3. Examples of discrete probability spaces

Example 17. Toss n coins. We saw this before, but assumed that the coins are fair. Now we do not. The sample space is

$$\Omega = \{0,1\}^n = \{\underline{\omega} = (\omega_1, \ldots, \omega_n) : \omega_i = 0 \text{ or } 1 \text{ for each } i \le n\}.$$

Further we assign $p_{\underline{\omega}} = \alpha_{\omega_1}^{(1)} \ldots \alpha_{\omega_n}^{(n)}$. Here $\alpha_0^{(j)}$ and $\alpha_1^{(j)}$ are supposed to indicate the probabilities that the jth coin falls tails up or heads up, respectively. Why did we take the product of $\alpha^{(j)}$s and not some other combination? This is a non-mathematical question about what model is suited for the given real-life example. For now, the only justification is that empirically the above model seems to capture the real life situation accurately.

In particular, if the n coins are identical, we may write $p = \alpha_1^{(j)}$ (for any j) and the elementary probabilities become $p_{\underline{\omega}} = p^{\sum_i \omega_i} q^{n - \sum_i \omega_i}$ where $q = 1 - p$.

Fix $0 \le k \le n$ and let $B_k = \{\underline{\omega} : \sum_{i=1}^n \omega_i = k\}$ be the event that we see exactly k heads out of n tosses. Then $\mathbf{P}(B_k) = \binom{n}{k} p^k q^{n-k}$. If A_k is the event that there are at least k heads, then $\mathbf{P}(A_k) = \sum_{\ell=k}^n \binom{n}{\ell} p^\ell q^{n-\ell}$.

Example 18. Toss a coin n times. Again

$$\Omega = \{0,1\}^n = \{\underline{\omega} = (\omega_1, \ldots, \omega_n) : \omega_i = 0 \text{ or } 1 \text{ for each } i \le n\},$$
$$p_{\underline{\omega}} = p^{\sum_i \omega_i} q^{n - \sum_i \omega_i}.$$

This is the same probability space that we got for the tossing of n identical looking coins. Implicit is the assumption that once a coin is tossed, for the next toss it is as good as a different coin but with the same p. It is possible to imagine a world where coins retain the memory of what happened before (or as explained before, we can make a "coin" that remembers previous tosses!), in which case this would not be a good model for the given situation. We don't believe that this is the case for coins in our world, and this can be verified empirically.

Example 19. Shuffle a deck of 52 cards. $\Omega = S_{52}$, the set of all permutations[3] of $[52]$ and $p_\pi = \frac{1}{52!}$ for each $\pi \in S_{52}$.

Example 20. "Psychic" guesses a deck of cards. The sample space is $\Omega = S_{52} \times S_{52}$ and $p_{(\pi,\sigma)} = 1/(52!)^2$ for each pair (π, σ) of permutations. In a pair (π, σ), the permutation π denotes the actual order of cards in the shuffled deck, and σ denotes the order guessed by the psychic. If the guesses are purely random, then the probabilities are as we have written.

An interesting random variable is the number of correct guesses. This is the function $X : \Omega \to \mathbb{R}$ defined by $X(\pi, \sigma) = \sum_{i=1}^{52} \mathbf{1}_{\pi_i = \sigma_i}$. Correspondingly we have the events $A_k = \{(\pi, \sigma) : X(\pi, \sigma) \ge k\}$.

---

[3]We use the notation $[n]$ to denote the set $\{1, 2, \ldots, n\}$. A permutation of $[n]$ is a vector $(i_1, i_2, \ldots, i_n)$ where $i_1, \ldots, i_n$ are distinct elements of $[n]$, in other words, they are $1, 2, \ldots, n$ but in some order. Mathematically, we may define a permutation as a bijection $\pi : [n] \to [n]$. Indeed, for a bijection π, the numbers $\pi(1), \ldots, \pi(n)$ are just $1, 2, \ldots, n$ in some order.

Example 21. Toss a coin till a head turns up. $\Omega = \{1, 01, 001, 0001, \ldots\} \cup \{\overline{0}\}$. Let us write $0^k 1 = 0 \ldots 01$ as a short form for k zeros (tails) followed by 1 and $\overline{0}$ stands for the sequence of all tails. Let $p \in [0, 1]$. Then, we set $p_{0^k 1} = q^k p$ for each $k \in \mathbb{N}$. We also set $p_{\overline{0}} = 0$ if $p > 0$ and $p_{\overline{0}} = 1$ if $p = 0$. This is forced on us by the requirement that elementary probabilities add to 1.

Let $A = \{0^k 1 : k \geq n\}$ be the event that at least n tails fall before a head turns up. Then $\mathbf{P}(A) = q^n p + q^{n+1} p + \ldots = q^n$.

Example 22. Place r distinguishable balls in m distinguishable urns at random. We saw this before (the words "labelled" and "distinguishable" mean the same thing here). The sample space is $\Omega = [m]^r = \{\underline{\omega} = (\omega_1, \ldots, \omega_r) : 1 \leq \omega_i \leq m\}$ and $p_{\underline{\omega}} = m^{-r}$ for every $\underline{\omega} \in \Omega$. Here ω_i indicates the urn number into which the ith ball goes.

Example 23. Place r indistinguishable balls in m distinguishable urns at random. Since the balls are indistinguishable, we can only count the number of balls in each urn. The sample space is

$$\Omega = \{(\ell_1, \ldots, \ell_m) : \ell_i \geq 0, \ \ell_1 + \ldots + \ell_m = r\}.$$

We give two proposals for the elementary probabilities.

(1) Let $p^{\text{MB}}_{(\ell_1, \ldots, \ell_m)} = \frac{m!}{\ell_1! \ell_2! \ldots \ell_m!} \frac{1}{m^r}$. These are the probabilities that result if we place r labelled balls in m labelled urns, and then erase the labels on the balls.

(2) Let $p^{\text{BE}}_{(\ell_1, \ldots, \ell_m)} = \frac{1}{\binom{m+r-1}{r-1}}$ for each $(\ell_1, \ldots, \ell_m) \in \Omega$. Elementary probabilities are chosen so that all distinguishable configurations are equally likely.

That these are legitimate probability spaces depend on two combinatorial facts.

Exercise 24. (1) Let $(\ell_1, \ldots, \ell_m) \in \Omega$. Show that $\#\{\underline{\omega} \in [m]^r : \sum_{j=1}^r \mathbf{1}_{\omega_j = i} = \ell_i$ for each $i \in [m]\} = \frac{n!}{\ell_1! \ell_2! \ldots \ell_m!}$. Hence or directly, show that $\sum_{\omega \in \Omega} p^{\text{MB}}_\omega = 1$.

(2) Show that $\#\Omega = \binom{m+r-1}{r-1}$. Hence, $\sum_{\omega \in \Omega} p^{\text{BE}}_\omega = 1$.

The two models are clearly different. Which one captures reality? We can arbitrarily label the balls for our convenience, and then erase the labels in the end. This clearly yields elementary probabilities p^{MB}. Or to put it another way, pick the balls one by one and assign them randomly to one of the urns. This suggests that p^{MB} is the "right one".

This leaves open the question of whether there is a natural mechanism of assigning balls to urns so that the probabilities p^{BE} shows up. No such mechanism has been found. But this probability space does occur in the physical world. If r photons ("indistinguishable balls") are to occupy m energy levels ("urns"), then empirically it has been verified that the correct probability space is the second one![4]

Example 25. Sampling with replacement from a population. Define $\Omega = \{\underline{\omega} \in [N]^k : \omega_i \in [N]$ for $1 \leq i \leq k\}$ with $p_{\underline{\omega}} = 1/N^k$ for each $\underline{\omega} \in \Omega$. Here $[N]$ is the population (so the size

---

[4]The probabilities p^{MB} and p^{BE} are called Maxwell-Boltzmann statistics and Bose-Einstein statistics. There is a third kind, called Fermi-Dirac statistics which is obeyed by electrons. For general $m \geq r$, the sample space is $\Omega_{\text{FD}} = \{(\ell_1, \ldots, \ell_m) : \ell_i = 0 \text{ or } 1 \text{ and } \ell_1 + \ldots + \ell_m = r\}$ with equal probabilities for each element. In words, all distinguishable configurations are equally likely, with the added constraint that at most one electron can occupy each energy level.

of the population is N) and the size of the sample is k. Often the language used is of a box with N coupons from which k are drawn with replacement.

Example 26. Sampling without replacement from a population. Now we take $\Omega = \{\underline{\omega} \in [N]^k : \omega_i \text{ are distinct elements of } [N]\}$ with $p_{\underline{\omega}} = 1/N(N-1)\ldots(N-k+1)$ for each $\underline{\omega} \in \Omega$.

Fix $m < N$ and define the random variable $X(\underline{\omega}) = \sum_{i=1}^{k} \mathbf{1}_{\omega_i \leq m}$. If the population $[N]$ contains a subset, say $[m]$, (could be the subset of people having a certain disease), then $X(\underline{\omega})$ counts the number of people in the sample who have the disease. Using X one can define events such as $A = \{\underline{\omega} : X(\underline{\omega}) = \ell\}$ for some $\ell \leq m$. If $\underline{\omega} \in A$, then ℓ of the ω_i must be in $[m]$ and the rest in $[N] \setminus [m]$. Hence

$$\#A = \binom{k}{\ell} m(m-1)\ldots(m-\ell+1)(N-m)(N-m-1)\ldots(N-m-(k-\ell)+1).$$

As the probabilities are equal for all sample points, we get

$$\begin{aligned}
\mathbf{P}(A) &= \frac{\binom{k}{\ell}m(m-1)\ldots(m-\ell+1)(N-m)(N-m-1)\ldots(N-m-(k-\ell)+1)}{N(N-1)\ldots(N-k+1)} \\
&= \frac{1}{\binom{N}{k}}\binom{m}{\ell}\binom{N-m}{k-\ell}.
\end{aligned}$$

This expression arises whenever the population is subdivided into two parts and we count the number of samples that fall in one of the sub-populations.

Example 27. Gibbs measures. Let Ω be a finite set and let $\mathcal{H} : \Omega \to \mathbb{R}$ be a function. Fix $\beta \geq 0$. Define $Z_\beta = \sum_\omega e^{-\beta \mathcal{H}(\omega)}$ and then set $p_\omega = \frac{1}{Z_\beta}e^{-\beta \mathcal{H}(\omega)}$. This is clearly a valid assignment of probabilities.

This is a class of examples from statistical physics. In that context, Ω is the set of all possible states of a system and $\mathcal{H}(\omega)$ is the energy of the state ω. In mechanics a system settles down to the state with the lowest possible energy, but if there are thermal fluctuations (meaning the ambient temperature is not absolute zero), then the system may also be found in other states, but higher energies are less and less likely. In the above assignment, for two states ω and ω', we see that $p_\omega/p_{\omega'} = e^{\beta(\mathcal{H}(\omega')-\mathcal{H}(\omega))}$ showing that higher energy states are less probable. When $\beta = 0$, we get $p_\omega = 1/|\Omega|$, the uniform distribution on Ω. In statistical physics, β is equated to $1/\kappa T$ where T is the temperature and κ is Boltzmann's constant.

Different physical systems are defined by choosing Ω and $\mathcal{H}$ differently. Hence this provides a rich class of examples which are of great importance in probability.

It may seem that probability is trivial, since the only problem is to find the sum of p_ω for ω belonging to event of interest. This is far from the case. The following example is an illustration.

Example 28. Percolation. Fix m, n and consider a rectangle in $\mathbb{Z}^2$, $R = \{(i, j) \in \mathbb{Z}^2 : 0 \leq i \leq n, \ 0 \leq j \leq m\}$. Draw this on the plane along with the grid lines. We see $(m+1)n$ horizontal edges and $(n+1)m$ vertical edges. Let E be the set of $N = (m+1)n + (n+1)m$ edges and let Ω be the set of all subsets of E. Then $|\Omega| = 2^N$. Let $p_\omega = 2^{-N}$ for each

$\omega \in \Omega$. An interesting event is

$$A = \{\omega \in \Omega : \text{ the subset of edges in } \omega$$

$$\text{connect the top side of } R \text{ to the bottom side of } R\}.$$

This may be thought of as follows. Imagine that each edge is a pipe through which water can flow. However each tube may be blocked or open. ω is the subset of pipes that are open. Now pour water at the top of the rectangle R. Will water trickle down to the bottom? The answer is yes if and only if ω belongs to A.

Finding $\mathbf{P}(A)$ is a very difficult problem. When n is large and $m = 2n$, it is expected that $\mathbf{P}(A)$ converges to a specific number, but proving it is an open problem as of today![5]

We now give two non-examples.

Example 29. Pick a natural number uniformly at random. The sample space is clearly $\Omega = \mathbb{N} = \{1, 2, 3, \ldots\}$. The phrase "uniformly at random" suggests that the elementary probabilities should be the same for all elements. That is $p_i = p$ for all $i \in \mathbb{N}$ for some p. If $p = 0$, then $\sum_{i \in \mathbb{N}} p_i = 0$ whereas if $p > 0$, then $\sum_{i \in \mathbb{N}} p_i = \infty$. This means that there is no way to assign elementary probabilities so that each number has the same chance to be picked.

This appears obvious, but many folklore puzzles and paradoxes in probability are based on the faulty assumption that it is possible to pick a natural number at random. For example, when asked a question like "What is the probability that a random integer is odd?", many people answer $1/2$. We want to emphasize that the probability space has to be defined first, and only then can probabilities of events be calculated. Thus, the question does not make sense to us and we do not have to answer it![6]

Example 30. A non-example. A dart is thrown at a circular dart board. We assume that the dart does hit the board but were it hits is "random" in the same sense in which we say the a coin toss is random. Intuitively this appears to make sense. However our framework is not general enough to incorporate this example. Let us see why.

The dart board can be considered to be the disk $\Omega = \{(x, y) : x^2 + y^2 \le r^2\}$ of given radius r. This is an uncountable set. We cannot assign elementary probabilities $p_{(x,y)}$ for each $(x, y) \in \Omega$ in any reasonable way. In fact the only reasonable assignment would be to set $p_{(x,y)} = 0$ for each (x, y) but then what is $\mathbf{P}(A)$ for a subset A? Uncountable sums are not well defined.

We need a branch of mathematics called *measure theory* to make proper sense of uncountable probability spaces. This will not be done in this course although we shall later

---

[5]In a very similar problem on a triangular lattice, it was proved by Stanislav Smirnov (2001) for which he won a fields medal. Proof that computing probabilities is not always trivial!

[6]For those interested, there is one way to make sense of such questions. It is to consider a sequence of probability spaces $\Omega^{(n)} = \{1, 2, \ldots, n\}$ with elementary probabilities $p_i^{(n)} = 1/n$ for each $i \in \Omega_n$. Then, for a subset $A \subseteq \mathbb{Z}$, we consider $\mathbf{P}_n(A \cap \Omega_n) = \#(A \cap [n])/n$. If these probabilities converge to a limit x as $n \to \infty$, then we could say that A has asymptotic probability x. In this sense, the set of odd numbers does have asymptotic probability $1/2$, the set of numbers divisible by 7 has asymptotic probability $1/7$ and the set of prime numbers has asymptotic probability 0. However, this notion of asymptotic probability has many shortcomings. Many subsets of natural numbers will not have an asymptotic probability, and even sets which do have asymptotic probability fail to satisfy basic rules of probability that we shall see later. Hence, we shall keep such examples out of our system.

say a bit about the difficulties involved. The same difficulty shows up in the following "random experiments" also.

(1) **Draw a number at random from the interval** $[0,1]$. $\Omega = [0,1]$ which is uncountable.

(2) **Toss a fair coin infinitely many times**. $\Omega = \{0,1\}^{\mathbb{N}} := \{\underline{\omega} = (\omega_1, \omega_2, \ldots) : \omega_i = 0 \text{ or } 1\}$. This is again an uncountable set.

Remark 31. In one sense, the first non-example is almost irredeemable but the second non-example can be dealt with, except for technicalities beyond this course. We shall later give a set of working rules to work with such "continuous probabilities". Fully satisfactory development will have to wait for a course in measure theory.