

# Chapter 4 Probability Theory

Chances are that you will face at least a couple of probability problems in most quantitative interviews. Probability theory is the foundation of every aspect of quantitative finance. As a result, it has become a popular topic in quantitative interviews.

Although good intuition and logic can help you solve many of the probability problems, having a thorough understanding of basic probability theory will provide you with clear and concise solutions to most of the problems you are likely to encounter. Furthermore, probability theory is extremely valuable in explaining some of the seemingly-counterintuitive results. Armed with a little knowledge, you will find that many of the interview problems are no more than disguised textbook problems.

So we dedicate this chapter to reviewing basic probability theory that is not only broadly tested in interviews but also likely to be helpful for your future career.<sup>1</sup> The knowledge is applied to real interview problems to demonstrate the power of probability theory. Nevertheless, the necessity of knowledge in no way downplays the role of intuition and logic. Quite the contrary, common sense and sound judgment are always crucial for analyzing and solving either interview or real-life problems. As you will see in the following sections, all the techniques we discussed in Chapter 2 still play a vital role in solving many of the probability problems.

Let's have some fun playing the odds.

## 4.1 Basic Probability Definitions and Set Operations

First let's begin with some basic definitions and notations used in probability. These definitions and notations may seem dry without examples—which we will present momentarily—yet they are crucial to our understanding of probability theory. In addition, it will lay a solid ground for us to systematically approach probability problems.

**Outcome ( $\omega$ ):** the outcome of an experiment or trial.

**Sample space/Probability space ( $\Omega$ ):** the set of all possible outcomes of an experiment.

---

<sup>1</sup> As I have emphasized in Chapter 3, this book does not teach probability or any other math topics due to the space limit—it is not my goal to do so, either. The book gives a summary of the frequently-tested knowledge and shows how it can be applied to a wide range of real interview problems. The knowledge used in this chapter is covered by most introductory probability books. It is always helpful to pick up one or two classic probability books in case you want to refresh your memory on some of the topics. My personal favorites are *First Course in Probability* by Sheldon Ross and *Introduction to Probability* by Dimitri P. Bertsekas and John N. Tsitsiklis.

## Probability Theory

$P(\omega)$ : Probability of an outcome ( $P(\omega) \geq 0, \forall \omega \in \Omega, \sum_{\omega \in \Omega} P(\omega) = 1$ ).

**Event:** A set of outcomes and a subset of the sample space.

$P(A)$ : Probability of an event  $A$ ,  $P(A) = \sum_{\omega \in A} P(\omega)$ .

$A \cup B$ : Union  $A \cup B$  is the set of outcomes in event  $A$  or in event  $B$  (or both).

$A \cap B$  or  $AB$ : Intersection  $A \cap B$  (or  $AB$ ) is the set of outcomes in both  $A$  and  $B$ .

$A^c$ : The complement of  $A$ , which is the event “not  $A$ ”.

**Mutually Exclusive:**  $A \cap B = \Phi$  where  $\Phi$  is an empty set.

For any mutually exclusive events  $E_1, E_2, \dots, E_N$ ,  $P\left(\bigcup_{i=1}^N E_i\right) = \sum_{i=1}^N P(E_i)$ .

**Random variable:** A function that maps each outcome ( $\omega$ ) in the sample space ( $\Omega$ ) into the set of real numbers.

Let's use the rolling of a six-sided dice to explain these definitions and notations. A roll of a dice has 6 possible outcomes (mapped to a random variable): 1, 2, 3, 4, 5, or 6. So the sample space  $\Omega$  is  $\{1, 2, 3, 4, 5, 6\}$  and the probability of each outcome is  $1/6$  (assuming a fair dice). We can define an event  $A$  representing the event that the outcome is an odd number  $A = \{1, 3, 5\}$ , then the complement of  $A$  is  $A^c = \{2, 4, 6\}$ . Clearly  $P(A) = P(1) + P(3) + P(5) = 1/2$ . Let  $B$  be the event that the outcome is larger than 3:  $B = \{4, 5, 6\}$ . Then the union is  $A \cup B = \{1, 3, 4, 5, 6\}$  and the intersection is  $A \cap B = \{5\}$ . One popular random variable called indicator variable (a binary dummy variable) for event  $A$  is defined as the following:

$I_A = \begin{cases} 1, & \text{if } x \in \{1, 3, 5\} \\ 0, & \text{if } x \notin \{1, 3, 5\} \end{cases}$ . Basically  $I_A = 1$  when  $A$  occurs and  $I_A = 0$  if  $A^c$  occurs. The expected value of  $I_A$  is  $E[I_A] = P(A)$ .

Now, time for some examples.

## Coin toss game

Two gamblers are playing a coin toss game. Gambler  $A$  has  $(n+1)$  fair coins;  $B$  has  $n$  fair coins. What is the probability that  $A$  will have more heads than  $B$  if both flip all their coins?<sup>2</sup>

*Solution:* We have yet to cover all the powerful tools probability theory offers. What do we have now? Outcomes, events, event probabilities, and surely our reasoning capabilities! The one extra coin makes  $A$  different from  $B$ . If we remove a coin from  $A$ ,  $A$  and  $B$  will become symmetric. Not surprisingly, the symmetry will give us a lot of nice properties. So let's remove the last coin of  $A$  and compare the number of heads in  $A$ 's first  $n$  coins with  $B$ 's  $n$  coins. There are three possible outcomes:

$E_1$ :  $A$ 's  $n$  coins have more heads than  $B$ 's  $n$  coins;

$E_2$ :  $A$ 's  $n$  coins have equal number of heads as  $B$ 's  $n$  coins;

$E_3$ :  $A$ 's  $n$  coins have fewer heads than  $B$ 's  $n$  coins.

By symmetry, the probability that  $A$  has more heads is equal to the probability that  $B$  has more heads. So we have  $P(E_1) = P(E_3)$ . Let's denote  $P(E_1) = P(E_3) = x$  and  $P(E_2) = y$ .

Since  $\sum_{\omega \in \Omega} P(\omega) = 1$ , we have  $2x + y = 1$ . For event  $E_1$ ,  $A$  will always have more heads than  $B$  no matter what  $A$ 's  $(n+1)th$  coin's side is; for event  $E_3$ ,  $A$  will have no more heads than  $B$  no matter what  $A$ 's  $(n+1)th$  coin's side is. For event  $E_2$ ,  $A$ 's  $(n+1)th$  coin does make a difference. If it's a head, which happens with probability 0.5, it will make  $A$  have more heads than  $B$ . So the  $(n+1)th$  coin increases the probability that  $A$  has more heads than  $B$  by  $0.5y$  and the total probability that  $A$  has more heads is  $x + 0.5y = x + 0.5(1 - 2x) = 0.5$  when  $A$  has  $(n+1)$  coins.

## Card game

A casino offers a simple card game. There are 52 cards in a deck with 4 cards for each value 2, 3, 4, 5, 6, 7, 8, 9, 10,  $J$ ,  $Q$ ,  $K$ ,  $A$ . Each time the cards are thoroughly shuffled (so each card has equal probability of being selected). You pick up a card from the deck and the dealer picks another one without replacement. If you have a larger number, you win; if the numbers are equal or yours is smaller, the house wins—as in all other casinos, the house always has better odds of winning. What is your probability of winning?

---

<sup>2</sup> Hint: What are the possible results (events) if we compare the number of heads in  $A$ 's first  $n$  coins with  $B$ 's  $n$  coins? By making the number of coins equal, we can take advantage of symmetry. For each event, what will happen if  $A$ 's last coin is a head? Or a tail?

*Solution:* One answer to this problem is to consider all 13 different outcomes of your card. The card can have a value  $2, 3, \dots, A$  and each has  $1/13$  of probability. With a value of 2, the probability of winning is  $0/51$ ; with a value of 3, the probability of winning is  $4/51$  (when the dealer picks a 2); ...; with a value of  $A$ , the probability of winning is  $48/51$  (when the dealer picks a 2, 3, ..., or  $K$ ). So your probability of winning is

$$\frac{1}{13} \times \left( \frac{0}{51} + \frac{4}{51} + \dots + \frac{48}{51} \right) = \frac{4}{13 \times 51} \times (0+1+\dots+12) = \frac{4}{13 \times 51} \times \frac{12 \times 13}{2} = \frac{8}{17}.$$

Although this is a straightforward solution and it elegantly uses the sum of an integer sequence, it is not the most efficient way to solve the problem. If you have got the core spirits of the coin tossing problem, you may approach the problem by considering three different outcomes:

$E_1$  : Your card has a number larger than the dealer's;

$E_2$  : Your card has a number equal to the dealer's;

$E_3$  : Your card has a number lower than the dealer's.

Again by symmetry,  $P(E_1) = P(E_3)$ . So we only need to figure out  $P(E_2)$ , the probability that two cards have equal value. Let's say you have randomly selected a card. Among the remaining 51 cards, only 3 cards will have the same value as your card. So the probability that the two cards have equal value is  $3/51$ . As a result, the probability that you win is  $P(E_1) = (1 - P(E_2))/2 = (1 - 3/51)/2 = 8/17$ .

## Drunk passenger

A line of 100 airline passengers are waiting to board a plane. They each hold a ticket to one of the 100 seats on that flight. For convenience, let's say that the  $n$ -th passenger in line has a ticket for the seat number  $n$ . Being drunk, the first person in line picks a random seat (equally likely for each seat). All of the other passengers are sober, and will go to their proper seats unless it is already occupied; In that case, they will randomly choose a free seat. You're person number 100. What is the probability that you end up in your seat (i.e., seat #100)?<sup>3</sup>

*Solution:* Let's consider seats #1 and #100. There are two possible outcomes:

---

<sup>3</sup> Hint: If you are trying to use complicated conditional probability to solve the problem, go back and think again. If you decide to start with a simpler version of the problem, starting with two passengers and increasing the number of passengers to show a pattern by induction, you can solve the problem more efficiently. But the problem is much simpler than that. Focus on events and symmetry and you will have an intuitive answer.

$E_1$ : Seat #1 is taken before #100;

$E_2$ : Seat #100 is taken before #1.

If any passenger takes seat #100 before #1 is taken, surely you will not end up in your own seat. But if any passenger takes #1 before #100 is taken, you will definitely end up in your own seat. By symmetry, either outcome has a probability of 0.5. So the probability that you end up in your seat is 50%.

In case this over-simplified version of reasoning is not clear to you, consider the following detailed explanation: If the drunk passenger takes #1 by chance, then it's clear all the rest of the passengers will have the correct seats. If he takes #100, then you will not get your seat. The probabilities that he takes #1 or #100 are equal. Otherwise assume that he takes the  $n$ -th seat, where  $n$  is a number between 2 and 99. Everyone between 2 and  $(n-1)$  will get his own seat. That means the  $n$ -th passenger essentially becomes the new "drunk" guy with designated seat #1. If he chooses #1, all the rest of the passengers will have the correct seats. If he takes #100, then you will not get your seat. (The probabilities that he takes #1 or #100 are again equal.) Otherwise he will just make another passenger down the line the new "drunk" guy with designated seat #1 and each new "drunk" guy has equal probability of taking #1 or #100. Since at all jump points there's an equal probability for the "drunk" guy to choose seat #1 or 100, by symmetry, the probability that you, as the 100th passenger, will seat in #100 is 0.5.

## N points on a circle

Given  $N$  points drawn randomly on the circumference of a circle, what is the probability that they are all within a semicircle?<sup>4</sup>

*Solution:* Let's start at one point and clockwise label the points as  $1, 2, \dots, N$ . The probability that all the remaining  $N-1$  points from 2 to  $N$  are in the clockwise semicircle starting at point 1 (That is, if point 1 is at 12:00, points 2 to  $N$  are all between 12:00 and 6:00) is  $1/2^{N-1}$ . Similarly the probability that a clockwise semicircle starting at any point  $i$ , where  $i \in \{2, \dots, N\}$  contains all the other  $N-1$  points is also  $1/2^{N-1}$ .

Claim: the events that all the other  $N-1$  points are in the clockwise semicircle starting at point  $i$ ,  $i = 1, 2, \dots, N$  are mutually exclusive. In other words, if we, starting at point  $i$  and proceeding clockwise along the circle, sequentially encounters points  $i+1, i+2, \dots, N, 1, \dots, i-1$  in half a circle, then starting at any other point  $j$ , we cannot encounter all

<sup>4</sup> Hint: Consider the events that starting from a point  $n$ , you can reach all the rest of the points on the circle clockwise,  $n \in \{1, \dots, N\}$  in a semicircle. Are these events mutually exclusive?

other points within a clockwise semicircle. Figure 4.1 clearly demonstrates this conclusion. If starting at point  $i$  and proceeding clockwise along the circle, we sequentially encounter points  $i+1, i+2, \dots, N, 1, \dots, i-1$  within half a circle, the clockwise arc between  $i-1$  and  $i$  must be no less than half a circle. If we start at any other point, in order to reach all other points clockwise, the clockwise arc between  $i-1$  and  $i$  are always included. So we cannot reach all points within a clockwise semicircle starting from any other points. Hence, all these events are mutually exclusive and we have

$$P\left(\bigcup_{i=1}^N E_i\right) = \sum_{i=1}^N P(E_i) \Rightarrow P\left(\bigcup_{i=1}^N E_i\right) = N \times 1/2^{N-1} = N/2^{N-1}$$

The same argument can be extended to any arcs that have a length less than half a circle. If the ratio of the arc length to the circumference of the circle is  $x$  ( $x \leq 1/2$ ), then the probability of all  $N$  points fitting into the arc is  $N \times x^{N-1}$ .

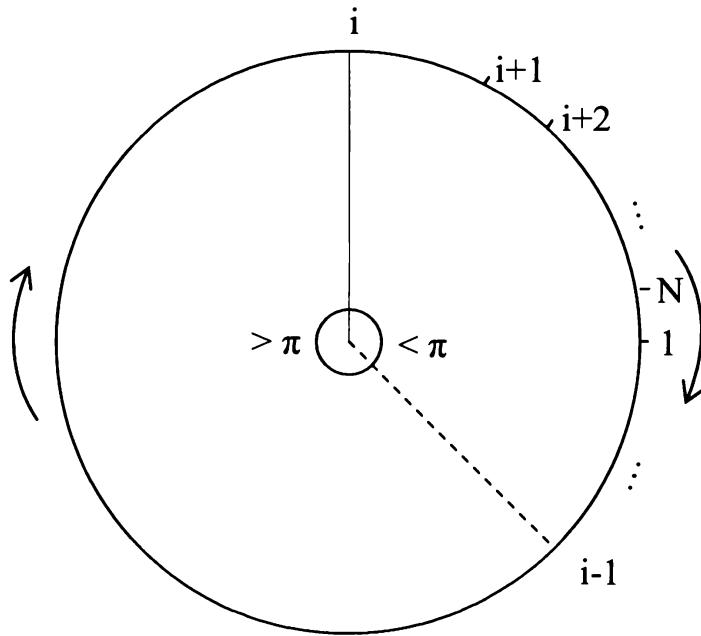


Figure 4.1  $N$  points fall in a clockwise semicircle starting from  $i$

## 4.2 Combinatorial Analysis

Many problems in probability theory can be solved by simply counting the number of different ways that a certain event can occur. The mathematic theory of counting is often referred to as combinatorial analysis (or combinatorics). In this section, we will cover the basics of combinatorial analysis.

**Basic principle of counting:** Let  $S$  be a set of length- $k$  sequences. If there are

- $n_1$  possible first entries,
- $n_2$  possible second entries for each first entry,
- $n_3$  possible third entries for each combination of first and second entries, etc.

Then there are a total of  $n_1 \cdot n_2 \cdots n_k$  possible outcomes.

**Permutation:** A rearrangement of objects into distinct sequence (i.e., order matters).

Property: There are  $\frac{n!}{n_1!n_2!\dots n_r!}$  different permutations of  $n$  objects, of which  $n_1$  are alike,  $n_2$  are alike,  $\dots$ ,  $n_r$  are alike.

**Combination:** An unordered collection of objects (i.e., order doesn't matter).

Property: There are  $\binom{n}{r} = \frac{n!}{(n-r)!r!}$  different combinations of  $n$  distinct objects taken  $r$  at a time.

**Binomial theorem:**  $(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$

**Inclusion-Exclusion Principle:**  $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1E_2)$

$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1E_2) - P(E_1E_3) - P(E_2E_3) + P(E_1E_2E_3)$

and more generally,

$$P(E_1 \cup E_2 \cup \dots \cup E_N) = \sum_{i=1}^N P(E_i) - \sum_{i_1 < i_2} P(E_{i_1}E_{i_2}) + \dots + (-1)^{r+1} \sum_{i_1 < i_2 < \dots < i_r} P(E_{i_1}E_{i_2}\dots E_{i_r}) + \dots + (-1)^{N+1} P(E_1E_2\dots E_N)$$

where  $\sum_{i_1 < i_2 < \dots < i_r} P(E_{i_1}E_{i_2}\dots E_{i_r})$  has  $\binom{N}{r}$  terms.

## Poker hands

Poker is a card game in which each player gets a hand of 5 cards. There are 52 cards in a deck. Each card has a value and belongs to a suit. There are 13 values, 2, 3, 4, 5, 6, 7, 8, 9, 10, jack, queen, king, ace, and four suits, spade, club, heart, diamond.

What are the probabilities of getting hands with four-of-a-kind (four of the five cards with the same value)? Hands with a full house (three cards of one value and two cards of another value)? Hands with two pairs?

**Solution:** The number of different hands of a five-card draw is the number of 5-element subsets of a 52-element set, so total number of hands =  $\binom{52}{5} = 2,598,960$ .

**Hands with a four-of-a-kind:** First we can choose the value of the four cards with the same value, there are 13 choices. The 5th card can be any of the rest 48 cards (12 choices for values and 4 choices for suits). So the number of hands with four-of-a kind is  $13 \times 48 = 624$ .

**Hands with a Full House:** In sequence we need to choose the value of the triple, 13 choices; the suits of the triple,  $\binom{4}{3}$  choices; the value of the pair, 12 choices; and the suits of the pair,  $\binom{4}{2}$  choices. So the number of hands with full house is  $13 \times \binom{4}{3} \times 12 \times \binom{4}{2} = 13 \times 4 \times 12 \times 6 = 3,744$ .

**Hands with Two Pairs:** In sequence we need to choose the values of the two pairs,  $\binom{13}{2}$  choices; the suits of the first pair,  $\binom{4}{2}$  choices; the suits of the second pair,  $\binom{4}{2}$  choices; and the remaining card, 44 ( $52 - 4 \times 2$ , since the last cards can not have the same value as either pair) choices. So the number of hands with two pairs is

$$\binom{13}{2} \times \binom{4}{2} \times \binom{4}{2} \times 44 = 78 \times 6 \times 6 \times 44 = 123,552.$$

To calculate the probability of each, we only need to divide the number of hands of each kind by the total possible number of hands.

## Hopping rabbit

A rabbit sits at the bottom of a staircase with  $n$  stairs. The rabbit can hop up only one or two stairs at a time. How many different ways are there for the rabbit to ascend to the top of the stairs?<sup>5</sup>

---

<sup>5</sup> Hint: Consider an induction approach. Before the final hop to reach the  $n$ -th stair, the rabbit can be at either the  $(n-1)$ th stair or the  $(n-2)$ th stair assuming  $n > 2$ .

*Solution:* Let's begin with the simplest cases and consider solving the problem for any number of stairs using induction. For  $n=1$ , there is only one way and  $f(1)=1$ . For  $n=2$ , we can have one 2-stair hop or two 1-stair hops. So  $f(2)=2$ . For any  $n > 2$ , there are always two possibilities for the last hop, either it's a 1-stair hop or a 2-stair hop. In the former case, the rabbit is at  $(n-1)$  before reaching  $n$ , and it has  $f(n-1)$  ways to reach  $(n-1)$ . In the latter case, the rabbit is at  $(n-2)$  before reaching  $n$ , and it has  $f(n-2)$  ways to reach  $(n-2)$ . So we have  $f(n)=f(n-2)+f(n-1)$ . Using this function we can calculate  $f(n)$  for  $n=3, 4, \dots$ <sup>6</sup>

## Screwy pirates 2

Having peacefully divided the loot (in chapter 2), the pirate team goes on for more looting and expands the group to 11 pirates. To protect their hard-won treasure, they gather together to put all the loot in a safe. Still being a democratic bunch, they decide that only a majority – any majority – of them ( $\geq 6$ ) together can open the safe. So they ask a locksmith to put a certain number of locks on the safe. To access the treasure, every lock needs to be opened. Each lock can have multiple keys; but each key only opens one lock. The locksmith can give more than one key to each pirate.

What is the smallest number of locks needed? And how many keys must each pirate carry?<sup>7</sup>

*Solution:* This problem is a good example of the application of combinatorial analysis in information sharing and cryptography. A general version of the problem was explained in a 1979 paper “*How to Share a Secret*” by Adi Shamir. Let's randomly select 5 pirates from the 11-member group; there must be a lock that none of them has the key to. Yet any of the other 6 pirates must have the key to this lock since any 6 pirates can open all locks. In other words, we must have a “special” lock to which none of the 5 selected pirates has a key and the other 6 pirates all have keys. Such 5-pirate groups are randomly selected. So for each combination of 5 pirates, there must be such a “special” lock. The

minimum number of locks needed is  $\binom{11}{5} = \frac{11!}{5!6!} = 462$  locks. Each lock has 6 keys,

which are given to a unique 6-member subgroup. So each pirate must have  $\frac{462 \times 6}{11} = 252$  keys. That's surely a lot of locks to put on a safe and a lot of keys for each pirate to carry.

<sup>6</sup> You may have recognized that the sequence is a sequence of Fibonacci numbers.

<sup>7</sup> Hint: every subgroup of 6 pirates should have the same key to a unique lock that the other 5 pirates do not have.

## Chess tournament

A chess tournament has  $2^n$  players with skills  $1 > 2 > \dots > 2^n$ . It is organized as a knockout tournament, so that after each round only the winner proceeds to the next round. Except for the final, opponents in each round are drawn at random. Let's also assume that when two players meet in a game, the player with better skills always wins. What's the probability that players 1 and 2 will meet in the final?<sup>8</sup>

*Solution:* There are at least two approaches to solve the problem. The standard approach applies multiplication rule based on conditional probability, while a counting approach is far more efficient. (We will cover conditional probability in detail in the next section.)

Let's begin with the conditional probability approach, which is easier to grasp. Since there are  $2^n$  players, the tournament will have  $n$  rounds (including the final). For round 1, players  $2, 3, \dots, 2^n$  each have  $\frac{1}{2^n - 1}$  probability to be 1's rival, so the probability that

1 and 2 do not meet in round 1 is  $\frac{2^n - 2}{2^n - 1} = \frac{2 \times (2^{n-1} - 1)}{2^n - 1}$ . Condition on that 1 and 2 do not

meet in round 1,  $2^{n-1}$  players proceed to the 2nd round and the conditional probability

that 1 and 2 will not meet in round 2 is  $\frac{2^{n-1} - 2}{2^{n-1} - 1} = \frac{2 \times (2^{n-2} - 1)}{2^{n-1} - 1}$ . We can repeat the same

process until the  $(n-1)th$  round, in which there are  $2^2 (= 2^n / 2^{n-2})$  players left and the conditional probability that 1 and 2 will not meet in round  $(n-1)$  is

$$\frac{2^2 - 2}{2^2 - 1} = \frac{2 \times (2^{2-1} - 1)}{2^2 - 1}.$$

Let  $E_1$  be the event that 1 and 2 do not meet in round 1;

$E_2$  be the event that 1 and 2 do not meet in rounds 1 and 2;

...

$E_{n-1}$  be the event that 1 and 2 do not meet in round  $1, 2, \dots, n-1$ .

Apply the multiplication rule, we have

$$\begin{aligned} P(1 \text{ and } 2 \text{ meet in the } n\text{th game}) &= P(E_1) \times P(E_2 | E_1) \times \dots \times P(E_{n-1} | E_1 E_2 \dots E_{n-2}) \\ &= \frac{2 \times (2^{n-1} - 1)}{2^n - 1} \times \frac{2 \times (2^{n-2} - 1)}{2^{n-1} - 1} \times \dots \times \frac{2 \times (2^{2-1} - 1)}{2^2 - 1} = \frac{2^{n-1}}{2^n - 1} \end{aligned}$$

<sup>8</sup> Hint: Consider separating the players to two  $2^{n-1}$  subgroups. What will happen if player 1 and 2 in the same group? Or not in the same group?

Now let's move on to the counting approach. Figure 4.2A is the general case of what happens in the final. Player 1 always wins, so he will be in the final. From the figure, it is obvious that  $2^n$  players are separated to two  $2^{n-1}$ -player subgroups and each group will have one player reaching the final. As shown in Figure 4.2B, for player 2 to reach the final, he/she must be in a different subgroup from 1. Since any of the remaining players in  $2, 3, \dots, 2^n$  are likely to be one of the  $(2^{n-1} - 1)$  players in the same subgroup as player 1 or one of the  $2^{n-1}$  players in the subgroup different from player 1, the probability that 2 is in a different subgroup from 1 and that 1 and 2 will meet in the final is simply  $\frac{2^{n-1}}{2^n - 1}$ . Clearly, the counting approach provides not only a simpler solution but also more insight to the problem.

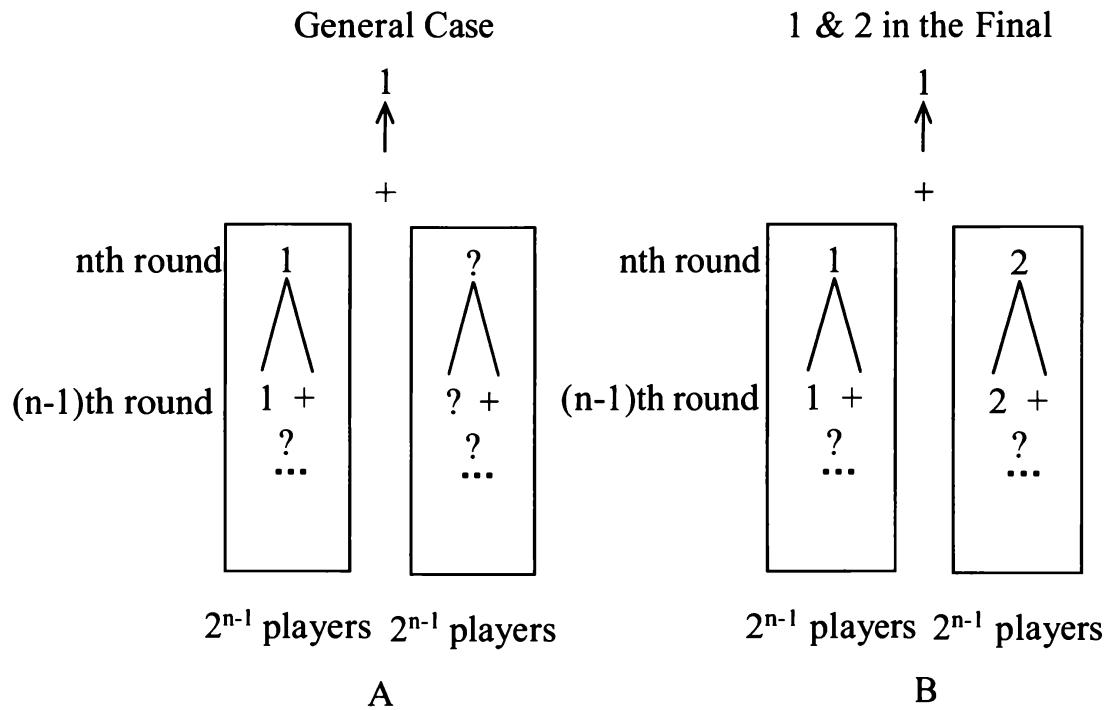


Figure 4.2A The general case of separating  $2^n$  players into  $2^{n-1}$ -player subgroups;  
4.2B The special case with players 1 and 2 in different groups

## Application letters

You're sending job applications to 5 firms: Morgan Stanley, Lehman Brothers, UBS, Goldman Sachs, and Merrill Lynch. You have 5 envelopes on the table neatly typed with names and addresses of people at these 5 firms. You even have 5 cover letters personalized to each of these firms. Your 3-year-old tried to be helpful and stuffed each cover letter into each of the envelopes for you. Unfortunately she randomly put letters

into envelopes without realizing that the letters are personalized. What is the probability that all 5 cover letters are mailed to the wrong firms?<sup>9</sup>

*Solution:* This problem is a classic example for the Inclusion-Exclusion Principle. In fact, a more general case is an example in Ross' textbook *First Course in Probability*.

Let's denote by  $E_i, i = 1, \dots, 5$  the event that the  $i$ -th letter has the correct envelope. Then

$P\left(\bigcup_{i=1}^5 E_i\right)$  is the probability that at least one letter has the correct envelope and

$1 - P\left(\bigcup_{i=1}^5 E_i\right)$  is the probability that all letters have the wrong envelopes.  $P\left(\bigcup_{i=1}^5 E_i\right)$  can

be calculated using the Inclusion-Exclusion Principle:

$$P\left(\bigcup_{i=1}^5 E_i\right) = \sum_{i=1}^5 P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} E_{i_2}) + \dots + (-1)^6 P(E_1 E_2 \dots E_5)$$

It's obvious that  $P(E_i) = \frac{1}{5}, \forall i = 1, \dots, 5$ . So  $\sum_{i=1}^5 P(E_i) = 1$ .

$P(E_{i_1} E_{i_2})$  is the event that both letter  $i_1$  and letter  $i_2$  have the correct envelope. The probability that  $i_1$  has the correct envelope is  $1/5$ ; Conditioned on that  $i_1$  has the correct envelope, the probability that  $i_2$  has the correct envelope is  $1/4$  (there are only 4 envelopes left). So  $P(E_{i_1} E_{i_2}) = \frac{1}{5} \times \frac{1}{4} = \frac{(5-2)!}{5!}$ .

There are  $\binom{5}{2} = \frac{5!}{2!(5-2)!}$  members of  $P(E_{i_1} E_{i_2})$  in  $\sum_{i_1 < i_2} P(E_{i_1} E_{i_2})$ , so we have

$$\sum_{i_1 < i_2} P(E_{i_1} E_{i_2}) = \frac{(5-2)!}{5!} \times \frac{5!}{2!(5-2)!} = \frac{1}{2!}$$

Similarly we have  $\sum_{i_1 < i_2 < i_3} P(E_{i_1} E_{i_2} E_{i_3}) = \frac{1}{3!}, \sum_{i_1 < i_2 < i_3 < i_4} P(E_{i_1} E_{i_2} E_{i_3} E_{i_4}) = \frac{1}{4!}$ , and

$$P(E_1 E_2 \dots E_5) = \frac{1}{5!}$$

---

<sup>9</sup> Hint: The complement is that at least one letter is mailed to the correct firm.

$$\therefore P\left(\bigcup_{i=1}^5 E_i\right) = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \frac{1}{5!} = \frac{19}{30}$$

So the probability that all 5 letters are mailed to the wrong firms is  $1 - P\left(\bigcup_{i=1}^5 E_i\right) = \frac{11}{30}$ .

## Birthday problem

How many people do we need in a class to make the probability that two people have the same birthday more than  $1/2$ ? (For simplicity, assume 365 days a year.)

*Solution:* The number is surprisingly small: 23. Let's say we have  $n$  people in the class. Without any restrictions, we have 365 possibilities for each individual's birthday. The basic principle of counting tells us that there are  $365^n$  possible sequences.

We want to find the number of those sequences that have no duplication of birthdays. For the first individual, we can choose any of the 365 days; but for the second, only 364 remaining choices left, ..., for the  $r$ th individual, there are  $365 - r + 1$  choices. So for  $n$  people there are  $365 \times 364 \times \dots \times (365 - n + 1)$  possible sequences where no two individuals have the same birthday. We need to have  $\frac{365 \times 364 \times \dots \times (365 - n + 1)}{365^n} < 1/2$  for the odds to be in our favor. The smallest such  $n$  is 23.

## 100th digit

What is the 100th digit to the right of the decimal point in the decimal representation of  $(1 + \sqrt{2})^{3000}$ ?<sup>10</sup>

*Solution:* If you still have not figure out the solution from the hint, here is one more hint:  $(1 + \sqrt{2})^n + (1 - \sqrt{2})^n$  is an integer when  $n = 3000$ .

Applying the binomial theorem for  $(x + y)^n$ , we have

$$(1 + \sqrt{2})^n = \sum_{k=0}^n \binom{n}{k} 1^{n-k} \sqrt{2}^k = \sum_{\substack{k=2j, 0 \leq j \leq \frac{n}{2}}} \binom{n}{k} 1^{n-k} \sqrt{2}^k + \sum_{\substack{k=2j+1, 0 \leq j < \frac{n}{2}}} \binom{n}{k} 1^{n-k} \sqrt{2}^k$$

---

<sup>10</sup> Hint:  $(1 + \sqrt{2})^2 + (1 - \sqrt{2})^2 = 6$ . What will happen to  $(1 - \sqrt{2})^{2n}$  as  $n$  becomes large?

$$(1-\sqrt{2})^n = \sum_{k=0}^n \binom{n}{k} 1^{n-k} (-\sqrt{2})^k = \sum_{\substack{k=2j, 0 \leq j \leq \frac{n}{2}}} \binom{n}{k} 1^{n-k} \sqrt{2}^k - \sum_{\substack{k=2j+1, 0 \leq j < \frac{n}{2}}} \binom{n}{k} 1^{n-k} \sqrt{2}^k$$

So  $(1+\sqrt{2})^n + (1-\sqrt{2})^n = 2 \sum_{\substack{k=2j, 0 \leq j \leq \frac{n}{2}}} \binom{n}{k} 1^{n-k} \sqrt{2}^k$ , which is always an integer. It is easy to see that  $0 < (1-\sqrt{2})^{3000} << 10^{-100}$ . So the 100th digit of  $(1+\sqrt{2})^n$  must be 9.

## Cubic of integer

Let  $x$  be an integer between 1 and  $10^{12}$ , what is the probability that the cubic of  $x$  ends with 11?<sup>11</sup>

*Solution:* All integers can be expressed as  $x = a + 10b$ , where  $a$  is the last digit of  $x$ . Applying the binomial theorem, we have  $x^3 = (a + 10b)^3 = a^3 + 30a^2b + 300ab^2 + 1000b^3$ .

The unit digit of  $x^3$  only depends on  $a^3$ . So  $a^3$  has a unit digit of 1. Only  $a = 1$  satisfies this requirement and  $a^3 = 1$ . Since  $a^3 = 1$ , the tenth digit only depends on  $30a^2b = 30b$ . So we must have that  $3b$  ends in 1, which requires the last digit of  $b$  to be 7. Consequently, the last two digits of  $x$  should be 71, which has a probability of 1% for integers between 1 and  $10^{12}$ .

## 4.3 Conditional Probability and Bayes' formula

Many financial transactions are responses to probability adjustments based on new—and most likely incomplete—information. Conditional probability surely is one of the most popular test subjects in quantitative interviews. So in this section, we focus on basic conditional probability definitions and theorems.

**Conditional probability  $P(A|B)$ :** If  $P(B) > 0$ , then  $P(A|B) = \frac{P(AB)}{P(B)}$  is the fraction of  $B$  outcomes that are also  $A$  outcomes.

**Multiplication Rule:**  $P(E_1 E_2 \cdots E_n) = P(E_1)P(E_2 | E_1)P(E_3 | E_1 E_2) \cdots P(E_n | E_1 \cdots E_{n-1})$ .

---

<sup>11</sup> Hint: The last two digits of  $x^3$  only depend on the last two digits of  $x$ .

**Law of total probability:** for any mutually exclusive events  $\{F_i\}$ ,  $i = 1, 2, \dots, n$ , whose union is the entire sample space ( $F_i \cap F_j = \emptyset, \forall i \neq j; \bigcup_{i=1}^n F_i = \Omega$ ), we have

$$\begin{aligned} P(E) &= P(EF_1) + P(EF_2) + \dots + P(EF_n) = \sum_{i=1}^n P(E|F_i)P(F_i) \\ &= P(E|F_1)P(F_1) + P(E|F_2)P(F_2) + \dots + P(E|F_n)P(F_n) \end{aligned}$$

**Independent events:**  $P(EF) = P(E)P(F) \Rightarrow P(EF^C) = P(E)P(F^C)$ .

Independence is a symmetric relation: X is independent of Y  $\Leftrightarrow$  Y is independent of X.

**Bayes' Formula:**  $P(F_j | E) = \frac{P(E | F_j)P(F_j)}{\sum_{i=1}^n P(E | F_i)P(F_i)}$  if  $F_i$ ,  $i = 1, \dots, n$ , are mutually exclusive events whose union is the entire sample space.

As the following examples will demonstrate, not all conditional probability problems have intuitive solutions. Many demand logical analysis instead.

## Boys and girls

Part A. A company is holding a dinner for working mothers with at least one son. Ms. Jackson, a mother with two children, is invited. What is the probability that both children are boys?

*Solution:* The sample space of two children is given by  $\Omega = \{(b,b), (b,g), (g,b), (g,g)\}$  (e.g.,  $(g,b)$  means the older child is a girl and the younger child a boy), and each outcome has the same probability. Since Ms. Jackson is invited, she has at least one son. Let  $B$  be the event that at least one of the children is a boy and  $A$  be the event that both children are boys, we have

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{(b,b)\})}{P(\{(b,b), (b,g), (g,b)\})} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Part B. Your new colleague, Ms. Parker is known to have two children. If you see her walking with one of her children and that child is a boy, what is the probability that both children are boys?

*Solution:* the other child is equally likely to be a boy or a girl (independent of the boy you've seen), so the probability that both children are boys is 1/2.

Notice the subtle difference between part *A* and part *B*. In part *A*, the problem essentially asks given there is at least one boy in two children, what is the conditional probability that both children are boys. Part *B* asks that given one child is a boy, what is the conditional probability that the other child is also a boy. For both parts, we need to assume that each child is equal likely to be a boy or a girl.

## All-girl world?

In a primitive society, every couple prefers to have a baby girl. There is a 50% chance that each child they have is a girl, and the genders of their children are mutually independent. If each couple insists on having more children until they get a girl and once they have a girl they will stop having more children, what will eventually happen to the fraction of girls in this society?

*Solution:* It was surprising that many interviewees—include many who studied probability—have the misconception that there will be more girls. Do not let the word “prefer” and a wrong intuition misguide you. The fraction of baby girls are driven by nature, or at least the *X* and *Y* chromosomes, not by the couples’ preference. You only need to look at the key information: 50% and independence. Every new-born child has equal probability of being a boy or a girl regardless of the gender of any other children. So the fraction of girls born is always 50% and the fractions of girls in the society will stay stable at 50%.

## Unfair coin

You are given 1000 coins. Among them, 1 coin has heads on both sides. The other 999 coins are fair coins. You randomly choose a coin and toss it 10 times. Each time, the coin turns up heads. What is the probability that the coin you choose is the unfair one?

*Solution:* This is a classic conditional probability question that uses Bayes’ theorem. Let *A* be the event that the chosen coin is the unfair one, then  $A^c$  is the event that the chosen coin is a fair one. Let *B* be the event that all ten tosses turn up heads. Apply Bayes’ theorem we have  $P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$ .

The priors are  $P(A) = 1/1000$  and  $P(A^c) = 999/1000$ . If the coin is unfair, it always turns up heads, so  $P(B|A) = 1$ . If the coin is fair, each time it has 1/2 probability turning

up heads. So  $P(B | A^c) = (1/2)^{10} = 1/1024$ . Plug in all the available information and we have the answer:

$$P(A | B) = \frac{P(B | A)P(A)}{P(B | A)P(A) + P(B | A^c)P(A^c)} = \frac{1/1000 \times 1}{1/1000 \times 1 + 999/1000 \times 1/1024} \approx 0.5.$$

## Fair probability from an unfair coin

If you have an unfair coin, which may bias toward either heads or tails at an unknown probability, can you generate even odds using this coin?

*Solution:* Unlike fair coins, we clearly can not generate even odds with one toss using an unfair coin. How about using 2 tosses? Let  $p_H$  be the probability the coin will yield head, and  $p_T$  be the probability the coin will yield tails ( $p_H + p_T = 1$ ). Consider two independent tosses. We have four possible outcomes  $HH$ ,  $HT$ ,  $TH$  and  $TT$  with probabilities  $P(HH) = p_H p_H$ ,  $P(HT) = p_H p_T$ ,  $P(TH) = p_T p_H$ , and  $P(TT) = p_T p_T$ .

So we have  $P(HT) = P(TH)$ . By assigning  $HT$  to winning and  $TH$  to losing, we can generate even odds.<sup>12</sup>

## Dart game

Jason throws two darts at a dartboard, aiming for the center. The second dart lands farther from the center than the first. If Jason throws a third dart aiming for the center, what is the probability that the third throw is farther from the center than the first? Assume Jason's skillfulness is constant.

*Solution:* A standard answer directly applies the conditional probability by enumerating all possible outcomes. If we rank the three darts' results from the best (A) to the worst (C), there are 6 possible outcomes with equal probability:

---

<sup>12</sup> I should point out that this simple approach is not the most efficient approach since I am disregarding the cases  $HH$  and  $TT$ . When the coin has high bias (one side is far more likely than the other side to occur), the method may take many runs to generate one useful result. For more complex algorithm that increasing efficiency, please refer to *Tree Algorithms for Unbiased Coin Tossing with a Biased Coin* by Quentin F. Stout and Bette L. Warren, Annals of Probability 12 (1984), pp. 212-222.

Outcome	1	2	3	4	5	6
1st throw	A	B	A	C	B	C
2nd throw	B	A	C	A	C	B
3rd throw	C	C	B	B	A	A

The information from the first two throws eliminates outcomes 2, 4 and 6. Conditioned on outcomes 1, 3, and 5, the outcomes that the 3rd throw is worse than the 1st throw are outcomes 1 and 3. So there is  $2/3$  probability that the third throw is farther from the center than the first.

This approach surely is reasonable. Nevertheless, it is not an efficient approach. When the number of darts is small, we can easily enumerate all outcomes. What if it is a more complex version of the original problem:

Jason throws  $n$  ( $n \geq 5$ ) darts at a dartboard, aiming for the center. Each subsequent dart is farther from the center than the first dart. If Jason throws the  $(n+1)$ th dart, what is the probability that it is also farther from the center than his first?

This question is equivalent to a simple question: what is the probability that the  $(n+1)$ th throw is not the best among all  $(n+1)$  throws? Since the 1st throw is the best among the first  $n$  throws, essentially I am saying the event that the  $(n+1)$ th throw is the best of all  $(n+1)$  throws (let's call it  $A_{n+1}$ ) is independent of the event that the 1st throw is the best of the first  $n$  throws (let's call it  $A_1$ ). In fact,  $A_{n+1}$  is independent of the order of the first  $n$  throws. Are these two events really independent? The answer is a resounding yes. If it is not obvious to you that  $A_{n+1}$  is independent of the order of the first  $n$  throws, let's look at it another way: the order of the first  $n$  throws is independent of  $A_{n+1}$ . Surely this claim is conspicuous. But independence is symmetric! Since the probability of  $A_{n+1}$  is  $1/(n+1)$ , the probability that  $(n+1)$ th throw is not the best is  $n/(n+1)$ .<sup>13</sup>

For the original version, three darts are thrown independently, they each have a  $1/3$  chance of being the best throw. As long as the third dart is not the best throw, it will be worse than the first dart. Therefore the answer is  $2/3$ .

## Birthday line

At a movie theater, a whimsical manager announces that she will give a free ticket to the first person in line whose birthday is the same as someone who has already bought a ticket. You are given the opportunity to choose any position in line. Assuming that you

---

<sup>13</sup> Here you can again use symmetry argument: each throw is equally likely to be the best.

don't know anyone else's birthday and all birthdays are distributed randomly throughout the year (assuming 365 days in a year), what position in line gives you the largest chance of getting the free ticket?<sup>14</sup>

*Solution:* If you have solved the problem that no two people have the same birthday in an  $n$ -people group, this new problem is just a small extension. Assume that you choose to be the  $n$ -th person in line. In order for you to get the free ticket, all of the first  $n-1$  individuals in line must have different birthdays and your birthday needs to be the same as one of those  $n-1$  individuals.

$$p(n) = p(\text{first } n-1 \text{ people have no same birthday}) \times p(\text{yours among those } n-1 \text{ birthdays}) \\ = \frac{365 \times 364 \times \dots \times (365-n+2)}{365^{n-1}} \times \frac{n-1}{365}$$

It is intuitive to argue that when  $n$  is small, increasing  $n$  will increase your chance of getting the free ticket since the increase of  $p(\text{yours among those } n-1 \text{ birthdays})$  is more significant than the decrease in  $p(\text{first } n-1 \text{ people have no same birthday})$ . So when  $n$  is small, we have  $P(n) > P(n-1)$ . As  $n$  increases, gradually the negative impact of  $p(\text{first } n-1 \text{ people have no same birthday})$  will catch up and at a certain point we will have  $P(n+1) < P(n)$ . So we need to find such an  $n$  that satisfies  $P(n) > P(n-1)$  and  $P(n) > P(n+1)$ .

$$P(n-1) = \frac{365}{365} \times \frac{364}{365} \times \dots \times \frac{365-(n-3)}{365} \times \frac{n-2}{365}$$

$$P(n) = \frac{365}{365} \times \frac{364}{365} \times \dots \times \frac{365-(n-2)}{365} \times \frac{n-1}{365}$$

$$P(n+1) = \frac{365}{365} \times \frac{364}{365} \times \dots \times \frac{365-(n-2)}{365} \times \frac{365-(n-1)}{365} \times \frac{n}{365}$$

$$\left. \begin{aligned} P(n) > P(n-1) &\Rightarrow \frac{365-(n-2)}{365} \times \frac{n-1}{365} > \frac{n-2}{365} \\ \text{Hence, } P(n) > P(n+1) &\Rightarrow \frac{n-1}{365} > \frac{365-(n-1)}{365} \times \frac{n}{365} \end{aligned} \right\} \Rightarrow \begin{cases} n^2 - 3n - 363 < 0 \\ n^2 - n - 365 > 0 \end{cases} \Rightarrow n = 20$$

You should be the 20th person in line.

---

<sup>14</sup> Hint: If you are the  $n$ -th person in line, to get the free ticket, the first  $(n-1)$  people in line must not have the same birthday and you must have the same birthday as one of them.

## Dice order

We throw 3 dice one by one. What is the probability that we obtain 3 points in strictly increasing order?<sup>15</sup>

*Solution:* To have 3 points in strictly increasing order, first all three points must be different numbers. Conditioned on three different numbers, the probability of strictly increasing order is simply  $1/3! = 1/6$  (one specific sequence out of all possible permutations). So we have

$$\begin{aligned} P &= P(\text{different numbers in all three throws}) \times P(\text{increasing order} | 3 \text{ different numbers}) \\ &= (1 \times \frac{5}{6} \times \frac{4}{6}) \times \frac{1}{6} = 5 / 54 \end{aligned}$$

## Monty Hall problem

Monty Hall problem is a probability puzzle based on an old American show *Let's Make a Deal*. The problem is named after the show's host. Suppose you're on the show now, and you're given the choice of 3 doors. Behind one door is a car; behind the other two, goats. You don't know ahead of time what is behind each of the doors.

You pick one of the doors and announce it. As soon as you pick the door, Monty opens one of the other two doors that he knows has a goat behind it. Then he gives you the option to either keep your original choice or switch to the third door. Should you switch? What is the probability of winning a car if you switch?

*Solution:* If you don't switch, whether you win or not is independent of Monty's action of showing you a goat, so your probability of winning is  $1/3$ . What if you switch? Many would argue that since there are only two doors left after Monty shows a door with goat, the probability of winning is  $1/2$ . But is this argument correct?

If you look at the problem from a different perspective, the answer becomes clear. Using a switching strategy, you win the car if and only if you originally pick a door with a goat, which has a probability of  $2/3$  (You pick a door with a goat, Monty shows a door with another goat, so the one you switch to must have a car behind it). If you originally picked the door with the car, which has a probability of  $1/3$ , you will lose by switching. So your probability of winning by switching is actually  $2/3$ .

---

<sup>15</sup> Hint: To obtain 3 points in strictly increasing order, the 3 points must be different. For 3 different points in a sequence, strictly increasing order is one of the possible permutations.

## Amoeba population

There is a one amoeba in a pond. After every minute the amoeba may die, stay the same, split into two or split into three with equal probability. All its offspring, if it has any, will behave the same (and independent of other amoebas). What is the probability the amoeba population will die out?

*Solution:* This is just another standard conditional probability problem once you realize we need to derive the probability conditioned on what happens to the amoeba one minute later. Let  $P(E)$  be the probability that the amoeba population will die out and apply the law of total probability conditioned on what happens to the amoeba one minute later:

$$P(E) = P(E | F_1)P(F_1) + P(E | F_2)P(F_2) + \dots + P(E | F_n)P(F_n).$$

For the original amoeba, as stated in the question, there are four possible mutually exclusive events each with probability  $1/4$ . Let's denote  $F_1$  as the event the amoeba dies;  $F_2$  as the event that it stays the same;  $F_3$  as the event that it splits into two;  $F_4$  as the event that it splits into three. For event  $F_1$ ,  $P(E | F_1) = 1$  since no amoeba is left.  $P(E | F_2) = P(E)$  since the state is the same as the beginning. For  $F_3$ , there are two amoebas; either behaves the same as the original one. The total amoeba population will die only if both amoebas die out. Since they are independent, the probability that they both will die out is  $P(E)^2$ . Similarly we have  $P(F_4) = P(E)^3$ . Plug in all the numbers, the equation becomes  $P(E) = 1/4 \times 1 + 1/4 \times P(E) + 1/4 \times P(E)^2 + 1/4 \times P(E)^3$ . Solve this equation with the restriction  $0 < P(E) < 1$ , and we will get  $P(E) = \sqrt{2} - 1 \approx 0.414$  (The other two roots of the equation are 1 and  $-\sqrt{2} - 1$  ).

## Candies in a jar

You are taking out candies one by one from a jar that has 10 red candies, 20 blue candies, and 30 green candies in it. What is the probability that there are at least 1 blue candy and 1 green candy left in the jar when you have taken out all the red candies?<sup>16</sup>

*Solution:* At first look, this problem appears to be a combinatorial one. However, a conditional probability approach gives a much more intuitive answer. Let  $T_r$ ,  $T_b$  and  $T_g$

---

<sup>16</sup> Hint: If there are at least 1 blue candy and 1 green candy left, the last red candy must have been removed before the last blue candy and the last green candy in the sequence of 60 candies. What is the probability that the blue candy is the last one in the 60-candy sequence? Conditioned on that, what is the probability that the last green candy is the last one in the 30-candy sequence (10 red, 20 green)? What if the green candy is the last one in the 60-candy sequence?

be the number that the last red, blue, and green candies are taken out respectively. To have at least 1 blue candy and 1 green candy left when all the red candies are taken out, we need to have  $T_r < T_b$  and  $T_r < T_g$ . In other words, we want to derive  $P(T_r < T_b \cap T_r < T_g)$ . There are two mutually exclusive events that satisfy  $T_r < T_b$  and  $T_r < T_g$ :  $T_r < T_b < T_g$  and  $T_r < T_g < T_b$ .

$$\therefore P(T_r < T_b \cap T_r < T_g) = P(T_r < T_b < T_g) + P(T_r < T_g < T_b)$$

$T_r < T_b < T_g$  means that the last candy is green ( $T_g = 60$ ). Since each of the 60 candies are equally likely to be the last candy and among them 30 are green ones, we have  $P(T_g = 60) = \frac{30}{60}$ . Conditioned on  $T_g = 60$ , we need  $P(T_r < T_b | T_g = 60)$ . Among the 30 red and blue candies, each candy is again equally likely to be the last candy and there are 20 blue candies, so  $P(T_r < T_b | T_g = 60) = \frac{20}{30}$  and  $P(T_r < T_b < T_g) = \frac{30}{60} \times \frac{20}{30}$ . Similarly,

$$\text{we have } P(T_r < T_g < T_b) = \frac{20}{60} \times \frac{30}{40}.$$

Hence,

$$P(T_r < T_b \cap T_r < T_g) = P(T_r < T_b < T_g) + P(T_r < T_g < T_b) = \frac{30}{60} \times \frac{20}{30} + \frac{20}{60} \times \frac{30}{40} = \frac{7}{12}.$$

## Coin toss game

Two players,  $A$  and  $B$ , alternatively toss a fair coin ( $A$  tosses the coin first, then  $B$  tosses the coin, then  $A$ , then  $B$ ...). The sequence of heads and tails is recorded. If there is a head followed by a tail ( $HT$  subsequence), the game ends and the person who tosses the tail wins. What is the probability that  $A$  wins the game?<sup>17</sup>

*Solution:* Let  $P(A)$  be the probability that  $A$  wins; then the probability that  $B$  wins is  $P(B) = 1 - P(A)$ . Let's condition  $P(A)$  on  $A$ 's first toss, which has 1/2 probability of  $H$  (heads) and 1/2 probability of  $T$  (tails).

$$P(A) = \frac{1}{2}P(A|H) + \frac{1}{2}P(A|T)$$

If  $A$ 's first toss is  $T$ , then  $B$  essentially becomes the first to toss (An  $H$  is required for the  $HT$  subsequence). So we have  $P(A|T) = P(B) = 1 - P(A)$ .

If  $A$ 's first toss ends in  $H$ , let's further condition on  $B$ 's first toss.  $B$  has 1/2 probability of getting  $T$ , in that case  $A$  loses. For the 1/2 probability that  $B$  gets  $H$ ,  $B$  essentially

---

<sup>17</sup> Hint: condition on the result of  $A$ 's first toss and use symmetry.

becomes the first one to toss an  $H$ . In that case,  $A$  has  $(1 - P(A|H))$  probability of winning. So  $P(A|H) = 1/2 \times 0 + 1/2(1 - P(A|H)) \Rightarrow P(A|H) = 1/3$

Combining all the available information, we have

$$P(A) = 1/2 \times 1/3 + 1/2(1 - P(A)) \Rightarrow P(A) = 4/9.$$

Sanity check: we can see that  $P(A) < 1/2$ , which is reasonable since  $A$  cannot win in his first toss, yet  $B$  has  $1/4$  probability to win in her first toss.

## Russian roulette series

Let's play a traditional version of Russian roulette. A single bullet is put into a 6-chamber revolver. The barrel is randomly spun so that each chamber is equally likely to be under the hammer. Two players take turns to pull the trigger—with the gun unfortunately pointing at one's own head—without further spinning until the gun goes off and the person who gets killed loses. If you, one of the players, can choose to go first or second, how will you choose? And what is your probability of loss?

*Solution:* Many people have the wrong impression that the first person has higher probability of loss. After all, the first player has a  $1/6$  chance of getting killed in the first round before the second player starts. Unfortunately, this is one of the few times that intuition is wrong. Once the barrel is spun, the position of the bullet is fixed. If you go first, you lose if and only if the bullet is in chamber 1, 3 and 5. So the probability that you lose is the same as the second player,  $1/2$ . In that sense, whether to go first or second does not matter.

Now, let's change the rule slightly. We will spin the barrel again after every trigger pull. Will you choose to be the first or the second player? And what is your probability of loss?

*Solution:* The difference is that each run now becomes independent. Assume that the first player's probability of losing is  $p$ , then the second player's probability of losing is  $1 - p$ . Let's condition the probability on the first person's first trigger pull. He has  $1/6$  probability of losing in this run. Otherwise, he essentially becomes the second player in the game with new (conditional) probability of losing  $1 - p$ . That happens with probability  $5/6$ . That gives us  $p = 1 \times 1/6 + (1 - p) \times 5/6 \Rightarrow p = 6/11$ . So you should choose to be the second player and have  $5/11$  probability of losing.

If instead of one bullet, two bullets are randomly put in the chamber. Your opponent played the first and he was alive after the first trigger pull. You are given the option whether to spin the barrel. Should you spin the barrel?

*Solution:* if you spin the barrel, the probability that you will lose in this round is 2/6. If you don't spin the barrel, there are only 5 chambers left and your probability of losing in this round (conditioned on that your opponent survived) is 2/5. So you should spin the barrel.

What if the two bullets are randomly put in two consecutive positions? If your opponent survived his first round, should you spin the barrel?

*Solution:* Now we have to condition our probability on the fact that the positions of the two bullets are consecutive. As shown in Figure 4.3, let's label the empty chambers as 1, 2, 3 and 4; label the ones with bullets 5 and 6. Since your opponent survived the first round, the possible position he encountered is 1, 2, 3 or 4 with equal probability. With 1/4 chance, the next one is a bullet (the position was 4). So if you don't spin, the chance of survival is 3/4. If you spin the barrel, each position has equal probability of being chosen, and your chance of survival is only 2/3. So you should not spin the barrel.

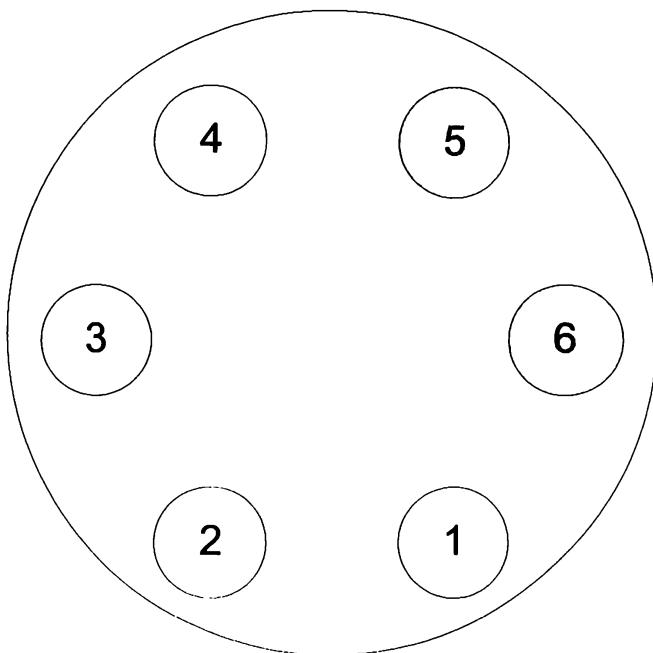


Figure 4.3 Russian roulette with two consecutive bullets.

## Aces

Fifty-two cards are randomly distributed to 4 players with each player getting 13 cards. What is the probability that each of them will have an ace?

*Solution:* The problem can be answered using standard counting methods. To distribute 52 cards to 4 players with 13 cards each has  $\frac{52!}{13!13!13!13!}$  permutations. If each player

needs to have one ace, we can distribute the aces first, which has  $4!$  ways. Then we distribute the rest 48 cards to 4 players with 12 cards each, which has  $\frac{48!}{12!12!12!12!}$  permutations. So the probability that each of them will have an Ace is

$$4! \times \frac{48!}{12!12!12!12!} \div \frac{52!}{13!13!13!13!} = \frac{52}{52} \times \frac{39}{51} \times \frac{26}{50} \times \frac{13}{49}.$$

The logic becomes clearer if we use a conditional probability approach. Let's begin with any one of the four aces; it has probability  $52/52=1$  of belonging to a pile. The second ace can be any of the remaining 51 cards, among which 39 belong to a pile different from the first ace. So the probability that the second ace is not in the pile of the first ace is  $39/51$ . Now there are 50 cards left, among which 26 belong to the other two piles. So the conditional probability that the third ace is in one of the other 2 piles given the first two aces are already in different piles is  $26/50$ . Similarly, the conditional probability that the fourth ace is in the pile different from the first three aces given that the first three aces are in different piles is  $13/49$ . So the probability that each pile has an ace is  $1 \times \frac{39}{51} \times \frac{26}{50} \times \frac{13}{49}$ .

## Gambler's ruin problem

A gambler starts with an initial fortune of  $i$  dollars. On each successive game, the gambler wins \$1 with probability  $p$ ,  $0 < p < 1$ , or loses \$1 with probability  $q = 1 - p$ . He will stop if he either accumulates  $N$  dollars or loses all his money. What is the probability that he will end up with  $N$  dollars?

*Solution:* This is a classic textbook probability problem called the Gambler's Ruin Problem. Interestingly, it is still widely used in quantitative interviews.

From any initial state  $i$  (the dollars the gambler has),  $0 \leq i \leq N$ , let  $P_i$  be the probability that the gambler's fortune will reach  $N$  instead of 0. The next state is either  $i+1$  with probability  $p$  or  $i-1$  with probability  $q$ . So we have

$$P_i = pP_{i+1} + qP_{i-1} \Rightarrow P_{i+1} - P_i = \frac{q}{p}(P_i - P_{i-1}) = \left(\frac{q}{p}\right)^2(P_{i-1} - P_{i-2}) = \dots = \left(\frac{q}{p}\right)^i(P_1 - P_0)$$

We also have the boundary probabilities  $P_0 = 0$  and  $P_N = 1$ .

So starting from  $P_2$ , we can successively evaluate  $P_i$  as an expression of  $P_1$ :

$$P_1 = pP_2 + qP_0 \Rightarrow P_2 = \frac{1}{p}P_1 = \left[1 + \frac{q}{p}\right]P_1$$

$$P_3 = \left[1 + \frac{q}{p} + \left(\frac{q}{p}\right)^2\right]P_1$$

...

$$P_i = \left[1 + \frac{q}{p} + \dots + \left(\frac{q}{p}\right)^{i-1}\right]P_1$$

Extending this expression to  $P_N$ , we have

$$P_N = 1 = \left[1 + \frac{q}{p} + \dots + \left(\frac{q}{p}\right)^{N-1}\right]P_1 = \begin{cases} \frac{1 - (q/p)^N}{1 - q/p} P_1, & \text{if } q/p \neq 1 \\ NP_1, & \text{if } q/p = 1 \end{cases}$$

$$\Rightarrow P_1 = \begin{cases} \frac{1 - q/p}{1 - (q/p)^N}, & \text{if } q/p \neq 1 \\ 1/N, & \text{if } q/p = 1 \end{cases} \Rightarrow P_i = \begin{cases} \frac{1 - (q/p)^i}{1 - (q/p)^N} P_1, & \text{if } p \neq 1/2 \\ i/N, & \text{if } p = 1/2 \end{cases}$$

## Basketball scores

A basketball player is taking 100 free throws. She scores one point if the ball passes through the hoop and zero point if she misses. She has scored on her first throw and missed on her second. For each of the following throw the probability of her scoring is the fraction of throws she has made so far. For example, if she has scored 23 points after the 40th throw, the probability that she will score in the 41th throw is 23/40. After 100 throws (including the first and the second), what is the probability that she scores exactly 50 baskets?<sup>18</sup>

*Solution:* Let  $(n, k)$ ,  $1 \leq k \leq n$ , be the event that the player scores  $k$  baskets after  $n$  throws and  $P_{n,k} = P((n, k))$ . The solution is surprisingly simple if we use an induction approach starting with  $n = 3$ . The third throw has 1/2 probability of scoring. So we have  $P_{3,1} = 1/2$  and  $P_{3,2} = 1/2$ . For the case when  $n = 4$ , let's apply the law of total probability

---

<sup>18</sup> Hint: Again, do not let the number 100 scares you. Start with smallest  $n$ , solve the problem; try to find a pattern by increasing  $n$ ; and prove the pattern using induction.

$$\begin{cases} P_{4,1} = P((4,1)|(3,1)) \times P_{3,1} + P((4,1)|(3,2)) \times P_{3,2} = \frac{2}{3} \times \frac{1}{2} + 0 \times \frac{1}{2} = \frac{1}{3} \\ P_{4,2} = P((4,2)|(3,1)) \times P_{3,1} + P((4,2)|(3,2)) \times P_{3,2} = \frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times \frac{1}{2} = \frac{1}{3} \\ P_{4,3} = P((4,3)|(3,1)) \times P_{3,1} + P((4,3)|(3,2)) \times P_{3,2} = 0 \times \frac{1}{2} + \frac{2}{3} \times \frac{1}{2} = \frac{1}{3} \end{cases}$$

The results indicate that  $P_{n,k} = \frac{1}{n-1}$ ,  $\forall k = 1, 2, \dots, n-1$ , and give the hint that the law of total probability can be used in the induction step.

Induction step: given that  $P_{n,k} = \frac{1}{n-1}$ ,  $\forall k = 1, 2, \dots, n-1$ , we need to prove

$P_{n+1,k} = \frac{1}{(n+1)-1} = \frac{1}{n}$ ,  $\forall k = 1, 2, \dots, n$ . To show it, simply apply the law of total probability:

$$\begin{aligned} P_{n+1,k} &= P(\text{miss} | (n, k)) P_{n,k} + P(\text{score} | (n, k-1)) P_{n,k-1} \\ &= \left(1 - \frac{k}{n}\right) \frac{1}{n-1} + \frac{k-1}{n} \frac{1}{n-1} = \frac{1}{n} \end{aligned}$$

The equation is also applicable to the  $P_{n+1,1}$  and  $P_{n+1,n}$ , although in these cases  $\frac{k-1}{n} = 0$

and  $\left(1 - \frac{k}{n}\right) = 0$ , respectively. So we have  $P_{n,k} = \frac{1}{n-1}$ ,  $\forall k = 1, 2, \dots, n-1$  and  $\forall n \geq 2$ .

Hence,  $P_{100,50} = 1/99$ .

## Cars on road

If the probability of observing at least one car on a highway during any 20-minute time interval is 609/625, then what is the probability of observing at least one car during any 5-minute time interval? Assume that the probability of seeing a car at any moment is uniform (constant) for the entire 20 minutes.

*Solution:* We can break down the 20-minute interval into a sequence of 4 non-overlapping 5-minute intervals. Because of constant default probability (of observing a car), the probability of observing a car in any 5-minute interval is constant. Let's denote the probability to be  $p$ , then the probability that in any 5-minute interval we do not observe a car is  $1-p$ .

The probability that we do not observe any car in all four of such independent 5-minute intervals is  $(1-p)^4 = 1 - 609/625 = 16/625$ , which gives  $p = 3/5$ .

## 4.4 Discrete and Continuous Distributions

In this section, we review a variety of distribution functions for random variables that are widely used in quantitative modeling. Although it may not be necessary to memorize the properties of these distributions, having an intuitive understanding of the distributions and having the ability to quickly derive important properties are valuable skills in practice. As usual, let's begin with the theories:

### Common function of random variables

Table 4.1 summarizes how the basic properties of discrete and continuous random variables are defined or calculated. These are the basics you should commit to memory.

Random variable ( $X$ )	Discrete	Continuous <sup>19</sup>
Cumulative distribution function/cdf	$F(a) = P\{X \leq a\}$	$F(a) = \int_{-\infty}^a f(x)dx$
Probability mass function /pmf Probability density function /pdf	pmf: $p(x) = P\{X = x\}$	pdf: $f(x) = \frac{d}{dx} F(x)$
Expected value/ $E[X]$	$\sum_{x:p(x)>0} xp(x)$	$\int_{-\infty}^{\infty} xf(x)dx$
Expected value of $g(X)$ / $E[g(X)]$	$\sum_{x:p(x)>0} g(x)p(x)$	$\int_{-\infty}^{\infty} g(x)f(x)dx$
Variance of $X$ / $var(X)$	$E[(X - E[X])^2] = E[X^2] - (E[X])^2$	
Standard deviation of $X$ / $std(X)$	$\sqrt{var(X)}$	

Table 4.1 Basic properties of discrete and continuous random variables

### Discrete random variables

Table 4.2 includes some of the most widely-used discrete distributions. Discrete uniform random variable represents the occurrence of a value between number  $a$  and  $b$  when all values in the set  $\{a, a+1, \dots, b\}$  have equal probability. Binomial random variable represents the number of successes in a sequence of  $n$  experiments when each trial is

---

<sup>19</sup> For continuous random variables,  $P(X = x) = 0$ ,  $\forall x \in (-\infty, \infty)$ , so  $P\{X \leq x\} = P\{X < x\}$ .

independently a success with probability  $p$ . Poisson random variable represents the number of events occurring in a fixed period of time with the expected number of occurrences  $\lambda t$  when events occur with a known average rate  $\lambda$  and are independent of the time since the last event. Geometric random variable represents the trial number ( $n$ ) to get the first success when each trial is independently a success with probability  $p$ . Negative Binomial random variable represents the trial number to get to the  $r$ -th success when each trial is independently a success with probability  $p$ .

Name	Probability mass function (pmf)	$E[X]$	$\text{var}(X)$
Uniform	$P(x) = \frac{1}{b-a+1}, \quad x = a, a+1, \dots, b$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2 - 1}{12}$
Binomial	$P(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n$	$np$	$np(1-p)$
Poisson	$P(x) = \frac{e^{-\lambda t} (\lambda t)^x}{x!}, \quad x = 0, 1, \dots^{20}$	$\lambda t$	$\lambda t$
Geometric	$P(x) = (1-p)^{x-1} p, \quad x = 1, 2, \dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative Binomial	$P(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r, r+1, \dots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

Table 4.2 Probability mass function, expected value and variance of discrete random variables

### Continuous random variables

Table 4.3 includes some of the commonly encountered continuous distributions. Uniform distribution describes a random variable uniformly distributed over the interval  $[a, b]$ . Because of the central limit theorem, normal distribution/Gaussian distribution is by far the most popular continuous distribution. Exponential distribution models the arrival time of an event if it has a constant arrival rate  $\lambda$ . Gamma distribution with parameters  $(\alpha, \lambda)$  often arises, in practice, as the distribution of the amount of time one has to wait until a total of  $n$  events occur. Beta distributions are used to model events

---

<sup>20</sup> Here we use the product of arrival rate  $\lambda$  and time  $t$  to define the parameter (expected value) since it is the definition used in many Poisson process studies.

that are constrained within a defined interval. By adjusting the shape parameters  $\alpha$  and  $\beta$ , it can model different shapes of probability density functions.<sup>21</sup>

Name	Probability density function (pdf)	$E[X]$	$\text{var}(X)$
Uniform	$\frac{1}{b-a}, \quad a \leq x \leq b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal	$\frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad x \in (-\infty, \infty)$	$\mu$	$\sigma^2$
Exponential	$\lambda e^{-\lambda x}, \quad x \geq 0$	$1/\lambda$	$1/\lambda^2$
Gamma	$\frac{\lambda e^{-\lambda x} (\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, \quad x \geq 0, \quad \Gamma(a) = \int_0^\infty e^{-y} y^{a-1}$	$\alpha/\lambda$	$\alpha/\lambda^2$
Beta	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 < x < 1$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$

Table 4.3 Probability density function, expected value and variance of continuous random variables

## Meeting probability

Two bankers each arrive at the station at some random time between 5:00 am and 6:00 am (arrival time for either banker is uniformly distributed). They stay exactly five minutes and then leave. What is the probability they will meet on a given day?

*Solution:* Assume banker  $A$  arrives  $X$  minutes after 5:00 am and  $B$  arrives  $Y$  minutes after 5:00 am.  $X$  and  $Y$  are independent uniform distribution between 0 and 60. Since both only stay exactly five minutes, as shown in Figure 4.4,  $A$  and  $B$  meet if and only if  $|X - Y| \leq 5$ .

So the probability that  $A$  and  $B$  will meet is simply the area of the shadowed region divided by the area of the square (the rest of the region can be combined to a square with size length 55):  $\frac{60 \times 60 - 2 \times (1/2 \times 55 \times 55)}{60 \times 60} = \frac{(60+55) \times (60-55)}{60 \times 60} = \frac{23}{144}$ .

---

<sup>21</sup> For example, beta distribution is widely used in modeling loss given default in risk management. If you are familiar with Bayesian statistics, you will also recognize it as a popular conjugate prior function.

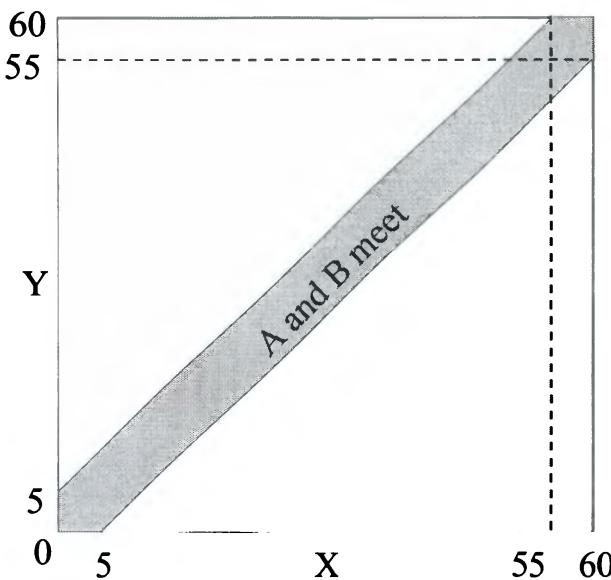


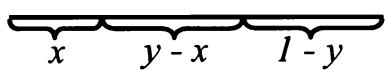
Figure 4.4 Distributions of Banker A's and Banker B's arrival times

### Probability of triangle

A stick is cut twice randomly (each cut point follows a uniform distribution on the stick), what is the probability that the 3 segments can form a triangle?<sup>22</sup>

*Solution:* Without loss of generality, let's assume that the length of the stick is 1. Let's also label the point of the first cut as  $x$  and the second cut as  $y$ .

If  $x < y$ , then the three segments are  $x$ ,  $y-x$  and  $1-y$ . The conditions to form a triangle are



$$x + (y - x) > 1 - y \Rightarrow y > 1/2$$

$$x + (1 - y) > y - x \Rightarrow y < 1/2 + x$$

$$(y - x) + (1 - y) > x \Rightarrow x < 1/2$$

The feasible area is shown in Figure 4.5. The case for  $x < y$  is the left gray triangle. Using symmetry, we can see that the case for  $x > y$  is the right gray triangle.

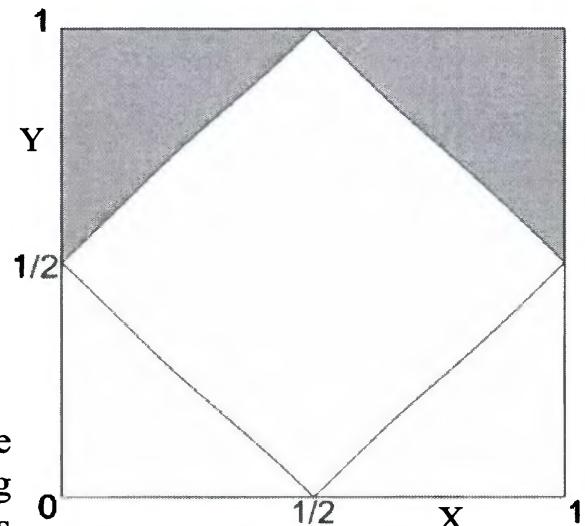


Figure 4.5 Distribution of cuts X and Y

<sup>22</sup> Hint: Let the first cut point be  $x$ , the second one be  $y$ , use the figure to show the distribution of  $x$  and  $y$ .

The total shadowed area represents the region where 3 segments can form a triangle, which is 1/4 of the square. So the probability is 1/4.

## Property of Poisson process

You are waiting for a bus at a bus station. The buses arrive at the station according to a Poisson process with an average arrival time of 10 minutes ( $\lambda = 0.1/\text{min}$ ). If the buses have been running for a long time and you arrive at the bus station at a random time, what is your expected waiting time? On average, how many minutes ago did the last bus leave?

*Solution:* Considering the importance of jump-diffusion processes in derivative pricing and the role of Poisson processes in studying jump processes, let's elaborate more on exponential random variables and the Poisson process. Exponential distribution is widely used to model the time interval between independent events that happen at a constant

average rate (arrival rate)  $\lambda$ :  $f(t) = \begin{cases} \lambda e^{-\lambda t} & (t \geq 0) \\ 0 & (t < 0) \end{cases}$ . The expected arrival time is  $1/\lambda$

and the variance is  $1/\lambda^2$ . Using integration, we can calculate the cdf of an exponential distribution to be  $F(t) = P(\tau \leq t) = 1 - e^{-\lambda t}$  and  $P(\tau > t) = e^{-\lambda t}$ , where  $\tau$  is the random variable for arrival time. One unique property of exponential distribution is memorylessness:  $P\{\tau > s+t | \tau > s\} = P(\tau > t)$ .<sup>23</sup> That means if we have waited for  $s$  time units, the extra waiting time has the same distribution as the waiting time when we start at time 0.

When the arrivals of a series of events each independently follow an exponential distribution with arrival rate  $\lambda$ , the number of arrivals between time 0 and  $t$  can be

modeled as a Poisson process  $P(N(t) = x) = \frac{e^{-\lambda t} \lambda^x}{x!}$ ,  $x = 0, 1, \dots$ <sup>24</sup> The expected

number of arrivals is  $\lambda t$  and the variance is also  $\lambda t$ . Because of the memoryless nature of exponential distribution, the number of arrivals between time  $s$  and  $t$  is also a Poisson process  $P(N(t-s) = x) = \frac{e^{-\lambda(t-s)} (\lambda(t-s))^x}{x!}$ .

Taking advantage of the memoryless property of exponential distribution, we know that the expected waiting time is  $1/\lambda = 10\text{ min}$ . If you look back in time, the memoryless property still applies. So on average, the last bus arrived 10 minutes ago as well.

---

<sup>23</sup>  $P\{\tau > s+t | \tau > s\} = e^{-\lambda(s+t)} / e^{-\lambda s} = e^{-\lambda t} = P(\tau > t)$

<sup>24</sup> More rigorously,  $N(t)$  is defined as a right-continuous function.

This is another example that your intuition may misguide you. You may be wondering that if the last bus on average arrived 10 minutes ago and the next bus on average will arrive 10 minutes later, shouldn't the average arrival time be 20 minutes instead of 10? The explanation to the apparent discrepancy is that when you arrive at a random time, you are more likely to arrive in a long time interval between two bus arrivals than in a short one. For example, if one interval between two bus arrivals is 30 minutes and another is 5 minutes, you are more likely to arrive at a time during that 30-minute interval rather than 5-minute interval. In fact, if you arrive at a random time, the expected residual life (the time for the next bus to arrive) is  $\frac{E[X^2]}{2E[X]}$  for a general distribution.<sup>25</sup>

## Moments of normal distribution

If  $X$  follows standard normal distribution ( $X \sim N(0, 1)$ ), what is  $E[X^n]$  for  $n = 1, 2, 3$  and 4?

*Solution:* The first to fourth moments of the standard normal distribution are essentially the mean, the variance, the skewness and the kurtosis. So you probably have remembered that the answers are 0, 1, 0 (no skewness), and 3, respectively.

Standard normal distribution has pdf  $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ . Using simple symmetry we

have  $E[x^n] = \int_{-\infty}^{\infty} x^n \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0$  when  $n$  is odd. For  $n = 2$ , integration by parts are

often used. To solve  $E[X^n]$  for any integer  $n$ , an approach using **moment generating functions** may be a better choice. Moment generating functions are defined as

$$M(t) = E[e^{tX}] = \begin{cases} \sum_x e^{tx} p(x), & \text{if } x \text{ is discrete} \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx, & \text{if } x \text{ is continuous} \end{cases}$$

Sequentially taking derivative of  $M(t)$ , we get one frequently-used property of  $M(t)$ :

$$M'(t) = \frac{d}{dt} E[e^{tX}] = E[Xe^{tX}] \Rightarrow M'(0) = E[X],$$

$$M''(t) = \frac{d}{dt} E[Xe^{tX}] = E[X^2 e^{tX}] \Rightarrow M''(0) = E[X^2],$$

---

<sup>25</sup> The residual life is explained in Chapter 3 of “**Discrete Stochastic Process**” by Robert G. Gallager.

and  $M^n(0) = E[X^n]$ ,  $\forall n \geq 1$  in general.

We can use this property to solve  $E[X^n]$  for  $X \sim N(0, 1)$ . For standard normal distribution  $M(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x-t)^2/2} dx = e^{t^2/2}$ .

$(\frac{1}{\sqrt{2\pi}} e^{-(x-t)^2/2})$  is the pdf of normal distribution  $X \sim N(t, 1)$ , so  $\int_{-\infty}^{\infty} f(x)dx = 1$ .

Taking derivatives, we have

$$M'(t) = te^{t^2/2} \Rightarrow M'(0) = 0, M''(t) = e^{t^2/2} + t^2 e^{t^2/2} \Rightarrow M''(0) = e^0 = 1,$$

$$M^3(t) = te^{t^2/2} + 2te^{t^2/2} + t^3 e^{t^2/2} = 3te^{t^2/2} + t^3 e^{t^2/2} \Rightarrow M^3(0) = 0,$$

$$\text{and } M^4(t) = 3e^{t^2/2} + 3t^2 e^{t^2/2} + 3t^2 e^{t^2/2} + 3t^4 e^{t^2/2} \Rightarrow M^4(0) = 3e^0 = 3.$$

## 4.5 Expected Value, Variance & Covariance

Expected value, variance and covariance are indispensable in estimating returns and risks of any investments. Naturally, they are a popular test subject in interviews as well. The basic knowledge includes the following:

If  $E[x_i]$  is finite for all  $i = 1, \dots, n$ , then  $E[X_1 + \dots + X_n] = E[X_1] + \dots + E[X_n]$ . The relationship holds whether the  $x_i$ 's are independent of each other or not.

If  $X$  and  $Y$  are independent, then  $E[g(X)h(Y)] = E[g(x)]E[h(Y)]$ .

**Covariance:**  $Cov(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$ .

**Correlation:**  $\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$

If  $X$  and  $Y$  are independent,  $Cov(X, Y) = 0$  and  $\rho(X, Y) = 0$ .<sup>26</sup>

**General rules of variance and covariance:**

$$Cov(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j Cov(X_i, Y_j)$$

$$Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i) + 2 \sum_{i < j} Cov(X_i, X_j)$$

---

<sup>26</sup> The reverse is not true.  $\rho(X, Y) = 0$  only means  $X$  and  $Y$  are uncorrelated; they may well be dependent.

## Conditional expectation and variance

For discrete distribution:  $E[g(X) | Y = y] = \sum_x g(x)p_{X|Y}(x | y) = \sum_x g(x)p(X = x | Y = y)$

For continuous distribution:  $E[g(X) | Y = y] = \int_{-\infty}^{\infty} g(x)f_{X|Y}(x | y)dx$

### Law of total expectation:

$$E[X] = E[E[X | Y]] = \begin{cases} \sum_y E[X | Y = y]p(Y = y), & \text{for discrete } Y \\ \int_{-\infty}^{\infty} E[X | Y = y]f_Y(y)dy, & \text{for continuous } Y \end{cases}$$

## Connecting noodles

You have 100 noodles in your soup bowl. Being blindfolded, you are told to take two ends of some noodles (each end on any noodle has the same probability of being chosen) in your bowl and connect them. You continue until there are no free ends. The number of loops formed by the noodles this way is stochastic. Calculate the expected number of circles.

*Solution:* Again do not be frightened by the large number 100. If you have no clue how to start, let's begin with the simplest case where  $n = 1$ . Surely you have only one choice (to connect both ends of the noodle), so  $E[f(1)] = 1$ . How about 2 noodles? Now you have 4 ends ( $2 \times 2$ ) and you can connect any two of them. There are  $\binom{4}{2} = \frac{4 \times 3}{2} = 6$

combinations. Among them, 2 combinations will connect both ends of the same noodle together and yield 1 circle and 1 noodle. The other 4 choices will yield a single noodle. So the expected number of circles is

$$E[f(2)] = 2/6 \times (1 + E[f(1)]) + 4/6 \times E[f(1)] = 1/3 + E[f(1)] = 1/3 + 1.$$

We now move on to 3 noodles with  $\binom{6}{2} = \frac{6 \times 5}{2} = 15$  choices. Among them, 3 choices will yield 1 circle and 2 noodles; the other 12 choices will yield 2 noodles only, so

$$E[f(3)] = 3/15 \times (1 + E[f(2)]) + 12/15 \times E[f(2)] = 1/5 + E[f(2)] = 1/5 + 1/3 + 1.$$

See the pattern? For any  $n$  noodles, we will have  $E[f(n)] = 1 + 1/3 + 1/5 + \dots + 1/(2n-1)$ , which can be easily proved by induction. Plug 100 in, we will have the answer.

Actually after the 2-noodle case, you probably have found the key to this question. If you start with  $n$  noodles, among  $\binom{2n}{2} = n(2n - 1)$  possible combinations, we have

$\frac{n}{n(2n - 1)} = \frac{1}{2n - 1}$  probability to yield 1 circle and  $n - 1$  noodles and  $\frac{2n - 2}{2n - 1}$  probability to yield  $n - 1$  noodles only, so  $E[f(n)] = E[f(n - 1)] + \frac{1}{2n - 1}$ . Working backward, you can get the final solution as well.

## Optimal hedge ratio

You just bought one share of stock  $A$  and want to hedge it by shorting stock  $B$ . How many shares of  $B$  should you short to minimize the variance of the hedged position? Assume that the variance of stock  $A$ 's return is  $\sigma_A^2$ ; the variance of  $B$ 's return is  $\sigma_B^2$ ; their correlation coefficient is  $\rho$ .

*Solution:* Suppose that we short  $h$  shares of  $B$ , the variance of the portfolio return is  $\text{var}(r_A - hr_B) = \sigma_A^2 - 2\rho h\sigma_A\sigma_B + h^2\sigma_B^2$

The best hedge ratio should minimize  $\text{var}(r_A - hr_B)$ . Take the first order partial derivative with respect to  $h$  and set it to zero:  $\frac{\partial \text{var}}{\partial h} = -2\rho\sigma_A\sigma_B + 2h\sigma_B^2 = 0 \Rightarrow h = \rho \frac{\sigma_A}{\sigma_B}$ .

To confirm it's the minimum, we can also check the second-order partial derivative:

$\frac{\partial^2 \text{var}}{\partial h^2} = 2\sigma_B^2 > 0$ . So Indeed when  $h = \rho \frac{\sigma_A}{\sigma_B}$ , the hedge portfolio has the minimum variance.

## Dice game

Suppose that you roll a dice. For each roll, you are paid the face value. If a roll gives 4, 5 or 6, you can roll the dice again. Once you get 1, 2 or 3, the game stops. What is the expected payoff of this game?

*Solution:* This is an example of the law of total expectation. Clearly your payoff will be different depending on the outcome of first roll. Let  $E[X]$  be your expected payoff and  $Y$  be the outcome of your first throw. You have  $1/2$  chance to get  $Y \in \{1, 2, 3\}$ , in which case the expected value is the expected face value 2, so  $E[X | Y \in \{1, 2, 3\}] = 2$ ; you have

1/2 chance to get  $Y \in \{4, 5, 6\}$ , in which case you get expected face value 5 and extra throw(s). The extra throw(s) essentially means you start the game again and have an extra expected value  $E[X]$ . So we have  $E[X | Y \in \{4, 5, 6\}] = 5 + E[X]$ . Apply the law of total expectation, we have  $E[X] = E[E[X | Y]] = \frac{1}{2} \times 2 + \frac{1}{2} \times (5 + E[X]) \Rightarrow E[X] = 7$ .<sup>27</sup>

## Card game

What is the expected number of cards that need to be turned over in a regular 52-card deck in order to see the first ace?

*Solution:* There are 4 aces and 48 other cards. Let's label them as card 1, 2, ..., 48. Let

$$X_i = \begin{cases} 1, & \text{if card } i \text{ is turned over before 4 aces} \\ 0, & \text{otherwise} \end{cases}$$

The total number of cards that need to be turned over in order to see the first ace is  $X = 1 + \sum_{i=1}^{48} X_i$ , so we have  $E[X] = 1 + \sum_{i=1}^{48} E[X_i]$ . As shown in the following sequence, each card  $i$  is equally likely to be in one of the five regions separated by 4 aces:

1 A 2 A 3 A 4 A 5

So the probability that card  $i$  appears before all 4 aces is  $1/5$ , and we have  $E[X_i] = 1/5$ .

$$\text{Therefore, } E[X] = 1 + \sum_{i=1}^{48} E[X_i] = 1 + 48/5 = 10.6.$$

This is just a special case for random ordering of  $m$  ordinary cards and  $n$  special cards.

$$\text{The expected position of the first special card is } 1 + \sum_{i=1}^m E[X_i] = 1 + \frac{m}{n+1}.$$

## Sum of random variables

Assume that  $X_1, X_2, \dots, X_n$  are independent and identically-distributed (IID) random variables with uniform distribution between 0 and 1. What is the probability that  $S_n = X_1 + X_2 + \dots + X_n \leq 1$ ?<sup>28</sup>

---

<sup>27</sup> You will also see that the problem can be solved using Wald's equality in Chapter 5.

<sup>28</sup> Hint: start with the simplest case where  $n = 1, 2$ , and  $3$ . Try to find a general formula and prove it using induction.

*Solution:* This problem is a rather difficult one. The general principle to start with the simplest cases and try to find a pattern will again help you approach the problem; even though it may not give you the final answer. When  $n=1$ ,  $P(S_1 \leq 1)$  is 1. As shown in Figure 4.6, when  $n=2$ , the probability that  $X_1 + X_2 \leq 1$  is just the area under  $X_1 + X_2 \leq 1$  within the square with side length 1 (a triangle). So  $P(S_2 \leq 1) = 1/2$ . When  $n=3$ , the probability becomes the tetrahedron ABCD under the plane  $X_1 + X_2 + X_3 \leq 1$  within the cube with side length 1. The volume of tetrahedron ABCD is  $1/6$ .<sup>29</sup> So  $P(S_3 \leq 1) = 1/6$ . Now we can guess that the solution is  $1/n!$ . To prove it, let's again resort to induction. Assume  $P(S_n \leq 1) = 1/n!$ . We need to prove that  $P(S_{n+1} \leq 1) = 1/(n+1)!$ .

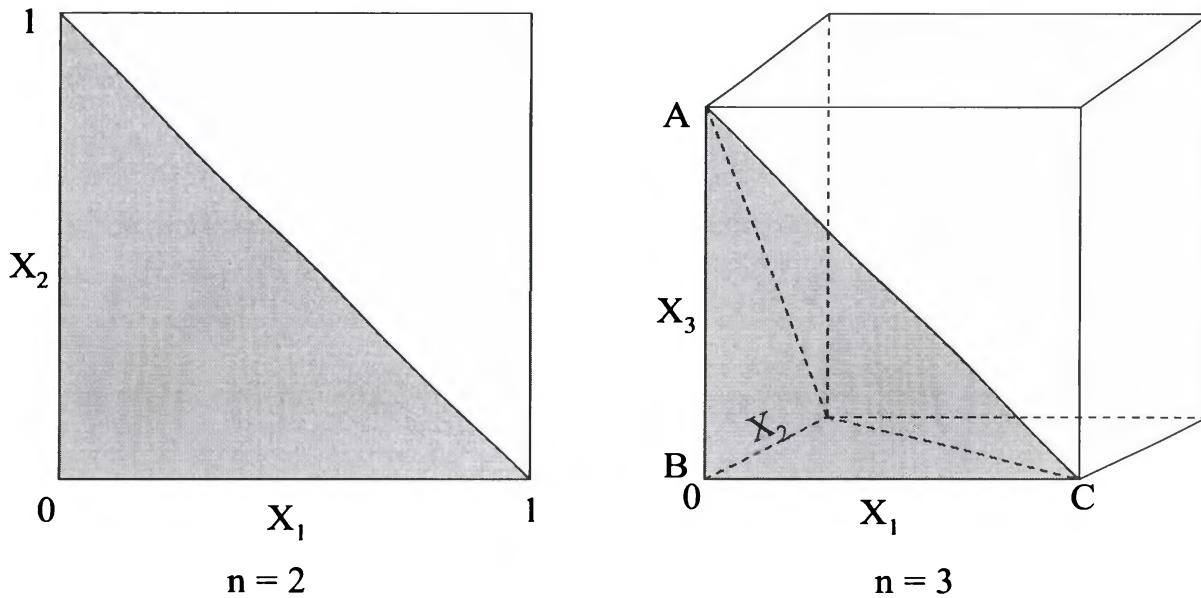


Figure 4.6 Probability that  $S_n \leq 1$  when  $n = 2$  or  $n = 3$ .

Here we can use probability by conditioning. Condition on the value of  $X_{n+1}$ , we have  $P(S_{n+1} \leq 1) = \int_0^1 f(X_{n+1})P(S_n \leq 1 - X_{n+1})dX_{n+1}$ , where  $f(X_{n+1})$  is the probability density function of  $X_{n+1}$ , so  $f(X_{n+1}) = 1$ . But how do we calculate  $P(S_n \leq 1 - X_{n+1})$ ? The cases of  $n=2$  and  $n=3$  have provided us with some clue. For  $S_n \leq 1 - X_{n+1}$  instead of  $S_n \leq 1$ , we essentially need to shrink every dimension of the  $n$ -dimensional simplex<sup>30</sup> from 1 to

<sup>29</sup> You can derive it by integration:  $\int_0^1 A(z)dz = \int_0^1 1/2z^2 dz = 1/6$ , where  $A(z)$  is the cross-sectional area.

<sup>30</sup> An  $n$ -Simplex is the  $n$ -dimensional analog of a triangle.

$1 - X_{n+1}$ . So its volume should be  $\frac{(1 - X_{n+1})^n}{n!}$  instead of  $\frac{1}{n!}$ . Plugging in these results,

$$\text{we have } P(S_{n+1} \leq 1) = \int_0^1 \frac{(1 - X_{n+1})^n}{n!} dX_{n+1} = \frac{1}{n!} \left[ -\frac{(1 - X_{n+1})^{n+1}}{n+1} \right]_0^1 = \frac{1}{n!} \times \frac{1}{n+1} = \frac{1}{(n+1)!}.$$

So the general result is true for  $n+1$  as well and we have  $P(S_n \leq 1) = 1/n!$ .

## Coupon collection

There are  $N$  distinct types of coupons in cereal boxes and each type, independent of prior selections, is equally likely to be in a box.

A. If a child wants to collect a complete set of coupons with at least one of each type, how many coupons (boxes) on average are needed to make such a complete set?

B. If the child has collected  $n$  coupons, what is the expected number of distinct coupon types?<sup>31</sup>

*Solution:* For part A, let  $X_i$ ,  $i = 1, 2, \dots, N$ , be the number of additional coupons needed to obtain the  $i$ -th type after  $(i-1)$  distinct types have been collected. So the total number of coupons needed is  $X = X_1 + X_2 + \dots + X_N = \sum_{i=1}^N X_i$ .

For any  $i$ ,  $i-1$  distinct types of coupons have already been collected. It follows that a new coupon will be of a different type with probability  $1 - (i-1)/N = (N-i+1)/N$ . Essentially to obtain the  $i$ -th distinct type, the random variable  $X_i$  follows a geometric distribution with  $p = (N-i+1)/N$  and  $E[X_i] = N/(N-i+1)$ . For example, if  $i=1$ , we simply have  $X_i = E[X_i] = 1$ .

$$\therefore E[X] = \sum_{i=1}^N E[X_i] = \sum_{i=1}^N \frac{N}{N-i+1} = N \left( \frac{1}{N} + \frac{1}{N-1} + \dots + \frac{1}{1} \right).$$

---

<sup>31</sup> Hint: For part A, let  $X_i$  be the number of extra coupons collected to get the  $i$ -th distinct coupon after  $i-1$  types of distinct coupons have been collected. Then the total expected number of coupons to collect all distinct types is  $E[X] = \sum_{i=1}^N E[X_i]$ . For part B, which is the expected probability (P) that the  $i$ -th coupon type is not in the  $n$  coupons?

For part *B*, let  $Y$  be the number of distinct types of coupons in the set of  $n$  coupons. We introduce indicator **random variables**  $I_i, i = 1, 2, \dots, N$ , where

$$\begin{cases} I_i = 1, & \text{if at least one coupon of the } i\text{-th type is in the set of } n \text{ coupons} \\ I_i = 0, & \text{otherwise} \end{cases}$$

So we have  $Y = I_1 + I_2 + \dots + I_N = \sum_{i=1}^N I_i$

For each collected coupon, the probability that it is not the  $i$ -th coupon type is  $\frac{N-1}{N}$ .

Since all  $n$  coupons are independent, the probability that none of the  $n$  coupons is the  $i$ -th coupon type is  $P(I_i = 0) = \left(\frac{N-1}{N}\right)^n$  and we have  $E[I_i] = P(I_i = 1) = 1 - \left(\frac{N-1}{N}\right)^n$ .

$$\therefore E[Y] = \sum_{i=1}^N E[I_i] = N - N \left(\frac{N-1}{N}\right)^n. ^{32}$$

## Joint default probability

If there is a 50% probability that bond *A* will default next year and a 30% probability that bond *B* will default. What is the range of probability that at least one bond defaults and what is the range of their correlation?

*Solution:* The range of probability that at least one bond defaults is easy to find. To have the largest probability, we can assume whenever *A* defaults, *B* does not default; whenever *B* defaults, *A* does not default. So the maximum probability that at least one bond defaults is  $50\% + 30\% = 80\%$ . (The result only applies if  $P(A) + P(B) \leq 1$ ). For the minimum, we can assume whenever *A* defaults, *B* also defaults. So the minimum probability that at least one bond defaults is 50%.

To calculate the corresponding correlation, let  $I_A$  and  $I_B$  be the indicator for the event that bond A/B defaults next year and  $\rho_{AB}$  be their correlation. Then we have  $E[I_A] = 0.5$ ,  $E[I_B] = 0.3$ ,  $\text{var}(I_A) = p_A \times (1 - p_A) = 0.25$ ,  $\text{var}(I_B) = 0.21$ .

---

<sup>32</sup> A similar question: if you randomly put 18 balls into 10 boxes, what is the expected number of empty boxes?

$$\begin{aligned}
P(A \text{ or } B \text{ defaults}) &= E[I_A] + E[I_B] - E[I_A I_B] \\
&= E[I_A] + E[I_B] - (E[I_A]E[I_B] - \text{cov}(I_A, I_B)) \\
&= 0.5 + 0.3 - (0.5 \times 0.3 - \rho_{AB}\sigma_A\sigma_B) \\
&= 0.65 - \sqrt{0.21}/2\rho_{AB}
\end{aligned}$$

For the maximum probability, we have  $0.65 - \sqrt{0.21}/2\rho_{AB} = 0.8 \Rightarrow \rho_{AB} = -\sqrt{3/7}$ .

For the minimum probability, we have  $0.65 - \sqrt{0.21}/2\rho_{AB} = 0.5 \Rightarrow \rho_{AB} = \sqrt{3/7}$ .

In this problem, do not start with  $P(A \text{ or } B \text{ defaults}) = 0.65 - \sqrt{0.21}/2\rho_{AB}$  and try to set  $\rho_{AB} = \pm 1$  to calculate the maximum and minimum probability since the correlation cannot be  $\pm 1$ . The range of correlation is restricted to  $[-\sqrt{3/7}, \sqrt{3/7}]$ .

## 4.6 Order Statistics

Let  $X$  be a random variable with cumulative distribution function  $F_X(x)$ . We can derive the distribution function for the minimum  $Y_n = \min(X_1, X_2, \dots, X_n)$  and for the maximum  $Z_n = \max(X_1, X_2, \dots, X_n)$  of  $n$  IID random variables with cdf  $F_X(x)$  as

$$P(Y_n \geq x) = (P(X \geq x))^n \Rightarrow 1 - F_{Y_n}(x) = (1 - F_X(x))^n \Rightarrow f_{Y_n}(x) = nF_X(x)(1 - F_X(x))^{n-1}$$

$$P(Z_n \leq x) = (P(X \leq x))^n \Rightarrow F_{Z_n}(x) = (F_X(x))^n \Rightarrow f_{Z_n}(x) = nf_X(x)(F_X(x))^{n-1}$$

### Expected value of max and min

Let  $X_1, X_2, \dots, X_n$  be IID random variables with uniform distribution between 0 and 1. What are the cumulative distribution function, the probability density function and expected value of  $Z_n = \max(X_1, X_2, \dots, X_n)$ ? What are the cumulative distribution function, the probability density function and expected value of  $Y_n = \min(X_1, X_2, \dots, X_n)$ ?

*Solution:* This is a direct test of textbook knowledge. For uniform distribution on  $[0,1]$ ,  $F_X(x) = x$  and  $f_X(x) = 1$ . Applying  $F_X(x)$  and  $f_X(x)$  to  $Z_n = \max(X_1, X_2, \dots, X_n)$  we have

$$\begin{aligned}
P(Z_n \leq x) &= (P(X \leq x))^n \Rightarrow F_{Z_n}(x) = (F_X(x))^n = x^n \\
\Rightarrow f_{Z_n}(x) &= nf_X(x)(F_X(x))^{n-1} = nx^{n-1}
\end{aligned}$$

$$\text{and } E[Z_n] = \int_0^1 xf_{Z_n}(x)dx = \int_0^1 nx^n dx = \frac{n}{n+1} \left[ x^{n+1} \right]_0^1 = \frac{n}{n+1}.$$

Applying  $F_X(x)$  and  $f_X(x)$  to  $Y_n = \min(X_1, X_2, \dots, X_n)$  we have

$$\begin{aligned} P(Y_n \geq x) &= (P(X \geq x))^n \Rightarrow F_{Y_n}(x) = 1 - (1 - F_X(x))^n = 1 - (1 - x)^n \\ &\Rightarrow f_{Y_n}(x) = n f_X(x) (1 - F_X(x))^{n-1} = n(1-x)^{n-1} \end{aligned}$$

$$\text{and } E[Y_n] = \int_0^1 nx(1-x)^{n-1} dx = \int_0^1 n(1-y)y^{n-1} dy = \left[ y^n \right]_0^1 - \frac{n}{n+1} \left[ y^{n+1} \right]_0^1 = \frac{1}{n+1}.$$

## Correlation of max and min

Let  $X_1$  and  $X_2$  be IID random variables with uniform distribution between 0 and 1,  $Y = \min(X_1, X_2)$  and  $Z = \max(X_1, X_2)$ . What is the probability of  $Y \geq y$  given that  $Z \leq z$  for any  $y, z \in [0, 1]$ ? What is the correlation of  $Y$  and  $Z$ ?

*Solution:* This problem is another demonstration that a figure is worth a thousand words. As shown in Figure 4.7, the probability that  $Z \leq z$  is simply the square with side length  $z$ . So  $P(Z \leq z) = z^2$ . Since  $Z = \max(X_1, X_2)$  and  $Y = \min(X_1, X_2)$ , we must have  $Y \leq Z$  for any pair of  $X_1$  and  $X_2$ . So if  $y > z$ ,  $P(Y \geq y | Z \leq z) = 0$ . For  $y \leq z$ , that  $X_1$  and  $X_2$  satisfies  $Y \geq y$  and  $Z \leq z$  is the square with vertices  $(y, y), (z, y), (z, z)$ , and  $(y, z)$ , which has an area  $(z - y)^2$ . So  $P(Y \geq y \cap Z \leq z) = (z - y)^2$ . Hence

$$P(Y \geq y | Z \leq z) = \begin{cases} (z - y)^2 / z^2, & \text{if } 0 \leq z \leq 1 \text{ and } 0 \leq y \leq z \\ 0, & \text{otherwise} \end{cases}.$$

Now let's move on to calculate the correlation of  $Y$  and  $Z$ .

$$\text{corr}(Y, Z) = \frac{\text{cov}(Y, Z)}{\text{std}(Y) \times \text{std}(Z)} = \frac{E[YZ] - E[Y]E[Z]}{\sqrt{E[Y^2] - E[Y]^2} \times \sqrt{E[Z^2] - E[Z]^2}}$$

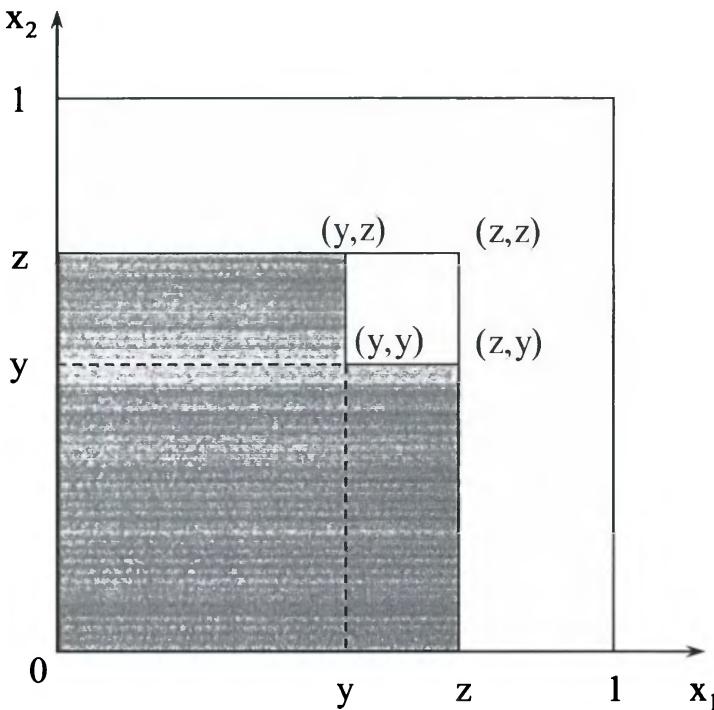


Figure 4.7 Distribution of  $X_1, X_2$ , their maximum and minimum.

Using previous problem's conclusions, we have  $E[Y] = \frac{1}{2+1} = \frac{1}{3}$ ,  $E[Z] = \frac{2}{2+1} = \frac{2}{3}$ .

From the pdfs of  $Y$  and  $Z$ ,  $f_{Y_n}(x) = n(1-x)^{n-1} = 2(1-x)$  and  $f_Z(z) = nz^{n-1} = 2z$ , we can also get  $E[Y_n^2] = \int_0^1 2(1-y)y^2 dy = \frac{2}{3} - \frac{2}{4} = \frac{1}{6}$  and  $E[Z^2] = \int_0^1 2z^3 dz = \frac{2}{4}$ , which give us the variances:  $\text{var}(Y) = E[Y^2] - E[Y]^2 = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}$  and  $\text{var}(Z) = \frac{2}{4} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$ .<sup>33</sup>

To calculate  $E[YZ]$ , we can use  $E[YZ] = \int_0^1 \int_0^z yzf(y,z) dy dz$ . To solve this equation, we need  $f(y,z)$ . Let's again go back to Figure 4.7. From the figure we can see that when  $0 \leq z \leq 1$  and  $0 \leq y \leq z$ ,  $F(y,z)$  is the shadowed area with probability

$$F(y,z) = P(Y \leq y \cap Z \leq z) = P(Z \leq z) - P(Y \geq y \cap Z \leq z) = z^2 - (z-y)^2 = 2zy - y^2$$

$$\therefore f(y,z) = \frac{\partial}{\partial y} F(y,z) = 2 \text{ and } E[YZ] = \int_0^1 \int_0^z 2yz dy dz = \int_0^1 z[y^2]_0^z dz = \int_0^1 z^3 dz = \frac{1}{4}.$$

---

<sup>33</sup> You may have noticed that  $\text{var}(Y) = \text{var}(Z)$  and wonder whether it is a coincidence for  $n = 2$ . It is actually true for all integer  $n$ . You may want to think about why that is true without resorting to calculation. Hint:  $\text{var}(x) = \text{var}(1-x)$  for any random variable  $x$ .

An alternative and simpler approach to calculate  $E[YZ]$  is again to take advantage of symmetry. Notice that no matter  $x_1 \leq x_2$  or  $x_1 > x_2$ , we always have  $yz = x_1 x_2$  ( $z = \max(x_1, x_2)$  and  $y = \min(x_1, x_2)$ ).

$$\therefore E[YZ] = \int_0^1 \int_0^1 x_1 x_2 dx_1 dx_2 = E[X_1]E[X_2] = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}.$$

$$\text{Hence } \text{cov}(Y, Z) = E[YZ] - E[Y]E[Z] = \frac{1}{36} \text{ and } \text{corr}(Y, Z) = \frac{\text{cov}(Y, Z)}{\sqrt{\text{var}(Y)} \times \sqrt{\text{var}(Z)}} = \frac{1}{2}.$$

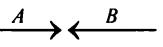
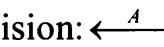
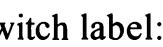
Sanity check: That  $Y$  and  $Z$  have positive autocorrelation make sense since when  $Y$  becomes large,  $Z$  tends to become large as well ( $Z \geq Y$ ).

## Random ants

500 ants are randomly put on a 1-foot string (independent uniform distribution for each ant between 0 and 1). Each ant randomly moves toward one end of the string (equal probability to the left or right) at constant speed of 1 foot/minute until it falls off at one end of the string. Also assume that the size of the ant is infinitely small. When two ants collide head-on, they both immediately change directions and keep on moving at 1 foot/min. What is the expected time for all ants to fall off the string?<sup>34</sup>

*Solution:* This problem is often perceived to be a difficult one. The following components contribute to the complexity of the problem: The ants are randomly located; each ant can go either direction; an ant needs to change direction when it meets another ant. To solve the problem, let's tackle these components.

When two ants collide head-on, both immediately change directions. What does it mean? The following diagram illustrates the key point:

Before collision:  ; After collision:  ; switch label: 

When an ant  $A$  collides with another ant  $B$ , both switch direction. But if we exchange the ants' labels, it's like that the collision never happens.  $A$  continues to move to the right and  $B$  moves to the left. Since the labels are randomly assigned anyway, collisions make no difference to the result. So we can assume that when two ants meet, each just keeps on going in its original direction. What about the random direction that each ant chooses? Once the collision is removed, we can use symmetry to argue that it makes no difference which direction that an ant goes either. That means if an ant is put at the  $x$ -th foot, the

---

<sup>34</sup> Hint: If we switch the label of two ants that collide with each other, it's like that the collision never happened.

expected value for it to fall off is just  $x$  min. If it goes in the other direction, simply set  $x$  to  $1 - x$ . So the original problem is equivalent to the following:

What is the expected value of the maximum of 500 IID random variables with uniform distribution between 0 and 1?

Clearly the answer is  $\frac{499}{500}$  min, which is the expected time for all ants to fall off the string.