MATH 3235: Probability Theory

Frank Qiang
Instructor: Christian Houdre

Georgia Institute of Technology
Fall 2024

Contents

Chapter 1

Events and Probabilities

1.1 Probability Spaces

Definition 1.1. A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ where

- Ω is called the *sample space* (the set of all possible outcomes of a random experiment);
- $\mathcal{F} \subseteq \mathcal{P}(\Omega)$, called the *event space*,[1] is nonempty and must satisfy:
  (i) if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$,
  (ii) if $A_1, A_2, \dots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$;
- $\mathbb{P}$ is a probabilty measure on $(\Omega, \mathcal{F})$ (to be defined later).

Remark. In general, when Ω is finite or countably infinite, one takes $\mathcal{F} = \mathcal{P}(\Omega)$.

Proposition 1.1. *We always have* $\varnothing, \Omega \in \mathcal{F}$.

Proof. Since $\mathcal{F} \neq \varnothing$, there exists some event $A \in \mathcal{F}$. Then we get $A^c \in \mathcal{F}$ and $\Omega = A \cup A^c \in \mathcal{F}$ by the complement and union properties of $\mathcal{F}$. Finally $\varnothing = \Omega^c \in \mathcal{F}$ by the complement property. □

Definition 1.2. A *probability measure* on $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \to [0, \infty)$ such that

(i) $\mathbb{P}(\Omega) = 1$,

(ii) and $\mathbb{P}(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mathbb{P}(A_k)$ whenever $A_1, A_2, \dots \in \mathcal{F}$ are pairwise disjoint.[2]

Proposition 1.2. *The following properties hold for any probability measure* $\mathbb{P}$ *on* $(\Omega, \mathcal{F})$*:*

(1) For any $A \in \mathcal{F}$*, we have* $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(2) Let $A, B \in \mathcal{F}$ *with* $A \subseteq B$*. Then* $\mathbb{P}(A) \leq \mathbb{P}(B)$.

(3) Let $A, B, C \in \mathcal{F}$*. Then*
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$
This is the principle of inclusion-exclusion.

Proof. (1) Observe that $A \cup A^c = \Omega$ and $A \cap A^c = \varnothing$, so $1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$.

[1] The elements of $\mathcal{F}$ are called *events*. Events with cardinality 1 are called *elementary*.
[2] i.e. $A_i \cap A_j = \varnothing$ whenever $i \neq j$.

(2) Write $B = A \cup (B \setminus A)$.[3] Since $A \cap (B \setminus A) = \varnothing$, we have $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$.[4]

(3) Left as an exercise. Follow similar ideas as in (2). □

Remark. Observe that property (2) implies $\mathbb{P}(A) \leq \mathbb{P}(\Omega) = 1$ since any $A \subseteq \Omega$.

Example 1.2.1. Pick a point uniformly at random from the unit square $\Omega = [0,1] \times [0,1]$ and record its coordinates. Then the probability of the point being inside a fixed shape $S \subseteq \Omega$ is $|S|$, the area of S.

Remark. Note that $\mathbb{P}$ only satisfies *countable* additivity. For instance let $\Omega = [0,1]$ and $\mathbb{P}$ be the uniform measure on Ω. Then $\Omega = \bigcup_{x \in [0,1]} \{x\}$ and $\mathbb{P}(\{x\}) = 0$ for every $x \in [0,1]$, but $\mathbb{P}(\Omega) = 1$. This is because the union $\bigcup_{x \in [0,1]} \{x\}$ is uncountable.

Definition 1.3. Let Ω be finite and $\mathcal{F} = \mathcal{P}(\Omega)$. The uniform probability on $(\Omega, \mathcal{F})$ is the one such that

$$\mathbb{P}(\{\omega\}) = \frac{1}{\operatorname{card} \Omega} \quad \text{for all } \omega \in \Omega.$$

Proposition 1.3. *Let $\mathbb{P}$ be the uniform probability on a finite set Ω and let $A \in \mathcal{F}$. Then*

$$\mathbb{P}(A) = \frac{\operatorname{card} A}{\operatorname{card} \Omega}.$$

Proof. Note that A is finite since Ω is and so we may enumerate its elements as $A = \{\omega_1, \omega_2, \ldots, \omega_n\}$, where $n = \operatorname{card} A$. Then the sets $\{\omega_i\}_{i=1}^n$ are pairwise disjoint and thus we have

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i=1}^n \{\omega_i\}\right) = \sum_{i=1}^n \mathbb{P}(\{\omega_i\}) = \sum_{i=1}^n \frac{1}{\operatorname{card} \Omega} = \frac{n}{\operatorname{card} \Omega} = \frac{\operatorname{card} A}{\operatorname{card} \Omega},$$

which is the desired result. □

1.2 Conditional Probability

Definition 1.4. Let $B \in \mathcal{F}$ such that $\mathbb{P}(B) > 0$. Then the *conditional probability* of A given B, written $\mathbb{P}(A|B)$, is given by

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

Remark. The intuition is that the extra information gained by knowing the occurrence of B should update our computation of the probability of A.

Remark. Another way to think about conditional probability is a restriction of the sample space to B.

Example 1.4.1. Suppose a family has two children, one of which is a girl. What is the probability the other is a girl? Define the sample space to be

$$\Omega = \{(B,G), (B,B), (G,G), (G,B)\}.$$

[3] Note that $B \setminus A \in \mathcal{F}$ since $B \setminus A = B \cap A^c = (B^c \cup A)^c$.

[4] Since $\mathbb{P} : \mathcal{F} \to [0, \infty)$, we have $\mathbb{P}(B \setminus A) \geq 0$.

Note that each elementary event is equally likely, i.e.

$$\mathbb{P}(\{(B,G)\}) = \mathbb{P}(\{(G,B)\}) = \mathbb{P}(\{(B,B)\}) = \mathbb{P}(\{(G,G)\}) = \frac{1}{4}.$$

Let $A = \{\text{both of them are } G\} = \{(G,G)\}$ and $B = \{\text{one is a girl}\} = \{(G,B),(B,G),(G,G)\}$. Then

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\{(G,G)\})}{\mathbb{P}(\{(G,B),(B,G),(G,G)\})} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Remark. If we instead condition on the event that one of them is a girl born on a Monday, then the probability changes! Carry out the calculation, and it should be 3/7 for a 7-day week.

Example 1.4.2. A drug test is 98% accurate, i.e. a drug user tests positive 98% of the time and a non-drug user tests negative 98% of the time. Among a given population, it is known 2% of people use drugs. Suppose I pick a person at random in the population and this person tests positive. What is the probability that the person is a drug user? Define the events

$$A = \text{the person is a drug user} \quad \text{and} \quad B = \text{the person tests positive}.$$

Then the goal is to compute $\mathbb{P}(A|B)$. The 98% accuracy assumption implies that $\mathbb{P}(B|A) = 0.98$. Now

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

Note that $B = (B\cap A)\cup(B\cap A^c)$ and this is a disjoint union, so

$$\begin{aligned}\mathbb{P}(B) = \mathbb{P}(B\cap A) + \mathbb{P}(B\cap A^c) &= \mathbb{P}(B|A)\mathbb{P}(A)\\ &+\mathbb{P}(B|A^c)\mathbb{P}(A^c) = 0.98(0.02) + 0.02(0.98) = 2(0.98)(0.02).\end{aligned}$$

Here we noted that the 98% accuracy of the test also implies that $\mathbb{P}(B|A^c) = 0.02$. Thus we get

$$\mathbb{P}(A|B) = \frac{0.98(0.02)}{2(0.98)(0.02)} = \frac{1}{2}.$$

Compute as an exercise that $\mathbb{P}(A^c|B^c) = 0.996$.

Remark. This test is designed clear non-drug users, not to identify drug users.

1.3 Bayes' Theorem

Definition 1.5. A *partition* of Ω is a collection or sequence of events $\{B_k\}_{k=1}^{\infty}$ such that

$$B_i \cap B_j = \varnothing \text{ for } i \neq j \quad \text{and} \quad \Omega = \bigcup_{k=1}^{\infty} B_k.$$

Remark. For any event A, observe that

$$A = A\cap\Omega = A\cap\left(\bigcup_{k=1}^{\infty} B_k\right) = \bigcup_{k=1}^{\infty}(A\cap B_k).$$

Then we get

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{k=1}^{\infty}(A \cap B_k)\right) = \sum_{k=1}^{\infty}\mathbb{P}(A \cap B_k) = \sum_{k=1}^{\infty}\mathbb{P}(A|B_k)\mathbb{P}(B_k).$$

This is the *partition theorem* in the book (Grimmett and Welsh). Now observe that

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(B_i \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B_i \cap A)}{\sum_{k=1}^{\infty}\mathbb{P}(A|B_k)\mathbb{P}(B_k)} = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{k=1}^{\infty}\mathbb{P}(A|B_k)\mathbb{P}(B_k)}$$

for each $i = 1, 2, \ldots$. This is *Bayes' theorem*, which relates posterior probabilities to prior probabilities.

1.4 Conditional Independence

Proposition 1.4. *Let B be such that $\mathbb{P}(B) > 0$. Then $Q : \mathcal{F} \to [0, 1]$ given by $A \mapsto Q(A) = \mathbb{P}(A|B)$ is a probability measure.*

Proof. (i) Observe that

$$Q(\Omega) = \mathbb{P}(\Omega|B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$$

(ii) Let $\{A_k\}_{k=1}^{\infty} \subseteq \mathcal{F}$ be pairwise disjoint. Then observe that we have

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

Thus Q is indeed a probability measure. □

Definition 1.6. Two events A and B are called *independent*, written $A \perp\!\!\!\perp B$, if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Example 1.6.1. Let A, B be events with $\mathbb{P}(A) = 0.6$ and $\mathbb{P}(B) = 0.8$. Then $0.4 \leq \mathbb{P}(A \cap B) \leq 0.6$. This is because $A \cap B \subseteq A$ implies $\mathbb{P}(A \cap B) \leq \mathbb{P}(A) = 0.6$. Also noting that $A \cap B \subseteq \Omega$ and so $\mathbb{P}(A \cap B) \leq \mathbb{P}(\Omega) = 1$ implies that

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B) \geq 0.6 + 0.8 - 1 = 0.4.$$

Note that if A and B were independent, then we can immediately conclude $\mathbb{P}(A \cap B) = 0.6(0.8) = 0.48$.

Proposition 1.5. *Assume A and B are events with $0 < \mathbb{P}(A), \mathbb{P}(B) < 1$. The following are equivalent:*

1. *A and B are independent,*
2. *$\mathbb{P}(A|B) = \mathbb{P}(A)$,*
3. *$\mathbb{P}(B|A) = \mathbb{P}(B)$,*
4. *$\mathbb{P}(A^c|B) = \mathbb{P}(A^c)$,*
5. *and $\mathbb{P}(B^c|A) = \mathbb{P}(B^c)$.*

Proof. Note that $\mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(A \cap B)$ and $A \perp\!\!\!\perp B$ if and only if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Then use cancellation since $\mathbb{P}(B) \neq 0$. Work out the rest as an exercise. □

Definition 1.7. We say that three events A, B, C are *independent* if A, B, C are pairwise independent[5] and $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$.

Remark. Pairwise independence does not imply independence. Consider flipping a fair coin twice. Let

$$A = \{\text{first flip is } T\}, \quad B = \{\text{second flip is } H\}, \quad C = \{\text{both flips are the same}\}.$$

Then A, B, C are pairwise independent but $\mathbb{P}(A \cap B \cap C) = 0 \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = 1/8$.

1.5 Homework Problems

Problems #1, 2, 9, 10, 14, 16, 17, 19 from Grimmett and Welsh.

[5] i.e. $\mathbb{P}(X \cap Y) = \mathbb{P}(X)\mathbb{P}(Y)$ for any $X, Y \in \{A, B, C\}$.