Chapter 1. Introduction to Probability
1.3. The Probability Set Function—Proofs of Theorems
Theorem 1.3.1

Theorem 1.3.1. For each event $A \in \mathcal{B}$, $P(A) = 1 - P(A^c)$.

Proof. We have $C = A \cup A^c$. So by Definition 1.3.1,

$$1 = P(C) = P(A \cup A^c) = P(A) + P(A^c),$$

so that $P(A) = 1 - P(A^c)$, as claimed. 

\qed
Theorem 1.3.1. For each event $A \in \mathcal{B}$, $P(A) = 1 - P(A^c)$.

Proof. We have $C = A \cup A^c$. So by Definition 1.3.1,

$$1 = P(C) = P(A \cup A^c) = P(A) + P(A^c),$$

so that $P(A) = 1 - P(A^c)$, as claimed. \qed
Theorem 1.3.2. The probability of the null set is zero; that is, $P(\emptyset) = 0$.

Proof. With $A = \emptyset$ we have $A^c = C$, so by Theorem 1.3.1 we have

$$P(\emptyset) = 1 - P(C) = 1 - 1 = 0,$$

as claimed.
Theorem 1.3.2. The probability of the null set is zero; that is, $P(\emptyset) = 0$.

**Proof.** With $A = \emptyset$ we have $A^c = C$, so by Theorem 1.3.1 we have

$$P(\emptyset) = 1 - P(C) = 1 - 1 = 0,$$

as claimed. $\square$
Theorem 1.3.3

**Theorem 1.3.3.** If $A$ and $B$ are events such that $A \subset B$, then $P(A) \leq P(B)$ (in measure theory, this is called *monotonicity*).

**Proof.** We have $B = A \cup (A^c \cap B)$, so by Definition 1.3.1(3) (countable additivity) $P(B) = P(A) + P(A^c \cap B)$ and by Definition 1.3.1(1), $P(A^c \cap B) \geq 0$ so that

$$P(B) = P(A) + P(A^c \cap B) \geq P(A),$$

or $P(A) \leq P(B)$, as claimed.
Theorem 1.3.3. If $A$ and $B$ are events such that $A \subset B$, then $P(A) \leq P(B)$ (in measure theory, this is called monotonicity).

Proof. We have $B = A \cup (A^c \cap B)$, so by Definition 1.3.1(3) (countable additivity) $P(B) = P(A) + P(A^c \cap B)$ and by Definition 1.3.1(1), $P(A^c \cap B) \geq 0$ so that

$$P(B) = P(A) + P(A^c \cap B) \geq P(A),$$

or $P(A) \leq P(B)$, as claimed.
Theorem 1.3.4

For each event $A \in \mathcal{B}$ we have $0 \leq P(A) \leq 1$.

Proof. Since $\emptyset \subset A \subset C$ then by Theorem 1.3.2, Theorem 1.3.3, and Definition 1.3.1(2),

$$0 = P(\emptyset) \leq P(A) \leq P(C) = 1,$$

or $0 \leq P(A) \leq 1$, as claimed.
Theorem 1.3.4. For each event $A \in \mathcal{B}$ we have $0 \leq P(A) \leq 1$.

Proof. Since $\emptyset \subset A \subset \mathcal{C}$ then by Theorem 1.3.2, Theorem 1.3.3, and Definition 1.3.1(2),

$$0 = P(\emptyset) \leq P(A) \leq P(\mathcal{C}) = 1,$$

or $0 \leq P(A) \leq 1$, as claimed.
Theorem 1.3.5

**Theorem 1.3.5.** If $A$ and $B$ are events in $\mathcal{C}$, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

**Proof.** We have $A \cup B = A \cup (A^c \cap B)$ and $B = (A \cap B) \cup (A^c \cap B)$. Hogg, McKean, and Craig justify these set equalities with the following Venn diagrams:
Theorem 1.3.5. If $A$ and $B$ are events in $\mathcal{C}$, then
\[ P(A \cup B) = P(A) + P(B) - P(A \cap B). \]

Proof. We have $A \cup B = A \cup (A^c \cap B)$ and $B = (A \cap B) \cup (A^c \cap B)$. Hogg, McKean, and Craig justify these set equalities with the following Venn diagrams:
Theorem 1.3.5. If $A$ and $B$ are events in $\mathcal{C}$, then
\[ P(A \cup B) = P(A) + P(B) - P(A \cap B). \]

Proof. We have $A \cup B = A \cup (A^c \cap B)$ and $B = (A \cap B) \cup (A^c \cap B)$. Hogg, McKean, and Craig justify these set equalities with the following Venn diagrams:

By Definition 1.3.1(3) (countable additivity),
\[ P(A \cup B) = P(A) + P(A^c \cap B) \] and \[ P(B) = P(A \cap B) + P(A^c \cap B). \] So (from the second equation) \[ P(A^c \cap B) = P(B) - P(A \cap B) \] and (from the first equation)
\[ P(A \cup B) = P(A) + (P(B) - P(A \cap B)) = P(A) + P(B) = P(A \cap B), \]
as claimed.
**Theorem 1.3.5.** If $A$ and $B$ are events in $\mathcal{C}$, then 

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

**Proof.** We have $A \cup B = A \cup (A^c \cap B)$ and $B = (A \cap B) \cup (A^c \cap B)$. Hogg, McKean, and Craig justify these set equalities with the following Venn diagrams:

By Definition 1.3.1(3) (countable additivity), 

$$P(A \cup B) = P(A) + P(A^c \cap B)$$

and 

$$P(B) = P(A \cap B) + P(A^c \cap B).$$

So (from the second equation) 

$$P(A^c \cap B) = P(B) - P(A \cap B)$$

and (from the first equation)

$$P(A \cup B) = P(A) + (P(B) - P(A \cap B)) = P(A) + P(B) = P(A \cap B),$$

as claimed.
Exercise 1.3.4. If the sample space is \( C = C_1 \cup C_2 \) and if \( P(C_1) = 0.8 \) and \( P(C_2) = 0.5 \), find \( P(C_1 \cap C_2) \).

**Solution.** With \( A = C_1 \) and \( B = C_2 \), we have from Theorem 1.3.4 that

\[
P(C = P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2)
\]

or \( 1 = (0.8) + (0.5) - P(C_1 \cap C_2) \) or \( P(C_1 \cap C_2) = 0.3 \). 

\[\square\]
Exercise 1.3.4. If the sample space is $\mathcal{C} = C_1 \cup C_2$ and if $P(C_1) = 0.8$ and $P(C_2) = 0.5$, find $P(C_1 \cap C_2)$.

Solution. With $A = C_1$ and $B = C_2$, we have from Theorem 1.3.4 that

$$P(C) = P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2)$$

or

$$1 = (0.8) + (0.5) - P(C_1 \cap C_2) \quad \text{or} \quad P(C_1 \cap C_2) = 0.3.$$
Exercise 1.3.6

**Exercise 1.3.6.** If the sample space is $C = \{ c \mid -\infty < c < \infty \}$ and if $C \subset C$ is a set for which the integral $\int_C e^{-|x|} \, dx$ exists, show that this set function is not a probability set function. What constant do we multiply the integral by to make it a probability function?

**Solution.** With $C = C = \mathbb{R}$ we have

$$
\int_\mathbb{R} e^{-|x|} \, dx = \int_\mathbb{R} e^{-|x|} \, dx
$$

$$
= 2 \int_0^\infty e^{-|x|} \, dx \text{ since } e^{-|x|} \text{ is an even function}
$$

$$
= 2 \int_0^\infty e^{-x} \, dx \text{ since } x \geq 0 \text{ here}
$$

$$
= 2 \lim_{b \to \infty} \left( \int_0^b e^{-x} \, dx \right) = 2 \lim_{b \to \infty} \left( -e^{-x} \big|_0^b \right)
$$

$$
= 2 \lim_{b \to \infty} (-e^{-b} + 1) = 2(0 + 1) = 2.
$$
Exercise 1.3.6

Exercise 1.3.6. If the sample space is \( C = \{ c \mid -\infty < c < \infty \} \) and if \( C \subset C \) is a set for which the integral \( \int_C e^{-|x|} \, dx \) exists, show that this set function is not a probability set function. What constant do we multiply the integral by to make it a probability function?

Solution. With \( C = C = \mathbb{R} \) we have

\[
\int_{\mathbb{R}} e^{-|x|} \, dx = \int_{-\infty}^{\infty} e^{-|x|} \, dx \\
= 2 \int_{0}^{\infty} e^{-x} \, dx \quad \text{since } e^{-|x|} \text{ is an even function} \\
= 2 \int_{0}^{\infty} e^{-x} \, dx \quad \text{since } x \geq 0 \text{ here} \\
= 2 \lim_{b \to \infty} \left( \int_{0}^{b} e^{-x} \, dx \right) = 2 \lim_{b \to \infty} \left( -e^{-x} \big|_{0}^{b} \right) \\
= 2 \lim_{b \to \infty} (-e^{-b} + 1) = 2(0 + 1) = 2.
\]
Exercise 1.3.6 (continued 1)

Solution (continued). So $\int_C e^{-|x|} \, dx$ is not a probability set function because applying it to $C = C$ does not yield a probability of a (in violation of Definition 1.3.1(2)). If we define $P(C) = \frac{1}{2} \int_C e^{-|x|} \, dx$ then we have $P(|\mathcal{C}|) = 1$ and Definition 1.3.1(2) is then satisfied. We should feel comfortable with the claim that $P(\emptyset) = \int_{\emptyset} e^{-|x|} \, dx = 0$ (though this is never technically defined for Riemann integrals), so that Definition 1.3.1(1) is satisfied.

But justifying Definition 1.3.1(3), countable additivity, is more complicated. If the integral is a Riemann integral then there are a lot of restrictions on the collection $\mathcal{B}$ of events. If the integral is a Lebesgue integral then the collection of events $\mathcal{B}$ is the $\sigma$-field (or $\sigma$-algebra) of Lebesgue measurable sets, which includes lots of sets of real numbers (probably every subset of $\mathbb{R}$ you can think of... certainly every subset that I can think of... almost...).
Solution (continued). So $\int_{C} e^{-|x|} \, dx$ is not a probability set function because applying it to $C = \cal{C}$ does not yield a probability of a (in violation of Definition 1.3.1(2)). If we define $P(C) = \frac{1}{2} \int_{C} e^{-|x|} \, dx$ then we have $P(\cal{C}) = 1$ and Definition 1.3.1(2) is then satisfied. We should feel comfortable with the claim that $P(\emptyset) = \int_{\emptyset} e^{-|x|} \, dx = 0$ (though this is never technically defined for Riemann integrals), so that Definition 1.3.1(1) is satisfied.

But justifying Definition 1.3.1(3), countable additivity, is more complicated. If the integral is a Riemann integral then there are a lot of restrictions on the collection $\cal{B}$ of events. If the integral is a Lebesgue integral then the collection of events $\cal{B}$ is the $\sigma$-field (or $\sigma$-algebra) of Lebesgue measurable sets, which includes lots of sets of real numbers (probably every subset of $\mathbb{R}$ you can think of... certainly every subset that I can think of... almost...).
Exercise 1.3.6 (continued 2)

**Exercise 1.3.6.** If the sample space is $\mathcal{C} = \{c \mid -\infty < c < \infty\}$ and if $\mathcal{C} \subset \mathcal{C}$ is a set for which the integral $\int_{\mathcal{C}} e^{-|x|} \, dx$ exists, show that this set function is not a probability set function. What constant do we multiply the integral by to make it a probability function?

**Solution (continued).** One of the properties of the Lebesgue integral is countable additivity:

$$\int_{\bigcup_{n=1}^{\infty} A_n} e^{-|x|} = \sum_{n=1}^{\infty} \left( \int_{A_n} e^{-|x|} \right)$$

so that $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$, and Definition 1.3.1(3) is satisfied. For more details on properties of Lebesgue integrals, see my online notes for Real Analysis 1.
Exercise 1.3.9. Determine the probability of being dealt a full house, i.e., three-of-a-kind and two-of-a-kind.

Solution. The suit of the three-of-a-kind can be chosen in \( \binom{13}{1} = 13 \) ways and the suit of the two-of-a-kind can then be chose in \( \binom{12}{1} = 12 \) ways. The three cards in the three-of-a-kind can then be chosen in \( \binom{4}{3} \) ways and the two cards in the two-of-a-kind can then be chosen in \( \binom{4}{2} \) ways. So the probability of being dealt a full house is

\[
\frac{\binom{13}{1} \binom{12}{1} \binom{4}{3} \binom{4}{2}}{\binom{52}{5}} = \frac{(13)(12)(4)(6)}{2,598,960} \approx 0.00144.
\]

\[\square\]
Exercise 1.3.9. Determine the probability of being dealt a full house, i.e., three-of-a-kind and two-of-a-kind.

Solution. The suit of the three-of-a-kind can be chosen in \( \binom{13}{1} = 13 \) ways and the suit of the two-of-a-kind can then be chose in \( \binom{12}{1} = 12 \) ways. The three cards in the three-of-a-kind can then be chosen in \( \binom{4}{3} \) ways and the two cards in the two-of-a-kind can then be chosen in \( \binom{4}{2} \) ways. So the probability of being dealt a full house is

\[
\frac{\binom{13}{1} \binom{12}{1} \binom{4}{3} \binom{4}{2}}{\binom{52}{5}} = \frac{(13)(12)(4)(6)}{2,598,960} \approx 0.00144.
\]
Theorem 1.3.6

Theorem 1.3.6. Continuity of the Probability Functions.
Let $\{C_n\}$ be a nondecreasing sequence of events. Then

$$\lim_{n \to \infty} P(C_n) = P\left( \lim_{n \to \infty} C_n \right) = P\left( \bigcup_{n=1}^{\infty} C_n \right).$$

Let $\{C_n\}$ be a nonincreasing sequence of sets. Then

$$\lim_{n \to \infty} P(C_n) = P\left( \lim_{n \to \infty} C_n \right) = P\left( \bigcap_{n=1}^{\infty} C_n \right).$$

Proof. First, we consider the proof for a nondecreasing sequence.
Theorem 1.3.6

Theorem 1.3.6. Continuity of the Probability Functions.
Let \( \{C_n\} \) be a nondecreasing sequence of events. Then

\[
\lim_{n \to \infty} P(C_n) = P \left( \lim_{n \to \infty} C_n \right) = P \left( \bigcup_{n=1}^{\infty} C_n \right).
\]

Let \( \{C_n\} \) be a nonincreasing sequence of sets. Then

\[
\lim_{n \to \infty} P(C_n) = P \left( \lim_{n \to \infty} C_n \right) = P \left( \bigcap_{n=1}^{\infty} C_n \right).
\]

Proof. First, we consider the proof for a nondecreasing sequence.

Define \( R_1 = C_1 \) and \( R_n = C_n \cap C_{n-1}^c \), for \( n \geq 2 \). Notice that since the events are in a \( \sigma \)-field then \( R_n \) is also an event. Then \( R_m \cap R_n = \emptyset \) for \( m \neq n \) (since, with \( m < n \) say, \( R_m \subset C_m \) but \( R_n \subset C_{n-1}^c \) and since the sequence is nondecreasing then \( C_m \subset C_{n-1} \), here \( m \leq n - 1 \), and so \( C_m \cap C_{n-1}^c = \emptyset \)).
Theorem 1.3.6

Theorem 1.3.6. Continuity of the Probability Functions.
Let \( \{ C_n \} \) be a nondecreasing sequence of events. Then

\[
\lim_{n \to \infty} P(C_n) = P \left( \lim_{n \to \infty} C_n \right) = P \left( \bigcup_{n=1}^{\infty} C_n \right).
\]

Let \( \{ C_n \} \) be a nonincreasing sequence of sets. Then

\[
\lim_{n \to \infty} P(C_n) = P \left( \lim_{n \to \infty} C_n \right) = P \left( \bigcap_{n=1}^{\infty} C_n \right).
\]

Proof. First, we consider the proof for a nondecreasing sequence.

Define \( R_1 = C_1 \) and \( R_n = C_n \cap C_{n-1}^c \), for \( n \geq 2 \). Notice that since the events are in a \( \sigma \)-field then \( R_n \) is also an event. Then \( R_m \cap R_n = \emptyset \) for \( m \neq n \) (since, with \( m < n \) say, \( R_m \subset C_m \) but \( R_n \subset C_{n-1}^c \) and since the sequence is nondecreasing then \( C_m \subset C_{n-1} \), here \( m \leq n - 1 \), and so \( C_m \cap C_{n-1}^c = \emptyset \)).
Proof (continued). Also, $\bigcup_{n=1}^{\infty} R_n = \bigcup_{n=1}^{\infty} C_n$ since $R_n \subset C_n$ for $n \geq 1$ and any $x \in \bigcup_{n=1}^{\infty} C_n$ is in some $C_N$ for a smallest value of $N \in \mathbb{N}$ so that $x \in R_N = C_N \cap C_{N-1}^c$ (since $N$ is the smallest such value then $x \notin X_{N-1}$ and so $x \in C_{N-1}^c$; we need $X_0 = \emptyset$ here). Since $R_n = C_n \cap C_{n-1}^c$ then

$$R_n \cup C_{n-1}^c = (C_n \cap C_{n-1}^c) \cup C_{n-1}$$

or $R_n \cup C_{n-1} = C_n$ and so by Definition 1.3.1(3), countable additivity,

$$P(R_n \cup C_{n-1}) = P(R_n) + P(C_{n-1}) = P(C_n)$$

or $P(R_n) = P(C_n) - P(C_{n-1})$. So for any $N \in \mathbb{N}$ we have

$$\sum_{n=1}^{N} P(R_n) = \sum_{n=1}^{N} (P(C_n) - P(C_{n-1})) = P(C_N) - P(C_0) = P(C_N).$$
Theorem 1.3.6 (continued 1)

**Proof (continued).** Also, $\bigcup_{n=1}^{\infty} R_n = \bigcup_{n=1}^{\infty} C_n$ since $R_n \subset C_n$ for $n \geq 1$ and any $x \in \bigcup_{n=1}^{\infty} C_n$ is in some $C_N$ for a smallest value of $N \in \mathbb{N}$ so that $x \in R_N = C_N \cap C_{N-1}^c$ (since $N$ is the smallest such value then $x \notin X_{N-1}$ and so $x \in C_{N-1}^c$; we need $X_0 = \emptyset$ here). Since $R_n = C_n \cap C_{n-1}^c$ then

$$R_n \cup C_{n-1}^c = (C_n \cap C_{n-1}^c) \cup C_{n-1}$$

or $R_n \cup C_{n-1} = C_n$ and so by Definition 1.3.1(3), countable additivity,

$$P(R_n \cup C_{n-1}) = P(R_n) + P(C_{n-1}) = P(C_n)$$

or $P(R_n) = P(C_n) - P(C_{n-1})$. So for any $N \in \mathbb{N}$ we have

$$\sum_{n=1}^{N} P(R_n) = \sum_{n=1}^{N} (P(C_n) - P(C_{n-1})) = P(C_N) - P(C_0) = P(C_N).$$
Theorem 1.3.6 (continued 2)

Proof (continued). . .

\[ \sum_{n=1}^{N} P(R_n) = \sum_{n=1}^{N} (P(C_n) - P(C_{n-1})) = P(C_N) - P(C_0) = P(C_N). \]

So

\[
P\left( \lim_{n \to \infty} C_n \right) = P\left( \bigcup_{n=1}^{\infty} C_n \right) = P\left( \bigcup_{n=1}^{\infty} R_n \right)
\]

\[ = \sum_{n=1}^{\infty} P(R_n) \text{ by Definition 1.3.1(3), countable additivity} \]

\[ = \lim_{N \to \infty} \left( \sum_{n=1}^{N} P(R_n) \right) = \lim_{N \to \infty} P(C_N) = \lim_{n \to \infty} P(C_n), \]

as claimed. \qed
Theorem 1.3.7. Boole’s Inequality/Countable Subadditivity.
Let \( \{C_n\} \) be an arbitrary sequence of events. Then

\[
P \left( \bigcup_{n=1}^{\infty} C_n \right) \leq \sum_{n=1}^{\infty} P(C_n).
\]

**Proof.** Define \( D_n = \bigcup_{i=1}^{n} C_i \). Then \( \{D_n\} \) is an increasing sequence of events that converge to \( \bigcup_{n=1}^{\infty} C_n \). Also \( D_j = D_{j-1} \cup C_j \) for all \( j \geq 2 \). So by Theorem 1.3.5,

\[
P(D_j) = P(D_{j-1} \cup C_j) \leq P(D_{j-1}) + P(C_j),
\]

or \( P(D_j) - P(D_{j-1}) \leq P(C_j) \).
Theorem 1.3.7. Boole’s Inequality/Countable Subadditivity.

Let \( \{C_n\} \) be an arbitrary sequence of events. Then

\[
P \left( \bigcup_{n=1}^{\infty} C_n \right) \leq \sum_{n=1}^{\infty} P(C_n).
\]

**Proof.** Define \( D_n = \bigcup_{i=1}^{n} C_i \). Then \( \{D_n\} \) is an increasing sequence of events that converge to \( \bigcup_{n=1}^{\infty} C_n \). Also \( D_j = D_{j-1} \cup C_j \) for all \( j \geq 2 \). So by Theorem 1.3.5,

\[
P(D_j) = P(D_{j-1} \cup C_j) \leq P(D_{j-1}) + P(C_j),
\]

or \( P(D_j) - P(D_{j-1}) \leq P(C_j) \).
Theorem 1.3.7. Boole’s Inequality/Countable Subadditivity.
Let \( \{ C_n \} \) be an arbitrary sequence of events. Then

\[
P(\bigcup_{n=1}^{\infty} C_n) \leq \sum_{n=1}^{\infty} P(C_n).
\]

Proof (continued). So by Theorem 1.3.1,

\[
P(\bigcup_{i=1}^{\infty} C_n) = (\bigcup_{i=1}^{\infty} D_c) = \lim_{n \to \infty} P(D_n)
\]

\[
= \lim_{n \to \infty} \left( P(D_1) + \sum_{j=2}^{n} (P(D_j) - P(D_{j-1})) \right) \leq \lim_{n \to \infty} \left( P(C_1) + \sum_{j=2}^{\infty} P(C_j) \right)
\]

\[
= \sum_{n=1}^{\infty} P(C_n),
\]

as claimed.