

MATH 3235: Probability Theory

Frank Qiang
Instructor: Christian Houdre

Georgia Institute of Technology
Fall 2024

Contents

| | | |
|----------|--|----------|
| 1 | Events and Probabilities | 2 |
| 1.1 | Probability Spaces | 2 |
| 1.2 | Conditional Probability | 3 |
| 1.3 | Bayes' Theorem | 4 |
| 1.4 | Conditional Independence | 5 |
| 1.5 | Continuity of Probability Measures | 6 |
| 1.6 | Homework Problems | 7 |
| 2 | Discrete Random Variables | 8 |
| 2.1 | Probability Mass Functions | 8 |

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 probability measure on (Ω, \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 $\emptyset, \Omega \in \mathcal{F}$.*

Proof. Since $\mathcal{F} \neq \emptyset$, 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 $\emptyset = \Omega^c \in \mathcal{F}$ by the complement property. \square

Definition 1.2. A *probability measure* on (Ω, \mathcal{F}) is a function $\mathbb{P} : \mathcal{F} \rightarrow [0, \infty)$ 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 (Ω, \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 $A \cap A^c = \emptyset$, so $1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$.

¹The elements of \mathcal{F} are called *events*. Events with cardinality 1 are called *elementary*.

²i.e. $A_i \cap A_j = \emptyset$ whenever $i \neq j$.

(2) Write $B = A \cup (B \setminus A)$.³ Since $A \cap (B \setminus A) = \emptyset$, 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). \square

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 (Ω, \mathcal{F}) is the one such that

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

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

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

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

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

which is the desired result. \square

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

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

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

⁴Since $\mathbb{P} : \mathcal{F} \rightarrow [0, \infty)$, 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

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

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

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

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

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

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 $i = 1, 2, \dots$. 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 $Q : \mathcal{F} \rightarrow [0, 1]$ 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

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

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. \square

Definition 1.7. We say that three events A, B, C are *independent* if A, B, C are pairwise independent⁵ 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 Continuity of Probability Measures

Remark. We want a system of probability that can say if I flip a fair coin infinitely many times, then

$$\mathbb{P}(\text{never get heads}) = 0.$$

For this experiment, we can set the sample space to be

$$\Omega = \{\text{all sequences like } (H, T, T, \dots)\}.$$

The event space \mathcal{F} is a little complicated, but it includes events like $\{\text{heads on the } n\text{th throw}\}$ and their complements and countable unions. We would like to show that $\mathbb{P}(\{(T, T, T, \dots)\}) = 0$. We know that

$$\mathbb{P}(A_n) = \mathbb{P}(\text{no heads in first } n \text{ tosses}) = \frac{1}{2^n}.$$

As $n \rightarrow \infty$, we can see that $2^{-n} \rightarrow 0$. If we have $\mathbb{P}(A_n) \rightarrow \mathbb{P}(\{(T, T, T, \dots)\})$, then we can conclude that

$$\mathbb{P}(\{(T, T, T, \dots)\}) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \lim_{n \rightarrow \infty} \frac{1}{2^n} = 0.$$

Notice that $A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots$ and $\bigcap_{n=1}^{\infty} A_n = \{(T, T, T, \dots)\}$. For sake of convenience, we will take complements and work with unions: We set

$$B_n = A_n^c = \{\text{at least one heads in first } n \text{ tosses}\},$$

then $B_1 \subseteq B_2 \subseteq B_3 \subseteq \dots$ and $\bigcup_{n=1}^{\infty} B_n = \Omega \setminus \{(T, T, T, \dots)\}$. But this union is not disjoint. To fix this, set $C_i = B_i \setminus B_{i-1}$, then

$$B_1 \cup \bigcup_{n=2}^{\infty} C_n = \Omega \setminus \{(T, T, T, \dots)\},$$

which is now a disjoint union. Taking probabilities, we can use countable additivity to get

$$\mathbb{P}(B_1 \cup C_2 \cup C_3 \cup \dots) = \mathbb{P}(B_1) + \mathbb{P}(C_2) + \mathbb{P}(C_3) + \dots \quad (*)$$

First, note that $B_1 \cup C_2 \cup C_3 \cup \dots = B_1 \cup B_2 \cup B_3 \cup \dots = \Omega \setminus \{(T, T, T, \dots)\}$. Thus in (*), the LHS is $1 - \mathbb{P}(\{(T, T, T, \dots)\})$. Now observe that in the RHS, we have $\mathbb{P}(C_i) = \mathbb{P}(B_i) - \mathbb{P}(B_{i-1})$. Then

$$\begin{aligned} \mathbb{P}(B_1) + \mathbb{P}(C_2) + \mathbb{P}(C_3) + \dots &= \lim_{n \rightarrow \infty} [\mathbb{P}(B_1) + \mathbb{P}(C_2) + \dots + \mathbb{P}(B_n)] \\ &= \lim_{n \rightarrow \infty} [\mathbb{P}(B_1) + (\mathbb{P}(B_2) - \mathbb{P}(B_1)) + \dots + (\mathbb{P}(B_n) - \mathbb{P}(B_{n-1}))] \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(B_n) = \lim_{n \rightarrow \infty} [1 - \mathbb{P}(A_n)] = \lim_{n \rightarrow \infty} \left[1 - \frac{1}{2^n}\right] = 1. \end{aligned}$$

Thus matching the LHS and RHS, we see that $1 - \mathbb{P}(\{(T, T, T, \dots)\}) = 1$ and so $\mathbb{P}(\{(T, T, T, \dots)\}) = 0$.

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

Theorem 1.1 (Continuity of probability measures). *If $\{B_i\}_{i=1}^\infty$ are nested events,⁶ then*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \lim_{n \rightarrow \infty} \mathbb{P}(B_n).$$

Proof. This was essentially the argument in the previous remark. □

1.6 Homework Problems

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

⁶i.e. $B_1 \subseteq B_2 \subseteq B_3 \subseteq \dots$

Chapter 2

Discrete Random Variables

2.1 Probability Mass Functions

Example 2.0.1. Consider the following game: Flip a fair coin 10 times and roll a fair die. I give you
(number of heads) \times (number on die) dollars.

This is a simple game, but it is kind of painful to write in terms of events (e.g. $\mathbb{P}(\text{win} \geq \$10)$). We would have to set

$$\Omega = \{\text{all sequences like } (H, T, H, H, T, T, T, T, T, H, 4)\}$$

and $\mathcal{F} = \mathcal{P}(\Omega)$. It is also not immediately obviously which sequences are in $\{\text{win} \geq \$10\}$. Instead, we would prefer something like

“Let H be the number of heads in 10 fair coin tosses and let R be the outcome of a roll of a fair die. Then you get HR dollars.”

How do we do this in our axiomatic framework? What are H, R ? Here are some observations:

- H, R are real numbers,
- and they are determined by the outcome of the experiment.

Thus we should think of H, R as functions from Ω to \mathbb{R} . These are examples of *discrete random variables*.

Remark. The name “random variable” is just historic. Really, H, R are non-random functions.

Remark. Can every function $X : \Omega \rightarrow \mathbb{R}$ be a discrete random variable? Note that we want to talk about probabilities like $\mathbb{P}(X = 17)$. This indicates that the event

$$\{X = 17\} = \{\omega \in \Omega : X(\omega) = 17\}$$

has to be in \mathcal{F} . So we require that X is *measurable*, i.e. for every $x \in \mathbb{R}$, we have $\{x \in \Omega : X(\omega) = x\} \in \mathcal{F}$. Also H, R must have special properties, for instance they can only take on finitely many values.

Definition 2.1. A function $X : \Omega \rightarrow \mathbb{R}$ is a *discrete random variable* if

- (i) for every $x \in \mathbb{R}$, we have $\{X = x\} \in \mathcal{F}$,
- (ii) and $X(\Omega) = \{x \in \mathbb{R} : x = X(\omega) \text{ for some } \omega\}$ is finite or countably infinite.

Remark. Often, we only care about what values X can take and with what probabilities. We store this data in a special function called the *probability mass function*.

Definition 2.2. Let X be a discrete random variable. Then its *probability mass function (pmf)* is

$$p_X : \mathbb{R} \rightarrow [0, 1] \quad \text{defined by} \quad p_X(s) = \mathbb{P}(X = s).$$

Example 2.2.1. Let X be the outcome of the roll of a fair die. Then

$$p_X(s) = \begin{cases} 1/6 & \text{if } s \in \{1, 2, 3, 4, 5, 6\}, \\ 0 & \text{otherwise.} \end{cases}$$

Remark. Another sentence we want to say is:

“A discrete random variable X takes values $\{1, 7, 9\}$ with probabilities $1/2, 1/3, 1/6$, respectively if and only if

$$p_X(s) = \begin{cases} 1/2 & \text{if } s = 1, \\ 1/3 & \text{if } s = 7, \\ 1/6 & \text{if } s = 9, \\ 0 & \text{otherwise.} \end{cases}$$

How do we know this exists? In other words, does there exist $(\Omega, \mathcal{F}, \mathbb{P})$ and $X : \Omega \rightarrow \mathbb{R}$ with this pmf?

Theorem 2.1. Let $S = \{s_i : i \in I\}$ be a countable subset of \mathbb{R} and let $\{\pi_i : i \in I\}$ be a collection of numbers such that $\pi_i \geq 0$ and

$$\sum_{i \in I} \pi_i = 1.$$

Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a discrete random variable $X : \Omega \rightarrow \mathbb{R}$ such that

$$p_X(s) = \begin{cases} \pi_i & \text{if } s = s_i, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Take $\Omega = S$ and $\mathcal{F} = \mathcal{P}(S)$. Set

$$\mathbb{P}(A) = \sum_{i: s_i \in A} \pi_i$$

and define $X : \Omega \rightarrow \mathbb{R}$ given by $X(\omega) = \omega$. Then one can check that X has the desired pmf. \square

Remark. This allows us to just say

“Let X be a discrete random variable taking these values with these probabilities”

without worrying about the underlying $(\Omega, \mathcal{F}, \mathbb{P})$.

Example 2.2.2. Some common examples of discrete random variables are:

1. Bernoulli random variables: For $p \in [0, 1]$, we say that $X \sim \text{Bernoulli}(p)$ if

$$X = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } q = 1 - p. \end{cases}$$

This models a possibly unfair coin flip. The Bernoulli random variable X has pmf

$$p_X(s) = \begin{cases} p & \text{if } s = 1, \\ 1 - p & \text{if } s = 0, \\ 0 & \text{otherwise.} \end{cases}$$