1. Probability Theory

Probability theory is a branch of pure mathematics, and forms the theoretical basis of statistics. In itself, probability theory has some basic objects and their relations (like real numbers, addition, etc. for analysis) and it makes no pretense of saying anything about the real world. Axioms are given and theorems are then deduced about these objects, just as in any other part of mathematics.

But, a very important aspect of probability is that it is *applicable*. In other words, there are many real world situations in which it is reasonable to take a model in probability and it turns out to reasonably replicate features of the real world situation.

2. Discrete probability spaces

Definition 1. Let Ω be a finite or countable set. Let $p : \Omega \to [0,1]$ 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 $X : \Omega \to \mathbb{R}$ is called a *random variable*.

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 2. $\Omega = \{0,1\}$ 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 3. $\Omega = \{0,1\}$. Fix a number $0 \le p \le 1$ 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 4. Fix a positive integer n. Let

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

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 [1]. One example is $A_k = \{\underline{\omega} : \underline{\omega} \in \Omega \text{ and } \omega_1 + \cdots + \omega_n = k\}$, 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,

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

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.

[1] $\#\Omega$ denotes the cardinality of the set Ω.

Example 5. Fix two positive integers r and m. Let

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

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 $\#S/m^r$.

- Counting A: We have m choices for each of $\omega_1, \ldots, \omega_{r-1}$. 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 $\mathbf{P}(B) = \frac{(m-1)^r}{m^r} = (1 - \frac{1}{m})^r$.
- Counting C: We must choose a distinct value for each $\omega_1, \ldots, \omega_r$. 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, $\#C = m(m-1)\cdots(m-r+1)$. 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, $\mathbf{P}(C) = \frac{m(m-1)\cdots(m-r+1)}{m^r}$.

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 model[2].

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

[2]Roughly 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.

examples of probability spaces arise. Describing the probability space is the first step in any probability problem.

Example 6. 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 2 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 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 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 3 is a good model for a physical coin toss. What physical situations are captured by the probability spaces in Example 4 and Example 5?

Example 4: 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 5: 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 $\underline{\omega} = (\omega_1, \ldots, \omega_n)$. 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 4, 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.

Example 7. Toss n coins. We saw this before, but assumed that the coins are fair. Now we do not. The sample space is

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

Further, we assign $p_{\underline{\omega}} = \alpha^{(1)}_{\omega_1} \dots \alpha^{(n)}_{\omega_n}$. Here, $\alpha^{(j)}_0$ and $\alpha^{(j)}_1$ are supposed to indicate the probabilities that the j^{th} coin falls tails up or heads up, respectively. Why did we take the product of $\alpha^{(j)}_{\cdot}$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 $p = \alpha^{(j)}_1$ (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 $\mathbf{P}(A_k) = \sum\limits_{\ell=k}^{n} \binom{n}{\ell} p^\ell q^{n-\ell}$.

Example 8. Toss a coin n times. Again

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

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 9. 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 $\pi \in S_{52}$. More generally, we can make a model for a deck of n cards, in which case the sample space is S_n and elementary probabilities are $1/n!$ for each n.

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

As an illustration, when $n = 3$, the sample space is

$$S_3 = \{(1, 2, 3), (1, 3, 2), (2, 3, 1), (2, 1, 3), (3, 1, 2), (3, 2, 1)\},$$

where $(2, 3, 1)$ denotes the deck where the top card is 2, the next one is 3 and the bottom card is 1. The elementary probabilities are all $1/6$ in this case.

Example 10. Toss a coin till a head turns up. $\Omega = \{1, 01, 001, 0001, \ldots\} \cup \{\bar{0}\}$. Let us write $0^k1 = 0 \ldots 01$ as a short form for k zeros (tails) followed by 1 and $\bar{0}$ stands for the sequence of all tails. Let $p \in [0, 1]$. Then, we set $p_{0^k1} = q^k p$ for each $k \in \mathbb{N}$. We also set $p_{\bar{0}} = 0$ if $p > 0$ and $p_{\bar{0}} = 1$ if $p = 0$. This is forced on us by the requirement that elementary probabilities add to 1.

Let $A = \{0^k1 : k \geq n\}$ be the event that at least n tails fall before a head turns up. Then $\mathbf{P}(A) = q^n p + q^{n+1} p + \cdots = q^n$.

Example 11. 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 i^{th} ball goes.

Example 12. Birthday "paradox" There are n people at a party. What is the chance that two of them have the same birthday?

This can be thought of as a balls in bin problem, where the bins are labelled $1, 2, \ldots, 365$ (days of the year), and the balls are labelled $1, 2, \ldots, n$ (people). The sample space can be copied from the previous example with $r = n$ and $m = 365$. In doing this, we have omitted the possibility of February 29th, which simplifies our life a little bit. What are the elementary probabilities? Let us assume that all sample points have equal probability[4].

The event of interest is that there is at least one bin that has at least two balls. In other words,

$$A = \{\underline{\omega} = (\omega_1, \ldots, \omega_n) : \omega_i = \omega_j \text{ for some } i < j\}.$$

In this case, it is easier to calculate the probability of

$$A^c = \{\underline{\omega} = (\omega_1, \ldots, \omega_n) : \omega_i \neq \omega_j \text{ for all } i < j\}.$$

Indeed, the cardinality of A^c is $m(m-1)\cdots(n-m+1)$ (each person has to be assigned a different birthday), whence

$$\mathbf{P}(A) = 1 - \frac{m(m-1)\cdots(m-n+1)}{m^r}.$$

[4]This assumption is not entirely realistic, for example, Figure 3 shows actual data from a specific geographic location over a specific period of time. Seasonal variation is apparent! In addition there are complications such as the possibility of a pair of twins attending the party.

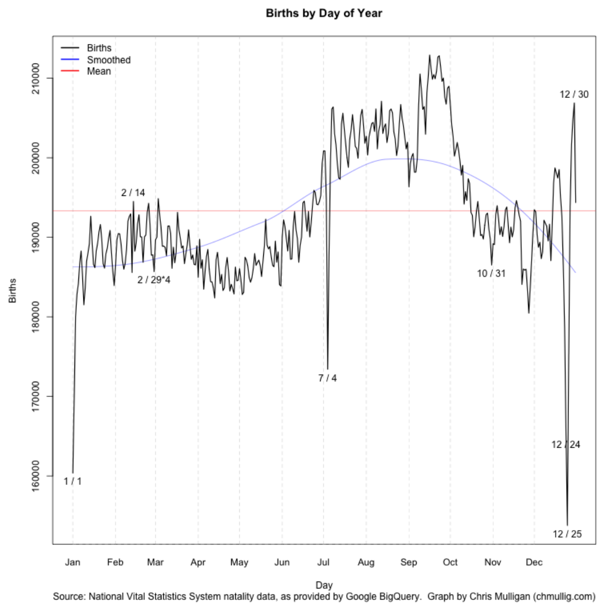

FIGURE 1. Frequencies of birthdays in the United States of America from 1969 to 1988. Data taken from Andrew Gelman.

If $n > m$, the probability is obviously zero. The reason this is called a "paradox" is that even for n much smaller than m, the probability becomes significantly large. For example, for $m = 365$, here are the probabilities for a few values of n:

n	5	15	25	35	45
$\mathbf{P}(A)$	0.027	0.253	0.569	0.814	0.941

Example 13. 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

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

We give two proposals for the elementary probabilities.

(1) Let $p^{\text{MB}}_{(\ell_1,\ldots,\ell_m)} = \frac{m!}{\ell_1!\ell_2!\cdots\ell_m!}\frac{1}{m^r}$. 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.

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.

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

Exercise 14. (1) Let $(\ell_1, \ldots, \ell_m) \in \Omega$. Show that $\#\{\underline{\omega} \in [m]^r : \sum_{j=1}^r \mathbf{1}_{\omega_j = i} = \ell_i \text{ for each } i \in [m]\} = \frac{n!}{\ell_1!\ell_2!\cdots\ell_m!}$. Hence or directly, show that $\sum\limits_{\omega\in\Omega} p_\omega^{MB} = 1$.

(2) Show that $\#\Omega = \binom{m+r-1}{r-1}$. Hence, $\sum\limits_{\omega\in\Omega} p_\omega^{BE} = 1$.

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}. 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![5]

Example 15. Sampling with replacement from a population. Define $\Omega = \{\underline{\omega} \in [N]^k : \omega_i \in [N] \text{ for } 1 \le i \le k\}$ with $p_{\underline{\omega}} = 1/N^k$ for each $\underline{\omega} \in \Omega$. Here, $[N]$ is the population (so the size 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 16. Sampling without replacement from a population. Now we take

$$\Omega = \left\{\underline{\omega} \in [N]^k : \omega_i \text{ are distinct elements of } [N]\right\},$$

$$p_{\underline{\omega}} = \frac{1}{N(N-1)\cdots(N-k+1)} \text{ for each } \underline{\omega} \in \Omega.$$

Fix $m < N$ and define the random variable $X(\underline{\omega}) = \sum_{i=1}^k \mathbf{1}_{\omega_i \le m}$. 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

[5]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 $m \ge r$, the sample space is $\Omega_{\text{FD}} = \{(\ell_1, \ldots, \ell_m) : \ell_i = 0 \text{ or } 1 \text{ and } \ell_1 + \cdots + \ell_m = r\}$ 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.

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

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

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

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

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.

We now give two non-examples.

Example 17. A non-example - 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 18. Another non-example - Throwing darts. 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

[6]For those interested, there is one way to make sense of such questions. It is to consider a sequence of probability spaces $\Omega^{(n)} = [n]$ 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.

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 $\Omega = \{(x,y) : x^2 + y^2 \leq r^2\}$ 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? [7] 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 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 19. 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.

[7] Some probability problems are geometric in nature. What this means is that our desired outcome space is in fact some area called the feasible region. To find the probability of this outcome, we find the ratio feasible region/sample space. A classic example of this type of problem is a dart board.
Let us say that $r = 10$ here. Furthermore, the bulls-eye is exactly in the center and has a 1 in diameter. If I throw a dart randomly and it hits the board, what is the probability that it hits the bulls-eye?