4. Countable and uncountable

Definition 32. An set Ω is said to be *finite* if there is an $n \in \mathbb{N}$ and a bijection from Ω onto $[n]$. An infinite set Ω is said to be countable if there is a bijection from $\mathbb{N}$ onto Ω.

Generally, the word countable also includes finite sets. If Ω is an infinite countable set, then using any bijection $f : \mathbb{N} \to \Omega$, we can list the elements of Ω as a sequence

$$f(1), f(2), f(3) \dots$$

so that each element of Ω occurs exactly once in the sequence. Conversely, if you can write the elements of Ω as a sequence, it defines an injective function from natural numbers onto Ω (send 1 to the first element of the sequence, 2 to the second element etc).

Example 33. The set of integers $\mathbb{Z}$ is countable. Define $f : \mathbb{N} \to \mathbb{Z}$ by

$$f(n) = \begin{cases} \frac{1}{2}n & \text{if } n \text{ is even.} \\ -\frac{1}{2}(n-1) & \text{if } n \text{ is odd.} \end{cases}$$

It is clear that f maps $\mathbb{N}$ into $\mathbb{Z}$. Check that it is one-one and onto. Thus, we have found a bijection from $\mathbb{N}$ onto $\mathbb{Z}$ which shows that $\mathbb{Z}$ is countable. This function is a formal way of saying the we can list the elements of $\mathbb{Z}$ as

$$0, +1, -1, +2, -2, +3, -3, \dots.$$

It is obvious, but good to realize there are wrong ways to try writing such a list. For example, if you list all the negative integers first, as $-1, -2, -3, \dots$, then you will never arrive at 0 or 1, and hence the list is incomplete!

Example 34. The set $\mathbb{N} \times \mathbb{N}$ is countable. Rather than give a formula, we list the elements of $\mathbb{Z} \times \mathbb{Z}$ as follows.

$$(1,1),\ (1,2),(2,1),\ (1,3),(2,2),(3,1),\ (1,4),(2,3),(3,2),(4,1),\ \dots$$

The pattern should be clear. Use this list to define a bijection from $\mathbb{N}$ onto $\mathbb{N} \times \mathbb{N}$ and hence show that $\mathbb{N} \times \mathbb{N}$ is countable.

Example 35. The set $\mathbb{Z} \times \mathbb{Z}$ is countable. This follows from the first two examples. Indeed, we have a bijection $f : \mathbb{N} \to \mathbb{Z}$ and a bijection $g : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$. Define a bijection $F : \mathbb{N} \times \mathbb{N} \to \mathbb{Z} \times \mathbb{Z}$ by composing them, i.e., $F(n,m) = f(g(n))$. Then, F is one-one and onto. This shows that $\mathbb{Z} \times \mathbb{Z}$ is indeed countable.

Example 36. The set of rational numbers $\mathbb{Q}$ is countable. Recall that rational numbers other than 0 can be written uniquely in the form p/q where p is a non-zero integer and q is a strictly positive integer, and there are no common factors of p and q (this is called the *lowest form* of the rational number r). Consider the map $f : \mathbb{Q} \to \mathbb{Z} \times \mathbb{Z}$ defined by

$$f(r) = \begin{cases} (0,1) & \text{if } r = 0 \\ (p,q) & \text{if } r = \frac{p}{q} \text{ in the lowest form.} \end{cases}$$

Clearly, f is injective and hence, it appears that $\mathbb{Z} \times \mathbb{Z}$ is a "bigger set" than $\mathbb{Q}$. Next define the function $g : \mathbb{Z} \to \mathbb{Q}$ by setting $g(n) = n$. This is also injective and hence we may say that "$\mathbb{Q}$ is a bigger set than $\mathbb{N}$.

But we have already seen that $\mathbb{N}$ and $\mathbb{Z} \times \mathbb{Z}$ are in bijection with each other, in that sense, they are of equal size. Since $\mathbb{Q}$ is sandwiched between the two it ought to be true that $\mathbb{Q}$ has the same size as $\mathbb{N}$, and thus countable.

This reasoning is not incorrect, but an argument is needed to make it an honest proof. This is indicated in the Schröder-Bernstein theorem stated later. Use that to fill the gap in the above argument, or alternately, try to directly find a bijection between $\mathbb{Q}$ and $\mathbb{N}$.

Example 37. The set of real numbers $\mathbb{R}$ is not countable. The extraordinarily proof of this fact is due to Cantor, and the core idea, called the *diagonalization trick* is one that can be used in many other contexts.

Consider any function $f : \mathbb{N} \to [0,1]$. We show that it is not onto, and hence not a bijection. Indeed, use the decimal expansion to write a number $x \in [0,1]$ as $0.x_1x_2x_3\ldots$ where $x_i \in \{0,1,\ldots,9\}$. Write the decimal expansion for each of the numbers $f(1), f(2), f(3), \ldots$ as follows.

$$\begin{aligned} f(1) &= 0.X_{1,1}X_{1,2}X_{1,3}\ldots \\ f(2) &= 0.X_{2,1}X_{2,2}X_{2,3}\ldots \\ f(3) &= 0.X_{3,1}X_{3,2}X_{3,3}\ldots \\ &\ldots\ldots\ \ldots\ \ldots \end{aligned}$$

Let $Y_1, Y_2, Y_3, \ldots$ be any numbers in $\{0,1,\ldots,9\}$ with the only condition that $Y_i \neq X_{i,i}$. Clearly it is possible to choose Y_i like this. Now consider the number $y = 0.Y_1Y_2Y_3\ldots$ which is a number in $[0,1]$. However, it does not occur in the above list. Indeed, y disagrees with $f(1)$ in the first decimal place, disagrees with $f(2)$ in the second decimal place etc. Thus, $y \neq f(i)$ for any $i \in \mathbb{N}$ which means that f is not onto $[0,1]$.

Thus, no function $f : \mathbb{N} \to [0,1]$ is onto, and hence there is no bijection from $\mathbb{N}$ onto $[0,1]$ and hence $[0,1]$ is not countable. Obviously, if there is no onto function onto $[0,1]$, there cannot be an onto function onto $\mathbb{R}$. Thus, $\mathbb{R}$ is also uncountable.

Example 38. Let $A_1, A_2, \ldots$ be subsets of a set Ω. Suppose each A_i is countable (finite is allowed). Then $\cup_i A_i$ is also countable. We leave it as an exercise. [*Hint:* If each A_i is countably infinite and pairwise disjoint, then $\cup A_i$ can be thought of as $\mathbb{N} \times \mathbb{N}$].

Lemma 39 (Schröder-Bernstein)**.** *Let A, B be two sets and suppose there exist injective functions $f : A \to B$ and $g : B \to A$. Then, there exists a bijective function $h : A \to B$.*

We omit the proof as it is irrelevant to the rest of the course[7].

5. On infinite sums

There were some subtleties in the definition of probabilities which we address now. The definition of $\mathbf{P}(A)$ for an event A and $\mathbf{E}[X]$ for a random variable X involve infinite sums (when Ω is countably infinite). In fact, in the very definition of probability space, we had the condition that $\sum_\omega p_\omega = 1$, but what is the meaning of this sum when Ω is infinite?

[7]For those interested, we describe the idea of the proof somewhat informally. Consider the two sets A and B (assumed to have no common elements) and draw a blue arrow from each $x \in A$ to $f(x) \in B$ and a red arrow from each $y \in B$ to $g(y) \in A$. Start at any $x \in A$ or $y \in B$ and follow the arrows in the forward and backward directions. There are only three possibilities

(1) The search closes, and we discover a cycle of alternating blue and red arrows.
(2) The backward search ends after finitely many steps and the forward search continues forever.
(3) Both the backward and forward searches continue forever.

The injectivity of f and g is used in checking that these are the only possibilities. In the first and third case, just use the blue arrows to define the function h. In the second case, if the first element of the chain is in A, use the blue arrows, and if the first element is in B use the red arrows (but in reverse direction) to define the function h. Check that the resulting function is a bijection!

In this section, we make precise the notion of infinite sums. In fact we shall give two methods of approach, it suffices to consider only the first.

5.1. First approach. Let Ω be a countable set, and let $f : \Omega \to \mathbb{R}$ be a function. We want to give a meaning to the infinite sum $\sum_{\omega\in\Omega} f(\omega)$. First we describe a natural attempt and then address the issues that it leaves open.

The idea: By definition of countability, there is a bijection $\varphi : \mathbb{N} \to \Omega$ which allows us to list the elements of Ω as $\omega_1 = \varphi(1), \omega_2 = \varphi(2), \ldots$. Consider the partial sums $x_n = f(\omega_1) + f(\omega_2) + \ldots + f(\omega_n)$. Since f is non-negative, these numbers are non-decreasing, i.e., $x_1 \leq x_2 \leq x_3 \leq \ldots$. Hence, they converge to a finite number or to $+\infty$ (which is just another phrase for saying that the partial sums grow without bound). We would like to simply define the sum $\sum_{\omega\in\Omega} f(\omega)$ as the limit $L = \lim_{n\to\infty}(f(\omega_1) + \ldots + f(\omega_n)$, which may be finite or $+\infty$.

The problem is that this may depend on the bijection Ω chosen. For example, if $\psi : \mathbb{N} \to \Omega$ is a different bijection, we would write the elements of Ω in a different sequence $\omega'_1 = \psi(1), \omega'_2 = \psi(2), \ldots$, the partial sums $y_n = f(\omega'_1) + \ldots + f(\omega'_n)$ and then define $\sum_{\omega\in\Omega} f(\omega)$ as the limit $L' = \lim_{n\to\infty}(f(\omega'_1) + \ldots + f(\omega'_n)$.

Is it necessarily true that $L = L'$?

Case I - Non-negative f: We claim that for any two bijections φ and ψ as above, the limits are the same (this means that the limits are $+\infty$ in both cases, or the same finite number in both cases). Indeed, fix any n and recall that $x_n = f(\omega_1) + \ldots + f(\omega_n)$. Now, ψ is surjective, hence there is some m (possibly very large) such that $\{\omega_1, \ldots, \omega_n\} \subseteq \{\omega'_1, \ldots, \omega'_m\}$. Now, we use the non-negativity of f to observe that

$$f(\omega_1) + \ldots + f(\omega_n) \leq f(\omega'_1) + \ldots + f(\omega'_m).$$

This is the same as $x_n \leq y_m$. Since y_k are non-decreasing, it follows that $x_n \leq y_m \leq y_{m+1} \leq y_{m+2} \ldots$, which implies that $x_n \leq L'$. Now let $n \to \infty$ and conclude that $L \leq L'$. Repeat the argument with the roles of φ and ψ reversed to conclude that $L' \leq L$. Hence $L = L'$, as desired to show.

In conclusion, for non-negative functions f, we can assign an unambiguous meaning to $\sum_\omega f(\omega)$ by setting it equal to $\lim_{n\to\infty}(f(\varphi(1) + \ldots + f(\varphi(n)))$, where $\varphi : \mathbb{N} \to \Omega$ is any bijection (the point being that the limit does not depend on the bijection chosen), and the limit here may be allowed to be $+\infty$ (in which case we say that the sum does not converge).

Case II - General $f : \Omega \to \mathbb{R}$: The above argument fails if f is allowed to take both positive and negative values (why?). In fact, the answers L and L' from different bijections may be completely different. An example is given later to illustrate this point. For now, here is how we deal with this problem.

For a real number x we introduce the notations, $x_+ = x \vee 0$ and $x_- = (-x) \vee 0$. Then $x = x_+ - x_-$ while $|x| = x_+ + x_-$. Define the non-negative functions $f_+, f_- : \Omega \to \mathbb{R}_+$ by $f_+(\omega) = (f(\omega))_+$ and $f_-(\omega) = (f(\omega))_-$. Observe that $f_+(\omega) - f_-(\omega) = f(\omega)$ while $f_+(\omega) + f_-(\omega) = |f(\omega)|$, for all $\omega \in \Omega$.

Example 40. Let $\Omega = \{a, b, c, d\}$ and let $f(a) = 1$, $f(b) = -1$, $f(c) = -3$, $f(4) = -0.3$. Then, $f_+(a) = 1$ and $f_+(b) = f_+(c) = f_+(d) = 0$ while $f_-(1) = 0$ and $f_-(b) = 1$, $f_-(c) = 3$, $f_-(d) = 0.3$.

Since f_+ and f_- are non-negative functions, we know how to define their sums. Let $S_+ = \sum_\omega f_+(\omega)$ and $S_- = \sum_\omega f_-(\omega)$. Recall that one or both of S_+, S_- could be equal to $+\infty$, in which case we say that $\sum_\omega f(\omega)$ *does not converge absolutely* and do not assign it any value. If both S_+ and S_- are finite, then we define $\sum_\omega f(\omega) = S_+ - S_-$. In this case we say that $\sum f$ *converges absolutely*.

This completes our definition of absolutely convergent sums. A few exercises to show that when working with absolutely convergent sums, the usual rules of addition remain valid. For example, we can add the numbers in any order.

Exercise 41. Show that $\sum_\omega f(\omega)$ converges absolutely if and only if $\sum_\omega |f(\omega)|$ is finite (since $|f(\omega)|$ is a non-negative function, this latter sum is always defined, and may equal $+\infty$).

For non-negative f, we can find the sum by using any particular bijection and then taking limits of partial sums. What about general f?

Exercise 42. Let $f : \Omega \to \mathbb{R}$. Suppose $\sum_{\omega\in\Omega} f(\omega)$ be summable and let the sum be S. Then, for any bijection $\varphi : \mathbb{N} \to \Omega$, we have $\lim_{n\to\infty}(f(\varphi(1)) + \ldots + f(\varphi(n))) = S$.

Conversely, if $\lim_{n\to\infty}(f(\varphi(1)) + \ldots + f(\varphi(n)))$ exists and is the same finite number for any bijection $\varphi : \mathbb{N} \to \mathbb{R}$, then f must be absolutely summable and $\sum_{\omega\in\Omega} f(\omega)$ is equal to this common limit.

The usual properties of summation without which life would not be worth living, are still valid.

Exercise 43. Let $f, g : \Omega \to \mathbb{R}_+$ and $a, b \in \mathbb{R}$. If $\sum f$ and $\sum g$ converge absolutely, then $\sum(af + bg)$ converges absolutely and $\sum(af + bg) = a\sum f + b\sum g$. Further, if $f(\omega) \le g(\omega)$ for all $\omega \in \Omega$, then $\sum f \le \sum g$.

Example 44. This example will illustrate why we refuse to assign a value to $\sum_\omega f(\omega)$ in some cases. Let $\Omega = \mathbb{Z}$ and define $f(0) = 0$ and $f(n) = 1/n$ for $n \neq 0$. At first one may like to say that $\sum_{n\in\mathbb{Z}} f(n) = 0$, since we can cancel $f(n)$ and $f(-n)$ for each n. However, following our definitions

$$f_+(n) = \begin{cases} \frac{1}{n} & \text{if } n \ge 1 \\ 0 & \text{if } n \le 0, \end{cases} \qquad f_-(n) = \begin{cases} \frac{1}{n} & \text{if } n \le -1 \\ 0 & \text{if } n \ge 0. \end{cases}$$

Hence S_+ and S_- are both $+\infty$ which means our definition does not assign any value to the sum $\sum_\omega f(\omega)$.

Indeed, by ordering the numbers appropriately, we can get any value we like! For example, here is how to get 10. We know that $1 + \frac{1}{2} + \ldots + \frac{1}{n}$ grows without bound. Just keep adding these positive number till the sum exceeds 10 for the first time. Then start adding the negative numbers $-1 - \frac{1}{2} - \ldots - \frac{1}{m}$ till the sum comes below 10. Then add the positive numbers $\frac{1}{n+1} + \frac{1}{n+2} + \ldots + \frac{1}{n'}$ till the sum exceeds 10 again, and then negative numbers till the sum falls below 10 again, etc. Using the fact that the individual terms in the series are going to zero, it is easy to see that the partial sums then converge to 10. There is nothing special about 10, we can get any number we want!

One last remark on why we assumed Ω to be countable.

Remark 45. What if Ω is uncountable? Take any $f : \Omega \to \mathbb{R}_+$. Define the sets $A_n = \{\omega : f(\omega) \ge 1/n\}$. For some n, if A_n has infinitely many elements, then clearly the only reasonable value that we can assign to $\sum f(\omega)$ is $+\infty$ (since the sum over elements of A_n

itself is larger than any finite number). Therefore, for $\sum f(\omega)$ to be a finite number it is essential that A_n is a finite set for each set.

Now, a countable union of finite sets is countable (or finite). Therefore $A = \bigcup_n A_n$ is a countable set. But note that A is also the set $\{\omega : f(\omega) > 0\}$ (since, if $f(\omega) > 0$ it must belong to some A_n). Consequently, even if the underlying set Ω is uncountable, our function will have to be equal to zero except on a countable subset of Ω. In other words, we are reduced to the case of countable sums!

5.2. Second approach. In the first approach, we assumed that you are already familiar with the notion of limits and series and used them to define countable sums. In the second approach, we start from scratch and define infinite sums. The end result is exactly the same. For the purposes of this course, you may ignore the rest of the section.

Definition 46. If Ω is a countable set and $f : \Omega \to \mathbb{R}_+$ is a non-negative function, then we define

$$\sum_{\omega} f(\omega) := \sup \left\{ \sum_{\omega \in A} f(\omega) : A \subseteq \Omega \text{ is finite} \right\}$$

where the supremum takes values in $\overline{\mathbb{R}}_+ = \mathbb{R}_+ \cup \{+\infty\}$. We say that $\sum f(\omega)$ converges if the supremum has a finite value.

Exercise 47. Show that if $f, g : \Omega \to \mathbb{R}_+$ and $a, b \in \mathbb{R}_+$, then $\sum(af + bg) = a\sum f + b\sum g$. Further, if $f(\omega) \leq g(\omega)$ for all $\omega \in \Omega$, then $\sum f \leq \sum g$.

Next, we would like to remove the condition of non-negativity. For a real number x we write $x_+ = x \vee 0$ and $x_- = (-x) \vee 0$. Then $x = x_+ - x_-$ while $|x| = x_+ + x_-$.

Definition 48. Now suppose $f : \Omega \to \mathbb{R}$ takes both positive and negative values. Then we first define the non-negative functions $f_+, f_- : \Omega \to \mathbb{R}_+$ by $f_+(\omega) = (f(\omega))_+$ and $f_-(\omega) = (f(\omega))_-$ and set $S_+ = \sum_\omega f_+(\omega)$ and $S_- = \sum_\omega f_-(\omega)$. If both S_+ and S_- are finite, then we define $\sum_\omega f(\omega) = S_+ - S_-$.

Remark 49. The condition that S_+ and S_- are both finite is the same as the condition that $\sum_\omega |f(\omega)|$ is finite. If these happen, we say that the sum $\sum f(\omega)$ *converges absolutely*.

Remark 50. Sometimes it is convenient to set $\sum f(\omega)$ to $+\infty$ if $S_+ = \infty$ and $S_- < \infty$ and set $\sum f(\omega)$ to $-\infty$ if $S_+ < \infty$ and $S_- = \infty$. But there is no reasonable value to assign if both the sums are infinite.

Exercise 51. Show that the two approaches give the same answers.

6. Basic rules of probability

So far we have defined the notion of probability space and probability of an event. But most often, we do not calculate probabilities from the definition. This is like in integration, where one defined the integral of a function as a limit of Riemann sums, but that definition is used only to find integrals of x^n, $\sin(x)$ and a few such functions. Instead, integrals of complicated expressions such as $x\sin(x) + 2\cos^2(x)\tan(x)$ are calculated by various rules, such as substitution rule, integration by parts etc. In probability we need some similar rules relating probabilities of various combinations of events to the individual probabilities.

Proposition 52. *Let $(\Omega, p.)$ be a discrete probability space.*

(1) For any event A, we have $0 \leq \mathbf{P}(A) \leq 1$. Also, $\mathbf{P}(\emptyset) = 0$ and $\mathbf{P}(\Omega) = 1$.

(2) Finite additivity of probability: *If $A_1,\ldots,A_n$ are pairwise disjoint events, then $\mathbf{P}(A_1\cup\ldots\cup A_n)=\mathbf{P}(A_1)+\ldots+\mathbf{P}(A_n)$. In particular, $\mathbf{P}(A^c)=1-\mathbf{P}(A)$ for any event A.*

(3) Countable additivity of probability: *If $A_1,A_2,\ldots$ is a countable collection of pairwise disjoint events, then $\mathbf{P}(\cup A_i)=\sum_i \mathbf{P}(A_i)$.*

All of these may seem obvious, and indeed they would be totally obvious if we stuck to finite sample spaces. But the sample space could be countable, and then probability of events may involve infinite sums which need special care in manipulation. Therefore we must give a proof. In writing a proof, and in many future contexts, it is useful to introduce the following notation.

Notation: Let $A\subseteq\Omega$ be an event. Then, we define a function $\mathbf{1}_A:\Omega\to\mathbb{R}$, called the *indicator function of A*, as follows.

$$\mathbf{1}_A(\omega)=\begin{cases}1 & \text{if } \omega\in A,\\ 0 & \text{if } \omega\notin A.\end{cases}$$

Since a function from Ω to $\mathbb{R}$ is called a random variable, the indicator of any event is a random variable. All information about the event A is in its indicator function (meaning, if we know the value of $\mathbf{1}_A(\omega)$, we know whether or not ω belongs to A). For example, we can write $\mathbf{P}(A)=\sum_{\omega\in\Omega}\mathbf{1}_A(\omega)p_\omega$.

Now we prove the proposition.

PROOF. (1) By definition of probability space $\mathbf{P}(\Omega)=1$ and $\mathbf{P}(\emptyset)=0$. If A is any event, then $\mathbf{1}_\emptyset(\omega)p_\omega\le\mathbf{1}_A(\omega)p_\omega\le\mathbf{1}_\Omega(\omega)p_\omega$. By Exercise 43, we get

$$\sum_{\omega\in\Omega}\mathbf{1}_\emptyset(\omega)p_\omega\le\sum_{\omega\in\Omega}\mathbf{1}_A(\omega)p_\omega\le\sum_{\omega\in\Omega}\mathbf{1}_\Omega(\omega)p_\omega.$$

As observed earlier, these sums are just $\mathbf{P}(\emptyset)$, $\mathbf{P}(A)$ and $\mathbf{P}(\Omega)$, respectively. Thus, $0\le\mathbf{P}(A)\le 1$.

(2) It suffices to prove it for two sets (why?). Let A,B be two events such that $A\cap B=\emptyset$. Let $f(\omega)=p_\omega\mathbf{1}_A(\omega)$ and $g(\omega)=p_\omega\mathbf{1}_B(\omega)$ and $h(\omega)=p_\omega\mathbf{1}_{A\cup B}(\omega)$. Then, the disjointness of A and B implies that $f(\omega)+g(\omega)=h(\omega)$ for all $\omega\in\Omega$. Thus, by Exercise 43, we get

$$\sum_{\omega\in\Omega}f(\omega)+\sum_{\omega\in\Omega}g(\omega)=\sum_{\omega\in\Omega}h(\omega).$$

But the three sums here are precisely $\mathbf{P}(A)$, $\mathbf{P}(B)$ and $\mathbf{P}(A\cup B)$. Thus, we get $\mathbf{P}(A\cup B)=\mathbf{P}(A)+\mathbf{P}(B)$.

(3) This is similar to finite additivity but needs a more involved argument. We leave it as an exercise for the interested reader. ■

Exercise 53. Adapt the proof to prove that for a countable family of events A_k in a common probability space (no disjointness assumed), we have

$$\mathbf{P}(\cup_k A_k)\le\sum_k\mathbf{P}(A_k).$$

Definition 54 (Limsup and liminf of sets)**.** If A_k, $k\ge 1$, is a sequence of subsets of Ω, we define

$$\limsup A_k=\bigcap_{N=1}^{\infty}\bigcup_{k=N}^{\infty}A_k,\qquad\text{and}\qquad\liminf A_k=\bigcup_{N=1}^{\infty}\bigcap_{k=N}^{\infty}A_k.$$

In words, $\limsup A_k$ is the set of all ω that belong to infinitely many of the A_ks, and $\liminf A_k$ is the set of all ω that belong to all but finitely many of the A_ks.

Two special cases are of increasing and decreasing sequences of events. This means $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ and $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$. In these cases, the limsup and liminf are the same (so we refer to it as the limit of the sequence of sets). It is $\cup_k A_k$ in the case of increasing events and $\cap_k A_k$ in the case of decreasing events.

Exercise 55. Events below are all contained in a discrete probability space. Use countable additivity of probability to show that

(1) If A_k are increasing events with limit A, show that $\mathbf{P}(A)$ is the increasing limit of $\mathbf{P}(A_k)$.
(2) If A_k are decreasing events with limit A, show that $\mathbf{P}(A)$ is the decreasing limit of $\mathbf{P}(A_k)$.

Now we re-write the basic rules of probability as follows.

The basic rules of probability:

(1) $\mathbf{P}(\emptyset) = 0, \mathbf{P}(\Omega) = 1$ and $0 \leq \mathbf{P}(A) \leq 1$ for any event A.
(2) $\mathbf{P}\left(\bigcup_k A_k\right) \leq \sum_k \mathbf{P}(A_k)$ for any countable collection of events A_k.
(3) $\mathbf{P}\left(\bigcup_k A_k\right) = \sum_k \mathbf{P}(A_k)$ if A_k is a countable collection of pairwise disjoint events.

7. Inclusion-exclusion formula

In general, there is no simple rule for $\mathbf{P}(A \cup B)$ in terms of $\mathbf{P}(A)$ and $\mathbf{P}(B)$. Indeed, consider the probability space $\Omega = \{0,1\}$ with $p_0 = p_1 = \frac{1}{2}$. If $A = \{0\}$ and $B = \{1\}$, then $\mathbf{P}(A) = \mathbf{P}(B) = \frac{1}{2}$ and $\mathbf{P}(A \cup B) = 1$. However, if $A = B = \{0\}$, then $\mathbf{P}(A) = \mathbf{P}(B) = \frac{1}{2}$ as before, but $\mathbf{P}(A \cup B) = \frac{1}{2}$. This shows that $\mathbf{P}(A \cup B)$ cannot be determined from $\mathbf{P}(A)$ and $\mathbf{P}(B)$. Similarly for $\mathbf{P}(A \cap B)$ or other set constructions.

However, it is easy to see that $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$. This formula is not entirely useless, because in special situations we shall later see that the probability of the intersection is easy to compute and hence we may compute the probability of the union. Generalizing this idea to more than two sets, we get the following surprisingly useful formula.

Proposition 56 (Inclusion-Exclusion formula)**.** *Let (Ω, p) be a probability space and let $A_1, \ldots, A_n$ be events. Then,*

$$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) = S_1 - S_2 + S_3 - \ldots + (-1)^{n-1} S_n$$

where

$$S_k = \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} \mathbf{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}).$$

We give two proofs, but the difference is only superficial. It is a good exercise to reason out why the two arguments are basically the same.

FIRST PROOF. For each $\omega \in \Omega$ we compute its contribution to the two sides. If $\omega \notin \bigcup_{i=1}^n A_i$, then p_ω is not counted on either side. Suppose $\omega \in \bigcup_{i=1}^n A_i$ so that p_ω is counted

once on the left side. We count the number of times p_ω is counted on the right side by splitting into cases depending on the exact number of A_is that contain ω.

Suppose ω belongs to exactly one of the A_is. For simplicity let us suppose that $\omega \in A_1$ but $\omega \in A_i^c$ for $2 \le i \le n$. Then p_ω is counted once in S_1 but not counted in $S_2, \ldots, S_n$.

Suppose ω belongs to A_1 and A_2 but not any other A_i. Then p_ω is counted twice in S_1 (once for $\mathbf{P}(A_1)$ and once for $\mathbf{P}(A_2)$) and subtracted once in S_2 (in $\mathbf{P}(A_1 \cap A_2)$). Thus, it is effectively counted once on the right side. The same holds if ω belongs to A_i and A_j but not any other A_ks.

If ω belongs to $A_1, \ldots, A_k$ but not any other A_i, then on the right side, p_ω is added k times in S_1, subtracted $\binom{k}{2}$ times in S_2, added $\binom{k}{3}$ times in S_k and so on. Thus p_ω is effectively counted

$$\binom{k}{1} - \binom{k}{2} + \binom{k}{3} - \ldots + (-1)^{k-1}\binom{k}{k}$$

times. By the Binomial formula, this is just the expansion of $1-(1-1)^k$ which is 1. ■

SECOND PROOF. Use the definition to write both sides of the statement. Let $A = \cup_{i=1}^n A_i$.

$$\text{LHS} = \sum_{\omega \in A} p_\omega = \sum_{\omega \in \Omega} \mathbf{1}_A(\omega) p_\omega.$$

Now we compute the right side. For any $i_1 < i_2 < \ldots < i_k$, we write

$$\mathbf{P}\left(A_{i_1} \cap \ldots \cap A_{i_k}\right) = \sum_{\omega \in \Omega} p_\omega \mathbf{1}_{A_{i_1} \cap \ldots \cap A_{i_k}}(\omega) = \sum_{\omega \in \Omega} p_\omega \prod_{\ell=1}^k \mathbf{1}_{A_{i_\ell}}(\omega).$$

Hence, the right hand side is given by adding over $i_1 < \ldots < i_k$, multiplying by $(-1)^{k-1}$ and then summing over k from 1 to n.

$$\begin{aligned}
\text{RHS} &= \sum_{k=1}^n (-1)^{k-1} \sum_{1 \le i_1 < \ldots < i_k \le n} \sum_{\omega \in \Omega} p_\omega \prod_{\ell=1}^k \mathbf{1}_{A_{i_\ell}}(\omega) \\
&= \sum_{\omega \in \Omega} \sum_{k=1}^n (-1)^{k-1} \sum_{1 \le i_1 < \ldots < i_k \le n} p_\omega \prod_{\ell=1}^k \mathbf{1}_{A_{i_\ell}}(\omega) \\
&= -\sum_{\omega \in \Omega} p_\omega \sum_{k=1}^n \sum_{1 \le i_1 < \ldots < i_k \le n} \prod_{\ell=1}^k (-\mathbf{1}_{A_{i_\ell}}(\omega)) \\
&= -\sum_{\omega \in \Omega} p_\omega \left(\prod_{j=1}^n (1 - \mathbf{1}_{A_j}(\omega)) \; - \; 1 \right) \\
&= \sum_{\omega \in \Omega} p_\omega \mathbf{1}_A(\omega).
\end{aligned}$$

because the quantity $\prod_{j=1}^n (1 - \mathbf{1}_{A_j}(\omega))$ equals -1 if ω belongs to at least one of the A_is, and is zero otherwise. Thus the claim follows. ■

As we remarked earlier, it turns out that in many settings it is possible to compute the probabilities of intersections. We give an example now.

Example 57. Let $\Omega = S_{52} \times S_{52}$ with $p_\omega = \frac{1}{(52!)^2}$ for all $\omega \in \Omega$. Consider the event $A = \{(\pi, \sigma) : \pi(i) \neq \sigma(i) \ \forall i\}$. Informally, we imagine two shuffled decks of cards kept side by side (or perhaps one shuffled deck and another permutation denoting a "psychic's

predictions" for the order in which the cards occur). Then A is the event that there are no matches (or correct guesses).

Let $A_i = \{(\pi,\sigma) : \pi(i) = \sigma(i)\}$ so that $A^c = A_1 \cup \ldots \cup A_{52}$. It is easy to see that $\mathbf{P}(A_{i_1} \cap A_{i_2} \ldots \cap A_{i_k}) = \frac{1}{52(52-1)\ldots(52-k+1)}$ for any $i_1 < i_2 < \ldots < i_k$ (why?). Therefore, by the inclusion-exclusion formula, we get

$$\begin{aligned}\mathbf{P}(A^c) &= \binom{52}{1}\frac{1}{52} - \binom{52}{2}\frac{1}{(52)(51)} + \ldots + (-1)^{51}\binom{52}{52}\frac{1}{(52)(51)\ldots(1)} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \ldots - \frac{1}{52!} \\ &\approx 1 - \frac{1}{e} \approx 0.6321\end{aligned}$$

by the expansion $e^{-1} = 1 - \frac{1}{2!} + \frac{1}{3!} - \ldots$. Hence $\mathbf{P}(A) \approx e^{-1} \approx 0.3679$.

Example 58. Place n distinguishable balls in r distinguishable urns at random. Let A be the event that some urn is empty. The probability space is $\Omega = \{\underline{\omega} = (\omega_1,\ldots,\omega_n) : 1 \le \omega_i \le r\}$ with $p_{\underline{\omega}} = r^{-n}$. Let $A_\ell = \{\underline{\omega} : \omega_i \neq \ell\}$ for $\ell = 1,2\ldots,r$. Then, $A = A_1 \cup \ldots \cup A_{r-1}$ (as A_r is empty, we could include it or not, makes no difference).

It is easy to see that $\mathbf{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = (r-k)^n r^{-n} = (1-\frac{k}{r})^n$. We could use the inclusion-exclusion formula to write the expression

$$\mathbf{P}(A) = r\left(1-\frac{1}{r}\right)^n - \binom{r}{2}\left(1-\frac{2}{r}\right)^n + \ldots + (-1)^{r-2}\binom{r}{r-1}\left(1-\frac{r-1}{r}\right)^n.$$

The last term is zero (since all urns cannot be empty). I don't know if this expression can be simplified any more.

We mention two useful formulas that can be proved on lines similar to the inclusion-exclusion principle. If we say "at least one of the events $A_1, A_2, \ldots, A_n$ occurs", we are talking about the union, $A_1 \cup A_2 \cup \ldots \cup A_n$. What about "at least m of the events $A_1, A_2, \ldots, A_n$ occur", how to express it with set operations. It is not hard to see that this set is precisely

$$B_m = \bigcup_{1 \le i_1 < i_2 < \ldots < i_m \le n} (A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_m}).$$

The event that "exactly m of the events $A_1, A_2, \ldots, A_n$ occur" can be written as

$$C_m = B_m \setminus B_{m+1} = \bigcup_{\substack{S \subseteq [n] \\ |S| = m}} \left(\bigcap_{i \in S} A_i\right) \bigcap \left(\bigcap_{i \notin S} A_i^c\right).$$

Exercise 59. Let $A_1,\ldots,A_n$ be events in a probability space (Ω,p) and let $m\leq n$. Let B_m and C_m be as above. Show that

$$\begin{aligned}\mathbf{P}(B_m) &= \sum_{k=m}^{n}(-1)^{k-m}\binom{k-1}{k-m}S_k\\ &= S_m-\binom{m}{1}S_{m+1}+\binom{m+1}{2}S_{m+2}-\binom{m+2}{3}S_{m+3}+\ldots\\ \mathbf{P}(C_m) &= \sum_{k=m}^{n}(-1)^{k-m}\binom{k}{m}S_k\\ &= S_m-\binom{m+1}{1}S_{m+1}+\binom{m+2}{2}S_{m+2}-\binom{m+3}{3}S_{m+3}+\ldots\end{aligned}$$

Exercise 60. Return to the setting of exercise 57 but with n cards in a deck, so that $\Omega = S_n\times S_n$ and $p_{(\pi,\sigma)}=\frac{1}{(n!)^2}$. Let A_m be the event that there are exactly m matches between the two decks.

(1) For fixed $m\geq 0$, show that $\mathbf{P}(A_m)\to e^{-1}\frac{1}{m!}$ as $n\to\infty$.
(2) Assume that the approximations above are valid for $n=52$ and $m\leq 10$. Find the probability that there are at least 10 matches.

8. Bonferroni's inequalities

Inclusion-exclusion formula is nice when we can calculate the probabilities of intersections of the events under consideration. Things are not always this nice, and sometimes that may be very difficult. Even if we could find them, summing them with signs according to the inclusion-exclusion formula may be difficult as the example 58 demonstrates. The *idea* behind the inclusion-exclusion formula can however be often used to compute *approximate values of probabilities*, which is very valuable in most applications. That is what we do next.

We know that $\mathbf{P}(A_1\cup\ldots\cup A_n)\leq\mathbf{P}(A_1)+\ldots+\mathbf{P}(A_n)$ for any events $A_1,\ldots,A_n$. This is an extremely useful inequality, often called the *union bound*. Its usefulness is in the fact that there is no assumption made about the events A_is (such as whether they are disjoint or not). The following inequalities generalize the union bound, and gives both upper and lower bounds for the probability of the union of a bunch of events.

Lemma 61 (Bonferroni's inequalities)**.** *Let $A_1,\ldots,A_n$ be events in a probability space (Ω,p) and let $A=A_1\cup\ldots\cup A_n$. We have the following upper and lower bounds for $\mathbf{P}(A)$.*

$$\mathbf{P}(A)\leq\sum_{k=1}^{m}(-1)^{k-1}S_k,\quad \textit{for any odd } m.$$
$$\mathbf{P}(A)\geq\sum_{k=1}^{m}(-1)^{k-1}S_k,\quad \textit{for any even } m.$$

PROOF. We shall write out the proof for the cases $m=1$ and $m=2$. When $m=1$, the inequality is just the union bound

$$\mathbf{P}(A)\leq\mathbf{P}(A_1)+\ldots+\mathbf{P}(A_n)$$

which we know. When $m=2$, the inequality to be proved is

$$\mathbf{P}(A)\geq\sum_k\mathbf{P}(A_k)-\sum_{k<\ell}\mathbf{P}(A_k\cap A_\ell)$$

To see this, fix $\omega \in \Omega$ and count the contribution of p_ω to both sides. Like in the proof of the inclusion-exclusion formula, for $\omega \notin A_1 \cup \ldots \cup A_n$, the contribution to both sides is zero. On the other hand, if ω belongs to exactly r of the sets for some $r \geq 1$, then it is counted once on the left side and $r - \binom{r}{2}$ times on the right side. Not that $r - \binom{r}{2} = \frac{1}{2}r(3-r)$ which is always non-positive (one if $r = 1$, zero if $r = 2$ and non-positive if $r \geq 3$). Hence we get LHS $\geq$ RHS.

Similarly, one can prove the other inequalities in the series. We leave it as an exercise. The key point is that $r - \binom{r}{2} + \ldots + (-1)^{k-1}\binom{r}{k}$ is non-negative if k is odd and non-positive if k is even (prove this). Here as always $\binom{x}{y}$ is interpreted as zero if $y > x$. ■

Here is an application of these inequalities.

Example 62. Return to Example 58. We obtained an exact expression for the answer, but that is rather complicated. For example, what is the probability of having at least one empty urn when $n = 40$ balls are placed at random in $r = 10$ urns? It would be complicated to sum the series. Instead, we could use Bonferroni's inequalities to get the following bounds.

$$r\left(1-\frac{1}{r}\right)^n - \binom{r}{2}\left(1-\frac{2}{r}\right)^n \leq \mathbf{P}(A) \leq r\left(1-\frac{1}{r}\right)^n.$$

If we take $n = 40$ and $r = 10$, the bounds we get are $0.1418 \leq \mathbf{P}(A) \leq 0.1478$. Thus, we get a pretty decent approximation to the probability. By experimenting with other numbers you can check that the approximations are good when n is large compared to r but not otherwise. Can you reason why?

9. Independence - a first look

We remarked in the context of inclusion-exclusion formulas that often the probabilities of intersections of events is easy to find, and then we can use them to find probabilities of unions etc. In many contexts, this is related to one of the most important notions in probability.

Definition 63. Let A, B be events in a common probability space. We say that A and B are *independent* is $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$.

Example 64. Toss a fair coin n times. Then $\Omega = \{\underline{\omega} : \underline{\omega} = (\omega_1, \ldots, \omega_n),\ \omega_i \text{ is 0 or 1}\}$ and $p_{\underline{\omega}} = 2^{-n}$ for each $\underline{\omega}$. Let $A = \{\underline{\omega} : \omega_1 = 0\}$ and let $B = \{\underline{\omega} : \omega_2 = 0\}$. Then, from the definition of probabilities, we can see that $\mathbf{P}(A) = 1/2$, $\mathbf{P}(B) = 1/2$ (because the elementary probabilities are equal, and both the sets A and B contain exactly 2^{n-1} elements). Further, $A \cap B = \{\underline{\omega} : \omega_1 = 1, \omega_2 = 0\}$ has 2^{n-2} elements, whence $\mathbf{P}(A \cap B) = 1/4$. Thus, $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$ and hence A and B are independent.

If two events are independent, then the probability of their intersection can be found from the individual probabilities. How do we check if two events are independent? By checking if the probability of the event is equal to the product of the individual probabilities! It seems totally circular and useless! There are many reasons why it is not an empty notion as we shall see.

Firstly, in physical situationsdependence is related to a basic intuition we have about whether two events are related or not. For example, suppose you are thinking of betting Rs.1000 on a particular horse in a race. If you get the news that your cousin is getting married, it will perhaps not affect the amount you plan to bet. However, if you get the news that one of the other horses has been injected with undetectable drugs, it might affect the bet you want to place. In other words, certain events (like marriage of a cousin) have

no bearing on the probability of the event of interest (the event that our horse wins) while other events (like the injection of drugs) do have an impact. This intuition is often put into the very definition of probability space that we have.

For example, in the above example of tossing a fair coin n times, it is our intuition that a coin does not remember how it fell previous times, and that chance of its falling head in any toss is just $1/2$, irrespective of how many heads or tails occured before[8] And this intuition was used in defining the elementary probabilities as 2^{-n} each. Since we started with the intuitive notion of independence, and put that into the definition of the probability space, it is quite expected that the event that the first toss is a head should be independent of the event that the second toss is a tail. That is the calculation shown in above.

But how is independence useful mathematically if the conditions to check independence are the very conclusions we want?! The answer to this lies in the following fact (to be explained later). When certain events are independent, then many other collections of events that can be made out of them also turn out to be independent. For example, if A, B, C, D are independent (we have not yet defined what this means!), then $A \cup B$ and $C \cup D$ are also independent. Thus, starting from independence of certain events, we get independence of many other events. For example, any event depending on the first four tosses is independent of eny event depending on the next five tosses.

[8]It may be better to attribute this to experience rather than intuition. There have been reasonable people in history who believed that if a coin shows heads in ten tosses in a row, then on the next toss it is more likely to show tails (to 'compensate' for the overabundance of heads)! Clearly this is also someone's intuition, and different from ours. Only experiment can decide which is correct, and any number of experiments with real coins show that our intuition is correct, and coins have no memory.