

MA 528 Measure Theoretic Probability Notes

Hubert J. Farnsworth

September 24, 2021

Chapter 2 Axioms of Probability

Notes

Let Ω be an abstract space and 2^Ω the power set of Ω . Let $\mathcal{A} \subset 2^\Omega$.

Definition 2.1

\mathcal{A} is an algebra if it satisfies (1), (2), and (3) below. \mathcal{A} is a σ -algebra if it satisfies (1), (2), and (4) below.

1. $\emptyset \in \mathcal{A}$ and $\Omega \in \mathcal{A}$
2. If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$
3. \mathcal{A} If $A_1, \dots, A_n \in \mathcal{A}$, then $\bigcup_{i=1}^n A_i \in \mathcal{A}$ and $\bigcap_{i=1}^n A_i \in \mathcal{A}$
4. If the countable sequence $A_1, A_2, \dots \in \mathcal{A}$, then $\bigcup_{i=1}^\infty A_i \in \mathcal{A}$ and $\bigcap_{i=1}^\infty A_i \in \mathcal{A}$

Note If (2) holds then $\emptyset \in \mathcal{A}$ implies $\Omega \in \mathcal{A}$ and $\Omega \in \mathcal{A}$ implies $\emptyset \in \mathcal{A}$. If (1) and (4) are satisfied then (3) is satisfied (every σ -algebra is an algebra).

Definition 2.2

If $\mathcal{C} \subset 2^\Omega$, the σ -algebra generated by \mathcal{C} , and written $\sigma(\mathcal{C})$, is the smallest σ -algebra containing \mathcal{C} .

Note $\sigma(\mathcal{C})$ always exists. See Exercise 2.2.

Theorem 2.1

The Borel σ -algebra of \mathbb{R} , $\mathfrak{B}(\mathbb{R})$, which is the smallest σ -algebra containing the open sets in \mathbb{R} (or equivalently containing the closed sets in \mathbb{R}), is generated by intervals of the form $(-\infty, a]$ where $a \in \mathbb{Q}$.

Proof: Let \mathcal{C} denote the set of all open intervals. Since every open set in \mathbb{R} is the countable union of open intervals, $\sigma(\mathcal{C}) = \mathfrak{B}$.

Let \mathcal{D} denote the set of all intervals of the form $(-\infty, a]$, $a \in \mathbb{Q}$. Let $(a, b) \in \mathcal{C}$. Let $(a_n)_{n \geq 1}$ be sequence of rational numbers decreasing strictly to a and $(b_n)_{n \geq 1}$ a sequence of rational numbers increasing strictly to b . Then

$$(a, b) = \bigcup_{n=1}^\infty (a_n, b_n] = \bigcup_{n=1}^\infty ((-\infty, a_n]^c \cap (-\infty, b_n]) \implies \mathcal{C} \subset \sigma(\mathcal{D}) \implies \sigma(\mathcal{C}) \subset \sigma(\mathcal{D}).$$

Every $D \in \mathcal{D}$ is a closed set since D^c is open. Since \mathcal{D} is a subset of the set of all closed sets in \mathbb{R} , $\sigma(\mathcal{D})$ is contained in the sigma algebra generated by set of closed sets in \mathbb{R} . That is, $\sigma(\mathcal{D}) \subset \mathfrak{B}$.

$$\mathfrak{B} = \sigma(\mathcal{C}) \subset \sigma(\mathcal{D}) \subset \mathfrak{B} \implies \sigma(\mathcal{D}) = \mathfrak{B}.$$

Definition 2.3

A probability measure defined on a σ -algebra \mathcal{A} is a function $P : \mathcal{A} \rightarrow [0, 1]$ that satisfies:

1. $P(\Omega) = 1$
2. For every countable sequence $(A_n)_{n \geq 1}$ of elements of \mathcal{A} , pairwise disjoint,

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

Condition (2) is called countable additivity. The number $P(A)$ is called the probability of event A . The more rudimentary property that $A, B \in \mathcal{A}, A \cap B = \emptyset \implies P(A \cup B) = P(A) + P(B)$ is called additivity. Additivity with respect to two sets (A and B) implies additivity with respect to any finite collection of disjoint sets (A_1, \dots, A_m) .

Theorem 2.2

If P is a probability measure on (Ω, \mathcal{A}) , then:

- (i) $P(\emptyset) = 0$
- (ii) P is additive.

Proof: To prove (i), use (2) of definition 2.3 and the fact that the codomain of P is $[0, 1]$.

$$P(\emptyset) = P\left(\bigcup_{n=1}^{\infty} \emptyset\right) = \sum_{n=1}^{\infty} P(\emptyset) = P(\emptyset) \sum_{n=1}^{\infty} 1 \implies P(\emptyset) = 0.$$

To prove (ii), suppose $A, B \in \mathcal{A}, A \cap B = \emptyset$. Let $A_1 = A, A_2 = B$, and let $A_n = \emptyset$ for $n \geq 3$.

$$P(A \cup B) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) = P(A) + P(B) + \sum_{n=1}^{\infty} P(\emptyset) = P(A) + P(B).$$

Note It follows from Theorem 2.2 that if $A, C \in \mathcal{A}$ with $A \subset C$ then $P(A) \leq P(C)$. To prove this, let $B = C \setminus A$ so that $A \cap B = \emptyset$ and $P(A) \leq P(A) + P(B) = P(A \cup B) = P(C)$.

Theorem 2.3

Suppose $P : \mathcal{A} \rightarrow [0, 1]$ satisfies (1) of definition 2.3 and P is (finitely) additive. The following are equivalent.

- (i) Axiom (2) of definition 2.3.
- (ii) $A_n \downarrow \emptyset \implies P(A_n) \downarrow 0$.
- (iii) $A_n \downarrow A \implies P(A_n) \downarrow P(A)$.
- (iv) $A_n \uparrow \Omega \implies P(A_n) \uparrow 1$.
- (v) $A_n \uparrow A \implies P(A_n) \uparrow P(A)$.

Proof:

(iii) \iff (v)

Assume (iii) and suppose $A_n \uparrow A$. Then $A_n^c \downarrow A^c$ so $P(A_n^c) \downarrow P(A^c)$. But then $P(A_n) = (1 - P(A_n^c)) \uparrow (1 - P(A^c)) = P(A)$. Proving the reverse is similar.

(ii) \iff (iv)

Let $A = \Omega$ so that $A^c = \emptyset$ and apply the previous result.

(iv) \iff (v)

Assuming (v) holds, $A_n \uparrow A = \Omega \implies P(A_n) \uparrow P(\Omega) = 1$. Therefore (v) \implies (iv). Now assume (iv) and suppose $A_n \uparrow A$. Define $B_n = A_n \cup A^c$ so that $B_n \uparrow \Omega$. Since $A_n \cap A^c = \emptyset$ for all n , $P(B_n) = P(A_n) + P(A^c)$ for all n . Since $A_n \subset A_{n+1}$ for each n , $P(A_n) \uparrow P(A)$.

$$1 = \lim P(B_n) = P(A^c) + \lim P(A_n) \implies \lim P(A_n) = 1 - P(A^c) = P(A) \implies P(A_n) \uparrow P(A).$$

(i) \iff (v)

Assume (v) holds and suppose $(A_n)_{n \geq 1}$ are pairwise disjoint. Define $B_n = \bigcup_{p=1}^n A_p$ and $B = \bigcup_{n=1}^{\infty} A_n$. We have $P(B_n) = \sum_{p=1}^n P(A_p)$ for each n by finite additivity. By (v) $P(B_n) \uparrow P(B)$.

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P(B) = \lim P(B_n) = \lim \sum_{p=1}^n P(A_p) = \sum_{p=1}^{\infty} P(A_p).$$

Therefore (v) \implies (i). Next assume (i) and suppose $A_n \uparrow A$. Define the sequence of disjoint set $(B_n)_{n \geq 1}$:

$$\begin{aligned} B_1 &= A_1 \\ B_2 &= A_2 \setminus A_1 \\ &\vdots \\ B_n &= A_n \setminus A_{n-1} \end{aligned}$$

Then $\bigcup_{n=1}^{\infty} B_n = A$ and $P(A) = \sum_{n=1}^{\infty} P(B_n)$ by (i). Since $A_n \subset A_{n+1}$, $P(A_n) \leq P(A_{n+1})$ for each n .

$$\lim P(A_n) = \lim \sum_{p=1}^n P(B_p) = \sum_{p=1}^{\infty} P(B_p) = P(A) \implies P(A_n) \uparrow P(A).$$

Theorem 2.4

Let P be a probability measure on \mathcal{A} and A_n a sequence of sets in \mathcal{A} with $A_n \rightarrow A$. Then $A \in \mathcal{A}$ and $\lim P(A_n) = P(A)$.

Proof:

$$\begin{aligned} \limsup A_n &:= \bigcap_{n=1}^{\infty} \bigcup_{m \geq n} A_m \\ \liminf A_n &:= \bigcup_{n=1}^{\infty} \bigcap_{m \geq n} A_m \end{aligned}$$

Since \mathcal{A} is a σ -algebra and thus closed under countable union and closed under countable intersection, $\limsup A_n \in \mathcal{A}$ and $\liminf A_n \in \mathcal{A}$.

By hypothesis $A_n \rightarrow A$ so $\lim 1_{A_n}(\omega) = 1_A(\omega)$ for each $\omega \in \Omega$. This is equivalent to saying $A = \limsup A_n = \liminf A_n$. Therefore $A \in \mathcal{A}$.

Let $B_n = \cap_{m \geq n} A_m$ and $C_n = \cup_{m \geq n} A_m$. Then $B_n \uparrow A$ and $C_n \downarrow A$ so that $\lim P(B_n) = \lim P(C_n) = P(A)$ by Theorem 2.3. Since $B_n \subset A_n \subset C_n$ for each n , $P(B_n) \leq P(A_n) \leq P(C_n)$ for each n .

$$P(A) = \lim P(B_n) \leq \lim P(A_n) \leq \lim P(C_n) = P(A) \implies \lim P(A_n) = P(A).$$

Exercises

Exercise 2.1

Let Ω be a finite set. Show that the set of all subsets of Ω , 2^Ω , is also finite and that it is a σ -algebra.

Answer:

Claim: If $|\Omega| = n$ for some nonnegative integer n , $|2^\Omega| = 2^{|\Omega|}$.

Proof (Induction): If $\Omega = \emptyset$, \emptyset is the only subset of Ω and $|2^\Omega| = 1 = 2^0 = 2^{|\Omega|}$. Assume the claim holds for a set of cardinality n , $n \geq 0$, and consider the case of $|\Omega| = n + 1$. Select an element $\omega \in \Omega$ and consider all $A \subset \Omega$ such that $\omega \notin A$. By the inductive hypothesis there are 2^n such subsets of Ω . For each of these subsets, we build a new subset of Ω by including ω . In this way we find another 2^n subsets of Ω . Since for any subset A of Ω , either $\omega \in A$ or $\omega \notin A$, conclude that $|2^\Omega| = 2 \cdot 2^n = 2^{n+1} = 2^{|\Omega|}$.

By the claim above, if Ω is a finite set then $|2^\Omega| = 2^{|\Omega|} < +\infty$.

To show that 2^Ω is a σ -algebra, check that 2^Ω satisfies axioms (1), (2), and (4) from Definition 2.1.

1. Since $\emptyset \subseteq \Omega$ and $\Omega \subseteq \Omega$, $\emptyset, \Omega \in 2^\Omega$.
2. Suppose $A \in 2^\Omega$. Then $A \subseteq \Omega$ and $A^c = \{\omega \in \Omega : \omega \notin A\} \subseteq \Omega$. Therefore $A^c \in 2^\Omega$ as well.
4. Suppose A_1, A_2, \dots is a countable sequence of events in 2^Ω . Since each A_k is a subset of Ω ,

$$\begin{aligned} \bigcup_{k=1}^{\infty} A_k &= \{\omega \in \Omega : \omega \in A_k \text{ for some } k\} \subseteq \Omega \implies \bigcup_{k=1}^{\infty} A_k \in 2^\Omega, \\ \bigcap_{k=1}^{\infty} A_k &= \{\omega \in \Omega : \omega \in A_k \text{ for all } k\} \subseteq \Omega \implies \bigcap_{k=1}^{\infty} A_k \in 2^\Omega. \end{aligned}$$

Exercise 2.2

Let $(G_\alpha)_{\alpha \in A}$ be an arbitrary family of σ -algebras defined on an abstract space Ω . Show that $H = \cap_{\alpha \in A} G_\alpha$ is also a σ -algebra.

Answer:

1. Since each G_α is a σ -algebra, $\emptyset, \Omega \in G_\alpha$ for each $\alpha \in A$. Thus $\emptyset, \Omega \in H$.
2. Suppose $A \in H$. Then $A \in G_\alpha$ for each α so that $A^c \in G_\alpha$ for each α . Thus $A^c \in H$.
4. Suppose A_1, A_2, \dots is a countable sequence of events in H . For each $\alpha \in A$, A_1, A_2, \dots is a countable sequence of events in G_α . This means

$$\bigcup_{k=1}^{\infty} A_k, \bigcap_{k=1}^{\infty} A_k \in G_\alpha \text{ for each } \alpha \in A \implies \bigcup_{k=1}^{\infty} A_k, \bigcap_{k=1}^{\infty} A_k \in H.$$

Exercise 2.4

Let \mathcal{A} be a σ -algebra and $(A_n)_{n \geq 1}$ a sequence of events in \mathcal{A} . Show that

$$\liminf_{n \rightarrow \infty} A_n \in \mathcal{A}; \quad \limsup_{n \rightarrow \infty} A_n \in \mathcal{A}; \quad \text{and} \quad \liminf_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} A_n.$$

Answer: Recall the definitions

$$\begin{aligned} \liminf_{n \rightarrow \infty} A_n &= \bigcup_{n=1}^{\infty} \bigcap_{m \geq n} A_m, \\ \limsup_{n \rightarrow \infty} A_n &= \bigcap_{n=1}^{\infty} \bigcup_{m \geq n} A_m. \end{aligned}$$

For each positive integer n , $(A_m)_{m \geq n}$ is a countable sequence of events in \mathcal{A} . By the definition of a σ -algebra, this means both $\bigcap_{m \geq n} A_m$ and $\bigcup_{m \geq n} A_m$ belong to \mathcal{A} as \mathcal{A} is closed under countable intersections and unions. But then $(\bigcap_{m \geq n} A_m)_{n \geq 1}$ and $(\bigcup_{m \geq n} A_m)_{n \geq 1}$ are each countable sequences of events in \mathcal{A} so that again by the definition of a σ -algebra

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{m \geq n} A_m \in \mathcal{A} \quad \text{and} \quad \limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m \geq n} A_m.$$

Suppose $a \in \liminf_{n \rightarrow \infty} A_n$. Then there exists a positive integer n such that $a \in \bigcap_{m \geq n} A_m$. Since $a \in A_m$ for every $m \geq n$, $a \in \bigcup_{i \geq k} A_i$ for each k (no matter how large we choose k , there is an $m \geq n$ such that $m \geq k$ so that $a \in A_m \subseteq \bigcup_{i \geq k} A_i$). Thus $a \in \bigcap_{i=1}^{\infty} \bigcup_{k \geq i} A_k = \limsup_{n \rightarrow \infty} A_n$, which establishes $\liminf_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} A_n$.

Exercise 2.5

Let $(A_n)_{n \geq 1}$ be a sequence of sets. Show that

$$\limsup_{n \rightarrow \infty} 1_{A_n} - \liminf_{n \rightarrow \infty} 1_{A_n} = 1_{\limsup_n A_n \setminus \liminf_n A_n}.$$

Answer: Assume that whenever we index in what follows, n comes from the set of positive integers.

Lemma: For all $\omega \in \Omega$,

$$\begin{aligned} \liminf_{n \rightarrow \infty} 1_{A_n}(\omega) &= 1_{\liminf_n A_n}, \\ \limsup_{n \rightarrow \infty} 1_{A_n}(\omega) &= 1_{\limsup_n A_n}. \end{aligned}$$

Proof: Note that $1_{\bigcap_n B_n} = \inf_n 1_{B_n}$ and $1_{\bigcup_n B_n} = \sup_n 1_{B_n}$. This follows from,

$$\begin{aligned} 1_{\bigcap_n B_n}(\omega) = 1 &\iff \omega \in \bigcap_n B_n \\ &\iff \forall n, \omega \in B_n \\ &\iff \forall n, 1_{B_n}(\omega) = 1 \\ &\iff \inf_n 1_{B_n}(\omega) = 1. \end{aligned}$$

$$\begin{aligned} 1_{\bigcup_n B_n}(\omega) = 1 &\iff \omega \in \bigcup_n B_n \\ &\iff \exists n, \omega \in B_n \\ &\iff \exists n, 1_{B_n}(\omega) = 1 \\ &\iff \sup_n 1_{B_n}(\omega) = 1. \end{aligned}$$

With only minor changes to the above, we see that $1_{\cap_{m \geq n} B_m} = \inf_{m \geq n} 1_{B_m}$ and $1_{\cup_{m \geq n} B_m} = \sup_{m \geq n} 1_{B_m}$ as well. Therefore,

$$\begin{aligned} 1_{\liminf_n A_n} &= 1_{\cup_n \cap_{m \geq n} A_m} = \sup_n 1_{\cap_{m \geq n} A_m} = \sup_n \inf_{m \geq n} 1_{A_m} = \liminf_{n \rightarrow \infty} 1_{A_n}, \\ 1_{\limsup_n A_n} &= 1_{\cap_n \cup_{m \geq n} A_m} = \inf_n 1_{\cup_{m \geq n} A_m} = \inf_n \sup_{m \geq n} 1_{A_m} = \limsup_{n \rightarrow \infty} 1_{A_n}. \end{aligned}$$

Lemma: For $A, B \subset \Omega$, $1_{A \setminus B} = 1_A - 1_{A \cap B}$.

Proof: For any $\omega \in \Omega$,

$$\begin{aligned} 1_{A \setminus B}(\omega) = 1 &\iff \omega \in A, \omega \notin B \\ &\iff 1_A(\omega) = 1 \text{ and } 1_{A \cap B}(\omega) = 0 \\ &\iff 1_A(\omega) - 1_{A \cap B}(\omega) = 1. \end{aligned}$$

Using the two lemmas and the result $\liminf_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} A_n$ from Exercise 2.4,

$$\begin{aligned} \limsup_{n \rightarrow \infty} 1_{A_n} - \liminf_{n \rightarrow \infty} 1_{A_n} &= 1_{\limsup_n A_n} - 1_{\liminf_n A_n} \\ &= 1_{\limsup_n A_n} - 1_{\liminf_n A_n \cap \limsup_n A_n} \\ &= 1_{\limsup_n A_n \setminus \liminf_n A_n} \end{aligned}$$

Exercise 2.6

Let \mathcal{A} be a σ -algebra of subsets of Ω and let $B \in \mathcal{A}$. Show that $\mathcal{F} = \{A \cap B : A \in \mathcal{A}\}$ is a σ -algebra of subsets of B . Is it still true when B is a subset of Ω that does not belong to \mathcal{A} ?

Answer: To prove that $\mathcal{F} \subseteq 2^B$ is a σ -algebra of subsets of B , verify axioms (1), (2), and (4) of Definition 2.1.

1. To prove that \mathcal{F} is a σ -algebra of subsets of B , check that $\emptyset, B \in \mathcal{F}$ (no need to check $\Omega \in \mathcal{F}$). Since $\emptyset, B \in \mathcal{A}$, $\emptyset = \emptyset \cap B \in \mathcal{F}$ and $B = B \cap B \in \mathcal{F}$.
2. Let $F \in \mathcal{F}$ with $F = A \cap B$ for some $A \in \mathcal{A}$. Since $A, B \in \mathcal{A}$, $F \in \mathcal{A}$ and so $B \setminus F = B \cap F^c \in \mathcal{A}$ as well. Since $F \subseteq B$, the complement of F relative to B is $F^c = B \setminus F = (B \setminus F) \cap B \in \mathcal{F}$.
4. Let $(F_n)_{n \geq 1}$ be a sequence of sets in \mathcal{F} with $F_n = A_n \cap B$ for $A_n \in \mathcal{A}$. Because \mathcal{A} is closed under countable unions and intersections,

$$\begin{aligned} \bigcup_{k=1}^{\infty} F_k &= \bigcup_{k=1}^{\infty} (A_k \cap B) = \left(\bigcup_{k=1}^{\infty} A_k \right) \cap B \in \mathcal{F}, \\ \bigcap_{k=1}^{\infty} F_k &= \bigcap_{k=1}^{\infty} (A_k \cap B) = \left(\bigcap_{k=1}^{\infty} A_k \right) \cap B \in \mathcal{F}. \end{aligned}$$

Exercise 2.7

Let f be a function mapping Ω to another space E with a σ -algebra \mathcal{E} . Let $\mathcal{A} = \{A \subset \Omega : \exists B \in \mathcal{E}, A = f^{-1}(B)\}$. Show that \mathcal{A} is a σ -algebra on Ω .

1. $\emptyset \in \mathcal{E}$ since \mathcal{E} is a σ -algebra. To see that $f^{-1}(\emptyset) = \emptyset$ suppose instead $f^{-1}(\emptyset) = A \neq \emptyset$. This would mean there is $a \in A \subseteq \Omega$ such that $f(a) \in \emptyset$, contradicting the definition of \emptyset . Thus $\emptyset \in \mathcal{A}$. Also $\Omega = \emptyset^c = \Omega \setminus \emptyset \in \mathcal{A}$ by (2), which is proved below.
2. Suppose $A \in \mathcal{A}$ with $A = f^{-1}(B)$. Then $A^c = (f^{-1}(B))^c = f^{-1}(B^c) \in \mathcal{A}$ since $B^c \in \mathcal{E}$ and

$$x \in (f^{-1}(B))^c \iff x \notin f^{-1}(B) \iff f(x) \notin B \iff f(x) \in B^c \iff x \in f^{-1}(B^c).$$

4. Let $(A_n)_{n \geq 1}$ be a sequence of sets in \mathcal{A} with $A_n = f^{-1}(B_n)$.

$$\bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} f^{-1}(B_k) = f^{-1}\left(\bigcup_{k=1}^{\infty} B_k\right) \in \mathcal{A}$$

as $\bigcup_{k=1}^{\infty} B_k \in \mathcal{E}$ and

$$x \in f^{-1}\left(\bigcup_{k=1}^{\infty} B_k\right) \iff f(x) \in \bigcup_{k=1}^{\infty} B_k \iff \exists k, f(x) \in B_k \iff \exists k, x \in f^{-1}(B_k) \iff x \in \bigcup_{k=1}^{\infty} f^{-1}(B_k).$$

Using this result and the fact that \mathcal{A} is closed under complement by (2), $\bigcap_{k=1}^{\infty} A_k \in \mathcal{A}$ as well.

Exercise 2.8

Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a continuous function, and let $\mathcal{A} = \{A \subseteq \mathbb{R} : \exists B \in \mathfrak{B}, A = f^{-1}(B)\}$ where \mathfrak{B} are the Borel subsets of the range space \mathbb{R} . Show that $\mathcal{A} \subset \mathfrak{B}$, the Borel subsets of the domain space \mathbb{R} .

Answer: Suppose $A \in \mathcal{A}$ so that $A = f^{-1}(B)$ for some $B \in \mathfrak{B}$. Since $B \in \mathfrak{B}$, B is the result of applying a countable number of complements, unions, and/or intersections to a collection of open intervals in \mathbb{R} . Since f^{-1} commutes with these set operations, A is the result of applying countably many set operations to the inverse images of open intervals in \mathbb{R} . Since f is continuous, the inverse image of an open interval is also an open interval. Applying countably many set operations to a collection of open intervals leaves a Borel set. Thus $A \in \mathfrak{B}$.

Exercise 2.15

Let \mathcal{A} be a σ -algebra on the space Ω and P a probability defined on (Ω, \mathcal{A}) . Let $A_i \in \mathcal{A}$ be a sequence of events. Show that

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i) \quad \forall n,$$

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

Answer:

Prove the first inequality (the finite case) by induction. For only one set $A \in \mathcal{A}$ equality holds and thus also inequality. Assume that the inequality holds and consider a sequence A_1, \dots, A_{n+1} . Let $A'_{n+1} = A_{n+1} \setminus (A_1 \cup \dots \cup A_n)$ for some $n \geq 1$. Then $A'_{n+1} \cap (A_1 \cup \dots \cup A_n) = \emptyset$ and $A'_{n+1} \subseteq A_{n+1}$.

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= P\left(\left(\bigcup_{i=1}^n A_i\right) \cup A'_{n+1}\right) \\ &= \sum_{i=1}^n P(A_i) + P(A'_{n+1}) \\ &\leq \sum_{i=1}^n P(A_i) + P(A_{n+1}) \\ &= \sum_{i=1}^{n+1} P(A_i) \end{aligned}$$

To prove countable subadditivity, let

$$\begin{aligned} E_1 &:= A_1 \\ E_2 &:= A_2 \setminus E_1 \\ E_3 &:= A_3 \setminus (E_1 \cup E_2) \\ E_4 &:= A_4 \setminus (E_1 \cup E_2 \cup E_3) \\ &\vdots \\ E_n &:= A_n \setminus \left(\bigcup_{i=1}^{n-1} E_i\right) \qquad \qquad \qquad \vdots \end{aligned}$$

The E_i are disjoint with $\bigcup_{i=1}^{\infty} E_i = \bigcup_{i=1}^{\infty} A_i$. To see that these unions are equal, first note that $E_i \subseteq A_i$ for each i so $\bigcup_{i=1}^{\infty} E_i \subseteq \bigcup_{i=1}^{\infty} A_i$. If $x \in \bigcup_{i=1}^{\infty} A_i$ we may choose by the Well-Ordering Principle the least index i such that $x \in A_i$. Then $x \in A_i$ and $x \notin A_j$ for $j < i$. Thus $x \in E_i \subseteq \bigcup_{i=1}^{\infty} E_i$. Since $E_i \subseteq A_i$ for each i , $P(E_i) \leq P(A_i)$ for each i and

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

Exercise 2.17

Suppose that Ω is an infinite set (countable or not), and let \mathcal{A} be the family of all subsets which are either finite or have a finite complement. Show that \mathcal{A} is an algebra, but not a σ -algebra.

1. Both $\emptyset \in \mathcal{A}$ and $\Omega \in \mathcal{A}$ as $\Omega^c = \emptyset$ is finite.
2. Suppose $A \in \mathcal{A}$. If A is finite, then $A^c \in \mathcal{A}$ since $(A^c)^c = A$. If A is infinite, then A^c must be finite so $A^c \in \mathcal{A}$.
3. Suppose $A_1, \dots, A_n \in \mathcal{A}$. If all of the A_i are finite, then the finite union of finite sets $\bigcup_{i=1}^n A_i$ is finite. If there is a set A_k , $1 \leq k \leq n$ such that A_k is infinite then $\bigcup_{i=1}^n A_i$ is not finite. However, A_k^c must be finite and $(\bigcup_{i=1}^n A_i)^c = \bigcap_{i=1}^n A_i^c \subset A_k^c$. This shows that $(\bigcup_{i=1}^n A_i)^c$ is finite so $\bigcup_{i=1}^n A_i \in \mathcal{A}$. Since \mathcal{A} is closed under complement and finite union, $\bigcap_{i=1}^n A_i \in \mathcal{A}$ as well.

Since \mathcal{A} satisfies axioms (1), (2), and (3), \mathcal{A} is an algebra. However \mathcal{A} is not a σ -algebra since it fails axiom (4):

4. \mathcal{A} is not necessarily closed under countable union. Either Ω is countably infinite or uncountable.

- If Ω is countably infinite, we can list the elements of $\Omega = \{\omega_1, \omega_2, \dots\}$. Let $A_i = \{\omega_{2i}\}$ for each positive integer i . Then both $\bigcup_{i=1}^{\infty} A_i = \{\omega_2, \omega_4, \dots\}$ is infinite and $(\bigcup_{i=1}^{\infty} A_i)^c = \{\omega_1, \omega_3, \dots\}$ is infinite so $\bigcup_{i=1}^{\infty} A_i \notin \mathcal{A}$.
- If Ω is uncountable, let $(A_i)_{i \geq 1}$ be a sequence of pairwise disjoint singleton sets. Then $\bigcup_{i=1}^{\infty} A_i$ has countably infinitely many elements and $(\bigcup_{i=1}^{\infty} A_i)^c$ must be uncountable (since Ω is uncountable). Since neither of $\bigcup_{i=1}^{\infty} A_i$, $(\bigcup_{i=1}^{\infty} A_i)^c$ is finite, $\bigcup_{i=1}^{\infty} A_i \notin \mathcal{A}$.

Chapter 3 Conditional Probability and Independence

Notes

Definition 3.1

1. Two events A and B are independent if $P(A \cap B) = P(A)P(B)$.
2. A (possibly infinite) collection of events $(A_i)_{i \in I}$ is an independent collection if for every finite subset J of I ,

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i).$$

The collection $(A_i)_{i \in I}$ is said to be mutually independent.

Theorem 3.1

If A and B are independent, so also are A and B^c , A^c and B , A^c and B^c .

Proof: For A and B^c ,

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A \cap B) \quad (\text{Exercise 2.12}) \\ &= P(A) - P(A)P(B) \quad (\text{Definition 3.1}) \\ &= P(A)(1 - P(B)) \\ &= P(A)P(B^c) \quad (\text{Exercise 2.11}). \end{aligned}$$

To prove that B and A^c are independent, switch A with B and B^c with A^c and repeat the previous argument. For A^c and B^c ,

$$\begin{aligned} P(A^c \cap B^c) &= P((A \cup B)^c) \\ &= 1 - P(A \cup B) \\ &= 1 - (P(A) + P(B) - P(A \cap B)) \quad (\text{Exercise 2.10}) \\ &= 1 - P(A) - P(B) + P(A)P(B) \\ &= (1 - P(B)) - P(A)(1 - P(B)) \\ &= P(B^c) - P(A)P(B^c) \quad (\text{Exercise 2.11}) \\ &= (1 - P(A))P(B^c) \\ &= P(A^c)P(B^c). \end{aligned}$$

Example Let $\Omega = \{1, 2, 3, 4\}$, and $A = 2^\Omega$. Let $P(i) = \frac{1}{4}$, where $i = 1, 2, 3, 4$. Let $A = \{1, 2\}$, $B = \{1, 3\}$, $C = \{2, 3\}$. Then A, B, C are pairwise independent but are not independent.

$$\begin{aligned} P(A \cap B) &= P(\{1\}) = \frac{1}{4} = \frac{1}{2} \frac{1}{2} = P(A)P(B), \\ P(A \cap C) &= P(\{2\}) = \frac{1}{4} = \frac{1}{2} \frac{1}{2} = P(A)P(C), \\ P(B \cap C) &= P(\{3\}) = \frac{1}{4} = \frac{1}{2} \frac{1}{2} = P(B)P(C), \\ P(A \cap B \cap C) &= P(\emptyset) = 0 \neq \left(\frac{1}{2}\right)^3 = P(A)P(B)P(C). \end{aligned}$$

Definition 3.2

Let A, B be events, $P(B) > 0$. The conditional probability of A given B is $P(A|B) = P(A \cap B)/P(B)$.

Theorem 3.2

Suppose $P(B) > 0$.

1. A and B are independent if and only if $P(A|B) = P(A)$.
2. The operation $A \rightarrow P(A|B)$ from $\mathcal{A} \rightarrow [0, 1]$ defines a new probability measure on \mathcal{A} , called the conditional probability measure given B .

Proof:

1. If A and B are independent, $P(A|B) = P(A \cap B)/P(B) = P(A)P(B)/P(B) = P(A)$. If $P(A|B) = P(A)$, $P(A \cap B)/P(B) = P(A) \implies P(A \cap B) = P(A)P(B)$, so A and B are independent.
2. Let $Q(A) := P(A|B)$. Verify that Q satisfies Definition 2.3.

$$Q(\Omega) = P(\Omega|B) = P(\Omega \cap B)/P(B) = P(B)/P(B) = 1.$$

If $(A_n)_{n \geq 1}$ is a countable sequence of pairwise disjoint elements of \mathcal{A} then $(A_n \cap B)_{n \geq 1}$ is also a sequence of pairwise disjoint elements of \mathcal{A} (If $i \neq j$, $(A_i \cap B) \cap (A_j \cap B) = A_i \cap A_j \cap B = \emptyset \cap B = \emptyset$).

$$\begin{aligned} Q\left(\left(\bigcup_{k=1}^{\infty} A_k\right) \cap B\right) &= \frac{P((\bigcup_{k=1}^{\infty} A_k) \cap B)}{P(B)} \\ &= \frac{P(\bigcup_{k=1}^{\infty} (A_k \cap B))}{P(B)} \\ &= \frac{\sum_{k=1}^{\infty} P(A_k \cap B)}{P(B)} \\ &= \sum_{k=1}^{\infty} \frac{P(A_k \cap B)}{P(B)} \\ &= \sum_{k=1}^{\infty} P(A_k|B) \\ &= \sum_{k=1}^{\infty} Q(A_k). \end{aligned}$$

Theorem 3.3

If $A_1, \dots, A_n \in \mathcal{A}$ with $P(A_1 \cap \dots \cap A_{n-1}) > 0$,

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|(A_1 \cap A_2)) \dots P(A_n|(A_1 \cap \dots \cap A_{n-1})).$$

Proof (Induction): For $n = 2$, the equality holds by Definition 3.2. Suppose the theorem holds for n events, $n \geq 2$. Let $B = A_1 \cap \dots \cap A_n$.

$$\begin{aligned}
& P(A_1 \cap \dots \cap A_n \cap A_{n+1}) \\
&= P(A_{n+1} \cap B) \\
&= P(A_{n+1}|B)P(B) \\
&= P(A_{n+1}|B)P(A_1)P(A_2|A_1)P(A_3|(A_1 \cap A_2)) \dots P(A_n|(A_1 \cap \dots \cap A_{n-1})) \\
&= P(A_1)P(A_2|A_1)P(A_3|(A_1 \cap A_2)) \dots P(A_n|(A_1 \cap \dots \cap A_{n-1}))P(A_{n+1}|(A_1 \cap \dots \cap A_n))
\end{aligned}$$

Theorem 3.4 (Partition Equation)

A collection of events (E_n) , $E_n \in \mathcal{A}$, is called a partition of Ω if they are pairwise disjoint, $P(E_n) > 0$ for each n , and $\cup_n E_n = \Omega$. Let $(E_n)_{n \geq 1}$ be a finite or countable partition of Ω . If $A \in \mathcal{A}$,

$$P(A) = \sum_n P(A|E_n)P(E_n).$$

Proof: Since the E_n are pairwise disjoint, the $A \cap E_n$ are also pairwise disjoint.

$$P(A) = P(A \cap \Omega) = P(A \cap (\cup_n E_n)) = P(\cup_n (A \cap E_n)) = \sum_n P(A \cap E_n) = \sum_n P(A|E_n)P(E_n).$$

Theorem 3.5 (Baye's Theorem)

Let $(E_n)_{n \geq 1}$ be a finite or countable partition of Ω and $P(A) > 0$.

$$P(E_n|A) = \frac{P(A|E_n)P(E_n)}{\sum_m P(A|E_m)P(E_m)}.$$

Proof: By Theorem 3.4,

$$\frac{P(A|E_n)P(E_n)}{\sum_m P(A|E_m)P(E_m)} = \frac{P(A|E_n)P(E_n)}{P(A)} = \frac{P(A \cap E_n)}{P(A)} = P(E_n|A).$$

Exercises

Exercise 3.1

Show that if $A \cap B = \emptyset$, then A and B cannot be independent unless $P(A) = 0$ or $P(B) = 0$.

Answer: Unless one or both of $P(A), P(B)$ is zero, $P(A)P(B) \neq 0 = P(A \cap B)/P(B) = P(A|B)$, meaning that A and B are not independent.

Exercise 3.2

Let $P(C) > 0$. Show that $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$.

Answer:

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

Exercise 3.6

Donated blood is screened for AIDS. Suppose the test has 99% accuracy, and that one in ten thousand people in your age group are HIV positive. The test has a 5% false positive rating, as well. Suppose the test screens you as positive. What is the probability you have AIDS?

Answer: Let A be the event that you have AIDS and B the event that you test HIV Positive. The events A, A^c are a finite partition of the probability space. By Baye's Theorem, the probability that you have AIDS given that you have tested positive is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} = \frac{(.99)(.0001)}{(.99)(.0001) + (.05)(.9999)} \approx 0.001976.$$

Exercise 3.7

Let $(A_n)_{n \geq 1}, (B_n)_{n \geq 1}$ with $A_n, B_n \in \mathcal{A}$ for each n , $A_n \rightarrow A$, $B_n \rightarrow B$, $P(B) > 0$, and $P(B_n) > 0$ for each n .

1. $\lim_{n \rightarrow \infty} P(A_n|B) = P(A|B)$,
2. $\lim_{n \rightarrow \infty} P(A|B_n) = P(A|B)$,
3. $\lim_{n \rightarrow \infty} P(A_n|B_n) = P(A|B)$.

Answer:

1. Since $A_n \rightarrow A$, $A_n \cap B \rightarrow A \cap B$. By Theorem 2.4, $\lim_{n \rightarrow \infty} P(A_n \cap B) = P(A \cap B)$.

$$\lim_{n \rightarrow \infty} P(A_n|B) = \lim_{n \rightarrow \infty} \frac{P(A_n \cap B)}{P(B)} = \frac{P(A \cap B)}{P(B)}.$$

2. Since $B_n \rightarrow B$, $A \cap B_n \rightarrow A \cap B$. By Theorem 2.4, $\lim_{n \rightarrow \infty} P(A \cap B_n) = P(A \cap B)$.

$$\lim_{n \rightarrow \infty} P(A|B_n) = \lim_{n \rightarrow \infty} \frac{P(A \cap B_n)}{P(B_n)} = \frac{P(A \cap B)}{P(B)}.$$

3. Since $A_n \rightarrow A$, $B_n \rightarrow B$, $A_n \cap B_n \rightarrow A \cap B$. By Theorem 2.4, $\lim_{n \rightarrow \infty} P(A_n \cap B_n) = P(A \cap B)$.

$$\lim_{n \rightarrow \infty} P(A_n|B_n) = \lim_{n \rightarrow \infty} \frac{P(A_n \cap B_n)}{P(B_n)} = \frac{P(A \cap B)}{P(B)}.$$

Exercise 3.11

(Polya's Urn) An urn contains r red balls and b blue balls. A ball is chosen at random from the urn, its color is noted, and it is returned together with d more balls of the same color. This is repeated indefinitely. What is the probability that

1. The second ball drawn is blue?
2. The first ball drawn is blue given that the second ball drawn is blue?

Answer: Let B_n be the event that the n th ball drawn is blue and R_n the event that the n th ball drawn is red.

$$1. P(B_2) = P(B_2|B_1)P(B_1) + P(B_2|R_1)P(R_1) = \frac{b+d}{b+r+d} \frac{b}{b+r} + \frac{b}{b+r+d} \frac{r}{b+r} = \frac{b}{b+r} \left(\frac{b+d+r}{b+r+d} \right) = \frac{b}{b+r}.$$

$$2. P(B_1|B_2) = \frac{P(B_1 \cap B_2)}{P(B_2)} = \frac{P(B_2|B_1)P(B_1)}{P(B_2)} = \frac{b+d}{b+r+d} \frac{b}{b+r} \frac{b+r}{b} = \frac{b+d}{b+r+d}$$

Exercise 3.12

Consider the framework of Exercise 3.11. Let B_n denote the event that the n th ball drawn is blue. Show that $P(B_n) = P(B_1)$ for all $n \geq 1$.

Answer: Prove $P(B_n) = P(B_1)$ for all $n \geq 1$ by induction. Exercise 3.11 showed $P(B_2) = P(B_1) = b/(b+r)$. Assume that $P(B_n) = P(B_1)$ for some $n \geq 1$. Let b_n, r_n stand respectively for the number of blue and red balls in the urn during the n th draw.

$$\begin{aligned} P(B_{n+1}) &= P(B_{n+1}|B_n)P(B_n) + P(B_{n+1}|R_n)P(R_n) \\ &= \frac{b_n + d}{b_n + r_n + d} \frac{b_n}{b_n + r_n} + \frac{b_n}{b_n + r_n + d} \frac{r_n}{b_n + r_n} \\ &= \frac{b_n}{b_n + r_n} \left(\frac{b_n + d}{b_n + r_n + d} + \frac{r_n}{b_n + r_n + d} \right) \\ &= \frac{b_n}{b_n + r_n} \\ &= P(B_n) \\ &= P(B_1). \end{aligned}$$

Exercise 3.13

Consider the framework of Exercise 3.11. Find the probability that the first ball is blue given that the n subsequent drawn balls are all blue. Find the limit of this probability as $n \rightarrow \infty$.

Answer:

$$\begin{aligned}
P(B_1|B_2 \cap \dots \cap B_{n+1}) &= \frac{P(B_2 \cap \dots \cap B_{n+1}|B_1)P(B_1)}{P(B_2 \cap \dots \cap B_{n+1})} \\
&= \frac{P(B_2 \cap \dots \cap B_{n+1}|B_1)P(B_1)}{P(B_2 \cap \dots \cap B_{n+1}|B_1)P(B_1) + P(B_2 \cap \dots \cap B_{n+1}|R_1)P(R_1)} \\
&= \frac{\frac{b+d}{b+r+d} \frac{b+2d}{b+r+2d} \dots \frac{b+nd}{b+r+nd} \frac{b}{b+r}}{\frac{b+d}{b+r+d} \frac{b+2d}{b+r+2d} \dots \frac{b+nd}{b+r+nd} \frac{b}{b+r} + \frac{b}{b+r+d} \frac{b+d}{b+r+2d} \dots \frac{b+(n-1)d}{b+r+nd} \frac{r}{b+r}} \\
&= \frac{(b+d)(b+2d) \dots (b+nd)b}{[(b+d)(b+2d) \dots (b+nd)b] + [b(b+d)(b+2d) \dots (b+(n-1)d)r]} \\
&= \frac{b+nd}{b+nd+r} \\
&= \frac{b+nd}{b+r+nd}.
\end{aligned}$$

$$\lim_{n \rightarrow \infty} P(B_1|B_2 \cap \dots \cap B_{n+1}) = \lim_{n \rightarrow \infty} \frac{b+nd}{b+r+nd} = \lim_{n \rightarrow \infty} \frac{b/n+d}{b/n+r/n+d} = 1.$$

Chapter 4 Probabilities on a Finite or Countable Space

For Chapter 4, we assume Ω is finite or countable, and we take the σ -algebra $\mathcal{A} = 2^\Omega$ (the class of all subsets of Ω).

Notes

Theorem 4.1

1. A probability on the finite or countable set Ω is characterized by its values on the atoms: $p_\omega = P(\{\omega\}), \omega \in \Omega$.
2. Let $(p_\omega)_{\omega \in \Omega}$ be a family of real numbers indexed by the finite or countable set Ω . Then there exists a unique probability P such that $P(\{\omega\}) = p_\omega$ if and only if $p_\omega \geq 0$ and $\sum_{\omega \in \Omega} p_\omega = 1$.

Proof:

1. Let $A \in \mathcal{A}$. Since $A \subseteq \Omega$, A is finite or countable and can be written $A = \cup_{\omega \in A} \{\omega\}$, a finite or countable union of pairwise disjoint sets. If P is a probability,

$$P(A) = P(\cup_{\omega \in A} \{\omega\}) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} p_\omega.$$

That is, the value of P evaluated at any set A in 2^Ω is determined by the values it takes on for each singleton $\{\omega\}$. This is what it means for P to be characterized by its values on the atoms.

2. Suppose there is a unique probability P such that $P(\{\omega\}) = p_\omega$. By definition of a probability measure, $p_\omega \geq 0$ and $1 = P(\Omega) = P(\cup_{\omega \in \Omega} \{\omega\}) = \sum_{\omega \in \Omega} P(\{\omega\}) = \sum_{\omega \in \Omega} p_\omega$.

Suppose $p_\omega \geq 0$ and $\sum_{\omega \in \Omega} p_\omega = 1$. Define a probability P by $P(A) \equiv \sum_{\omega \in A} p_\omega$ with the convention that if summing over no elements (an 'empty' sum), the sum is equal to 0. Then $P(\emptyset) = 0$ and $P(\Omega) = \sum_{\omega \in \Omega} p_\omega = 1$.

If $(A_k)_{k=1}^n$ is collection of finitely many events, pairwise disjoint,

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{\omega \in \bigcup_{k=1}^n A_k} p_\omega = \sum_{k=1}^n \sum_{\omega \in A_k} p_\omega = \sum_{k=1}^n P(A_k).$$

If $(A_i)_{i \in I}$ is collection of countably infinitely many events, pairwise disjoint,

$$P\left(\bigcup_{i \in I} A_i\right) = \sum_{\omega \in \bigcup_{i \in I} A_i} p_\omega = \sum_{i \in I} \sum_{\omega \in A_i} p_\omega = \sum_{i \in I} P(A_i).$$

Definition 4.1

A probability P on the finite set Ω is called uniform if $p_\omega = P(\{\omega\})$ does not depend on ω .

Note If P is uniform, $p_\omega = c$ for some constant c for any $\omega \in \Omega$.

$$P(A) = \sum_{\omega \in A} p_\omega = c \sum_{\omega \in A} 1 = c\#(A), \quad 1 = P(\Omega) = \sum_{\omega \in \Omega} p_\omega = c \sum_{\omega \in \Omega} 1 = c\#(\Omega) \implies P(A) = \frac{\#(A)}{\#(\Omega)}.$$

The Hypergeometric Distribution

An urn contains N white balls and M black balls. One draws n balls without replacement, so $n \leq N + M$. One gets X white balls and $n - X$ black balls, $X \in \{0, 1, \dots, n\}$. Consider an outcome to be a subset (containing n elements) of the set $\{1, 2, \dots, N + M\}$ of all $N + M$ balls (which can be assumed to be numbered from 1 to $N + M$). That is, Ω is the family of all subsets of $\{1, \dots, N + M\}$ with n points and $\#(\Omega) = \binom{N+M}{n}$.

Consider the case where P is the uniform probability on Ω . When an event ω is known, the number $X(\omega)$ of white balls drawn is also known. The set $\{X = x\} := X^{-1}(\{x\})$ contains $\binom{N}{x} \binom{M}{n-x}$ points for $x \leq N$ and $n - x \leq M$ (and is otherwise empty). By the note following Definition 4.1,

$$P(X = x) = \begin{cases} \frac{\binom{N}{x} \binom{M}{n-x}}{\binom{N+M}{n}} & \text{if } 0 \leq x \leq N, 0 \leq n-x \leq M \\ 0 & \text{otherwise} \end{cases}$$

As x varies, we obtain the distribution of the random variable X , called the hypergeometric distribution.

The Binomial Distribution

From the same urn as above, we draw n balls, but each time a ball is drawn we put it back, hence n can be as big as one wishes. We again want the probability $P(X = x)$, where x is an integer between 0 and n . Let the probability space be the Cartesian product $\Omega = \{1, 2, \dots, N + M\}^n$ with again the uniform probability. The cardinality of Ω is $\#(\Omega) = (N + M)^n$. Consider the set $X^{-1}(\{x\}) = \{X = x\}$. Choose x of the n draws to be white. For each of the x draws that retrieve a white ball, there are N choices. For each of the $n - x$ draws that retrieve a black ball, there are M choices. The cardinality of $\{X = x\}$ is then $\#(X = x) = \binom{n}{x} N^x M^{n-x}$ and the probability that $X = x$ is

$$\begin{aligned} P(X = x) &= \frac{\#(X = x)}{\#(\Omega)} \\ &= \frac{\binom{n}{x} N^x M^{n-x}}{(N + M)^n} \\ &= \binom{n}{x} \left(\frac{N}{N + M} \right)^x \left(\frac{M}{N + M} \right)^{n-x}, \quad x \in \{0, 1, \dots, n\} \\ &= \binom{n}{x} p^x (1 - p)^{n-x}, \quad p = \frac{N}{N + M}, \quad x \in \{0, 1, \dots, n\} \end{aligned}$$

This formula gives the Binomial distribution with size n and parameter $p \in [0, 1]$, denoted $B(n, p)$.

The Binomial Distribution as a Limit of Hypergeometric Distributions

If n is held fixed for the Hypergeometric distribution and N, M increase to $+\infty$ such that $\frac{N}{N+M} \rightarrow p \in [0, 1]$,

$$\begin{aligned} \frac{\binom{N}{x} \binom{M}{n-x}}{\binom{N+M}{n}} &= \frac{\frac{N!}{x!(N-x)!} \frac{M!}{(n-x)!(M-n+x)!}}{\frac{(N+M)!}{n!(N+M-n)!}} \\ &= \frac{N! M! n! (N + M - n)!}{x! (N - x)! (n - x)! (M - n + x)! (N + M)!} \\ &= \frac{n!}{x! (n - x)!} \frac{N!}{(N - x)!} \frac{M!}{(M - n + x)!} \frac{(N + M - n)!}{(N + M)!} \end{aligned}$$

$$\begin{aligned}
&= \binom{n}{x} \frac{[N(N-1)\dots(N-x+1)][M(M-1)\dots(M-(n-x)+1)]}{(N+M)(N+M-1)\dots(N+M-n+1)} \\
&= \binom{n}{x} \frac{[N(N-1)\dots(N-x+1)][M(M-1)\dots(M-(n-x)+1)]}{(N+M)(N+M-1)\dots(N+M-x-(n-x)+1)} \\
&= \binom{n}{x} \frac{N(N-1)\dots(N-x+1)}{(N+M)(N+M-1)\dots(N+M-x+1)} \\
&\quad \cdot \frac{M(M-1)\dots(M-(n-x)+1)}{(N+M-x)(N+M-x-1)\dots(N+M-x-(n-x)+1)} \\
&\approx \binom{n}{x} \left(\frac{N}{N+M}\right)^x \left(\frac{M}{N+M}\right)^{n-x} \\
&\rightarrow \binom{n}{x} p^x (1-p)^{n-x} \quad \text{as} \quad \frac{N}{N+M} \rightarrow p.
\end{aligned}$$

Exercises

Exercise 4.1 (Poisson Approximation to the Binomial)

Let P be a binomial probability with probability of success p and number of trials n . Let $\lambda = pn$. . Show that

$$P(k \text{ successes}) = \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left\{ \frac{n}{n} \frac{n-1}{n} \dots \frac{n-k+1}{n} \right\} \left(1 - \frac{\lambda}{n}\right)^{-k}.$$

Let $n \rightarrow \infty$ and let p change so that λ remains constant. Conclude that for small p and large n ,

$$P(k \text{ successes}) \approx \frac{\lambda^k}{k!} e^{-\lambda}, \quad \text{where } \lambda = pn.$$

Answer:

$$\begin{aligned}
P(X = k) &= \binom{n}{k} p^k (1-p)^{n-k} \\
&= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \frac{n(n-1)\dots(n-k+1)}{n^k} \\
&= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \frac{n}{n} \frac{n-1}{n} \dots \frac{n-k+1}{n} \\
&= \frac{\lambda^k}{k!} \left(1 + \left(-\frac{\lambda}{n}\right)\right)^n \left\{ \frac{n}{n} \frac{n-1}{n} \dots \frac{n-k+1}{n} \right\} \left(1 - \frac{\lambda}{n}\right)^{-k} \\
&\approx \frac{\lambda^k}{k!} e^{-n\lambda/n} \left\{ \frac{n}{n} \frac{n-1}{n} \dots \frac{n-k+1}{n} \right\} \left(1 - \frac{\lambda}{n}\right)^{-k} \quad ((1+a)^n \approx e^{na} \text{ for small } a)^* \\
&= \frac{\lambda^k}{k!} e^{-\lambda} \left\{ \frac{n}{n} \frac{n-1}{n} \dots \frac{n-k+1}{n} \right\} \left(1 - \frac{\lambda}{n}\right)^{-k} \\
&\rightarrow \frac{\lambda^k}{k!} e^{-\lambda} \cdot 1 \cdot 1^{-k} = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{as } n \rightarrow \infty.
\end{aligned}$$

*The approximation $(1+a) \approx e^{na}$ for 'small a ' warrants discussion.

$$\begin{aligned}
\frac{1}{1+a} &= 1 - a + a^2 - a^3 + \dots, \quad |a| < 1 \\
\ln(1+a) &= a - \frac{a^2}{2} + \frac{a^3}{3} - \dots \\
1+a &= e^a e^{-a^2/2} e^{a^3/3} \dots \\
(1+a)^n &= e^{na} e^{-na^2/2} e^{na^3/3} \dots
\end{aligned}$$

If $na \ll 1$, $(1+a)^n \approx e^0 = 1$. In our calculation, $n(-\lambda/n) = \lambda$ does not satisfy this criterion. If $na^2 \ll 1$, $(1+a)^n \approx e^{na}$. In our calculation $n(-\lambda/n)^2 = \lambda/n \ll 1$ for large n assuming that we let p change so that λ remains constant as $n \rightarrow \infty$.

Exercise 4.2 (Poisson Approximation to the Binomial continued)

In the setting of Exercise 4.1, let $p_k = P(\{k\})$ and $q_k = 1 - p_k$. Show that the q_k are the probabilities of singletons for a Binomial distribution $B(1-p, n)$. Deduce a Poisson approximation of the Binomial when n is large and p is close to 1.

Answer: For $X \sim B(n, p)$ and $Y \sim B(n, 1-p)$,

$$\begin{aligned}
P(\{k\}) &= P(X = k) \\
&= \binom{n}{k} p^k (1-p)^{n-k} \\
&= \binom{n}{n-k} p^k (1-p)^{n-k} \\
&= \binom{n}{j} p^{n-j} (1-p)^j, \quad j = n-k \\
&= \binom{n}{j} (1-p)^j p^{n-j} \\
&= \binom{n}{j} q^j (1-q)^{n-j} \\
&= P(Y = j) = P(\{j\})
\end{aligned}$$

$$\lambda = pn \implies n - \lambda = n(1-p) = nq.$$

$$P(Y = j) \approx \frac{(n-\lambda)^j}{j!} e^{\lambda-n} \quad \text{for large } n$$

Exercise 4.3

We consider the setting of the hypergeometric distribution, except that we have m colors and N_i balls of color i . Set $N = N_1 + \dots + N_m$, and call X_i the number of balls of color i drawn among n balls. Of course $X_1 + \dots + X_m = n$. Show that

$$P(X_1 = x_1, \dots, X_m = x_m) = \begin{cases} \frac{\binom{N_1}{x_1} \dots \binom{N_m}{x_m}}{\binom{N}{n}} & \text{if } x_1 + \dots + x_m = n \\ 0 & \text{otherwise} \end{cases}$$

Answer: Consider an outcome to be a subset (containing n elements) of the set $\{1, 2, \dots, N\}$ of all $N = N_1 + \dots + N_m$ balls (which can be assumed to be numbered from 1 to N). That is, Ω is the family of all subsets of $\{1, \dots, N\}$ with n points and $\#(\Omega) = \binom{N}{n}$.

Consider the case where P is the uniform probability on Ω . The set $\{X_1 = x_1, \dots, X_m = x_m\} = X^{-1}(\{(x_1, \dots, x_m)\})$ contains $\binom{N_1}{x_1} \binom{N_2}{x_2} \dots \binom{N_m}{x_m}$ points for $x_1 + x_2 + \dots + x_m = n$. If $x_1 + \dots + x_m \neq n$, $\{X_1 = x_1, \dots, X_m = x_m\} = \emptyset$ so $\#(X_1 = x_1, \dots, X_m = x_m) = 0$. Therefore,

$$P(X_1 = x_1, \dots, X_m = x_m) = \frac{\#(X_1 = x_1, \dots, X_m = x_m)}{\#(\Omega)} = \begin{cases} \frac{\binom{N_1}{x_1} \binom{N_2}{x_2} \dots \binom{N_m}{x_m}}{\binom{N}{n}} & \text{if } x_1 + \dots + x_m = n \\ 0 & \text{otherwise} \end{cases}$$

Chapter 5 Random Variables on a Countable Space

Notes

In Chapter 5 we again assume Ω is countable and $\mathcal{A} = 2^\Omega$. A random variable X in this case is defined to be a function from $X : \Omega \rightarrow T$. Even if the set T (the state space or range space) of X is uncountable, the image T' of Ω under X (T' is the set of all points $\{i\}$ in T for which there exists and $\omega \in \Omega$ such that $X(\omega) = i$) must be finite or countably infinite since Ω is countable. Define the distribution of X on the range space T' by

$$P^X(A) = P(\{\omega : X(\omega) \in A\}) = P(X^{-1}(A)) = P(X \in A).$$

This formula defines a probability measure on T' with the σ -algebra $2^{T'}$.

$$1. P^X(T') = P(\Omega) = 1.$$

2.

$$\begin{aligned} P^X\left(\bigcup_{n=1}^{\infty} A_n\right) &= P(\{\omega : X(\omega) \in \bigcup_{n=1}^{\infty} A_n\}) \\ &= P\left(\bigcup_{n=1}^{\infty} \{\omega : X(\omega) \in A_n\}\right) \\ &= \sum_{n=1}^{\infty} P(\{\omega : X(\omega) \in A_n\}) \\ &= \sum_{n=1}^{\infty} P^X(A_n). \end{aligned}$$

The probability is completely determined by

$$p_j^X = P(X = j) = \sum_{\{\omega : X(\omega) = j\}} p_\omega.$$

We have

$$\begin{aligned} P^X(A) &= P(\{\omega : X(\omega) \in A\}) \\ &= P(\bigcup_{j \in A} \{\omega : X(\omega) \in \{j\}\}) \\ &= \sum_{j \in A} P(\{\omega : X(\omega) \in \{j\}\}) \\ &= \sum_{j \in A} P(X = j) \\ &= \sum_{j \in A} p_j^X. \end{aligned}$$

Definition 5.1

Let X be a real-valued random variable on a countable space Ω . The expectation of X , denoted $E\{X\}$, is defined to be

$$E\{X\} = \sum_{\omega} X(\omega)p_\omega,$$

provided this sum makes sense: this is the case when Ω is finite; this is also the case when Ω is countable and the series is absolutely convergent or Ω is countable and $X \geq 0$ always (in this last case the sum and hence the $E\{X\}$ may take the value $+\infty$).

\mathcal{L}^1 : Define the space \mathcal{L}^1 to be the space of real valued random variables on (Ω, \mathcal{A}, P) which have finite expectation.

- (i) \mathcal{L}^1 is a vector space and the expectation operator E is linear.
- (ii) The expectation operator is positive: If $X \in \mathcal{L}^1$ and $X \geq 0$ then $E\{X\} \geq 0$. More generally, if $X, Y \in \mathcal{L}^1$ and $X \leq Y$ then $E\{X\} \leq E\{Y\}$.
- (iii) \mathcal{L}^1 contains all bounded random variables. If Y is a bounded random variable there is a random variable $X \equiv a$ such that $-X \leq Y \leq X$ so $-a = E\{-X\} \leq E\{Y\} \leq E\{X\} = a$.
- (iv) If $X \in \mathcal{L}^1$ and T' is the range of X , $E\{X\} = \sum_{j \in T'} jP(X = j)$.
- (v) If $X = 1_A$ is the indicator function of an event A , then $E\{X\} = \sum_{\omega \in \Omega} X(\omega)p_\omega = \sum_{\omega \in A} p_\omega = P(A)$.

If $\sum_{\omega} (X(\omega))^2 p_\omega$ is absolutely convergent,

$$\sum_{\omega} |X(\omega)| p_\omega \leq \sum_{|X(\omega)| < 1} X(\omega) p_\omega + \sum_{|X(\omega)| \geq 1} X(\omega) p_\omega \leq \sum_{|X(\omega)| < 1} p_\omega + \sum_{|X(\omega)| \geq 1} (X(\omega))^2 p_\omega < \infty \implies X \in \mathcal{L}^1.$$

Theorem 5.1

Let $h : \mathbb{R} \rightarrow [0, \infty)$ be a nonnegative function and let X be a real valued random variable. Then

$$P(\{\omega : h(X(\omega)) \geq a\}) \leq \frac{E\{h(X)\}}{a}, \quad \forall a > 0.$$

Proof: Since X is a random variable, so is $Y = h(X)$. Let

$$A = Y^{-1}([a, \infty)) = \{\omega : h(X(\omega)) \geq a\} = \{h(x) \geq a\}.$$

Then $h(X) \geq a1_A$ since $h(X(\omega)) \geq a = a1_A(\omega)$ if $\omega \in A$ and $h(X(\omega)) \geq 0 = a1_A(\omega)$ if $\omega \notin A$.

$$E\{h(X)\} \geq E\{a1_A\} = aP(A) \implies \frac{E\{h(X)\}}{a} \geq P(A) = P(\{\omega : h(X(\omega)) \geq a\})$$

Corollary 5.1 (Markov's Inequality)

$$P(\{|X| \geq a\}) \leq \frac{E\{|X|\}}{a}.$$

Proof: Let $h : \mathbb{R} \rightarrow [0, \infty)$ be defined by $h(x) = |x|$. By Theorem 5.1,

$$P(\{|X| \geq a\}) = P(\{\omega : |X(\omega)| \geq a\}) = P(\{\omega : h(X(\omega)) \geq a\}) \leq \frac{E\{h(X)\}}{a} = \frac{E\{|X|\}}{a}$$

Definition 5.2

Let X be a random variable with $X^2 \in \mathcal{L}^1$. The Variance of X is defined to be

$$\sigma^2 = \sigma_X^2 \equiv E\{(X - E(X))^2\}.$$

The standard deviation of X , σ_X is the nonnegative square root of the variance.

Corollary 5.2 (Chebyshev's Inequality)

If $X^2 \in \mathcal{L}^1$,

$$(a) \ P(\{|X| \geq a\}) \leq \frac{E\{X^2\}}{a^2}, \quad a > 0.$$

$$(b) \ P(\{|X - E\{X\}| \geq a\}) \leq \frac{\sigma_X^2}{a^2}, \quad a > 0.$$

Proof: To prove (a), let $h(x) = x^2$. By Theorem 5.1,

$$P(\{|X| \geq a\}) = P(\{X^2 \geq a^2\}) = P(\{h(X) \geq a^2\}) \leq \frac{E\{h(X)\}}{a^2} = \frac{E\{X^2\}}{a^2}.$$

To prove (b), let $Y = |X - E\{X\}|$. Again by Theorem 5.1,

$$P(\{|X - E\{X\}| \geq a\}) = P(\{Y \geq a\}) = P(\{Y^2 \geq a^2\}) \leq \frac{E\{Y^2\}}{a^2} = \frac{E\{(X - E\{X\})^2\}}{a^2} = \frac{\sigma_X^2}{a^2}.$$

Exercises**Exercise 5.1**

Let $g : [0, \infty) \rightarrow [0, \infty)$ be strictly increasing and nonnegative. Show that

$$P(\{|X| \geq a\}) \leq \frac{E\{g(|X|)\}}{g(a)}, \quad a > 0.$$

Answer: For $a > 0$, $\{|X| \geq a\} = \{\omega : |X(\omega)| \geq a\} = \{\omega : g(|X(\omega)|) \geq g(a)\} = \{g(|X|) \geq g(a)\}$ since g is strictly increasing and nonnegative. Also $|X|$ is random variable since X is a random variable. Since g is strictly increasing and has domain $[0, \infty)$, the only possible x such that $g(x) = 0$ is $x = 0$. We must have $g(a) > 0$ because $a > 0$. By Theorem 5.1,

$$P(\{|X| \geq a\}) = P(\{g(|X|) \geq g(a)\}) \leq \frac{E\{g(|X|)\}}{g(a)}.$$

Exercise 5.6

Let X be Binomial $B(p, n)$. For what value of j is $P(X = j)$ the greatest?

Answer:

For $k \geq 1$, $P(X = k - 1) > 0$ and

$$\frac{P(X = k)}{P(X = k - 1)} = \frac{\binom{n}{k} p^k (1 - p)^{n - k}}{\binom{n}{k - 1} p^{k - 1} (1 - p)^{n - k + 1}} = \frac{n!}{k!(n - k)!} \frac{(k - 1)!(n - k + 1)!}{n!} \frac{p}{1 - p} = \frac{n - k + 1}{k} \frac{p}{1 - p}.$$

We have $P(X = k) \geq P(X = k - 1)$ for $P(X = k)/P(X = k - 1) \geq 1$.

$$\frac{n - k + 1}{k} \frac{p}{1 - p} \geq 1 \iff (n + 1)p \geq k.$$

Let $j = [(n + 1)p]$, where $[x]$ is the floor function. For any $k < j$, $k \leq j - 1$ so $P(X = j) \geq P(X = k)$. For any $k > j$, $P(X = j) < P(X = k)$. Conclude that $P(X = j)$ is greatest for $j = [(n + 1)p]$.

Exercise 5.7

Let X be Binomial $B(p, n)$. Find the probability X is even.

Answer: If $p = 0$ then X , there will always be 0 successes so the probability that X is even is 1. If $p = 1$, then $X = n$ and so the probability that X is even is 1 if n is even and 0 if n is odd.

Let P_n be the probability that $X \sim B(p, n), p \in (0, 1)$ is an even number. To calculate the probability P_{n+1} consider for $X \sim B(p, n+1)$ whether there are an even number of successes or an odd number of successes from n trials. Then $P_{n+1} = (1-p)P_n + p(1-P_n) = p + (1-2p)P_n$. Motivated by the idea that P_n approaches $1/2$ as n approaches $+\infty$, let $x_n = P_n - \frac{1}{2}$.

$$x_{n+1} = P_{n+1} - \frac{1}{2} = p + (1-2p)(x_n + \frac{1}{2}) + \frac{1}{2} = (1-2p)x_n.$$

Since the probability that X is even for $X \sim B(0, p)$ is 1, $x_0 = P_0 - \frac{1}{2} = \frac{1}{2}$. Thus (x_n) is a geometric sequence with common ratio $(1-2p)$ so we write $x_n = \frac{1}{2}(1-2p)^n$. For $0 < p < 1$, $|1-2p| < 1$ so $x_n \rightarrow 0$ as $n \rightarrow \infty$. This means that the probability that $X \sim B(n, p)$ is even is $P_n = \frac{1}{2} + x_n = \frac{1}{2}(1 + (1-2p)^n)$. Checking P_n for $p = 0$ and $p = 1$ shows that this formula holds for any $p \in [0, 1]$.

Exercise 5.9

Let X be Poisson (λ) . What value of j maximizes $P(X = j)$?

Answer: Consider that $P(X = k) \geq P(X = k-1)$ for $P(X = k)/P(X = k-1) \geq 1$.

$$\frac{P(X = k)}{P(X = k-1)} = \frac{\frac{\lambda^k}{k!} e^{-\lambda}}{\frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda}} = \frac{\lambda}{k}.$$

This shows that $P(X = k) \geq P(X = k-1)$ for $k \leq \lambda$. Let $j = [\lambda]$, where $[x]$ is the floor function. By the reasoning given in Exercise 5.6, j maximizes $P(X = j)$.

Exercise 5.12

Let X be Binomial $B(p, n)$. Show that for $\lambda > 0$ and $\epsilon > 0$,

$$P(X - np > n\epsilon) \leq E\{\exp(\lambda(X - np - n\epsilon))\}.$$

Partial Answer:

$$P(X - np > n\epsilon) \leq P(|X - np| \geq n\epsilon) \leq \frac{E\{\exp(\lambda|X - np|)\}}{\exp(\lambda n\epsilon)} = E\{\exp(\lambda(|X - np| - n\epsilon))\}$$

Exercise 5.13

Let X_n be Binomial $B(p, n)$ with $p > 0$ fixed. Show that for any fixed $b > 0$, $P(X_n \leq b)$ tends to 0.

Exercise 5.16

Let X be Geometric. Show that for $i, j > 0$,

$$P(X > i + j | X > i) = P(X > j).$$

Answer:

$$\begin{aligned}
P(X > i + j | X > i) &= \frac{P(\{X > i + j\} \cap \{X > i\})}{P(X > i)} \\
&= \frac{P(X > i + j)}{P(X > i)} \\
&= \frac{P(X = i + j + 1) + P(X = i + j + 2) + \dots}{P(X = i + 1) + \dots + P(X = i + j + 1) + \dots} \\
&= \frac{p(1-p)^{i+j+1} + p(1-p)^{i+j+2} + \dots}{p(1-p)^{i+1} + p(1-p)^{i+2} + \dots + p(1-p)^{i+j+1} + \dots} \\
&= \frac{(1-p)^j + (1-p)^{j+1} + \dots}{1 + (1-p) + \dots + (1-p)^j + \dots} \\
&= \frac{(1-p)^j(1 + (1-p) + \dots)}{1 + (1-p) + \dots} \\
&= (1-p)^j \\
&= p(1-p)^j \frac{1}{1 - (1-p)} \\
&= p(1-p)^j(1 + (1-p) + (1-p)^2 + \dots) \\
&= p(1-p)^{j+1} + p(1-p)^{j+2} + \dots \\
&= P(X = j + 1) + P(X = j + 2) + \dots \\
&= P(X > j).
\end{aligned}$$

Exercise 5.17

Let X be Geometric (p). Show

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

Partial Answer:

$$\begin{aligned}
E \left\{ \frac{1}{1+X} \right\} &= \sum_{k=0}^{\infty} \frac{1}{1+k} P(X = k) \\
&= \sum_{k=0}^{\infty} \frac{1}{1+k} p(1-p)^k \\
&= \sum_{j=1}^{\infty} \frac{1}{j} p(1-p)^{j-1} \\
&= \frac{p}{1-p} \sum_{j=1}^{\infty} \frac{(1-p)^j}{j} \\
&= -\frac{p}{1-p} \log(1 - (1-p)) \\
&= \frac{p}{p-1} \log(p) \\
&= \log \left(p^{\frac{p}{p-1}} \right)
\end{aligned}$$

Exercise 5.19

Show that for a sequence of events $(A_n)_{n \geq 1}$,

$$E \left\{ \sum_{n=1}^{\infty} 1_{A_n} \right\} = \sum_{n=1}^{\infty} P(A_n).$$

Answer: Using properties (i) and (v) of \mathcal{L}^1 given on page 28,

$$E \left\{ \sum_{n=1}^{\infty} 1_{A_n} \right\} = \sum_{n=1}^{\infty} E\{1_{A_n}\} = \sum_{n=1}^{\infty} P(A_n) .$$

Exercise 5.20

Suppose X takes all its values in $\mathbb{N} = \{0, 1, 2, 3, \dots\}$. Show that

$$E\{X\} = \sum_{n=0}^{\infty} P(X > n).$$

Answer:

$$\begin{aligned} \sum_{n=0}^{\infty} P(X > n) &= \sum_{n=0}^{\infty} \sum_{k=n+1}^{\infty} P(X = k) \\ &= \sum_{n=1}^{\infty} nP(X = n) \\ &= \sum_{n=0}^{\infty} nP(X = n) \\ &= E\{X\}. \end{aligned}$$

Chapter 6 Construction of a Probability Measure

We assume a given Ω and a σ -algebra $\mathcal{A} \subset 2^\Omega$ but no longer assume that Ω is countable. (Ω, \mathcal{A}) is called a *measurable space* and we want to construct probability measures on \mathcal{A} . For Ω uncountable a typical probability P will have $P(\{\omega\}) = 0$. Thus the family of all numbers $P(\{\omega\}), \omega \in \Omega$ does not characterize the probability P in general.

Suppose \mathcal{A}_0 is an algebra and that $\mathcal{A} = \sigma(\mathcal{A}_0)$. Suppose further that we are given a probability P on the algebra \mathcal{A}_0 . That is, a function $P : \mathcal{A}_0 \rightarrow [0, 1]$ satisfying

1. $P(\Omega) = 1$.
2. For any sequence (A_n) of elements of \mathcal{A}_0 , pairwise disjoint, and such that $\cup_n A_n \in \mathcal{A}_0$, we have $P(\cup_n A_n) = \sum_n P(A_n)$.

Theorem 6.1

Each probability P defined on the algebra \mathcal{A}_0 has a unique extension (also called P) on \mathcal{A} .

Proof (Uniqueness): Suppose P, Q are probabilities defined on \mathcal{A}_0 with extensions P, Q on \mathcal{A} that agree on \mathcal{A}_0 . By definition of an algebra \mathcal{A}_0 is closed under finite intersections. Since $\sigma(\mathcal{A}_0) = \mathcal{A}$, we have $P = Q$ by Corollary 6.1 (below).

Definition 6.1

A class \mathcal{C} of subsets of Ω is closed under finite intersections if for any n (arbitrary but finite) $A_1 \cap \dots \cap A_n$ when $A_1, \dots, A_n \in \mathcal{C}$. A class \mathcal{C} is closed under increasing limits if whenever $A_1 \subset A_2 \subset \dots \subset A_n \subset \dots$ is a sequence of events in \mathcal{C} , then $\cup_{n=1}^\infty A_n \in \mathcal{C}$ as well. A class \mathcal{C} is closed under differences if whenever $A, B \in \mathcal{C}$ with $A \subset B$, then $B \setminus A \in \mathcal{C}$.

Theorem 6.2 (Monotone Class Theorem)

Let \mathcal{C} be a class of subsets of Ω , closed under finite intersections and containing Ω . Let \mathcal{B} be the smallest class containing \mathcal{C} which is closed under increasing limits and by difference. Then $\mathcal{B} = \sigma(\mathcal{C})$.

Proof: First note that the intersection of classes of sets closed under increasing limits and differences is again a class of that type. If $A_1 \subset \dots \subset A_n$ is a sequence of events in this intersection, then this sequence is in each class in the intersection and so $\cup_{n=1}^\infty A_n$ is in each class in the intersection. Similar with differences. So, by taking the intersection of all such classes, there always exists a smallest class \mathcal{B} containing \mathcal{C} . For each set B , denote by \mathcal{B}_B the collection of sets A such that $A \in \mathcal{B}$ and $A \cap B \in \mathcal{B}$. Suppose $A_1 \subset \dots \subset A_n$ is a sequence in \mathcal{B}_B . Then each $A_i \in \mathcal{B}$ and each $A_i \cap B \in \mathcal{B}$ (this is also an increasing sequence) so that $\cup_{n=1}^\infty A_n \in \mathcal{B}$ and $B \cap (\cup_n A_n) = \cup_n (B \cap A_n) \in \mathcal{B}$ by the properties of \mathcal{B} . This means \mathcal{B}_B is closed under increasing limits. Similarly if $A_1, A_2 \in \mathcal{B}_B$ with $A_1 \subset A_2$ then $A_1, A_2 \in \mathcal{B}$ and since $A_1 \setminus A_2 \in \mathcal{B}$ so is $(A_1 \setminus A_2) \cap B$. This means \mathcal{B}_B is closed by difference.

Let $B \in \mathcal{C}$. For each $C \in \mathcal{C}$ on has $B \cap C \in \mathcal{C} \subset \mathcal{B}$ and $C \in \mathcal{B}$. Thus $C \in \mathcal{B}_B$ and $\mathcal{C} \subset \mathcal{B}_B \subset \mathcal{B}$. Since \mathcal{B} is the smallest class containing \mathcal{C} closed under increasing limits and differences, $\mathcal{B} \subset \mathcal{B}_B$. Therefore $B = \mathcal{B}_B$ for this particular $B \in \mathcal{C}$.

Now let $B \in \mathcal{B}$. For each $C \in \mathcal{C}$. Since $B \in \mathcal{B}$, $C \in \mathcal{B}$ (as $\mathcal{C} \subset \mathcal{B}$), and $B \cap C \in \mathcal{B}$ we have $B \in \mathcal{B}_C$. Hence $C \in \mathcal{B}_B$ and $C \subset \mathcal{B}_B \subset \mathcal{B}$ so that $\mathcal{B} = \mathcal{B}_B$. Since $B \in \mathcal{B}_B$ was arbitrary, we have established that $B = \mathcal{B}_B$ for any $B \in \mathcal{B}$. This means that \mathcal{B} is closed by finite intersections. Since $\Omega \in \mathcal{C}$, $\Omega \in \mathcal{B}$. Since $\Omega \in \mathcal{B}$ and \mathcal{B} is closed under difference, \mathcal{B} is also closed under complementation. Since \mathcal{B} is closed under increasing limits

(established earlier), conclude that \mathcal{B} is a σ -algebra. Since \mathcal{B} was the smallest class containing \mathcal{C} and we showed that this smallest class is a σ -algebra, $\mathcal{B} = \sigma(\mathcal{C})$.

Corollary 6.1

Let P and Q be two probabilities defined on \mathcal{A} , and suppose P and Q agree on a class $\mathcal{C} \subset \mathcal{A}$ which is closed under finite intersections. If $\sigma(\mathcal{C}) = \mathcal{A}$, we have $P = Q$.

Proof: $\Omega \in \mathcal{A}$ because \mathcal{A} is a σ -algebra and $P(\Omega) = Q(\Omega)$ since they are both probabilities. We assume $\Omega \subset \mathcal{C}$. Let $\mathcal{B} = \{A \in \mathcal{A} : P(A) = Q(A)\}$. By the definition of a probability measure and Theorem 2.3, \mathcal{B} is closed by difference and by increasing limits. Since \mathcal{B} contains all sets on which P and Q agree, $\mathcal{C} \subset \mathcal{B}$. Since $\sigma(\mathcal{C}) = \mathcal{A}$, $\mathcal{B} = \mathcal{A}$ by Theorem 6.2.

Definition 6.2

Let P be a probability on \mathcal{A} . A null set for P is a set $A \subset \Omega$ such that there exists a $B \in \mathcal{A}$ satisfying $A \subset B$ and $P(B) = 0$.

Chapter 7 Construction of a Probability Measure on \mathbb{R}

This chapter assumes $\Omega = \mathbb{R}$ (a special case of what we dealt with in Chapter 6). Let $\mathcal{B} = \sigma(\mathcal{O})$, be the Borel σ -algebra of \mathbb{R} (\mathcal{O} are the open subsets of \mathbb{R}).

Definition 7.1

The distribution function F induced by a probability P on $(\mathbb{R}, \mathcal{B})$ is the function

$$F(x) = P((-\infty, x]).$$

Theorem 7.1

The distribution function F characterizes the probability.

Proof: We want to show that if there is another probability Q such that $G(x) = Q((-\infty, x])$ for $x \in \mathbb{R}$, and if $F = G$, then also $P = Q$.

Let \mathcal{B}_0 be the set of finite disjoint unions of intervals of the form $(x, y]$ with $-\infty \leq x \leq y \leq \infty$ (use the convention that $(x, \infty] = (x, \infty)$). If $x = y$, then $(x, y] = \emptyset$ and if $x = -\infty, y = \infty$ then $(x, y] = \mathbb{R} = \Omega$. If $A \in \mathcal{B}_0$ then A is a disjoint union $\cup_{i=1}^n (x_i, y_i]$ and $A^c = (-\infty, x_1] \cup (y_1, x_2] \cup \dots \cup (y_{n-1}, x_n] \cup (y_n, \infty)$ if $x_1 \neq -\infty$ and $y_n \neq \infty$. If $x_1 = -\infty$ the first interval is the empty set and if $y_n = \infty$ the last interval is the empty set. So A^c is also in \mathcal{B}_0 . A drawing helps to see that \mathcal{B}_0 is also closed under finite intersection and union (but I do not write out the details here). So \mathcal{B}_0 is an algebra. Since $(a, b) = \cup_{n=N}^{\infty} (a, b - 1/n]$ for appropriately large N , $\sigma(\mathcal{B}_0)$ contains all open intervals. All open sets on the real line can be written as a countable union of open intervals. So $\sigma(\mathcal{B}_0)$ contains all the open subsets of \mathbb{R} and $\sigma(\mathcal{B}_0) \supset \mathcal{B}$. Since $(a, b] = \cap_{n=1}^{\infty} (a, b + \frac{1}{n})$, $\mathcal{B}_0 \subset \mathcal{B}$ and so $\sigma(\mathcal{B}_0) \subset \mathcal{B}$. Thus $\sigma(\mathcal{B}_0) = \mathcal{B}$.

The definition of F implies $P((x, y]) = F(y) - F(x)$ and if $A \in \mathcal{B}_0$ with $A = \cup_{i=1}^n (x_i, y_i]$ with $y_i < x_{i+1}$ for $i = 1, \dots, n-1$ then

$$P(A) = \sum_{i=1}^n (F(y_i) - F(x_i)).$$

If Q is another probability measure $(\mathbb{R}, \mathcal{B})$ such that $F(x) = Q((-\infty, x])$ then $P = Q$ on \mathcal{B}_0 . By Theorem 6.1 $P = Q$ on \mathcal{B} as well. That is, P and Q are the same probability measure.

Theorem 7.2

A function F is the distribution function of a (unique) probability on $(\mathbb{R}, \mathcal{B})$ if and only if one has:

1. F is nondecreasing.
2. F is right continuous ($\lim_{y \downarrow x} F(y) = F(x)$ for all $x \in \mathbb{R}$).
3. $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow \infty} F(x) = 1$.

Corollary 7.1

Let F be the distribution function of the probability P on \mathbb{R} . For all $x < y$ we have:

1. $P((x, y]) = F(y) - F(x)$.
2. $P([x, y]) = F(y) - F(x-)$.

3. $P([x, y)) = F(y-) - F(x-)$.
4. $P((x, y)) = F(y-) - F(x)$.
5. $P(\{x\}) = F(x) - F(x-)$.

Exercises

Exercise 7.1

Let $(A_n)_{n \geq 1}$ be a sequence of pairwise disjoint events and P a probability. Show that $\lim_{n \rightarrow \infty} P(A_n) = 0$.

Exercise 7.6

Show that the maximum of the Lognormal density occurs at $x = e^\mu e^{-\sigma^2}$.

Answer: The Lognormal distribution with parameters μ, σ^2 ($-\infty < \mu < \infty$, $0 < \sigma^2 < \infty$) is

$$f(x) = \begin{cases} \frac{1}{x} g_{\mu, \sigma^2}(\log x) & x > 0 \\ 0 & x \leq 0 \end{cases},$$

where $g_{\mu, \sigma^2}(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(z-\mu)^2/(2\sigma^2)}$ is the normal distribution with parameters μ, σ^2 . Fix the parameters μ, σ^2 and let $g(z) = g_{\mu, \sigma^2}(z)$, $\alpha = 1/(\sqrt{2\pi}\sigma)$. Since $g > 0$, $f(x) > 0$ for $x > 0$. So the maximum of f on \mathbb{R} is the maximum of f on $(0, \infty)$.

$$\begin{aligned} 0 &= f'(x) = -\frac{1}{x^2} g(\log x) + \frac{1}{x^2} g'(\log x) \\ g(\log x) &= g'(\log x) \\ \alpha e^{-(\log x - \mu)^2/(2\sigma^2)} &= \alpha e^{-(z - \mu)^2/(2\sigma^2)} \frac{\mu - \log x}{\sigma^2} \\ 1 &= \frac{\mu - \log x}{\sigma^2} \\ \log x &= \mu - \sigma^2 \\ x &= e^\mu e^{-\sigma^2} \end{aligned}$$

Exercise 7.11

Let $P(A) = \int_{-\infty}^{\infty} 1_A(x) f(x) dx$ for a nonnegative function f with $\int_{-\infty}^{\infty} f(x) dx = 1$. Let $A = \{x_0\}$, a singleton (that is, the set A consists of one single point on the real line). Show that A is a Borel set and also a null set (that is, $P(A) = 0$).

Answer: We can write A as

$$\begin{aligned} (x_0 - 1, x_0] &\cap [x_0, x_0 + 1) \\ (x_0 - 1, x_0] &= \cap_{n=1}^{\infty} (x_0 - 1, x_0 + 1/n), \quad [x_0, x_0 + 1) = \cap_{n=1}^{\infty} (x_0 - 1/n, x_0) \end{aligned}$$

This shows that A can be written as the intersection of two Borel sets, which must also be a Borel set and

$$P(A) = \int_{-\infty}^{\infty} 1_{\{x_0\}}(x) f(x) dx = \int_{x_0}^{x_0} f(x) dx = 0.$$

Exercise 7.14

Let $(A_i)_{i \geq 1}$ be a sequence of null sets. Show that $B = \cup_{i=1}^{\infty} A_i$ is also a null set.

Answer: For each A_i , there is a $B_i \in \mathcal{A}$ (the σ -algebra on Ω) such that $A_i \subset B_i$ and $P(B_i) = 0$. Let $B' = \cup_{i=1}^{\infty} B_i$. Then $B \subset B'$ and $0 \leq P(B') \leq \sum_{i=1}^{\infty} P(B_i) = 0$. This shows that $P(B') = 0$ and therefore B is a null set for P .

Exercise 7.15

Let X be a r.v. defined on a countable Probability space. Suppose $E\{|X|\} = 0$. Show that $X = 0$ except possibly on a null set. Is it possible to conclude, in general, that $X = 0$ everywhere (i.e., for all ω)?

Answer: If $0 = E\{|X|\} = \sum_{\omega \in \Omega} |X(\omega)|P(\{\omega\})$ implies that for each ω either $|X(\omega)| = 0$ or $P(\{\omega\}) = 0$. If $|X(\omega)| = 0$ for all ω we are done. Otherwise let B be the union of all singletons $\{\omega\}$ such that $P(\{\omega\}) = 0$. Since Ω is countable, this is either a finite or countably infinite union. By the previous exercise, B is also a null set. It is not possible to conclude that $X = 0$ everywhere.

Exercise 7.16

Let F be a distribution function. Show that in general F can have an infinite number of jump discontinuities, but that there can be at most countably many.

Answer:

Exercise 7.17

Suppose a distribution function F is given by

$$F(x) = \frac{1}{4}1_{[0, \infty)} + \frac{1}{2}1_{[1, \infty)} + \frac{1}{4}1_{[2, \infty)}.$$

Let P be given by $P((-\infty, x]) = F(x)$. Then find the probabilities of the following events:

- a) $A = (-\frac{1}{2}, \frac{1}{2})$
- b) $B = (-\frac{1}{2}, \frac{3}{2})$
- c) $C = (\frac{2}{3}, \frac{5}{2})$
- d) $D = [0, 2)$
- e) $E = (3, \infty)$.

Exercise 7.18

Suppose a function F is given by

$$F(x) = \sum_{i=1}^{\infty} \frac{1}{2^i} 1_{[\frac{1}{i}, \infty)}.$$

Show that F is a distribution on \mathbb{R} . Let us define P by $P((-\infty, x]) = F(x)$. Find the probabilities of the following events:

- a) $A = [1, \infty)$
- b) $B = [\frac{1}{10}, \infty)$

c) $C = \{0\}$

d) $D = [0, \frac{1}{2})$

e) $E = (-\infty, 0)$

f) $G = (0, \infty)$

Chapter 8 Random Variables

Consider an abstract space (Ω, \mathcal{A}, P) (countable or not). Suppose X maps Ω into a state space (F, \mathcal{F}) . We want to calculate the probability that X takes on values in a given subset of the state space, where such a subset is an element of the σ -algebra \mathcal{F} of subsets of F . This can be written $P(\{\omega : X(\omega) \in A\}) = P(X \in A) = P(X^{-1}(A))$. In order to find $P(X^{-1}(A))$ it is necessary that $X^{-1}(A) \in \mathcal{A}$.

Definition 8.1

- a) Let (E, \mathcal{E}) and (F, \mathcal{F}) be two measurable spaces. A function $X : E \rightarrow F$ is called measurable (relative to \mathcal{E} and \mathcal{F}) if $X^{-1}(\Lambda) \in \mathcal{E}$ for all $\Lambda \in \mathcal{F}$ ($X^{-1}(\mathcal{F}) \subset \mathcal{E}$).
- b) When $(E, \mathcal{E}) = (\Omega, \mathcal{A})$, a measurable function X is called a random variable.
- c) When $F = \mathbb{R}$ assume that we take $\mathcal{F} = \mathcal{B}$, the Borel σ -algebra of \mathbb{R} .

Theorem 8.1

Let \mathcal{C} be a class of subsets of F such that $\sigma(\mathcal{C}) = \mathcal{F}$. Then $X : E \rightarrow F$ is measurable (relative to \mathcal{E} and \mathcal{F}) iff $X^{-1}(\mathcal{C}) \subset \mathcal{E}$.

Proof: Suppose that X is measurable. Since $\mathcal{C} \subset \mathcal{F}$ and $X^{-1}(\mathcal{F}) \subset \mathcal{E}$, it follows that $X^{-1}(\mathcal{C}) \subset \mathcal{E}$.

Suppose that $X^{-1}(C) \in \mathcal{E}$ for any $C \in \mathcal{C}$. Define $\mathcal{B} = \{A \in \mathcal{F} : X^{-1}(A) \in \mathcal{E}\}$. Then for any $C \in \mathcal{C}$ we have $C \in \mathcal{F}$ and $X^{-1}(C) \in \mathcal{E}$ which means $\mathcal{C} \subset \mathcal{B}$. Since X^{-1} commutes with countable set operations and both \mathcal{E}, \mathcal{F} are σ -algebras, it follows that \mathcal{B} is also a σ -algebra. Therefore $\sigma(\mathcal{C}) \subset \sigma(\mathcal{B}) = \mathcal{B}$. But also $\mathcal{B} \subset \mathcal{F} = \sigma(\mathcal{C})$ so it must be the case that $\mathcal{B} = \mathcal{F}$. Since $X^{-1}(\mathcal{C}) \subset \mathcal{E}$, $\sigma(X^{-1}(\mathcal{C})) \subset \mathcal{E}$. Thus $X^{-1}(\mathcal{F}) \subset \sigma(X^{-1}(\mathcal{C})) \subset \mathcal{E}$.

Corollary 8.1

Let $(F, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$ and let (E, \mathcal{E}) be an arbitrary measurable space. Let X, X_n be real valued functions on E .

- a) X is measurable iff $\{X \leq a\} = \{\omega : X(\omega) \leq a\} = X^{-1}((-\infty, a]) \in \mathcal{E}$ for each $a \in \mathbb{R}$ (or iff $\{X < a\} \in \mathcal{E}$).
- b) If X_n are measurable, $\sup X_n, \inf X_n, \limsup X_n, \liminf X_n$ are measurable.
- c) If X_n are measurable and $X_n \rightarrow X$ pointwise, then X is measurable.

Theorem 8.2

Let $X : (E, \mathcal{E}) \rightarrow (F, \mathcal{F})$ be measurable and $Y : (F, \mathcal{F}) \rightarrow (G, \mathcal{G})$.

Proof: Let $A \in \mathcal{G}$. Since Y is measurable there exists a $B \in \mathcal{F}$ such that $Y^{-1}(A) = B$. Since X is measurable there exists a $C \in \mathcal{E}$ such that $X^{-1}(B) = C$. Then $(Y \circ X)^{-1}(A) = X^{-1}(Y^{-1}(A)) = X^{-1}(B) = C \in \mathcal{E}$. Since A was arbitrary conclude that $(Y \circ X)^{-1}(\mathcal{G}) \subset \mathcal{E}$, which means $Y \circ X$ is measurable (relative to \mathcal{E} and \mathcal{G}).

Definition A topological space is an abstract space with a collection of open sets. A “collection of open sets” is a collection of sets such that any union of sets in the collection is also in the collection, and any finite intersection of open sets in the collection is also in the collection. The collection of open sets is called the topology of the space.

Theorem 8.3

Let (E, \mathcal{U}) and (F, \mathcal{V}) be two topological spaces, and let \mathcal{E}, \mathcal{F} be their Borel σ -algebras. Every continuous function X from E into F is then measurable (also called “Borel”).

Theorem 8.4

Let $(F, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$, and (E, \mathcal{E}) be any measurable space.

- a) An indicator 1_A on E is measurable if and only if $A \in \mathcal{E}$.
- b) If X_1, \dots, X_n are real-valued measurable functions on (E, \mathcal{E}) , and if f is Borel on \mathbb{R}^n , then $f(X_1, \dots, X_n)$ is measurable.
- c) If X, Y are measurable, so also are $X+Y, XY, X \vee Y$ (a short-hand for $\max(X, Y)$), $X \wedge Y$ (a short-hand for $\min(X, Y)$), and X/Y (if $Y \neq 0$.)

Theorem 8.5

The distribution of X , (or the law of X), is a probability measure on (E, \mathcal{E}) .