Answers to Exercises

Exercise 2.1

1. Let $\mathcal{F}$ be a field over Ω. Show that
   (a) $\emptyset \in \mathcal{F}$,

   Answer: $\mathcal{F}$ is closed under complementation, and $\emptyset = \Omega^c$ where $\Omega \in \mathcal{F}$, so $\emptyset = \Omega^c$.

   (b) $\mathcal{F}$ is closed under set differences,

   Answer: Let $A, B \in \mathcal{F}$. Then $A \setminus B = A \cap B^c = (A^c \cup B)^c$ (De Morgan's laws). Hence $A \setminus B \in \mathcal{F}$ because $\mathcal{F}$ is closed under complementation and pairwise unions.

   (c) $\mathcal{F}$ is closed under pairwise intersections,

   Answer: Let $A, B \in \mathcal{F}$. Then $A \cap B = (A^c \cup B^c)^c$ (De Morgan's laws). Hence $A \cap B \in \mathcal{F}$ because $\mathcal{F}$ is closed under complementation and pairwise unions.

   (d) $\mathcal{F}$ is closed under finite unions,

   Answer: Proof by induction. Suppose that $\mathcal{F}$ is closed under unions of n sets (where $n \geq 2$). Let $A_1, A_2, \ldots, A_{n+1} \in \mathcal{F}$. By the inductive hypothesis, $\cup_{i=1}^{n} \in \mathcal{F}$, so $\cup_{i=1}^{n+1} A_i = \left[\cup_{i=1}^{n} A_i\right] \cup A_{n+1} \in \mathcal{F}$ because $\mathcal{F}$ is closed under pairwise unions.

   (e) $\mathcal{F}$ is closed under finite intersections.

   Answer: Let $A_1, A_2, \ldots, A_n \in \mathcal{F}$. Then $\cap_{i=1}^{n} A_i = \left[\cup_{i=1}^{n} A_i^c\right]^c$ (De Morgan's laws). Hence $\cap_{i=1}^{n} A_i \in \mathcal{F}$ because $\mathcal{F}$ is closed under complementation and finite unions.

2. Let $\mathcal{F}$ be a σ-field over Ω. Show that
   (a) $\mathcal{F}$ is closed under finite unions,

   Answer: Let $A_1, A_2, \ldots, A_n \in \mathcal{F}$. Since $\mathcal{F}$ is closed under countable unions and $\emptyset \in \mathcal{F}$,
   $$\cup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \ldots \cup A_n \cup \emptyset \cup \emptyset \ldots \in \mathcal{F}.$$

   (b) $\mathcal{F}$ is closed under finite intersections.

   Answer: Let $A_1, A_2, \ldots, A_n \in \mathcal{F}$. Since $\mathcal{F}$ is closed under complementation and finite unions,
   $$\cap_{i=1}^{n} A_i = A_1 \cap \ldots \cap A_n = (A_1^c \cup \ldots \cup A_n^c)^c \in \mathcal{F}.$$

   (c) $\mathcal{F}$ is closed under countable intersections.

Answer: Let $A_1, A_2, \ldots \in \mathcal{F}$. Since $\mathcal{F}$ is closed under complementation and countable unions,

$$\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c\right)^c \in \mathcal{F}.$$

Exercise 2.2

1. Let $\Omega = \{1,2,3,4,5,6\}$.
   (a) What is the smallest σ-field containing the event $A = \{1,2\}$?

   Answer: A σ-field must contain $\emptyset$ and Ω, and be closed under complementation and countable unions.
   The smallest σ-field containing $A = \{1,2\}$ is therefore

   $$\mathcal{F} = \{\emptyset, \{1,2\}, \{3,4,5,6\}, \Omega\}$$

   (b) What is the smallest σ-field containing the events $A = \{1,2\}$, $B = \{3,4\}$ and $C = \{5,6\}$?

   Answer:
   $$\mathcal{F} = \{\emptyset, \{1,2\}, \{3,4\}, \{5,6\}, \{1,2,3,4\}, \{1,2,5,6\}, \{3,4,5,6\}, \Omega\}$$

2. Let $\mathcal{F}$ and $\mathcal{G}$ be σ-fields over Ω.
   (a) Show that $\mathcal{H} = \mathcal{F} \cap \mathcal{G}$ is a σ-field over Ω.

   Answer: $\mathcal{H}$ is a σ-field because:
   - $\emptyset \in \mathcal{F}$ and $\emptyset \in \mathcal{G}$ so $\emptyset \in \mathcal{H}$;
   - if A belongs to both $\mathcal{F}$ and $\mathcal{G}$, then A^c belongs to both $\mathcal{F}$ and $\mathcal{G}$, so $\mathcal{H}$ is closed under complementation;
   - if $A_1, A_2, \ldots$ all belong to both $\mathcal{F}$ and $\mathcal{G}$, then their union also lies in both $\mathcal{F}$ and $\mathcal{G}$, so $\mathcal{H}$ is closed under countable unions.

   (b) Find a counterexample to show that $\mathcal{H} = \mathcal{F} \cup \mathcal{G}$ is not necessarily a σ-field over Ω.

   Answer: Let $\Omega = \{a,b,c\}$, $\mathcal{G} = \big\{\emptyset, \{a\}, \{b,c\}, \Omega\big\}$ and $\mathcal{G} = \big\{\emptyset, \{a,b\}, \{c\}, \Omega\big\}$. Then

   $$\mathcal{H} = \mathcal{F} \cup \mathcal{G} = \big\{\emptyset, \{a\}, \{c\}, \{a,b\}, \{b,c\}, \Omega\big\}.$$

   Hence $\{a\} \in \mathcal{H}$ and $\{c\} \in \mathcal{H}$, but $\{a,c\} \notin \mathcal{H}$ so $\mathcal{H}$ is not a σ-field.

Exercise 3.1

1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $A, B, C \in \mathcal{F}$. Show that

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

   This is called the *inclusion-exclusion principle.*

Answer: Set union is an associative operator: $A \cup B \cup C = (A \cup B) \cup C$, so by the addition rule,

$$\begin{aligned}\mathbb{P}(A \cup B \cup C) &= \mathbb{P}\big((A \cup B) \cup C\big) \\ &= \mathbb{P}(A \cup B) + \mathbb{P}(C) - \mathbb{P}\big((A \cup B) \cap C\big).\end{aligned}$$

Set intersection is distributive over set union: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$, so by the addition rule,

$$\begin{aligned}\mathbb{P}\big((A \cup B) \cap C\big) &= \mathbb{P}\big((A \cap C) \cup (B \cap C)\big) \\ &= \mathbb{P}(A \cap C) + \mathbb{P}(B \cap C) - \mathbb{P}\big((A \cap C) \cap (B \cap C)\big) \\ &= \mathbb{P}(A \cap C) + \mathbb{P}(B \cap C) - \mathbb{P}(A \cap B \cap C).\end{aligned}$$

2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
   (a) Show that $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$ for all $A, B \in \mathcal{F}$. This is called *subadditivity*.

   Answer: TODO

   (b) Show that for any sequence $A_1, A_2, \ldots$ of events in $\mathcal{F}$,

   $$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

   This is called *countable subadditivity*.

   Answer: TODO

Exercise 3.2

1. Let A and B be events with probabilities $\mathbb{P}(A) = 3/4$ and $\mathbb{P}(B) = 1/3$.
   (a) Show that $\frac{1}{12} \leq \mathbb{P}(A \cap B) \leq \frac{1}{3}$, and construct examples to show that both extremes are possible.

   Answer:
   - Lower bound: $\mathbb{P}(A \cup B) \leq 1$ so $\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1 = \frac{1}{12}$.
   - Upper bound: $A \cap B \subseteq A$ and $A \cap B \subseteq B$, so $\mathbb{P}(A \cap B) \leq \min\{\mathbb{P}(A), \mathbb{P}(B)\} = \frac{1}{3}$.

   Example: let $\Omega = \{1, 2, \ldots, 12\}$ with each outcome equally likely, and let $A = \{1, 2, \ldots, 9\}$.
   - Let $B = \{9, 10, 11, 12\}$. Then $\mathbb{P}(A \cap B) = \mathbb{P}(\{9\}) = \frac{1}{12}$.
   - Let $B = \{1, 2, 3, 4\}$. Then $\mathbb{P}(A \cap B) = \mathbb{P}(\{1, 2, 3, 4\}) = \frac{1}{3}$.

   (b) Find corresponding bounds for $\mathbb{P}(A \cup B)$.

   Answer:
   - Upper bound: $\mathbb{P}(A \cup B) \leq \min\{\mathbb{P}(A) + \mathbb{P}(B), 1\} = 1$.
   - Lower bound: $\mathbb{P}(A \cup B) \geq \max\{\mathbb{P}(A), \mathbb{P}(B)\} = 3/4$.

   These bounds are attained in the above example.

2. A roulette wheel consists of 37 slots of equal size. The slots are numbered from 0 to 36, with odd-numbered slots coloured red, even-numbered slots coloured black, and the slot labelled 0 coloured green. The wheel is spun in one direction and a ball is rolled in the opposite direction along a track running around the circumference of the wheel. The ball eventually falls on to the wheel and into one of the 37 slots. A player bets on the event that the ball stops in a red slot, and another player bets on the event that the ball stops in a black slot.
   (a) Define a suitable sample space Ω for this random experiment, and identify the events of interest.

Answer: A suitable sample space for the experiment is $\Omega = \{0, 1, 2, \dots, 36\}$. The events of interest are $G = \{0\}$, $R = \{1, 3, \dots, 35\}$ and $B = \{2, 4, \dots, 36\}$.

(b) Find the smallest field $\mathcal{F}$ over Ω that contains the events of interest.

Answer: The smallest field of sets containing the events G, R and B is

$$\mathcal{F} = \{\emptyset, G, R, B, G \cup R, G \cup B, R \cup B, \Omega\}.$$

$\mathcal{F}$ is indeed a field of sets, because

- $\Omega \in \mathcal{F}$,
- $\mathcal{F}$ is closed under complementation,
  - $\emptyset^c = \Omega \in \mathcal{F}$ and $\Omega^c = \emptyset \in \mathcal{F}$,
  - $G^c = R \cup B \in \mathcal{F}$, $R^c = B \cup G \in \mathcal{F}$ and $B^c = R \cup G \in \mathcal{F}$,
  - $(G \cup R)^c = B \in \mathcal{F}$, $(G \cup B)^c = R \in \mathcal{F}$ and $(R \cup B)^c = G \in \mathcal{F}$
- $\mathcal{F}$ is closed under pairwise unions, for example
  - $R \cup \emptyset = R \in \mathcal{F}$ and $R \cup \Omega = \Omega \in \mathcal{F}$,
  - $R \cup B \in \mathcal{F}$ and $R \cup G \in \mathcal{F}$,
  - $R \cup (R \cup B) = R \cup B \in \mathcal{F}$,
  - $R \cup (R \cup G) = R \cup G \in \mathcal{F}$,
  - $R \cup (B \cup G) = \Omega \in \mathcal{F}$.

  and so on.

(c) Define a suitable probability measure $(\Omega, \mathcal{F})$ to represent the game.

Answer: A suitable probability measure over $(\Omega, \mathcal{F})$ is given by

$$\begin{aligned}
&\mathbb{P}(\emptyset) = 0, \\
&\mathbb{P}(R) = 18/37,\ \mathbb{P}(B) = 18/37,\ \mathbb{P}(G) = 1/37, \\
&\mathbb{P}(R \cup B) = 36/37,\ \mathbb{P}(R \cup G) = 19/37,\ \mathbb{P}(B \cup G) = 19/37, \\
&\mathbb{P}(\Omega) = 1.
\end{aligned}$$

This is indeed a probability measure, because

- $\mathbb{P}(\emptyset) = 0$,
- $\mathbb{P}(\Omega) = 1$, and
- $\mathbb{P}$ is additive over $\mathcal{F}$; for example,
  - $\frac{36}{37} = \mathbb{P}(R \cup B) = \mathbb{P}(R) + \mathbb{P}(B) = \frac{18}{37} + \frac{18}{37} = \frac{36}{37}$,
  - $\frac{19}{37} = \mathbb{P}(R \cup G) = \mathbb{P}(R) + \mathbb{P}(G) = \frac{18}{37} + \frac{1}{37} = \frac{19}{37}$,
  - $\frac{19}{37} = \mathbb{P}(B \cup G) = \mathbb{P}(B) + \mathbb{P}(G) = \frac{18}{37} + \frac{1}{37} = \frac{19}{37}$,

  and so on.

Exercise 3.3

1. A biased coin has probability p of showing heads. The coin is tossed repeatedly until a head occurs. Describe a suitable probability space for this experiment.

Answer: The sample space is the set of all finite sequences of tails followed by a head, together with the infinite sequence of tails:

$$\Omega = \{T^n H : n \geq 0\} \cup \{T^\infty\}.$$

The σ-field can be taken to be the power set of Ω, and the probability measure can be defined on the elementary events by

$$\begin{aligned}\mathbb{P}(T^nH) &= (1-p)^n p,\\ \mathbb{P}(T^\infty) &= \lim_{n\to\infty}(1-p)^n = 0 \text{ if } p \neq 0.\end{aligned}$$

2. A fair coin is tossed repeatedly.
   (a) Show that a head eventually occurs with probability one.

Answer: Let A_n be the event that no heads occur in the first n tosses, and let A be the event that no heads occur at all. Then $A_1, A_2, \dots$ is a decreasing sequence ($A_{n+1} \subset A_n$), with $A = \cap_{i=1}^\infty A_n$. Hence by the continuity property of probability measures,

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcap_{n=1}^\infty A_n\right) = \lim_{n\to\infty}\mathbb{P}(A_n) = \lim_{n\to\infty}\left(\frac{1}{2}\right)^n = 0,$$

or alternatively,

$$\mathbb{P}(\text{no heads}) = \lim_{n\to\infty}\mathbb{P}(\text{no heads in the first } n \text{ tosses}) = \lim_{n\to\infty}\left(\frac{1}{2}\right)^n = 0.$$

Thus we are certain of eventually observing a head.

   (b) Show that a sequence of 10 consecutive tails eventually occurs with probability one.

Answer: Let us think of the first $10n$ tosses as disjoint groups of consecutive outcomes, each group of length 10. The probability any one of the n groups consists of 10 consecutive tails is 2^{-10}, independently of the other groups. The event that one of the groups consists of 10 consecutive tails is a subset of the event that a sequence of 10 consecutive tails appears anywhere in the first $10n$ tosses. Hence, using the continuity of probability measures,

$$\begin{aligned}\mathbb{P}(10T \text{ eventually appears}) &= \lim_{n\to\infty}\mathbb{P}(10T \text{ occurs somewhere in the first } 10n \text{ tosses})\\ &\geq \lim_{n\to\infty}\mathbb{P}(10T \text{ occurs as one of the first } n \text{ groups of } 10)\\ &= 1 - \lim_{n\to\infty}\mathbb{P}(10T \text{ does not occur as one of the first } n \text{ groups of } 10)\\ &= 1 - \lim_{n\to\infty}\left(1 - \frac{1}{2^{10}}\right)^n = 1.\end{aligned}$$

Thus we are certain of eventually observing sequence of 10 consecutive tails.

   (c) Show that any finite sequence of heads and tails eventually occurs with probability one.

Answer: Let s be a fixed sequence of length k. As in the previous part, we think of the first kn tosses as n distinct groups of length k. The event that the one of these groups is exactly equal to s is a subset of the event that first kn tosses contains at least one instance of s. Hence

$$\begin{aligned}\mathbb{P}(s \text{ eventually appears}) &= \lim_{n\to\infty}\mathbb{P}(s \text{ occurs somewhere in the first } kn \text{ tosses})\\ &\geq \lim_{n\to\infty}\mathbb{P}(s \text{ occurs as one of the first } n \text{ groups of } k)\\ &= 1 - \lim_{n\to\infty}\mathbb{P}(s \text{ does not occur as one of the first } n \text{ groups of } k)\\ &= 1 - \lim_{n\to\infty}\left(1 - \frac{1}{2^k}\right)^n = 1.\end{aligned}$$

Thus we are certain of eventually observing the sequence s.

- In an infinite sequence of coin tosses, anything that can happen, does happen!

Exercise 4.1 [Revision]

1. Let Ω be a sample space, and let $A_1, A_2, \ldots$ be a partition of Ω with the property that $\mathbb{P}(A_i) > 0$ for all i.
   (a) Show that $\mathbb{P}(B) = \displaystyle\sum_{i=1}^{\infty} \mathbb{P}(B|A_i)\mathbb{P}(A_i)$.

   Answer: Bookwork: this is the *partition theorem*, also known as the *law of total probability.*

   (b) Show that $\mathbb{P}(A_i|B) = \dfrac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{\infty} \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$.

   Answer: Bookwork: this is *Bayes' formula.*

Exercise 4.2

1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, and consider the function $\mathbb{Q} : \mathcal{F} \to [0, 1]$ defined by $\mathbb{Q}(A) = \mathbb{P}(A|B)$.
   (a) Show that $(\Omega, \mathcal{F}, \mathbb{Q})$ is a probability space.

   Answer:
   - $\mathbb{Q}(\Omega) = \mathbb{P}(\Omega|B) = 1$.
   - Let $\{A_i\}_{i=1}^{\infty}$ be a countable collection of pairwise disjoint events in $\mathcal{F}$.
     Since $\mathcal{F}$ is a σ-field, $\{A_i \cap B\}_{i=1}^{\infty}$ is also a countable collection of pairwise disjoint events in $\mathcal{F}$. Hence

$$\mathbb{Q}(\cup_i A_i) = \frac{\mathbb{P}\big[(\cup_i A_i) \cap B\big]}{\mathbb{P}(B)} = \frac{\mathbb{P}\big[\cup_i (A_i \cap B)\big]}{\mathbb{P}(B)} = \frac{\sum_i \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_i \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_i \mathbb{Q}(A_i).$$

   (b) If $C \in \mathcal{F}$ and $\mathbb{Q}(C) > 0$, show that $\mathbb{Q}(A|C) = \mathbb{P}(A|B \cap C)$.

   Answer: Since $\mathbb{Q}$ is a probability measure,

$$\mathbb{Q}(A|C) = \frac{\mathbb{Q}(A \cap C)}{\mathbb{Q}(C)} = \frac{\mathbb{P}(A \cap C|B)}{\mathbb{P}(C|B)} = \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(B \cap C)} = \mathbb{P}(A|B \cap C).$$

   This shows that the order in which we impose the conditions B and C does not matter.

2. A random number N of dice are rolled. Let A_k be the event that $N = k$, and suppose that $\mathbb{P}(A_k) = 2^{-k}$ for $k \in \{1, 2, \ldots\}$ (and zero otherwise). Let S be the sum of the scores shown on the dice. Find the probability that:
   (a) $N = 2$ given that $S = 4$,

   Answer:

$$\begin{aligned}
\mathbb{P}(N = 2|S = 4) &= \frac{\mathbb{P}(\{N = 2\} \cap \{S = 4\})}{\mathbb{P}(\{S = 4\})} \\
&= \frac{\mathbb{P}(S = 4|N = 2)\mathbb{P}(N = 2)}{\sum_{k=1}^{n} \mathbb{P}(S = 4|N = k)\mathbb{P}(N = k)} \\
&= \frac{1/12 \times 1/4}{(1/6 \times 1/2) + (1/12 \times 1/4) + (3/6^3 \times 1/8) + (1/6^4 \times 1/16)} \\
&=
\end{aligned}$$

(b) $S = 4$ given that N is even,

Answer:

$$\begin{aligned}\mathbb{P}(S=4|N \text{ even}) &= \frac{\mathbb{P}(S=4|N=2)\times(1/4)+\mathbb{P}(S=4|N=4)\times(1/16)}{\mathbb{P}(N \text{ even})} \\ &= \frac{(1/12\times 1/4)+(1/1296\times 1/16)}{1/4+1/16+1/64+\cdots} \\ &= \end{aligned}$$

(c) $N = 2$ given that $S = 4$ and the first die shows 1,

Answer: Let D be the score on the first die.

$$\begin{aligned}\mathbb{P}(N=2|S=2,D=1) &= \frac{\mathbb{P}(N=2,S=4,D=1}{\mathbb{P}(S=4,D=1)} \\ &= \frac{1/6\times 1/6\times 1/4}{(1/6\times 1/6\times 1/4)+(1/6\times 2/36\times 1/8)+(1/6^4\times 1/16)} \\ &= \end{aligned}$$

(d) the largest number shown by any dice is r (where S is unknown).

Answer: Let M be the maximum number shown on the dice. For $r \in \{1,2,3,4,5,6\}$,

$$\begin{aligned}\mathbb{P}(M\leq r) &= \sum_{k=1}^{\infty}\mathbb{P}(M\leq r|N=k)\frac{1}{2^k} \\ &= \sum_{k=1}^{\infty}\left(\frac{r}{6}\right)^k\frac{1}{2^k} \\ &= \frac{r}{12}\left(1-\frac{r}{12}\right)^{-1} \\ &= \frac{r}{12-r}.\end{aligned}$$

3. Let $\Omega = \{1,2,\ldots,p\}$ where p is a prime number. Let $\mathcal{F}$ be the power set of Ω, and let $\mathbb{P} : \mathcal{F} \to [0,1]$ be the probability measure on $(\Omega,\mathcal{F})$ defined by $\mathbb{P}(A) = |A|/p$, where $|A|$ denotes the cardinality of A. Show that if A and B are independent events, then at least one of A and B is either $\emptyset$ or Ω.

Answer: Let A and B be independent events with $|A| = a$, $|B| = b$ and $|A\cap B| = c$.

- By independence, $\mathbb{P}(A\cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
- This means that $(a/p)(b/p) = (c/p)$ and therefore $ab = pc$.
- If $ab \neq 0$, then p divides ab.
- Since p is prime, either p divides a, or p divides b (by the fundamental theorem of arithmetic).
- Hence $a = p$ or $b = p$ (or both).
- Thus follows that $A = \Omega$ or $B = \Omega$ (or both).

Exercise 5.1

1. Let Ω be the sample space of some random experiment, and let $\mathcal{F}$ be a σ-field over Ω.
   (a) For any $A \in \mathcal{F}$, show that the function $X : \Omega \to \mathbb{R}$, defined by

$$X(\omega) = \begin{cases}1 & \text{if } \omega\in A,\\ 0 & \text{if } \omega\notin A,\end{cases}$$

is a random variable on $(\Omega, \mathcal{F})$.

Answer: For any $B \in \mathcal{B}$,
- if $1 \in B$, then $\{\omega : X(\omega) \in B\} = A$, which is contained in $\mathcal{F}$;
- if $1 \notin B$, then $\{\omega : X(\omega) \in B\} = \emptyset$, which is also contained in $\mathcal{F}$.

(b) Let $A_1, A_2, \ldots, A_n \in \mathcal{F}$ be a partition of Ω and let $a_1, a_2, \ldots, a_n \in \mathbb{R}$. Show that the function $X : \Omega \to \mathbb{R}$, defined by

$$X(\omega) = \sum_{i=1}^{n} a_i I_{A_i}(\omega) \qquad \text{where} \qquad I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A, \end{cases}$$

is a random variable on $(\Omega, \mathcal{F})$.

Answer: For any $B \in \mathcal{B}$,

$$\{\omega : X(\omega) \in B\} = \cup\{A_i : a_i \in B\} \in \mathcal{F},$$

because $\mathcal{F}$ is closed under finite unions.

Exercise 6.1

1. Let F and G be CDFs, and let $0 < \lambda < 1$ be a constant. Show that $H = \lambda F + (1-\lambda)G$ is also a CDF.

Answer: Let $H(x) = \lambda F(x) + (1-\lambda)G(x)$. It is easy to show that H has the following properties:
- if $x < y$ then $H(x) \leq H(y)$,
- $H(x) \to 0$ as $x \to -\infty$,
- $H(x) \to 1$ as $x \to +\infty$, and
- $H(x + \epsilon) \to H(x)$ as $\epsilon \downarrow 0$.

Thus H is a distribution function.

2. Let X_1 and X_2 be the numbers observed in two independent rolls of a fair die. Find the PMF of each of the following random variables:

(a) $Y = 7 - X_1$,

Answer: $P(Y = k) = 1/6$ for $k = 1, \ldots, 6$.

(b) $U = \max(X_1, X_2)$,

Answer: Let $U = \max\{X_1, X_2\}$. Then since $\{X_1 \leq k\}$ and $\{X_2 \leq k\}$ are independent events,

$$\begin{aligned} P(U \leq k) &= P(X_1 \leq k \text{ and } X_2 \leq k) \\ &= P(X_1 \leq k)P(X_2 \leq k) \\ &= (k/6) \cdot (k/6) = k^2/36 \end{aligned}$$

Thus,

$$P(U = k) = P(U \leq k) - P(U \leq k-1) = \frac{(k^2 - (k-1)^2)}{36} = \frac{(2k-1)}{36}$$

(c) $V = X_1 - X_2$.

Answer: The values of $V = X_1 - X_2$ at each point of the sample space $\Omega = \{(i,j) : 1 \le i, j \le 6\}$ are

		j = 1	2	3	4	5	6
	1	0	1	2	3	4	5
	2	-1	0	1	2	3	4
i	3	-2	-1	0	1	2	3
	4	-3	-2	-1	0	1	2
	5	-4	-3	-2	-1	0	1
	6	-5	-4	-3	-2	-1	0

The required probabilities are obtained by counting the number of outcomes that give the same value of $V = X_1 - X_2$:

v	-5	-4	-3	-2	-1	0	1	2	3	4	5
$P(V = v)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

(d) $W = |X_1 - X_2|$.

Answer:

w	0	1	2	3	4	5
$P(W = w)$	6/36	10/36	8/36	6/36	4/36	2/36

3. The PDF of a continuous random variable X is given by $f(x) = \begin{cases} cx^2 & 1 \le x \le 2, \\ 0 & \text{otherwise.} \end{cases}$

(a) Find the value of the constant c, and sketch the PDF of X.

Answer: The PDF must integrate to 1:

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_1^2 cx^2\,dx = \left[\frac{cx^3}{3}\right]_1^2 = \frac{7c}{3} = 1$$

so $c = 3/7$. (The sketch is a quadratic curve between $x = 1$ and $x = 2$.)

(b) Find the value of $P(X > 3/2)$.

Answer:

$$P(X > 3/2) = \int_{3/2}^2 \frac{3x^2}{7}\,dx = \left[\frac{x^3}{7}\right]_{3/2}^2 = \frac{37}{56}$$

(c) Find the CDF of X.

Answer: For $1 \le x \le 2$,

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_1^x \frac{3x^2}{7}\,dx = \left[\frac{x^3}{7}\right]_1^x = \frac{x^3 - 1}{7}$$

so the CDF of X is

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{7}(x^3 - 1) & 1 \le x < 2 \\ 1 & x \ge 2 \end{cases}$$

4. The PDF of a continuous random variable X is given by $f(x) = \begin{cases} cx^{-d} & \text{for } x > 1, \\ 0 & \text{otherwise.} \end{cases}$

(a) Find the range of values of d for which $f(x)$ is a probability density function.

Answer: The function $f(x) = cx^{-d}$ is only integrable if $d > 1$, in which case

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{1}^{\infty} \frac{c}{x^d}\,dx = \left[\frac{-c}{(d-1)x^{d-1}}\right]_1^{\infty} = \frac{c}{d-1}$$

(b) If $f(x)$ is a density function, find the value of c, and the corresponding CDF.

Answer: If $f(x)$ is a probability density function, we require that $\int_{-\infty}^{\infty} f(x)\,dx = 1$, so we must have that $c = d-1$. The corresponding distribution function is

$$F(x) = \int_{-\infty}^{x} f(u)\,du = \int_{1}^{\infty} \frac{d-1}{u^d}\,du = \left[\frac{-1}{x^{d-1}}\right]_1^{x} = 1 - \frac{1}{x^{d-1}}$$

for $x > 1$, and zero otherwise.

5. Let $f(x) = \dfrac{ce^x}{(1+e^x)^2}$ be a PDF, where c is a constant. Find the value of c, and the corresponding CDF.

Answer: By inspection, $f(x) = F'(x)$ where $F(x) = \frac{ce^x}{1+e^x}$. Writing this as $F(x) = \frac{c}{e^{-x}+1}$ it is easy to see that $F(x) \to c$ as $x \to \infty$, so we must have that $c = 1$.

6. Let $X_1, X_2, \ldots$ be independent and identically distributed observations, and let F denote their common CDF. If F is unknown, describe and justify a way of estimating F, based on the observations. [Hint: consider the indicator variables of the events $\{X_j \leq x\}$.]

Answer: Let X be a random variable with same CDF, and let $I_j(x)$ the indicator variable of the event $\{X_j \leq x\}$. Then

$$\mathbb{P}(X \leq x) \approx \frac{1}{n}\sum_{j=1}^{n} I_j(x).$$

The RHS yields the proportion of observations that are at most equal to x.

Exercise 7.1

1. Let X be a discrete random variable, with PMF $f_X(-2) = 1/3$, $f_X(0) = 1/3$, $f_X(2) = 1/3$, and zero otherwise. Find the distribution of $Y = X + 3$.

Answer: The function $g(x) = x + 3$ is injective, with $g^{-1}(y) = y - 3$, so

$$f_Y(1) = 1/3, \quad f_Y(3) = 1/3, \quad f_Y(5) = 1/3.$$

Note that $\text{supp}(f_X) = \{-2, 0, 2\}$, and $\text{supp}(f_Y) = \{g(x) : x \in \text{supp}(f_X)\} = \{1, 3, 5\}$.

2. Let $X \sim \text{Binomial}(n, p)$ and define $g(x) = n - x$. Show that $g(X) \sim \text{Binomial}(n, 1-p)$.

Answer: $g(x) = n - x$ is a decreasing function on $[0, n]$: its (unique) inverse is $g^{-1}(y) = n - y$. By Theoem **??**, the PMF of $Y = g(X)$ is

$$f_Y(y) = f_X\left[g^{-1}(y)\right] = f_X(n-y) = \binom{n}{n-y} p^{n-y}(1-p)^{n-(n-y)} = \binom{n}{y}(1-p)^y(1-(1-p))^{n-y},$$

which is the PMF of the Binomial$(n, 1-p)$ distribution.

3. Let X be a random variable, and let F_X denote its CDF. Find the CDF of $Y = X^2$ in terms of F_X.

Answer:

$$\begin{aligned} F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(X^2 \le y) \\ &= \mathbb{P}(-\sqrt{y} \le X \le \sqrt{y}) \\ &= \mathbb{P}(X \le \sqrt{y}) - \mathbb{P}(X < -\sqrt{y}) \\ &= \begin{cases} F_X(\sqrt{y}) - F_X(-\sqrt{y}) & y > 0, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

4. Let X be a random variable with the following CDF:

$$F_X(x) = \begin{cases} 1 - \dfrac{1}{x^3} & \text{for } x \ge 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find the CDF of the random variable $Y = 1/X$, and describe how a pseudo-random sample from the distribution of Y can be obtained using an algorithm that generates uniformly distributed pseudo-random numbers in the range $[0, 1]$.

Answer: Let $g(x) = 1/x$ denote the transformation.

- $\text{supp}(f_X) = [1, \infty] \Rightarrow \text{supp}(f_Y) = [0, 1]$.
- The inverse transformation: $g^{-1}(y) = 1/y$.

Because $g(x)$ is a decreasing function over $\text{supp}(f_X)$,

$$F_Y(y) = 1 - F_X\left[g^{-1}(y)\right] = 1 - F_X\left(\frac{1}{y}\right) = \begin{cases} 0 & y < 0 \\ y^3 & 0 \le y \le 1 \\ 1 & y > 1. \end{cases}$$

To find a pseudo-random sample from the distribution of Y, we use the fact that $F_Y(Y) \sim \text{Uniform}(0, 1)$. Let $u = F_Y(y)$. Then

$$y = F_Y^{-1}(u) = u^{1/3}.$$

The required sample is obtained by generating a pseudo-random sample $u_1, u_2, \ldots, u_n$ from the Uniform$(0, 1)$ distribution, then computing

$$y_i = u_i^{1/3} \quad \text{for } i = 1, 2, \ldots, n.$$

Exercise 8.1

1. Let $X \sim \text{Uniform}(-1, 1)$. Find the CDF and PDF of X^2.

Answer: The PDF of X is

$$f_X(x) = \begin{cases} 1/2 & -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

For $x \in [-1, 1]$,

$$\mathbb{P}(X \le x) = \int_{-\infty}^{x} f_X(t)\, dt = \int_{-1}^{x} \frac{1}{2}\, dt = \left[\frac{t}{2}\right]_{-1}^{x} = \frac{1}{2}(x + 1).$$

The CDF of X is:

$$F(x) = \begin{cases} 0 & x < -1, \\ \frac{1}{2}(x+1) & -1 \leq x \leq 1, \\ 1 & x > 1. \end{cases}$$

Let $Y = X^2$. For $0 \leq y \leq 1$ we have

$$\begin{aligned} \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) &= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \mathbb{P}(X \leq \sqrt{y}) - \mathbb{P}(X \leq -\sqrt{y}) \\ &= \sqrt{y}. \end{aligned}$$

Hence the CDF of Y is

$$F_Y(y) = \begin{cases} 0 & y < 0, \\ \sqrt{y} & 0 \leq y \leq 1, \\ 1 & y > 1. \end{cases}$$

and the PDF of Y is

$$f_Y(y) = \begin{cases} \frac{1}{2}y^{-1/2} & 0 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

2. Let X have exponential distribution with rate parameter $\lambda > 0$. The PDF of X is

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{for } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Find the PDFs of $Y = X^2$ and $Z = e^X$.

Answer:

(1) The transformation $g(x) = x^2$ is monotonic increasing over $[0, \infty)$; its inverse function is

$$g^{-1}(y) = \sqrt{y}, \quad \text{which has first derivative} \quad \frac{d}{dy}g^{-1}(y) = \frac{1}{2\sqrt{y}}.$$

Since $\operatorname{supp}(f_X) = [0, \infty)$ it follows immediately that $\operatorname{supp}(f_Y) = [0, \infty)$.
For $y > 0$,

$$f_Y(y) = f_X\left[g^{-1}(y)\right]\left|\frac{d}{dy}g^{-1}(y)\right| = \lambda\exp(\lambda\sqrt{y})\left|\frac{1}{2\sqrt{y}}\right| = \frac{\lambda}{2\sqrt{y}}\exp(-\lambda\sqrt{y}).$$

Hence the PDF of $Y = X^2$ is given by

$$f_Y(y) = \begin{cases} \dfrac{\lambda}{2\sqrt{y}}\exp(-\lambda\sqrt{y}) & y \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

(2) The transformation $g(x) = e^x$ is a monotonic increasing function over $[0, \infty)$; its inverse function is

$$g^{-1}(z) = \log y \quad \text{and} \quad \frac{d}{dy}g^{-1}(z) = \frac{1}{z}.$$

Since $\operatorname{supp}(f_X) = [0, \infty)$ it follows immediately that $\operatorname{supp}(f_Z) = [1, \infty)$.
For $z \geq 1$,

$$f_Z(z) = f_X\left[g^{-1}(z)\right]\left|\frac{d}{dz}g^{-1}(z)\right| = \lambda\exp(-\lambda\log z)\left|\frac{1}{z}\right| = \lambda z^{-(\lambda+1)}.$$

Hence the PDF of $Z = e^X$ is given by

$$f_Z(z) = \begin{cases} \lambda z^{-(\lambda+1)} & z \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

3. Let $X \sim \text{Pareto}(1,2)$. Find the PDF of $Y = 1/X$.

Answer: $X \sim \text{Pareto}(1,2)$ has PDF

$$f_X(x) = \begin{cases} \dfrac{2}{x^3} & x > 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $g(x) = 1/x$.

- $g(x)$ is monotonic decreasing over $x > 1$; the inverse transformation is $g^{-1}(y) = 1/y$.
- $\text{supp}(f_Y) = \{x^{-1} : x > 1\} = (0,1)$.

Hence the PDF of Y is given by

$$f_Y(y) = f_X\left[g^{-1}(y)\right]\left|\frac{d}{dy}g^{-1}(y)\right| = f_X\left(\frac{1}{y}\right)\left|-\frac{1}{y^2}\right| = \begin{cases} 2y & \text{for } 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

4. A continuous random variable U has PDF

$$f(u) = \begin{cases} 12u^2(1-u) & \text{for} \quad 0 < u < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find the PDF of $V = (1-U)^2$.

Answer:

- The transformation $g(u) = (1-u)^2$ is monotonic decreasing over $[0,1]$.
- The inverse transformation is $g^{-1}(v) = 1 - v^{1/2}$, for which $\dfrac{d}{dv}g^{-1}(v) = -\dfrac{1}{2v^{1/2}}$.
- Since $\text{supp}(f_U) = (0,1)$ it follows that $\text{supp}(f_V) = (0,1)$.

Hence for $0 < v < 1$ the PDF of V is

$$\begin{aligned} f_V(v) &= f_U\left[g^{-1}(v)\right]\left|\frac{d}{dv}g^{-1}(v)\right| \\ &= 12(1-v^{1/2})^2 v^{1/2}\left|-\frac{1}{2v^{1/2}}\right| \\ &= 6(1-v^{1/2})^2, \end{aligned}$$

and zero otherwise.

5. The continuous random variable U has PDF

$$f_U(u) = \begin{cases} 1+u & -1 < u \leq 0, \\ 1-u & 0 < u \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find the PDF of $V = U^2$. (Note that the transformation is not injective over $\text{supp}(f_U)$, so you should first compute the CDF of V, then derive its PDF by differentiation.)

Answer: Let $g(u) = u^2$. This is not injective over $\text{supp}(f_U) = (-1,1)$, and does not therefore have a unique inverse over this interval. Instead we will compute the CDF of V, then obtain the PDF by differentiation.

For $0 < v < 1$,

$$\begin{aligned}
F_V(v) = P(V \leq v) &= P(U^2 \leq v) \\
&= P(-\sqrt{v} \leq U \leq \sqrt{v}) \\
&= \int_{-\sqrt{v}}^{+\sqrt{v}} f_U(u)\,du \\
&= \int_{-\sqrt{v}}^{0} (1+u)\,du + \int_{0}^{+\sqrt{v}} (1-u)\,du \\
&= \left[u + \frac{u^2}{2}\right]_{-\sqrt{v}}^{0} + \left[u - \frac{u^2}{2}\right]_{0}^{\sqrt{v}} \\
&= \sqrt{v} - \frac{v}{2} + \sqrt{v} - \frac{v}{2} \\
&= 2\sqrt{v} - v.
\end{aligned}$$

The CDF is therefore

$$F_V(u) = \begin{cases} 0 & v \leq 0, \\ 2\sqrt{v} - v & 0 < v < 1, \\ 1 & v \geq 1. \end{cases}$$

The PDF is then found by differentiation with respect to v:

$$f_V(u) = \begin{cases} v^{-1/2} - 1 & \text{for } 0 \leq v < 1, \\ 0 & \text{otherwise.} \end{cases}$$

6. Let X have exponential distribution with scale parameter $\theta > 0$. This has PDF

$$f(x) = \begin{cases} \frac{1}{\theta}\exp(-x/\theta) & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Find the PDF of $Y = X^{1/\gamma}$ where $\gamma > 0$.

Answer: Let $g(x) = x^{1/\gamma}$.

- g a monotonic increasing function over $\text{supp}(f_X) = \{x : x > 0\}$, so its inverse exists:
- The inverse transformation is $g^{-1}(y) = y^{\gamma}$, for which $\frac{d}{dy}g^{-1}(y) = \gamma y^{\gamma-1}$.
- $\text{supp}(f_X) = \{x : x > 0\}$ means that $\text{supp}(f_Y) = \{y : y > 0\}$.

Since $f_Y(y) = f_X\left[g^{-1}(y)\right]\left|\frac{d}{dy}g^{-1}(y)\right|$, we obtain

$$f_Y(y) = \begin{cases} (\gamma/\theta)y^{\gamma-1}\exp(-y^{\gamma}/\theta) & \text{for } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

This is called the *Weibull* distribution (with scale parameter θ and shape parameter γ).

7. Suppose that X has the *Beta Type I* distribution, with parameters $\alpha, \beta > 0$. This has PDF

$$f_X(x) = \begin{cases} \dfrac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{for } 0 \leq x \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\,dt$ is the so-called *beta function*. Show that the random variable $Y = \dfrac{X}{1-X}$ has the *Beta Type II* distribution, which has PDF

$$f_Y(y) = \begin{cases} \dfrac{1}{B(\alpha,\beta)} \dfrac{y^{\alpha-1}}{(1+y)^{\alpha+\beta}} & \text{for } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Answer: Let $g(x) = x/(1-x)$

- $g(x)$ is monotonic increasing on $\text{supp}(f_X) = [0,1]$.
- The inverse transformation is $g^{-1}(y) = \dfrac{y}{1+y}$, which has derivative $\dfrac{d}{dy}g^{-1}(y) = \dfrac{1}{(1+y)^2}$.
- Since $\text{supp}(f_X) = [0,1]$, we see that $\text{supp}(f_Y) = [0,\infty)$.

Thus for $y > 0$, the PDF of Y is

$$\begin{aligned} f_Y(y) &= f_X\left[g^{-1}(y)\right]\left|\frac{d}{dy}g^{-1}(y)\right| \\ &= \frac{1}{B(\alpha,\beta)}\left(\frac{y}{1+y}\right)^{\alpha-1}\left(\frac{1}{1+y}\right)^{\beta-1}\left|\frac{1}{(1+y)^2}\right| \\ &= \frac{1}{B(\alpha,\beta)}\frac{y^{\alpha-1}}{(1+y)^{\alpha+\beta}}, \end{aligned}$$

and zero otherwise.

Exercise 10.1

1. Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, and let $0 \leq X_1 \leq X_2 \leq \ldots$ be an increasing sequence of non-negative random variables over $(\Omega,\mathcal{F})$ such that $X_n(\omega) \uparrow X(\omega)$ ans $n \to \infty$ for all $\omega \in \Omega$. Show that X is a random variable on $(\Omega,\mathcal{F})$.

Answer: Let $x \in \mathbb{R}$. Since the X_n are random variables, we have (by definition) that $\{X_n \leq x\} \in \mathcal{F}$ for every $n \in \mathbb{N}$. Since $\mathcal{F}$ is closed under countable intersections,

$$\{X \leq x\} = \bigcap_{n=1}^{\infty}\{X_n \leq x\} \in \mathcal{F}$$

so X is a random variable.

2. Let X be an integrable random variable. Show that $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.

Answer: Since $|X| = X^+ + X^-$, by the triangle inequality

$$|\mathbb{E}(X)| = |\mathbb{E}(X^+) - \mathbb{E}(X^-)| \leq \mathbb{E}(X^+) + \mathbb{E}(X^-) = \mathbb{E}(|X|),$$

3. If $X \leq Y$ then $X^+ \leq Y^+$ and $X^- \geq Y^-$ so

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-) \leq \mathbb{E}(Y^+) - \mathbb{E}(Y^-) = \mathbb{E}(Y),$$

4. Let X and Y be integrable random variables. Show that $aX + bY$ is integrable.

Answer: To show that $aX + bY$ is integrable, first we have by the triangle inequality that

$$|aX + bY| \leq |a||X| + |b||Y|.$$

By the linearity and monotonicity of expectation for non-negative random variables,

$$\mathbb{E}(|aX + bY|) \leq |a|\mathbb{E}(|X|) + |b|\mathbb{E}(|Y|)$$

and since $\mathbb{E}(|X|) < \infty$ and $\mathbb{E}(|Y|) < \infty$, it follows that $\mathbb{E}(|aX + bY|) < \infty$, so $aX + bY$ is integrable.

Exercise 11.1

1. Let X be the score on a fair die, and let $g(x) = 3x - x^2$. Find the expected value and variance of the random variable $Y = g(X)$.

Answer: The expectation of $Y = 3X - X^2$ is determined by the distribution of X,

$$\begin{aligned}\mathbb{E}(Y) &= \sum_{x=1}^{6} y(x)f(x) = \sum_{x=1}^{6}(3x - x^2) \times \frac{1}{6} \\ &= \frac{1}{6}\left(3\sum_{x=1}^{6} x - \sum_{x=1}^{6} x^2\right) = \frac{-14}{3}\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}(Y^2) &= \sum_{x=1}^{6} y^2(x)f(x) = \sum_{x=1}^{6}(3x - x^2)^2 \times \frac{1}{6} \\ &= \frac{1}{6}\left(9\sum_{x=1}^{6} x^2 - 6\sum_{x=1}^{6} x^3 + \sum_{x=1}^{6} x^4\right) = \frac{448}{6}\end{aligned}$$

Hence

$$\text{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = \frac{448}{6} - \frac{196}{9} = \frac{476}{9}$$

2. A long line of athletes $k = 0, 1, 2, \ldots$ make throws of a javelin to distances $X_0, X_1, X_2, \ldots$ respectively. The distances are independent and identically distributed random variables, and the probability that any two throws are exactly the same distance is equal to zero. Let Y be the index of the first athlete in the sequence who throws further than distance X_0. Show that the expected value of Y is infinite.

Answer: Y is a discrete random variable, taking values in the set $\{1, 2, \ldots\}$.

- The event $\{Y > k\}$ means that out of the first $k+1$ throws, the initial throw was the furthest.

Because the distances $X_0, X_1, \ldots, X_k$ are identically distributed, it follows that

$$\mathbb{P}(Y > k) = \frac{1}{k+1}.$$

Thus,

$$\mathbb{P}(Y = k) = \mathbb{P}(Y > k-1) - \mathbb{P}(Y > k) = \frac{1}{k} - \frac{1}{k+1} = \frac{1}{k(k+1)}$$

so

$$\mathbb{E}(Y) = \sum_{n=0}^{\infty} k\mathbb{P}(Y = k) = \sum_{n=0}^{\infty} \frac{1}{k+1} = \sum_{n=1}^{\infty} \frac{1}{k} = \infty.$$

3. Consider the following game. A random number X is chosen uniformly from $[0, 1]$, then a sequence $Y_1, Y_2, \ldots$ of random numbers are chosen independently and uniformly from $[0, 1]$. Let Y_n be the first number in the sequence for which $Y_n > X$. When this occurs, the game ends and the player is paid $(n-1)$ pounds. Show that the expected win is infinite.

Answer: Let Z be the amount won.

$$\begin{aligned}\mathbb{P}(Z = k|X = x) &= \mathbb{P}(Y_1 \leq x, Y_2 \leq x, \ldots, Y_k \leq x, Y_{k+1} > x) \\ &= \mathbb{P}(Y_1 \leq x)\mathbb{P}(Y_2 \leq x)\ldots\mathbb{P}(Y_k \leq x)\mathbb{P}(Y_{k+1} > x) \qquad \text{(by independence)} \\ &= x^k(1-x)\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{P}(Z=k) &= \int_0^1 x^k(1-x)\,dx \\ &= \left[\frac{1}{k+1}x^{k+1} - \frac{1}{k+2}x^{k+2}\right]_0^1 \\ &= \frac{1}{k+1} - \frac{1}{k+2} \\ &= \frac{1}{(k+1)(k+2)}\end{aligned}$$

Thus,

$$\mathbb{E}(Z) = \sum_{k=0}^{\infty} k\left(\frac{1}{(k+1)(k+2)}\right) = \infty.$$

4. Let X be a discrete random variable with PMF

$$f(k) = \begin{cases} \dfrac{3}{\pi^2k^2} & \text{if } \ k \in \{\pm1, \pm2, \ldots\} \\ 0 & \text{otherwise.}\end{cases}$$

Show that $\mathbb{E}(X)$ is undefined.

Answer: Let $X = X^+ - X^-$ where

$$\begin{aligned}X^+ &= \max\{X, 0\} = \begin{cases} X & \text{if } X \geq 0, \\ 0 & \text{otherwise.}\end{cases} \\ X^- &= \max\{-X, 0\} = \begin{cases} -X & \text{if } X < 0, \\ 0 & \text{otherwise.}\end{cases}\end{aligned}$$

Then

$$\begin{aligned}\mathbb{E}(X^+) &= \sum_{k=1}^{\infty} k\left(\frac{3}{\pi^2k^2}\right) && = \frac{3}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{k} = \infty \\ \mathbb{E}(X^-) &= \sum_{k=-\infty}^{-1} (-k)\left(\frac{3}{\pi^2k^2}\right) && = \frac{3}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{k} = \infty\end{aligned}$$

so $\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-)$ is undefined.

5. Let X be a continuous random variable having the Cauchy distribution, defined by the PDF

$$f(x) = \frac{1}{\pi(1+x^2)} \qquad x \in \mathbb{R}$$

Show that $\mathbb{E}(X)$ is undefined.

Answer: The expectation of X is

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(X_+) - \mathbb{E}(X_-) \\ &= \int_0^{\infty} xf(x)\,dx - \int_{-\infty}^{0} (-x)f(x)\,dx \\ &= \int_0^{\infty} \frac{x}{\pi(1+x^2)}\,dx - \int_0^{\infty} \frac{x}{\pi(1+x^2)}\,dx\end{aligned}$$

If $x > 1$ then $x^2 > 1$ and therefore $2x^2 > 1 + x^2$, so

$$\frac{x}{1+x^2} > \frac{1}{2x} \qquad \text{for all } x > 1$$

Consequently,

$$\int_0^\infty \frac{x}{1+x^2}\,dx > \int_1^\infty \frac{x}{1+x^2}\,dx > \frac{1}{2}\int_1^\infty \frac{1}{x}\,dx = \infty$$

Thus X is not integrable:

$$\mathbb{E}(|X|) = \mathbb{E}(X_+) + \mathbb{E}(X_-) = 2\int_0^\infty \frac{x}{\pi(1+x^2)}\,dx = \infty$$

and $\mathbb{E}(X)$ is not defined.

6. A coin is tossed until the first time a head is observed. If this occurs on the nth toss and n is odd, you win $2^n/n$ pounds, but if n is even then you lose $2^n/n$ pounds. Show that the expected win is undefined.

Answer: Let X represent the amount won. $\mathbb{P}(\text{First head occurs on } n\text{th toss}) = 1/2^n$, so

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{n=1}^\infty \left(\frac{(-1)^{n+1}2^n}{n} \times \frac{1}{2^n}\right) \\
&= \sum_{n=1}^\infty (-1)^{n+1}\frac{1}{n} \\
&= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots
\end{aligned}$$

This is the alternating harmonic series, which is not absolutely convergent. Hence the expected win is undefinned.

Remark. It is known that the alternating harmonic series is convergent:

$$\sum_{n=0}^\infty \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots = \log 2$$

However, the series is not absolutely convergent, because

$$\sum_{n=0}^\infty \left|\frac{(-1)^{n+1}}{n}\right| = \sum_{n=0}^\infty \frac{1}{n} = \infty$$

The Riemann rearrangement theorem says that if a series is convergent but not absolutely convergent, then its limit depends on the order in which its terms are added. For example

$$\begin{aligned}
\log 2 &= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \dots \\
&= 1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{6} - \frac{1}{8} + \frac{1}{5} - \frac{1}{10} + \dots \\
&= \left(1 - \frac{1}{2}\right) - \frac{1}{4} + \left(\frac{1}{3} - \frac{1}{6}\right) - \frac{1}{8} + \left(\frac{1}{5} - \frac{1}{10}\right) + \dots \\
&= 1 - \frac{1}{4} + \frac{1}{6} - \frac{1}{8} + \frac{1}{10} - \dots \\
&= \frac{1}{2}\left(1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \dots\right) \\
&= \frac{1}{2}\log 2
\end{aligned}$$

which is absurd, since $\log 2 \neq 0$. The expectation $\mathbb{E}(X) = \sum_x g(x)f(x)$ of a discrete random variable cannot be sensibly defined unless the series $\sum_x g(x)f(x)$ is absolutely convergent.

7. Let X be a continuous random variable with uniform density on the interval $[-1,1]$,

$$f(x) = \begin{cases} \frac{1}{2} & \text{if } x \in [-1,+1] \\ 0 & \text{otherwise.} \end{cases}$$

Compute $\mathbb{E}(X)$, $\mathbb{E}(X^2)$, $\mathbb{E}(X^3)$, $\mathbb{E}(1/X)$ and $\mathbb{E}(1/X^2)$.

Answer: Let F be the CDF of X, let $g:\mathbb{R}\to\mathbb{R}$, and recall the following:

- If $g(X)$ is non-negative random variable, its expectation with respect to F is

$$\mathbb{E}\big[g(X)\big] = \int_{-\infty}^{\infty} g(x)f(x)\,dx$$

(For non-negative random variables, we can accept that its expectation is infinite.)

- If $g(X)$ is a signed random variable, its expectation with respect to F is only defined if

$$\int_{-\infty}^{\infty} |g(x)|f(x)\,dx < \infty.$$

If this condition holds, the expectation is given by

$$\mathbb{E}\big[g(X)\big] = \int_{-\infty}^{\infty} g^+(x)f(x)\,dx - \int_{-\infty}^{\infty} g^-(x)f(x)\,dx$$

where $g^+(x)$ and $g^-(x)$ are respectively is the positive and negative parts of $g(x)$:

$$g^+(x) = \begin{cases} g(x) & \text{if} \quad g(x)\geq 0, \\ 0 & \text{if} \quad g(x) < 0, \end{cases} \qquad \text{and} \qquad g^-(x) = \begin{cases} 0 & \text{if} \quad g(x)\geq 0, \\ -g(x) & \text{if} \quad g(x) < 0. \end{cases}$$

(1) $g(x) = x$. In this case, $g(x)$ is a signed function. Since

$$|g(x)| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0, \end{cases}$$

we see that the expectation exists:

$$\int_{-\infty}^{\infty} |g(x)|f(x)\,dx = \frac{1}{2}\int_{-1}^{0}(-x)\,dx + \frac{1}{2}\int_0^1 x\,dx = \int_0^1 x\,dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2} < \infty.$$

The positive and negative parts of g are

$$g^+(x) = \begin{cases} x & \text{if} \quad x\geq 0, \\ 0 & \text{if} \quad x < 0, \end{cases} \qquad \text{and} \qquad g^-(x) = \begin{cases} 0 & \text{if} \quad x\geq 0, \\ -x & \text{if} \quad x < 0. \end{cases}$$

Thus we have

$$\begin{aligned}
\mathbb{E}(X) = \mathbb{E}\big[g(X)\big] &= \int_{-\infty}^{\infty} g^+(x)f(x)\,dx - \int_{-\infty}^{\infty} g^-(x)f(x)\,dx \\
&= \frac{1}{2}\int_0^1 x\,dx - \frac{1}{2}\int_{-1}^{0}(-x)\,dx \\
&= \frac{1}{2}\left[\frac{x^2}{2}\right]_0^1 - \frac{1}{2}\left[\frac{-x^2}{2}\right]_{-1}^{0} \\
&= \left(\frac{1}{4} - 0\right) - \left(0 + \frac{1}{4}\right) = 0.
\end{aligned}$$

Note that, if we regard an integral as the "area between a curve and the x-axis", the positive part gives the area above the x-axis (which has a positive sign), and the negative part gives the area below the x-axis (which has a negative sign): the integral is zero because these two areas are of equal magnitude.

(2) $g(x)=x^2$. In this case, $g(x)$ is a non-negative function, so

$$\mathbb{E}(X^2)=\mathbb{E}\big[g(X)\big]=\int_{-\infty}^{\infty} g(x)f(x)\,dx=\frac{1}{2}\int_{-1}^{1}\frac{x^2}{2}\,dx=\int_0^1 x^2\,dx=\left[\frac{x^3}{3}\right]_0^1=\frac{1}{3}.$$

(3) $g(x)=x^3$. In this case, $g(x)$ is a signed function. Since

$$|g(x)|=\begin{cases} x^3 & \text{if } x\geq 0,\\ -x^3 & \text{if } x<0,\end{cases}$$

we see that its expectation exists:

$$\int_{-\infty}^{\infty}|g(x)|f(x)\,dx=\frac{1}{2}\int_{-1}^{0}(-x^3)\,dx+\frac{1}{2}\int_0^1 x^3\,dx=\int_0^1 x^3\,dx=\left[\frac{x^4}{4}\right]_0^1=\frac{1}{4}<\infty.$$

The positive and negative parts of g are

$$g^+(x)=\begin{cases} x^3 & \text{if}\quad x^3\geq 0,\\ 0 & \text{if}\quad x^3<0,\end{cases}\qquad\text{and}\qquad g^-(x)=\begin{cases} 0 & \text{if}\quad x^3\geq 0,\\ -x^3 & \text{if}\quad x^3<0.\end{cases}$$

Since $x^3\geq 0$ if and only if $x\geq 0$, these can be written as:

$$g^+(x)=\begin{cases} x^3 & \text{if}\quad x\geq 0,\\ 0 & \text{if}\quad x<0,\end{cases}\qquad\text{and}\qquad g^-(x)=\begin{cases} 0 & \text{if}\quad x\geq 0,\\ -x^3 & \text{if}\quad x<0.\end{cases}$$

Thus we have

$$\begin{aligned}
\mathbb{E}(X^3)=\mathbb{E}\big[g(X)\big]&=\int_{-\infty}^{\infty} g^+(x)f(x)\,dx-\int_{-\infty}^{\infty} g^-(x)f(x)\,dx\\
&=\frac{1}{2}\int_0^1 x^3\,dx-\frac{1}{2}\int_{-1}^{0}(-x^3)\,dx\\
&=\frac{1}{2}\left[\frac{x^4}{4}\right]_0^1-\frac{1}{2}\left[\frac{-x^4}{4}\right]_{-1}^0\\
&=\left(\frac{1}{8}-0\right)-\left(0+\frac{1}{8}\right)=0.
\end{aligned}$$

(4) $g(x)=1/x$. In this case, $g(x)$ is a signed function. Since

$$|g(x)|=\begin{cases} 1/x & \text{if } x\geq 0,\\ -1/x & \text{if } x<0,\end{cases}$$

we see that its expectation does *not* exist:

$$\begin{aligned}
\int_{-\infty}^{\infty}|g(x)|f(x)\,dx&=\frac{1}{2}\int_{-1}^{0}\frac{-1}{x}\,dx+\frac{1}{2}\int_0^1\frac{1}{x}\,dx\\
&=\frac{1}{2}\int_0^1\frac{1}{x}\,dx+\frac{1}{2}\int_0^1\frac{1}{x}\,dx\\
&=\int_0^1\frac{1}{x}\,dx\\
&=\infty.
\end{aligned}$$

Another way of seeing that the expectation is undefined is to consider the positive and negative parts of g:

$$g^+(x)=\begin{cases} 1/x & \text{if}\quad 1/x\geq 0,\\ 0 & \text{if}\quad 1/x<0,\end{cases}\qquad\text{and}\qquad g^-(x)=\begin{cases} 0 & \text{if}\quad 1/x\geq 0,\\ -1/x & \text{if}\quad 1/x<0.\end{cases}$$

Thus we have

$$\begin{aligned}\mathbb{E}\left(\frac{1}{X}\right) = \mathbb{E}[g(X)] &= \int_{-\infty}^{\infty} g^+(x)f(x)\,dx - \int_{-\infty}^{\infty} g^-(x)f(x)\,dx \\ &= \frac{1}{2}\int_0^1 \frac{1}{x}\,dx - \frac{1}{2}\int_{-1}^0 \frac{-1}{x}\,dx \\ &= \frac{1}{2}\int_0^1 \frac{1}{x}\,dx - \frac{1}{2}\int_0^1 \frac{1}{x}\,dx \\ &= \infty - \infty.\end{aligned}$$

so $\mathbb{E}(1/X)$ is undefined.

(5) $g(x) = 1/x^2$. In this case, $g(x)$ is a non-negative function, so

$$\mathbb{E}\left(\frac{1}{X^2}\right) = \mathbb{E}_F(g) = \int_{-\infty}^{\infty} g(x)f(x)\,dx = \frac{1}{2}\int_{-1}^1 \frac{1}{x^2}\,dx = \int_0^1 \frac{1}{x^2}\,dx = \infty.$$

so $\mathbb{E}(1/X^2)$ is infinite (which is acceptable because $1/X^2$ is non-negative).

8. Let X be a random variable with the following CDF:

$$F(x) = \begin{cases} 0 & \text{for} \quad x \leq 1 \\ 1 - 1/x^2 & \text{for} \quad x \geq 1 \end{cases}$$

Compute $\mathbb{E}(X)$, $\mathbb{E}(X^2)$, $\mathbb{E}(1/X)$ and $\mathbb{E}(1/X^2)$.

Answer:

$$f(x) = \begin{cases} \frac{2}{x^3} & \text{if } x \geq 1, \\ 0 & \text{otherwisee.} \end{cases}$$

$$\begin{aligned}\mathbb{E}(X) &= \int_1^\infty x\left(\frac{2}{x^3}\right)dx = 2\int_1^\infty \frac{1}{x^2}\,dx = 2\left[-\frac{1}{x}\right]_1^\infty = 2 \\ \mathbb{E}(X^2) &= \int_1^\infty x^2\left(\frac{2}{x^3}\right)dx = 2\int_1^\infty \frac{1}{x}\,dx = \infty \\ \mathbb{E}\left(\frac{1}{X}\right) &= \int_1^\infty \frac{1}{x}\left(\frac{2}{x^3}\right)dx = 2\int_1^\infty \frac{1}{x^4}\,dx = 2\left[-\frac{1}{3x^3}\right]_1^\infty = \frac{2}{3} \\ \mathbb{E}\left(\frac{1}{X^2}\right) &= \int_1^\infty \frac{1}{x^2}\left(\frac{2}{x^3}\right)dx = 2\int_1^\infty \frac{1}{x^5}\,dx = 2\left[-\frac{1}{4x^4}\right]_1^\infty = \frac{1}{2}\end{aligned}$$

9. Let X be a continuous random variable with the following PDF:

$$f(x) = \begin{cases} 1 - |x| & \text{if } x \in [-1, 1] \\ 0 & \text{otherwise.} \end{cases}$$

Find the range of integer values $\alpha \in \mathbb{Z}$ for which $\mathbb{E}(X^\alpha)$ exists.

Answer: For $\alpha > 0$,

$$\mathbb{E}(X^\alpha) = \int_{-1}^0 x^\alpha(1+x)\,dx + \int_0^1 x^\alpha(1-x)\,dx < \infty$$

Let $\alpha < 0$. If α is even then X^α is non-negative, so

$$\mathbb{E}(X^\alpha) = \mathbb{E}\big((X^+)^\alpha\big) = +\infty$$

If α is odd,

$$\mathbb{E}(X^\alpha) = \mathbb{E}\big((X^+)^\alpha\big) - \mathbb{E}\big((X^-)^\alpha\big) = \infty - \infty$$

so in this case the moment $\mathbb{E}(X^\alpha)$ does not exist.

Exercise 12.1

1. Let $X \sim \text{Uniform}[0, 20]$ be a continuous random variable.
   (1) Use Chebyshev's inequality to find an upper bound on the probability $\mathbb{P}(|X - 10| \geq z)$.
   (2) Find the range of z for which Chebyshev's inequality gives a non-trivial bound.
   (3) Find the value of z for which $\mathbb{P}(|X - 10| \geq z) \leq 3/4$.

Answer:
(1) By Chebyshev's inequality, $\mathbb{P}(|X - 10| \geq z) \leq \dfrac{\text{Var}(X)}{z^2} = \dfrac{100}{3z^2}$.
(2) For a non trivial bound, we need that $\mathbb{P}(|X - 10| \geq z) \leq \dfrac{100}{3z^2} < 1$ and hence $z^2 > \dfrac{100}{3}$.
We reject the case $z = -10/\sqrt{3}$ because $\mathbb{P}(|X - 10| > -10/\sqrt{3}) = 1$.
Thus we conclude that $z > 10/\sqrt{3}$.
(3) This time we need that $\mathbb{P}(|X - 10| \geq z) \leq \dfrac{100}{3z^2} < \dfrac{3}{4}$ and hence $z^2 > \dfrac{400}{9}$.
As before, we reject the case $z = -20/3$ because $\mathbb{P}(|X - 10| > -20/3) = 1$.
Thus we conclude that $z > 20/3$.

2. Let X be a discrete random variable, taking values in the range $\{1, 2, \ldots, n\}$, and suppose that $\mathbb{E}(X) = \text{Var}(X) = 1$. Show that $\mathbb{P}(X \geq k + 1) \leq k^2$ for any integer k.

Answer: Using the fact that $X - 1 \geq 0$,

$$\mathbb{P}(X \geq k + 1) = \mathbb{P}(X - 1 \geq k) = \mathbb{P}(|X - 1| \geq k).$$

By Chebyshev's inequality, with $\mathbb{E}(X) = 0$ and $\text{Var}(X) = 1$,

$$\mathbb{P}(|X - 1| \geq k) \leq \frac{\text{Var}(X)}{k^2} = \frac{1}{k^2}$$

3. Let $k \in \mathbb{N}$. Show that Markov's inequality is tight (i.e. cannot be improved) by finding a non-negative random variable X such that

$$\mathbb{P}\big[X \geq k\mathbb{E}(X)\big] = \frac{1}{k}.$$

Answer: Let X be a random variable taking values in the set $\{0, k\}$, such that $\mathbb{P}(X = k) = 1/k$ and $\mathbb{P}(X = 0) = 1 - 1/k$. Then $\mathbb{E}(X) = 1$ and $\mathbb{P}\big(X \geq k\mathbb{E}(X)\big) = \mathbb{P}\big(X \geq k\big) = \mathbb{P}\big(X = k\big) = 1/k$ as required.

4. What does the Chebyshev inequality tell us about the probability that the value taken by a random variable deviates from its expected value by six or more standard deviations?

Answer: For any random variable X with finite variance σ^2,

$$\mathbb{P}(|X-\mu| \geq 6\sigma) \leq \frac{\sigma^2}{(6\sigma)^2} = \frac{1}{36}.$$

5. Let S_n be the number of successes in n Bernoulli trials with probability p of success on each trial. Use Chebyshev's Inequality to show that, for any $\epsilon > 0$, the upper bound

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq \frac{1}{4n\epsilon^2}$$

is valid for any p.

Answer: For the Binomial(n,p) distribution, Chebyshev's inequality yields

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq \frac{p(1-p)}{n\epsilon^2}$$

The result then follows by the fact that for any p,

$$p(1-p) = \frac{1}{4} - \left(\frac{1}{4} - p + p^2\right) = \frac{1}{4} - \left(\frac{1}{2} - p\right)^2 \leq \frac{1}{4}$$

6. Let $X \sim N(0,1)$.
   (1) Use Chebyshev's Inequality to find upper bounds for the probabilities $\mathbb{P}(|X| \geq 1)$, $\mathbb{P}(|X| \geq 2)$ and $\mathbb{P}(|X| \geq 3)$.
   (2) Use statistical tables to find the area under the standard normal curve over the intervals $[-1,1]$, $[-2,2]$ and $[-3,3]$.
   (3) Compare the bounds computed in part (a) with the exact values found in part (b). How good is the Chebyshev inequality in this case?

Answer:

(1) $\mathbb{P}(|X| \geq 1) \leq 1$, $\mathbb{P}(|X| \geq 2) \leq 1/4$ and $\mathbb{P}(|X| \geq 3) \leq 1/9$.

(2) From tables, $\mathbb{P}(|X| \geq 1) = 0.3173$, $\mathbb{P}(|X| \geq 2) = 0.0455$ and $\mathbb{P}(|X| \geq 3) = 0.0027$.

(3) Chebyshev's inequality provides only crude bounds on the tail probabilities of the standard normal distribution.

7. Let X be a random variable with mean $\mu \neq 0$ and variance σ^2, and define the *relative deviation* of X from its mean by $D = \left|\frac{X-\mu}{\mu}\right|$. Show that

$$\mathbb{P}(D \geq a) \leq \left(\frac{\sigma}{\mu a}\right)^2.$$

Answer: By Chebyshev's inequality,

$$\mathbb{P}(D \geq a) = \mathbb{P}\left(\left|\frac{X-\mu}{\mu}\right| \geq a\right) = \mathbb{P}\big(|X-\mu| \geq |\mu|a\big) \leq \frac{\sigma^2}{\mu^2 a^2}$$

Exercise 13.1

1. Let $X \sim \text{Binomial}(m,p)$ and $Y \sim \text{Binomial}(n,p)$. Show that $X + Y \sim \text{Binomial}(m+n,p)$,

Answer: The PGFs of X and Y are

$$G_X(t) = (1-p+pt)^m \quad \text{and} \quad G_Y(t) = (1-p+pt)^n$$

Using the properties of PGFs,

$$G_{X+Y}(t) = G_X(t)G_Y(t) = (1-p+pt)^m(1-p+pt)^n = (1-p+pt)^{m+n},$$

which we recognise as the PGF of the Binomial$(m+n,p)$ distribution.

2. Show that a discrete distribution on the non-negative integers is uniquely determined by its PGF, in the sense that if two such random variables X and Y have PGFs $G_X(t)$ and $G_Y(t)$ respectively, then $G_X(t) = G_Y(t)$ if and only if $\mathbb{P}(X=k) = \mathbb{P}(Y=k)$ for all $k = 0,1,2,\ldots$.

Answer: The PGFs of X and Y are

$$G_X(t) = \sum_{k=1}^{\infty} \mathbb{P}(X=k)t^k \quad \text{and} \quad G_Y(t) = \sum_{k=1}^{\infty} \mathbb{P}(Y=k)t^k$$

Clearly, if $\mathbb{P}(X=k) = \mathbb{P}(Y=k)$ for all $k = 0,1,2,\ldots$, then $G_X(t) = G_Y(t)$. Conversely, $G_X(1) = 1$ implies that its power series expansion (about the origin) is unique, and likewise for G_Y. Thus if $G_X = G_Y$, their power series must have identical coefficients, so $\mathbb{P}(X=k) = \mathbb{P}(Y=k)$ for all $k = 0,1,2,\ldots$, as required

3. The PGF of a random variable is given by $G(t) = 1/(2-t)$. What is its PMF?

Answer: To find the PMF, we need to express $G(t)$ as a power series:

$$G_X(t) = \frac{1}{2-t} = \frac{1}{2}\left(1-\frac{s}{2}\right)^{-1} = \frac{1}{2}\left(1+\frac{t}{2}+\left(\frac{t}{2}\right)^2+\ldots\right) - \sum_{k=0}^{\infty}\left(\frac{1}{2}\right)^{k+1} t^k$$

Thus the PMF of X

$$\mathbb{P}(X=k) = \left(\frac{1}{2}\right)^{k+1} \quad \text{for } k = 0,1,2,\ldots.$$

4. Let $X \sim$ Binomial(n,p). Using the PGF of X, show that

$$\mathbb{E}\left(\frac{1}{1+X}\right) = \frac{1-(1-p)^{n+1}}{(n+1)p}.$$

Answer: Let $G(t)$ be the PGF of X. Then $G(t) = \mathbb{E}(t^X) = (q+pt)^n$ where $q = 1-p$. Now

$$\int_0^1 t^x\,dt = \left[\frac{t^{1+x}}{1+x}\right]_0^1 = \frac{1}{1+x},$$

so

$$\mathbb{E}\left(\frac{1}{1+X}\right) = \mathbb{E}\left(\int_0^1 t^X\,dt\right) = \int_0^1 \mathbb{E}(t^X)\,dt = \int_0^1 (q+pt)^n\,dt = \frac{1-q^{n+1}}{(n+1)p}$$

Exercise 14.1

1. Let X be a discrete random variable, taking values in the set $\{-3,-2,-1,0,1,2,3\}$ with uniform probability, and let $M(t)$ denote the MGF of X.

(1) Show that $M(t) = \frac{1}{7}(e^{-3t} + e^{-2t} + e^{-t} + 1 + e^{t} + e^{2t} + e^{3t})$.

(2) Use $M(t)$ to compute the mean and variance of X.

Answer:

(1) $M(t) = \mathbb{E}(e^{Xt}) = \sum_k e^{tk}p(k) = \frac{1}{7}(e^{-3t} + e^{-2t} + e^{-t} + 1 + e^{t} + e^{2t} + e^{3t})$

(2) Using the power series expansion $e^x = \displaystyle\sum_{k=0}^{\infty}\frac{x^k}{k!}$, the MGF of X can be written as

$$\begin{aligned} M(t) &= \frac{1}{7}\left[\left(1 - 3t + \frac{9t^2}{2} + \dots\right) + \left(1 - 2t + \frac{4t^2}{2} + \dots\right) + \left(1 - t + \frac{t^2}{2} + \dots\right)\right. \\ &\qquad \left. +1 + \left(1 + t + \frac{t^2}{2} + \dots\right) + \left(1 + 2t + \frac{4t^2}{2} + \dots\right) + \left(1 + 3t + \frac{9t^2}{2} + \dots\right)\right] \\ &= 1 + \frac{4t^2}{2} + \dots \end{aligned}$$

- The mean is the coefficient of t, so $\mu = 0$.
- The second non-central moment is the coefficient of $\frac{1}{2}t^2$, so $\mu_2' = 4$.
- The variance is therefore $\sigma^2 = \mu_2' - \mu^2 = 4$.

A more direct method is to use the fact that $\mu = M_X'(0)$ and $\mu_2' = M_X''(0)$. Since

$$\begin{aligned} M'(t) &= \frac{1}{7}(-3e^{-3t} - 2e^{-2t} - e^{-t} + e^{t} + 2e^{2t} + 3e^{3t}), \\ M''(t) &= \frac{1}{7}(9e^{-3t} + 4e^{-2t} + e^{-t} + e^{t} + 4e^{2t} + 9e^{3t}) \end{aligned}$$

it follows that

$$\begin{aligned} M'(0) &= \frac{1}{7}(-3 - 2 - 2 + 1 + 2 + 3) = 0, \\ M''(0) &= \frac{1}{7}(9 + 4 + 1 + 1 + 4 + 9) = 4. \end{aligned}$$

2. A continuous random variable X has MGF given by $M(t) = \exp(t^2 + 3t)$. Find the distribution of X.

Answer:

$$M_X(t) = \exp(t^2 + 3t) = \exp(3t + \frac{1}{2}2t^2)).$$

The MGF is of the form $\exp(\mu t + \frac{1}{2}\sigma^2 t^2)$, which is the MGF of a normal distribution. The mean μ is 3 and the variance σ^2 is 2, so $X \sim N(3, 2)$. Note that under reasonable conditions, a random variable can be uniquely identified by its MGF: only normally distributed random variables have MGFs of the form $\exp(\mu t + \frac{1}{2}\sigma^2 t^2)$.

3. Let X be a discrete random variable with probability mass function

$$\mathbb{P}(X = k) = \begin{cases} q^k p & k = 0, 1, 2, \dots \\ 0 & \text{otherwise,} \end{cases}$$

where $0 < p < 1$ and $q = 1 - p$.

(1) Show that the MGF of X is given by $M(t) = \dfrac{p}{1 - qe^t}$ for $t < -\log q$.

(2) Find the PGF of X.

(3) Use the PGF of X to find the PMF of $Y = X + 1$.

(4) Use $M(t)$ to find the mean and variance of X.

Answer:

(1) The MGF of X is

$$M(t) = \mathbb{E}(e^{Xt}) = \sum_{k=0}^{\infty} e^{kt} q^k p = p \sum_{k=0}^{\infty} \left[qe^t\right]^k = \frac{p}{1-qe^t}$$

provided that $|qe^t| < 1$, or equivalently that $t < -\log q$.

(2) Since $M(t) = G(e^t)$, the PGF is

$$G(t) = M(\log t) = \frac{p}{1-qt}.$$

Note that $G(0) = 0$ and $G(1) = 1$.

(3) To derive the PMF of $Y = X+1$, we have

$$G_Y(t) = \mathbb{E}(t^Y) = \mathbb{E}(t^{X+1}) = \mathbb{E}(t^X.t) = t\mathbb{E}(t^X) = tG_(t)$$

so by part (i),

$$G_Y(t) = \frac{pt}{1-qt} = pt(1+qt+q^2t^2+\ldots$$

Comparing the coefficients in this expression with those of $G_Y(t)$ expressed as a power series,

$$G_Y(t) = \sum_{k=0}^{\infty} \mathbb{P}(Y=k)t^k = \mathbb{P}(Y=0) + \mathbb{P}(Y=1)t + \mathbb{P}(Y=2)t^2 + \ldots$$

we see that

$$\begin{array}{lll} \mathbb{P}(Y=0) & = \text{constant term} & = 0 \\ \mathbb{P}(Y=1) & = \text{coefficient of } t & = p \\ \mathbb{P}(Y=2) & = \text{coefficient of } t^2 & = p(1-p) \\ \mathbb{P}(Y=3) & = \text{coefficient of } t^3 & = p(1-p)^2 \text{ etc.} \end{array}$$

As we might expect, $\mathbb{P}(Y=k) = \mathbb{P}(X=k-1)$ for $k = 0, 1, \ldots$.

(4)

$$M'(t) = \frac{d}{dt}\left(\frac{p}{1-qe^t}\right) = \frac{pqe^t}{(1-qe^t)^2}$$

so

$$\mu = M'(0) = \frac{pq}{(1-q)^2} = \frac{q}{p} = \frac{1-p}{p}.$$

Similarly,

$$M''_X(t) = \frac{d}{dt}\left(\frac{pqe^t}{(1-qe^t)^2}\right) = \frac{pqe^t(1+qe^t)}{(1-qe^t)^3}$$

so

$$\mu'_2 = M''_X(0) = \frac{q(1+q)}{p^2} = \frac{(1-p)(2-p)}{p}$$

and hence

$$\mathrm{Var}(X) = \mu'_2 - \mu^2 = \frac{q}{p^2} = \frac{1-p}{p^2}.$$

4. Let $M(t)$ denote the MGF of the normal distribution $\mathrm{N}(0,\sigma^2)$. By exanding $M(t)$ as a power series in t, show that the moments μ_k of the $\mathrm{N}(0,\sigma^2)$ distribution are zero if k is odd, and equal to

$$\mu_{2m} = \frac{\sigma^{2m}(2m)!}{2^m m!} \qquad \text{if } k = 2m \text{ is even.}$$

Answer: The MGF of the $N(\mu,\sigma^2)$ distribution is $\exp\left(\mu t+\frac{1}{2}\sigma^2t^2\right)$, so

$$M(t)=\exp\left(\frac{\sigma^2t^2}{2}\right).$$

Expand this as a power series in t:

$$M(t)=1+\frac{\sigma^2t^2}{2}+\frac{\sigma^4t^4}{4\times 2!}+\frac{\sigma^6t^6}{8\times 3!}+\ldots+\frac{\sigma^{2m}t^{2m}}{2^m\times m}+\ldots$$

The kth moment μ_k is the coefficient of $t^k/k!$ in this expansion. In particular, the skewness $\gamma_1=\mu_3/\sigma^3$ is zero, and the excess kurtosis is

$$\gamma_2=\frac{\mu_4}{\sigma^4}-3=\frac{\sigma^4 4!}{4\times 2!}\cdot\frac{1}{\sigma^4}-3=0.$$

5. Let $X\sim\text{Exponential}(\theta)$ where θ is a scale parameter.
   (1) Show that the MGF of X is $M(t)=\dfrac{1}{1-\theta t}$.
   (2) By expanding this expression as a power series in t, find the first four non-central moments of X.
   (3) Find the skewness γ_1 and the excess kurtosis γ_2 of X.

Answer:
(1)

$$\begin{aligned}
M(t)=\mathbb{E}(e^{Xt}) &= \int e^{xt}\frac{1}{\theta}e^{-x/\theta}\,dx\\
&=\frac{1}{\theta}\int_0^\infty e^{-x(1-\theta t)/\theta}\,dx\\
&=-\frac{1}{1-\theta t}\int_0^\infty \frac{d}{dx}e^{-x(1-\theta t)/\theta}\,dx\\
&=-\frac{1}{1-\theta t}\left[e^{-x(1-\theta t)/\theta}\right]_0^\infty\\
&=\frac{1}{1-\theta t}.
\end{aligned}$$

(2) Using the binomial expansion,

$$M(t)=(1-\theta t)^{-1}=1+\theta t+\theta^2t^2+\theta^3t^3+\theta^4t^4+\ldots$$

The non-central moment μ_k is the coefficient of $\frac{t^k}{k!}$ in this power series expansion, so $\mu=\theta$, $\mu_2=2\theta^2$, $\mu_3=6\theta^3$ and $\mu_4=24\theta^4$.

(3) The skewness and excess kurtosis are

$$\begin{aligned}
\gamma_1&=\frac{\mu_3}{\sigma^3}=\frac{2\theta^3}{\theta^3}=2,\quad\text{and}\\
\gamma_2&=\frac{\mu_4}{\sigma^4}-3=\frac{9\theta^4}{\theta^4}-3=6,
\end{aligned}$$

respectively.

6. Let $X_1,X_2,\ldots$ be independent and identically distributed random variables, with each $X_i\sim\mathrm{N}(\mu,\sigma^2)$.
   (1) Find the MGF of the random variable $\bar{X}=\dfrac{1}{n}\sum_{i=1}^n X_i$.
   (2) Show that $\bar{X}$ has a normal distribution, and find its mean and variance.

Answer: Let $M(t)$ denote the common MGF of the random variables X_i:

$$M(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

Moment generating functions have the following properties:

- $M_{X+Y}(t) = M_X(t)M_Y(t)$
- If $Y = a + bX$ then $M_Y(t) = e^{at}M_X(bt)$.

In view of these, the MGF of $\bar{X}$ is

$$\begin{aligned}
M_{\bar{X}}(t) &= M_{\frac{1}{n}(X_1+X_2+\ldots+X_n)}(t) \\
&= M_{X_1+X_2+\ldots+X_n}\left(\frac{t}{n}\right) \\
&= \left[M\left(\frac{t}{n}\right)\right]^n = \exp\left[\frac{\mu t}{n} + \frac{\sigma^2 t^2}{2n^2}\right]^n \\
&= \exp\left[\mu t + \frac{1}{2}\left(\frac{\sigma^2}{n}\right)t^2\right]
\end{aligned}$$

which is the MGF of a normal distribution with mean μ and variance σ^2/n.

7. Let $X_1 \sim \text{Gamma}(k_1, \theta)$ and $X_2 \sim \text{Gamma}(k_2, \theta)$ be independent random variables. Use the MGFs of X_1 and X_2 to find the distribution of the random variable $Y = X_1 + X_2$.

Answer: The *MGFs* of X_1 and X_2 are:

$$M_{X_1}(t) = \frac{1}{(1-\theta t)^{k_1}} \quad \text{and} \quad M_{X_2}(t) = \frac{1}{(1-\theta t)^{k_2}}.$$

Since X_1 and X_2 are independent,

$$M_{X_1+X_2}(t) = M_{X_1}(t)M_{X_2}(t) = \frac{1}{(1-\theta t)^{k_1+k_2}}$$

This is the MGF of a $\text{Gamma}(k_1 + k_2, \theta)$ random variable.

8. A coin has probability p of showing heads. The coin is tossed repeatedly until exactly k heads occur. Let N be the number of times the coin is tossed. Using the continuity theoram for characteristic functions, show that the distribution of the random variable $X = 2pN$ converges to a gamma distribution as $p \to 0$.

Answer: Let $N = Z_1 + Z_2 + \ldots + Z_k$ with $Z_i \sim \text{Geometric}(p)$ independent. Let $\phi_Z(t)$ denote the common characteristic function of the Z_i:

$$\phi_Z(t) = \mathbb{E}(e^{itZ}) = \frac{pe^{it}}{1-qe^{it}} \qquad \text{where } Z \sim \text{Geometric}(p) \text{ and } q = 1-p.$$

By the properties of characteristic functions, the characteristic function of N is

$$\phi_N(t) = \big[\phi_Z(t)\big]^k = \left[\frac{pe^{it}}{1-qe^{it}}\right]^k$$

and the characteristic function of $X = 2pN$ is

$$\phi_X(t) = \phi_N(2pt) = \left[\frac{pe^{2pit}}{1-(1-p)e^{2pit}}\right] = \left[\frac{1+2pit+o(1)}{1-2it+o(1)}\right] \to \frac{1}{1-2it} \qquad \text{as} \quad p \to 0,$$

where $o(1)$ respresnts a quantity that tends to zero as $p \to 0$. This is the characteristic function of the Gamma$(k, \frac{1}{2})$ distribution. The result then follows by the continuity theorem for characteristic functions.

9. Let X and Y be independent and identically distributed random variables, with means equal to 0 and variances equal to 1. Let $\phi(t)$ denote their common characteristic function, and suppose that the random variables $X+Y$ and $X-Y$ are independent. Show that $\phi(2t) = \phi(t)^3\phi(-t)$, and hence deduce that X and Y must be independent standard normal variables.

Answer: Let $U = X+Y$ and $V = X-Y$. Since U and V are independent, we have $\phi_{U+V}(t) = \phi_U\phi_V$, or equivalently $\phi_{2X} = \phi_{X+Y}\phi_{X-Y}$. Thus, since $\phi_{2X}(t) = \phi(2t)$,$\phi_{X+Y} = \phi(t)^2$ and $\phi X - Y(t) = \phi(t)\phi(-t)$, it follows that

$$\phi(2t) = \phi(t)^3\phi(-t)$$

To show that $X, Y \sim \mathrm{N}(0,1)$, it is sufficient to show that $\phi(t) = e^{-\frac{1}{2}t^2}$ (by the inversion theorem).
It can be shown that characteristic functions are symmetric: $\phi(t) = \phi(-t)$ for all t.
Thus we have $\phi(2t) = \phi(t)^4$, and hence

$$\phi(t) = \phi\left(\frac{1}{2}t\right)^4 = \phi\left(\frac{1}{4}t\right)^{16} = \ldots = \phi\left(\frac{1}{2^n}t\right)^{2^{2^n}} \qquad \text{for } n = 0, 1, 2, \ldots.$$

Hence,

$$\phi(t) = \left[1 - \frac{1}{2}\left(\frac{t}{2^n}\right)^2 + \ldots\right]^{2^{2^n}} \to e^{-\frac{1}{2}t^2} \qquad \text{as } n \to \infty.$$

Exercise 15.1

1. Let c be a constant, and let $X_1, X_2, \ldots$ be a sequence of random variables with $\mathbb{E}(X_n) = c$ and $\mathrm{Var}(X_n) = 1/\sqrt{n}$ for each n. Show that the sequence converges to c in probability as $n \to \infty$.

Answer: Let $\epsilon > 0$. By Chebyshev's inequality,

$$\mathbb{P}\big(|X_n - c| \geq \epsilon\big) \leq \frac{\mathrm{Var}(X_n)}{\epsilon^2} = \frac{1}{\epsilon^2\sqrt{n}}$$

for all $n \in \mathbb{N}$. Thus $\lim_{n\to\infty} \mathbb{P}\big(|X_n - c| \geq \epsilon\big) = 0$, so the sequence converges to c in probability.

2. A fair coin is tossed n times. Does the law of large numbers ensure that the observed number of heads will not deviate from $n/2$ by more than 100 with probability of at least 0.99, provided that n is sufficiently large?

Answer: Yes, because the indicator variable has finite mean and variance.

Exercise 16.1

1. The continuous uniform distribution on (a,b) has the following PDF:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b, \\ 0 & \text{otherwise.} \end{cases}$$

Use the central limit theorem to deduce the approximate distribution of the sample mean of n independent observations from this distribution when n is large.

Answer: The mean is

$$\mu = \int_a^b \frac{x}{b-a}\,dx = \frac{a+b}{2},$$

and the second moment is

$$\mu_2 = \int_a^b \frac{x^2}{b-a}\,dx = \frac{a^2+ab+b^2}{3},$$

so the variance is

$$\sigma^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{(b-a)^2}{12}$$

By the central limit theorem, if X is a random variable with mean μ and variance σ^2, the distribution of the sample mean $\bar{X}$ of a random sample of n independent observations is approximately $N(\mu, \frac{\sigma^2}{n})$, the approximation being better for larger n. In this case, the approximate distribution of $\bar{X}$ is $N\left(\frac{a+b}{2}, \frac{(b-a)^2}{12n}\right)$.

2. The exponential distribution with scale parameter $\theta > 0$ has the following PDF:

$$f(x) = \begin{cases} \frac{1}{\theta}e^{-x/\theta} & x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Use the central limit theorem to deduce the approximate distribution of the sample mean of n independent observations from this distribution when n is large.

Answer:

$$\begin{aligned} \mathbb{E}(X) &= \frac{1}{\theta}\int_0^\infty xe^{-x/\theta}\,dx = \theta, \\ \mathbb{E}(X^2) &= \frac{1}{\theta}\int_0^\infty x^2e^{-x/\theta}\,dx = 2\theta^2 \\ \mathrm{Var}(Y) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \theta^2. \end{aligned}$$

By the CLT, the distribution of $\bar{X}$ is approximately $\mathrm{N}(\mu, \frac{\sigma^2}{n})$, the approximation being better for larger n. In this case, the approximate distribution of $\bar{X}$ is $\mathrm{N}\left(\theta, \theta^2/n\right)$.

3. Let $X \sim \text{Binomial}(n_1, p_1)$ and $X_2 \sim \text{Binomial}(n_2, p_2)$ be independent random variables.

(1) Use the central limit theorem to find the approximate distribution of $Y = X_1 - X_2$ when n_1 and n_2 are both large.

(2) Let $Y_1 = X_1/n_1$ and $Y_2 = X_2/n_2$. Show that $Y_1 - Y_2$ is approximately normally distributed with mean $p_1 - p_2$ and variance $\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}$ when n_1 and n_2 are both large.

(3) Show that when n_1 and n_2 are both large,

$$\frac{(Y_1 - Y_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim N(0,1) \qquad \text{approx.}$$

Answer:

(1) The mean and variance of X_1 are respectively n_1p_1 and $n_1p_1q_1$ where $q_1 = 1 - p_1$. The mean and variance of X_2 are respectively n_2p_2 and $n_2p_2q_2$ where $q_2 = 1 - p_2$. Since $Y = X_1 - X_2$ is

a linear combination of random variables,

$$\mathbb{E}(Y) = \mathbb{E}(X_1) - \mathbb{E}(X_2) = n_1p_1 - n_2p_2$$

and since X_1 and X_2 are independent,

$$\text{Var}(Y) = \text{Var}(X_1) + \text{Var}(X_2) = n_1p_1q_1 + n_2p_2q_2.$$

Because both X_1 and X_2 are the sums of Bernoulli random variables, the CLT applies, so the approximate distribution of Y is

$$Y \sim N(n_1p_1 - n_2p_2, n_1p_1q_1 + n_2p_2q_2)$$

(2) For large n_1 and n_2, by the CLT the distribution of X_1 is approximately $N(n_1p_1, n_1p_1(1-p_1))$ for n_1 large, and the distribution of X_2 is approximately $N(n_2p_2, n_2p_2(1-p_2))$ for n_2 large. Thus the distributions of Y_1 and Y_2 are approximately $N\left(p_1, \frac{p_1q_1}{n_1}\right)$ and $N\left(p_2, \frac{p_2q_2}{n_2}\right)$ respectively, and the distribution of $Y_1 - Y_2$ is therefore approximately $N\left(p_1 - p_2, \frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}\right)$ for large n_1 and n_2.

(3) The usual standardization for the normal distribution (subtract the mean and divide by the standard deviation) yields the result. This is used in devising approximate tests and confidence intervals for the difference of proportions.

4. 5% of items produced by a factory production line are defective. Items are packed into boxes of 2000 items. As part of a quality control exercise, a box is chosen at random and found to contain 120 defective items. Use the central limit theorem to estimate the probability of finding at least this number of defective items when the production line is operating properly.

Answer: Let X be the number of defective items in a box. Then $X \sim \text{Binomial}(n,p)$ with $n = 2000$ and $p = 0.05$. Since n is large, X has approximately normal distribution with mean equal to $np(1-p) = 100$, and variance equal to $npq = 95$. The standardized variable $Z = (X - 100)/\sqrt{95}$ has therefore approximately the standard normal distribution $N(0,1)$. Thus

$$\mathbb{P}(X \geq 120) = \mathbb{P}\left(Z \geq \frac{120 - 100}{\sqrt{95}}\right) = \mathbb{P}(Z \geq 2.0520) \approx 0.0202$$

where the probability $\mathbb{P}(Z \geq 2.0520) \approx 0.0202$ can be obtained from statistical tables.

5. Use the central limit theorem to prove the law of large numbers.

Answer: Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables, and define $S_n = \sum_{i=1}^n X_i$. To prove the (weak) law of large numbers, we need to show that

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \to 0 \qquad \text{as} \quad n \to \infty$$

Now,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = \mathbb{P}\left(\left|\frac{S_n - n\mu}{\sigma\sqrt{n}}\right| \geq \frac{n\epsilon}{\sigma\sqrt{n}}\right) = \mathbb{P}\left(\left|\frac{S_n - n\mu}{\sigma\sqrt{n}}\right| \geq \frac{\sqrt{n}\epsilon}{\sigma}\right)$$

By the central limit theorem, $\dfrac{S_n - n\mu}{\sigma\sqrt{n}}$ is approximately distributed according to $N(0,1)$, so this probability is approximated by the area under the standard normal curve between $\dfrac{\sqrt{n}\epsilon}{\sigma}$ and infinity, which approaches zero as $n \to \infty$.

6. We perform a sequence of independent Bernoulli trials, each with probability of success p, until a fixed number r of successes is obtained. The total number of failures Y (up to the rth succes) has the *negative*

binomial distribution with parameters r and p, so the PMF of Y is

$$\mathbb{P}(Y = k) = \binom{k+r-1}{k}(1-p)^k p^r, \qquad k = 0, 1, 2, \ldots$$

Using the fact that Y can be written as the sum of r independent geometric random variables, show that this distribution can be approximated by a normal distribution when r is large.

Answer: If $Y \sim \mathrm{NB}(r, p)$, we can write

$$Y = X_1 + X_2 + \ldots + X_r \qquad \text{where} \quad X_i \sim \mathrm{Geometric}(p).$$

Let $X \sim \mathrm{Geometric}(p)$. Since $\mathrm{Var}(X) < \infty$, it follows by the central limit theorem that

$$\frac{Y - r\mathbb{E}(X)}{\sqrt{r\mathrm{Var}(X)}} \to \mathrm{N}(0,1) \quad \text{in distribution as } r \to \infty.$$

In fact, since $\mathbb{E}(X) = (1-p)/p$ and $\mathrm{Var}(X) = (1-p)/p^2$, we see that Y can be approximated by the $\mathrm{N}\left(\frac{r(1-p)}{p}, \frac{r(1-p)}{p^2}\right)$ distribution as $r \to \infty$.