aligned}$$

Using part (a) then shows that

$$\mathbf{E}[X] = \frac{1}{3}(3 + 7 + \mathbf{E}[X] + 9 + \mathbf{E}[X])$$

Solving for $\mathbf{E}[X]$ yields

$$\frac{1}{3}\mathbf{E}[X] = \frac{19}{3}$$

so $\mathbf{E}[X] = 19$. □

Exercise 6.8. Let X, Y, Z be independent geometric random variables with the same parameter $p \in (0, 1)$. Compute $\mathbf{P}(X = k | X + Y + Z = n)$ for k, n positive integers.

Solution. Since X, Y, Z take values in the positive integers, the event $\{X + Y + Z = n\}$ has probability zero if $n < 3$, in which case the conditional probability is undefined. So assume $n \geq 3$.

By the definition of conditional probability,

$$\mathbf{P}(X = k | X + Y + Z = n) = \frac{\mathbf{P}(X = k, X + Y + Z = n)}{\mathbf{P}(X + Y + Z = n)}$$

If $k \geq n - 1$ then $\mathbf{P}(X = k, X + Y + Z = n) = 0$, and if $1 \leq k \leq n - 2$ then

$$\begin{aligned} \mathbf{P}(X = k, X + Y + Z = n) &= \mathbf{P}(X = k) \mathbf{P}(Y + Z = n - k) \\ &= p(1 - p)^{k-1} \sum_{y=1}^{n-k-1} \mathbf{P}(Y = y) \mathbf{P}(Z = n - k - y) \\ &= p(1 - p)^{k-1} \sum_{y=1}^{n-k-1} (1 - p)^{y-1} p(1 - p)^{n-k-y-1} p = p^3(1 - p)^{n-3} \sum_{y=1}^{n-k-1} 1 \\ &= (n - k - 1)p^3(1 - p)^{n-3} \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbf{P}(X + Y + Z = n) &= \sum_{\substack{x, y, z \geq 1 \\ x + y + z = n}} p(1 - p)^{x-1} p(1 - p)^{y-1} p(1 - p)^{z-1} \\ &= p^3(1 - p)^{n-3} \sum_{\substack{x, y, z \geq 1 \\ x + y + z = n}} 1 = p^3(1 - p)^{n-3} \sum_{x=1}^{n-2} \sum_{y=1}^{n-x-1} 1 \\ &= p^3(1 - p)^{n-3} \sum_{x=1}^{n-2} (n - x - 1) = p^3(1 - p)^{n-3} \sum_{j=1}^{n-2} j = \frac{p^3(1 - p)^3(n - 1)(n - 2)}{2} \end{aligned}$$

Therefore

$$\frac{\mathbf{P}(X = k) \mathbf{P}(X + Y = n - k)}{\mathbf{P}(X + Y + Z = n)} = \frac{2(n - k - 1)}{(n - 1)(n - 2)}$$

for $n \geq 3$ and $1 \leq k \leq n - 2$.

There is another, less computational, way to determine $\mathbf{P}(Y + Z = n - k)$ and $\mathbf{P}(X + Y + Z = n)$: A geometric random variable counts the number of trials until the first success, so the sum of two independent geometric random variables counts the number of trials until the second success.

Thus if $Y + Z = n - k$ then the last trial was a success, and of the first $n - k - 1$ trials exactly one was a success. Therefore

$$\mathbf{P}(Y + Z = n - k) = \binom{n - k - 1}{1} p^2(1 - p)^{n-k-2} = (n - k - 1)p^2(1 - p)^{n-k-2}$$

Similarly, $X + Y + Z$ counts the number of trials until the third success, hence

$$\mathbf{P}(X + Y + Z = n) = \binom{n-1}{2} p^3 (1-p)^{n-3} = \frac{(n-1)(n-2)}{2} p^3 (1-p)^{n-3}$$

Combining these results with $\mathbf{P}(X = k) = p(1-p)^{k-1}$ yields the same answer as the first method. $\square$

7. HOMEWORK 7

Exercise 7.1.

- Give an example (with proof) of two random variables that are independent.
- Give an example (with proof) of two random variables that are not independent.
- Find two random variables X, Y such that $\mathbf{E}[XY] \neq \mathbf{E}[X]\mathbf{E}[Y]$.

Solution. (a) Let $\Omega = \{HH, HT, TH, TT\}$ with the uniform probability law, and let $X = 1$ if the first character is H and $X = 0$ otherwise, and $Y = 1$ if the second character is H and $Y = 0$ otherwise.

Then

$$\mathbf{P}(X = 1, Y = 1) = \mathbf{P}(HH) = \frac{1}{4} = \mathbf{P}(X = 1)\mathbf{P}(Y = 1)$$

$$\mathbf{P}(X = 1, Y = 0) = \mathbf{P}(HT) = \frac{1}{4} = \mathbf{P}(X = 1)\mathbf{P}(Y = 0)$$

$$\mathbf{P}(X = 0, Y = 1) = \mathbf{P}(TH) = \frac{1}{4} = \mathbf{P}(X = 0)\mathbf{P}(Y = 1)$$

and

$$\mathbf{P}(X = 0, Y = 0) = \mathbf{P}(TT) = \frac{1}{4} = \mathbf{P}(X = 0)\mathbf{P}(Y = 0)$$

so X and Y are independent.

- Let X be a random variable with $\mathbf{P}(X = 0) = \mathbf{P}(X = 1) = \frac{1}{2}$, and let $Y = -X$. Then

$$\mathbf{P}(X = 1, Y = 0) = 0 \neq \frac{1}{2} \cdot \frac{1}{2} = \mathbf{P}(X = 1)\mathbf{P}(Y = 0)$$

Therefore X and Y are not independent.

- Let X be a random variable with $\mathbf{P}(X = -1) = \mathbf{P}(X = 1) = \frac{1}{2}$, and let $Y = X$. Then $\mathbf{E}[X] = \mathbf{E}[Y] = 0$, but $XY = X^2 = 1$, so $\mathbf{E}[XY] = 1$. $\square$

Exercise 7.2. Does there exist a random variable which is independent of itself? Either find such an X , or prove that no such X can exist.

Solution. Let X be a random variable with $\mathbf{P}(X = c) = 1$ for some constant c . Then

$$\mathbf{P}(X = c) = 1 = \mathbf{P}(X = c)^2$$

and if at least one of x_1, x_2 is not equal to c , then

$$\mathbf{P}(X = x_1, X = x_2) = 0 = \mathbf{P}(X = x_1)\mathbf{P}(X = x_2)$$

Therefore X is independent of itself.

It also turns out that any random variable which is independent of itself has the above form for some constant c . $\square$

Exercise 7.3. Let n be a positive integer and $p \in (0, 1)$. Suppose that $X_1, \dots, X_n$ are pairwise independent Bernoulli random variables with parameter p , and let $S_n = \frac{X_1 + \dots + X_n}{n}$.

Compute $\mathbf{E}[S_n]$ and $\text{var}(S_n)$. What does the variance computation tell you as $n \rightarrow \infty$?

Solution.

The linearity of expectation implies that

$$\mathbf{E}[S_n] = \frac{1}{n} \left(\mathbf{E}[X_1] + \dots + \mathbf{E}[X_n] \right) = \frac{np}{n} = p$$

and since the X_i are pairwise independent it follows that

$$\text{var}(S_n) = \frac{1}{n^2} (\text{var}(X_1) + \dots + \text{var}(X_n)) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

Observe that $\text{var}(S_n) \rightarrow 0$ as $n \rightarrow \infty$. □

Exercise 7.4. Let X and Y be independent random variables taking on finitely many values. Show that

$$\mathbf{E}[f(X)g(Y)] = \mathbf{E}[f(X)]\mathbf{E}[g(Y)]$$

for any functions $f, g : \mathbb{R} \rightarrow \mathbb{R}$.

Solution. Since X and Y are independent,

$$\begin{aligned} \mathbf{E}[f(X)g(Y)] &= \sum_{x,y \in \mathbb{R}} f(x)g(y) \mathbf{P}(X=x, Y=y) = \sum_{x,y \in \mathbb{R}} f(x)g(y) \mathbf{P}(X=x) \mathbf{P}(Y=y) \\ &= \left[\sum_{x \in \mathbb{R}} f(x) \mathbf{P}(X=x) \right] \left[\sum_{y \in \mathbb{R}} g(y) \mathbf{P}(Y=y) \right] = \mathbf{E}[f(X)] \mathbf{E}[g(Y)] \end{aligned}$$

□

Exercise 7.5. Find three random variables X_1, X_2, X_3 which are pairwise independent but not independent.

Solution. Let A_1, A_2, A_3 be the events from problem 6 on Homework 3, and for $i = 1, 2, 3$ define a random variable X_i which is equal to 1 on A_i and 0 otherwise. Then the X_i are pairwise independent but not independent. □

Exercise 7.6. Let $X_1, \dots, X_n$ be independent Bernoulli random variables with parameter $0 < p < 1$.

Solution. a. Show that $\mathbf{E}[e^{tX_i}] = (1-p) + pe^t$ for $1 \leq i \leq n$.

X_i takes the value 1 with probability p and 0 with probability $1-p$, hence

$$\mathbf{E}[e^{tX_i}] = e^0(1-p) + e^t p = (1-p) + pe^t$$

b. Let $S_n = X_1 + \dots + X_n$. Show that $\mathbf{E}[e^{tS_n}] = [(1-p) + pe^t]^n$.

Using part (a) and the generalization of problem 6 to n independent random variables shows that

$$\mathbf{E}[e^{tS_n}] = \mathbf{E}[e^{tX_1} \dots e^{tX_n}] = \mathbf{E}[e^{tX_1}] \dots \mathbf{E}[e^{tX_n}] = [(1-p) + pe^t]^n$$

c. Use part (b) to compute $\mathbf{E}[S_n]$ and $\mathbf{E}[S_n^2]$.

Since $\mathbf{E}[e^{tS_n}] = [(1-p) + pe^t]^n$, it follows that

$$\mathbf{E}[S_n] = \frac{d}{dt}[(1-p) + pe^t]^n \Big|_{t=0} = np$$

Similarly,

$$\mathbf{E}[S_n^2] = \frac{d^2}{dt^2}[(1-p) + pe^t]^n \Big|_{t=0} = np + n(n-1)p^2$$

□

Exercise 7.7. Let $X_1, \dots, X_n$ be independent discrete random variables. Show that

$$\mathbf{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n \mathbf{P}(X_i \leq x_i)$$

for all $x_1, \dots, x_n \in \mathbb{R}$.

Solution. Since $X_1, \dots, X_n$ are independent, we have

$$\begin{aligned} \mathbf{P}(X_1 \leq x_1, \dots, X_n \leq x_n) &= \sum_{y_1 \leq x_1, \dots, y_n \leq x_n} \mathbf{P}(X_1 = y_1, \dots, X_n = y_n) \\ &= \sum_{y_1 \leq x_1, \dots, y_n \leq x_n} \mathbf{P}(X_1 = y_1) \cdots \mathbf{P}(X_n = y_n) = \left[\sum_{y_1 \leq x_1} \mathbf{P}(X_1 = y_1) \right] \cdots \left[\sum_{y_n \leq x_n} \mathbf{P}(X_n = y_n) \right] \\ &= \prod_{i=1}^n \mathbf{P}(X_i \leq x_i) \end{aligned}$$

□

Exercise 7.8. Verify that $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1$. (Hint: let $T = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$. It suffices to show that $T^2 = 1$, since $T > 0$.)

Solution. Using polar coordinates,

$$\begin{aligned} T^2 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} dx dy \\ &= \frac{1}{2\pi} \int_{r=0}^{\infty} \int_{\theta=0}^{2\pi} r e^{-r^2/2} d\theta dr = \int_{r=0}^{\infty} r e^{-r^2/2} dr = \lim_{N \rightarrow \infty} [-e^{-r^2/2}]_{r=0}^{r=N} = \lim_{N \rightarrow \infty} [1 - e^{-N^2/2}] = 1. \end{aligned}$$

□

8. HOMEWORK 8

Exercise 8.1. Let X be a continuous random variable with distribution function $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \forall x \in \mathbb{R}$. Show that $\text{var}(X) = 1$.

Solution. $\mathbf{E}[X] = 0$ since $f_X(x)$ is an even function, so it is enough to show that $\mathbf{E}[X^2] = 1$. By definition,

$$\mathbf{E}[X^2] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx$$

Integrating by parts with $u = -x$ and $dv = -xe^{-\frac{x^2}{2}}$ and noting that the boundary terms vanish, we obtain

$$\begin{aligned}\mathbf{E}[X^2] &= \lim_{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-n}^n (-x) \frac{d}{dx} e^{-\frac{x^2}{2}} dx \\ &= \lim_{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \left([-xe^{-x^2/2}]_{x=-n}^{x=n} + \int_{-n}^n e^{-\frac{x^2}{2}} dx \right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = 1\end{aligned}$$

using Exercise 7.8. □

Exercise 8.2. Let X be a random variable such that $f_X(x) = x$ when $0 \leq x \leq \sqrt{2}$ and $f_X(x) = 0$ otherwise. Compute $\mathbf{E}X^2$ and $\mathbf{E}X^3$.

Solution. By definition,

$$\mathbf{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^{\sqrt{2}} x^3 dx = \frac{x^4}{4} \Big|_0^{\sqrt{2}} = 1$$

Similarly,

$$\mathbf{E}[X^3] = \int_0^{\sqrt{2}} x^4 dx = \frac{x^5}{5} \Big|_0^{\sqrt{2}} = \frac{2^{\frac{5}{2}}}{5}$$

□

Exercise 8.3 (Numerical Integration). In computer graphics in video games, etc., various integrations are performed in order to simulate lighting effects. Here is a way to use random sampling to integrate a function in order to quickly and accurately render lighting effects. Let $\Omega = [0, 1]$, and let $\mathbf{P}$ be the uniform probability law on Ω , so that if $0 \leq a < b \leq 1$, we have $\mathbf{P}([a, b]) = b - a$. Let $X_1, \dots, X_n$ be independent random variables such that $\mathbf{P}(X_i \in [a, b]) = b - a$ for all $0 \leq a < b \leq 1$, for all $i \in \{1, \dots, n\}$. Let $f: [0, 1] \rightarrow \mathbb{R}$ be a continuous function we would like to integrate. Instead of integrating f directly, we instead compute the quantity

$$\frac{1}{n} \sum_{i=1}^n f(X_i).$$

Show that

$$\begin{aligned}\lim_{n \rightarrow \infty} \mathbf{E} \left(\frac{1}{n} \sum_{i=1}^n f(X_i) \right) &= \int_0^1 f(t) dt. \\ \lim_{n \rightarrow \infty} \text{var} \left(\frac{1}{n} \sum_{i=1}^n f(X_i) \right) &= 0.\end{aligned}$$

That is, as n becomes large, $\frac{1}{n} \sum_{i=1}^n f(X_i)$ is a good estimate for $\int_0^1 f(t) dt$.

Solution. By definition of X_i we have $\mathbf{E}f(X_i) = \int_0^1 f(t) dt$ for all $i \geq 1$ so that $\mathbf{E} \left(\frac{1}{n} \sum_{i=1}^n f(X_i) \right) = \frac{1}{n} \int_0^1 f(t) dt = \int_0^1 f(t) dt$. Also, by independence we have

$$\text{var} \left(\frac{1}{n} \sum_{i=1}^n f(X_i) \right) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(f(X_i)) = \frac{1}{n} \text{var}(f(X_1)).$$

This quantity goes to zero as $n \rightarrow \infty$. (Since f is continuous on $[0, 1]$, f is bounded by some constant c on $[0, 1]$, i.e. $|f(t)| \leq c$ for all $t \in [0, 1]$, so $|f(X_1)| \leq c$, so $\text{var} f(X_i) \leq \mathbf{E}[f(X_i)]^2 \leq c^2$ for all $i \geq 1$.) $\square$

Exercise 8.4. Let X be a random variable such that $X = 1$ with probability 1. Show that X is not a continuous random variable. That is, there does not exist a probability density function f such that $\mathbf{P}(X \leq a) = \int_{-\infty}^a f(x)dx$ for all $x \in \mathbb{R}$. (Hint: if X were continuous, then the function $g(t) = \int_{-\infty}^t f(x)dx$ would be continuous, by the Fundamental Theorem of Calculus.)

Solution. As suggested, if X is continuous, then the CDF of X is a continuous function. But the CDF of X is discontinuous at 1 by assumption, a contradiction. $\square$

Exercise 8.5. Verify that a Gaussian random variable X with mean μ and variance σ^2 actually has mean μ and variance σ^2 .

Let $a, b \in \mathbb{R}$ with $a \neq 0$. Show that $aX + b$ is a normal random variable with mean $a\mu + b$ and variance $a^2\sigma^2$.

In particular, conclude that $(X - \mu)/\sigma$ is a standard normal.

Solution. Since $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $\forall x \in \mathbb{R}$, we have

$$\begin{aligned} \mathbf{E}X &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} (x + \mu) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} dx = \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \mu, \end{aligned}$$

where we used Exercise 7.8 in the last step, and a few changes of variables. Similarly,

$$\begin{aligned} \mathbf{E}X^2 &= \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} (x + \mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} (x^2 + \mu^2) e^{-\frac{x^2}{2\sigma^2}} dx = \mu^2 + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} x^2 e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \mu^2 + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} x \left[- (d/dx) \sigma^2 e^{-\frac{x^2}{2\sigma^2}} \right] dx \\ &= \mu^2 + \sigma^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} dx = \mu^2 + \sigma^2. \end{aligned}$$

So, $\text{var}(X) = \mathbf{E}X^2 - (\mathbf{E}X)^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2$.

Suppose $\mu = \sigma = 1$. Then

$$\begin{aligned} \mathbf{P}(aX + b \leq t) &= \mathbf{P}(X \leq (t - b)/a) = \int_{-\infty}^{(t-b)/a} e^{-x^2/2} dx / \sqrt{2\pi} = \int_{-\infty}^{t-b} a e^{-x^2/[2a^2]} dx / \sqrt{2\pi} \\ &= \int_{-\infty}^t a e^{-(x+b)^2/[2a^2]} dx / \sqrt{2\pi}. \end{aligned}$$

That is, $aX + b$ has a Gaussian density with mean b and variance a^2 . The case of general μ, σ follows from this case. $\square$

Exercise 8.6. Using the De Moivre-Laplace Theorem, estimate the probability that 1,000,000 coin flips of fair coins will result in more than 501,000 heads. (Some of the following integrals may be relevant: $\int_{-\infty}^0 e^{-t^2/2} dt / \sqrt{2\pi} = 1/2$, $\int_{-\infty}^1 e^{-t^2/2} dt / \sqrt{2\pi} \approx .8413$, $\int_{-\infty}^2 e^{-t^2/2} dt / \sqrt{2\pi} \approx .9772$, $\int_{-\infty}^3 e^{-t^2/2} dt / \sqrt{2\pi} \approx .9987$.)

Casinos do these kinds of calculations to make sure they make money and that they do not go bankrupt. Financial institutions and insurance companies do similar calculations for similar reasons.

Solution. Let X be the number of heads in 1,000,000 flips. The De Moivre-Laplace limit theorem states that if S_n is the sum of n independent Bernoulli random variables with parameter $\frac{1}{2}$, then

$$\frac{S_n - \frac{n}{2}}{\sqrt{\frac{n}{4}}}$$

can be closely approximated by a normal random variable. Taking $n = 10^6$, we have

$$\begin{aligned} \mathbf{P}(X \geq 501,000) &= \mathbf{P}(X - 500,000 \geq 1000) = \mathbf{P}\left(\frac{X - 500000}{500} \geq 2\right) \\ &= 1 - \mathbf{P}\left(\frac{X - 500000}{500} < 2\right) \approx 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^2 e^{-\frac{x^2}{2}} dx \approx 0.0228 \end{aligned}$$

□

Exercise 8.7. Let X be a uniformly distributed random variable on $[-1, 1]$. Let $Y := X^2$. Find f_Y .

Solution. First, note that for any $0 < a < 1$, the definition of X implies that

$$\mathbf{P}(-a \leq X \leq a) = \int_{-a}^a f_X(s) ds = \frac{1}{2} \int_{-a}^a ds = \frac{2a}{2} = a. \quad (*)$$

By differentiating the CDF and using the definition of Y and X , we have for any $0 \leq t \leq 1$,

$$\begin{aligned} f_Y(t) &= \frac{d}{dt} \mathbf{P}(Y \leq t) = \frac{d}{dt} \mathbf{P}(X^2 \leq t) = \frac{d}{dt} \mathbf{P}(|X| \leq \sqrt{t}) = \frac{d}{dt} \mathbf{P}(-\sqrt{t} \leq X \leq \sqrt{t}) \\ &\stackrel{(*)}{=} \frac{d}{dt} 2\sqrt{t} = t^{-1/2}. \end{aligned}$$

So, $f_Y(t) = t^{-1/2}$ for any $0 \leq t \leq 1$, and $f_Y(t) = 0$ for any other t . □

Exercise 8.8. Let X be a uniformly distributed random variable on $[0, 1]$. Let $Y := 4X(1 - X)$. Find f_Y .

Solution. Using the quadratic formula, note that $4x(1 - x) = t$ occurs when $-4x^2 + 4x - t = 0$, i.e. when $x^2 - x + t/4 = 0$, i.e. when $x = [1 \pm \sqrt{1 - t}]/2$. So, $4x(1 - x) \leq t$ and $0 < x < 1$ when $0 \leq x \leq [1 - \sqrt{1 - t}]/2$ and $[1 + \sqrt{1 - t}]/2 \leq x \leq 1$. So, using set disjointness and the definition of X ,

$$\begin{aligned} \mathbf{P}(4X(1 - X) \leq t) &= \mathbf{P}(0 \leq X \leq [1 - \sqrt{1 - t}]/2 \text{ or } [1 + \sqrt{1 - t}]/2 \leq X \leq 1) \\ &= \mathbf{P}(0 \leq X \leq [1 - \sqrt{1 - t}]/2) + \mathbf{P}([1 + \sqrt{1 - t}]/2 \leq X \leq 1) \\ &= [1 - \sqrt{1 - t}]/2 + 1 - [1 + \sqrt{1 - t}]/2 = 1 - \sqrt{1 - t}. \quad (*) \end{aligned}$$

Then, by differentiating the CDF, we have for any $t > 0$,

$$\begin{aligned} f_Y(t) &= \frac{d}{dt} \mathbf{P}(Y \leq t) = \frac{d}{dt} \mathbf{P}(4X(1-X) \leq t) = \frac{d}{dt} \mathbf{P}(4X(1-X) \leq t) \\ &\stackrel{(*)}{=} \frac{d}{dt} (1 - \sqrt{1-t}) = \frac{1}{2} (1-t)^{-1/2}. \end{aligned}$$

So, $f_Y(t) = \frac{1}{2}(1-t)^{-1/2}$ for any $0 < t < 1$, and $f_Y(t) = 0$ for any other t . $\square$

Exercise 8.9. Let X be a uniformly distributed random variable on $[0, 1]$. Find the PDF of $-\log(X)$.

Solution. First, note that for any $0 < a < 1$, the definition of X implies that

$$\mathbf{P}(X \geq a) = \int_a^1 f_X(s) ds = \int_a^1 ds = 1 - a. \quad (*)$$

By differentiating the CDF, we have for any $t > 0$,

$$\begin{aligned} f_{-\log X}(t) &= \frac{d}{dt} \mathbf{P}(-\log X \leq t) = \frac{d}{dt} \mathbf{P}(X \geq e^{-t}) \\ &\stackrel{(*)}{=} \frac{d}{dt} (1 - e^{-t}) = e^{-t}. \end{aligned}$$

So, $f_{-\log X}(t) = e^{-t}$ for any $t > 0$, and $f_{-\log X}(t) = 0$ for any other t . $\square$

Exercise 8.10. Let X be a standard normal random variable. Find the PDF of e^X .

Solution. First, note that for any $0 < a < 1$, the definition of X implies that

$$\mathbf{P}(X \leq a) = \int_{-\infty}^a f_X(s) ds = \int_{-\infty}^a e^{-s^2/2} ds / \sqrt{2\pi}. \quad (*)$$

By differentiating the CDF and using the chain rule, we have for any $t > 0$,

$$\begin{aligned} f_{e^X}(t) &= \frac{d}{dt} \mathbf{P}(e^X \leq t) = \frac{d}{dt} \mathbf{P}(X \leq \log t) \\ &\stackrel{(*)}{=} \frac{d}{dt} \int_{-\infty}^{\log t} e^{-s^2/2} ds / \sqrt{2\pi} = \frac{1}{\sqrt{2\pi}} e^{-(\log t)^2/2} \frac{d}{dt} \log t = \frac{1}{\sqrt{2\pi}} e^{-(\log t)^2/2} \frac{1}{t} \end{aligned}$$

So, $f_{e^X}(t) = \frac{1}{\sqrt{2\pi}} e^{-(\log t)^2/2} \frac{1}{t}$ for any $t > 0$, and $f_{e^X}(t) = 0$ for any other t . $\square$

9. HOMEWORK 9

Exercise 9.1. Let X, Y be random variables with joint PDF $f_{X,Y}$. Let $a, b \in \mathbb{R}$. Using the definition of expected value, show that $\mathbf{E}(aX + bY) = a\mathbf{E}X + b\mathbf{E}Y$.

Solution. Using the joint pdf, we have

$$\begin{aligned} \mathbf{E}[aX + bY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (ax + by) f(x, y) dx dy \\ &= a \int_{-\infty}^{\infty} x \left[\int_{-\infty}^{\infty} f(x, y) dy \right] dx + b \int_{-\infty}^{\infty} y \left[\int_{-\infty}^{\infty} f(x, y) dx \right] dy \\ &= a \int_{-\infty}^{\infty} x f_X(x) dx + b \int_{-\infty}^{\infty} y f_Y(y) dy = a\mathbf{E}[X] + b\mathbf{E}[Y] \end{aligned}$$

$\square$

Exercise 9.2. Let X_1, Y_1 be random variables with joint PDF f_{X_1, Y_1} . Let X_2, Y_2 be random variables with joint PDF f_{X_2, Y_2} . Let $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ and let $S: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ so that $ST(x, y) = (x, y)$ and $TS(x, y) = (x, y)$ for every $(x, y) \in \mathbb{R}^2$. Let $J(x, y)$ denote the determinant of the Jacobian of S at (x, y) . Assume that $(X_2, Y_2) = T(X_1, Y_1)$. Using the change of variables formula from multivariable calculus, show that

$$f_{X_2, Y_2}(x, y) = f_{X_1, Y_1}(S(x, y)) |J(x, y)|.$$

Solution. According to the change of variables theorem, if U is a “nice” subset of $\mathbb{R}^2$ and ϕ is an injective differentiable function on U , then

$$\int_{\phi(U)} f(u, v) \, dudv = \int_U f(\phi(x, y)) |\det D\phi(x, y)| \, dxdy$$

where $D\phi(x, y)$ is the Jacobian of ϕ at (x, y) . Since $(X_2, Y_2) = T(X_1, Y_1)$, it follows that

$$\begin{aligned} \mathbf{P}((X_2, Y_2) \in U) &= \mathbf{P}((X_1, Y_1) \in S(U)) = \int_{S(U)} f_1(u, v) \, dudv \\ &= \int_U f_1(S(x, y)) |J(x, y)| \, dxdy \end{aligned}$$

On the other hand,

$$\mathbf{P}((X_2, Y_2) \in U) = \int_U f_2(x, y) \, dxdy$$

by the definition of a pdf. Therefore we have shown that

$$\int_U f_2(x, y) \, dxdy = \int_U f_1(S(x, y)) |J(x, y)| \, dxdy$$

for all “nice” subsets $U \subseteq \mathbb{R}^2$, which implies that $f_2(x, y) = f_1(S(x, y)) |J(x, y)|$, at least outside of some negligible set of points. $\square$

Exercise 9.3. Let X and Y be nonnegative random variables. Recall that we can define

$$\mathbf{E}X := \int_0^\infty \mathbf{P}(X > t) dt.$$

Assume that $X \leq Y$. Conclude that $\mathbf{E}X \leq \mathbf{E}Y$.

More generally, if X satisfies $\mathbf{E}|X| < \infty$, we define $\mathbf{E}X := \mathbf{E} \max(X, 0) - \mathbf{E} \max(-X, 0)$. If X, Y are any random variables with $X \leq Y$, $\mathbf{E}|X| < \infty$ and $\mathbf{E}|Y| < \infty$, show that $\mathbf{E}X \leq \mathbf{E}Y$.

Exercise 9.4. Let X, Y, Z be independent standard Gaussian random variables. Find the PDF of $\max(X, Y, Z)$.

Solution. We have, for any $t \in \mathbb{R}$, using the definition of maximum, independence, and the definition of a Gaussian,

$$\mathbf{P}(\max(X, Y, Z) \leq t) = \mathbf{P}(X \leq t, Y \leq t, Z \leq t) = \mathbf{P}(X \leq t)^3 = \left(\int_{-\infty}^t e^{-s^2/2} ds / \sqrt{2\pi} \right)^3.$$

So, by the chain rule,

$$f_{\max(X, Y, Z)}(t) = \frac{d}{dt} \mathbf{P}(\max(X, Y, Z) \leq t) = 3 \left(\int_{-\infty}^t e^{-s^2/2} ds / \sqrt{2\pi} \right)^2 e^{-t^2/2} \frac{1}{\sqrt{2\pi}}.$$

□

Exercise 9.5. Let X be a random variable uniformly distributed in $[0, 1]$ and let Y be a random variable uniformly distributed in $[0, 2]$. Suppose X and Y are independent. Find the PDF of X/Y^2 .

Solution. We have, for any $t > 0$, using the definition of joint PDF and of X and Y ,

$$\begin{aligned}
\mathbf{P}(X/Y^2 \leq t) &= \mathbf{P}(X \leq tY^2) = \iint_{\{(x,y) \in \mathbb{R}^2 : x \leq ty^2\}} f_{X,Y}(x,y) dx dy \\
&= \int_{x=0}^{x=\min(1,4t)} \int_{y=\min(2, \sqrt{x/t})}^{y=2} f_X(x) f_Y(y) dy dx = \int_{x=0}^{x=\min(1,4t)} \int_{y=\min(2, \sqrt{x/t})}^{y=2} \frac{1}{2} dy dx \\
&= \frac{1}{2} \int_{x=0}^{x=\min(1,4t)} [2 - \min(2, \sqrt{x/t})] dx \\
&= \begin{cases} \frac{1}{2} \int_{x=0}^{x=4t} [2 - \min(2, \sqrt{x/t})] dx & , \text{ if } 0 < t < 1/4 \\ \frac{1}{2} \int_{x=0}^{x=1} [2 - \min(2, \sqrt{x/t})] dx & , \text{ if } 1/4 < t \end{cases} \\
&= \begin{cases} \frac{1}{2} \int_{x=0}^{x=4t} [2 - \sqrt{x/t}] dx & , \text{ if } 0 < t < 1/4 \\ \frac{1}{2} \int_{x=0}^{x=1} [2 - \sqrt{x/t}] dx & , \text{ if } 1/4 < t \end{cases} \\
&= \begin{cases} \frac{1}{2} [8t - (2/3)(4t)^{3/2} t^{-1/2}] & , \text{ if } 0 < t < 1/4 \\ \frac{1}{2} [2 - (2/3)t^{-1/2}] & , \text{ if } 1/4 < t \end{cases} \\
&= \begin{cases} 4t - (8/3)t & , \text{ if } 0 < t < 1/4 \\ 1 - (1/3)t^{-1/2} & , \text{ if } 1/4 < t \end{cases} \\
&= \begin{cases} (4/3)t & , \text{ if } 0 < t < 1/4 \\ 1 - (1/3)t^{-1/2} & , \text{ if } 1/4 < t. \end{cases}
\end{aligned}$$

So,

$$f_{X/Y^2}(t) = \frac{d}{dt} \mathbf{P}(X/Y^2 \leq t) = \begin{cases} 4/3 & , \text{ if } 0 < t < 1/4 \\ \frac{1}{6} t^{-3/2} & , \text{ if } 1/4 < t. \end{cases}$$

□

Exercise 9.6. Let X, Y be independent random variables with joint PDF $f_{X,Y}$. Show that

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y).$$

Solution. $f(x, y) = f_X(x) f_Y(y)$ since X and Y are independent, hence

$$\begin{aligned}
\text{var}(X + Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (X + Y - \mathbf{E}[X] - \mathbf{E}[Y])^2 f(x, y) dx dy \\
&= \int_{-\infty}^{\infty} (x - \mathbf{E}[X])^2 f_X(x) f_Y(y) dx dy + \int_{-\infty}^{\infty} (y - \mathbf{E}[Y])^2 f_X(x) f_Y(y) dx dy \\
&\quad + 2 \int_{-\infty}^{\infty} (x - \mathbf{E}[X])(y - \mathbf{E}[Y]) f_X(x) f_Y(y) dx dy
\end{aligned}$$

$$\begin{aligned}
&= \text{var}(X) + \text{var}(Y) + 2 \left[\int_{-\infty}^{\infty} (x - \mathbf{E}[X]) f_X(x) dx \right] \left[\int_{-\infty}^{\infty} (y - \mathbf{E}[Y]) f_Y(y) dy \right] \\
&= \text{var}(X) + \text{var}(Y)
\end{aligned}$$

□

Exercise 9.7. Let X and Y be uniformly distributed random variables on $[0, 1]$. Assume that X and Y are independent. Compute the following probabilities:

- $\mathbf{P}(X > 3/4)$
- $\mathbf{P}(Y < X)$
- $\mathbf{P}(X + Y < 1/2)$
- $\mathbf{P}(\max(X, Y) > 1/2)$
- $\mathbf{P}(XY < 1/3)$.

Solution.

a. $\mathbf{P}(X > \frac{3}{4})$.

Solution: $\mathbf{P}(X > \frac{3}{4}) = 1 - \mathbf{P}(X \leq \frac{3}{4}) = 1 - \frac{3}{4} = \frac{1}{4}$.

b. $\mathbf{P}(Y < X)$.

Solution: The joint pdf of X and Y is $f(x, y) = 1$ if $0 \leq x, y \leq 1$ and $f(x, y) = 0$ otherwise, hence

$$\mathbf{P}(Y < X) = \int_0^1 \int_0^x dy dx = \int_0^1 x dx = \frac{1}{2}$$

c. $\mathbf{P}(X + Y < \frac{1}{2})$.

Solution: Using the same pdf as in part (b), we have

$$\mathbf{P}\left(X + Y < \frac{1}{2}\right) = \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}-x} dy dx = \int_0^{\frac{1}{2}} \frac{1}{2} - x dx = \frac{x - x^2}{2} \Big|_0^{\frac{1}{2}} = \frac{1}{8}$$

d. $\mathbf{P}(\max\{X, Y\} > \frac{1}{2})$.

Solution: Since X and Y are independent,

$$\begin{aligned}
\mathbf{P}\left(\max\{X, Y\} > \frac{1}{2}\right) &= 1 - \mathbf{P}\left(\max\{X, Y\} \leq \frac{1}{2}\right) = 1 - \mathbf{P}\left(X \leq \frac{1}{2}, Y \leq \frac{1}{2}\right) \\
&= 1 - \mathbf{P}\left(X \leq \frac{1}{2}\right) \mathbf{P}\left(Y \leq \frac{1}{2}\right) = 1 - \frac{1}{4} = \frac{3}{4}
\end{aligned}$$

Solution: If $X < \frac{1}{3}$, then $XY < \frac{1}{3}$ regardless of the value of Y . If $\frac{1}{3} \leq X \leq 1$, then we must have $Y < \frac{1}{3X}$. Therefore

$$\mathbf{P}\left(XY < \frac{1}{3}\right) = \frac{1}{3} + \int_{\frac{1}{3}}^1 \int_0^{\frac{1}{3x}} dy dx = \frac{1}{3} + \int_{\frac{1}{3}}^1 \frac{dx}{3x} = \frac{1}{3} \left(1 - \ln\left(\frac{1}{3}\right)\right) = \frac{1 + \ln(3)}{3}$$

□

Exercise 9.8. Let X, Y be random variables with $\mathbf{E}X^2 < \infty$ and $\mathbf{E}Y^2 < \infty$. Prove the **Cauchy-Schwarz inequality**:

$$\mathbf{E}(XY) \leq (\mathbf{E}X^2)^{1/2} (\mathbf{E}Y^2)^{1/2}.$$

Then, deduce the following when X, Y both have finite variance:

$$|\text{cov}(X, Y)| \leq (\text{var}(X))^{1/2} (\text{var}(Y))^{1/2}.$$

(Hint: in the case that $\mathbf{E}Y^2 > 0$, expand out the product $\mathbf{E}(X - Y\mathbf{E}(XY)/\mathbf{E}Y^2)^2$.)

Exercise 9.9. Suppose you go to the bus stop, and the time T between successive arrivals of the bus is anything between 0 and 30 minutes, with all arrival times being equally likely.

Suppose you get to the bus stop, and the bus just leaves as you arrive. How long should you expect to wait for the next bus? What is the probability that you will have to wait at least 15 minutes for the next bus to arrive?

On a different day, suppose you go to the bus stop and someone says the last bus came 10 minutes ago. How long should you expect to wait for the next bus? What is the probability that you will have to wait at least 10 minutes for the next bus to arrive?

Solution. The pdf of T is $f_T(t) = \frac{1}{30}$ if $0 \leq t \leq 30$, and $f_T(t) = 0$ otherwise. Therefore

$$\mathbf{E}[T] = \int_{-\infty}^{\infty} t f_T(t) dt = \frac{1}{30} \int_0^{30} t dt = \frac{t^2}{60} \Big|_0^{30} = 15$$

and

$$\mathbf{P}(T > 15) = \frac{1}{30} \int_{15}^{30} dt = \frac{1}{2}$$

Let $A = \{T \geq 10\}$. Then the conditional pdf of T given A is $f_{T|A}(t) = \frac{1}{20}$ if $10 \leq t \leq 30$, and $f_{T|A}(t) = 0$ otherwise. Therefore

$$\mathbf{E}[T|A] = \frac{1}{20} \int_{10}^{30} t dt = \frac{t^2}{40} \Big|_{10}^{30} = 20$$

which means that you should expect to wait an additional 10 minutes for the bus. Similarly,

$$\mathbf{P}(T > 20|A) = \frac{1}{20} \int_{20}^{30} dt = \frac{1}{2}$$

□

Exercise 9.10. Let $A_1, A_2, \dots$ be disjoint events such that $\mathbf{P}(A_i) = 2^{-i}$ for each $i \geq 1$. Assume that $\cup_{i=1}^{\infty} A_i = \Omega$. Let X be a random variable such that $\mathbf{E}(X|A_i) = (-1)^{i+1}$ for each $i \geq 1$. Compute $\mathbf{E}X$.

Solution. By the law of total expectation,

$$\mathbf{E}[X] = \sum_{i=1}^{\infty} \mathbf{E}[X|A_i] \mathbf{P}(A_i) = \sum_{i=1}^{\infty} (-1)^{i+1} 2^{-i}$$

Beginning with the geometric series

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}$$

for $|x| < 1$ and multiplying by x shows that

$$\sum_{i=1}^{\infty} x^i = \frac{x}{1-x}$$

Replacing x with $-x$ and multiplying by -1 yields

$$\sum_{i=1}^{\infty} (-1)^{i+1} x^i = \frac{x}{1+x}$$

so setting $x = \frac{1}{2}$, we obtain

$$\mathbf{E}[X] = \frac{\frac{1}{2}}{1 + \frac{1}{2}} = \frac{1}{3}$$

□

Exercise 9.11. Let X, Y be random variables. For any $y \in \mathbb{R}$, assume that $\mathbf{E}(X|Y = y) = e^{-|y|}$. Also, assume that Y has an exponential distribution with parameter $\lambda = 2$. Compute $\mathbf{E}X$.

Solution. By the law of total expectation,

$$\mathbf{E}[X] = \int_0^{\infty} \mathbf{E}[X|Y = y] f_Y(y) dy = 2 \int_0^{\infty} e^{-3y} dy = \frac{2}{3}$$

□

10. HOMEWORK 10

Exercise 10.1. Let X be a binomial random variable with parameters $n = 2$ and $p = 1/2$. So, $\mathbf{P}(X = 0) = 1/4$, $\mathbf{P}(X = 1) = 1/2$ and $\mathbf{P}(X = 2) = 1/4$. And X satisfies $\mathbf{E}X = 1$ and $\mathbf{E}X^2 = 3/2$.

Let Y be a geometric random variable with parameter $1/2$. So, for any positive integer k , $\mathbf{P}(Y = k) = 2^{-k}$. And Y satisfies $\mathbf{E}Y = 2$ and $\mathbf{E}Y^2 = 6$.

Let Z be a Poisson random variable with parameter 1. So, for any nonnegative integer k , $\mathbf{P}(Z = k) = \frac{1}{e} \frac{1}{k!}$. And Z satisfies $\mathbf{E}Z = 1$ and $\mathbf{E}Z^2 = 2$.

Let W be a discrete random variable such that $\mathbf{P}(W = 0) = 1/2$ and $\mathbf{P}(W = 4) = 1/2$, so that $\mathbf{E}W = 2$ and $\mathbf{E}W^2 = 8$.

Assume that X, Y, Z and W are all independent. Compute

$$\text{var}(X + Y + Z + W).$$

Solution. Since the variance of the sum is the sum of the variances for independent random variables, we have

$$\text{var}(X + Y + Z + W) = \text{var}(X) + \text{var}(Y) + \text{var}(Z) + \text{var}(W) = 1/2 + 2 + 1 + 4.$$

□

Exercise 10.2. Let $X_1, \dots, X_n$ be random variables with finite variance. Define an $n \times n$ matrix A such that $A_{ij} = \text{cov}(X_i, X_j)$ for any $1 \leq i, j \leq n$. Show that the matrix A is positive semidefinite. That is, show that for any $y = (y_1, \dots, y_n) \in \mathbb{R}^n$, we have

$$y^T A y = \sum_{i,j=1}^n y_i y_j A_{ij} \geq 0.$$

Solution.

$$\sum_{i,j=1}^n y_i y_j A_{ij} = \sum_{i,j=1}^n y_i y_j \text{cov}(X_i, X_j) = \sum_{i,j=1}^n \text{cov}(y_i X_i, y_j X_j) = \text{var}\left(\sum_{i=1}^n y_i X_i\right) \geq 0.$$

The last line used $\text{var}(Z) \geq 0$ for any random variable Z . $\square$

Exercise 10.3. Using the definition of convergence, show that the sequence of numbers $1, 1/2, 1/3, 1/4, \dots$ converges to 0.

Solution. Let $\varepsilon > 0$. We need to find $m \geq 0$ such that, for all $n \geq m$, we have $|1/n - 0| = 1/n \leq \varepsilon$. So, choose m to be any integer larger than $1/\varepsilon$, i.e. $m > 1/\varepsilon$ so that $1/m < \varepsilon$. If $n \geq m$, then $1/n \leq 1/m < \varepsilon$, as desired. $\square$

Exercise 10.4 (Uniqueness of limits). Let $x_1, x_2, \dots$ be a sequence of real numbers. Let $x, y \in \mathbb{R}$. Assume that $x_1, x_2, \dots$ converges to x . Assume also that $x_1, x_2, \dots$ converges to y . Prove that $x = y$. That is, a sequence of real numbers cannot converge to two different real numbers.

Solution. Let $\varepsilon > 0$. By assumption, there exists $m \geq 0$ such that, for all $n \geq m$, $|x_n - x| < \varepsilon$. By assumption, there exists $p \geq 0$ such that, for all $n \geq p$, $|x_n - y| < \varepsilon$. So, if $n \geq \max(m, p)$, we have

$$|x_n - x| < \varepsilon, \quad \text{and} \quad |x_n - y| < \varepsilon.$$

So, using the triangle inequality, for any $\varepsilon > 0$, if $n \geq \max(m, p)$, we have

$$|x - y| = |x - x_n + x_n - y| \leq |x - x_n| + |x_n - y| < \varepsilon + \varepsilon = 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we conclude that $x = y$. $\square$

Exercise 10.5. Let X be a random variable. Assume that $M_X(t)$ exists for all $t \in \mathbb{R}$, and assume we can differentiate under the expected value any number of times. For any positive integer n , show that

$$\frac{d^n}{dt^n} M_X(t) \Big|_{t=0} = \mathbf{E}(X^n).$$

So, in principle, all moments of X can be computed just by taking derivatives of the moment generating function.

Exercise 10.6. Let X be a standard Gaussian random variable. Compute an explicit formula for the moment generating function of X . (Hint: completing the square might be helpful.) From this explicit formula, compute an explicit formula for all moments of the Gaussian random variable. (The $2n^{\text{th}}$ moment of X should be something resembling a factorial.)

Solution.

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

If n is odd, then $\mathbf{E}X^n$ is zero, since $\mathbf{E}X^n = \int_{-\infty}^{\infty} x^n e^{-x^2/2} \frac{dx}{\sqrt{2\pi}} = 0$, since the integrand is odd. 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

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

□

Exercise 10.7. Construct two random variables $X, Y: \Omega \rightarrow \mathbb{R}$ such that $X \neq Y$ but $M_X(t), M_Y(t)$ exist for all $t \in \mathbb{R}$, and such that $M_X(t) = M_Y(t)$ for all $t \in \mathbb{R}$.

Solution. Let X be a mean zero standard Gaussian and let $Y = -X$. Then $X \neq Y$, but $M_X(t) = M_Y(t) = e^{t^2/2}$. □

Exercise 10.8. Unfortunately, there exist random variables X, Y such that $\mathbf{E}X^n = \mathbf{E}Y^n$ for all $n = 1, 2, 3, \dots$, but such that X, Y do not have the same CDF. First, explain why this does not contradict the Lévy Continuity Theorem, Weak Form. Now, let $-1 < a < 1$, and define a density

$$f_a(x) := \begin{cases} \frac{1}{x\sqrt{2\pi}} e^{-\frac{(\log x)^2}{2}} (1 + a \sin(2\pi \log x)) & , \text{ if } x > 0 \\ 0 & , \text{ otherwise.} \end{cases}$$

Suppose X_a has density f_a . If $-1 < a, b < 1$, show that $\mathbf{E}X_a^n = \mathbf{E}X_b^n$ for all $n = 1, 2, 3, \dots$ (Hint: write out the integrals, and make a change of variables $s = \log(x) + n$.)

Solution.

$$\begin{aligned} \mathbf{E}X^n &= \int_{-\infty}^{\infty} x^n f_a(x) dx = \int_0^{\infty} x^n \frac{1}{x\sqrt{2\pi}} e^{-\frac{(\log x)^2}{2}} (1 + a \sin(2\pi \log x)) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{(s-n)n} e^{-\frac{(s-n)^2}{2}} (1 + a \sin(2\pi(s-n))) ds \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{sn} e^{-\frac{s^2}{2}} (1 + a \sin(2\pi(s-n))) ds \\ &= \frac{e^{n^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(s-n)^2}{2}} (1 + a \sin(2\pi(s-n))) ds \\ &= \frac{e^{n^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{s^2}{2}} (1 + a \sin(2\pi s)) ds = e^{n^2/2}. \end{aligned}$$

The last equality used that the function $e^{-\frac{s^2}{2}} \sin(2\pi s)$ is odd, so its integral is zero. Since $\mathbf{E}X^n$ does not depend on a , the result follows. □

11. HOMEWORK 11

Exercise 11.1. Compute the characteristic function of a uniformly distributed random variable on $[-1, 1]$. (Some of the following formulas might help to simplify your answer: $e^{it} = \cos(t) + i \sin(t)$, $\cos(t) = [e^{it} + e^{-it}]/2$, $\sin(t) = [e^{it} - e^{-it}]/[2i]$, $t \in \mathbb{R}$.)

Solution.

$$\mathbf{E}e^{itX} = \frac{1}{2} \int_{-1}^1 e^{itx} dx = \frac{1}{2} \frac{1}{it} e^{itx} \Big|_{x=-1}^{x=1} = \frac{1}{2it} [e^{it} - e^{-it}] = \frac{\sin(t)}{t}.$$

□

Exercise 11.2. Let X be a random variable. Assume we can differentiate under the expected value of $\mathbf{E}e^{itX}$ any number of times. For any positive integer n , show that

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

So, in principle, all moments of X can be computed just by taking derivatives of the characteristic function.

Solution.

$$\frac{d^n}{dt^n} \Big|_{t=0} \phi_X(t) = \mathbf{E} \frac{d^n}{dt^n} \Big|_{t=0} e^{itX} = \mathbf{E}(iX)^n e^{itX} \Big|_{t=0} = i^n \mathbf{E}(X^n).$$

□

Exercise 11.3. Let X be a random variable such that $\mathbf{E}|X|^3 < \infty$. Prove that for any $t \in \mathbb{R}$,

$$\mathbf{E}e^{itX} = 1 + it\mathbf{E}X - t^2\mathbf{E}X^2/2 + o(t^2).$$

That is,

$$\lim_{t \rightarrow 0} t^{-2} |\mathbf{E}e^{itX} - [1 + it\mathbf{E}X - t^2\mathbf{E}X^2/2]| = 0$$

(Hint: it may be helpful to use Jensen's inequality, to first justify that $\mathbf{E}|X| < \infty$ and $\mathbf{E}X^2 < \infty$. Then, use the Taylor expansion with error bound: $e^{iy} = 1 + iy - y^2/2 - (i/2) \int_0^y (y-s)^2 e^{is} ds$, which is valid for any $y \in \mathbb{R}$.)

Actually, this same bound holds only assuming $\mathbf{E}X^2 < \infty$, but the proof of that bound requires things we have not discussed.

Solution. Using $y = tX$ in the Taylor expansion with error bound,

$$e^{itX} = 1 + itX - (tX)^2/2 - (i/2) \int_0^{tX} (tX - s)^2 e^{is} ds.$$

Taking expected values, rearranging, then taking the absolute values,

$$\begin{aligned} |\mathbf{E}e^{itX} - [1 + it\mathbf{E}X - t^2\mathbf{E}X^2/2]| &\leq \frac{1}{2} \mathbf{E} \left| \int_0^{tX} (tX - s)^2 e^{is} ds \right| \\ &\leq \frac{1}{2} \mathbf{E} \left| \int_0^{tX} (tX - s)^2 ds \right| \\ &\leq \frac{1}{2} \mathbf{E} |(1/3)(tX)^3| \leq \frac{1}{6} t^3 \mathbf{E}|X|^3. \end{aligned}$$

Therefore,

$$\lim_{t \rightarrow 0} t^{-2} |\mathbf{E}e^{itX} - [1 + it\mathbf{E}X - t^2\mathbf{E}X^2/2]| = (1/6) \mathbf{E}|X|^3 \lim_{t \rightarrow 0} t = 0.$$

□

Exercise 11.4. Let X, Y, Z be independent and uniformly distributed on $[0, 1]$. Note that f_X is not a continuous function.

Using convolution, compute f_{X+Y} . Draw f_{X+Y} . Note that f_{X+Y} is a continuous function, but it is not differentiable at some points.

Using convolution, compute f_{X+Y+Z} . Draw f_{X+Y+Z} . Note that f_{X+Y+Z} is a differentiable function, but it does not have a second derivative at some points.

Make a conjecture about how many derivatives $f_{X_1+\dots+X_n}$ has, where $X_1, \dots, X_n$ are independent and uniformly distributed on $[0, 1]$. You do not have to prove this conjecture. The idea of this exercise is that convolution is a kind of average of functions. And the more averaging you do, the more derivatives $f_{X_1+\dots+X_n}$ has.

Solution.

$$f_{X+Y}(t) = f_X * f_Y(t) = \int_{\mathbb{R}} 1_{[0,1]}(x) 1_{[0,1]}(t-x) dx = \int_{x=0}^{x=1} 1_{[0,1]}(t-x) dx = \int_{x=0}^{x=1} 1_{x \in [t-1, t]} dx$$

So, if $t < 0$, $f_{X+Y}(t) = 0$. If $t > 2$, $f_{X+Y}(t) = 0$. And if $t \in [0, 1]$, then $f_{X+Y} = \int_{x=0}^{x=t} dx = t$. And if $t \in [1, 2]$, then $f_{X+Y} = \int_{x=t-1}^{x=1} dx = 2 - t$. In summary,

$$f_{X+Y}(t) = \begin{cases} 0 & , \text{ if } t < 0 \\ t & , \text{ if } 0 \leq t < 1 \\ 2 - t & , \text{ if } 1 \leq t < 2 \\ 0 & , \text{ if } t \geq 2. \end{cases}$$

That is, f_{X+Y} is piecewise linear and continuous, and symmetric about the point $t = 1$. Similarly breaking into various cases, we have

$$\begin{aligned} f_{X+Y+Z}(t) &= f_{X+Y} * f_Z(t) = \int_{\mathbb{R}} f_{X+Y}(x) 1_{[0,1]}(t-x) dx = \int_{x=0}^{x=2} f_{X+Y}(x) 1_{x \in [t-1, t]} dx \\ &= \begin{cases} 0 & , \text{ if } t < 0 \\ \int_0^t s ds & , \text{ if } 0 \leq t < 1 \\ \int_{1-t}^1 s ds + \int_1^t (2-s) ds & , \text{ if } 1 \leq t < 2 \\ \int_{t-1}^2 (2-s) ds & , \text{ if } 2 \leq t < 3 \\ 0 & , \text{ if } t \geq 3. \end{cases} \\ &= \begin{cases} 0 & , \text{ if } t < 0 \\ t^2/2 & , \text{ if } 0 \leq t < 1 \\ (1/2)(1 - (1-t)^2) + (2t - t^2/2) - 3/2 & , \text{ if } 1 \leq t < 2 \\ (2s - s^2/2)4 - 2 - 2(t-1) + (t-1)^2/2 & , \text{ if } 2 \leq t < 3 \\ 0 & , \text{ if } t \geq 3. \end{cases} \\ &= \begin{cases} 0 & , \text{ if } t < 0 \\ t^2/2 & , \text{ if } 0 \leq t < 1 \\ -t^2 + 3t - 3/2 & , \text{ if } 1 \leq t < 2 \\ (1/2)(t-3)^2 & , \text{ if } 2 \leq t < 3 \\ 0 & , \text{ if } t \geq 3. \end{cases} \end{aligned}$$

That is, f_{X+Y+Z} is piecewise quadratic, continuous, and symmetric about the point $t = 3/2$.

When n is large, $f_{X_1+\dots+X_n}$ looks more and more like a Gaussian density. This observation agrees with the Central Limit Theorem. $\square$

Exercise 11.5. Construct two random variables X, Y such that X and Y are each uniformly distributed on $[0, 1]$, and such that $\mathbf{P}(X + Y = 1) = 1$.

Then construct two random variables W, Z such that W and Z are each uniformly distributed on $[0, 1]$, and such that $W + Z$ is uniformly distributed on $[0, 2]$.

(Hint: there is a way to do each of the above problems with about one line of work. That is, there is a way to solve each problem without working very hard.)

Solution. Let X be uniformly distributed in $[0, 1]$ and define $Y := 1 - X$.

Let W be uniformly distributed in $[0, 1]$ and let $Z := W$. Then $W + Z = 2W$ is uniformly distributed in $[0, 2]$, since $\mathbf{P}(2W \leq t) = \mathbf{P}(W \leq t/2) = t/2$ for all $0 \leq t \leq 2$. $\square$

Exercise 11.6. Let X be a standard Gaussian random variable. Let $t > 0$ and let n be a positive even integer. Show that

$$\mathbf{P}(X > t) \leq \frac{(n-1)(n-3) \cdots (3)(1)}{t^n}.$$

That is, the function $t \mapsto \mathbf{P}(X > t)$ decays faster than any monomial.

Solution. This follows by combining Markov's inequality with Exercise 10.6, noting also that $\mathbf{P}(X > t) \leq \mathbf{P}(|X| > t)$. $\square$

Exercise 11.7. Let X be a random variable. Let $t > 0$. Show that

$$\mathbf{P}(|X| > t) \leq \frac{\mathbf{E}X^4}{t^4}.$$

Solution. This follows from Markov's inequality with $n = 4$. $\square$

Exercise 11.8 (The Chernoff Bound). Let X be a random variable and let $r > 0$. Show that, for any $t > 0$,

$$\mathbf{P}(X > r) \leq e^{-tr} M_X(t).$$

Consequently, if $X_1, \dots, X_n$ are independent random variables with the same CDF, and if $r, t > 0$,

$$\mathbf{P}\left(\frac{1}{n} \sum_{i=1}^n X_i > r\right) \leq e^{-trn} (M_{X_1}(t))^n.$$

For example, if $X_1, \dots, X_n$ are independent Bernoulli random variables with parameter $0 < p < 1$, and if $r, t > 0$,

$$\mathbf{P}\left(\frac{X_1 + \cdots + X_n}{n} - p > r\right) \leq e^{-trn} (e^{-tp} [pe^t + (1-p)])^n.$$

And if we choose t appropriately, then the quantity $\mathbf{P}\left(\frac{1}{n} \left|\sum_{i=1}^n (X_i - p)\right| > r\right)$ becomes exponentially small as either n or r become large. That is, $\frac{1}{n} \sum_{i=1}^n X_i$ becomes very close to its mean. Importantly, the Chernoff bound is much stronger than either Markov's or Chebyshev's inequality, since they only respectively imply that

$$\mathbf{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - p\right| > r\right) \leq \frac{2p(1-p)}{r}, \quad \mathbf{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - p\right| > r\right) \leq \frac{p(1-p)}{nr^2}.$$

Solution. Since the exponential function is increasing, we have, for any $t > 0$ that the events $\{X > r\}$ and $\{e^{tX} > e^{tr}\}$ are the same, for any $r > 0$. So, their probabilities are the same, and applying Markov's inequality gives

$$\mathbf{P}(X > r) = \mathbf{P}(e^{tX} > e^{tr}) \leq \frac{\mathbf{E}e^{tX}}{e^{tr}} = e^{-tr} M_X(t).$$

□

12. HOMEWORK 12

Exercise 12.1. Let $X_1, X_2, \dots$ be independent random variables, each with exponential distribution with parameter $\lambda = 1$. For any $n \geq 1$, let $Y_n := \max(X_1, \dots, X_n)$. Let $0 < a < 1 < b$. Show that $\mathbf{P}(Y_n \leq a \log n) \rightarrow 0$ as $n \rightarrow \infty$, and $\mathbf{P}(Y_n \leq b \log n) \rightarrow 1$ as $n \rightarrow \infty$. Conclude that $Y_n / \log n$ converges to 1 in probability as $n \rightarrow \infty$.

Exercise 12.2. We say that random variables $X_1, X_2, \dots$ converge to a random variable X in L_2 if

$$\lim_{n \rightarrow \infty} \mathbf{E} |X_n - X|^2 = 0.$$

Show that, if $X_1, X_2, \dots$ converge to X in L_2 , then $X_1, X_2, \dots$ converges to X in probability.

Is the converse true? Prove your assertion.

Solution. From Markov's inequality, we have, for any $\varepsilon > 0$,

$$\mathbf{P}(|X_n - X| > \varepsilon) \leq \frac{\mathbf{E} |X_n - X|^2}{\varepsilon^2}.$$

The right quantity converges to 0 as $n \rightarrow \infty$ by assumption. We therefore conclude that $X_1, X_2, \dots$ converges in probability to X .

The converse is false in general. We can use the same example from class, where $X_n = n1_{[0, 1/n]}$ for all $n \geq 1$, and $\mathbf{P}$ is uniform on $[0, 1]$. Then $X_1, X_2, \dots$ converges in probability to 0, but $\mathbf{E}X_n^2 = n \rightarrow \infty$ as $n \rightarrow \infty$. □

Exercise 12.3. Let $X_1, X_2, \dots$ be independent, identically distributed random variables such that $\mathbf{E}|X_1| < \infty$ and $\text{var}(X_1) < \infty$. For any $n \geq 1$, define

$$Y_n := \frac{1}{n} \sum_{i=1}^n X_i^2.$$

Show that $Y_1, Y_2, \dots$ converges in probability. Express the limit in terms of $\mathbf{E}X_1$ and $\text{var}(X_1)$.

Solution. The weak law of large numbers implies that Y_n converges in probability to $\mathbf{E}X_1^2$ as $n \rightarrow \infty$, since $X_1^2, \dots, X_n^2$ are i.i.d. with mean $\mathbf{E}X_1^2$. Since $\mathbf{E}X_1^2 = \text{var}(X_1) + (\mathbf{E}X_1)^2$, Y_n converges in probability to $\text{var}(X_1) + (\mathbf{E}X_1)^2$ as $n \rightarrow \infty$. □

Exercise 12.4. Let $f, g, h: \mathbb{R} \rightarrow \mathbb{R}$. We use the notation $f(t) = o(g(t)) \forall t \in \mathbb{R}$ to denote $\lim_{t \rightarrow 0} \left| \frac{f(t)}{g(t)} \right| = 0$. For example, if $f(t) = t^3 \forall t \in \mathbb{R}$, then $f(t) = o(t^2)$, since $\lim_{t \rightarrow 0} \left| \frac{f(t)}{t^2} \right| = \lim_{t \rightarrow 0} |t| = 0$. Show: (i) if $f(t) = o(g(t))$ and if $h(t) = o(g(t))$, then $(f + h)(t) = o(g(t))$. (ii) If c is any nonzero constant, then $o(cg(t)) = o(g(t))$. (iii) $\lim_{t \rightarrow 0} g(t)o(1/g(t)) = 0$. (iv) $\lim_{t \rightarrow 0} o(g(t))/g(t) = 0$. (v) $o(g(t) + o(g(t))) = o(g(t))$.

Solution. (i) We have $|f(t) + h(t)| / |g(t)| \leq |f(t)/g(t)| + |h(t)/g(t)|$, so that $f + h(t) = o(g(t))$.

(ii) We have $|cg(t)| = |c| |g(t)|$, so $o(cg(t)) = o(g(t))$.

(iii) If $f(t) = o(1/g(t))$, then by definition, $|f(t)/(1/g(t))| = |f(t)g(t)| \rightarrow 0$ as $t \rightarrow 0$.

(iv) If $f(t) = o(g(t))$, then by definition, $|f(t)/g(t)| \rightarrow 0$ as $t \rightarrow 0$.

(v) If $f(t) = o(g(t))$, then by definition, $|f(t)/g(t)| \rightarrow 0$ as $t \rightarrow 0$. So, if $h(t) = o(g(t) + o(g(t))) = o(g(t) + f(t))$, then $\lim_{t \rightarrow 0} |h(t)/(g(t) + f(t))| = \lim_{t \rightarrow 0} \frac{|h(t)/g(t)|}{1 + f(t)/g(t)}$. The bottom goes to 1 as $t \rightarrow 0$, so that $\lim_{t \rightarrow 0} |h(t)/(g(t) + f(t))| = \lim_{t \rightarrow 0} |h(t)/g(t)|$. That is, $o(g(t) + o(g(t))) = o(g(t))$. $\square$

Exercise 12.5. This exercise demonstrates that geometry in high dimensions is different than geometry in low dimensions.

Let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Let $\|x\| := \sqrt{x_1^2 + \dots + x_n^2}$. Let $\varepsilon > 0$. Show that for all sufficiently large n , “most” of the cube $[-1, 1]^n$ is contained in the annulus

$$A := \{x \in \mathbb{R}^n : (1 - \varepsilon)\sqrt{n/3} \leq \|x\| \leq (1 + \varepsilon)\sqrt{n/3}\}.$$

That is, if $X_1, \dots, X_n$ are each independent and identically distributed in $[-1, 1]$, then for n sufficiently large

$$\mathbf{P}((X_1, \dots, X_n) \in A) \geq 1 - \varepsilon.$$

(Hint: apply the weak law of large numbers to $X_1^2, \dots, X_n^2$.)

Exercise 12.6 (Confidence Intervals, Optional). Among 625 members of a bank chosen uniformly at random among all bank members, it was found that 25 had a savings account. Give an interval of the form $[a, b]$ where $0 \leq a, b \leq 625$ are integers, such that with about 95% certainty, if we sample 625 bank members independently and uniformly at random (from a very large bank membership), then the number of these people with savings accounts lies in the interval $[a, b]$. (Hint: if Y is a standard Gaussian random variable, then $\mathbf{P}(-2 \leq Y \leq 2) \approx .95$.)

Solution. Let X_i be the indicator random variable which is 1 if the i^{th} member had a savings account and 0 if not, for all $i \in \{1, 2, \dots, 625\}$. Then we are assuming the $X_1, X_2, \dots$ are i.i.d. with $\mathbf{P}(X_1 = 1) = p = 25/625 = 1/25$ and thus $\mathbf{E}[X_1] = p$ $\text{var}(X_1) = p(1 - p) = 24/625$. Then by the central limit theorem, we have

$$\mathbf{P}\left(-2 \leq \frac{X_1 + \dots + X_{625} - 625p}{\sqrt{625p(1 - p)}} \leq 2\right) \approx .95$$

That is,

$$\mathbf{P}\left(625p - 2\sqrt{625p(1 - p)} \leq X_1 + \dots + X_{625} \leq 625p + 2\sqrt{625p(1 - p)}\right) \approx .95$$

Using $p = 1/25$,

$$\mathbf{P}\left(25 - 2\sqrt{24} \leq X_1 + \dots + X_{625} \leq 25 + 2\sqrt{24}\right) \approx .95$$

So, with about 95% certainty, the number of the 625 chosen bank members with savings accounts lies in the interval $[15.2, 34.8]$. Since the number of bank members is an integer, this interval can be chosen as $[15, 35]$. $\square$

Exercise 12.7 (Hypothesis Testing, Optional). Suppose we run a casino, and we want to test whether or not a particular roulette wheel is biased. Let p be the probability that red results from one spin of the roulette wheel. Using statistical terminology, “ $p = 18/38$ ” is the null hypothesis, and “ $p \neq 18/38$ ” is the alternative hypothesis. (On a standard roulette wheel, 18 of the 38 spaces are red.) For any $i \geq 1$, let $X_i = 1$ if the i^{th} spin is red, and let $X_i = 0$ otherwise.

Let $\mu := \mathbf{E}X_1$ and let $\sigma := \sqrt{\text{var}(X_1)}$. If the null hypothesis is true, and if Y is a standard Gaussian random variable

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\left| \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \right| \geq 2 \right) = \mathbf{P}(|Y| \geq 2) \approx .05.$$

To test the null hypothesis, we spin the wheel n times. In our test, we reject the null hypothesis if $|X_1 + \cdots + X_n - n\mu| > 2\sigma\sqrt{n}$. Rejecting the null hypothesis when it is true is called a type I error. In this test, we set the type I error percentage to be 5%. (The type I error percentage is closely related to the p-value.)

Suppose we spin the wheel $n = 3800$ times and we get red 1868 times. Is the wheel biased? That is, can we reject the null hypothesis with around 95% certainty?

Solution. Assume that the null hypothesis is true. Then $\sigma = \sqrt{p(1-p)} = \sqrt{(18/38)(20/38)}$. Then, plug in the sample values to get

$$|X_1 + \cdots + X_n - np| / (\sigma\sqrt{n}) = |(1868 - 3800(18/38))| / [\sqrt{(18/38)(20/38)}\sqrt{3800}] \approx 2.2 > 2.$$

Thus we reject the null hypothesis with around .95% certainty. $\square$

Exercise 12.8. Suppose random variables $X_1, X_2, \dots$ converge in probability to a random variable X . Prove that $X_1, X_2, \dots$ converge in distribution to X .

Then, show that the converse is false.

ALL EXERCISES BELOW ARE OPTIONAL. THEY WILL NOT BE GRADED.

Exercise 12.9 (Optional). Let $X_1, X_2, \dots$ be independent identically distributed random variables with $\mathbf{P}(X_1 = 1) = \mathbf{P}(X_1 = -1) = 1/2$. For any $n \geq 1$, define

$$S_n := \frac{X_1 + \cdots + X_n}{\sqrt{n}}.$$

The Central Limit Theorem says that S_n converges in distribution to a standard Gaussian random variable. We show that S_n does not converge in probability to any random variable. The intuition here is that if S_n did converge in probability to a random variable Z , then when n is large, S_n is close to Z , $Y_n := \frac{\sqrt{2}S_{2n} - S_n}{\sqrt{2-1}}$ is close to Z , but S_n and Y_n are independent. And this cannot happen.

Proceed as follows. Assume that S_n converges in probability to Z .

- Let $\varepsilon > 0$. For n very large (depending on ε), we have $\mathbf{P}(|S_n - Z| > \varepsilon) < \varepsilon$ and $\mathbf{P}(|Y_n - Z| > \varepsilon) < \varepsilon$.
- Show that $\mathbf{P}(S_n > 0, Y_n > 0)$ is around $1/4$, using independence and the Central Limit Theorem.
- From the first item, show $\mathbf{P}(S_n > 0 | Z > \varepsilon) > 1 - \varepsilon$, $\mathbf{P}(Y_n > 0 | Z > \varepsilon) > 1 - \varepsilon$, so $\mathbf{P}(S_n > 0, Y_n > 0 | Z > \varepsilon) > 1 - 2\varepsilon$.
- Without loss of generality, for ε small, we have $\mathbf{P}(Z > \varepsilon) > 4/9$.

- By conditioning on $Z > \varepsilon$, show that $\mathbf{P}(S_n > 0, Y_n > 0)$ is at least $3/8$, when n is large.

Exercise 12.10 (Optional). Let $X_1, X_2, \dots$ be random variables that converge almost surely to a random variable X . That is,

$$\mathbf{P}(\lim_{n \rightarrow \infty} X_n = X) = 1.$$

Show that $X_1, X_2, \dots$ converges in probability to X in the following way.

- For any $\varepsilon > 0$ and for any positive integer n , let

$$A_{n,\varepsilon} := \bigcup_{m=n}^{\infty} \{\omega \in \Omega : |X_m(\omega) - X(\omega)| > \varepsilon\}.$$

Show that $A_{n,\varepsilon} \supseteq A_{n+1,\varepsilon} \supseteq A_{n+2,\varepsilon} \supseteq \dots$.

- Show that $\mathbf{P}(\cap_{n=1}^{\infty} A_{n,\varepsilon}) = 0$.
- Using Continuity of the Probability Law, deduce that $\lim_{n \rightarrow \infty} \mathbf{P}(A_{n,\varepsilon}) = 0$.

Now, show that the converse is false. That is, find random variables $X_1, X_2, \dots$ that converge in probability to X , but where $X_1, X_2, \dots$ do not converge to X almost surely.

Exercise 12.11 (Renewal Theory, Optional). Let $t_1, t_2, \dots$ be positive, independent identically distributed random variables. Let $\mu \in \mathbb{R}$. Assume $\mathbf{E}t_1 = \mu$. For any positive integer j , we interpret t_j as the lifetime of the j^{th} lightbulb (before burning out, at which point it is replaced by the $(j+1)^{\text{st}}$ lightbulb). For any $n \geq 1$, let $T_n := t_1 + \dots + t_n$ be the total lifetime of the first n lightbulbs. For any positive integer t , let $N_t := \min\{n \geq 1 : T_n \geq t\}$ be the number of lightbulbs that have been used up until time t . Show that N_t/t converges almost surely to $1/\mu$ as $t \rightarrow \infty$. (Hint: by definition of N_t , we have $T_{N_t-1} < t \leq T_{N_t}$. Now divide the inequalities by N_t and apply the Strong Law.)

Solution. From the Strong Law of Large Numbers,

$$\mathbf{P}(\lim_{n \rightarrow \infty} T_n/n = \mu) = 1. \quad (*)$$

In particular,

$$\mathbf{P}(\lim_{t \rightarrow \infty} N_t = \infty) = 1. \quad (**)$$

By definition of N_t , $T_{N_t-1} < t \leq T_{N_t}$. Dividing this by N_t , we get

$$\frac{N_t - 1}{N_t} \frac{T_{N_t-1}}{N_t - 1} = \frac{T_{N_t-1}}{N_t} \leq \frac{t}{N_t} \leq \frac{T_{N_t}}{N_t}. \quad (***)$$

Letting $t \rightarrow \infty$, $(**)$ implies that $\lim_{t \rightarrow \infty} \frac{N_t-1}{N_t} = 1$ with probability one. The combination of $(*)$ and $(**)$ implies that, with probability one,

$$\lim_{t \rightarrow \infty} \frac{T_{N_t-1}}{N_t - 1} = \lim_{t \rightarrow \infty} \frac{T_{N_t}}{N_t} = \mu.$$

So, $(***)$ implies that, with probability one,

$$\lim_{t \rightarrow \infty} \frac{t}{N_t} = \mu.$$

So, with probability one,

$$\lim_{t \rightarrow \infty} \frac{N_t}{t} = \frac{1}{\mu}.$$

□

Exercise 12.12 (Playing Monopoly Forever, Optional). Let $t_1, t_2, \dots$ be independent random variables, all of which are uniform on $\{1, 2, 3, 4, 5, 6\}$. For any positive integer j , we think of t_j as the result of rolling a single fair six-sided die. For any $n \geq 1$, let $T_n = t_1 + \dots + t_n$ be the total number of spaces that have been moved after the n^{th} roll. (We think of each roll as the amount of moves forward of a game piece on a very large Monopoly game board.) For any positive integer t , let $N_t := \min\{n \geq 1: T_n \geq t\}$ be the number of rolls needed to get t spaces away from the start. Using Exercise 12.11, show that N_t/t converges almost surely to $2/7$ as $t \rightarrow \infty$.

Solution. Apply Exercise 12.11 with $\mu = \mathbf{E}t_1 = 7/2$. □

Exercise 12.13 (Random Numbers are Normal, Optional). Let X be a uniformly distributed random variable on $(0, 1)$. Let X_1 be the first digit in the decimal expansion of X . Let X_2 be the second digit in the decimal expansion of X . And so on.

- Show that the random variables $X_1, X_2, \dots$ are uniform on $\{0, 1, 2, \dots, 9\}$ and independent.
- Fix $m \in \{0, 1, 2, \dots, 9\}$. Using the Strong Law of Large Numbers, show that with probability one, the fraction of appearances of the number m in the first n digits of X converges to $1/10$ as $n \rightarrow \infty$.

(Optional): Show that for any ordered finite set of digits of length k , the fraction of appearances of this set of digits in the first n digits of X converges to 10^{-k} as $n \rightarrow \infty$. (You already proved the case $k = 1$ above.) That is, a randomly chosen number in $(0, 1)$ is normal. On the other hand, if we just pick some number such that $\sqrt{2} - 1$, then it may not be easy to say whether or not that number is normal.

(As an optional exercise, try to explicitly write down a normal number. This may not be so easy to do, even though a random number in $(0, 1)$ satisfies this property!)

Solution. Fix $x_1, \dots, x_n \in \{0, 1, 2, \dots, 9\}$. If we specify the first n decimals of a number in $(0, 1)$, then the fraction of numbers in $(0, 1)$ with those specified decimals is 10^{-n} . That is,

$$\mathbf{P}(X_1 = x_1, \dots, X_n = x_n) = 10^{-n}.$$

By similar reasoning,

$$\mathbf{P}(X_i = x_i) = 10^{-1}, \quad \forall 1 \leq i \leq n. \quad (*)$$

That is, if we specify the i^{th} decimal of a number in $(0, 1)$, then the fraction of numbers in $(0, 1)$ with that specified decimal is 10^{-1} . Therefore,

$$\mathbf{P}(X_1 = x_1, \dots, X_n = x_n) = 10^{-n} = \mathbf{P}(X_1 = x_1) \cdots \mathbf{P}(X_n = x_n), \quad \forall x_1, \dots, x_n \in \{0, 1, \dots, 9\}.$$

That is, $X_1, \dots, X_n$ are independent. Also (*) implies that $X_1, X_2, \dots$ are uniformly distributed in $\{0, 1, \dots, 9\}$.

Fix $m \in \{0, 1, \dots, 9\}$. For any $i \geq 1$, let $Y_i = 1_{\{X_i = m\}}$. Since $X_1, X_2, \dots$ are i.i.d., $Y_1, Y_2, \dots$ are i.i.d. as well. The quantity $(Y_1 + \dots + Y_n)/n$ is the fraction of the number of appearances of m in the first n digits of X . Also by definition of Y_i , $\mathbf{E}Y_i = \mathbf{P}(X_i = m) = 1/10$ by (*). So the Strong Law says

$$\mathbf{P}\left(\lim_{n \rightarrow \infty} \frac{Y_1 + \dots + Y_n}{n} = \frac{1}{10}\right) = 1.$$

□

Exercise 12.14 (Optional). Using the Central Limit Theorem, prove the Weak Law of Large Numbers.

Exercise 12.15 (Optional). Let $X_1, X_2, \dots$ be random variables with mean zero and variance one. The Strong Law of Large Numbers says that $\frac{1}{n}(X_1 + \dots + X_n)$ converges almost surely to zero. The Central Limit Theorem says that $\frac{1}{\sqrt{n}}(X_1 + \dots + X_n)$ converges in distribution to a standard Gaussian random variable. But what happens if we divide by some other power of n ? This Exercise gives a partial answer to this question.

Let $\varepsilon > 0$. Show that

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

converges to zero almost surely as $n \rightarrow \infty$. (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 .)

USC DEPARTMENT OF MATHEMATICS, LOS ANGELES, CA
E-mail address: `stevenmheilman@gmail.com`