Introduction to Discrete Probability

Probability theory has its origins in gambling — analyzing card games, dice, roulette wheels. Today it is an essential tool in engineering and the sciences. No less so in EECS, where its use is widespread in algorithms, systems, signal processing, learning theory and control/AI.

Here are some typical statements that you might see concerning probability:

1. The chance of getting a flush in a 5-card poker hand is about 2 in 1000.
2. The chance that this randomized primality testing algorithm outputs "prime" when the input is not prime is at most one in a trillion.
3. In this load-balancing scheme, the probability that any processor has to deal with more than 12 requests is negligible.
4. The average time between system failures is about 3 days.
5. There is a 30% chance of a magnitude 8.0 earthquake in Northern California before 2030.

Implicit in all such statements is the notion of an underlying *probability space*. This may be the result of a random experiment that we have ourselves constructed (as in 1, 2 and 3 above), or some model we build of the real world (as in 4 and 5 above). None of these statements makes sense unless we specify the probability space we are talking about: for this reason, statements like 5 (which are typically made without this context) are almost content-free.

In this note, we will try to understand all this more clearly. The first important notion here is that of a *random experiment*. We will start by introducing the space of all possible outcomes of the experiment, called a sample space. Each element of the sample space is assigned a probability which tells us how likely the outcome is to occur when we actually perform the experiment.

1 Random Experiments

In general, a random experiment consists of drawing a sample of k elements from a set S of cardinality n. The possible outcomes of such an experiment are exactly the objects that we counted in the last note. Recall from the last note that we considered four possible scenarios for counting, depending upon whether we sampled with or without replacement, and whether the order in which the k elements are chosen does or does not matter. The same will be the case for our random experiments. The outcome of a random experiment is called a *sample point*, and the *sample space* (often denoted by Ω) is the set of all possible outcomes of the experiment.

An example of such an experiment is tossing a coin 4 times. In this case, $S = \{H,T\}$ and we are drawing 4 elements with replacement. $HTHT$ is an example of a sample point and the sample space Ω has 16 elements, as illustrated in the following picture:

EECS 70 离散数学与概率论 2020秋季

Note 13

离散概率导论

概率论起源于赌博——分析纸牌游戏、骰子、轮盘赌等。如今，它已成为工程和科学领域不可或缺的工具。在EECS中同样如此，广泛应用于算法、系统、信号处理、学习理论以及控制/AI等领域。

以下是关于概率的一些典型陈述：

1. 在5张牌的扑克牌手中得到同花的概率大约是千分之二。
2. 当输入不是素数时，这个随机素性检测算法输出“素数”的概率最多为万亿分之一。
3. 在这种负载均衡方案中，任何处理器需要处理超过12个请求的概率可以忽略不计。
4. 系统故障之间的平均时间约为3天。
5. 在2030年之前，北加州发生8.0级地震的可能性为30%。

所有这些陈述背后都隐含着一个基本的概率空间概念。这可能是我们自己构建的随机实验的结果（如上面的1、2和3），或者是对现实世界建立的某种模型（如上面的4和5）。除非我们明确指出所讨论的概率空间，否则这些陈述都没有意义：正因为如此，像5这样的陈述（通常在没有上下文的情况下做出）几乎是空洞无物的。

在本讲义中，我们将尝试更清晰地理解这些概念。第一个重要概念是随机实验。我们将从引入实验所有可能结果的空间开始，称为样本空间。样本空间中的每个元素都被赋予一个概率，告诉我们当实际执行实验时该结果发生的可能性。

1 随机实验

一般来说，一个随机实验包括从大小为n的集合S中抽取k个元素作为样本。这类实验的可能结果正是我们在上一讲义中计数过的对象。回顾上一讲义，我们曾考虑过四种可能的计数情形，取决于我们是有放回还是无放回抽样，以及选出的k个元素的顺序是否重要。我们的随机实验也将面临同样的情况。随机实验的结果称为样本点，样本空间（通常记作Ω）是实验所有可能结果构成的集合。

此类实验的一个例子是将一枚硬币抛掷4次。在这种情况下，S = {H,T}，我们从中抽取4个元素并允许重复。HTHT是一个样本点的例子，样本空间Ω包含16个元素，如下图所示：

$$\Omega = \begin{Bmatrix} HHHH & HTHH & THHH & TTHH \\ HHHT & HTHT & THHT & TTHT \\ HHTH & HTTH & THTH & TTTH \\ HHTT & HTTT & THTT & TTTT \end{Bmatrix}$$

How do we determine the chance of each particular outcome, such as $HHTH$, of our experiment? In order to do this, we need to define the probability for each sample point, as described below.

2 Probability Spaces

A probability space is a sample space Ω, together with a probability $\mathbb{P}[\omega]$ (often also denoted as $\Pr[\omega]$) for each sample point ω, such that

- (Non-negativity): $0 \leq \mathbb{P}[\omega] \leq 1$ for all $\omega \in \Omega$.
- (Total one): $\sum_{\omega \in \Omega} \mathbb{P}[\omega] = 1$, i.e., the sum of the probabilities over all outcomes is 1.

The easiest way to assign probabilities to sample points is to do it *uniformly*: if $|\Omega| = N$, then $\mathbb{P}[\omega] = \frac{1}{N}$, $\forall \omega \in \Omega$. For example, if we toss a fair coin 4 times, each of the 16 sample points (as pictured above) is assigned probability $\frac{1}{16}$. We will see examples of non-uniform probability distributions soon.

After performing an experiment, we are often interested in knowing whether a certain event occurred. For example, we might be interested in the event that there were "exactly 2 H's in four tosses of the coin". How do we formally define the concept of an event in terms of the sample space Ω? The answer is to identify the event "exactly 2 H's in four tosses of the coin" with the set of all those outcomes in which there are exactly two H's: $\{HHTT, HTHT, HTTH, THHT, THTH, TTHH\} \subset \Omega$. Hence, formally an event A is just a subset of the sample space Ω, i.e., $A \subseteq \Omega$.

How should we define the probability of an event A? Naturally, we should just *add up* the probabilities of the sample points in A. For any event $A \subseteq \Omega$, we define the probability of A to be

$$\mathbb{P}[A] = \sum_{\omega \in A} \mathbb{P}[\omega].$$

Note that $0 \leq \mathbb{P}[A] \leq 1$ for all $A \subseteq \Omega$, and $\mathbb{P}[\Omega] = 1$. The probability of getting exactly two H's in four coin tosses can be calculated using this definition as follows. Event A consists of all sequences that have exactly two H's, and so $|A| = \binom{4}{2} = 6$. For this example, there are $2^4 = 16$ possible outcomes for flipping four coins. Thus, each sample point $\omega \in A$ has probability $\frac{1}{16}$; and since there are six sample points in A, we obtain $\mathbb{P}[A] = 6 \cdot \frac{1}{16} = \frac{3}{8}$.

3 Examples

We will now look at examples of random experiments and their corresponding sample spaces, along with possible probability spaces and events.

$$\Omega = \left\{ \begin{array}{cccc} HHHH & HTHH & THHH & TTHH \\ HHHT & HTHT & THHT & TTHT \\ HHTH & HTTH & THTH & TTTH \\ HHTT & HTTT & THTT & TTTT \end{array} \right\}$$

我们如何确定实验中每个特定结果（例如HHTH）的可能性？为了做到这一点，我们需要为每个样本点定义概率，如下所述。

2 概率空间

概率空间是一个样本空间Ω，加上为每个样本点ω定义的概率P[ω]（也常记作Pr[ω]），满足以下条件

- （非负性）：对所有ω∈Ω，0 ⩽P[ω] ⩽1。
- （总和为一）：$\sum_{\omega\in\Omega}$ P[ω] = 1，即所有结果的概率之和为1。

给样本点分配概率最简单的方法是均匀分配：如果|Ω| = N，则P[ω] = 1 $\overline{N}$，∀ω∈Ω。例如，如果我们抛一枚公平硬币4次，每一个16个样本点（如上图所示）都被赋予概率116。我们很快就会看到非均匀概率分布的例子。

进行实验后，我们常常关心某个特定事件是否发生。例如，我们可能关心在四次抛硬币中恰好出现两次正面的事件。我们如何用样本空间Ω来形式化定义“事件”这一概念？答案是将“四次抛硬币恰好出现两次正面”这一事件与所有恰好包含两个正面的结果集合对应起来：{HHTT,HTHT,HTTH,THHT,THTH,TTHH} ⊂Ω。因此，形式上一个事件A只是样本空间Ω的一个子集，即A ⊆Ω。

我们应如何定义事件A的概率？自然地，我们应该把A中所有样本点的概率加起来。对于任意事件A ⊆Ω，我们定义A的概率为

$$\mathbb{P}[A] = \sum_{\omega\in A} \mathrm{P}[\omega]。$$

注意对于所有A ⊆Ω都有0 ⩽P[A] ⩽1，且P[Ω] = 1。可以通过如下定义计算四次抛硬币恰好得到两个正面的概率。事件A由所有恰好包含两个正面的序列组成，因此|A| = 4 $\binom{}{2}$ = 6。在这个例子中，抛四枚硬币共有24 = 16种可能的结果。因此，每个样本点 ω∈A 的概率是1 $\overline{16}$；并且由于A中有六个样本点，我们得到 $\mathbb{P}[A] = 6 \cdot \frac{1}{16} = \frac{3}{8}$.

3 例子

我们现在将查看一些随机实验及其对应的样本空间，以及可能的概率空间和事件的例子。

3.1 Coin Tosses

Suppose we have a coin with $\mathbb{P}(H) = p$ and $\mathbb{P}(T) = 1 - p$, and our experiment consists of flipping the coin 4 times. The sample space Ω consists of the sixteen possible sequences of H's and T's shown in the figure on the previous page.

The probability space depends on p. If $p = \frac{1}{2}$ the probabilities are assigned uniformly; the probability of each sample point is $\frac{1}{16}$. What if $p = \frac{2}{3}$? Then the probabilities are different. For example, $\mathbb{P}[HHHH] = \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{16}{81}$, while $\mathbb{P}[TTHH] = \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{4}{81}$. (Note: We have simply multiplied probabilities here; we will explain later why this is the correct thing to do for this example. It is NOT always OK to multiply probabilities like this.)

What type of events can we consider in this setting? Let A be the event that all four coin tosses are the same. Then $A = \{HHHH, TTTT\}$. $HHHH$ has probability $\left(\frac{2}{3}\right)^4$ and $TTTT$ has probability $\left(\frac{1}{3}\right)^4$. Thus, $\mathbb{P}[A] = \mathbb{P}[HHHH] + \mathbb{P}[TTTT] = \left(\frac{2}{3}\right)^4 + \left(\frac{1}{3}\right)^4 = \frac{17}{81}$.

Next, consider the event B that there are exactly two heads. The probability of any particular outcome with two heads (such as $HTHT$) is $\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^2$. How many such outcomes are there? There are $\binom{4}{2} = 6$ ways of choosing the positions of the heads, and these choices completely specify the sequence. So, $\mathbb{P}[B] = 6\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^2 = \frac{24}{81} = \frac{8}{27}$.

More generally, if we flip the coin n times, we get a sample space Ω of cardinality 2^n. The sample points are all possible length-n sequences of H's and T's. If the coin has $\mathbb{P}(H) = p$, and if we consider any sequence of n coin flips with exactly r H's, then the probability of this sequence is $p^r(1-p)^{n-r}$.

Now consider the event C that we get exactly r H's when we flip the coin n times. This event consists of exactly $\binom{n}{r}$ sample points and each has probability $p^r(1-p)^{n-r}$. So, $\mathbb{P}[C] = \binom{n}{r}p^r(1-p)^{n-r}$.

Biased coin tossing sequences show up in many contexts: for example, they might model the behavior of n trials of a faulty system, which fails each time with probability p.

3.2 Rolling Dice

Consider rolling two fair dice. In this experiment, $\Omega = \{(i,j) : 1 \le i,j \le 6\}$. The probability space is uniform, i.e., all sample points have the *same* probability $\frac{1}{|\Omega|} = \frac{1}{36}$. Hence, the probability of any event A is

$$\mathbb{P}[A] = \frac{\text{\# of sample points in } A}{\text{\# of sample points in } \Omega} = \frac{|A|}{|\Omega|}.$$

So, for uniform spaces, computing probabilities reduces to *counting* sample points!

Now consider two events: the event A that the sum of the dice is at least 10 and the event B that there is at least one 6. By enumerating the sample points contained in each event, it can be easily shown that $|A| = 6$ and $|B| = 11$. Then, by the observation above, it follows that $\mathbb{P}[A] = \frac{6}{36} = \frac{1}{6}$ and $\mathbb{P}[B] = \frac{11}{36}$.

3.3 Card Shuffling

Consider a random experiment of shuffling a deck of standard playing cards. Here, Ω is equal to the set of the 52! permutations of the deck. We assume that the probability space is uniform. (Note that we are really talking about an idealized mathematical model of shuffling here; in real life, there will always be a bit of bias in our shuffling. However, the mathematical model is close enough to be useful.)

3.1 抛硬币

假设我们有一枚硬币，满足P(H) = p且P(T) = 1−p，我们的实验包括将这枚硬币抛掷4次。样本空间Ω由十六种可能的H和T序列组成，如前一页图中所示。

概率空间取决于p。如果p = 1，则概率被均匀分配；每个样本点的概率是116。如果p = 2呢？那么 — 2 概率就不同了。例如，P[HHHH] =
$\frac{2}{3}\times\frac{2}{3}\times\frac{2}{3}\times\frac{2}{3}=\frac{16}{81}$，而P[TTHH] = 1 $_{3}\times\frac{1}{3}\times\frac{2}{3}\times\frac{2}{3}=\frac{4}{81}$。（注意：我们只是简单地将概率相乘
这里；我们将在后面解释为什么在这个例子中这样做是正确的。（注意：这种概率相乘的方法并不总是适用。）

在这种情况下我们可以考虑哪些类型的事件？设A为四次抛币结果都相同的事件。那么A = {HHHH,TTTT}。HHHH的概率是2 4，TTTT的概率是1 4 $\left(_{3}\right)$。因此，P[A] = P[HHHH]+P[TTTT] = 2 $\left(_{3}\right)$
4 + 1 4 = 17 $\left(_{3}\right)$ $\left(_{3}\right)$ $_{81}$.

接下来，考虑事件B，即恰好有两个正面。任意一个有两个正面的特定结果（例如HTHT）的概率是2 $\left(_{3}\right)^{2}\left(\frac{1}{3}\right)$ 2。这样的结果有多少种？有42 = 6种方式 $\left(\right)$
选择正面位置的方式数量，而这些选择完全确定了序列。因此，P[B] = 6 2
$\left(_{3}\right)^{2}\left(\frac{1}{3}\right)^{2}=\frac{24}{81}=\frac{8}{27}$.

更一般地，如果我们抛掷硬币n次，会得到一个大小为2^n的样本空间Ω。样本点是所有可能的长度为n的H和T序列。如果硬币出现正面的概率为P(H) = p，并且我们考虑任意一个包含恰好r个正面的n次抛掷序列，那么该序列的概率为p^r(1−p)^(n−r)。

现在考虑事件C，即我们抛掷硬币n次时恰好得到r个正面。这个事件恰好包含n $\left(_{r}\right)$ 样本点，每个的概率为 p^r(1−p)^(n−r)。因此，P[C] = C(n,r) $\left(_{r}\right)$ p^r(1−p)^(n−r)

有偏硬币抛掷序列出现在许多情境中：例如，它们可能模拟一个故障系统进行 n 次试验的行为，每次试验以概率 p 失败。

3.2 掷骰子

考虑掷两个公平的骰子。在此实验中，Ω = {(i, j) : 1 ⩽ i, j ⩽ 6}。概率空间是均匀的，即所有样本点具有相同的概率 1/36 $_{|\Omega|}=\frac{1}{36}$。因此，任何事件 A 的概率为

$$\mathbb{P}[A]=\frac{\text{事件 A 中的样本点数量 / 整}}{\text{个样本空间中的样本点数量}}=\frac{|A|}{|\Omega|}.$$

因此，对于均匀空间，计算概率就归结为计算样本点的数量！

现在考虑两个事件：事件 A 表示骰子点数之和至少为 10，事件 B 表示至少有一个骰子显示 6。通过枚举每个事件包含的样本点，可以很容易地证明 |A| = 6 且 |B| = 11。根据上述观察，可得 P[A] = 6 $_{36}=\frac{1}{6}$ 且 P[B] = 11 $_{36}$.

3.3 洗牌

考虑一个随机实验：洗一副标准扑克牌。这里，Ω 等于这副牌所有 52! 种排列的集合。我们假设概率空间是均匀的。（请注意，我们在这里讨论的是关于洗牌的理想化数学模型；在现实中，我们的洗牌总会有一些偏差。然而，这个数学模型已经足够接近实际而具有实用性。）

3.4 Poker Hands

Here's another experiment: shuffling a deck of cards and dealing a poker hand. In this case, S is the set of 52 cards and our sample space $\Omega = \{\text{all possible poker hands}\}$, which corresponds to choosing $k = 5$ objects without replacement from a set of size $n = 52$ where order does not matter. Hence, as we saw in Note 11, $|\Omega| = \binom{52}{5} = \frac{52\times 51\times 50\times 49\times 48}{5\times 4\times 3\times 2\times 1} = 2,598,960$. Assuming that the deck is well shuffled, the probability of each outcome is equally likely and we are therefore dealing with a uniform probability space.

Let A be the event that the poker hand is a flush. (For those who are not familiar with poker, a *flush* is a hand in which all cards have the same suit, say Hearts.) Since the probability space is uniform, computing $\mathbb{P}[A]$ reduces to simply computing $|A|$, the number of poker hands that are flushes. There are 13 cards in each suit, so the number of flushes in each suit is $\binom{13}{5}$. The total number of flushes is therefore $4 \times \binom{13}{5}$. Then we have

$$\mathbb{P}[\text{hand is a flush}] = \frac{4\times\binom{13}{5}}{\binom{52}{5}} = 4\times\frac{13!}{5!8!}\times\frac{5!47!}{52!} = 4\times\frac{13\cdot 12\cdot 11\cdot 10\cdot 9}{52\cdot 51\cdot 50\cdot 49\cdot 48} \approx 0.002.$$

3.5 Balls and Bins

In this experiment, we will throw 20 (labeled) balls into 10 (labeled) bins. Assume that each ball is equally likely to land in any bin, regardless of what happens to the other balls.

If you wish to understand this situation in terms of sampling a sequence of k elements from a set S of cardinality n: the set S consists of the 10 bins, and we are sampling with replacement $k = 20$ times. The order of sampling matters, since the balls are labeled.

The sample space Ω is equal to $\{(b_1, b_2, \ldots, b_{20}) : 1 \le b_i \le 10 \text{ for each } i = 1, \ldots, 20\}$, where the component b_i denotes the bin in which ball i lands. The cardinality $|\Omega|$ of the sample space is equal to 10^{20}, since each element b_i in the sequence has 10 possible choices and there are 20 elements in the sequence. More generally, if we throw m balls into n bins, we have a sample space of size n^m. The probability space is uniform; as we said earlier, each ball is equally likely to land in any bin.

Let A be the event that bin 1 is empty. Since the probability space is uniform, we simply need to count how many outcomes have this property. This is exactly the number of ways all 20 balls can fall into the remaining nine bins, which is 9^{20}. Hence, $\mathbb{P}[A] = \frac{9^{20}}{10^{20}} = \left(\frac{9}{10}\right)^{20} \approx 0.12$.

Let B be the event that bin 1 contains at least one ball. This event is the *complement* $\overline{A}$ of A, i.e., it consists of precisely those sample points which are not in A. So $\mathbb{P}[B] = 1 - \mathbb{P}[A] \approx 0.88$. More generally, if we throw m balls into n bins, we have:

$$\mathbb{P}[\text{bin 1 is empty}] = \left(\frac{n-1}{n}\right)^m = \left(1 - \frac{1}{n}\right)^m.$$

As we shall see, balls and bins is another probability space that shows up very often in Computer Science: for example, we can think of it as modeling a load balancing scheme, in which each job is sent to a random processor.

It is also a more general model for problems we have previously considered. For example, flipping a fair coin 3 times is a special case in which the number of balls (m) is 3 and the number of bins (n) is 2. Rolling two dice is a special case in which $m = 2$ and $n = 6$.

3.4 扑克手牌

另一个实验：洗一副牌并发出一手扑克牌。在这种情况下，S 是 52 张牌的集合，我们的样本空间 Ω = {所有可能的扑克手牌}，对应于从大小为 n = 52 的集合中无放回地选出 k = 5 个对象，且顺序无关。因此，正如我们在 Note 11 中所见，|Ω| = C(52,5)

$\binom{}{5} = \frac{52\times51\times50\times49\times48}{5\times4\times3\times2\times1}$ = 2,598,960。假设牌堆充分洗匀，则每个

每个结果出现的可能性相同，因此我们处理的是一个均匀概率空间。

设 A 为事件“扑克手牌是一副同花”。（对于不熟悉扑克的人来说，同花是指所有牌属于同一花色的手牌，比如红心。）由于概率空间是均匀的，计算 P[A] 就简化为计算 |A|，即构成同花的扑克手牌数量。每种花色有 13 张牌，因此每种花色的同花数量为 13

$\binom{}{5}$。因此同花总数为 4×13 $\binom{}{5}$。然后我们

有

$$\text{P[手牌是一副同花]} = \frac{4\times\binom{13}{5}}{\binom{52}{5}} = 4\times\frac{13!}{5!8!}\times\frac{5!47!}{52!} = 4\times\frac{13\cdot12\cdot11\cdot10\cdot9}{52\cdot51\cdot50\cdot49\cdot48} \approx 0.002。$$

3.5 球与箱子

在这个实验中，我们将把 20 个（带标签的）球扔进 10 个（带标签的）箱子。假设每个球落入任何一个箱子的可能性相等，不受其他球的影响。

如果你想用从大小为 n 的集合 S 中采样 k 个元素的序列来理解这种情况：集合 S 包含 10 个箱子，我们进行 k = 20 次有放回采样。采样的顺序很重要，因为球是有标签的。

样本空间 Ω 等于 {(b1,b2,...,b20) : 1 ≤ bi ≤ 10 对每个 i = 1,...,20}，其中分量 bi 表示第 i 个球落入的箱子。样本空间的基数 |Ω| 等于 10^20，因为序列中的每个元素 bi 有 10 种可能选择，并且序列中共有 20 个元素。更一般地，如果我们把 m 个球扔进 n 个箱子，则样本空间大小为 n^m。概率空间是均匀的；如前所述，每个球落入任一箱子的可能性相等。

设 A 为事件“箱子 1 为空”。由于概率空间是均匀的，我们只需计算具有此性质的结果数量。这正好等于所有 20 个球都落入其余九个箱子的方式数目，即 9^20。因此，P[A] = 9^20

$\overline{10^{20}} = \left(\frac{9}{10}\right)$20 ≈0.12。

设 B 为事件“箱子 1 至少包含一个球”。该事件是 A 的补事件 $\bar{A}$，即它恰好由那些不属于 A 的样本点组成。因此 P[B] = 1−P[A] ≈0.88。更一般地，如果我们把 m 个球扔进 n 个箱子，我们有：

$$\text{P[箱子 1 为空]} = \left(\frac{n-1}{n}\right)^m = \left(1-\frac{1}{n}\right)^m.$$

正如我们将看到的，球与箱子是计算机科学中经常出现的另一种概率空间：例如，我们可以将其视为一种负载均衡方案的建模，其中每个任务被随机发送到某个处理器。

它也是对我们之前考虑过的问题的一个更通用模型。例如，抛三次公平硬币是一个特例，其中球的数量（m）为 3，箱子的数量（n）为 2。掷两个骰子则是 m = 2 且 n = 6 的特例。

3.6 Birthday Paradox

The "birthday paradox" is a remarkable phenomenon that examines the chances that two people in a group have the same birthday. It is a "paradox" not because of a logical contradiction, but because it goes against intuition. For ease of calculation, we take the number of days in a year to be 365. Then $S = \{1, \ldots, 365\}$, and the random experiment consists of drawing a sample of n elements from S, where the elements are the birth dates of n people in a group. Then $|\Omega| = 365^n$. This is because each sample point is a sequence of possible birthdays for n people; so there are n points in the sequence and each point has 365 possible values.

Let A be the event that at least a pair of people have the same birthday. If we want to determine $\mathbb{P}[A]$, it might be simpler to instead compute the probability of the complement of A; i.e., $\mathbb{P}[\overline{A}]$, where $\overline{A}$ is the event that no two people have the same birthday. Since $\mathbb{P}[A] = 1 - \mathbb{P}[\overline{A}]$, we can then easily compute $\mathbb{P}[A]$.

We are again working in a uniform probability space, so we just need to determine $|\overline{A}|$. Equivalently, we are computing the number of ways for no two people to have the same birthday. There are 365 choices for the first person, 364 for the second, ..., $365 - n + 1$ choices for the n-th person, for a total of $365 \times 364 \times \cdots \times (365 - n + 1)$. This is simply an application of the First Rule of Counting from Note 11; we are sampling without replacement and the order matters.

In summary, we have

$$\mathbb{P}[\overline{A}] = \frac{|\overline{A}|}{|\Omega|} = \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n},$$

so $\mathbb{P}[A] = 1 - \mathbb{P}[\overline{A}] = 1 - \frac{365 \times 364 \times \cdots \times (365-n+1)}{365^n}$. This allows us to compute $\mathbb{P}[A]$ as a function of the number n of people. Of course, as n increases $\mathbb{P}[A]$ increases. In fact, with $n = 23$ people, you should be willing to bet that at least a pair of people have the same birthday, since then $\mathbb{P}[A]$ is larger than 50%. For $n = 60$ people, $\mathbb{P}[A]$ is over 99%.

3.7 The Monty Hall Problem

In an (in)famous 1970s game show hosted by Monty Hall, a contestant was shown three doors; behind one of the doors was a valuable prize (a car), and behind the other two were goats. The contestant picks a door (but does not open it). Then Hall's assistant (Carol), opens one of the other two doors, revealing a goat (since Carol knows where the prize is, she can always do this). The contestant is then given the option of sticking with his/her current door, or switching to the other unopened one. The contestant wins the prize if and only if their chosen door is the correct one. The question is: Does the contestant have a better chance of winning if he/she switches doors?

Intuitively, it seems obvious that since there are only two remaining doors after the host opens one, they must have equal probability. So you may be tempted to jump to the conclusion that it should not matter whether or not the contestant stays or switches. We will see that actually, the contestant has a better chance of picking the car if he or she uses the switching strategy. We will first give an intuitive pictorial argument, and then take a more rigorous probability approach to the problem.

To see why it is in the contestant's best interests to switch, consider the following. Initially when the contestant chooses the door, he or she has a $\frac{1}{3}$ chance of picking the car. This must mean that the other doors combined have a $\frac{2}{3}$ chance of winning. But after Carol opens a door with a goat behind it, how do the probabilities change? Well, the door the contestant originally chose still has a $\frac{1}{3}$ chance of winning, and the door that Carol opened has no chance of winning. What about the last door? It must have a $\frac{2}{3}$ chance of containing the car, and so the contestant has a higher chance of winning if he or she switches doors. This argument can be summed up nicely in the following picture:

3.6 生日悖论

生日悖论是一种研究群体中两人拥有相同生日可能性的显著现象。它被称为悖论并不是因为存在逻辑矛盾，而是因为它违背直觉。为了便于计算，我们将一年中的天数设为 365 天。于是 S = {1,...,365}，随机实验包括从 S 中抽取 n 个元素的样本，这些元素代表群体中 n 个人的出生日期。那么 |Ω| = 365^n。这是因为每个样本点是 n 个人可能生日的一个序列；序列中有 n 个位置，每个位置有 365 种可能取值。令 A 表示至少有一对人生日相同的事件。如果我们想确定 P[A]，或许更简单的方法是先计算 A 的补事件的概率，即 P[Ā]，其中 Ā 是指没有任何两个人生日相同的事件。由于 P[A] = 1−P[Ā]，我们就可以很容易地算出 P[A]。

我们再次处于一个均匀的概率空间中，因此我们只需确定 |Ā|。等价地，我们是在计算没有任何两个人生日相同的方案数。第一个人有 365 种选择，第二个人有 364 种，……，第 n 个人有 365−n+1 种选择，总共为 365×364×···× (365−n+1)。这仅仅是 Note 11 中计数第一法则的一个应用；我们进行的是无放回抽样且顺序重要。

总之，我们有

$$\mathbb{P}[\overline{A}] = \frac{|\overline{A}|}{|\Omega|} = \frac{365 \times 364 \times \cdots \times (365-n+1)}{365^n},$$

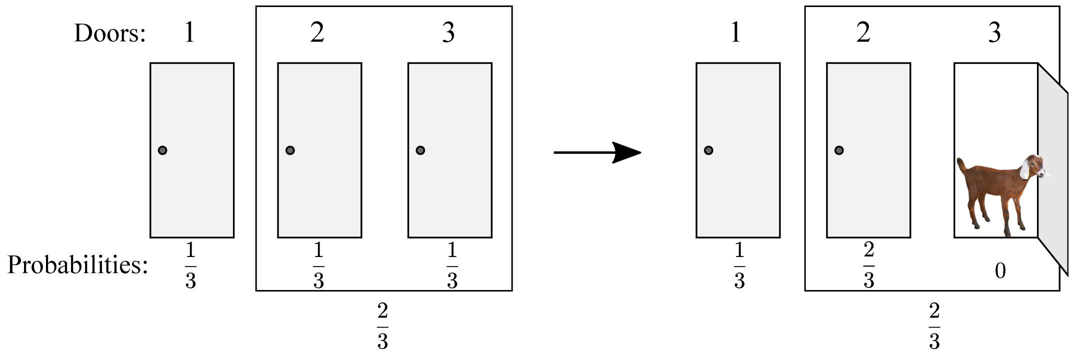
所以 P[A] = 1−P[Ā] = 1−365×364×···×(365−n+1) / 365^n。这使我们能够将P[A]表示为人数n的函数人的情况。当然，随着n增大，P[A]也会增大。事实上，当n=23时，你应该愿意打赌至少有一对人生日相同，因为此时P[A]已超过50%。当n=60时，P[A]超过99%。

3.7 蒙提霍尔问题

在一个（臭名昭著的）1970年代由蒙提·霍尔主持的游戏节目中，参赛者面前有三扇门；其中一扇门后有一件贵重奖品（一辆汽车），另外两扇门后是山羊。参赛者选择一扇门（但不打开）。然后主持人助手Carol打开另外两扇门中的一扇，露出一只山羊（由于Carol知道奖品的位置，她总能做到这一点）。接着参赛者可以选择坚持原来的选择，或者换到另一扇未打开的门。参赛者只有在其选择的门正确时才能赢得奖品。问题是：参赛者换门是否会提高获胜的机会？

直观上看，当主持人打开一扇门后只剩下两扇门，它们似乎应该有相等的概率。你可能因此倾向于认为参赛者是否换门并不重要。但我们将看到，实际上如果参赛者采用换门策略，他/她选中汽车的概率更高。我们先给出一个直观的图示论证，然后再用更严谨的概率方法来分析这个问题。

要理解为什么参赛者换门更有利，请考虑以下情况。最初当参赛者选择一扇门时，他/她只有1 3 的概率选中汽车。这意味着另外两扇门合起来有2 3 的获胜概率。但在Carol打开一扇后面是山羊的门之后，概率会发生什么变化？参赛者最初选择的那扇门仍然只有1 3 获胜的概率，而Carol打开的那扇门获胜的概率为零。那么剩下的那扇门呢？它有2/3的概率藏有汽车，因此如果参赛者换门，获胜的机会更高。这个论点可以用下面这张图清晰地总结出来：



What is the sample space here? Well, we can describe the outcome of the game (up to the point where the contestant makes his/her final decision) using a triple of the form (i, j, k), where $i, j, k \in \{1, 2, 3\}$. The values i, j, k respectively specify the location of the prize, the initial door chosen by the contestant, and the door opened by Carol. Note that some triples are not possible: e.g., $(1, 2, 1)$ is not, because Carol never opens the prize door. Thinking of the sample space as a tree structure, in which first i is chosen, then j, and finally k (depending on i and j), we see that there are exactly 12 sample points.

Assigning probabilities to the sample points here requires pinning down some assumptions:

- The prize is equally likely to be behind any of the three doors.
- Initially, the contestant is equally likely to pick any of the three doors.
- If the contestant happens to pick the prize door (so there are two possible doors for Carol to open), Carol is equally likely to pick either one. (Actually our calculation will have the same result no matter how Carol picks the door.)

From this, we can assign a probability to every sample point. For example, the point $(2, 1, 3)$ corresponds to the prize being placed behind door 2 (with probability $\frac{1}{3}$), the contestant picking door 1 (with probability $\frac{1}{3}$), and Carol opening door 3 (with probability 1, because she has no choice). So

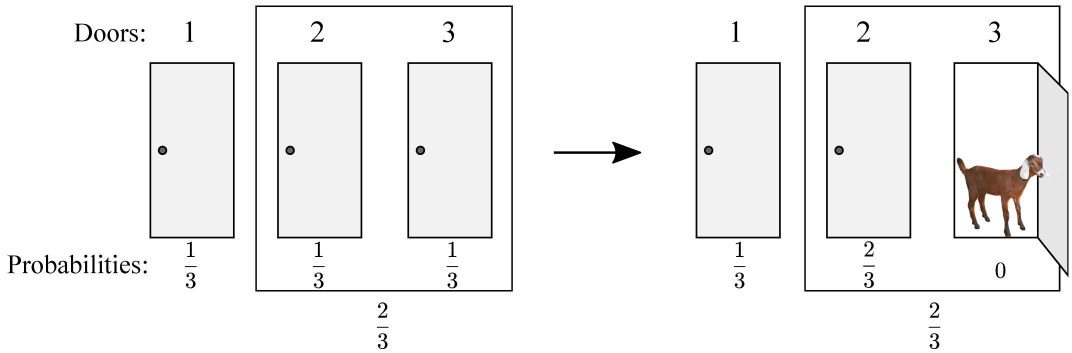
$$\mathbb{P}[(2,1,3)] = \frac{1}{3} \times \frac{1}{3} \times 1 = \frac{1}{9}.$$

[Note: Again we are multiplying probabilities here, without proper justification!] Note that there are six outcomes of this type, characterized by having $i \neq j$ (and hence k must be different from both). On the other hand, we have

$$\mathbb{P}[(1,1,2)] = \frac{1}{3} \times \frac{1}{3} \times \frac{1}{2} = \frac{1}{18}.$$

And there are six outcomes of this type, having $i = j$. These are the only possible outcomes, so we have completely defined our probability space. Just to check our arithmetic, we note that the sum of the probabilities of all outcomes is $(6 \times \frac{1}{9}) + (6 \times \frac{1}{18}) = 1$.

Let's return to the Monty Hall problem. Recall that we want to investigate the relative merits of the "sticking" strategy and the "switching" strategy. Let's suppose the contestant decides to switch doors. The event W we are interested in is the event that the contestant wins. Which sample points (i, j, k) are in W? Well, since the contestant is switching doors, their initial choice j cannot be equal to the prize door, which is i. And all outcomes of this type correspond to a win for the contestant, because Carol must open the second non-prize door, leaving the contestant to switch to the prize door. So W consists of all outcomes of the first type in our earlier analysis; recall that there are six of these, each with probability $\frac{1}{9}$. So $\mathbb{P}[W] = \frac{6}{9} = \frac{2}{3}$. That is,



这里的样本空间是什么？我们可以用一个三元组(i, j,k)来描述游戏的结果（直到参赛者做出最终决定为止），其中i, j,k ∈{1,2,3}。i, j,k分别表示奖品的位置、参赛者最初选择的门以及Carol打开的门。注意有些三元组是不可能的：例如(1,2,1)就不可能，因为Carol不会打开有奖品的门。将样本空间想象成一个树状结构，首先是i被选定，然后是j，最后是k（取决于i和j），我们会发现恰好有12个样本点。

为这里的样本点分配概率需要明确一些假设：

- 奖品出现在三扇门中任意一扇后面的概率是相等的。
- 最初，参赛者选择三扇门中任意一扇的可能性是相等的。
- 如果参赛者恰好选中了有奖品的门（因此Carol有两个门可选），她选择任一扇门的可能性是相等的。（实际上，无论Carol如何选择门，我们的计算结果都相同。）

从这一点出发，我们可以为每个样本点分配一个概率。例如，点(2,1,3)对应奖品放在2号门后（概率为$\frac{1}{3}$），参赛者选择1号门（概率为$\frac{1}{3}$），而Carol打开3号门（概率为1，因为她别无选择）。所以

$$\mathbb{P}[(2,1,3)] = \frac{1}{3} \times \frac{1}{3} \times 1 = \frac{1}{9}.$$

[注：这里我们再次在没有充分依据的情况下相乘概率！] 注意这类结果共有六个，其特征是i = j（因此k必须与两者都不同）。另一方面，我们有

$$\mathbb{P}[(1,1,2)] = \frac{1}{3} \times \frac{1}{3} \times \frac{1}{2} = \frac{1}{18}.$$

这类结果共有六个，其中i = j。这些是所有可能的结果，因此我们已经完整定义了我们的概率空间。为了验证计算，我们注意到所有结果的概率之和为$(6\times\frac{1}{9}) + (6 \times \frac{1}{18}) = 1$.

让我们回到蒙提霍尔问题。回想一下，我们想研究坚持原门和换门这两种策略的相对优劣。假设参赛者决定换门。我们关心的事件W是参赛者获胜的事件。哪些样本点(i, j,k)属于W？由于参赛者会换门，他们最初的选择j不能等于奖品所在的门i。所有这类结果都对应参赛者获胜，因为Carol必须打开另一扇没有奖品的门，使得参赛者最终换到有奖品的门。因此W包含了我们之前分析中第一类的所有结果；回忆一下，这类结果共有六个，每个的概率为$\frac{1}{9}$。所以 P[W] = $\frac{6}{9} = \frac{2}{3}$。也就是说，

using the switching strategy, the contestant wins with probability $\frac{2}{3}$. It should be intuitively clear (and easy to check formally — try it!) that under the sticking strategy their probability of winning is $\frac{1}{3}$. (In this case, the contestant is really just picking a single random door.) So by switching, the contestant actually improves their odds by a huge amount!

4 Summary

The Monty Hall example well illustrates the importance of doing probability calculations systematically, rather than "intuitively." Recall the key steps in all our calculations:

- What is the <u>sample space</u> (i.e., the experiment and its set of possible outcomes)?
- What is the <u>probability</u> of each outcome (sample point)?
- What is the <u>event</u> we are interested in (i.e., which subset of the sample space)?
- Finally, compute the probability of the event by adding up the probabilities of the sample points contained in it.

Whenever you meet a probability problem, you should always go back to these basics to avoid potential pitfalls. Even experienced researchers make mistakes when they forget to do this — witness many erroneous "proofs", submitted by mathematicians to newspapers at the time, of the claim that the switching strategy in the Monty Hall problem does not improve the odds.

使用换门策略，参赛者以2的概率获胜 3。这在直觉上应该是显而易见的（并且很容易正式验证——试试看！）在坚持原选择的策略下，他们获胜的概率是1 3。（在这种情况下，参赛者实际上只是随机选择了一扇门。）因此，通过换门，参赛者的获胜几率实际上大幅提高了！

4 总结

蒙提霍尔的例子很好地说明了进行概率计算时系统化的重要性，而非依赖直觉。回顾我们在所有计算中的关键步骤：

- 样本空间是什么（即实验及其所有可能结果的集合）？
- 每个结果（样本点）的概率是多少?
- 我们感兴趣的事件是什么（即样本空间中的哪个子集）？
- 最后，通过将事件所包含的样本点的概率相加，计算出该事件的概率。

每当你遇到一个概率问题时，都应该回到这些基本原则上，以避免潜在的陷阱。即使是经验丰富的研究人员，如果忘记这样做也可能会犯错——例如，在蒙提霍尔问题中，当时许多数学家向报纸提交了错误的证明，声称换门策略并不会提高获胜几率。