

Probability

ADAM KELLY

January 23, 2021

This set of notes is a work-in-progress account of the course ‘Probability’, originally lectured by Dr Perla Sousi in Lent 2020 at Cambridge. These notes are not a transcription of the lectures, but they do roughly follow what was lectured (in content and in structure).

These notes are my own view of what was taught, and should be somewhat of a superset of what was actually taught. I frequently provide different explanations, proofs, examples, and so on in areas where I feel they are helpful. Because of this, this work is likely to contain errors, which you may assume are my own. If you spot any or have any other feedback, I can be contacted at ak2316@cam.ac.uk.

Contents

1 Basic Concepts	4
1.1 Probability Space	4
1.2 Combinatorial Analysis	6
1.3 Stirling's Formula	7
1.4 Properties of Probability Measures	9
1.4.1 Countable Subadditivity	9

1 Basic Concepts

Most of the phenomena in everyday lives involve randomness. What we try to do in probability is model this randomness in a mathematical way. It's likely that you have studied some probability before, but the difference in the treatment here is that we will try to be somewhat more rigorous.

We will define the notion of a probability space, where 'our experiments take place'. Then we will discuss discrete and continuous random variables. In the discrete setting, we will find that there is no real subtleties, and we can be quite rigorous. In the continuous setting however we will have to take some things for granted (but rigour will return in the Part II course).

"Probability theory has a right and a left hand. On the right is the rigorous foundational work using the tools of measure theory. The left hand 'thinks probabilistically,' reduces problems to gambling situations, coin-tossing, motions of a physical particle."

In this course, we will need both hands.

§1.1 Probability Space

Probability is the mathematical formulation of randomness. So in order to study random phenomena in a rigorous way, we first need to set out a rigorous mathematical framework.

The first notion that we will define is that of a *probability space*.

Definition 1.1.1 (σ -Algebra)

Suppose Ω is a set and \mathcal{F} is a collection of subsets of Ω . We call \mathcal{F} a **σ -algebra** if the following properties are satisfied.

- (i) $\Omega \in \mathcal{F}$.
- (ii) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$, the compliment of A .
- (iii) For any countable collection A_1, A_2, \dots with $A_i \in \mathcal{F}$ for all i , we must also have that $\bigcup_{i \geq 1} A_i \in \mathcal{F}$.

Definition 1.1.2 (Probability Measure)

Suppose that \mathcal{F} is a σ -algebra on Ω . Then a function $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ is called a **probability measure** if the following are true.

- (i) $\mathbb{P}(\Omega) = 1$.
- (ii) For any countable disjoint collection A_1, A_2, \dots with $A_i \in \mathcal{F}$ for all i , we have

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

We say that $\mathbb{P}(A)$ is the **probability** of A .

Definition 1.1.3 (Probability Space)

If \mathcal{F} is a σ -algebra on Ω and \mathbb{P} is a probability measure, then $(\Omega, \mathcal{F}, \mathbb{P})$ is a **probability space**.

When the set Ω is countable, we take \mathcal{F} to be all subsets of Ω .

Definition 1.1.4 (Outcomes and Events)

The elements of Ω are called **outcomes**, and the elements of \mathcal{F} are called **events**.

Note that \mathbb{P} is defined on \mathcal{F} , so it is defined on the *events*, not the *outcomes*.

Let's look at some properties of the probability measure (which follow immediately from the definition).

Proposition 1.1.5 (Properties of \mathbb{P})

If \mathbb{P} is a probability measure then

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- $\mathbb{P}(\emptyset) = 0$
- If $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Proof Sketch. Check definitions. □

Example 1.1.6 (Examples of Probability Spaces)

Some examples of probability spaces are given below.

- *Rolling a fair die.* Consider rolling a fair die. Then $\Omega = \{1, 2, 3, 4, 5, 6\}$, and \mathcal{F} is all subsets of Ω . Then $\mathbb{P}(\{\omega\}) = \frac{1}{6}$ for all $\omega \in \Omega$ and if $A \subseteq \Omega$, then $\mathbb{P}(A) = \frac{|A|}{6}$.
- *Equally likely outcomes.* Let Ω be a finite set, $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, and \mathcal{F} be all subsets of Ω . Then define $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ by $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$. In classical probability, this models picking a random element of Ω . Note that $\mathbb{P}(\{\omega_i\}) = \frac{1}{|\Omega|}$ for all $\omega_i \in \Omega$.
- *Picking balls from a bag.* Suppose we have n balls with n labels from $\{1, \dots, n\}$ that are all indistinguishable by touch. Picking $k \leq n$ balls at random^a without replacement. Then we take $\Omega = \{A \subseteq \{1, 2, \dots, n\} \mid |A| = k\}$, and \mathcal{F} be all subsets of Ω . Then $|\Omega| = \frac{n!}{k!(n-k)!}$, and for $\omega \in \Omega$, $\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}$.
- *Deck of cards.* Take a well-shuffled deck of 52 cards. Then let Ω be the set of all permutations of the cards, and note $|\Omega| = 52!$. Then we have $\mathbb{P}(\text{top 2 cards are aces}) = \frac{4 \cdot 3 \cdot 50!}{52!} = \frac{1}{221}$.

- *Largest digit.* Consider a string of n random digits from $0, \dots, 9$. Then $\Omega = \{0, 1, \dots, 9\}^n$, and $|\Omega| = 10^n$. Now define $A_k = \{\text{no digit exceeds } k\}$ and $B_k = \{\text{largest digit is } k\}$. Then $\mathbb{P}(B_k) = \frac{|B_k|}{|\Omega|}$. Notice that $B_k = A_k \setminus A_{k-1}$, and $|A_k| = (k+1)^n$, so $|B_k| = (k+1)^n - k^n$, thus $\mathbb{P}(B_k) = \frac{(k+1)^n - k^n}{10^n}$.
- *Birthday Problem.* There are n people. What is the probability that at least two of them share the same birthday? We can assume nobody is born on 29/02, and that each birthday is equally likely. So $\Omega = \{1, \dots, 365\}^n$, and \mathcal{F} is all subsets of Ω . As we assumed all outcomes are equally likely, we take $\mathbb{P}(\{\omega\}) = \frac{1}{365^n}$ with $\omega \in \Omega$. Letting $A = \{\text{at least 2 people share a birthday}\}$, then $A^c = \{\text{all } n \text{ birthdays are different}\}$, and since $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$, it suffices to calculate $\mathbb{P}(A^c)$. Now $\mathbb{P}(A^c) = \frac{|A^c|}{|\Omega|} = \frac{365 \times 364 \times \dots \times (365 - n + 1)}{365^n}$, and hence $\mathbb{P}(A) = 1 - \frac{365 \times 364 \times \dots \times (365 - n + 1)}{365^n}$. If $n = 23$, then this probability is approximately 0.507.

^aThat is, with all outcomes equally likely.

§1.2 Combinatorial Analysis

Suppose we have some finite set Ω , and that $|\Omega| = n$. We want to partition Ω into k disjoint subsets $\Omega_1, \Omega_2, \dots, \Omega_k$ with $|\Omega_i| = n_i$ and $\sum_{i=1}^k n_i = n$. How many ways is there to do this?

If M is the number of ways, then

$$M = \binom{n}{n_1} \binom{n - n_1}{n_2} \dots \binom{n - (n_1 + \dots + n_{k-1})}{n_k} = \frac{n!}{n_1! n_2! \dots n_k!}.$$

Definition 1.2.1 (Multinomial Coefficient)

We define the **multinomial coefficient** $\binom{n}{n_1, \dots, n_k}$ to be the number of ways of partitioning a set with n elements into k subsets of size n_1, n_2, \dots, n_k . We have

$$\binom{n}{n_1, \dots, n_k} = \frac{n!}{n_1! n_2! \dots n_k!}.$$

Now let's think about the following question: How many strictly increasing and increasing functions are there between two sets?

If we have $f : \{1, \dots, k\} \rightarrow \{1, \dots, n\}$, we say it's *strictly increasing* if whenever $x < y$, then $f(x) < f(y)$. We say that it's *increasing* if $x < y$ implies $f(x) \leq f(y)$.

Any such function is uniquely determined by its range which is a subset of $\{1, \dots, n\}$ of size k . There are $\binom{n}{k}$ such subsets, and hence $\binom{n}{k}$ strictly increasing functions.

We define a bijection from $\{f : \{1, \dots, k\} \rightarrow \{1, \dots, n\} \mid f \text{ increasing}\}$ to $\{g : \{1, \dots, k\} \rightarrow \{1, \dots, n+k-1\} \mid f \text{ strictly increasing}\}$. For each f , we define $g(i) = f(i) + i - 1$. Then g is strictly increasing and takes values in $\{1, \dots, n+k-1\}$. Then g is strictly increasing and takes values in $\{1, \dots, n+k-1\}$.

So the total number of increasing functions $f : \{1, \dots, k\} \rightarrow \{1, \dots, n\}$ is $\binom{n+k-1}{k}$.

§1.3 Stirling's Formula

Frequently in probability it will help to have some bounds/an asymptotic expression for the factorial.

Notation. Let (a_n) and (b_n) be two sequences. We will write $a_n \sim b_n$ if $\frac{a_n}{b_n} \rightarrow 1$ as $n \rightarrow \infty$.

We will first prove a weak approximation of the factorial.

Proposition 1.3.1

$\log(n!) \sim n \log n$ as $n \rightarrow \infty$.

Proof. For $x \in \mathbb{R}$, we write $\lfloor x \rfloor$ for the *integer part* of x . Then we have $\log \lfloor x \rfloor \leq \log x \leq \log \lfloor x + 1 \rfloor$. Integrating this from 1 to n , we get

$$\begin{aligned} \sum_{k=1}^{n-1} \log k &\leq \int_1^n \log x \, dx \leq \sum_{k=1}^n \log k \\ \implies \log(n-1)! &\leq n \log n - n + 1 \leq \log n! \end{aligned}$$

Using this, we get that

$$n \log n - n + 1 \leq \log n! \leq (n+1) \log(n+1) - (n+1) + 1,$$

and dividing through by $n \log n$ we get

$$\frac{\log n!}{n \log n} \rightarrow 1 \quad n \rightarrow \infty.$$

□

Now let's prove Stirling's formula. Note that the proof is non-examinable.

Theorem 1.3.2 (Stirling Approximation)

$n! \sim n^n \sqrt{2\pi n} e^{-n}$ as $n \rightarrow \infty$.

Proof (Non-Examinable). For any function f that is twice differentiable, and $a < b$, then

$$\int_a^b f(x) \, dx = \frac{f(a) + f(b)}{2} (b - a) - \frac{1}{2} \int_a^b (x - a)(b - x) f''(x) \, dx.$$

This follows by integrating by parts.

Now take $f(x) = \log x$, and $a = k$, $b = k + 1$. Then substituting into the formula, we get

$$\begin{aligned} \int_k^{k+1} \log x \, dx &= \frac{\log k + \log(k+1)}{2} - \frac{1}{2} \int_k^{k+1} \frac{(x-k)(k+1-x)}{x^2} \, dx \\ &= \frac{\log k + \log(k+1)}{2} + \frac{1}{2} \int_0^1 \frac{x(1-x)}{(x+k)^2} \, dx. \end{aligned}$$

Then taking the sum for $k = 1, \dots, n-1$ of the equality above, we get

$$\begin{aligned} \int_1^n \log x \, dx &= \frac{\log(n-1)! + \log n!}{2} + \frac{1}{2} \sum_{k=1}^{n-1} \int_0^1 \frac{x(1-x)}{(x+k)^2} \, dx \\ \implies \log n - n + 1 &= \log(n!) - \frac{\log n}{2} + \sum_{k=1}^{n-1} a_k, \end{aligned}$$

where we set

$$a_k = \frac{1}{2} \int_0^1 \frac{x(1-x)}{(x+k)^2} \, dx.$$

But then

$$\begin{aligned} \log n! &= n \log n - n + \frac{\log n}{2} + 1 - \sum_{k=1}^{n-1} a_k \\ \implies n! &= n^n e^{-n} \cdot \sqrt{n} \exp \left(1 - \sum_{k=1}^{n-1} a_k \right). \end{aligned}$$

Note that $a_k \leq \frac{1}{2} \int_0^1 \frac{x(1-x)}{k^2} dx = \frac{1}{12k^2}$, so $\sum a_k < \infty$. We set $A = \exp(1 - \sum_{k=1}^{\infty} a_k)$. Then

$$n! = n^n \cdot e^{-n} \sqrt{n} \cdot A \cdot \exp \left(\sum_{k=n}^{\infty} a_k \right),$$

and as $\exp(\sum_{k=n}^{\infty} a_k) \rightarrow 1$ (since the argument goes to 0), we have proved that

$$\frac{n!}{n^n e^{-n} \sqrt{n}} \rightarrow A, \quad \text{as } n \rightarrow \infty,$$

which means that $n! \sim n^n e^{-n} \sqrt{n} \cdot A$ as $n \rightarrow \infty$.

To finish the proof, we need to show that $A = \sqrt{2\pi}$. Knowing that $n! \sim n^n e^{-n} \sqrt{n} \cdot A$ as $n \rightarrow \infty$, we have

$$2^{-2n} \cdot \binom{2n}{n} = 2^{-2n} \frac{(2n)!}{n! \cdot n!} \sim \frac{2^{-2n} \cdot (2n)^{2n} \cdot \sqrt{2n} \cdot A \cdot e^{-2n}}{n^n \cdot e^{-n} \cdot \sqrt{n} \cdot A \cdot n^n \cdot e^{-n} \cdot \sqrt{n} \cdot A} = \frac{\sqrt{2}}{A\sqrt{n}}.$$

Using a different method we will prove that

$$2^{2n} \binom{2n}{n} \sim \frac{1}{\sqrt{\pi n}},$$

which will force $A = \sqrt{2\pi}$.

Consider the integral

$$I_n = \int_0^{2\pi} (\cos \theta)^n \, d\theta, \quad n \geq 0.$$

So $I_0 = \pi/2$ and $I_1 = 1$. Then integrating by parts, we get $I_n = \frac{n-1}{n} I_{n-2}$, and thus

$$I_{2n} = \frac{2n-1}{2n} \cdot I_{2n-2} = \frac{(2n-1)(2n-3) \cdots 3 \cdot 1}{2n \cdot (2n-2) \cdots 2} I_0 = \frac{(2n)!}{2^{2n} n! \cdot n!} \cdot \frac{\pi}{2},$$

so

$$I_{2n} = 2^{-2n} \binom{2n}{n} \frac{\pi}{2}.$$

In the same way we get

$$I_{2n+1} = \frac{2n \cdots 4 \cdot 2}{(2n+1) \cdots 3 \cdot 1} I_1 = \frac{1}{2n+1} \left(2^{-2n} \binom{2n}{n} \right)^{-1}.$$

From $I_n = \frac{n-1}{n} I_{n-2}$ we get $\frac{I_n}{I_{n-2}} \rightarrow 1$ as $n \rightarrow \infty$, and what we want is $\frac{I_{2n}}{I_{2n+1}} \rightarrow 1$ as $n \rightarrow \infty$.

Note that I_n is a decreasing function of n , therefore

$$\frac{I_{2n}}{I_{2n+1}} \leq \frac{I_{2n-1}}{I_{2n+1}} \rightarrow 1,$$

and also

$$\frac{I_{2n}}{I_{2n+1}} \geq \frac{I_{2n}}{I_{2n-2}} \rightarrow 1,$$

thus $\frac{I_{2n}}{I_{2n+1}} \rightarrow 1$ as $n \rightarrow \infty$, which means

$$\begin{aligned} & \frac{2^{-2n} \binom{2n}{n} \frac{\pi}{2}}{\left(2^{-2n} \binom{2n}{n} \right)^{-1} \frac{1}{2n+1}} \rightarrow 1 \\ \implies & \left(2^{-2n} \binom{2n}{n} \right)^2 \frac{\pi}{2} (2n+1) \rightarrow 1, \end{aligned}$$

thus

$$\left(2^{-2n} \binom{2n}{n} \right)^2 \sim \frac{2}{\pi(2n+1)} \sim \frac{1}{\pi n},$$

thus $A = \sqrt{2\pi}$, which completes the proof. \square

§1.4 Properties of Probability Measures

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Recall that the probability measure \mathbb{P} is a function

$$\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$$

with $\mathbb{P}(\Omega) = 1$ which is *countable additive*, that is, for any countable disjoint collection A_1, A_2, \dots with $A_i \in \mathcal{F}$ for all i ,

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

§1.4.1 Countable Subadditivity

When the sequence is not necessarily disjoint, this equality becomes an inequality.

Proposition 1.4.1 (Countable Subadditivity)

Let (A_n) be a sequence of events with $A_i \in \mathcal{F}$ for all i . Then we have

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

Proof. Define $B_1 = A_1$ and $B_n = A_n \setminus (A_1 \cup \dots \cup A_{n-1})$ for all $n \geq 2$. Then (B_n) is a disjoint sequence of events in \mathcal{F} , and $\bigcup_{n \geq 1} B_n = \bigcup_{n \geq 1} A_n$. So $\mathbb{P}(\bigcup A_n) = \mathbb{P}(\bigcup B_n)$. By countable additivity for (B_n) ,

$$\mathbb{P}\left(\bigcup_{n \geq 1} B_n\right) = \sum_{n \geq 1} \mathbb{P}(B_n).$$

But $B_n \subseteq A_n$, so $\mathbb{P}(B_n) \leq \mathbb{P}(A_n)$ for all n . Therefore

$$\mathbb{P}\left(\bigcup A_n\right) = \mathbb{P}\left(\bigcup B_n\right) = \sum \mathbb{P}(B_n) \leq \sum_{n \geq 1} \mathbb{P}(A_n).$$

□