Chapter 1
Events, Probabilities and Random Variables

1 Introduction

The field of probability can be traced back to the 17th century when it arose out of the study of gambling games. Its range of applications extends beyond games into business decisions, insurance, law, medical tests, and the social sciences. The stock market, "the largest casino in the world", cannot do without it. The telephone network, call centres, and airline companies with their randomly fluctuating loads could not have been economically designed without probability theory.

The aim of this chapter is to explore some important uses of probabilistic arguments to a number of stylised problems. To this end we will make extensive use of a number of fundamental tools from probability theory to analyse them. Although we leave the formal definition of random processes to later chapters, all the examples introduced below can be seen as random processes, they deal with sequences of random instances.

The chapter is organised in sections each dealing with a specific setting of interest. Each section is generally self-contained. Sections 4, 6 and 9 review the notions that have been used in the preceding sections which you are likely to be already familiar with.

Sections 2 and 8 address the issue of *generating* some random variable given an available source of randomness. Section 3 gives a glimpse of the potential of applying probability to perform more *efficient computations* focusing on the example of matrix multiplication. Section 5 introduces and analyses the *coupon collector problem*, a natural problem that arises in a number of settings where one wants to make an inventory of the identities of events occurring at random (e.g. system failures, security breaches, file swarming in peer-to-peer file sharing systems like BitTorrent, load balancing). In Section 7, we briefly discuss *inequalities for bounding random variables* and illustrate their use in the context of the coupon collector problem of Section 5.

2 Tossing a biased coin

In a number of applications of probability theory, we need to generate random numbers. In general, we only dispose of a source that provides pseudo-random number generators which happen to be good enough for most applications. Yet when more secure sources of randomness are required as for generating cryptographic keys, these pseudo-random sources may be broken. This is what happened to the Netscape browser in the 1990s when people found out how their pseudo-numbers were created.

An alternative way to using pseudo-random number generators is to use a physical source of randomness as a starting point. However one might not be able to accurately determine the distribution underlying this randomness. As a matter of example assume that we dispose of source of randomness that generates two possible outcomes. For example a possibly biased coin that one can toss without knowing the probability of obtaining heads of tails that we will denote by p and $1-p$, $p \in (0,1)$.

Our aim in this introductory example is to come up with a procedure that will enable us to generate an unbiased random variable that is equal to 1 with probability $1/2$ and to 0 with probability $1/2$. How should we proceed?

As a matter of example, we could take advantage of the symmetries of the problem and toss the coin twice. If it comes first heads and tails second (HT) we call if 0, and if it comes first tails and heads second (TH) we call it 1 and we discard the two other possible outcomes, i.e. HH and TT and toss the coin again twice until we either get HT or TH.

It is clear that the probability of getting either HT or TH is given by $p(1-p)$ and by the definition of conditional probabilities

$$\mathbf{P}(HT \mid HT \text{ or } TH) = \frac{\mathbf{P}\left(HT \text{ and } (HT \text{ or } TH)\right)}{\mathbf{P}(HT \text{ or } TH)} = \frac{p(1-p)}{p(1-p)+p(1-p)} = \frac{1}{2}\,.$$

where we used the fact that $\mathbf{P}\left(HT \text{ and } (HT \text{ or } TH)\right) = \mathbf{P}(HT)$ and $\mathbf{P}(HT \text{ or } TH) = \mathbf{P}(HT) + \mathbf{P}(TH)$. Similarly one can show that $\mathbf{P}(TH \mid HT \text{ or } TH) = 1/2$. This solution was proposed by von Neumann.

Note that the fair (unbiased) bit is obtained after some average/expected number of flips t. How many flips does it take?

It we let z be the probability of succeeding in getting a fair bit, i.e. obtaining HT or TH then

$$t = 2/z = \frac{2}{2p(1-p)} = \frac{1}{p(1-p)}\,.$$

Alternatively, we know that we have to start with 2 flips and then we succeed with probability $2p(1-p)$ and if not we have to repeat the process all over again taking us an extra t flips on average, i.e.

$$t = 2 + [1 - 2p(1-p)]t \quad \Rightarrow \quad t = \frac{1}{p(1-p)}\,.$$

In the above setting, we do not need to know p to generate the fair bit but we may have to wait for sometime before getting it. Assume that $p = 2/3$ for which $t = 9/2$. Can one do better knowing that $p = 2/3$?

For $p = 2/3$, we know that

$$\mathbf{P}(HH) = 4/9,\ \mathbf{P}(TH) = \mathbf{P}(HT) = 2/9,\ \mathbf{P}(TT) = 1/9\,.$$

One could do better, i.e. find a quicker way on average to get a fair bit by saying that 1 corresponds to HH and 0 corresponds to obtaining either HT or TH both of which occur with the same probability $4/9$ requiring an average of $2/(8/9) = 9/4$ flips instead of $9/2$ with the initial procedure. Note though that we require the knowledge of p which might be too much to ask!

Let us now go back to our general setting and see if can generate more (independent) fair random bits for a biased source of randomness. To do this let us focus on an example. Assume that we tossed the coin a certain number of times and we obtained the following sequence

$$H\,H \mid T\,T \mid H\,T \mid H\,H \mid H\,T \mid H\,H \mid H\,T \mid T\,H$$

First we extract four independent random bits from occurrences of HT and TH which gives us $0, 0, 0, 1$ from the third, fifth, seventh and eighth pairs of flips. Let us now look at the remaining flips

$$H\,H \mid T\,T \mid H\,H \mid H\,H$$

We rewrite this sequence by clumping together the pairs of heads and the pairs of tails into one H_1 and one T_1 respectively to obtain

$$H_1\,T_1 \mid H_1 H_1$$

and we repeat the above process with this new sequence extracting independent fair bits when we encounter HT and TH and holding on to the HH and TT pairs for the next iteration of our procedure. Is there additional information we have not used from the initial sequence?

So far we extracted fair bits from HT and TH pairs and used the remaining tosses to create a new sequence of H and T. But we did not use the information that connects these two sequences that is to say whether the pair of flips are equal or different in the original sequence. Using this information, we create the following third sequence independent of the two others, where we denote by a H_A the pairs that are equal and T_A when they are distinct. The initial sequence gives rise to the following sequence

$$H_A\,H_A \mid T_A\,H_A \mid T_A\,H_A \mid T_A\,T_A$$

Subsequently we keep extracting from this sequence new sequences as before.

To sum up, we start from sequence

$$H\,H \mid T\,T \mid H\,T \mid H\,H \mid H\,T \mid H\,H \mid H\,T \mid T\,H$$

extract the bits $0, 0, 0, 1$ and define two new sequences

$$\text{Sequence 1} \quad H\,T \mid HH$$

and

$$\text{Sequence A} \quad H\,H \mid T\,H \mid T\,H \mid T\,T$$

- from Sequence 1 we extract the bit 0 and define two new sequences: sequence 2: H from which we cannot extract a new bit and sequence 1A: TH from which we extract 1,
- from sequence A, we extract the bits $1, 0$ and define two new sequences: sequence A1: HT from which we cannot extract a new bit and sequence B: $HT \mid TH$, and then repeat...

Let $A(p)$ be the expected number of unbiased flips per biased flip one can get when the bias is p

- First we get the mismatched pairs HT and TH from the original sequence of flips with probability $2p(1-p)$.
- Then we create the sequence 1 using pairs of flips that are the same which happens with probability $p^2 + (1-p)^2$ and where H_1 appears with probability

  $$\mathbf{P}(H_1) = \mathbf{P}(HH \mid HH \text{ or } TT) = \frac{p^2}{p^2 + (1-p)^2}$$

  and T_1 with probability

  $$\mathbf{P}(T_1) = \mathbf{P}(TT \mid HH \text{ or } TT) = \frac{(1-p)^2}{p^2 + (1-p)^2}$$

- and a new sequence A obtained by transforming two flips from the initial sequence into one flip where H_A appears with probability

  $$\mathbf{P}(H_A) = \mathbf{P}(HH \text{ or } TT) = p^2 + (1-p)^2$$

  and T_A with probability

  $$\mathbf{P}(T_A) = \mathbf{P}(TH \text{ or } TH) = 2p(1-p)$$

Putting all this together we have

$$A(p) = p(1-p) + \frac{1}{2}\left(p^2 + (1-p)^2\right) A\left(\frac{p^2}{p^2 + (1-p)^2}\right) + \frac{1}{2} A\left(p^2 + (1-p)^2\right)$$

Unsurprisingly $A(1/2) = 1$ as for $p = 1/2$ the coin is already unbiased.

Finally, you might have come across the notion of entropy where for the biased coin with bias p, the entropy is given by

$$H(p) = -p \log_2 p - (1-p) \log_2 (1-p)$$

After some tedious calculations (see below), one can verify that $H(p) = A(p)$.

To simplify the notations let $q = 1 - p$. First,

$$\begin{aligned}
\frac{1}{2} H(p^2 + q^2) &= -\frac{1}{2}(p^2 + q^2) \log_2((p^2 + q^2)) - \frac{1}{2}(1 - p^2 - q^2) \log_2\left(1 - p^2 - q^2\right) \\
&= -\frac{1}{2}(p^2 + q^2) \log_2\left(p^2 + q^2\right) - pq \log_2(2pq) \\
&= -\frac{1}{2}(p^2 + q^2) \log_2\left(p^2 + q^2\right) - pq - pq \log_2(p) - pq \log_2(q)\,,
\end{aligned}$$

since $1-p^2-q^2=2pq$, $\log_2(2pq)=\log_2 2+\log_2 p+\log_2 q$ and $\log_2 2=1$. Besides,

$$\begin{aligned}\frac{p^2+q^2}{2}H\left(\frac{p^2}{p^2+q^2}\right) &= -\frac{1}{2}p^2\log_2\left(\frac{p^2}{p^2+q^2}\right)-\frac{1}{2}q^2\log_2\left(\frac{q^2}{p^2+q^2}\right)\\ &= -\frac{1}{2}p^2\log_2 p^2+\frac{1}{2}p^2\log_2\left(p^2+q^2\right)-\frac{1}{2}q^2\log_2 q^2+\frac{1}{2}q^2\log_2\left(p^2+q^2\right)\\ &= -p^2\log_2 p-q^2\log_2 q+\frac{1}{2}(p^2+q^2)\log_2\left(p^2+q^2\right)\ .\end{aligned}$$

Finally, recall that $p+q=1$, so that

$$\begin{aligned}pq+\frac{1}{2}\left(p^2+q^2\right)H\left(\frac{p^2}{p^2+q^2}\right)+\frac{1}{2}H\left(p^2+q^2\right) &= -pq\log_2 p-pq\log_2 q-p^2\log_2 p-q^2\log_2 q\\ &= -p(p+q)\log_2 p-q(p+q)\log_2 q\\ &= -p\log_2 p-p\log_2 p=H(p)\ .\end{aligned}$$

Example We are given three coins and are told that two of them are fair and the third is biased landing heads with probability 2/3, but we are not told which of the three coins is biased. Assume that we take the three coins label them coin 1, coin 2 and coin 3. After flipping them, the first and second coin show heads and the third coin shows tails. What is the probability that the first coin is the biased one?

Remember that before the toss, each of the three coins is equally likely to be the biased one. Let E_i be the event that the ith coin is the biased one and let B be the event that the three coin flips show HHT. Before we flip the coins $\mathbf{P}(E_i)=1/3$. Moreover

$$\mathbf{P}(B\mid E_1)=\mathbf{P}(B\mid E_2)=\frac{2}{3}\frac{1}{2}\frac{1}{2}=\frac{1}{6},\quad \mathbf{P}(B\mid E_3)=\frac{1}{2}\frac{1}{2}\frac{1}{3}=\frac{1}{12}\ .$$

Applying Bayes' rule, we have

$$\mathbf{P}(E_1\mid B)=\frac{\mathbf{P}(E_1\cap B)}{\mathbf{P}(B)}=\frac{\mathbf{P}(B\mid E_1)\mathbf{P}(E_1)}{\sum_{i=1}^3\mathbf{P}(B\mid E_i)\mathbf{P}(E_i)}=\frac{2}{5}\ .$$

What this tells us is that the outcome of the three coin flips increases the likelihood that the first coin is the biased one from 1/3 to 2/5!

3 Verifying matrix multiplication

In this section, we will illustrate an instance where randomness can enable us to check a matrix identity. Let A,B,C three square n by n matrices with binary (0 or 1) entries. We would like to check if the identity

$$AB=C$$

holds. One simple way of proceeding is to compute the product AB at the cost of about $2n^3$ operations (additions and multiplications)[1]. Can one do better?

Alternatively, we could take a vector (or many vectors) $r\in\{0,1\}^n$ and check whether $A(Br)=Cr$ if it is not the case then we are guaranteed that the inequality does not hold if however $ABr=Cr$ it might be that we were lucky despite the fact that $AB\neq C$ but we do not know if the identity is true or false for sure.

The main advantage of doing so is that each of the matrix-vector mutiplications Br, $A(Br)$ and Cr takes about $2n^2$ operations with a total of about $6n^2$ operations instead of $2n^3$ which is very convenient if n is very large.

Let us assume that we generate a random vector $r\in\{0,1\}^n$ which we can do by generating each of the entries $r_1,r_2,\ldots,r_n$ of r through random fair bits as in the previous section say, i.e. $\mathbf{P}(r_i=1)=\mathbf{P}(r_i=0)=1/2$.

[1] There are more sophisticated techniques that take slightly less though.

For convenience let us define $D = AB - C$. If the identity $AB = C$ does not hold then $D \neq 0$ so it must have a nonzero entry, let that entry be d_{11}.

If we happen to choose a vector such that $Dr = 0$ then

$$\sum_{i=1}^{n} d_{1i} r_i = 0 \quad \Rightarrow \quad r_1 = -\frac{\sum_{i=2}^{n} d_{1i} r_i}{d_{11}} .$$

In particular the event $\{ABr = Cr\}$ implies that $\{r_1 = -\frac{\sum_{i=2}^{n} d_{1i} r_i}{d_{11}}\}$, i.e.

$$\{ABr = Cr\} \subset \{r_1 = -\frac{\sum_{i=2}^{n} d_{1i} r_i}{d_{11}}\} \quad \Rightarrow \quad \mathbf{P}(ABr = Cr) \leq \mathbf{P}\left(r_1 = -\frac{\sum_{i=2}^{n} d_{1i} r_i}{d_{11}} \right) .$$

Using law of total probability, we have that

$$\begin{aligned}
\mathbf{P}(ABr = Cr) &\leq \mathbf{P}\left(r_1 = -\frac{\sum_{i=2}^{n} d_{1i} r_i}{d_{11}} \right) \\
&= \sum_{(x_2,\ldots,x_n) \in \{0,1\}^{n-1}} \mathbf{P}\left(\left(r_1 = -\frac{\sum_{i=2}^{n} d_{1i} r_i}{d_{11}} \right) \cap (r_2 = x_2, \ldots, r_n = x_n) \right) \\
&= \sum_{(x_2,\ldots,x_n) \in \{0,1\}^{n-1}} \mathbf{P}\left(\left(r_1 = -\frac{\sum_{i=2}^{n} d_{1i} r_i}{d_{11}} \right) \mid (r_2 = x_2, \ldots, r_n = x_n) \right) \\
&\quad \mathbf{P}\left(r_2 = x_2, \ldots, r_n = x_n \right) \\
&= \sum_{(x_2,\ldots,x_n) \in \{0,1\}^{n-1}} \mathbf{P}\left(r_1 = -\frac{\sum_{i=2}^{n} d_{1i} x_i}{d_{11}} \right) \mathbf{P}\left(r_2 = x_2, \ldots, r_n = x_n \right)) \\
&= \sum_{(x_2,\ldots,x_n) \in \{0,1\}^{n-1}} \frac{1}{2} \mathbf{P}\left(r_2 = x_2, \ldots, r_n = x_n \right) = \frac{1}{2}
\end{aligned}$$

For the penultime equlity we used the independence of r_1 from the other r_is and the fact that the expression $\frac{\sum_{i=2}^{n} d_{1i} x_i}{d_{11}}$ is completely determined since the the x_is are not random, i.e. it is a number $\alpha \in \{0, 1\}$ and we know that $\mathbf{P}(r_1 = \alpha) = 1/2$.

So the probability of obtaining the equality $ABr = Cr$ despite having $AB \neq C$ is smaller than $1/2$. This already not too bad but this is still a rather large bound for an error probability. How can we reduce it?

We can run the algorithm multiple times drawing a fresh random vector r for each run and after k independent runs the probability of error becomes smaller than 2^{-k}. Note that repeated computations increase the number of operations to $6kn^2$. For $k = 100$ the error is very small without blowing up the computational budget which states of the order of n^2 operations.

Let us analyse our confidence in the algorithm as we repeat the randomised test. Let E be the event that the identity is correct and let B be the event that the test returns that the identity is correct. Initially $\mathbf{P}(E) = \mathbf{P}(E^c) = 1/2$ and we showed that $\mathbf{P}(B \mid E^c) \leq 1/2$ and we know that $\mathbf{P}(B \mid E) = 1$ since if the identity is verified then the test will always be correct.

By Bayes' rule,

$$\mathbf{P}(E \mid B) = \frac{\mathbf{P}(B \mid E)\mathbf{P}(E)}{\mathbf{P}(B \mid E)\mathbf{P}(E) + \mathbf{P}(B \mid E^c)\mathbf{P}(E^c)} \geq \frac{1/2}{1.1/2 + 1/2.1/2} = 2/3 .$$

After the first run, if the test returns that the identity is correct, we know that the identity is likely to be verified with probability more than $2/3$ and it has a probability of less than $1/3$ not to be verified.

Iterating the above argument 100 times and if all the 100 calls to the matrix identity test returns that the identity is correct, our confidence in the correctness of the identity is at least $1 - 1/(2^{100} + 1)$. Try this yourself.

4 Events and their probabilities

In the previous two sections, we introduced a number of important notions of probability theory, namely (i) Bernoulli random variables that take one of two values say 0 and 1 with probability p and $1-p$, (ii) Geometric random variables which are the time taken for the first appearance of 1 in a sequence of independent 0 and 1 trials, (iii) unions and intersections of events and their probabilities and (iv) conditional probabilities.

In this section we provide some mathematical definitions for the sake of completeness.

Definition 1 *A collection $\mathcal{F}$ of subsets of Ω is called a σ-field if it satisfies the following conditions:*

(*i*) $\emptyset \in \mathcal{F}$;

(*ii*) *If $A_1, A_2, \ldots$ are in $\mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$;*

(*iii*) *If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$*

A σ-field is, in other words, a collection of subsets of Ω that contains the empty set and that is "closed" under the operations of taking countable intersections and complements.

Definition 2 *A probability measure $\mathbf{P}$ on a probability space $(\Omega, \mathcal{F})$ is a function $\mathbf{P} : \mathcal{F} \to [0,1]$ that satisfies the following conditions:*

(*i*) $\mathbf{P}(\emptyset) = 0$ *and* $\mathbf{P}(\Omega) = 1$;

(*ii*) *If $A_1, A_2, \ldots$ is a collection of disjoint members of $\mathcal{F}$, i.e such that $A_i \cap A_j = \emptyset$, for all pairs i and j, $i \neq j$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$;*

$$\mathbf{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbf{P}(A_i) \, .$$

Finally, we state the law of total probability and Bayes' law.

Lemma 1 *For any events A and B such that $0 < \mathbf{P}(B) < 1$,*

$$\mathbf{P}(A) = \mathbf{P}(A \cap B) + \mathbf{P}(A \cap B^c) = \mathbf{P}(A \mid B)\mathbf{P}(B) + \mathbf{P}(A \mid B^c)\mathbf{P}(B^c) \, .$$

More generally, let let $B_1, \ldots, B_n$ a partition of Ω, i.e. for all $i \neq j$, $B_i \cap B_j = \emptyset$, and $\bigcup_{i=1}^{n} B_i = \Omega$, such that $\mathbf{P}(B_i) > 0$ for all i. then

$$\mathbf{P}(A) = \sum_{i=1}^{n} \mathbf{P}(A \mid B_i)\mathbf{P}(B_i) \, .$$

Moreover

$$\mathbf{P}(B_j \mid A) = \frac{\mathbf{P}(B_j \cap A)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A \mid B_j)\mathbf{P}(B_j)}{\sum_{i=1}^{n} \mathbf{P}(A \mid B_i)\mathbf{P}(B_i)} \, .$$

5 The coupon collector problem

We consider an urn containing n balls numbered $1, \ldots, n$. We proceed at discovering the balls from the urn by, each time, drawing a ball, writing its number and putting it back in the urn. We suppose that the draws are independent from each other. We stop the process when we see each ball (number) at least once. This process is known as the *coupon collector problem*. Despite its apparent simplicity this problem highlights a number of interesting probabilistic results, in particular for the analysis of randomised algorithms.

We want to characterise the time X it will take to see each of the balls at least once. Let X_i be the number of draws before we see i different balls once we have seen $i-1$ different balls, for $i = 1, \dots n$ with $X_1 = 1$. It is not difficult to see that

$$X = \sum_{i=1}^{n} X_i \, .$$

When $i-1$ different balls have been observed, the probability of obtaining a new ball in the following draw is

$$p_i = \frac{n-(i-1)}{n} = 1 - \frac{i-1}{n} \, ,$$

so that the random variable X_i is a geometric random variable with parameter p_i, i.e.

$$\mathbf{P}(X_i = k) = (1-p_i)^{k-1} p_i, \quad k = 1, 2, \dots \, .$$

and on average it takes

$$\mathbf{E}(X_i) = \sum_{k \geq 1} k \mathbf{P}(X_i = k) = \frac{1}{p_i} = \frac{n}{n-i+1}$$

draws before we see the ith number where $\mathbf{E}(X_i)$ is the expectation of X_i. Using the linearity of the expectation, we have that, on average the process takes $\mathbf{E}X$ to complete where

$$\mathbf{E}(X) = \mathbf{E}\left(\sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \mathbf{E} X_i = \sum_{i=1}^{n} \frac{n}{n-i+1} = n \sum_{i=1}^{n} \frac{1}{i} \, .$$

For n large, we have

$$\mathbf{E}(X) \approx n \log n \, .$$

Assume that we relax our objective to just observing half the balls. How long does it take on average?

Similarly, it is not difficult to see that it takes (let n be even for simplicity)

$$\begin{aligned} \sum_{i=1}^{n/2} \frac{n}{n-i+1} &= n \sum_{i=(n/2)+1}^{n} \frac{1}{i} \\ &= n \left(\sum_{i=1}^{n} \frac{1}{i} - \sum_{i=1}^{n/2} \frac{1}{i} \right) \\ &\approx n \log(n) - \log(n/2) \\ &= n \log(2) \, . \end{aligned}$$

6 Discrete random variables and their distributions

We review some important definitions related to discrete random variables

- A random variable X is discrete if it takes values in some countable set $\{x_1, x_2, \dots\}$. The probability (mass) function of X is the function $f_X : \{x_1, x_2, \dots\} \to [0,1]$ that to a value x that can be taken by X associates its corresponding probability $f_X(x) = \mathbf{P}(X = x)$.
- Two discrete random variables X and Y are independent if the events $\{X = x\}$ and $\{Y = y\}$ are independent for all the values x and y that can be taken by X and Y.
- The mean value, expectation or expected value of the random variable X is defined by $\mathbf{E}(X) = \sum_x x \mathbf{P}(X = x)$.
- The variance of X is defined by $\mathbf{Var}(X) = \mathbf{E}\left[(\mathbf{E}(X) - X)^2 \right]$.

Examples of discrete distributions We now present the most important discrete random variables' distributions.

- Bernoulli distribution, Ber(p), where $\mathbf{P}(X=0)=1-\mathbf{P}(X=1)=p$ Introduced in Section 6 Chapter 1,
$$\mathbf{E}X=p,\ \mathbf{Var}(X)=p(1-p)\,.$$
Similarly, for indicator I_A the indicator function of the event A, we have $\mathbf{E}[I_A]=\mathbf{P}(A)$, $\mathbf{Var}(I_A)=\mathbf{P}(A)(1-\mathbf{P}(A))$.

- Binomial distribution, Bin(n,p), counts the number of heads out of n independent tosses of biased coin,
$$\mathbf{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\quad \mathbf{E}X=np,\quad \mathbf{Var}(X)=np(1-p)\,.$$

- Poisson distribution, Po(λ), Let Y be a Bin(n,p) random variable when n becomes large while p tends to 0 in such a way that np approaches a non-zero constant λ, we have, for k fixed
$$\mathbf{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k}\approx\frac{1}{k!}\left(\frac{np}{1-p}\right)^k(1-p)^n\to\frac{\lambda^k}{k!}e^{-\lambda}\,.$$
Therefore
$$\mathbf{E}X=\mathbf{Var}(X)=\lambda\,.$$

- Geometric distribution, Ge(p), we have
$$\mathbf{P}(X=k)=p(1-p)^{k-1},\quad \mathbf{E}X=\frac{1}{p},\quad \mathbf{Var}(X)=\frac{1-p}{p^2}\,.$$
Moreover, $\mathbf{P}(X>n)=(1-p)^n$.

7 Bounding random variables

In the previous section we computed the average values of random variables, namely the time for the coupon collection to complete. Can one say more? In particular, given a random variable X we call the tail distribution of X the probability that X is larger than some constant x. Can we use the knowledge about the mean/average of X to bound this tail distribution, i.e. to say that the probability that the random variable is large is smaller than some constant?

For example suppose we are told that the average price of some object (a house, a share...) is 10. Can we evaluate the probability that it is higher that 100?

Note that
$$\mathbf{E}X=\sum_{k\geq 0}k\mathbf{P}(X=k)\leq\sum_{k\geq 100}k\mathbf{P}(X=k)\leq 100\sum_{k\geq 100}\mathbf{P}(X=k)=100\mathbf{P}(X\geq 100)$$
and the answer is that there is 10% chance that the price is higher than 100!

This is in fact the proof of a more general result known as *Markov's inequality*:

Theorem 1 *For $x>0$, we have*
$$\mathbf{P}(|X|\geq x)\leq\frac{\mathbf{E}|X|}{x}\,.$$

Let us apply this result in the context of the coupon collector problem of the previous section. Recall that X is the number of draws required before we see all the different balls whose expectation is

$$\mathbf{E}(X) = nH_n$$

where $H_n = \sum_{i=1}^{n} 1/i$. Markov's inequality yields

$$\mathbf{P}(|X - H_n| \geq nH_n) \leq \mathbf{P}(X \geq 2nH_n) \leq \frac{\mathbf{E}X}{2nH_n} \leq \frac{nH_n}{2nH_n} = \frac{1}{2}\,,$$

where in the first inequality we used the fact that $\mathbf{E}X = nH_n$ and the fact that

$$\{|X - nH_n| \geq nH_n\} = \{X - nH_n \geq nH_n\} \cup \{X - nH_n \leq -nH_n\}\ .$$

In this very setting the inequality is in fact an equality since $X > 0$ by definition and so the event $\{X - nH_n \leq -nH_n\} = \{X \leq 0\}$ is the empty event.

This does not seem like a great result. Can we do better?

There is a more refined version of Markov's inequality known as *Chebyshev's inequality*. It states that

Theorem 2 *For $x > 0$,*

$$\mathbf{P}(|X - \mathbf{E}X| \geq x) \leq \frac{VarX}{x^2}\,,$$

Proof Applying Markov's inequality to the random variable $(X - \mathbf{E}X)^2$ we get

$$\mathbf{P}(|X - \mathbf{E}X| \geq x) = \mathbf{P}((X - \mathbf{E}X)^2 \geq x^2) \leq \frac{\mathbf{E}\left((X - \mathbf{E}(X))^2\right)}{x^2} = \frac{\mathrm{Var}X}{x^2}\,,$$

QED

Let us apply it to the coupon collector problem.

Each of the random variables X_i is Geometric with parameter $p_i = \frac{n-i+1}{n}$. Since the X_is are independent, we have

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_{i=1}^{n} \mathrm{Var}(X_i) \\
&= \sum_{i=1}^{n} \frac{1 - p_i}{p_i^2} \\
&\leq \sum_{i=1}^{n} \frac{1}{p_i^2} \\
&= \sum_{i=1}^{n} n^2 \frac{1}{i^2} \\
&\leq \frac{\pi^2}{6} n^2
\end{aligned}$$

where, for the last inequality, we used the identity $\sum_{i \geq 1} \frac{1}{i^2} = \frac{\pi^2}{6}$.

Now applying Chebyshev's inequality we obtain that, for large n

$$\mathbf{P}(X \geq 2nH_n) \leq \mathbf{P}\left(|X - nH_n| \geq nH_n\right) \leq \frac{n^2\pi^2/6}{n^2 H_n^2} \approx \frac{\pi^2}{6\log^2 n}\,,$$

which is a much better bound than the one obtained by Markov's inequality as

$$\lim_{n\to\infty} \frac{\pi^2}{6\log^2 n} = 0\,.$$

Let us now examine the process in more detail to try to do even better. For this let us introduce $Y_{k,j}$ the indicator function that the ball j is not chosen in one of the first k draws. It is not difficult to see that

$$\mathbf{P}(Y_{k,j}=1) = (\mathbf{P}(\ j \text{ does not appear in a draw}))^k = \left(\frac{n-1}{n}\right)^k\,.$$

Now consider the probability of not obtaining the jth ball after $n\log n + dn$ draws. This probability is

$$\mathbf{P}(Y_{n\log n+dn,j}=1) = \left(1-\frac{1}{n}\right)^{n(\log n+d)} \leq e^{-(\log n+d)} = \frac{1}{e^d n} \tag{1}$$

where we used the identity $1-x \leq e^{-x}$ for $x \geq 0$.

Let

$$U_k = \sum_{i=1}^{n} Y_{k,i}$$

be the number of balls that are not drawn until the k-th draw. The probability that some ball has not been seen after $n\log n + dn$ draws is

$$\begin{aligned}
\mathbf{P}(U_{n\log n+dn} > 0) &= \mathbf{P}\left(\bigcup_{i=1}^{n}\{Y_{n\log n+dn,i}=1\}\right)\\
&\leq \sum_{i=1}^{n}\mathbf{P}\left(Y_{n\log n+dn,i}=1\right)\\
&\leq n\frac{1}{e^d n}\\
&= e^{-d}\,.
\end{aligned}$$

where we used equation (1) and the union bound, i.e.

$$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n}\mathbf{P}\left(A_i\right)\,.$$

If we let $d=\log n$, the above inequality tells us that the probability that not all the coupons have been collected after $2n\log n$ steps is at most $e^{-\log n} = 1/n$, i.e.

$$\mathbf{P}(X \geq 2n\log(n)) \leq 1/n$$

which is significantly better than what can be achieved with Chebyshev's inequality.

To finish the section, we examine what happens if we only wait for a linear number of draws. Recall that

$$\mathbf{P}(Y_{k,j}=1) = \left(\frac{n-1}{n}\right)^k\,.$$

so that

$$\mathbf{E}Y_{k,j} = \left(\frac{n-1}{n}\right)^k\,.$$

Also note that for distinct i an j

$$\mathbf{P}(Y_{k,i}Y_{k,j}=1)=(\mathbf{P}(\text{neither } i \text{ nor } j \text{ appear in a draw}))^k=\left(\frac{n-2}{n}\right)^k .$$

so that

$$\mathbf{E}(Y_{k,i}Y_{k,j})=\left(\frac{n-2}{n}\right)^k .$$

This implies that

$$\mathrm{Cov}(Y_{k,i},Y_{k,j})=\left(\frac{n-2}{n}\right)^k-\left(\frac{n-1}{n}\right)^{2k}<0$$

since $(1-1/n)^2-(1-2/n)=1/n^2>0$.

The fact that $\mathrm{Cov}(Y_{k,i},Y_{k,j})<0$ tells us that $Y_{k,i}$ and $Y_{k,j}$ are negatively correlated random variables. This is not surprising as knowing that ball j is not chosen increases the chance that ball i might be chosen. Recall that

$$U_k=\sum_{i=1}^{n}Y_{k,i}$$

is the number of balls that are not drawn until the k-th draw then

$$\mathbf{E}U_k=n\mathbf{E}Y_{k,1}=n\left(\frac{n-1}{n}\right)^k ,$$

and, as a consequence of this negative correlation of the $Y_{k,i}$s ,we have

$$\begin{aligned}\mathrm{Var}(U_k) &= \sum_{i=1}^{n}\mathrm{Var}(Y_{k,i})+\sum_{i=1}^{n}\sum_{j\neq i}\mathrm{Cov}(Y_{k,i},Y_{k,j})\\ &\leq n\mathrm{Var}(Y_{k,1})\\ &= n\mathbf{P}Y_{k,1}(1-\mathbf{P}Y_{k,1})\leq\mathbf{E}(Y_{k,1}) .\end{aligned}$$

Appying Chebyshev's inequality to U_k, for $k=cn, c>0$, for $\epsilon>0$, we have

$$\mathbf{P}(|U_{cn}-\mathbf{E}(U_{cn})|\geq n\epsilon)\leq\frac{\mathrm{Var}(U_{cn})}{n^2\epsilon^2}\leq\frac{\mathbf{E}(U_{cn})}{n^2\epsilon^2}=\frac{e^{-c}}{n\epsilon^2}$$

since we have, for n large

$$\mathbf{E}(U_{cn})=n\left(\frac{n-1}{n}\right)^{cn}\approx ne^{-c} .$$

Hence $\mathbf{P}(|U_{cn}-\mathbf{E}(U_{cn})|\geq n\epsilon)$ goes to 0 when n goes to infinity. What this tells us is that at time cn we are very likely to have only seen about $(1-e^{-c})n$ balls. For $c=\log 2$, this is indeed consistent with our earlier calculation for the time it takes to explore half the balls.

8 Simulation of random variables

We are interested in generating continuos random variables[2]. In practice one has access to a random variable with simple probability density function and wants to use this source of randomness to generate a more sophisticated distribution.

[2]In fact most the following analysis extends to discrete time random variables with some minor modifications.

8.1 Inversion method

Let us say we know how to generate a uniform random variable U on the interval $[0,1]$ whose *probability density function* (pdf) is given by

$$f_U(x) = \mathbf{1}_{x\in[0,1]} ,$$

which implies that its *cumulative distribution function* (cdf) is given by

$$F_U(x) = \mathbf{P}(U \leq x) = \begin{cases} 0, & \text{if } x \leq 0, \\ x, & \text{if } 0 < x \leq 1, \\ 1, & \text{if } x > 1 \end{cases}$$

and we want to generate an exponential distribution X with parameter $\lambda > 0$ such that for $x \geq 0$

$$F_X(x) = \mathbf{P}(X \leq x) = 1 - e^{-\lambda x}, \quad f_X(x) = \lambda e^{-\lambda x} .$$

Note that for $y \in [0,1]$ with $y = F_X(x) = 1 - e^{-\lambda x}$ we have

$$x = -\frac{1}{\lambda}\log(1-y) .$$

Therefore the inverse functin of F_X is given by

$$F_X^{-1}(x) = -\frac{1}{\lambda}\log(1-x) .$$

Now let us compute, for U a uniform random variable on $[0,1]$, the following

$$\mathbf{P}\left(-\frac{1}{\lambda}\log(1-U) \leq x\right) = \mathbf{P}\left(U \leq 1 - e^{-\lambda x}\right) = 1 - e^{-\lambda x} ,$$

which implies that the random variable $-\frac{1}{\lambda}\log(1-U)$ is distributed according to an exponential distribution with parameter λ.

This procedure can be generalised to any random variable X with cumulative distribution function F_X whose inverse function is F_X^{-1}.

To generate X given the access to a source for generating uniform random variables on $[0,1]$ what we need to do is apply the function F_X^{-1} to the value obtained from the uniform generator and we obtain a sample from the target distribution with cumulative distribution function F_X.

The above procedure is known as the *Inversion Method*:

> **Generate** U a uniform random variable on $[0,1]$
>
> **Return** $F_X^{-1}(U)$ where F_X^{-1} is the inversion function of the cdf of X.

Although this method appears rather simple at first glance, it requires the computation of the function F_X^{-1} which is often not possible. In particular if X is distributed according to a Normal distribution (a.k.a. Gaussian distribution) then we do not even have a closed-form expression for F_X. So how can one generate a Gaussian distribution?

8.2 Acceptance-Rejection method

We now analyse a more sophisticated and more general procedure known as the *Acceptance-Rejection method*. We assume that we know how to generate some random variable with probability density function g and we wish to find a method for generating a random variable with probability density function f. We suppose that there exists a constant $c > 1$ such that, for all $x \in \mathbb{R}$, we have

$$f(x) \leq cg(x) . \tag{2}$$

As it will appear clearly in what follows, we need to make sure that the above property is verified.

We also assume that we can generate uniformly distributed random variables on the interval $[0,1]$. The Acceptance-Rejection method goes as follows.

Generate U a uniform random variable on $[0,1]$ and a random variable Y with pdf g

If $U > \frac{f(Y)}{cg(Y)}$

Reject and Repeat

else **Accept and Return** Y

Let $U_1, U_2, \ldots$ be a sequence of independent and identically distributed random variables uniformly distributed on $[0,1]$ and let $Y_1, Y_2, \ldots$ be a sequence of independent and identically distributed random variables with probability density function g. We assume that the two sequences are independent.

To generate a random variable according to f, we generate the random variables Y_n and U_n until the instant τ which is the first (random) instant n where we have

$$U_n \leq \frac{f(Y_n)}{cg(Y_n)}\,.$$

In what follows we will show that $Z = Y_\tau$ has probability density function f which we want to generate.

We are interested in the sequence of events, for $n \geq 1$,

$$A_n = \left\{ U_n > \frac{f(Y_n)}{cg(Y_n)} \right\}\,.$$

that describes the failure of the procedure at the n-th step so that its complement A_n^c is the condition for the termination of the algorithm for generating the random variable with pdf f. It is not difficult to see that

$$\begin{aligned}
\mathbf{P}(A_n^c) = \mathbf{P}\left(U_n \leq \frac{f(Y_n)}{cg(Y_n)}\right) &= \int_{-\infty}^{+\infty} \mathbf{P}\left(U_n \leq \frac{f(Y_n)}{cg(Y_n)} \mid Y_n = y\right) g(y)dy \\
&= \int_{-\infty}^{+\infty} \mathbf{P}\left(U_n \leq \frac{f(y)}{cg(y)}\right) g(y)dy \\
&= \int_{-\infty}^{+\infty} \frac{f(y)}{cg(y)} g(y)dy \\
&= \frac{1}{c}\int_{-\infty}^{+\infty} f(y)dy = \frac{1}{c}\,,
\end{aligned}$$

where in the last inequality we used the fact that f is a probability density function and so $\int_{-\infty}^{+\infty} f(y)dy = 1$. Moreover

$$\begin{aligned}
\mathbf{P}(A_n^c, Y_n \leq x) &= \mathbf{P}\left(U_n \leq \frac{f(Y_n)}{cg(Y_n)},\, Y_n \leq x\right) \\
&= \int_{-\infty}^{x} \mathbf{P}\left(U_n \leq \frac{f(Y_n)}{cg(Y_n)} \mid Y_n = y\right) g(y)dy \\
&= \int_{-\infty}^{x} \mathbf{P}\left(U_n \leq \frac{f(y)}{cg(y)}\right) g(y)dy \\
&= \int_{-\infty}^{x} \frac{f(y)}{cg(y)} g(y)dy \\
&= \frac{1}{c}\int_{-\infty}^{x} f(y)dy\,.
\end{aligned}$$

and so

$$\begin{aligned}\mathbf{P}(A_1,\dots A_{n-1},A_n^c,Y_n\le x) &= \mathbf{P}(A_1)\mathbf{P}(A_2)\dots\mathbf{P}(A_{n-1})\mathbf{P}(A_n^c,Y_n\le x)\\ &= \left(1-\frac{1}{c}\right)^{n-1}\frac{1}{c}\int_{-\infty}^{x}f(y)dy\,.\end{aligned}$$

Now let us put the previous arguments together to show that $Z=Y_\tau$ where τ is the first time we have $U_n\le\frac{f(Y_n)}{cg(Y_n)}$ indeed has probability density function f.

$$\begin{aligned}\mathbf{P}(Z\le x)=\mathbf{P}(Y_\tau\le x) &= \sum_{n\ge1}\mathbf{P}(Y_\tau\le x \text{ and } \tau=n)\\ &= \sum_{n\ge1}\mathbf{P}(A_1,\dots A_{n-1},A_n^c,Y_n\le x)\\ &= \sum_{n\ge1}\left(1-\frac{1}{c}\right)^{n-1}\frac{1}{c}\int_{-\infty}^{x}f(y)dy\\ &= \int_{-\infty}^{x}f(y)dy\,.\end{aligned}$$

since $c>1$ it is not difficult to check that $\sum_{n\ge1}\left(1-\frac{1}{c}\right)^{n-1}=c$, which concludes the proof that Z has pdf f.

Note that it is not difficult to see that the time it takes the procedure to complete and output a ranom variable with pdf f is a geometric random variable with parameter $1/c$ which is the probability of success, and the averagre number of iterations is thus given by

$$\mathbf{E}(\tau)=c\,.$$

Example As an example let us apply the previous method to the case where X is a standard Gaussian distribution assuming that we have access to a generate of exponenetial random variables Y that we could obtain by means of the inversion method described in the begining of the section.

Recall that

$$f_X(x)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right),\ x\in\mathbb{R}\quad\text{ and }\quad f_Y(y)=e^{-y},\ y\ge0\,.$$

To simplify the analysis, we can restrict ourselves to generate a random variable Z such that

$$f_z(x)=\frac{2}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right),\ x\ge0\,, \tag{3}$$

and obtain X which is Gaussian by returning $S.Z$ where S is such that

$$\mathbf{P}(S=1)=\mathbf{P}(S=-1)=\frac{1}{2}\,.$$

In the remainder of the example we are interested in generating Z with pdf f_z given by equation (3). Let us first check that condition (2) is verified. For $x\ge0$

$$\frac{f_Z(x)}{f_Y(x)}=\frac{\frac{2}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)}{e^{-x}}=\sqrt{\frac{2}{\pi}}e^{x-x^2/2}=\sqrt{\frac{2}{\pi}}e^{-\frac{(x-1)^2}{2}}e^{1/2}=\sqrt{\frac{2e}{\pi}}e^{-\frac{(x-1)^2}{2}}\le\sqrt{\frac{2e}{\pi}}\approx1.32\,.$$

Applying the acceptance-rejection method with $c = \sqrt{\frac{2e}{\pi}}$ we have

Generate U uniform on $[0,1]$ and Y exponential with parameter 1

If $U > \frac{f_Z(Y)}{cf_Y(Y)} = e^{-\frac{(Y-1)^2}{2}}$

Reject and Repeat

else **Accept and Return** Y

Note that

$$U > e^{-\frac{(Y-1)^2}{2}}$$

is equivalen to

$$-\log U < \frac{(Y-1)^2}{2}$$

and that $-\log U$ is an exponential distribution with parameter 1 so we can transform the algorithm as follow

Generate$Y$ and Y' independent exponential random variables with parameter 1

If $Y' < \frac{(Y-1)^2}{2}$

Reject and Repeat

else **Accept and Return** Y

In fact taking advantage of the memoriless property of the exponential distribution (see below) one could come up with a more elegant algorithm...

9 Continuous random variables and distributions

In this section, we review some important notions related to continuous random variables.

- A random variable X is continuous if its (cumulative) distribution function of X $F_X : \mathbb{R} \to [0,1]$ can be written as
$$F_X(x) = \mathbf{P}(X \leq x) = \int_{-\infty}^{x} f_X(x)\, dx\,,$$
for some function $f_X : \mathbb{R} \to [0,\infty)$ called the probability density function (p.d.f) of X.
- Two continuous random variables X and Y are independent if the events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent for all the values x and y.
- The mean value, expectation or expected value of the random variable X is defined by $\mathbf{E}(X) = \int_{-\infty}^{+\infty} x f_X(x)\, dx$.
- The variance of X is defined by $\mathbf{Var}(X) = \mathbf{E}\left[(\mathbf{E}(X) - X)^2\right]$.

Examples of continuous distributions We now present the most important continuous random variables' distributions.

- Uniform distribution, $\mathrm{U}([a,b])$, $a < b$
$$F_{\mathrm{U}([a,b])}(x) = \begin{cases} 0, & \text{if } x \leq a, \\ \frac{x-a}{b-a}, & \text{if } a < x \leq b, \\ 1, & \text{if } x > b \end{cases}$$

$$f_{\mathrm{U}([a,b])}(x) = \begin{cases} 0, & \text{if } x \leq a, \\ \frac{1}{b-a}, & \text{if } a < x \leq b, \\ 0, & \text{if } x > b \end{cases}$$

$$\mathbf{E}X = \frac{b+a}{2}, \quad \mathbf{Var}(X) = \frac{(b-a)^2}{12}\,.$$

- Exponential distribution, $\mathrm{Exp}(\lambda)$,

$$F_{\mathrm{Exp}(\lambda)}(x) = 1 - e^{-\lambda x}, \quad f_{\mathrm{Exp}(\lambda)}(x) = \lambda e^{-\lambda x}, \mathbf{E}X = 1/\lambda,\ \mathbf{Var}(X) = 1/\lambda^2\,.$$

  Moreover $\mathbf{P}(X > x) = e^{-\lambda x}$ and $\mathbf{P}(X > x + y \mid X > x) = e^{-\lambda y}$. This is known as the *memoriless property* of the exponential distribution.

- Normal distribution, $\mathcal{N}(\mu, \sigma^2)$,

$$f_{\mathcal{N}(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad \mathbf{E}X = \mu, \quad \mathbf{Var}(X) = \sigma^2\,.$$

  In fact $\frac{X-\mu}{\sigma^2}$ has $\mathcal{N}(0,1)$ distribution[3].

[3]Recall that $\int_{-\infty}^{+\infty} e^{-x^2/2}\,dx = \sqrt{2\pi}$

10 Exercises

Exercise 1 Consider an experiment which includes flipping a fair coin. Suppose that this experiment succeeds if the outcome is "heads" and fails otherwise.

1. Let Y be a random variable such that $Y = 1$ if the experiment succeeds and $Y = 0$ otherwise. What is the expected value of Y?
2. What is the probability of having j successes (j heads) knowing that you have flipped the coin for n times?
3. Suppose that we flip a coin until the first heads comes up. What is the probability distribution of the number of flips?

Exercise 2 A traditional fair die is tossed twice, what is the probability that

1. A six turns up exactly once?
2. Both numbers are odd?
3. The sum of the scores is 4?
4. The sum of the scores is divisible by 3?

Exercise 3 A fair coin is thrown repeatedly. What is the probability that on the nth throw:

1. A head happens for the first time?
2. The number of heads and tails are equal?
3. Exactly two heads have appeared?
4. At least two heads have appeared?

Exercise 4 An urn contains 10 identical balls numbered $0, 1, ..., 9$. A random experiment involves selecting a ball from the urn and noting the number of the ball. Find the probability of the following events:

$A =$ "number of ball selected is odd,"
$B =$ "number of ball selected is a multiple of 3,"
$C =$ "number of ball selected is less than 5,"
and of A and $A \cup B \cup C$.
Hint: Use the following formula.

$$P\left[\bigcup_{k=1}^{n} A_k\right] = \sum_{j=1}^{n} P[A_j] - \sum_{j<k} P[A_j \cap A_k] + ... + (-1)^{n+1} P[A_1 \cap ... \cap A_n]$$

Exercise 5 A communication system transmits binary information over a channel that introduces random bit errors with probability $e = 10^{-3}$. The transmitter transmits each information bit three times, and a decoder takes a majority vote of the received bits to decide on what the transmitted bit was. Find the probability that the receiver will make an incorrect decision.

Exercise 6 Consider an experiment where we pick two numbers x and y uniformly at random between zero and one. Suppose that all pairs of numbers are equally likely to be selected. Find the probability of the following:

$A = \{x > 0.5\}$
$B = \{y > 0.5\}$
$C = \{x > y\}$
$D = \{x = 0.5\}$

Exercise 7 At the station there are three payphones which accept 20p pieces. One never works, another always works, while the third works with probability 1/2. On my way to the metropolis for the day, I wish to identify the reliable phone, so that I can use it on my return. The station is empty and I have just three 20p pieces. I try one phone and it does not work. I try another twice in succession and it works both times. What is the probability that this second phone is the reliable one?

Exercise 8. Poker hands Assume we have a deck of 52 cards consisting of 4 suits (Hearts ♡, Diamonds ♢, Clubs ♣, Spades ♠) with 13 ranks in each (2, 3, 4, 5, 6, 7, 8, 9, 10, Jack, Queen, King, Ace). We want to evaluate the frequency, the probability or the odds for each of the hands. It is not difficult to see that there are $\binom{52}{5} = 2\,598\,960$ possible hands. Compute the probability of having each of the following hands.

- *One Pair* is a poker hand such as, which contains 2 cards of the same rank out of the 13 ranks of any of the 4 suits, plus 3 other unmatched cards from any of the remaining 12 ranks and can have any of the 4 suits. Example: 1♡ 1♣ 2♣ K♠ 10♢.
- *Two Pair* contains any 2 of the 13 ranks with pair having 2 of 4 four suits. The remaining card can have any one of the 11 other ranks, and any suit. Example: 1♡ 1♣ 2♣ 2♠ 10♢.
- *Three-of-a-kind* contains any of thirteen ranks that form the three of a kind from any three of the four suits. The remaining cards may have any two of the remaining 12 ranks, and each can have any one of the four suits. Example: A♡ A♣ A♠ 10♢ 9♣.
- *Straight* consists of any 1 of the 10 possible sequences of 5 consecutive cards, from 5-4-3-2-A to A-K-Q-J-10. Each of these 5 cards may have any one of the 4 suits. Example: 3♡ 4♣ 5♠ 6♢ 7♣.
- *Flush* contains any 5 of the 13 ranks, all of which belong to 1 of the 4 suits, minus the 40 straight flushes. Example: 3♡ 4♡ 7♡ 9♡ J♡ .
- *Full house* comprises a triple (three of a kind) and a pair. Example A♡ A♣ A♠ 10♢ 10♣.
- *Four-of-a-kind* is formed of four cards of the same rank. Example A♡ A♣ A♠ A♢ 10♣.
- *Straight flush* is formed of 5 consecutive cards of the same suit. Example 3♡ 4♡ 5♡ 6♡ 7♡.

Exercises 9 Define the following events

- *A Die :* The event the outcome of the experiment of a die tossing is an even number.
- *Darts :* Let us assume that the target that player wants to reach corresponds to a disc of radius 1 centred at the point $(0, 0)$. Define the event of "reaching the target"
- *Coin tosses :* Suppose that we are taking part in a game where a player wins if he succeeds at getting a head in the first two tosses. Define the event of winning.

Exercise 10 Suppose that n (absent-minded) professors attend the opera one night.

They each leave their hat at the cloakroom at the beginning of the night, but by the end of the night they have all lost their cloakroom tickets.

They are the last n people to return to the cloakroom at the end of the opera, and none of them can recognize his own hat; so they decide to each pick a hat at random.

What is the probability that none of the professors leaves the opera wearing the same hat he came in with?

Exercise 11 Two Urns: Urn I contains 2 white and 3 blue balls, Urn 2 contains 3 white and 4 blue. Take a ball from Urn 1 and put it in urn 2. Pick a random ball from Urn 2. Probability that it is blue?

Exercise 12 Toss repeatedly a biased coin. Compute p_n the probability that an even number of heads has occurred after n tosses.

Exercise 13 A couple has three children. Events A "all children are of the same sex", B "There is at most one boy", C " the family includes a boy and a girl". Are A and B, B and C, and A and C independent?

Exercise 14 Starting with $\$k$ a gambler is determined to earn an extra $\$(n-k)$ playing the following game repeatedly: He tosses a fair coin and if it shows head he wins \$1 and he loses \$1 if it is tail. What is the probability that he is ultimately bankrupt?

Exercise 15 Let X_1, X_2 be two independent geometric random variables with parameters p_1, p_2. Compute the following probabilities

$$\mathbf{P}(X_1 > n+k \mid X_1 > n), \quad \mathbf{P}(X_1 = X_2), \quad \mathbf{P}(\min(X_1, X_2) = k), \quad \mathbf{P}(X_1 < X_2), \quad \mathbf{P}(X_1 \leq X_2)\,.$$

Then compute $\mathbf{E}[\max(X_1, X_2)]$, $\mathbf{E}(X_1 \mid X_1 \leq X_2)$.

Exercise 16. Rendez-vous Alice and Bob agree to meet in front of the science museem in South Kensington after their Wednesday lectures. They arrive at times that are independent and uniformly distributed between 12:00 and 13:00. Each is prepared to wait s minutes before leaving. Find a minimal s such that the probability that they meet is at least 1/2.

Exercise 17

1. Try to prove *Markov's Inequality* which is as following:
   Let X be a random variable that assumes only non-negative values. Then, for all $a > 0$,

   $$\mathbf{P}(X \geq a) \leq \frac{\mathbf{E}[X]}{a}$$

   Hint: Use the indicator function I:

   $$I = \begin{cases} 1 & \text{if } X \geq a, \\ 0 & \text{otherwise,} \end{cases}$$

2. Using Markov's Inequality prove

(a) *Chebyshev's Inequality* which is as follows:
For any $a > 0$,

$$\mathbf{P}(|X - \mathbf{E}[X]| \geq a) \leq \frac{Var[X]}{a^2}$$

Hint: Use the following

$$\mathbf{P}(|X - \mathbf{E}[X]| \geq a) = \mathbf{P}((X - \mathbf{E}[X])^2 \geq a^2)$$

(b) *Chernoff Bound* which is the following:

$$\mathbf{P}(X \geq a) \leq \min_{t>0} \frac{\mathbf{E}[e^{tX}]}{e^{ta}}$$

Exercise 18 Consider tossing a coin for n times. The variable X_i describes the outcome of the i-th toss: X_i=1 if heads shows and $X_i = 0$ if tail shows. Let $X = \sum_{i=1}^{n} X_i$.

1. State the distribution of X, and then compute its expectation and its variance.

2. (a) Show that
$$\mathbf{P}\left(X \geq \frac{3n}{4}\right) \leq \mathbf{P}\left(|X - \frac{n}{2}| \geq \frac{n}{4}\right).$$
   (b) Using Chebyshev's inequality, prove that
$$\mathbf{P}\left(X \geq \frac{3n}{4}\right) \leq \frac{4}{n}$$
   (c) Find the $\lim_{n\to\infty} \mathbf{P}(X \geq \frac{3n}{4})$

3. We now derive a tighter bound for convergence of $\mathbf{P}(X \geq \frac{3n}{4})$ as n goes to ∞.
   (a) Let $x, \theta \geq 0$. Combining Markov's inequality and the fact that
$$\{X \geq x\} = \{e^{\theta X} \geq e^{\theta x}\}$$
   prove that
$$\mathbf{P}(X \geq x) \leq \exp(\frac{n}{2}(e^{\theta} - 1) - \theta x)$$
   Hint: Use the inequality $1 + \alpha \leq e^{\alpha}$, for $\alpha \geq 0$.
   (b) Choose θ so that $\frac{1}{2}(e^{\theta} - 1) - \frac{3}{4}\theta \leq -0.01$.
   (c) Prove that for the choice of θ in the previuos question
$$\mathbf{P}\left(X \geq \frac{3n}{4}\right) \leq e^{-0.01n}.$$

Exercise 19 In a simple model of Foreign Exchange markets, the exchange rate for the British pound against the Euro is believed to change from day to day according to a simple probability model; from one day to the next, the rate increases by a constant multiplicative factor $a > 1$ with probability 1/2, and decreases by a factor $1/a < 1$ with probability 1/2.

1. Let Y_n be the logarithm of the exchange rate on day n starting from an exchange rate of 1 on day 0. Show that
$$Y_n = \sum_{i=1}^{n} X_i$$
where
$$X_i = \begin{cases} \log a, & \text{with probability } 1/2, \\ -\log a, & \text{with probability } 1/2, \end{cases}$$
2. Using the Central Limit Theorem, find an approximation of Y_n, for large n. Discuss the distribution of the actual exchange rate.
3. Compute the expectation and the variance of Y_n.
4. Using Chebyshev's inequality show that, for $k = 2, 3, 4, \ldots$
$$\mathbf{P}\left(|Y_n| \geq k\sqrt{n}\log(a)\right) \leq \frac{1}{k^2}\,.$$
Comment.

Exercise 20 Let $X_1, ..., X_n$ be independent poisson trials such that $\mathbb{P}(X_i = 1) = p_i$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}[X]$. Then the following Chernoff bounds hold:

For any $0 < \delta \leq 1$

$$\mathbb{P}\left(X \geq (1+\delta)\mu\right) \leq e^{-\mu\delta^2/3} \tag{4}$$

$$\mathbb{P}\left(X \leq (1-\delta)\mu\right) \leq e^{-\mu\delta^2/2} \tag{5}$$

We are interested in evaluating the probability of a particular gene mutation given n samples. As the test for determining the mutation is expensive, we would like to have a close enough estimation $\hat{p}$ of the actual mutation probability p.

To this end, we introduce the notion of *confidence interval*,

A $(1-\gamma)$ confidence interval for a parameter p is an interval $[\hat{p}-\delta, \hat{p}+\delta]$ such that

$$\mathbb{P}\left(p \in [\hat{p}-\delta, \hat{p}+\delta]\right) \geq 1-\gamma$$

Let $X_i = 1$ if we see a mutation in samples and $X_i = 0$ otherwise. Also, let $X = \sum_{i=1}^{n} X_i$.

1. Find the distribution of X and $\mathbb{E}(X)$.
2. Let $\hat{p} = \frac{X}{n}$.
   (a) Show that if $p < \hat{p} - \delta$, then $X > \mathbb{E}(X)\left(1 + \frac{\delta}{p}\right)$.
   (b) Show that if $p > \hat{p} + \delta$, then $X < \mathbb{E}(X)\left(1 - \frac{\delta}{p}\right)$.
3. Using Chernoff Bounds 4 and 5, show that
$$\mathbb{P}\left(p \notin [\hat{p}-\delta, \hat{p}+\delta]\right) < e^{-n\delta^2/2p} + e^{-n\delta^2/3p}$$

4. Note that $p \leq 1$. Define γ independent of p such that

$$\mathbb{P}\left(p \in [\hat{p} - \delta, \hat{p} + \delta]\right) \geq 1 - \gamma$$

5. In the above we assumed that the X_is were $\{0, 1\}$ and estimated $\sum_{i=1}^{n} \frac{X_i}{n}$ up to some confidence. Here, we assume that we do not have the distribution of the X_is and we only know that they are independent. We will define n the number of samples needed so that

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} \frac{X_i}{n} - \mathbb{E}(X)\right| \leq \delta \mathbb{E}(X)\right) \geq 1 - \gamma$$

Let $r = \frac{\sqrt{\mathrm{Var}(X)}}{\mathbb{E}(X)}$,

Using Chebyshev inequality show that $n = o\left(\frac{r^2}{\delta^2 \gamma}\right)$ is sufficient, i.e. $\exists c$ such that $n \leq \frac{cr^2}{\delta^2 \gamma}$.

Exercise 21. Secretary problem Imagine that your assistant is about to retire and that you want to hire a new assistant and you have n people to interview. You obviously want to hire the best candidate for the position. Assume that your company policy is to interview the candidates one by one and that when you interview some candidate you either offer the candicate the job before the next interview or you forever lose the chance to hire that candidate. We assume that the candidates are interviewed in a random order.

We consider the following strategy. First, interview m candidates but reject them all, these candidates give an idea of how strong the pool of candidates is. after the m-th candidate, hire the first candidate who is better than all the previous candidates you interviewed. In what follows we will evaluate this strategy.

1. Let E be the event of hiring the best candidate and let E_i that the i-th candidate is the best and he is hired. Determine $\mathbf{P}(E_i)$ and show that

$$\mathbf{P}(E) = \frac{m}{n} \sum_{j=m+1}^{n} \frac{1}{j-1}\,.$$

2. using the fact $\frac{1}{x} \leq \log(x) - \log(x-1) \leq \frac{1}{x-1}$, $x > 1$ show that

$$\frac{m}{n}(\log n - \log m) \leq \mathbf{P}(E) \leq \frac{m}{n}(\log(n-1) - \log(m-1))\,.$$

3. Show that $\frac{m}{n}(\log n - \log m)$ is maximised when $m = \frac{n}{e}$.
4. For $m = \frac{n}{e}$ show that $\mathbf{P}(E) \geq e^{-1}$.

Exercise 22. Bertrand's Paradox A chord has been chosen at random in a circle of radius r. What is the probability that it is longer than the side of the equilateral triangle inscribed in the circle?

Consider the following three cases.

1. the middle point of the chord is distributed uniformly inside the circle;
2. one endpoint is fixed and the second is uniformly distributed over the circumference;
3. The distance between the middle point of the chord and the centre of the circle is uniformly dustributed over the interval $[0, r]$.

Exercise 23. Verifyng polynomial identities We want to check whether two polynomials written in different forms are equal. For example, do we have

$$(x+1)(x-2)(x+3)(x-4)(x+5)(x-6) = x^6 - 7x^3 + 25\ ?$$

The easy way of checking this is to multiply together the terms on the left-hand side and check if the resulting polynomial matches the right-hand side.

In general to check if two polynomials F anf G are equal we convert each of them to its canonical form which is

$$\sum_{i=1}^{d} c_i x^i$$

and check if both canonical forms coincide. To simplify the analysis assume that

$$F(x) = \prod_{i=1}^{d}(x - a_i) \quad \text{and} \quad G(x) = \sum_{i=1}^{d} c_i x^i\ .$$

It is not difficult to see that transforming F to its canonical form requires some d^2 multiplications of coefficients.

In what follows we advocate a different way of proceeding similar in spirit to how we performed the verification of matrix multiplications. We choose an integer uniformly at random in the set $\{1, 2, 3, \ldots, 100d\}$ and check whether the numbers $F(r)$ and $G(r)$ are equal or not. The algorithm will then say the two polynomials are equal if $F(r) = G(r)$ and that they are distintc if $F(r) \neq G(r)$. Note that computing $F(r)$ and $G(r)$ only takes about d multiplications which is faster than the naive approach.

1. Analyse the outcomes in which the algorithm provides the wrong answer and show that

$$\mathbf{P}(\text{ algorithm fails}) \leq \frac{1}{100}\ .$$

   Justify your answer.

   Hint: remember that if F and G agree on d values then they must be equal. This is a standard property of polynomials

2. Assume that the algorithm is repeated and that the numbers are drawn with replacement, i.e. we do not exclude a number we have already drawn from $\{1, 2, 3, \ldots, 100d\}$. What is the probability of error of the algorithm when it is repeated k times?

3. Assume that the algorithm is repeated and that the numbers are drawn without replacement, i.e. we exclude a number we have already drawn from $\{1, 2, 3, \ldots, 100d\}$. What is the probability of error of the algorithm when it is repeated k times?

4. Discuss the above two procedures (with and without) replacement and compare their accuracy and efficiency.