

2. Discrete probability spaces

Definition 9. Let Ω be a finite or countable¹ set. Let $p : \Omega \rightarrow [0, 1]$ be a function such that $\sum_{\omega \in \Omega} p_\omega = 1$. Then (Ω, p) is called a *discrete probability space*. Ω is called the *sample space* and p_ω are called *elementary probabilities*.

- Any subset $A \subseteq \Omega$ is called an *event*. For an event A we define its *probability* as $\mathbf{P}(A) = \sum_{\omega \in A} p_\omega$.
- Any function $X : \Omega \rightarrow \mathbb{R}$ is called a *random variable*. For a random variable we define its *expected value* or *mean* as $\mathbf{E}[X] = \sum_{\omega \in \Omega} X(\omega)p_\omega$.

All of probability in one line: Take an (interesting) probability space (Ω, p) and an (interesting) event $A \subseteq \Omega$. Find $\mathbf{P}(A)$.

This is the mathematical side of the picture. It is easy to make up any number of probability spaces - simply take a finite set and assign non-negative numbers to each element of the set so that the total is 1.

Example 10. $\Omega = \{0, 1\}$ and $p_0 = p_1 = \frac{1}{2}$. There are only four events here, $\emptyset, \{0\}, \{1\}$ and $\{0, 1\}$. Their probabilities are, 0, $1/2$, $1/2$ and 1, respectively.

Example 11. $\Omega = \{0, 1\}$. Fix a number $0 \leq p \leq 1$ and let $p_1 = p$ and $p_0 = 1 - p$. The sample space is the same as before, but the probability space is different for each value of p . Again there are only four events, and their probabilities are $\mathbf{P}\{\emptyset\} = 0$, $\mathbf{P}\{0\} = 1 - p$, $\mathbf{P}\{1\} = p$ and $\mathbf{P}\{0, 1\} = 1$.

Example 12. Fix a positive integer n . Let

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

Let $p_{\underline{\omega}} = 2^{-n}$ for each $\underline{\omega} \in \Omega$. Since Ω has 2^n elements, it follows that this is a valid assignment of elementary probabilities.

There are $2^{\#\Omega} = 2^{2^n}$ events. One example is $A_k = \{\underline{\omega} : \underline{\omega} \in \Omega \text{ and } \omega_1 + \dots + \omega_n = k\}$ where k is some fixed integer. In words, A_k consists of those n -tuples of zeros and ones that have a total of k many ones. Since there are $\binom{n}{k}$ ways to choose where to place these ones, we see that $\#A_k = \binom{n}{k}$. Consequently,

$$\mathbf{P}\{A_k\} = \sum_{\underline{\omega} \in A_k} p_{\underline{\omega}} = \frac{\#A_k}{2^n} = \begin{cases} \binom{n}{k} 2^{-n} & \text{if } 0 \leq k \leq n, \\ 0 & \text{otherwise.} \end{cases}$$

It will be convenient to adopt the notation that $\binom{a}{b} = 0$ if a, b are positive integers and if $b > a$ or if $b < 0$. Then we can simply write $\mathbf{P}\{A_k\} = \binom{n}{k} 2^{-n}$ without having to split the values of k into cases.

Example 13. Fix two positive integers r and m . Let

$$\Omega = \{\underline{\omega} : \underline{\omega} = (\omega_1, \dots, \omega_r) \text{ with } 1 \leq \omega_i \leq m \text{ for each } i \leq r\}.$$

The cardinality of Ω is m^r (since each co-ordinate ω_i can take one of m values). Hence, if we set $p_{\underline{\omega}} = m^{-r}$ for each $\underline{\omega} \in \Omega$, we get a valid probability space.

Of course, there are 2^{m^r} many events, which is quite large even for small numbers like $m = 3$ and $r = 4$. Some interesting events are $A = \{\underline{\omega} : \omega_r = 1\}$, $B = \{\underline{\omega} : \omega_i \neq 1 \text{ for all } i\}$, $C = \{\underline{\omega} : \omega_i \neq \omega_j \text{ if } i \neq j\}$. The reason why these are interesting will be explained later. Because of equal elementary probabilities, the probability of an event S is just $\#S/m^r$.

¹For those unfamiliar with countable sets, it will be explained in some detail later.

- Counting *A*: We have m choices for each of $\omega_1, \dots, \omega_{r-1}$. There is only one choice for ω_r . Hence $\#A = m^{r-1}$. Thus, $\mathbf{P}(A) = \frac{m^{r-1}}{m^r} = \frac{1}{m}$.
- Counting *B*: We have $m-1$ choices for each ω_i (since ω_i cannot be 1). Hence $\#B = (m-1)^r$ and thus $\mathbf{P}(B) = \frac{(m-1)^r}{m^r} = \left(1 - \frac{1}{m}\right)^r$.
- Counting *C*: We must choose a distinct value for each $\omega_1, \dots, \omega_r$. This is impossible if $m < r$. If $m \geq r$, then ω_1 can be chosen as any of m values. After ω_1 is chosen, there are $(m-1)$ possible values for ω_2 , and then $(m-2)$ values for ω_3 etc., all the way till ω_r which has $(m-r+1)$ choices. Thus, $\#C = m(m-1)\dots(m-r+1)$. Note that we get the same answer if we choose ω_i in a different order (it would be strange if we did not!).

Thus, $\mathbf{P}(C) = \frac{m(m-1)\dots(m-r+1)}{m^r}$. Note that this formula is also valid for $m < r$ since one of the factors on the right side is zero.

2.1. Probability in the real world. In real life, there are often situations where there are several possible outcomes but which one will occur is unpredictable in some way. For example, when we toss a coin, we may get heads or tails. In such cases we use words such as *probability or chance, event or happening, randomness* etc. What is the relationship between the intuitive and mathematical meanings of words such as probability or chance?

In a given physical situation, we choose one out of all possible probability spaces that we think captures best the chance happenings in the situation. The chosen probability space is then called a *model* or a *probability model* for the given situation. Once the model has been chosen, calculation of probabilities of events therein is a mathematical problem. Whether the model really captures the given situation, or whether the model is inadequate and over-simplified is a non-mathematical question. Nevertheless that is an important question, and can be answered by observing the real life situation and comparing the outcomes with predictions made using the model².

Now we describe several “random experiments” (a non-mathematical term to indicate a “real-life” phenomenon that is supposed to involve chance happenings) in which the previously given examples of probability spaces arise. Describing the probability space is the first step in any probability problem.

Example 14. Physical situation: Toss a coin. Randomness enters because we believe that the coin may turn up head or tail and that it is inherently unpredictable.

The corresponding probability model: Since there are two outcomes, the sample space $\Omega = \{0, 1\}$ (where we use 1 for heads and 0 for tails) is a clear choice. What about elementary probabilities? Under the equal chance hypothesis, we may take $p_0 = p_1 = \frac{1}{2}$. Then we have a probability model for the coin toss.

If the coin was not fair, we would change the model by keeping $\Omega = \{0, 1\}$ as before but letting $p_1 = p$ and $p_0 = 1 - p$ where the parameter $p \in [0, 1]$ is fixed.

Which model is correct? If the coin looks very symmetrical, then the two sides are equally likely to turn up, so the first model where $p_1 = p_0 = \frac{1}{2}$ is reasonable. However, if the coin looks irregular, then theoretical considerations are usually inadequate to arrive at

²Roughly speaking we may divide the course into two parts according to these two issues. In the probability part of the course, we shall take many such models for granted and learn how to calculate or approximately calculate probabilities. In the statistics part of the course we shall see some methods by which we can arrive at such models, or test the validity of a proposed model.

the value of p . Experimenting with the coin (by tossing it a large number of times) is the only way.

There is always an approximation in going from the real-world to a mathematical model. For example, the model above ignores the possibility that the coin can land on its side. If the coin is very thick, then it might be closer to a cylinder which can land in three ways and then we would have to modify the model...

Thus we see that example 11 is a good model for a physical coin toss. What physical situations are captured by the probability spaces in example 12 and example 13?

Example 12: This probability space can be a model for tossing n fair coins. It is clear in what sense, so we omit details for you to fill in.

The same probability space can also be a model for the tossing of the same coin n times in succession. In this, we are implicitly assuming that the coin forgets the outcomes on the previous tosses. While that may seem obvious, it would be violated if our “coin” was a hollow lens filled with a semi-solid material like glue (then, depending on which way the coin fell on the first toss, the glue would settle more on the lower side and consequently the coin would be more likely to fall the same way again). This is a coin with memory!

Example 13: There are several situations that can be captured by this probability space. We list some.

- There are r labelled balls and m labelled bins. One by one, we put the balls into bins “at random”. Then, by letting ω_i be the bin-number into which the i^{th} ball goes, we can capture the full configuration by the vector $\underline{\omega} = (\omega_1, \dots, \omega_n)$. If each ball is placed completely at random then the probabilities are m^{-r} for each configuration $\underline{\omega}$.

In that example, A is the event that the last ball ends up in the first bin, B is the event that the first bin is empty and C is the event that no bin contains more than one ball.

- If $m = 6$, then this may also be the model for throwing a fair die r times. Then ω_i is the outcome on the i^{th} throw. Of course, it also models throwing r different (and distinguishable) fair dice.
- If $m = 2$ and $r = n$, this is same as Example 12, and thus models the tossing of n fair coins (or a fair coin n times).
- Let $m = 365$. Omitting the possibility of leap years, this is a model for choosing r people at random and noting their birthdays (which can be in any of 365 “bins”). If we assume that all days are equally likely as a birthday (is this really true?), then the same probability space is a model for this physical situation. In this example, C is the event that no two people have the same birthday.

The next example is more involved and interesting.

Example 15. Real-life situation: Imagine a man-woman pair. Their first child is random, for example, the sex of the child, or the height to which the child will ultimately grow, etc cannot be predicted with certainty. How to make a probability model that captures the situation?

A possible probability model: Let there be n genes in each human, and each of the genes can take two possible values (Mendel’s “factors”), which we denote as 0 or 1. Then, let $\Omega = \{0, 1\}^n = \{\mathbf{x} = (x_1, \dots, x_n) : x_i = 0 \text{ or } 1\}$. In this sense, each human being can be encoded as a vector in $\{0, 1\}^n$.

To assign probabilities, one must know the parents. Let the two parents have gene sequences $\mathbf{a} = (a_1, \dots, a_n)$ and $\mathbf{b} = (b_1, \dots, b_n)$. Then the possible offsprings gene sequences are in the set $\Omega_0 := \{\mathbf{x} \in \{0, 1\}^n : x_i = a_i \text{ or } b_i, \text{ for each } i \leq n\}$. Let $L := \#\{i : a_i \neq b_i\}$.

One possible assignment of probabilities is that each of these offsprings is equally likely. In that case we can capture the situation in the following probability models.

- (1) Let Ω_0 be the sample space and let $p_{\mathbf{x}} = 2^{-L}$ for each $\mathbf{x} \in \Omega_0$.
- (2) Let Ω be the sample space and let

$$p_{\mathbf{x}} = \begin{cases} 2^{-L} & \text{if } \mathbf{x} \in \Omega_0 \\ 0 & \text{if } \mathbf{x} \notin \Omega_0. \end{cases}$$

The second one has the advantage that if we change the parent pair, we don’t have to change the sample space, only the elementary probabilities. What are some interesting events? Hypothetically, the susceptibility to a disease X could be determined by the first ten genes, say the person is likely to get the disease if there are at-most four 1s among the first ten. This would correspond to the event that $A = \{\mathbf{x} \in \Omega_0 : x_0 + \dots + x_{10} \leq 4\}$. (Caution: As far as I know, reading the genetic sequence to infer about the phenotype is still an impractical task in general).

Reasonable model? There are many simplifications involved here. Firstly, genes are somewhat ill-defined concepts, better defined are nucleotides in the DNA (and even then there are two copies of each gene). Secondly, there are many “errors” in real DNA, even the total number of genes can change, there can be big chunks missing, a whole extra chromosome etc. Thirdly, the assumption that all possible gene-sequences in Ω_0 are equally likely is incorrect - if two genes are physically close to each other in a chromosome, then they are likely to both come from the father or both from the mother. Lastly, if our interest originally was to guess the eventual height of the child or its intelligence, then it is not clear that these are determined by the genes alone (environmental factors such as availability of food etc. also matter). Finally, in case of the problem that Solomon faced, the information about genes of the parents was not available, the model as written would be use.

Remark 16. We have discussed at length the reasonability of the model in this example to indicate the enormous effort needed to find a sufficiently accurate but also reasonably simple probability model for a real-world situation. Henceforth, we shall omit such caveats and simply switch back-and-forth between a real-world situation and a reasonable-looking probability model as if there is no difference between the two. However, thinking about the appropriateness of the chosen models is much encouraged.