

2. Probability

2.2 Sample Spaces and Events

The **sample space** Ω is the set of possible outcomes of an experiment. Points ω in Ω are called **sample outcomes** or **realizations**. **Events** are subsets of Ω .

Given an event A , let $A^c = \{\omega \in \Omega : \text{not } (\omega \in A)\}$ denote the complement of A . The complement of Ω is the empty set \emptyset . The union of events A and B is defined as $A \cup B = \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\}$. If A_1, A_2, \dots is a sequence of sets, then

$$\bigcup_{i=1}^{\infty} A_i = \{\omega \in \Omega : \omega \in A_i \text{ for some } i\}$$

The intersection of A and B is $A \cap B = \{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$. If A_1, A_2, \dots is a sequence of sets then

$$\bigcap_{i=1}^{\infty} A_i = \{\omega \in \Omega : \omega \in A_i \text{ for all } i\}$$

Let $A - B = \{\omega \in \Omega : \omega \in A \text{ and not } (\omega \in B)\}$. If every element of A is contained in B we write $A \subset B$ or $B \supset A$. If A is a finite set, let $|A|$ denote the number of elements in A .

notation	meaning
Ω	sample space
ω	outcome
A	event (subset of Ω)
$ A $	number of elements in A (if finite)
A^c	complement of A (not A)
$A \cup B$	union (A or B)
$A \cap B$ or AB	intersection (A and B)
$A - B$	set difference (points in A but not in B)
$A \subset B$	set inclusion (A is a subset of or equal to B)
\emptyset	null event (always false)
Ω	true event (always true)

We say that A_1, A_2, \dots are **disjoint** or **mutually exclusive** if $A_i \cap A_j = \emptyset$ whenever $i \neq j$. A **partition** of Ω is a sequence of disjoint sets A_1, A_2, \dots such that $\cup_{i=1}^{\infty} A_i = \Omega$. Given an event A , define the **indicator function of A** by

$$I_A(\omega) = I(\omega \in A) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases}$$

A sequence of sets A_1, A_2, \dots is **monotone increasing** if $A_1 \subset A_2 \subset \dots$, and we define $\lim_{n \rightarrow \infty} A_n = \cup_{i=1}^{\infty} A_i$. A sequence of sets A_1, A_2, \dots is **monotone decreasing** if $A_1 \supset A_2 \supset \dots$ and then we define $\lim_{n \rightarrow \infty} A_n = \cap_{i=1}^{\infty} A_i$. In either case, we will write $A_n \rightarrow A$.

2.3 Probability

A function \mathbb{P} that assign a real number $\mathbb{P}(A)$ to each event A is a **probability distribution** or a **probability measure** if it satisfies the following three axioms:

- **Axiom 1:** $\mathbb{P}(A) \geq 0$ for every A
- **Axiom 2:** $\mathbb{P}(\Omega) = 1$
- **Axiom 3:** If A_1, A_2, \dots are disjoint then

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

A few properties that can be derived from the axioms:

- $\mathbb{P}(\emptyset) = 0$
- $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$
- $0 \leq \mathbb{P}(A) \leq 1$
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- $A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

Lemma 2.6. For any events A and B , $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)$.

Proof.

$$\mathbb{P}(A \cup B) = \mathbb{P}((AB^c) \cup (AB) \cup (A^c B)) \quad (1)$$

$$= \mathbb{P}(AB^c) + \mathbb{P}(AB) + \mathbb{P}(A^c B) \quad (2)$$

$$= \mathbb{P}(AB^c) + \mathbb{P}(AB) + \mathbb{P}(A^c B) + \mathbb{P}(AB) - \mathbb{P}(AB) \quad (3)$$

$$= \mathbb{P}((AB^c) \cup (AB)) + \mathbb{P}((A^c B) \cup (AB)) - \mathbb{P}(AB) \quad (4)$$

$$= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB) \quad (5)$$

Theorem 2.8 (Continuity of Probabilities). If $A_n \rightarrow A$ then $\mathbb{P}(A_n) \rightarrow \mathbb{P}(A)$ as $n \rightarrow \infty$.

Proof. Suppose that A_n is monotone increasing, $A_1 \subset A_2 \subset \dots$. Let $B_1 = A_1$, and $B_{n+1} = A_{n+1} - A_n$ for $n > 1$. The B_i 's are disjoint by construction, and $A_n = \cup_{i=1}^n A_i = \cup_{i=1}^n B_i$ for all n . From axiom 3,

$$\mathbb{P}(A_n) = \mathbb{P}(\cup_{i=1}^n B_i) = \sum_{i=1}^n \mathbb{P}(B_i)$$

and so

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}(B_i) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \mathbb{P}(\cup_{i=1}^{\infty} B_i) = \mathbb{P}(A)$$

2.4 Probability on Finite Sample Spaces

If Ω is finite and each outcome is equally likely, then

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$$

which is called the **uniform probability distribution**.

We will need a few facts from counting theory later.

- Given n objects, the number of way or ordering these objects is $n! = n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1$. We define $0! = 1$.
- We define

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

read " n choose k ", which is the number of different ways of choosing k objects from n .

- Note that choosing a subset k objects can be mapped to choosing the complement set of $n - k$ objects, so

$$\binom{n}{k} = \binom{n}{n-k}$$

and that there is only one way of choosing the empty set, so

$$\binom{n}{0} = \binom{n}{n} = 1$$

2.5 Independent Events

- Two events A and B are **independent** if

$$\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$$

and we write $A \perp B$. A set of events $\{A_i : i \in I\}$ is independent if

$$\mathbb{P}(\cap_{i \in J} A_i) = \prod_{i \in J} \mathbb{P}(A_i)$$

for every finite subset J of I .

- Independence is sometimes assumed and sometimes derived.
- Disjoint events with positive probability are not independent.

2.6 Conditional Probability

- If $\mathbb{P}(B) > 0$ then the **conditional probability** of A given B is

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)}$$

- $\mathbb{P}(\cdot|B)$ satisfies the axioms of probability, for fixed B . In general, $\mathbb{P}(A|\cdot)$ does **not** satisfy the axioms of probability for fixed A .
- In general, $\mathbb{P}(B|A) \neq \mathbb{P}(A|B)$.

- A and B are independent if and only if $\mathbb{P}(A|B) = \mathbb{P}(A)$.

2.7 Bayes' Theorem

Theorem 2.15 (The Law of Total Probability). Let A_1, \dots, A_k be a partition of Ω . Then, for any event B ,

$$\mathbb{P}(B) = \sum_{i=1}^k \mathbb{P}(B|A_i)\mathbb{P}(A_i)$$

Proof. Let $C_j = BA_j$. Note that the C_j 's are disjoint and that $B = \cup_{i=1}^k C_j$. Hence

$$\mathbb{P}(B) = \sum_j \mathbb{P}(C_j) = \sum_j \mathbb{P}(BA_j) = \sum_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)$$

Theorem 2.16 (Bayes' Theorem). Let A_1, \dots, A_k be a partition of Ω such that $\mathbb{P}(A_i) > 0$ for each i . If $\mathbb{P}(B) > 0$, then, for each $i = 1, \dots, k$,

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$$

We call $\mathbb{P}(A_i)$ the **prior probability** of A_i and $\mathbb{P}(A_i|B)$ the **posterior probability** of A_i .

Proof. We apply the definition of conditional probability twice, followed by the law of total probability:

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$$

2.9 Technical Appendix

Generally, it is not feasible to assign probabilities to all the subsets of a sample space Ω . Instead, one restricts attention to a set of events called a **σ -algebra** or a **σ -field**, which is a class \mathcal{A} that satisfies:

- $\emptyset \in \mathcal{A}$
- If $A_1, A_2, \dots \in \mathcal{A}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$
- $A \in \mathcal{A}$ implies that $A^c \in \mathcal{A}$

The sets in \mathcal{A} are said to be **measurable**. We call (Ω, \mathcal{A}) a **measurable space**. If \mathbb{P} is a probability measure defined in \mathcal{A} then $(\Omega, \mathcal{A}, \mathbb{P})$ is a **probability space**. When Ω is the real line, we take \mathcal{A} to be the smallest σ -field that contains all off the open sets, which is called the **Borel σ -field**.

2.10 Exercises

Exercise 2.10.1. Fill in the details in the proof of Theorem 2.8. Also, prove the monotone decreasing case.

If $A_n \rightarrow A$ then $\mathbb{P}(A_n) \rightarrow \mathbb{P}(A)$ as $n \rightarrow \infty$.

Solution.

Suppose that A_n is monotone increasing, $A_1 \subset A_2 \subset \dots$. Let $B_1 = A_1$, and $B_{i+1} = A_{i+1} - A_i$ for $i > 1$.

The B_i 's are disjoint by construction: assuming without loss of generality $i < j$, $\omega \in B_i \cap B_j$ implies that ω is in A_j , A_i , but not in A_{j-1} , A_{i-1} , where $A_0 = \emptyset$. In particular, this means that $\omega \in A_i$ but not $\omega \in A_{j-1}$. Since $A_i \subset A_{j-1}$, this implies that no such ω can satisfy those properties, and so B_i and B_j are disjoint.

Note that $A_n = \cup_{i=1}^n A_i = \cup_{i=1}^n B_i$ for all n :

$$\cup_{i=1}^n B_i = \cup_{i=1}^n (A_i - A_{i-1}) \subset \cup_{i=1}^n A_i = A_n$$

Also note that $A_n \subset \cup_{i=1}^n B_i$, since, if $f(\omega) = \min\{k : \omega \in A_k\}$, then $\omega \in B_{f(\omega)}$, so all elements of A_n are in some B_k .

The proof follows as given; from axiom 3,

$$\mathbb{P}(A_n) = \mathbb{P}\left(\cup_{i=1}^n B_i\right) = \sum_{i=1}^n \mathbb{P}(B_i)$$

and so

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}(B_i) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \mathbb{P}\left(\cup_{i=1}^{\infty} B_i\right) = \mathbb{P}(A)$$

The monotone decreasing case can be obtained by looking at the complementary series A_1^c, A_2^c, \dots , which is monotone increasing. We get

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n^c) = \mathbb{P}(A^c) \quad (6)$$

$$\lim_{n \rightarrow \infty} 1 - \mathbb{P}(A_n^c) = 1 - \mathbb{P}(A^c) \quad (7)$$

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = 1 - \mathbb{P}(A^c) \quad (8)$$

Exercise 2.10.2. Prove the statements in equation (2.1).

- $\mathbb{P}(\emptyset) = 0$
- $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$
- $0 \leq \mathbb{P}(A) \leq 1$
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- $A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

Solution.

- By partitioning the event space Ω into disjoint partitions (Ω, \emptyset) we get

$$\mathbb{P}(\Omega) + \mathbb{P}(\emptyset) = \mathbb{P}(\Omega) \Rightarrow \mathbb{P}(\emptyset) = 0$$

- Assuming $A \subset B$ and partitioning B as $(A, B - A)$, we get

$$\mathbb{P}(A) + \mathbb{P}(B - A) = \mathbb{P}(B) \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$$

- $\mathbb{P}(A) \geq 0$ from axiom 1. By partitioning Ω as (A, A^c) , we get

$$\mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(\Omega) = 1 \Rightarrow \mathbb{P}(A) \leq 1$$

- By partitioning Ω as (A, A^c) , we get

$$\mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(\Omega) = 1 \Rightarrow \mathbb{P}(A) = 1 - \mathbb{P}(A^c)$$

- Assuming A, B are disjoint, we partition $A \cup B$ in (A, B) and get:

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

Exercise 2.10.3. Let Ω be a sample space and let A_1, A_2, \dots be events. Define $B_n = \bigcup_{i=n}^{\infty} A_i$ and $C_n = \bigcap_{i=n}^{\infty} A_i$

(a) Show that $B_1 \supset B_2 \supset \dots$ and $C_1 \subset C_2 \subset \dots$.

(b) Show that $\omega \in \bigcap_{n=1}^{\infty} B_n$ if and only if ω belongs to an infinite number of the events A_1, A_2, \dots .

(c) Show that $\omega \in \bigcup_{n=1}^{\infty} C_n$ if and only if ω belongs to all of the events A_1, A_2, \dots except possibly a finite number of those events.

Solution.

(a) By construction, $B_{n+1} = A_{n+1} \cup B_n$ and so $B_{n+1} \supset B_n$. Similarly, $C_{n+1} = A_{n+1} \cap C_n$ and so $C_{n+1} \subset C_n$.

(b)

- Assume ω belongs to an infinite number of the events, $\omega \in A_j$ for $j \in J(\omega)$. Then, for every n , there is a $m \geq n$ such that $m \in J(\omega)$, and so $\omega \in B_n$ for every n . This implies that $\omega \in \bigcap_{n=1}^{\infty} B_n$.
- Assume that $\omega \in \bigcap_{n=1}^{\infty} B_n$. Then, for every n , $\omega \in B_n$, so for every n there is a $m \geq n$ such that $\omega \in A_m$. This implies there is an infinite number of such events A_m .

(c)

Let's prove the contrapositive.

- Assume that ω does not belong to an infinite number of events A_i . Then, for every n , there is a $m \geq n$ such that $\omega \in A_m^c$, and so ω is not in C_n . Since ω is not in none of the C_n 's, it is not in the union of all C_n 's either.
- Assume that ω is not in the union of all C_n . This implies that ω is not in any event C_n . This implies that, for every n , there is a $m \geq n$ such that ω is not in A_m . This implies that there is an infinite number of such events A_m .

Exercise 2.10.4. Let $\{A_i : i \in I\}$ be a collection of events where I is an arbitrary index set. Show that

$$(\cup_{i \in I} A_i)^c = \cap_{i \in I} A_i^c \quad \text{and} \quad (\cap_{i \in I} A_i)^c = \cup_{i \in I} A_i^c$$

Hint: First prove this for $I = \{1, \dots, n\}$.

Solution.

We can prove the result directly by noting that every outcome ω belongs to or does not belong to both sides of each equality:

$$\omega \in (\cup_{i \in I} A_i)^c \tag{9}$$

$$\iff \text{not } (\omega \in \cup_{i \in I} A_i) \tag{10}$$

$$\iff \forall i \in I, \text{not } (\omega \in A_i) \tag{11}$$

$$\iff \forall i \in I, \omega \in A_i^c \tag{12}$$

$$\iff \omega \in \cap_{i \in I} A_i^c \tag{13}$$

and

$$\omega \in (\cap_{i \in I} A_i)^c \quad (14)$$

$$\iff \text{not } (\omega \in \cap_{i \in I} A_i) \quad (15)$$

$$\iff \text{not } (\forall i \in I, \omega \in A_i) \quad (16)$$

$$\iff \exists i \in I, \text{not } \omega \in A_i \quad (17)$$

$$\iff \exists i \in I, \omega \in A_i^c \quad (18)$$

$$\iff \omega \in \cup_{i \in I} A_i^c \quad (19)$$

Exercise 2.10.5. Suppose we toss a fair coin until we get exactly two heads. Describe the sample space S . What is the probability that exactly k tosses are required?

Solution. The sample space is a set of coin toss results sequences containing two heads, and ending in heads:

$$S = \left\{ (r_1, \dots, r_k) : r_i \in \{\text{head}, \text{tails}\}, |\{r_j = \text{head}\}| = 2, r_k = \text{head} \right\}$$

The probability of requiring exactly k tosses is 0 if $k < 2$, as there are no such sequences in the event space.

The probability of stopping after k tosses is the probability of obtaining exactly 1 head in the first $k - 1$ tosses, in a procedure that would not stop after any number of tosses, followed by the probability of getting a head in the k -th toss. This value is

$$\left((k-1) \left(\frac{1}{2} \right)^{k-1} \right) \left(\frac{1}{2} \right) = \frac{k-1}{2^k}$$

Note that, besides this combinatorial argument, we can verify that these probabilities do indeed add up to 1:

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k \quad (20)$$

$$\frac{d}{dx} \frac{1}{1-x} = \sum_{k=0}^{\infty} \frac{d}{dx} x^k \quad (21)$$

$$\frac{1}{(1-x)^2} = \sum_{k=0}^{\infty} kx^{k-1} \quad (22)$$

$$\frac{x}{(1-x)^2} = \sum_{k=0}^{\infty} kx^k \quad (23)$$

so, for $x = 1/2$, $\sum_{k=0}^{\infty} 2^{-k} k = 2$, and so

$$\sum_{k=0}^{\infty} \frac{k}{2^{k+1}} = 1$$

Exercise 2.10.6. Let $\Omega = \{1, 2, \dots\}$. Prove that there does not exist a uniform distribution on Ω , i.e. if $\mathbb{P}(A) = \mathbb{P}(B)$ whenever $|A| = |B|$ then \mathbb{P} cannot satisfy the axioms of probability.

Solution. Assume that such a distribution exists, and let $\mathbb{P}(\{1\}) = p$. Since the distribution is uniform, the probability associated with any set of size 1 is p , and the probability associated with any set of size n is np .

- If $p > 0$, then a finite set A of size $|A| = \lceil 2/p \rceil$ would have probability value $\mathbb{P}(A) = \lceil 2/p \rceil p \geq (2/p)p = 2$, which is greater than 1 -- a contradiction.
- If $p = 0$, then any finite set A must have $\mathbb{P}(A) = 0$. But then $\mathbb{P}(\Omega) = \sum_i \mathbb{P}(\{i\}) = \sum_i 0 = 0$, instead of 1 -- a contradiction.

Exercise 2.10.7. Let A_1, A_2, \dots be events. Show that

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

Hint: Define $B_n = A_n - \bigcup_{i=1}^{n-1} A_i$. Then show that the B_n are disjoint and that $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$.

Solution. Following the hint, let $B_n = A_n - \bigcup_{i=1}^{n-1} A_i$.

- Note that, for $i < j$, B_i and B_j are disjoint, since all elements of B_i must be elements of A_i , and all elements of A_i are explicitly excluded on the definition of B_j .
- Also note that $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$: $A_n = \bigcup_{i=1}^n B_i$ by construction, so $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} \bigcup_{i=1}^n B_i = \bigcup_{n=1}^{\infty} B_n$, since $B_i \cup B_i = B_i$ and we can include each B_i only once in the expression.

Now, we have:

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(B_n) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

since $B_n \cup \left(\bigcup_{i=1}^{n-1} A_i\right) = A_n$ and so $\mathbb{P}(B_n) \leq \mathbb{P}(A_n)$ for every n .

Exercise 2.10.8. Suppose that $\mathbb{P}(A_i) = 1$ for each i . Prove that

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = 1$$

Solution. Using the result from exercise 4,

$$\mathbb{P}(\cap_{i=1}^{\infty} A_i) = 1 - \mathbb{P}\left(\left(\cap_{i=1}^{\infty} A_i\right)^c\right) = 1 - \mathbb{P}\left(\cup_{i=1}^{\infty} A_i^c\right)$$

Using the result from exercise 7,

$$\mathbb{P}\left(\cup_{i=1}^{\infty} A_i^c\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i^c) = \sum_{i=1}^{\infty} (1 - \mathbb{P}(A_i)) = \sum_{i=1}^{\infty} 0 = 0$$

so the equality holds, since a probability is non-negative. Therefore,

$$\mathbb{P}(\cap_{i=1}^{\infty} A_i) = 1 - \mathbb{P}(\cup_{i=1}^{\infty} A_i^c) = 1 - 0 = 1$$

Exercise 2.10.9. For fixed B such that $\mathbb{P}(B) > 0$, show that $\mathbb{P}(\cdot|B)$ satisfies the axioms of probability.

Solution.

- Axiom 1: $\mathbb{P}(\cdot|B) = \frac{\mathbb{P}(\cdot \cap B)}{\mathbb{P}(B)} \geq 0$, since $\mathbb{P}(\cdot \cap B) \geq 0$.
- Axiom 2: $\mathbb{P}(\Omega|B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$.
- Axiom 3: Assuming A_1, A_2, \dots are disjoint,

$$\mathbb{P}\left(\cup_{i=1}^{\infty} A_i|B\right) = \frac{\mathbb{P}\left(B \cap \left(\cup_{i=1}^{\infty} A_i\right)\right)}{\mathbb{P}(B)} = \frac{\mathbb{P}\left(\cup_{i=1}^{\infty} (A_i \cap B)\right)}{\mathbb{P}(B)} = \frac{\sum_{i=1}^{\infty} \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_{i=1}^{\infty} \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_{i=1}^{\infty} \mathbb{P}(A_i|B)$$

Exercise 2.10.10. You have probably heard it before. Now you can solve it rigorously. It is called the "Monty Hall Problem". A prize is placed at random between one of three doors. You pick a door. To be concrete, let's suppose you always pick door 1. Now Monty Hall chooses one of the other two doors, opens it and shows to you that it is empty. He then gives you the opportunity to keep your door or switch to the other unopened door. Should you stay or switch? Intuition suggests it doesn't matter. The correct answer is that you should switch. Prove it. It will help to specify the sample space and the relevant events carefully. Thus write $\Omega = \{(\omega_1, \omega_2) : \omega_i \in \{1, 2, 3\}\}$ where ω_1 is where the prize is and ω_2 is the door Monty opens.

Solution. Following the provided notation, the event space is

$$\Omega = \{(1, 2), (1, 3), (2, 3), (3, 2)\}$$

The probability and the reward associated with switching for each outcome are:

ω	\mathbb{P}	R
$(1, 2)$	$\frac{1}{3} \frac{1}{2}$	0
$(1, 3)$	$\frac{1}{3} \frac{1}{2}$	0
$(2, 3)$	$\frac{1}{2}$	1
$(3, 2)$	$\frac{1}{2}$	1

Therefore,

$$\mathbb{P}[R|\omega_2 = 2] = \frac{\mathbb{P}(\{(3, 2)\})}{\mathbb{P}(\{(3, 2), (1, 2)\})} \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{3} \frac{1}{2}} = \frac{2}{3}$$

and, similarly, $\mathbb{P}[R|\omega_3 = 3]$, and so $\mathbb{P}[R] = \frac{2}{3}$.

Exercise 2.10.11. Suppose that A and B are independent events. Show that A^c and B^c are independent events.

Solution.

$$\mathbb{P}(A^c B^c) = \mathbb{P}((A \cup B)^c) = 1 - \mathbb{P}(A \cup B) \quad (24)$$

$$= 1 - (\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)) \quad (25)$$

$$= 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A)\mathbb{P}(B) \quad (26)$$

$$= 1 - (1 - \mathbb{P}(A^c)) - (1 - \mathbb{P}(B^c)) + (1 - \mathbb{P}(A^c))(1 - \mathbb{P}(B^c)) \quad (27)$$

$$= \mathbb{P}(A^c)\mathbb{P}(B^c) \quad (28)$$

Exercise 2.10.12. There are three cards. The first card is green on both sides, the second is red on both sides, and the third is green on one side and red on the other. We choose a card at random and we see one side (also chosen at random). If the side we see is green, what is the probability that the other side is also green? Many people intuitively answer 1/2. Show that the correct answer is 2/3.

Solution. There are 6 potential card sides to be chosen, all with equal probability, of which only 3 are green -- one belongs to the red / green card, and two belong to the green / green card. The probability that the other side is also green is the probability that the a side on the green / green card was chosen, which is 2 / 3.

Exercise 2.10.13. Suppose a fair coin is tossed repeatedly until both a head and a tail have appeared at least once.

(a) Describe the sample space Ω .

(b) What is the probability that three tosses will be required?

Solution.

(a). The sample space consists of the sequence of k identical coin toss results and a coin toss result with the opposite value,

$$\Omega = \{(r_1, \dots, r_k, r_{k+1}) : r_i \in \{\text{head}, \text{tails}\}, r_1 = \dots = r_k \neq r_{k+1}\}$$

(b) Exactly 3 tosses will be required if the first 3 results are (h, h, t) or (t, t, h) .

If we map all infinite coin toss sequences to Ω by truncating it whenever the stop condition occurs, the probability of a (single-outcome) event in Ω is the same as the probability of all outcomes mapped into it. In particular, the probability of a sequence with its first 3 symbols being a specific sequence is $1/8$, and so the probability of the desired outcome is $1/8 + 1/8 = 1/4$.

Exercise 2.10.14. Show that if $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$ then A is independent of every other event. Show that if A is independent of itself then $\mathbb{P}(A)$ is either 0 or 1.

Solution.

If $\mathbb{P}(A) = 0$, then $\mathbb{P}(AB) = \mathbb{P}(A) - \mathbb{P}(A - B) = 0 - \mathbb{P}(A - B) \leq 0$, and since probabilities are non-negative we must have $\mathbb{P}(AB) = 0$. Therefore $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B) = 0$ for all events B , and A is independent of every other event.

If $\mathbb{P}(A) = 1$, then $\mathbb{P}(A^c) = 0$, and so A^c and B are independent for every other event B . Then, from the result in exercise 10, A is also independent from every other event B^c -- which covers all potential events, since every event has a complement.

If A is independent of itself, $\mathbb{P}(AA) = \mathbb{P}(A)\mathbb{P}(A)$, so $\mathbb{P}(A) = \mathbb{P}(A)^2$ or $\mathbb{P}(A)(\mathbb{P}(A) - 1) = 0$. Therefore $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

Exercise 2.10.15. The probability that a child has blue eyes is $1/4$. Assume independence between children. Consider a family with 5 children.

(a) If it is known that at least one child has blue eyes, what is the probability that at least 3 children have blue eyes?

(b) If it is known that the youngest child has blue eyes, what is the probability that at least 3 children have blue eyes?

Solution.

(a) Represent the sample space as

$$\Omega = \{(x_1, x_2, x_3, x_4, x_5) : x_i \in \{0, 1\}\}$$

where $x_i = 1$ if the i -th child (youngest to oldest) has blue eyes.

- "At least one child has blue eyes" is the event $A = \Omega - \{(0, 0, 0, 0, 0)\}$.
- "At least 3 children have blue eyes" is the event B with 3 children with blue eyes, 4 children with blue eyes, or 5 children with blue eyes.
- The intersection of these events is $BA = B$.

Let $p = 1/4$ be the probability at a given child will have blue eyes. The desired probability is then:

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(BA)}{\mathbb{P}(A)} = \frac{\binom{5}{3}p^3(1-p)^2 + \binom{5}{4}p^4(1-p) + \binom{5}{5}p^5}{1 - (1-p)^5} = \frac{106}{781} \approx 0.1357$$

(b)

- "The youngest child has blue eyes" is the event $C = \{\omega = (1, x_2, x_3, x_4, x_5) : \omega \in \Omega\}$.
- The intersection of events B and C is BC , the set of outcomes starting with 1 and having the other 4 dimensions having 2, 3, or 4 values 1;
 $BC = \{\omega = (1, x_2, x_3, x_4, x_5) : \omega \in \Omega, x_2 + x_3 + x_4 + x_5 \geq 2\}$.

The desired probability is then

$$\mathbb{P}(B|C) = \frac{\mathbb{P}(BC)}{\mathbb{P}(C)} = \frac{p \left(\binom{4}{2}p^2(1-p)^2 + \binom{4}{3}p^3(1-p) + \binom{4}{4}p^4 \right)}{p} = \frac{67}{256} \approx 0.2617$$

Exercise 2.10.16. Show that

$$\mathbb{P}(ABC) = \mathbb{P}(A|BC)\mathbb{P}(B|C)\mathbb{P}(C)$$

Solution.

$$\mathbb{P}(A|BC)\mathbb{P}(B|C)\mathbb{P}(C) = \frac{\mathbb{P}(ABC)}{\mathbb{P}(BC)} \frac{\mathbb{P}(BC)}{\mathbb{P}(C)} \mathbb{P}(C) = \mathbb{P}(ABC)$$

Exercise 2.10.17. Suppose k events for a partition of the sample space Ω , i.e. they are disjoint and $\cup_{i=1}^k A_i = \Omega$. Assume that $\mathbb{P}(B) > 0$. Prove that if $\mathbb{P}(A_1|B) < \mathbb{P}(A_1)$ then $\mathbb{P}(A_i|B) > \mathbb{P}(A_i)$ for some $i = 2, \dots, k$.

Solution.

We have

$$B = B\Omega = B \left(\cup_{i=1}^k A_i \right) = \cup_{i=1}^k A_i B$$

and so

$$\mathbb{P}\left(\bigcup_{i=1}^k A_i|B\right) = \mathbb{P}(B) \iff \sum_{i=1}^k \frac{\mathbb{P}(A_i B)}{\mathbb{P}(B)} = 1 \iff \sum_{i=1}^k \mathbb{P}(A_i|B) = 1 \iff \sum_{i=1}^k \mathbb{P}(A_i|B) = \sum_{i=1}^k \mathbb{P}(A_i)$$

If we assume that $\mathbb{P}(A_i|B) \leq \mathbb{P}(A_i)$ for all i and $\mathbb{P}(A_1|B) < \mathbb{P}(A_1)$, then we must have $\sum_{i=1}^k \mathbb{P}(A_i|B) < \sum_{i=1}^k \mathbb{P}(A_i)$, a contradiction. Therefore the desired statement must hold.

Exercise 2.10.18. Suppose that 30% of computer users use a Macintosh, 50% use Windows and 20% use Linux. Suppose that 65% of the Mac users have succumbed to a computer virus, 82% of the Windows users get the virus and 50% of the Linux users get the virus. We select a person at random and learn that her system was infected with the virus. What is the probability that she is a Windows user?

Solution. The event space can be described as:

outcome	probability
Mac, no virus	30% * 35%
Mac, virus	30% * 65%
Windows, no virus	50% * 18%
Windows, virus	50% * 82%
Linux, no virus	20% * 50%
Linux, virus	20% * 50%

The desired conditional probability is

$$\mathbb{P}(\text{Windows}|\text{virus}) = \frac{\mathbb{P}(\text{Windows, virus})}{\mathbb{P}(\text{virus})} = \frac{0.50 \cdot 0.82}{0.30 \cdot 0.65 + 0.50 \cdot 0.82 + 0.20 \cdot 0.50} \approx 0.5816$$

Exercise 2.10.19. A box contains 5 coins and each has a different probability of showing heads. Let p_1, \dots, p_5 denote the probability of heads on each coin. Suppose that

$$p_1 = 0, \quad p_2 = 1/4, \quad p_3 = 1/2, \quad p_4 = 3/4, \quad \text{and } p_5 = 1$$

Let H denote "heads is obtained" and let C_i denote the event that coin i is selected.

(a) Select a coin at random and toss it. Suppose a head is obtained. What is the posterior probability that coin i was selected ($i = 1, \dots, 5$)? In other words, find $\mathbb{P}(C_i|H)$ for $i = 1, \dots, 5$.

(b) Toss the coin again. What is the probability of another head? In other words find $\mathbb{P}(H_2|H_1)$ where H_j means "heads on toss j ".

(c) Find $\mathbb{P}(C_i|B_4)$ where B_4 means "first head is obtained on toss 4".

Solution.

(a) We have:

$$\mathbb{P}(C_i|H) = \frac{\mathbb{P}(C_i H)}{\mathbb{P}(H)} = \frac{\mathbb{P}(C_i H)}{\sum_j \mathbb{P}(C_j H)} = \frac{\mathbb{P}(C_i) \mathbb{P}(H|C_i)}{\sum_j \mathbb{P}(C_j) \mathbb{P}(H|C_j)}$$

Assuming that the coin selection is uniformly random, $\mathbb{P}(C_i) = 1/5$ for $i = 1, \dots, 5$, and the above simplifies to

$$\frac{\mathbb{P}(H|C_i)}{\sum_j \mathbb{P}(H|C_j)} = \frac{p_i}{\sum_j p_j}$$

Therefore,

$$\mathbb{P}(C_1|H) = 0 \quad \mathbb{P}(C_2|H) = 1/10 \quad \mathbb{P}(C_3|H) = 1/5 \quad \mathbb{P}(C_4|H) = 3/10 \quad \mathbb{P}(C_5|H) = 2/5$$

(b) We have:

$$\mathbb{P}(H_2|H_1) = \frac{\mathbb{P}(H_2 H_1)}{\mathbb{P}(H_1)} = \frac{\sum_j \mathbb{P}(C_j H_1 H_2)}{\sum_j \mathbb{P}(C_j H_1)} = \frac{\sum_j (1/5) p_j^2}{\sum_j (1/5) p_j} = \frac{3}{16}$$

(c) We have:

$$\mathbb{P}(C_i|B_4) = \frac{\mathbb{P}(C_i B_4)}{\mathbb{P}(B_4)} = \frac{\mathbb{P}(C_i B_4)}{\sum_j \mathbb{P}(C_j B_4)}$$

But $\mathbb{P}(C_i B_4) = (1/5)(1 - p_i)^3 p_i$ -- selecting coin i , then obtaining 3 tails followed by a head on that coin -- so

$$\mathbb{P}(C_1|B_4) = 0 \quad \mathbb{P}(C_2|B_4) = \frac{27}{46} \quad \mathbb{P}(C_3|B_4) = \frac{8}{23} \quad \mathbb{P}(C_4|B_4) = \frac{3}{46} \quad \mathbb{P}(C_5|B_4) = 0$$

Exercise 2.10.20 (Computer Experiment). Suppose a coin has probability p of falling heads. If we flip the coin many times, we would expect the proportion of heads to be near p . We will make this formal later. Take $p = .3$ and $n = 1000$ and simulate n coin flips. Plot the proportion of heads as a function of n . Repeat for $p = .03$.

```
In [1]: import numpy as np

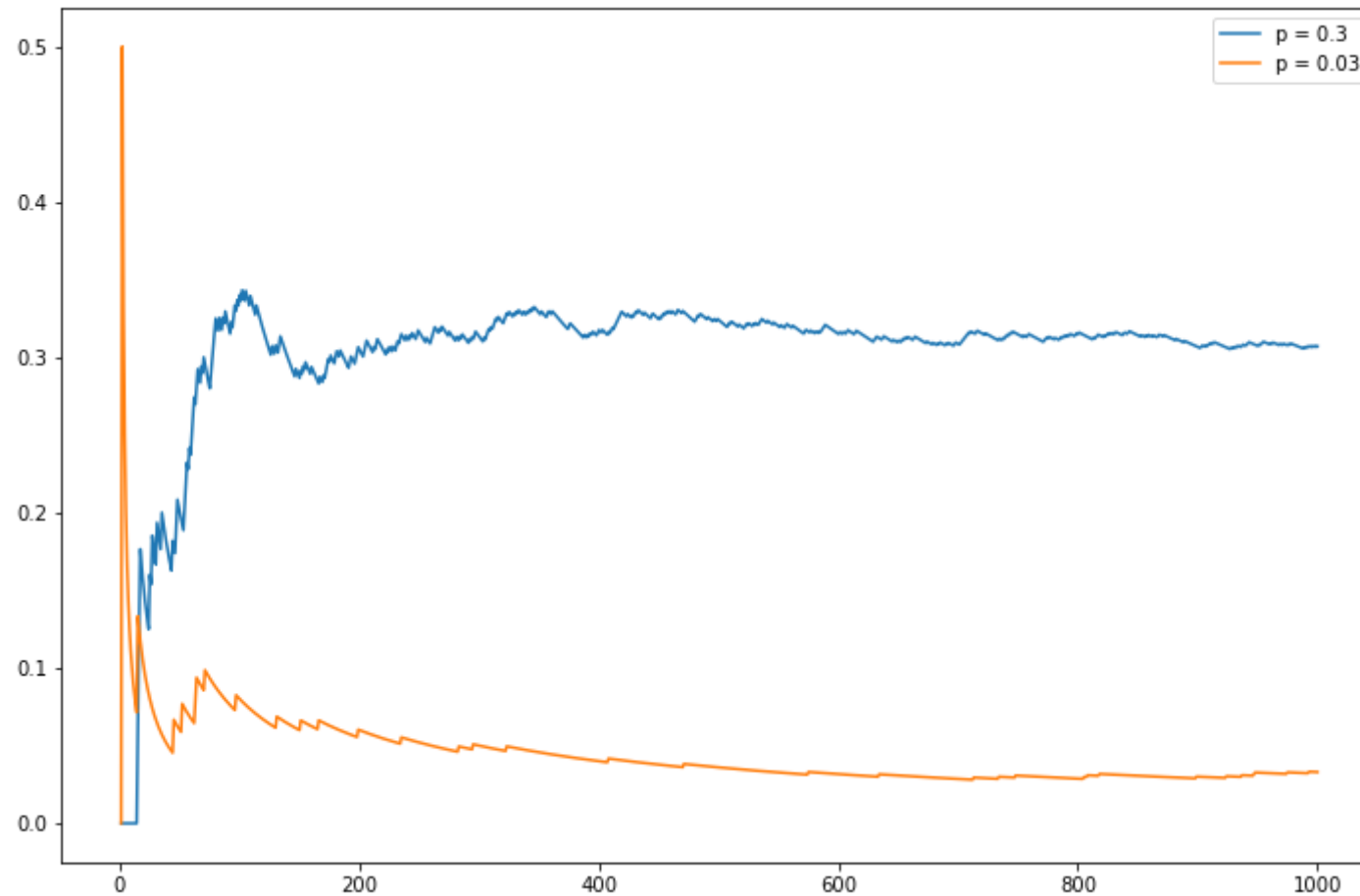
np.random.seed(0)

n = 1000
X1 = np.where(np.random.uniform(low=0, high=1, size=n) < 0.3, 1, 0)
X2 = np.where(np.random.uniform(low=0, high=1, size=n) < 0.03, 1, 0)
```

```
In [2]: import matplotlib.pyplot as plt
%matplotlib inline

nn = np.arange(1, n + 1)

plt.figure(figsize=(12, 8))
plt.plot(nn, np.cumsum(X1) / nn, label='p = 0.3')
plt.plot(nn, np.cumsum(X2) / nn, label='p = 0.03')
plt.legend(loc='upper right')
plt.show()
```



Exercise 2.10.21 (Computer Experiment). Suppose we flip a coin n times and let p denote the probability of heads. Let X be the number of heads. We call X a binomial random variable which is discussed in the next chapter. Intuition suggests that X will be close to np . To see if this is true, we can repeat this experiment many times and average the X values. Carry out a simulation and compare the averages of the X 's to np . Try this for $p = .3$ and $n = 10, 100, 1000$.

```
In [3]: import numpy as np

        B = 50000

        np.random.seed(0)

        X1 = np.random.binomial(n=10, p=0.3, size=B)
```

```
X2 = np.random.binomial(n=100, p=0.3, size=B)
X3 = np.random.binomial(n=1000, p=0.3, size=B)
```

```
In [4]: print('X1 mean: %.3f' % X1.mean())
print('X1 np:   %.3f' % (0.3 * 10))
print()
print('X2 mean: %.3f' % X2.mean())
print('X2 np:   %.3f' % (0.3 * 100))
print()
print('X3 mean: %.3f' % X3.mean())
print('X3 np:   %.3f' % (0.3 * 1000))
```

```
X1 mean: 2.989
X1 np:   3.000
```

```
X2 mean: 30.023
X2 np:   30.000
```

```
X3 mean: 299.903
X3 np:   300.000
```

```
In [5]: import matplotlib.pyplot as plt
%matplotlib inline

plt.figure(figsize=(12, 8))

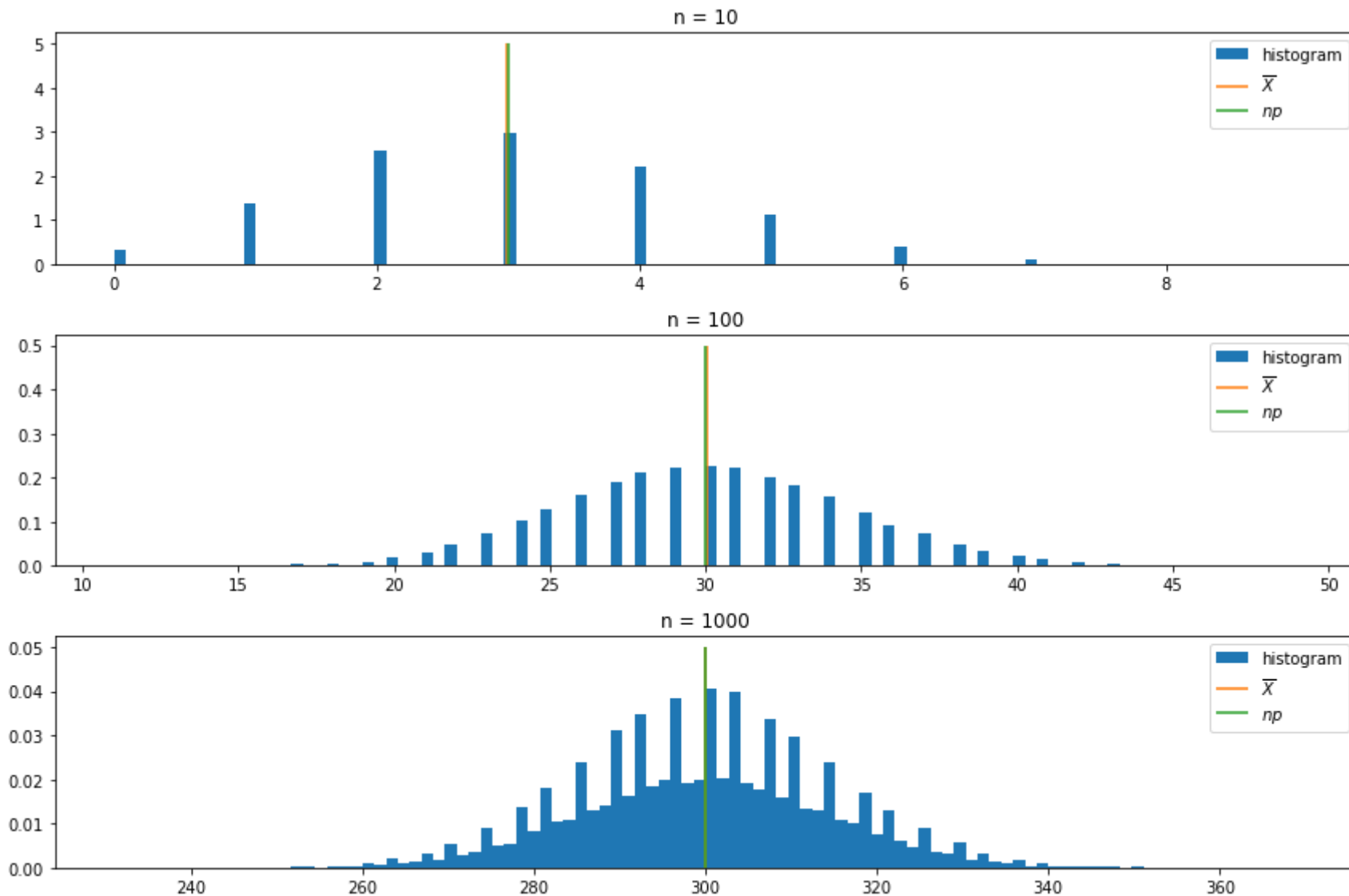
ax = plt.subplot(3, 1, 1)
ax.hist(X1, density=True, bins=100, label='histogram', color='C0')
ax.vlines(X1.mean(), ymin=0, ymax=5, label=r'$\overline{X}$', color='C1')
ax.vlines(0.3 * 10, ymin=0, ymax=5, label=r'$np$', color='C2')
ax.legend(loc='upper right')
ax.set_title('n = 10')

ax = plt.subplot(3, 1, 2)
ax.hist(X2, density=True, bins=100, label='histogram', color='C0')
ax.vlines(X2.mean(), ymin=0, ymax=0.5, label=r'$\overline{X}$', color='C1')
ax.vlines(0.3 * 100, ymin=0, ymax=0.5, label=r'$np$', color='C2')
ax.legend(loc='upper right')
ax.set_title('n = 100')

ax = plt.subplot(3, 1, 3)
ax.hist(X3, density=True, bins=100, label='histogram', color='C0')
ax.vlines(X3.mean(), ymin=0, ymax=0.05, label=r'$\overline{X}$', color='C1')
ax.vlines(0.3 * 1000, ymin=0, ymax=0.05, label=r'$np$', color='C2')
ax.legend(loc='upper right')
```

```
ax.set_title('n = 1000')

plt.tight_layout()
plt.show()
```



Exercise 2.10.22 (Computer Experiment). Here we will get some experience simulating conditional probabilities. Consider tossing a fair die. Let $A = \{2, 4, 6\}$ and $B = \{1, 2, 3, 4\}$. Then $\mathbb{P}(A) = 1/2$, $\mathbb{P}(B) = 2/3$ and $\mathbb{P}(AB) = 1/3$. Since $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$, the events A and B are

independent. Simulate draws from the sample space and verify that $\hat{P}(AB) = \hat{P}(A)\hat{P}(B)$ where \hat{P} is the proportion of times an event occurred in the simulation. Now find two events A and B that are not independent. Compute $\hat{P}(A)$, $\hat{P}(B)$ and $\hat{P}(AB)$. Compare the calculated values to their theoretical values. Report your results and interpret.

```
In [6]: import numpy as np

np.random.seed(0)

B = 10000
results = np.random.randint(low=1, high=7, size=B)
```

```
In [7]: A_hat = np.isin(results, [2, 4, 6])
B_hat = np.isin(results, [1, 2, 3, 4])
AB_hat = np.isin(results, [2, 4])
```

```
In [8]: import matplotlib.pyplot as plt
%matplotlib inline

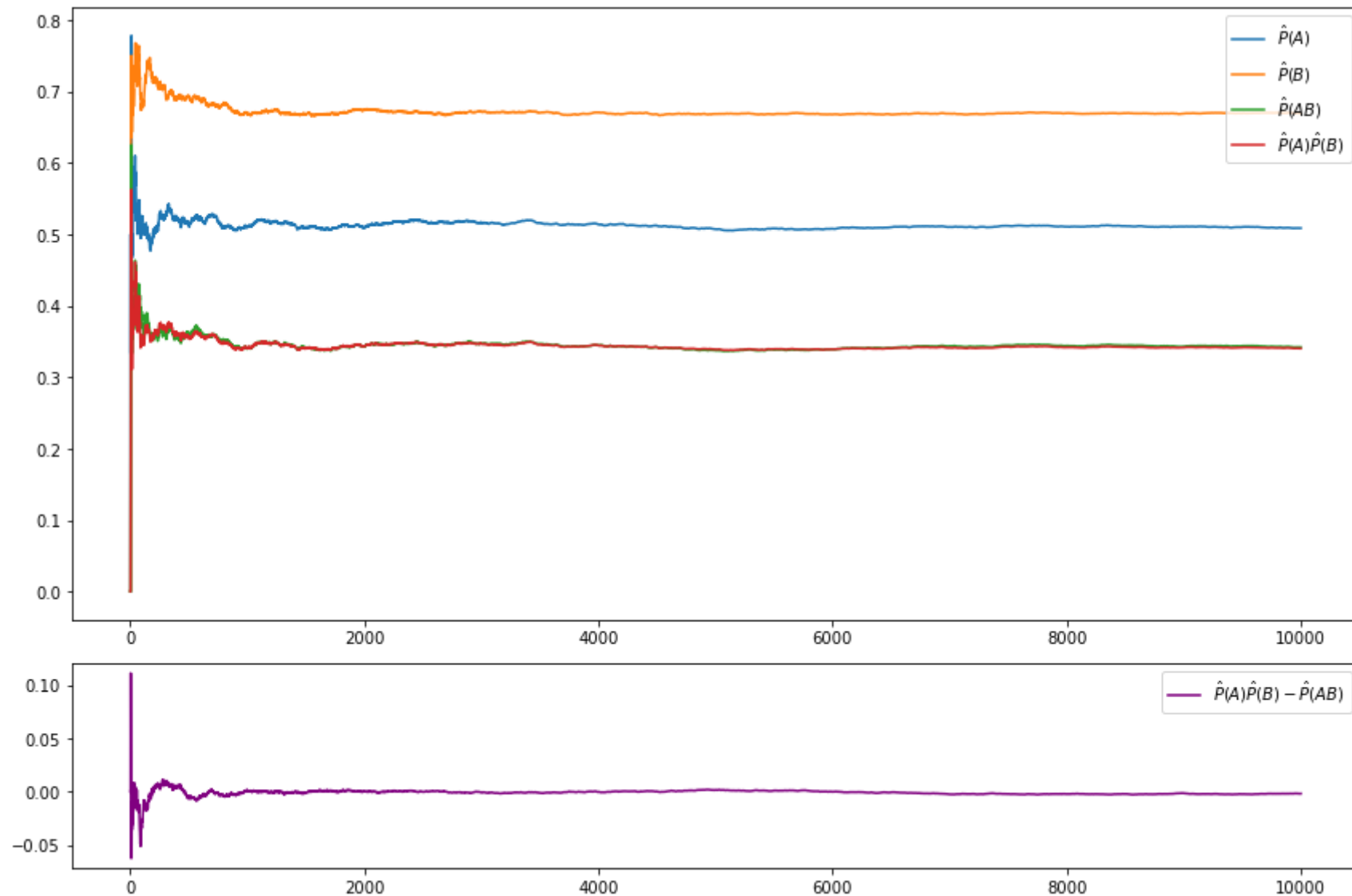
nn = np.arange(1, B + 1)

f, (a0, a1) = plt.subplots(2, 1, gridspec_kw={'height_ratios': [3, 1]}, figsize=(12, 8))

a0.plot(nn, np.cumsum(A_hat) / nn, label=r'$\hat{P}(A)$')
a0.plot(nn, np.cumsum(B_hat) / nn, label=r'$\hat{P}(B)$')
a0.plot(nn, np.cumsum(AB_hat) / nn, label=r'$\hat{P}(AB)$')
a0.plot(nn, np.cumsum(A_hat) * np.cumsum(B_hat) / (nn * nn), label=r'$\hat{P}(A) \hat{P}(B)$')
a0.legend(loc='upper right')

a1.plot(nn, np.cumsum(A_hat) * np.cumsum(B_hat) / (nn * nn) - np.cumsum(AB_hat) / nn,
        label=r'$\hat{P}(A) \hat{P}(B) - \hat{P}(AB)$', color='purple')
a1.legend(loc='upper right')

plt.tight_layout()
plt.show()
```



For our own choice of non-independent events, let $A = \{2, 4, 6\}$ and $B = \{2, 4, 5\}$. Then $\mathbb{P}(A) = \mathbb{P}(B) = 1/2$ but $\mathbb{P}(AB) = 1/3$.

```
In [9]: A_hat = np.isin(results, [2, 4, 6])
        B_hat = np.isin(results, [2, 4, 5])
        AB_hat = np.isin(results, [2, 4])
```

```
In [10]: import matplotlib.pyplot as plt
```

```

%matplotlib inline

nn = np.arange(1, B + 1)

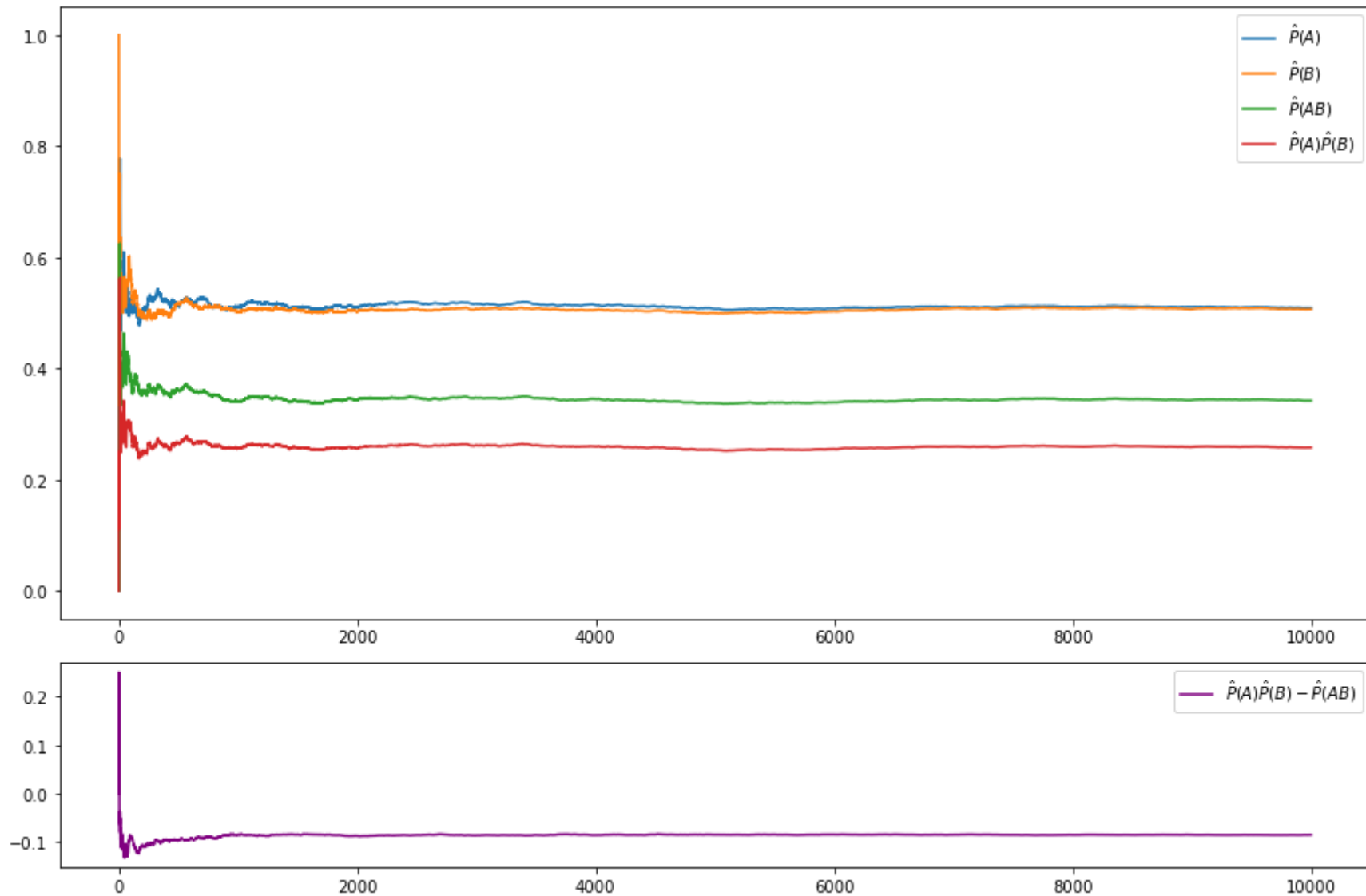
f, (a0, a1) = plt.subplots(2, 1, gridspec_kw={'height_ratios': [3, 1]}, figsize=(12, 8))

a0.plot(nn, np.cumsum(A_hat) / nn, label=r'$\hat{P}(A)$')
a0.plot(nn, np.cumsum(B_hat) / nn, label=r'$\hat{P}(B)$')
a0.plot(nn, np.cumsum(AB_hat) / nn, label=r'$\hat{P}(AB)$')
a0.plot(nn, np.cumsum(A_hat) * np.cumsum(B_hat) / (nn * nn), label=r'$\hat{P}(A) \hat{P}(B)$')
a0.legend(loc='upper right')

a1.plot(nn, np.cumsum(A_hat) * np.cumsum(B_hat) / (nn * nn) - np.cumsum(AB_hat) / nn,
        label=r'$\hat{P}(A) \hat{P}(B) - \hat{P}(AB)$', color='purple')
a1.legend(loc='upper right')

plt.tight_layout()
plt.show()

```



As noted, the estimates converge to the theoretical value -- and the product of the estimates only converges to the estimate of the joint event in the scenario where the events are independent.