All of Statistics

My proposed solutions to the book:

All of Statistics - A Concise Course in Statistical Inference.

I really needed a proper refresher on statistics.

Let me know if you find any mistakes!

Contents

1 Chapter 1 - Probability

Exercises

1.1
Proving the Continuity of Probabilities.

1.8 Theorem (Continuity of Probabilities). If $A_n \to A$ then

$$\mathbb{P}(A_n) \to \mathbb{P}(A)$$

as $n \to \infty$.

Proof. Suppose that A_n is monotone increasing so that $A_1 \subset A_2 \subset \dots$. Let $A = \lim_{n\to\infty} A_n = \bigcup_{i=1}^{\infty} A_i$. Define:

$$\begin{aligned} B_1 &= A_1 \\ B_2 &= \{\omega \in \Omega \ : \ \omega \in A_2, \omega \notin A_1\} \\ B_3 &= \{\omega \in \Omega \ : \ \omega \in A_3, \omega \notin A_2, \omega \notin A_1\} \\ &\vdots \end{aligned}$$

(1) Showing that B_i are disjoint sets.
We have $B_1 = A_1$. Since $B_2 = \{\omega \in \Omega \ : \ \omega \in A_2, \omega \notin A_1\}$, we can rewrite this as $B_2 = A_2 - A_1$ by definition of set difference.

Since $B_3 = \{\omega \in \Omega \ : \ \omega \in A_3, \omega \notin A_2, \omega \notin A_1\}$ and since $A_1 \subset A_2$, we can rewrite this as $B_3 = A_3 - (A_1 \bigcup A_2)$. In general; $B_k = A_k - (\bigcup_{i=1}^{k-1} A_i)$.

Assuming some random sets B_m, B_p for some $m, p \in \mathbb{N}$. Without loss of generality, we assume $m > p$. Then:

$$\begin{aligned} B_m \bigcap B_p &= \Big(A_m - (\bigcup_{i=1}^{m-1} A_i)\Big) \bigcap \Big(A_p - (\bigcup_{i=1}^{p-1} A_i)\Big) \\ &= \Big(A_m \bigcap (\bigcup_{i=1}^{m-1} A_i)^c\Big) \bigcap \Big(A_p \bigcap (\bigcup_{i=1}^{p-1} A_i)^c\Big) \end{aligned}$$

DeMorgan's law

$$= \Big(A_m \cap A_1^c \cap A_2^c \cap \dots \cap A_p^c \cap \dots \cap A_{m-1}^c\Big) \bigcap \Big(A_p \cap A_1^c \cap A_2^c \cap \dots \cap A_{p-1}^c\Big)$$

Reshuffling terms and repeated use of $C \cap C = C$

$$\begin{aligned} &= A_m \cap A_1^c \cap A_2^c \cap \dots \cap \underbrace{A_p^c \cap A_p}_{=\emptyset} \cap \dots \cap A_{m-1}^c\Big) \\ &= \emptyset \end{aligned}$$

Since $B_m \cap B_p = \emptyset$, they are disjoint. (*Quite certain this is a correct argument, but could have made it a bit easier by going directly to the $A_3 \cap A_2^c \cap A_1^c$ version of the sets.*)

(2) Showing that

$$A_n = \bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} B_i$$

for each n. Showing this with an induction argument. Note that $A_1 = B_1$, and $A_2 \cup A_1 = A_2$ since $A_1 \subset A_2$. Note also that:

$$\begin{aligned}
B_2 \cup B_1 &= (A_2 \cap A_1^c) \cup A_1 \\
&= (A_2 \cup A_1) \cap (A_1^c \cup A_1) \\
&= (A_2 \cup A_1) \cap (\Omega) \\
&= A_2 \cup A_1 \\
&= A_2
\end{aligned}$$

So this is true for $n = 2$. Assume that $A_k = \cup_{i=1}^k A_i = \cup_{i=1}^k B_i$ for $k \in \mathbb{N}$. Showing that this applies to $k + 1$.

$$\bigcup_{i=1}^{k+1} A_i = A_{k+1} \cup \Big(\bigcup_{i=1}^{k} A_i \Big) = A_{k+1} \cup A_k = A_{k+1}$$

$$\begin{aligned}
\bigcup_{i=1}^{k+1} B_i &= B_{k+1} \cup \Big(\bigcup_{i=1}^{k} B_i \Big) \\
&= B_{k+1} \cup A_k \\
&= \Big[A_{k+1} \cap \Big(A_k^c \cap \ldots \cap A_1^c \Big) \Big] \cup A_k \\
&= \Big[A_{k+1} \cap \Big(A_k \cup \ldots \cup A_1 \Big)^c \Big] \cup A_k \\
&= \Big[A_{k+1} \cap A_k^c \Big] \cup A_k \\
&= (A_{k+1} \cup A_k) \cap (A_k^c \cup A_k) \\
&= (A_{k+1} \cup A_k) \cap \Omega \\
&= A_{k+1} \cup A_k \\
&= A_{k+1}
\end{aligned}$$

Result verified by induction argument.

(3) Showing that

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i. \qquad (\diamondsuit)$$

By definition:

$$A = \lim_{n\to\infty} A_n = \bigcup_{i=1}^{\infty} A_i.$$

By step (2):

$$A = \lim_{n\to\infty} A_n = \lim_{n\to\infty} \bigcup_{i=1}^{n} A_i = \lim_{n\to\infty} \bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{\infty} B_i.$$

Hence, $(\diamondsuit)$ is satisfied.

1.2
Proving some well known results by using the axioms:

$$\mathbb{P}(A) \geq 0, \; \forall A \subset \Omega \tag{A.1}$$

$$\mathbb{P}(\Omega) = 1, \tag{A.2}$$

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) \tag{A.3}$$

where $A_1, A_2, \ldots$ are disjoint in $(A.3)$.

• **Claim**: $\mathbb{P}(\emptyset) = 0$.
Proof. Since $\Omega \cap \emptyset = \emptyset$ and the empty set is disjoint with itself, we can make set $E_1 = \Omega$ and $E_k = \emptyset$ for all $k \geq 2$. Now assume for contradiction that $\mathbb{P}(\emptyset) = a > 0$ for some $a \in \mathbb{R}$. Then by $(A.3)$:

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) \overset{A.3}{=} \sum_{i=1}^{\infty} \mathbb{P}(E_i) = \mathbb{P}(\Omega) + \sum_{i=2}^{\infty} \mathbb{P}(\emptyset) = 1 + \sum_{i=2}^{\infty} a = \infty$$

Now instead of using $(A.3)$ we use that the infinite set of E_i becomes Ω:

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \mathbb{P}(\Omega) \overset{A.2}{=} 1$$

We have reached a contradiction. This shows that $\mathbb{P}(\emptyset) = 0$. $\square$

• **Claim**: $A \cap B = \emptyset \implies \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
Proof. Set $E_1 = A$ and $E_2 = B$ and $E_k = \emptyset$ for all $k \geq 3$. Then:

$$\mathbb{P}(A \cup B) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) \overset{A.3}{=} \sum_{i=1}^{\infty} \mathbb{P}(E_i) = \mathbb{P}(A) + \mathbb{P}(B) + \sum_{k=3}^{\infty} \mathbb{P}(\emptyset) = \mathbb{P}(A) + \mathbb{P}(B) + 0 = \mathbb{P}(A) + \mathbb{P}(B)$$

$\square$

With this result, we can apply $(A.3)$ indirectly to any finite sum of disjoint sets. All other sets are set to $\emptyset$ and then (A.3) is applied. Not giving a formal argument, but if A, B, C are mutually disjoint we can set $E_k = \emptyset$ for all $k \geq 4$.

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) \overset{A.3}{=} \sum_{i=1}^{\infty} \mathbb{P}(E_i) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) + 0 = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C)$$

• **Claim**: $A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$.
Proof. Since $A \subset B$ we can split B into two disjoint parts: $B = A \cup B - A$. By axiom (A.3):

$$\mathbb{P}(B) = \mathbb{P}(A \cup B - A) \overset{A.3}{=} \mathbb{P}(A) + \mathbb{P}(B - A) \overset{A.1}{\geq} \mathbb{P}(A) \implies \mathbb{P}(A) \leq \mathbb{P}(B)$$ $\square$

• **Claim**: $0 \leq \mathbb{P}(A) \leq 1$.
Proof. Since $A \subset \Omega$ and by the previous proof: $\mathbb{P}(A) \leq \mathbb{P}(\Omega) = 1$ Combining this with axiom (A.1):

$$0 \overset{A.1}{\leq} \mathbb{P}(A) \leq \mathbb{P}(\Omega) = 1 \implies 0 \leq \mathbb{P}(A) \leq 1$$ $\square$

• **Claim**: $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
Proof. Since $A \cup A^c$ are disjoint, we get by finite version of (A.3):

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

Since $\Omega = A \cup A^c$ we also get by (A.2):

$$\mathbb{P}(A \cup A^c) = \mathbb{P}(\Omega) = 1$$

Putting them together:

$$\mathbb{P}(A) + \mathbb{P}(A^c) = 1 \implies \mathbb{P}(A^c) = 1 - \mathbb{P}(A) \qquad \square$$

1.3
Ω is a sample space and $A_1, A_2, \ldots$ are events. We define:

$$B_n = \bigcup_{i=n}^{\infty} A_i, \qquad C_n = \bigcap_{i=n}^{\infty} A_i$$

(a) **Claim**: $B_1 \supset B_2 \supset \ldots$
Proof. This follows directly from the definition of B_n. First we see that:

$$B_1 = \bigcup_{i=1}^{\infty} A_i = A_1 \cup \left(\bigcup_{i=2}^{\infty} A_i \right) \supset \bigcup_{i=2}^{\infty} A_i = B_2, \implies B_1 \supset B_2$$

The same argument can be used for some $k \in \mathbb{N}$.

$$B_k = \bigcup_{i=k}^{\infty} A_i = A_k \cup \left(\bigcup_{i=k+1}^{\infty} A_i \right) \supset \bigcup_{i=k+1}^{\infty} A_i = B_{k+1}, \implies B_k \supset B_{k+1}$$

From this, it follows that $B_1 \supset B_2 \supset \ldots$. $\square$

Claim: $C_1 \subset C_2 \subset \ldots$
Proof. This also follows directly from the definition. First note that:

$$C_1 = \bigcap_{i=1}^{\infty} A_i = A_1 \cap \left(\bigcap_{i=2}^{\infty} A_i \right) \subset \bigcap_{i=2}^{\infty} A_i = C_2 \implies C_1 \subset C_2$$

And, in general, for some $k \in \mathbb{N}$.

$$C_k = \bigcap_{i=k}^{\infty} A_i = A_k \cap \left(\bigcap_{i=k+1}^{\infty} A_i \right) \subset \bigcap_{i=k+1}^{\infty} A_i = C_{k+1} \implies C_k \subset C_{k+1}$$

This shows that $C_1 \subset C_2 \subset \ldots$. $\square$

(b) **Claim**:

$$\omega \in \bigcap_{n=1}^{\infty} B_n \iff \omega \in A_1, A_2, \ldots$$

Proof.
$\Rightarrow$) Assume $\omega \in \bigcap_{n=1}^{\infty} B_n$, which means $\omega \in B_k, \forall k \in \mathbb{N}$. Expanding the intersection:

$$\omega \in B_1 \cap B_2 \cap \ldots \cap B_k \cap B_{k+1} \cap \ldots$$

Now, from the definition of B_1, like we used above:

$$B_1 = \bigcup_{i=1}^{\infty} A_i = A_1 \cup \bigcup_{i=2}^{\infty} A_i = A_1 \cup B_2$$

So, we can write: $B_1 \cap B_2 = (A_1 \cup B_2) \cap B_2 = (A_1 \cap B_2) \cup (B_2 \cap B_2) = (A_1 \cap B_2) \cup B_2$. And in general:

$$B_k = \bigcup_{i=k}^{\infty} A_i = A_k \cup \bigcup_{i=k+1}^{\infty} A_i = A_k \cup B_{k+1},$$

so $B_k \cap B_{k+1} = (A_k \cup B_{k+1}) \cap B_{k+1} = (A_k \cap B_{k+1}) \cup B_{k+1}$. So, each B_i can be decomposed into $(A_i \cap B_{i+1}) \cup B_{i+1}$ and the B_{i+1} is rewritten as its own intersection. Ultimately, we get:

$$\begin{aligned}
&\omega \in B_1 \cap B_2 \cap B_3 \cap \ldots \cap B_k \cap B_{k+1} \cap \ldots \implies \\
&\omega \in \big[(A_1 \cap B_2) \cup B_2\big] \cap B_3 \cap \ldots \cap B_k \cap B_{k+1} \cap \ldots \implies \\
&\omega \in \big[(A_1 \cap B_2) \cup (A_2 \cap B_3) \cup B_3\big] \cap \ldots \cap B_k \cap B_{k+1} \cap \ldots \implies \\
&\omega \in (A_1 \cap B_2) \cup (A_2 \cap B_3) \cup \ldots \cup (A_k \cap B_{k+1}) \cup (A_{k+1} \cap B_{k+2}) \cup \ldots \implies \\
&\omega \in (A_k \cap B_{k+1}) \forall k \in \mathbb{N} \text{ by assumption that } \omega \in B_k \forall k \in \mathbb{N} \implies \\
&\omega \in A_k, \forall k \in \mathbb{N}
\end{aligned}$$

which means $\omega \in A_1, A_2, \ldots, A_k, A_{k+1}, \ldots$ proving the first implication.

$\Leftarrow$) Now assume $\omega \in A_1, A_2, \ldots, A_k, \ldots$ which means that $\omega \in A_k$ for any $k \in \mathbb{N}$. Then,

$$\omega \in A_k \implies \omega \in A_k \cup \left(\bigcup_{i=k+1}^{\infty} A_i \right) = \bigcup_{i=k}^{\infty} A_i = B_k \implies \omega \in B_k,$$

So ω is included in all B_k for $k \in \mathbb{N}$ which means ω is also in the intersection of all B_i.

$$\omega \in B_1, B_2, \ldots, B_k, \ldots \implies \omega \in \bigcap_{n=1}^{\infty} B_n$$

By implication both ways, we have proved the equivalency. $\square$

(c) *Skipped.*

1.4
Proving DeMorgan's laws. First in the case when $i = \{1, 2, \ldots, n\}$.

Claim:
$$\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c$$

Proof. Starting with $x \in (\cup_{i=1}^{n} A_i)^c$, we have the following equivalencies:

$$\begin{aligned} x \in \left(\bigcup_{i=1}^{n} A_i\right)^c &\iff x \notin \bigcup_{i=1}^{n} A_i \\ &\iff x \notin A_1 \text{ and } x \notin A_2 \text{ and } \ldots \text{ and } x \notin A_n \\ &\iff x \in A_1^c \text{ and } x \in A_2^c \text{ and } \ldots \text{ and } x \in A_n^c \\ &\iff x \in \bigcap_{i=1}^{n} A_i^c \end{aligned} \qquad \square$$

Claim:
$$\left(\bigcap_{i=1}^{n} A_i\right)^c = \bigcup_{i=1}^{n} A_i^c$$

Proof. Starting with $x \in (\cap_{i=1}^{n} A_i)^c$, we have the following equivalencies:

$$\begin{aligned} x \in \left(\bigcap_{i=1}^{n} A_i\right)^c &\iff x \notin \bigcap_{i=1}^{n} A_i \\ &\iff \exists p \in \{1, \ldots, n\}\ x \notin A_p \\ &\iff \exists p \in \{1, \ldots, n\}\ x \in A_p^c \\ &\iff x \in \bigcup_{i=1}^{n} A_i^c \end{aligned} \qquad \square$$

Proving these results for a random index set is essentially the same, except finding some way of numerating the index set with $i_1, \ldots, i_n$ (since it has to be countably infinite) which I won't bother doing.

1.5
Tossing a fair coin until we get exactly two heads. Description of the sample space:

$$\Omega = \{(\omega_1, \omega_2, \ldots, \omega_{k-1}, \omega_k) : \omega_i \in \{H, T\}, \omega_{k-1} = \omega_k = H, \omega_{p-1} = H \Rightarrow \omega_p = T, p < k\}$$

Now, to calculate the probability that there will be exactly k tosses. Let us look at a few specific examples:

$$\begin{aligned} k = 2 : \{H, H\} &\implies \mathbb{P}(2) = \frac{1}{2^2} = \frac{1}{4} = 0.25 \\ k = 3 : \{T, H, H\} &\implies \mathbb{P}(3) = \frac{1}{2^3} = \frac{1}{8} = 0.125 \end{aligned}$$

For all tosses after $k > 3$, the last three tosses have to be $\{T, H, H\}$.

$$k=4: \quad \left.\begin{matrix} H \\ T \end{matrix}\right|, T, H, H = 2\cdot\frac{1}{2^4} = \frac{1}{8} = 0.125$$

$$k=5: \quad \left.\begin{matrix} TT \\ TH \\ HT \end{matrix}\right|, T, H, H = 3\cdot\frac{1}{2^5} = \frac{3}{32} = 0.09375$$

$$k=6: \quad \left.\begin{matrix} TTT \\ TTH \\ THT \\ HTT \\ HTH \end{matrix}\right|, T, H, H = 5\cdot\frac{1}{2^6} = \frac{5}{64} = 0.078125$$

$$k=7 \quad \left.\begin{matrix} TTTT \\ TTTH \\ TTHT \\ THTT \\ HTTT \\ THTH \\ HTHT \\ HTTH \end{matrix}\right|, T, H, H = 8\cdot\frac{1}{2^7} = \frac{1}{2^4} = \frac{1}{16} = 0.0625$$

We can see a pattern emerging. The last three tosses are fixed, and for the first $k-3$ tosses, we have to count all the combinations that do not contain 2 simultaneous H. The details are quite interesting, and it turns out that the number in the $k-3$ first tosses follows the Fibonacci sequence (as we can see above). Defining:

$$F(1) = 2, F(2) = 3, F(3) = 5, \ldots$$

In general, we will then get:

$$\mathbb{P}(K=k) = \begin{cases} \dfrac{1}{2^k} & k = 2,3 \\ \dfrac{F(k-2)}{2^k} & k > 3 \end{cases}$$

1.6

Let the sample space be $\Omega = \{0, 1, 2, \ldots\}$ (which is $\{0\} \cup \mathbb{N}$). We are going to prove that there is no uniform distribution on this sample space. That is, if $\mathbb{P}(A) = \mathbb{P}(B)$ whenver $|A| = |B|$ (the cardinality of the sets) then $\mathbb{P}$ cannot satisfy the axioms of probability.

The first two properties are satisfied. If we take some $k \in \mathbb{N}$ and define $A_1 = \{1, \ldots, k\}$ in such a way that $P(A_1) > 0$, and generally define $A_m = \{mk+1, mk+2, \ldots, mk+k\}$, then $A_n \cap A_m = \emptyset$. (This can be checked by setting $k = 100$ and $n = 5$ and $m = 6$ for instance). Also, $|A_n| = |A_m|$ and $\mathbb{P}(A_n) = \mathbb{P}(A_m) > 0$. This will cause problems in the third axiom:

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

since we are summing an infinite number of positive values. The axioms of probability are not satisfied, so we cannot have a uniform distribution on $\{0, 1, 2, \ldots\}$.

1.7

For some events $A_1, A_2, \ldots$, we are going to show that:

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

Proof. Following the hint, we define:

$$B_n = A_n - \bigcup_{i=1}^{n-1} A_i.$$

Next, we show that any B_m and B_n are disjoint for $n, m \in \mathbb{N}$ and $m > n$.

$$\begin{aligned}
B_m \cap B_n &= \left(A_m - \bigcup_{i=1}^{m-1} A_i\right) \cap \left(A_n - \bigcup_{i=1}^{n-1} A_i\right) \\
&= \left(A_m \cap \left(\bigcup_{i=1}^{m-1} A_i\right)^c\right) \cap \left(A_n \cap \left(\bigcup_{i=1}^{n-1} A_i\right)^c\right) \\
&= \left(A_m \cap \bigcap_{i=1}^{m-1} A_i^c\right) \cap \left(A_n \cap \bigcap_{i=1}^{n-1} A_i^c\right) \\
&= \left(A_m \cap A_{m-1}^c \cap \ldots \cap A_n^c \cap \ldots \cap A_2^c \cap A_1^c\right) \cap \left(A_n \cap A_{n-1}^c \cap \ldots \cap A_2^c \cap A_1^c\right) \\
&= A_m \cap A_{m-1}^c \cap \ldots \cap A_2^c \cap A_1^c \cap A_{n-1}^c \cap \ldots \cap A_2^c \cap A_1^c \cap \underbrace{A_n^c \cap A_n}_{=\emptyset} \\
&= \emptyset
\end{aligned}$$

Hence, the sets B_i are mutually disjoint. Next, we show that:

$$\left(\bigcup_{n=1}^{\infty} A_n\right) = \left(\bigcup_{n=1}^{\infty} B_n\right).$$

We will use an induction argument. For the first element we have: $B_1 = A_1$, and:

$$\begin{aligned}
\cup_{i=1}^2 B_i &= B_1 \cup B_2 \\
&= A_1 \cup (A_2 - A_1) \\
&= A_1 \cup (A_2 \cap A_1^c) \\
&= (A_1 \cup A_2) \cap (A_1 \cup A_1^c) \\
&= (A_1 \cup A_2) \cap \Omega \\
&= (A_1 \cup A_2) \\
&= \cup_{i=1}^2 A_i
\end{aligned}$$

So, we assume that for $k \in \mathbb{N}$, we have:

$$\left(\bigcup_{n=1}^{k} A_n\right) = \left(\bigcup_{n=1}^{k} B_n\right),$$

and show that this applies to $k+1$.

$$
\begin{aligned}
\bigcup_{i=1}^{k+1} B_i &= B_{k+1} \cup \left(\bigcup_{i=1}^{k} B_i\right) \\
&= B_{k+1} \cup \left(\bigcup_{i=1}^{k} A_i\right) \\
&= \left(A_{k+1} - \bigcup_{i=1}^{k} A_i\right) \cup \left(\bigcup_{i=1}^{k} A_i\right) \\
&= \left(A_{k+1} \cap \left(\bigcup_{i=1}^{k} A_i\right)^c\right) \cup \left(\bigcup_{i=1}^{k} A_i\right) \\
&= \left(A_{k+1} \cup \bigcup_{i=1}^{k} A_i\right) \cap \left[\left(\bigcup_{i=1}^{k} A_i\right)^c \cup \left(\bigcup_{i=1}^{k} A_i\right)\right] \\
&= \left(\bigcup_{i=1}^{k+1} A_i\right) \cap \Omega \\
&= \bigcup_{i=1}^{k+1} A_i
\end{aligned}
$$

So, by induction, we have verfied the equality for $k+1$, so it follows that the sets are equal for all indexes.

By definition of B_n, we have $B_n \subset A_n$ which means $\mathbb{P}(B_n) \leq \mathbb{P}(A_n)$. Using this, we can prove the main statement:

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

1.8

Suppose that $\mathbb{P}(A_i) = 1$ for each A_i. We are going to prove that:

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

Proof. We will prove this by induction. For events A_1 and A_2, we know that $\mathbb{P}(A_1) = 1$ and $\mathbb{P}(A_2) = 1$. By the probability axioms:

$$
A_1 \subset A_1 \cup A_2 \subset \Omega \implies \mathbb{P}(A_1) \leq \mathbb{P}(A_1 \cup A_2) \leq \mathbb{P}(\Omega)
$$

Using that $\mathbb{P}(A_1) = 1$ and $\mathbb{P}(\Omega) = 1$:

$$
1 \leq \mathbb{P}(A_1 \cup A_2) \leq 1 \implies \mathbb{P}(A_1 \cup A_2) = 1
$$

By Lemma 1.6:

$$
\begin{aligned}
\mathbb{P}(A_1 \cup A_2) &= \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2) \\
1 &= 2 - \mathbb{P}(A_1 \cap A_2) \\
\mathbb{P}(A_1 \cap A_2) &= 1
\end{aligned}
$$

Assuming this holds for $k \in \mathbb{N}$, i.e.:

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

To simplify the notation, we call this intersection W_k, so $\mathbb{P}(W_k) = 1$. Now we will show it also applies to $k+1$.

$$A_1 \cap A_2 \cap \ldots \cap A_k \cap A_{k+1} = W_k \cap A_{k+1}$$

Again, from the probability axioms:

$$W_k \subset W_k \cup A_{k+1} \subset \Omega \implies \mathbb{P}(W_k) \leq \mathbb{P}(W_k \cup A_{k+1}) \leq \mathbb{P}(\Omega)$$

Using that $\mathbb{P}(W_k) = 1$ and $\mathbb{P}(\Omega) = 1$:

$$1 \leq \mathbb{P}(W_k \cup A_{k+1}) \leq 1 \implies \mathbb{P}(W_k \cup A_{k+1}) = 1$$

By Lemma 1.6, and using that $\mathbb{P}(A_i) = 1$ for each i:

$$\begin{aligned}
\mathbb{P}(W_k \cup A_{k+1}) &= \mathbb{P}(W_k) + \mathbb{P}(A_{k+1}) - \mathbb{P}(W_k \cap A_{k+1}) \\
1 &= 2 - \mathbb{P}(W_k \cap A_{k+1}) \\
\mathbb{P}(W_k \cap A_{k+1}) &= 1 \\
\mathbb{P}\left(\bigcap_{i=1}^{k} A_i \cap A_{k+1}\right) &= 1 \\
\mathbb{P}\left(\bigcap_{i=1}^{k+1} A_i\right) &= 1
\end{aligned}$$

Which verifies that equality holds for $k+1$. By induction, we can conclude that:

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

1.9

For a fixed B such that $\mathbb{P}(B) > 0$, we will show that $\mathbb{P}(\cdot|B)$ satisfies the axioms of probability. Recalling the definition of conditional probability:

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

• **Axiom 1**. $\mathbb{P}(A|B) \geq 0$ for all A.

Proof. The set $A \cap B$ is an event wrt. $\mathbb{P}(\cdot)$, so $\mathbb{P}(A \cap B) = a \geq 0$.

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{a}{\mathbb{P}(B)} \geq 0,$$

since $\mathbb{P}(B) > 0$. $\square$

• **Axiom 2**. $\mathbb{P}(\Omega|B) = 1$.
Proof.

$$\mathbb{P}(\Omega|B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1 \qquad \square$$

• **Axiom 3**. If $A_1, A_2, \ldots$ are disjoint, then:

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i \middle| B\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i|B)$$

Proof. First a supporting argument. If we define $C_i := A_i \cap B$ for each i, then $C_1, C_2, \ldots$ are mutually disjoint sets and by (A.3):

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

This means:

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

which proves the final axiom for the conditional probability. $\square$

1.10 Monty Hall
There are three doors, 1, 2, 3, and there is a prize between one of them. We don't know which door contains the prize, so the probability of selecting the correct door is evenly spread out:

$$P(1) = \frac{1}{3}, \quad P(2) = \frac{1}{3}, \quad P(3) = \frac{1}{3}$$

To simplify the calculations, we assume we pick door 1, and that the prize is not behind door 3. Now we will decide if it's better to switch or not. What is the probability that the host will open door III conditioned on where the prize is?

$$\begin{aligned}
&P(\mathrm{III}|1) = \frac{1}{2} &&\text{(Correct door is selected, door 3 is opened randomly)}\\
&P(\mathrm{III}|2) = 1 &&\text{(Host never reveals the prize)}\\
&P(\mathrm{III}|3) = 0 &&\text{(Host never reveals the prize)}
\end{aligned}$$

Now, what is the probability that the prize is behind door number 2, given that door III was opened? This can be calculated by Bayes Law.

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

In our case, probability of door 2 containing the prize, given that door III was opened:

$$\begin{aligned}
P(2|\text{III}) &= \frac{\mathbb{P}(\text{III}|2)\mathbb{P}(2)}{\mathbb{P}(\text{III}|1)\mathbb{P}(1) + \mathbb{P}(\text{III}|2)\mathbb{P}(2) + \mathbb{P}(\text{III}|3)\mathbb{P}(3)} \\
&= \frac{(1)(\frac{1}{3})}{(\frac{1}{2})(\frac{1}{3}) + (1)(\frac{1}{3}) + (0)(\frac{1}{3})} \\
&= \frac{\frac{1}{3}}{\frac{1}{6} + \frac{1}{3} + 0} = \frac{1}{3} / \frac{1}{2} \\
&= \frac{2}{3}
\end{aligned}$$

Probability that door 1 contains the prize, given that door III was opened:

$$\begin{aligned}
P(1|\text{III}) &= \frac{\mathbb{P}(\text{III}|1)\mathbb{P}(1)}{\mathbb{P}(\text{III}|1)\mathbb{P}(1) + \mathbb{P}(\text{III}|2)\mathbb{P}(2) + \mathbb{P}(\text{III}|3)\mathbb{P}(3)} \\
&= \frac{(\frac{1}{2})(\frac{1}{3})}{(\frac{1}{2})(\frac{1}{3}) + (1)(\frac{1}{3}) + (0)(\frac{1}{3})} \\
&= \frac{\frac{1}{6}}{\frac{1}{6} + \frac{1}{3} + 0} = \frac{1}{6} / \frac{1}{2} \\
&= \frac{1}{3}
\end{aligned}$$

This shows that we are better off switching doors then remaining in the one we initially selected.

1.11
Claim. If A and B are independent, then A^c and B^c are independent.

Proof. By definition of independence $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. By property of probabilities: $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ and $\mathbb{P}(B^c) = 1 - \mathbb{P}(B)$. Applying Lemma 1.6 to $A^c \cup B^c$:

$$\mathbb{P}(A^c \cup B^c) = \mathbb{P}(A^c) + \mathbb{P}(B^c) - \mathbb{P}(A^c \cap B^c) \tag{2.11.1}$$

By property of complements, and using DeMorgan's law:

$$\mathbb{P}(A \cap B) = 1 - \mathbb{P}([A \cap B]^c) = 1 - \mathbb{P}(A^c \cup B^c)$$

Which we can rewrite in the following way, and then apply the independence of A and B:

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

By property of complements, we can represent this as:

$$\begin{aligned}
\mathbb{P}(A^c \cup B^c) &= 1 - [1 - \mathbb{P}(A^c)][1 - \mathbb{P}(B^c)] \\
&= 1 - \left[1 - \mathbb{P}(B^c) - \mathbb{P}(A^c) + \mathbb{P}(A^c)\mathbb{P}(B^c)\right] \\
&= \mathbb{P}(A^c) + \mathbb{P}(B^c) - \mathbb{P}(A^c)\mathbb{P}(B^c) \qquad (2.11.2)
\end{aligned}$$

By setting (2.11.1) equal to (2.11.2) we can deduce that $\mathbb{P}(A^c)\mathbb{P}(B^c) = \mathbb{P}(A^c \cap B^c)$ which proves independence. $\square$

Alternative Proof. Slightly more compact argument, but it still relies on Lemma 1.6.

$$\begin{aligned}
\mathbb{P}(A^c)\mathbb{P}(B^c) &= \big[1-\mathbb{P}(A)\big]\big[1-\mathbb{P}(B)\big] \\
&= 1-\mathbb{P}(A)-\mathbb{P}(B)+\mathbb{P}(A)\mathbb{P}(B) \\
&= 1-\big[\mathbb{P}(A)+\mathbb{P}(B)-\mathbb{P}(A)\mathbb{P}(B)\big] \\
&\overset{IND}{=} 1-\big[\mathbb{P}(A)+\mathbb{P}(B)-\mathbb{P}(A\cap B)\big] \\
&\overset{L1.6}{=} 1-\mathbb{P}(A\cup B) \\
&= \mathbb{P}([A\cup B]^c) \\
&= \mathbb{P}(A^c\cap B^c)
\end{aligned}$$

□

1.12

Initially, we have three cards that we call G (Green both sides), R (Red both sides) and M (mixed colors). Next we define r as 'Red side' and g as 'Green side'. We can define the joint distribution as follows:

	G	M	R	
r		1/6	2/6	1/2
g	2/6	1/6		1/2
	1/3	1/3	1/3	

From the marginal distributions, we can see that the initial probabilities for selecting each card is:

$$\mathbb{P}(G)=\frac{1}{3},\quad \mathbb{P}(R)=\frac{1}{3},\quad \mathbb{P}(M)=\frac{1}{3}$$

And the probabilities for randomly viewing a colored card side:

$$\mathbb{P}(g)=\frac{1}{2},\quad \mathbb{P}(r)=\frac{1}{2}$$

Note that we do not have independene, since $\mathbb{P}(G\cap g)=2/6$ while $\mathbb{P}(G)\mathbb{P}(g)=(1/3)(1/2)=1/6$.

We want to find the probability that we have a green card given that we observe a green side. This can be calculated by:

$$\mathbb{P}(G|g)=\frac{\mathbb{P}(G\cap g)}{\mathbb{P}(g)}=\frac{2/6}{1/2}=4/6=2/3$$

1.13

A fair coin is thrown until we have both H and T.

(a) The sample space is:

$$\Omega=\{(\omega_1,\ldots,\omega_K):\omega_i\in\{H,T\},1\le i\le K \text{ and } \omega_{K-1}\neq\omega_K, K\ge 2\}$$

(b) The probability of getting 3 tosses can only happen in two ways (H,H,T) and (T,T,H). So:

$$P(K=3)=2\cdot\frac{1}{2^3}=\frac{1}{4}$$

(it would be more complicated if it were e.g. $K\le 3$).

1.14
Claim: If $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$, then A is independent of every other event.

Proof. First consider the case $\mathbb{P}(A) = 0$. Obviously, for any $E \in \Omega$:

$$\mathbb{P}(A)\mathbb{P}(E) = 0 \cdot \mathbb{P}(E) = 0.$$

By the axioms of probability: $A \cap E \subset A$ so $\mathbb{P}(A \cap E) \leq \mathbb{P}(A) = 0$ and since for any event: $0 \leq \mathbb{P}(A \cap E)$ we can conclude that $\mathbb{P}(A \cap E) = 0$. So:

$$\mathbb{P}(A \cap E) = 0 = \mathbb{P}(A)\mathbb{P}(E) \quad \Longrightarrow \quad \mathbb{P}(A \cap E) = \mathbb{P}(A)\mathbb{P}(E)$$

which proves independence.

Now consider $\mathbb{P}(A) = 1$. Then, for any event $E \in \Omega$:

$$\mathbb{P}(A)\mathbb{P}(E) = 1 \cdot \mathbb{P}(E) = \mathbb{P}(E).$$

Since $A \subset A \cup E \subset \Omega$, then $\mathbb{P}(A) \leq \mathbb{P}(A \cup E) \leq \mathbb{P}(\Omega)$. We know the probabilities of A and Ω, so $1 \leq \mathbb{P}(A \cup E) \leq 1$ means that $\mathbb{P}(A \cup E) = 1$. By Lemma 1.6:

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(E) - \mathbb{P}(A \cap E) \\ 1 &= 1 + \mathbb{P}(E) - \mathbb{P}(A \cap E) \\ \mathbb{P}(A \cap E) &= \mathbb{P}(E) \end{aligned}$$

In conclusion:

$$\mathbb{P}(A \cap E) = \mathbb{P}(E) = \mathbb{P}(A)\mathbb{P}(E) \quad \Longrightarrow \quad \mathbb{P}(A \cap E) = \mathbb{P}(A)\mathbb{P}(E)$$

which proves independence. □

Claim: If A is independent of itself, then $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

Proof. If A is independent of itself, we can write:

$$\begin{aligned} \mathbb{P}(A \cap A) &= \mathbb{P}(A)\mathbb{P}(A) \\ \mathbb{P}(A) &= \mathbb{P}(A)^2 \end{aligned}$$

since $A \cap A = A$. We can write this as an equation, by defining $p := \mathbb{P}(A)$.

$$\begin{aligned} p &= p^2 \\ p^2 - p &= 0 \\ p(p-1) &= 0 \end{aligned}$$

This equation is solved when $p = \mathbb{P}(A) = 1$ or $p = \mathbb{P}(A) = 0$. □

1.15

The probability that a child has blue eyes is 1/4, $\mathbb{P}(B) = 1/4$. We assume independence in the eye color of the children.

(a) Given that one child has blue eyes, what is the probability that 2 or more children have blue eyes? Since it's given that one child has blue eyes, we just need to find the probability that one, or both, of the remaining children have blue eyes. Define B_i: as the event that child i has blue eyes, and $\neg B_i$ the event that child i does NOT have blue eyes. We only need to sum up the following three probabilities:

$$\mathbb{P}(B_1 \cap B_2) = \mathbb{P}(B_1)\mathbb{P}(B_2) = \left(\frac{1}{4}\right)\left(\frac{1}{4}\right) = \frac{1}{16}$$
$$\mathbb{P}(B_1 \cap \neg B_2) = \mathbb{P}(B_1)\mathbb{P}(\neg B_2) = \left(\frac{1}{4}\right)\left(\frac{3}{4}\right) = \frac{3}{16}$$
$$\mathbb{P}(\neg B_1 \cap B_2) = \mathbb{P}(\neg B_1)\mathbb{P}(B_2) = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right) = \frac{3}{16}$$

By summing these up, we find that the probability of two or more children having blue eyes is $\frac{7}{16}$.

(b) Which child has blue eyes is irrelevant. The probability will be the same: $\frac{7}{16}$.

1.16

Proving Lemma 1.14. If A and B are independent events, then $\mathbb{P}(A|B) = \mathbb{P}(A)$. Also, for any pair of events A and B,

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

Proof. Suppose A and B are independent. By definition of independence, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. By definition of conditional probability:

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

which proves the first statement. Again, by definition of conditional probability:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad \Longrightarrow \quad \mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$$

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} \quad \Longrightarrow \quad \mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A),$$

since $A \cap B = B \cap A$ which proves the remaining two statements. $\square$

1.17
Show that
$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A|B \cap C)\mathbb{P}(B|C)\mathbb{P}(C).$$

Proof. Define $E := B \cap C$, so $A \cap B \cap C = A \cap E$. By Lemma 1.14:
$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \cap E) = \mathbb{P}(A|E)\mathbb{P}(E) = \mathbb{P}(A|B \cap C)\mathbb{P}(B \cap C) \tag{2.17.1}$$

By Lemma 1.14 again:
$$\mathbb{P}(B \cap C) = \mathbb{P}(B|C)\mathbb{P}(C),$$
which we can replace in equation (2.17.1) and get the desired result:
$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A|B \cap C)\mathbb{P}(B|C)\mathbb{P}(C)$$
□

1.18
Suppose k events form a partition of Ω, i.e. they are disjoint and $\bigcup_{i=1}^{k} A_i = \Omega$. Assume that $\mathbb{P}(B) > 0$. If $\mathbb{P}(A_1|B) < \mathbb{P}(A_1)$ then $\mathbb{P}(A_i|B) > \mathbb{P}(A_i)$ for some $i = 2, 3, \ldots, k$.

Proof. By summing over all A_i with and without conditioning on B, we get the following two equalities:
$$\sum_{i=1}^{k} \mathbb{P}(A_i|B) = 1, \qquad \sum_{i=1}^{k} \mathbb{P}(A_i) = 1$$

We can set these equal to each other:
$$\sum_{i=1}^{k} \mathbb{P}(A_i|B) = \sum_{i=1}^{k} \mathbb{P}(A_i)$$

Removing the terms for A_1 from the sums:
$$\mathbb{P}(A_1|B) + \sum_{i=2}^{k} \mathbb{P}(A_i|B) = \mathbb{P}(A_1) + \sum_{i=2}^{k} \mathbb{P}(A_i)$$

Since $\mathbb{P}(A_1|B) < \mathbb{P}(A_1)$, then $\mathbb{P}(A_1) - \mathbb{P}(A_1|B) = a > 0$ for some $a \in \mathbb{R}$. By subtracting $\mathbb{P}(A_1|B)$ from both sides, we can write:
$$\sum_{i=2}^{k} \mathbb{P}(A_i|B) = \mathbb{P}(A_1) - \mathbb{P}(A_1|B) + \sum_{i=2}^{k} \mathbb{P}(A_i)$$
$$\sum_{i=2}^{k} \mathbb{P}(A_i|B) = a + \sum_{i=2}^{k} \mathbb{P}(A_i) \; > \; \sum_{i=2}^{k} \mathbb{P}(A_i)$$

So, we get:
$$\sum_{i=2}^{k} \mathbb{P}(A_i|B) \; > \; \sum_{i=2}^{k} \mathbb{P}(A_i)$$
which means there exists some $i \in \{2, 3, \ldots, k\}$ such that $\mathbb{P}(A_i|B) > \mathbb{P}(A_i)$. □

1.19

Define M as Mac users, W as Windows users, and L as Linux users. We have:

$$\mathbb{P}(M) = 0.3, \quad \mathbb{P}(W) = 0.5, \quad \mathbb{P}(L) = 0.2$$

The computers are infected with a virus, which give the following probabilities:

$$\mathbb{P}(V|M) = 0.65, \quad \mathbb{P}(V|W) = 0.82, \quad \mathbb{P}(V|L) = 0.50$$

A person is selected at random and we learn that her computer has the virus. What is the probability that she is using Windows? Or, what is $\mathbb{P}(W|V)$? We calculate this with Bayes Law.

$$\begin{aligned}
\mathbb{P}(W|V) &= \frac{\mathbb{P}(V|W)\mathbb{P}(W)}{\mathbb{P}(V|M)\mathbb{P}(M) + \mathbb{P}(V|W)\mathbb{P}(W) + \mathbb{P}(V|L)\mathbb{P}(L)} \\
&= \frac{(0.82)(0.5)}{(0.65)(0.3) + (0.82)(0.5) + (0.5)(0.2)} \\
&= 0.5815
\end{aligned}$$

1.20

We have a box containing 5 coins, $C_1, \ldots, C_5$. Each coin is selected at random, so $\mathbb{P}(C_i) = 1/5$ for all coins. They have the following probabilities of getting H when flipped ($p_i = \mathbb{P}(H|C_i)$):

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

(a) We flip a coin and get a head. Calculate $\mathbb{P}(C_i|H)$ for all coins. First an intermediary calculation:

$$\begin{aligned}
\sum_{i=1}^{5} \mathbb{P}(H|C_i)\mathbb{P}(C_i) &= \mathbb{P}(H|C_1)\mathbb{P}(C_1) + \mathbb{P}(H|C_2)\mathbb{P}(C_2) + \mathbb{P}(H|C_3)\mathbb{P}(C_3) + \mathbb{P}(H|C_4)\mathbb{P}(C_4) + \mathbb{P}(H|C_5)\mathbb{P}(C_5) \\
&= (0)\left(\frac{1}{5}\right) + \left(\frac{1}{4}\right)\left(\frac{1}{5}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{5}\right) + \left(\frac{3}{4}\right)\left(\frac{1}{5}\right) + (1)\left(\frac{1}{5}\right) \\
&= 0 + \frac{1}{20} + \frac{2}{20} + \frac{3}{20} + \frac{4}{20} \\
&= \frac{1}{2}
\end{aligned}$$

Calculating the posterior for each C_i.

$$\mathbb{P}(C_1|H) = \frac{\mathbb{P}(H|C_1)\mathbb{P}(C_1)}{\sum_{i=1}^{5} \mathbb{P}(H|C_i)\mathbb{P}(C_i)} = \frac{0}{\frac{1}{2}} = 0$$

$$\mathbb{P}(C_2|H) = \frac{\mathbb{P}(H|C_2)\mathbb{P}(C_2)}{\sum_{i=1}^{5} \mathbb{P}(H|C_i)\mathbb{P}(C_i)} = \frac{\frac{1}{20}}{\frac{1}{2}} = 1/10$$

$$\mathbb{P}(C_3|H) = \frac{\mathbb{P}(H|C_3)\mathbb{P}(C_3)}{\sum_{i=1}^{5} \mathbb{P}(H|C_i)\mathbb{P}(C_i)} = \frac{\frac{2}{20}}{\frac{1}{2}} = 2/10$$

$$\mathbb{P}(C_4|H) = \frac{\mathbb{P}(H|C_4)\mathbb{P}(C_4)}{\sum_{i=1}^{5} \mathbb{P}(H|C_i)\mathbb{P}(C_i)} = \frac{\frac{3}{20}}{\frac{1}{2}} = 3/10$$

$$\mathbb{P}(C_5|H) = \frac{\mathbb{P}(H|C_5)\mathbb{P}(C_5)}{\sum_{i=1}^{5} \mathbb{P}(H|C_i)\mathbb{P}(C_i)} = \frac{\frac{4}{20}}{\frac{1}{2}} = 4/10$$

(b) We toss the coin again. Finding the probability $\mathbb{P}(H_2|H_1)$ (H_i: getting H on toss i). We are still using the same random coin, so we must calculate the probability for all coins together.

1.21

Proportion of H vs T for $p = 0.3$.

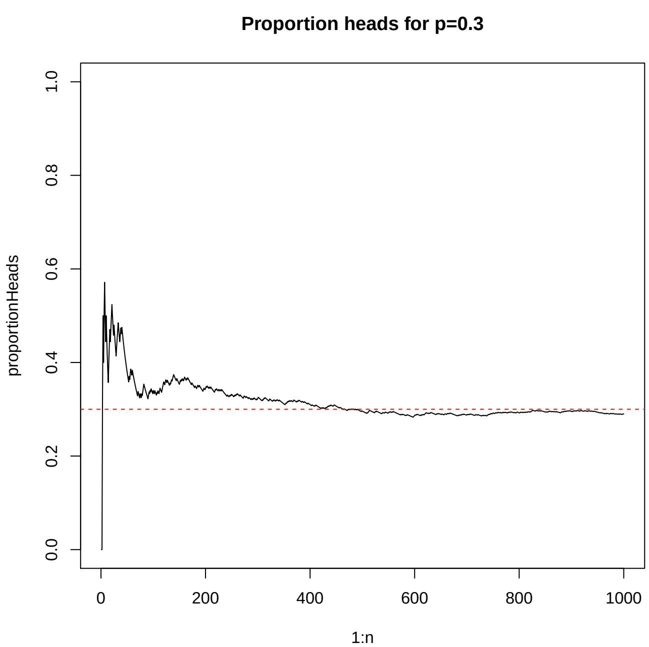
```
# 1.21 - Plotting proportion of H vs T
n = 1000
p = 0.3
coinTosses = sample(c("H","T"), prob = c(p, 1-p), size = n, replace = TRUE)
proportionHeads = rep(0, n)
headCount = 0
for(i in 1:n) {
    if (coinTosses[i] == "H") {
    headCount = headCount + 1
    }
    proportionHeads[i] = headCount/i
}

# PDF
pdf("~/ALLSTAT/ch2_2.21a.pdf")
plot(x = 1:n, y = proportionHeads, type = "l", ylim = c(0,1),
        main = paste0("Proportion heads for p=", p))
abline(h = p, col = "red", lty = "dashed")
dev.off()
```

R

Result

After some initial randomness due to few samples, we can see that the simulated results stabilizes aroud 0.3, as expected.



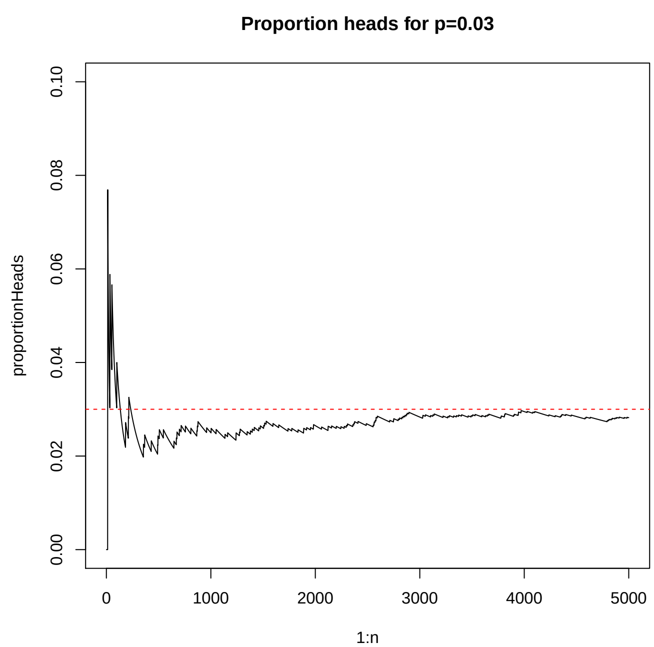
Proportion of H vs T for $p = 0.03$.

```
# 1.21 - Plotting proportion of H vs T
n = 5000
p = 0.03
coinTosses = sample(c("H","T"), prob = c(p, 1-p), size = n, replace = TRUE)
proportionHeads = rep(0, n)
headCount = 0
for(i in 1:n) {
    if (coinTosses[i] == "H") {
    headCount = headCount + 1
    }
    proportionHeads[i] = headCount/i
}

# PDF
pdf("~/ALLSTAT/ch2_2.21b.pdf")
plot(x = 1:n, y = proportionHeads, type = "l", ylim = c(0,1),
        main = paste0("Proportion heads for p=", p))
abline(h = p, col = "red", lty = "dashed")
dev.off()
```

R

Result
After some initial randomness due to few samples, we can see that the simulated results stabilizes aroud 0.03, as expected. Did a simulation of 5000 to make the 'convergence' clearer. Note that the y-axis only goes up to 0.1 in this plot.



1.22
Investigating some properties of Binomial random variables.

```r
# 1.22 - Binomial Random Variables
REP = 10  # Number of simulations per n
p = 0.3   # Probability of H
sim10 = rep(0, REP)
sim100 = rep(0, REP)
sim1000 = rep(0, REP)
# Simulating 10
for(i in 1:REP) {
    # 1 means head
    coinTosses = sample(c(1,0), prob = c(p, 1-p), size = 10, replace = TRUE)
    sim10[i] = sum(coinTosses)
}
# Simulating 100
for(i in 1:REP) {
    # 1 means head
    coinTosses = sample(c(1,0), prob = c(p, 1-p), size = 100, replace = TRUE)
    sim100[i] = sum(coinTosses)
}
# Simulating 1000
for(i in 1:REP) {
    # 1 means head
    coinTosses = sample(c(1,0), prob = c(p, 1-p), size = 1000, replace = TRUE)
    sim1000[i] = sum(coinTosses)
}
df = data.frame(
    SIM10 = sim10,
    SIM100 = sim100,
    SIM1000 = sim1000
)
# Output
df
apply(df, 2, mean)
```

R

As seen in the results, the mean of the 10 simulations is close to np which would be 3, 30 and 300.

```
> df
    SIM10 SIM100 SIM1000
1      1      32      282
2      2      30      298
3      3      33      319
4      3      28      290
5      3      29      284
6      2      34      286
7      4      37      310
8      2      30      294
9      2      30      297
10     2      31      326
> apply(df, 2, mean)
    SIM10  SIM100 SIM1000
    2.4    31.4   298.6
```

1.23
Simulating conditional probabilities. First we simulate an independent experiment.

```
# 1.23 - Simulating a fair die
options(digits=8)
numberOfTosses = 10000
A = c(2, 4, 6)
B = c(1, 2, 3, 4)
AandB = intersect(A, B) # c(2, 4)

dieTosses = sample(1:6, size = numberOfTosses, replace = TRUE)

# Calculating P(A), P(B), P(A)*P(B) and P(A cap B)
PA = sum(dieTosses %in% A)/numberOfTosses
PB = sum(dieTosses %in% B)/numberOfTosses
PAandB = sum(dieTosses %in% AandB)/numberOfTosses

# Output
PA
PB
PA*PB
PAandB
```

R

Results from the calculation. As we can see, the estimates for $\mathbb{P}(A) \approx 1/2$ and $\mathbb{P}(B) \approx 2/3$. Also $\mathbb{P}(A \cap B) \approx \mathbb{P}(A)\mathbb{P}(B) \approx 1/3$. Differences are probably due to rounding errors.

```
> # Output
> PA
[1] 0.5005
> PB
[1] 0.6695
> PA*PB
[1] 0.33508475
> PAandB
[1] 0.3344
```

Now we will construct an experiment with a conditional probability. We will use a fair coin and a die. When we get heads, this will correspond to 1 and the die is unchanged. If we get tails, this will be 2 and will double the die count; so if we get tails and we roll a 2, this will give us a 4.

	1	2	3	4	5	6	7	8	9	10	11	12
H	1/12	1/12	1/12	1/12	1/12	1/12						
T		1/12		1/12		1/12		1/12		1/12		1/12

Define the events: $A = \{2, 3, 4, 5, 6\}$ and $B = \{2, 4, 6, 8\}$ which will give $A \cap B = \{2, 4, 6\}$. The theoretical probabilities are:

$$\mathbb{P}(A) = 8/12 = 2/3 = 0.666$$
$$\mathbb{P}(B) = 7/12 \approx 0.5833$$
$$\mathbb{P}(A)\mathbb{P}(B) = 7/18 \approx 0.3888$$
$$\mathbb{P}(A \cap B) = 1/2 = 0.5$$

Simulating conditional probabilities.

```r
# 1.23 - Simulating a conditional probability
options(digits=8)
numberOfTosses = 10000
A = c(2, 3, 4, 5, 6)
B = c(2, 4, 6, 8)
AandB = intersect(A, B) # c(2, 4, 6)

dieTosses = sample(1:6, size = numberOfTosses, replace = TRUE)
coinTosses = sample(1:2, size = numberOfTosses, replace = TRUE)
jointToss = dieTosses*coinTosses

# Calculating P(A), P(B), P(A)*P(B) and P(A cap B)
PA = sum(jointToss %in% A)/numberOfTosses
PB = sum(jointToss %in% B)/numberOfTosses
PAandB = sum(jointToss %in% AandB)/numberOfTosses

# Output
PA
PB
PA*PB
PAandB
```

R

```
> # Output
> PA
[1] 0.6691
> PB
[1] 0.5825
> PA*PB
[1] 0.38975075
> PAandB
[1] 0.4995
```

Repeating the theoretical values from previous page:

$$\begin{aligned}
\mathbb{P}(A) &= 8/12 = 2/3 = 0.666 \\
\mathbb{P}(B) &= 7/12 \approx 0.5833 \\
\mathbb{P}(A)\mathbb{P}(B) &= 7/18 \approx 0.3888 \\
\mathbb{P}(A \cap B) &= 1/2 = 0.5
\end{aligned}$$

The simulated results are very close to the theoretical calculations. We can see that we do not have independence, as expected.

2 Chapter 2 - Random Variables

Exercises

2.1
Claim: $\mathbb{P}(X = x) = F(x^+) - F(x^-)$. (Discrete)
PROOF. By definition of the CDF:

$$F(x^+) = \lim_{z \downarrow x} F(z) = \lim_{z \downarrow x} \mathbb{P}(X \le z), \qquad F(x^-) = \lim_{y \uparrow x} F(y) = \lim_{y \uparrow x} \mathbb{P}(X \le y)$$

(so $y < x$ and $y \to x$, and $x < z$ and $x \leftarrow z$). By the right continuous property, we can deduce that $z > y$ and we can set $z = x$ and $y = x - 1$.

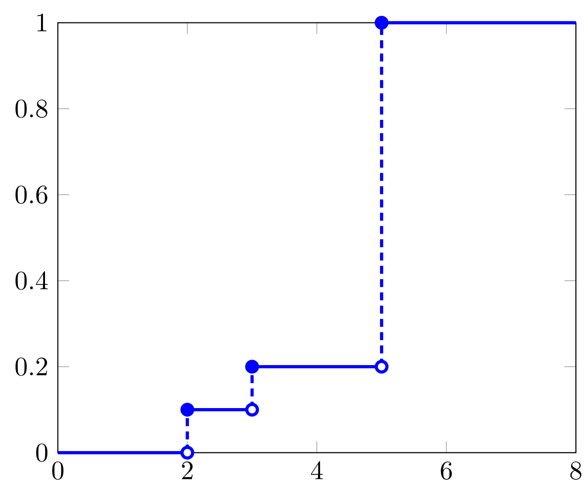
$$\mathbb{P}(X \le x^+) = \mathbb{P}(X \le x) = \mathbb{P}(X = x) + \mathbb{P}(X \le x - 1), \quad \mathbb{P}(X \le x^-) = \mathbb{P}(X \le x - 1)$$

So:

$$\begin{aligned}
\mathbb{P}(X = x) &= \mathbb{P}(X = x) + \mathbb{P}(X \le x - 1) - \mathbb{P}(X \le x - 1) \\
&= \mathbb{P}(X \le x) - \mathbb{P}(X \le x - 1) \\
&= \mathbb{P}(X \le x^+) - \mathbb{P}(X \le x^-) \\
&= F(x^+) - F(x^-)
\end{aligned}$$

□

2.2
Let X be such that $\mathbb{P}(X = 2) = \mathbb{P}(X = 3) = 1/10$ and $\mathbb{P}(X = 5) = 8/10$. Here is a plot of the CDF.



By reading the plot, we can see that:

$$\mathbb{P}(2 < X \le 4.8) = F(4.8) - F(2) = 2/10 - 1/10 = 1/10$$

$$\mathbb{P}(2 \le X \le 4.8) = F(4.8) = 2/10$$

2.3

Lemma 2.15 Let F be the CDF for a random variable X. Then:

1. $\mathbb{P}(X = x) = F(x) - F(x^-)$
2. $\mathbb{P}(x < X \leq y) = F(y) - F(x)$
3. $\mathbb{P}(X > x) = 1 - F(x)$
4. If X is continuous, then

$$F(b) - F(a) = \mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b)$$

Proof. We will prove each statement in turn. (1.) was proved in exercise **2.1**. Doing (3) first, since we need it to prove (2).

(3) By definition of complements of sets $A = \{X > x\}$ means $A^c = \{X \leq x\}$, and it follows that:

$$\mathbb{P}(X > x) = \mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \mathbb{P}(X \leq x) = 1 - F(x).$$

(2) Assume $x < y$. We will need that $\{X > x\} \cup \{X \leq y\} = \Omega$, and we will also use Lemma 1.6 (in reverse).

$$\begin{aligned}
\mathbb{P}(x < X \leq y) &= \mathbb{P}(\{X > x\} \cap \{X \leq y\}) \\
&= \mathbb{P}(X > x) + \mathbb{P}(X \leq y) - \mathbb{P}(\{X > x\} \cup \{X \leq y\}) \\
&= 1 - F(x) + F(y) - 1 \\
&= F(y) - F(x)
\end{aligned}$$

(4) Similar argument for all cases, so will just do one. We just need to turn the inequalities into strict inequalities. For continuous random variables, pointwise probabilities are 0. Again, we will need to use $\{X > a\} \cup \{X < b\} = \Omega$.

Define $A := \{a \leq X\}$ and $B := \{X < b\}$. First, we make the following observation:

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(\{a \leq X\}) \\
&= \mathbb{P}(\{a = X\} \cup \{a < X\}) \\
&= \mathbb{P}(\{a = X\}) + \mathbb{P}(\{a < X\}) + \mathbb{P}(\{a = X\} \cap \{a < X\}) \\
&= 0 + \mathbb{P}(A') + 0 \\
&= \mathbb{P}(A')
\end{aligned}$$

where $A' = \{a < X\}$. We get 0 for the pointwise probability, since this is continuous, and we get 0 because the sets are disjoint. We have shown that $\mathbb{P}(A) = \mathbb{P}(A')$ and can use this to conclude the proof.

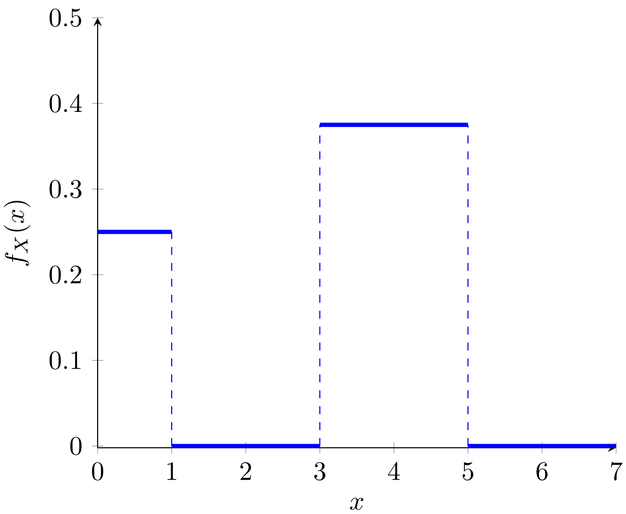
$$\begin{aligned}
\mathbb{P}(a \leq X < b) &= \mathbb{P}(\{a \leq X\} \cap \{X < b\}) \\
&= \mathbb{P}(A \cap B) \\
&= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(A \cup B) \\
&= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(\Omega) \\
&= \mathbb{P}(A') + \mathbb{P}(B) + \mathbb{P}(A' \cup B) \\
&= \mathbb{P}(A' \cap B) \\
&= \mathbb{P}(a < X < b)
\end{aligned}$$

□

2.4
X has the probability density (PDF):

$$f_X(x) = \begin{cases} 1/4 & 0 < x < 1 \\ 3/8 & 3 < x < 5 \\ 0 & \text{otherwise} \end{cases}$$

Plot of the PDF:



From the relatively simple structure, we can easily determine the area under the graph:

$$A = (1)\left(\frac{1}{4}\right) + (2)\left(\frac{3}{8}\right) = \frac{2}{8} + \frac{6}{8} = 1$$

(a) Finding the CDF by integrating the PDF. We will split up the integral in several parts. First for the case when $y \in (0, 1)$:

$$F_X(y) = \int_{-\infty}^{y} f_X(t)dt = \frac{1}{4}\int_0^y 1dt = \frac{1}{4}\Big[t\Big]_0^y = \frac{y}{4}$$

When $y = 1$ we have $F_X(1) = 1/4$. Next, we must consider the case $y \in (1, 3)$. Here the PDF is 0, so it doesn't increase. It remains constant at 1/4 (since the CDF doesn't decrease).

$$F_X(y) = \frac{1}{4}$$

Next is the case $y \in (3, 5)$. Consider the intermediary integral:

$$I_1 = \int_3^y \frac{3}{8}dt = \frac{3}{8}\Big[t\Big]_3^y = \frac{3y - 9}{8}$$

For values $y \in (3, 5)$ we start on 1/4, so the CDF in this region becomes:

$$F_X(y) = \frac{3y - 9}{8} + \frac{1}{4}$$

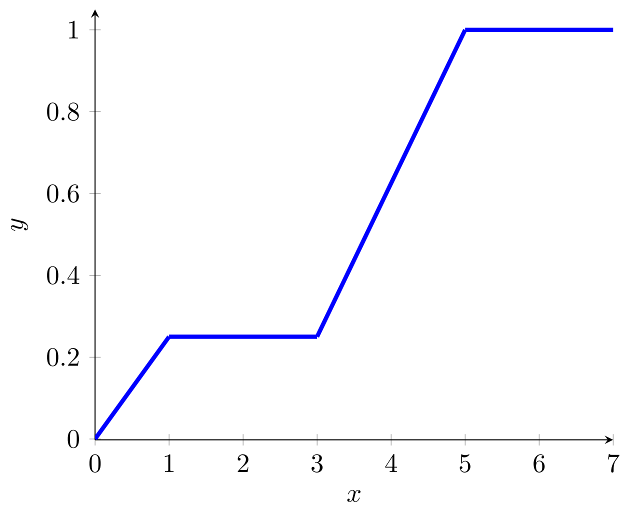
So, the full expression for the CDF becomes:

$$F_X(y) = \begin{cases} y/4 & y \in (0,1) \\ 1/4 & y \in (1,3) \\ \dfrac{3y-9}{8} + \dfrac{1}{4} & y \in (3,5) \\ 1 & y \geq 5 \end{cases}$$

Note that when $y = 5$ we get:

$$F_X(5) = \frac{3(5)-9}{8} + \frac{1}{4} = \frac{6}{8} + \frac{2}{8} = 1$$

Plot of the CDF:



(b) Defining $Y = 1/X$ and finding the PDF of Y. Following the hint we are given, we will consider the following three sets:

$$A_1 = \frac{1}{5} \leq y \leq \frac{1}{3}, \quad A_2 = \frac{1}{3} \leq y \leq 1, \quad A_3 = y \geq 1$$

Where A_1 corresponds to $(3,5)$, A_2 to $(1,3)$ and A_3 to $(0,1)$. We can express the CDF for $F_Y(y)$ in terms of $F_X(x)$:

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(\frac{1}{X} \leq y) \\ &= \mathbb{P}(X \geq \frac{1}{y}) \\ &= 1 - \mathbb{P}(X \leq \frac{1}{y}) \\ &= 1 - F_X(\frac{1}{y}) \end{aligned}$$

First, we consider $A_1 : y \in [1/5, 1/3]$, and when we input $1/y$ to $F_X(\cdot)$, it will be in $(3, 5)$. So:

$$\begin{aligned}
F_Y(y) &= 1 - F_X(1/y) \\
&= 1 - \left(\frac{3(\frac{1}{y}) - 9}{8} + \frac{1}{4} \right) \\
&= 1 - \frac{3 - 9y}{8y} - \frac{1}{4} \\
&= \frac{3}{4} + \frac{9y - 3}{8y} \\
&= \frac{15y - 3}{8y}
\end{aligned}$$

Next, we consider $A_2 : y \in [1/3, 1]$. The input to $F_X(\cdot)$ will be in $(1, 3)$:

$$\begin{aligned}
F_Y(y) &= 1 - F_X(1/y) \\
&= 1 - \frac{1}{4} \\
&= \frac{3}{4}
\end{aligned}$$

Next, we consider $A_3 : y \geq 1$. The input to $F_X(\cdot)$ will be in $(0, 1)$:

$$\begin{aligned}
F_Y(y) &= 1 - F_X(1/y) \\
&= 1 - \frac{\frac{1}{y}}{4} \\
&= 1 - \frac{1}{4y}
\end{aligned}$$

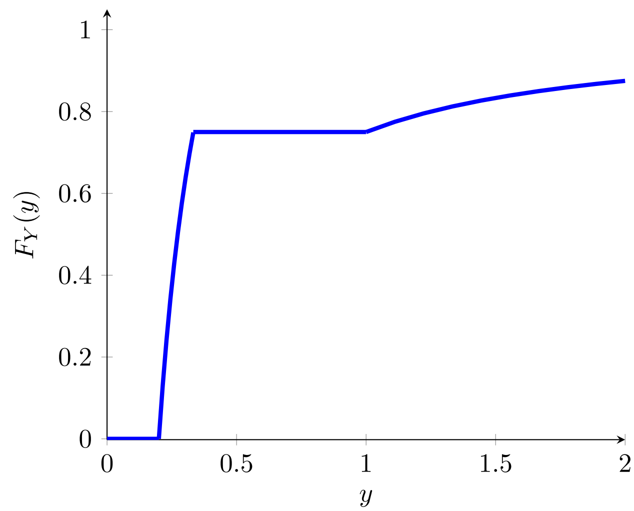
Also, whenever $y < 1/5$, then $1/y > 5$ which means $F_X(\cdot) = 1$, and so:

$$F_Y(y) = 1 - F_X(1/y) = 1 - 1 = 0.$$

This gives a full description of the CDF for $F_Y(y)$.

$$F_Y(y) = \begin{cases} 0 & y < 1/5 \\ \frac{15y - 3}{8y} & 1/5 \leq y \leq 1/3 \\ \frac{3}{4} & 1/3 \leq y \leq 1 \\ 1 - \frac{1}{4y} & y \geq 1 \end{cases}$$

Plot of CDF:



Finally, we can find the PDF of Y. We differentiate each of the parts in the CDF. When $y \in (1/5, 1/3)$:

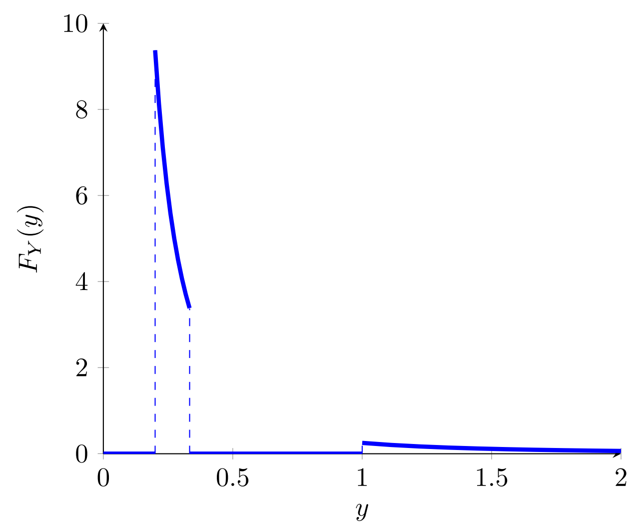
$$\frac{d}{dy}\left(\frac{15y-3}{8y}\right) = \frac{3}{8y^2}$$

When $y \geq 1$:

$$\frac{d}{dy}\left(1 - \frac{1}{4y}\right) = \frac{1}{4y^2}$$

(All other parts are constant, so they become 0). This gives us the PDF and its plot:

$$f_Y(y) = \begin{cases} 0 & y < 1/5 \\ \dfrac{3}{8y^2} & 1/5 \leq y \leq 1/3 \\ 0 & 1/3 < y < 1 \\ \dfrac{1}{4y^2} & y \geq 1 \end{cases}$$



2.5
Let X and Y be discrete RV. X and Y are independent if and only if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all x and y.
Proof.
$\Rightarrow$) Assume that X and Y are independent. That means that for any x, y, we have

$$\mathbb{P}(X = x \cap Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$$

Starting with the definition of the joint pdf:

$$\begin{aligned}
f_{X,Y}(x,y) &= \mathbb{P}(X = x, Y = y) \\
&= \mathbb{P}(X = x \cap Y = y) \\
&= \mathbb{P}(X = x)\mathbb{P}(Y = y) \\
&= f_X(x)f_Y(y)
\end{aligned}$$

Which shows that $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all x and y.

$\Leftarrow$) Assume that $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all x and y. By definition:

$$\begin{aligned}
f_{X,Y}(x,y) &= \mathbb{P}(X = x, Y = y) \\
&= \mathbb{P}(X = x \cap Y = y)
\end{aligned}$$

And,

$$f_X(x)f_Y(y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$$

From our assumption, these are equal, so $\mathbb{P}(X = x \cap Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$ which shows that X and Y are independent.
By implication both ways, the statement is proved. $\square$

2.6
Let X have distribution F and density f, and let A be a subset of the real line, e.g. $A = (a, b)$ for some $a, b \in \mathbb{R}$ and $a < b$. We have the indicator function

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

We will set $Y = I_A(X)$ and find the PDF and CDF of Y. The exercise asks for a probability mass function, but that cannot be correct. Since X has a density f, it is a continuous RV. If $X \sim U(0,1)$ and $A = (0,1)$, then $Y = X$ and it will be a uniform variable with a continuous distribution.

Given what we know, all we can conclude about the PDF of Y is the following:

$$f_Y(y) = \int_A f_X(t)dt$$