Solutions to "A Practical Guide to Quantitative Finance"

February 14, 2025

Huawei Wu

---

1 Probability

1.1 Combinatorial Analysis

1. (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?

> **Ans:** Let's consider a more general version of the problem. Suppose we have n envelopes and n corresponding cover letters, and we number them from left to right as $1, 2, \ldots, n$ (one, two, up to n). A way of placing these cover letters can be seen as a permutation of these cover letters. The case where no letter is placed in its correct envelope corresponds to a **derangement**—a permutation with no fixed points.
>
> There are $n!$ (n factorial) possible permutations in total. And we use D_n to denote the number of derangements of the n cover letters. Let's start with the simplest cases:
>
> - When $n = 1$, there is only one cover letter, and it is always placed in the correct envelope. So, we have $D_1 = 0$.
>
> - When $n = 2$, the only valid derangement is swapping the two cover letters. So, we have $D_2 = 1$.
>
> Now, let's consider the general case for $n \geq 3$. Since cover letter 1 cannot be placed in envelope 1, we have $n - 1$ choices for where to place it. Suppose

we put it in envelope k (where $2 \le k \le n$). Let $D_{n,k}$ represent the number of derangements where cover letter 1 is specifically placed in envelope k. Due to symmetry, we know that:

$$D_{n,2} = D_{n,3} = \cdots = D_{n,n}$$

This means that the total number of derangements of n cover letters can be expressed as:

$$D_n = (n-1) \cdot D_{n,2}$$

So, it is enough to analyze $D_{n,2}$. There are two cases to consider:

1. **Case 1:** Cover letter 2 is placed in envelope 1. In this scenario, the remaining elements are cover letters $3, 4, \ldots, n$ and envelopes $3, 4, \ldots, n$. Hence what we need to do next is to permute the remaining $n - 2$ cover letters with no fixed points. Consequently, the number of such derangements is D_{n-2}.

2. **Case 2:** Cover letter 2 is not placed in envelope 1. Since the original position of cover letter 2 (that is, envelope 2) has already been taken by cover letter 1, we can think of envelope 1 as its new "correct" position. Not placing it there means it is also misplaced, so this situation is equivalent to a derangement of the remaining $n - 1$ letters, which gives us D_{n-1}.

Thus, we derive the recurrence relation:

$$D_n = (n-1) \cdot (D_{n-1} + D_{n-2})$$

This formula allows us to compute the number of derangements recursively. We have
$$D_3 = 2 \cdot (D_2 + D_1) = 2 \cdot (1 + 0) = 2,$$
$$D_4 = 3 \cdot (D_3 + D_2) = 3 \cdot (2 + 1) = 9,$$
and
$$D_5 = 4 \cdot (D_4 + D_3) = 4 \cdot (9 + 2) = 44.$$
Hence the probability that all 5 cover letters are mailed to the wrong firms is

$$\frac{D_5}{5!} = \frac{44}{120} = \frac{11}{30}.$$

Remark 1 *neat /niːt/, adj.* 整齐的，简洁的.

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

> **Ans:** Assume that there are n people in class. Without any restrictions, we have 365 possibilities for each one's birthday. The basic principle of counting tells us that there 365^n possible ways to assign birthdays to n people.
>
> We want to find the number of ways in which no people share the same birthday. The first person can choose any of the 365 days as his/her birthday, while the second person must have a different birthday, i.e., he/she has 364 choices. Generally, for the r-th person, there are $365 - r + 1$ choices. Therefore, the number of ways to assign birthdays to n people such that no people share the same birthday is $365 \times 364 \times \cdots \times (365 - n + 1)$, and the probability that no people share the same birthday is:
>
> $$\frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n}.$$
>
> We want this probability to be less than $\frac{1}{2}$. The smallest integer n that satisfies this inequality is 23.

3. (100-th Digit) What is the 100-th digit to the right of the decimal point in the decimal representation of $(1 + \sqrt{2})^{3000}$?

Ans: Applying the binomial theorem, we have

$$(1 + \sqrt{2})^n = \sum_{k=0}^{n} \binom{n}{k} \sqrt{2}^k.$$

We can divide the above summation into two parts based on whether k is even or odd:

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

Similarly, we have

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

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

It follows that

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

which always a positive integer. Note that $1 - \sqrt{2}$ is approximately 0.4, which implies that

$$0 < (1 - \sqrt{2})^{3000} \approx 0.4^{3000} = (0.064)^{1000} < 0.1^{1000} = 10^{-10000} < 10^{-100}.$$

It follows that the 100-th digit to the right of the decimal point in the decimal representation of $(1 + \sqrt{2})^{3000}$ is 9.

4. (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.

Ans: We can express x as $x = 10b + a$, where a is the last digit of x. Applying the binomial theorem, we have

$$x^3 = (10b + a)^3 = 1000b^3 + 300a^2b^2 + 30a^2b + a^3,$$

which implies that

$$x^3 \equiv 30a^2b + a^3 \pmod{100}$$

and

$$x^3 \equiv a^3 \pmod{10}.$$

Hence the units digit of x^3 equals that of a^3. Since $0 \le a \le 9$, a brute-force search shows that the unit digit of a^3 is 1 if and only if $a = 1$.

If $a = 1$, then

$$x^3 \equiv 30b + 1 \pmod{100},$$

which implies that

$$\frac{x^3 - 1}{10} \equiv 3b \pmod{10}.$$

Note that if the units digit of x^3 is 1, then the tens digit of x^3 is equal to the units digit of $\frac{x^3-1}{10}$. Hence the cubic of x ends with 11 is equivalent to saying that $a = 1$ and the units digit of $3b$ is 1. A brute-force search shows that the latter condition is equivalent to saying that the units digit of b is 7. Since $x = 10b + 1$, we conclude that the cubic of x ends with 11 if and only if the last two digits of x should be 71, which occurs with probability $\frac{1}{100}$.

Remark 2 $x^3 \equiv a^3 \pmod{10}$ *can be read as "x cubed in congruent to a cubed modulo 10".*

1.2   Conditional Probability and Bayes' Formula

1. (Boys and Girls) A company is holding a dinner for working mothers with at least one son.

   (a) Ms. Jackson, a mother with two children, is invited. What is the probability that both children are boys?

   (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?

Ans: The sample space of the 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 of $\frac{1}{4}$.

(a) Since Ms. Jackson is invited, she has at least one son. Let B be the evenet that at least one of the children is a boy and A be the event that both children are boys, we have

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

(b) 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 $\frac{1}{2}$.

2. (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?

Ans: Do not let the word "prefer" and a wrong intuition misguide us. The fraction of baby girls are driven by nature, or at least the X and Y chromosomes (/ˈkroʊməsəʊm/), not by the couples' preference. We 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. Therefore, a couple having multiple children can be equivalently viewed as multiple couples each having only one child. As a consequence, the fraction of girls born is always 50% and the fractions of girls in the society will stay stable at 50%.

3. (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?

Ans: Let A be the event that the chosen coin is unfair and B be the event that all 10 tosses are heads. If the coin is fair, each time, the probability of getting a head is $\frac{1}{2}$. If the coin is unfair, the probability of getting a head is 1. By Bayes' formula, we have

$$
\begin{aligned}
P(A|B) &= \frac{P(A \cap B)}{P(B)} \\
&= \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)} \\
&= \frac{\frac{1}{1000} \times 1}{\frac{1}{1000} \times 1 + \frac{999}{1000} \times \left(\frac{1}{2}\right)^{10}} \\
&= \frac{2^{10}}{2^{10} + 999} \approx 0.506.
\end{aligned}
$$

Remark 3 *A^c can be read as "not A" or A complement (/ˈkɑːmplɪment/), and $P(A|B)$ can be read as "the probability of A given B" or "P of A given B".*

4. (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?

   Ans: Assume that the probability of getting heads is p. Then the probability of getting tails is $1 - p$. Consider two independent tosses. We have four possible outcomes HH, HT, TH and TT with probabilities

   $$
   \begin{aligned}
   P(HH) &= pp, & P(HT) &= p(1-p), \\
   P(TH) &= (1-p)p, & P(TT) &= (1-p)(1-p).
   \end{aligned}
   $$

   Hence $P(HT) = P(TH)$. By assigning HT to wining and TH to losing, we can generate even odds.

5. (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.

Ans: We rank the results of the three dart throws based in their distance from the center, from nearest (/ˈnɪrəst/) to farthest (/ˈfɑːrðɪst/), as A, B, C. Then there are 6 possible outcomes with equal probabilities:

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

Let X be the event that the second throw is farther from the center than the first, and Y be the event that the third throw is from the center than the first. Then the event X consists of outcomes 1, 3 and 5, and the event XY (X and Y) consists of 1 and 3. Hence we have

$$P(Y|X) = \frac{P(XY)}{P(X)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

6. (Generalized Dart Game) Jason throws n ($n \geq 3$) darts at a dartboard, aiming for the center. Assume that each subsequent dart is farther from the center than the first dart. If Jason throws a $(n+1)$-th dart aiming for the center, what is the probability that the $(n+1)$-th throw is farther from the center than the first?

Ans: Similarly, we rank the results of the $n + 1$ dart throws based in their distance from the center, from nearest to farthest, as A_1, A_2, $\cdots$, A_{n+1}. We can still enumerate all the possible outcomes which have equal probabilities (but we would not actually do it).

Let X be the event that the subsequent $n - 1$ throws are all farther from the center than the first, and Y be the event that the $(n+1)$-th throw is nearer from the center than the first.

When the event X occurs, either the first throw is the best one, or the first throw is the second best one while the $(n+1)$-th throw is the best one. In other words, the event X consists of the following two types of outcomes:

1. **Type 1:** The first throw is labelled as A_1. At this point, the rankings of the other n throws can be arbitrary. Hence there are $n!$ such outcomes.

2. **Type 2:** The first throw is labelled as A_2 while the $(n + 1)$-th throw is labelled as A_1. At this point, the rankings of the other $n - 1$ throws can be arbitrary. Hence there are $(n - 1)!$ such outcomes.

Hence the event X consists of $n! + (n - 1)!$ outcomes.

Similarly, it is easy to see that the event XY consists of the second type of outcomes above, i.e., the event XY consists of $(n - 1)!$ outcomes. Therefore, we have

$$P(Y|X) = \frac{P(XY)}{P(X)} = \frac{(n - 1)!}{n! + (n - 1)!} = \frac{1}{n + 1}.$$

The probability that the $(n+1)$-th throw is farther from the center than the first is

$$P(Y^c|X) = 1 - P(Y|X) = \frac{n}{n + 1}.$$

7. (Birthday Line) At a movie theater, a whimsical (/ˈwɪmzɪkəl/, 古怪的) 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 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?

Ans: Let A_n be the event that the n-th person in line gets the free ticket. Then it is clear that the event A_n is equivalent to the event that the first $n - 1$ people in line have different birthdays and the n-th person has the same birthday as one of the first $n - 1$ people. Let B_{n-1} be the event that the first $n - 1$ people in line have different birthdays. Then we have

$$P(B_{n-1}) = \frac{A_{365}^{n-1}}{365^{n-1}} = \frac{365 \times 364 \times \cdots \times (365 - n + 2)}{365^{n-1}}.$$

Let C_n be the event that the n-th person has the same birthday as one of the first $n - 1$ people. Then we have

$$P(C_n|B_{n-1}) = \frac{n - 1}{365},$$

that is, we can choose any of the first $n-1$ people's birthdays (which are pairwise

distinct) as the n-th person's birthday. Hence we have

$$P(A_n) = P(B_{n-1}C_n) = P(B_{n-1})P(C_n|B_{n-1})$$
$$= \frac{365 \times 364 \times \cdots \times (365 - n + 2) \times (n - 1)}{365^n}.$$

To determine for which n the probability $P(A_n)$ is maximized, we solve the following system of inequalities:

$$\begin{cases} P(A_n) \geq P(A_{n-1}), \Leftrightarrow \frac{A_{365}^{n-1} \times (n-1)}{365^{n-1}} \geq \frac{A_{365}^{n-2} \times (n-2)}{365^{n-2}}, \\ P(A_n) \geq P(A_{n+1}) \Leftrightarrow \frac{A_{365}^{n-1} \times (n-1)}{365^{n-1}} \geq \frac{A_{365}^{n} \times n}{365^{n}}, \end{cases}$$

which simplifies to

$$\begin{cases} (365 - n + 2)(n - 1) \geq 365(n - 2), \\ (365 - n + 1)n \leq 365(n - 1), \end{cases} \Leftrightarrow \begin{cases} n^2 - 3n - 363 \leq 0, \\ n^2 - n - 365 \geq 0. \end{cases}$$

The solution to the above system of inequalities is $n = 20$. Hence the best position in line is the 20-th.

Remark 4 A_{365}^{n-1} is the number of permutations of $n-1$ elements chosen from a set of 365 elements, which can be read as "365 permute $n-1$".

8. (Dice Order) We throw 3 dice one by one. What is the probability that we obtain 3 points in strictly increasing order?

Ans: The sample space of the three dice is given by

$$\Omega = \{(a_1, a_2, a_3) : a_1, a_2, a_3 \in \{1, 2, \cdots, 6\}\}$$

and each outcome has the same probability. We have $\#\Omega = 6^3$. Let X be the event that we obtain 3 points in strictly increasing order. If a_1, a_2, a_3 form a strictly increasing sequence, then they must be pairwise distinct. Moreover, any three distinct numbers from $\{1, 2, \cdots, 6\}$ can be uniquely arranged in strictly increasing order. Hence we have a bijection from the set of 3-subsets of $\{1, 2, \cdots, 6\}$ to the event X, which implies that $\#X = \binom{6}{3}$. Therefore, we have

$$P(X) = \frac{\#X}{\#\Omega} = \frac{\binom{6}{3}}{6^3} = \frac{5}{54}.$$

Remark 5 *Dice /daɪs/, n. 骰子.*

Remark 6 *#X can be read as "the cardinality (/kɑːrdɪˈnæləti/) of X" or "the number of elements in X".*

Remark 7 $\binom{6}{3}$ *can be read as "6 choose 3".*

Remark 8 $\{1, 2, \cdots, 6\}$ *can be read as "the set of integers from 1 to 6".*

9. (Monty Hall) 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?

   > **Ans:** If we don't switch, then whether we win or not is independent of Monty's action of showing us a goat, so our probability of wining is $\frac{1}{3}$. To calculate the probability of winning if we switch, we may assume that the car is behind door 1 without loss of generality. There are three fundamental outcomes, namely choosing door 1, door 2, or door 3, each with an equal probability.
   >
   > 1. If we choose door 1, then Monty will open one of the other two doors, say door 2. If we switch, then the new door we choose is door 3, and we loose.
   >
   > 2. If we choose door 2, then Monty will open door 3. If we switch, then the new door we choose is door 1, and we win.
   >
   > 3. If we choose door 3, then by symmetry, we will win.
   >
   > Hence, among the three fundamental outcomes, there are two that lead to winning, which implies that the probability of winning if we switch is $\frac{2}{3}$. It is larger than $\frac{1}{3}$, so we should switch.

10. (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?

Ans: Let E be the event that the amoeba population will die out. Let F_1 be the event of the original amoeba dying, F_2 be the event of the amoeba staying the same, F_3 be the event of the original amoeba splitting into two, and F_4 be the event of the original amoeba splitting into three. Applying the law of total probability, we have

$$P(E) = P(E|F_1)P(F_1) + P(E|F_2)P(F_2) + P(E|F_3)P(F_3) + P(E|F_4)P(F_4)$$
$$= \frac{1}{4}\Big(P(E|F_1) + P(E|F_2) + P(E|F_3) + P(E|F_4)\Big). \qquad (1)$$

1. If F_1 happens, then no amoeba is left, which implies that $P(E|F_1) = 1$.

2. If F_2 happens, then the new state is the same as the beginning, which implies that $P(E|F_2) = P(E)$.

3. If F_3 happens, then there are two amoebas left. Since the behavior of the two amoebas are independent, we can equivalently consider them as being in two completely independent ponds, making the extinction of the two amoeba populations independent events, which implies that $P(E|F_3) = P(E)^2$.

4. Similarly, we have $P(E|F_4) = P(E)^3$.

Substituting all the conditional probabilities into equation (1), we obtain that

$$P(E) = \frac{1}{4}\Big(1 + P(E) + P(E)^2 + P(E)^3\Big),$$

i.e.,

$$P(E)^3 + P(E)^2 - 3P(E) + 1 = 0. \qquad (2)$$

It is clear that $P(E) = 1$ is a solution to the above equation, and thus a factor of $P(E) - 1$ can be factored out from the cubic polynomial on the left-hand side of the equation. Indeed, we have

$$P(E)^3 + P(E)^2 - 3P(E) + 1 = (P(E) - 1)(P(E)^2 + 2P(E) - 1),$$

which implies that the three solutions of equation (2) are 1, $-1+\sqrt{2}$, and $-1-\sqrt{2}$. Since $P(E)$ is a probability, we have $P(E) = \sqrt{2} - 1$ or 1.

According to the theory of branching processes,

1. if the mean number of offspring produced by a single parent is less than or equal to one, then the ultimate extinction probability is one;

2. if the mean number of offspring produced by a single parent is greater than one, then the ultimate extinction probability is less than one.

This result can be proved using generating functions. In our case, the mean number of offspring produced by a single parent is $\frac{1}{4} \times 0 + \frac{1}{4} \times 1 + \frac{1}{4} \times 2 + \frac{1}{4} \times 3 = \frac{7}{4} > 1$, which implies that the ultimate extinction probability is $\sqrt{2} - 1$.

Remark 9 *amoeba /ə'miːbə/, n. 变形虫.*

Remark 10 *offspring /'ɔːfsprɪŋ/, n.(u) 后代，结果，产物.*

Remark 11 *ultimate /'ʌːltɪmət/, adj. 最终的，根本的.*

11. (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?

Ans: Let T_r, T_b and T_g be the total number of candies that have been taken out when the last red, blue and green candy is taken out, respectively. There are all random variables. What we want is exactly the conjunction of the event $T_r < T_b$ and the event $T_r < T_g$, which we denote by A.

The event A can be decomposed as the union of the following two mutually exclusive events: $T_r < T_b < T_g$ and $T_r < T_g < T_b$. Hence

$$P(A) = P(T_r < T_b < T_g) + P(T_r < T_g < T_b).$$

Note that the event $T_r < T_b < T_g$ is equivalent to the conjunction the event $T_r < T_b$ and the event $T_g = 60$. It is clear that the event $T_r < T_b$ and the event $T_g = 60$ are independent, which implies that

$$P(T_r < T_b < T_g) = P(T_r < T_b)P(T_g = 60).$$

Since each of the 60 candies are equally likely to be the last candy and there are 30 green ones, we have $P(T_g = 60) = \frac{30}{60} = \frac{1}{2}$. Let T_b' be the total number of red and blue candies that have been taken out when the last blue candy is taken out, respectively. Then the event $T_r < T_b$ is equivalent to the event $T_b' = 30$. Among the 30 red and blue candies, each candy is again equally likely to be the

---

last candy and there are 20 blue candies. Hence we have $P(T_r < T_b) = \frac{20}{30} = \frac{2}{3}$ and thus
$$P(T_r < T_b < T_g) = \frac{1}{2} \times \frac{2}{3} = \frac{1}{3}.$$

Similarly, we have
$$P(T_r < T_g < T_b) = \frac{20}{60} \times \frac{30}{40} = \frac{1}{4}.$$

Therefore, we have
$$P(A) = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}.$$

Remark 12 *jar /dʒɑːr/, n.* 玻璃罐，广口瓶.

Remark 13 *mutual /ˈmjuːtʃuəl/, adj.* 相互的，共同的.

Remark 14 *exclusive /ɪkˈskluːsɪv/, adj.* 独有的，专用的，排外的.

Remark 15 *conjunction /kənˈdʒʌŋkʃ(ə)n/, n.* 连接，联合.

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?

Ans: Let p be the probability that the first player (the one who tosses the coin first) wins, let q be the probability that the first player wins given that the first toss is H, and let X and Y be the event that A and B wins, respectively.

Let's condition $P(A)$ on A' first toss, which has a $\frac{1}{2}$ probability of being a head. By the law of total probability, we have

$$p = P(X) = \frac{1}{2}P(X|T) + \frac{1}{2}P(X|H). \tag{3}$$

1. If the first toss of A is T, then we can effectively treat the subsequent game as a complete restart, except that player B now tosses first. The event that A wins (which means B loses) in the original game, corresponds to the event that the first player in the restarted game losses. In other words, we have
$$P(X|T) = 1 - p.$$

Page 14

2. If the first toss of A is H, let's further condition $P(X|H)$ on B's first toss. By the law of total probability, we have

$$q = P(X|H) = \frac{1}{2}P(X|HT) + \frac{1}{2}P(X|HH). \tag{4}$$

If the first toss of B is T, then the game ends and A loesses, which implies that $P(X|HT) = 0$. If the first toss of B is H, we consider B's first toss and the subsequent process as a new game, where the first player in the new game has tossed a head. The event that A wins (which means B loses) in the original game, corresponds to the event that the first player in the restarted game losses. In other words, we have

$$P(X|HH) = 1 - q = 1 - P(X|H).$$

Substituting all the conditional probabilities into equation (4), we obtain that

$$q = \frac{1}{2}(1 - q),$$

which implies that $P(X|H) = q = \frac{1}{3}$.

Substituting all the conditional probabilities into equation (3), we obtain that

$$p = \frac{1}{2}(1 - p) + \frac{1}{2} \times \frac{1}{3},$$

which implies that

$$P(X) = p = \frac{4}{9}.$$

Remark 16 *We can see that $P(X) < \frac{1}{2}$. This is reasonable since A cannot win in this first toss, yet B has a $\frac{1}{4}$ probability to win in her first toss.*

13. (Russian Roulette Series)

   (a) 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?

(b) If the barrel is spun again after each trigger pull, then will you choose to be the first or the second player? And what is your probability of loss?

(c) 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?

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

Ans: We number the chambers as $1, 2, \cdots, 6$.

(a) The bullet has an equal probability of lying in each chamber, which is $\frac{1}{6}$. The first player loses if and only if the bullet is in chamber 1, 3 and 5. Hence the probability of his loss is $\frac{1}{2}$. It turns out whether to go first or second does not matter.

(b) 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 p on the first player's first trigger pull. He has a probability of $\frac{1}{6}$ of losing in this run. Otherwise, the game can be viewed as restarting with the second player going first. Hence we have

$$ p = \frac{1}{6} \times 1 + \frac{5}{6} \times (1 - p), $$

which implies that $p = \frac{6}{11}$. Therefore, it is better to go second and the probability of loss is $\frac{5}{11}$.

(c) If we spin the barrel, since each chamber is equally likely to be under the hammer and there are two containing a bullet, the probability that we will lose in this run is $\frac{2}{6}$. If we don't spin the barrel, there are only 5 chambers left, each equally likely to contain a bullet, and two of them do. Hence the probability that we will lose in this run is $\frac{2}{5}$. Since $\frac{2}{5} > \frac{2}{6}$, we should spin the barrel.

(d) We label the known empty chamber as 1. At the initial moment, the two bullets are equally likely to be positioned in the following locations: 12, 23, 34, 45, 56 and 61, each with a probability of $\frac{1}{6}$. If we spin the barrel, then we will lose when the two bullets are in the following positions: 12 and 61. Hence the probability that we will lose in this run is $\frac{1}{3}$. If we don't spin the barrel, then the possible positions of the two bullets are 23, 34, 45 and 56, each with a probability of $\frac{1}{4}$. We will lose when the two bullets are in position 23, and thus the probability that we will lose in this run is $\frac{1}{4}$. Since $\frac{1}{4} < \frac{1}{3}$, we should not spin the barrel.

Remark 17 *We can view the game (a) as a lottery process. We have the well-known result in probability theory that "the probability of winning a prize is independent of the drawing order".*

Remark 18 *In game (b), since each shot is an independent event with a $\frac{1}{6}$ chance of losing, the first player must face this risk immediately, while the second player only has to shoot if the first player survives. This effectively reduces the second player's overall exposure to risk.*

Remark 19 *roulette /ruː'let/, n.* 轮盘赌.

Remark 20 *bullet /'bʊlɪt/, n.* 子弹.

Remark 21 *revolver /rɪ'vɑlvə/, n.* 左轮手枪.

Remark 22 *barrel /'bærəl/, n.* 枪管.

Remark 23 *spin /spɪn/, v.* 旋转.

Remark 24 *hammer /'hæmə/, n.* 锤子，枪锤.

Remark 25 *trigger /'trɪgər/, n.* 扳机，触发器.

Remark 26 *go off* 突然爆炸或发出巨响.

Remark 27 *lottery /'lɑtəri/, n.* 彩票，抽奖，碰运气的事.

14. (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?

> **Ans:** We can assume that there are 52 positions, with the first 13 positions belonging to the first player, the next 13 positions to the second player, and so on. Distributing the cards among the four players is equivalent to arranging the 52 cards into these 52 positions. There are in total 52! ways to arrange the cards.
>
> We first arrange the four Aces. Given any specific arrangement of these four Aces (there are 4! = 24 such arrangements), we choose one of the first 13 positions to place the first Ace, one of the next 13 positions to place the second Ace, and so on. Once the four Aces are placed, the remaining 48 cards can be freely arranged

in the remaining 48 positions. Hence the total number of ways to distribute the cards such that each player has an Ace is

$$4! \times \binom{13}{1}^4 \times 58!$$

and the probability that each player has an Ace is

$$\frac{4! \times \binom{13}{1}^4 \times 48!}{52!} = \frac{4! \times 13^4}{52 \times 51 \times 50 \times 49} = \frac{39 \times 26 \times 13}{51 \times 50 \times 49}.$$

Remark 28 *ace /eɪs/, n.* （扑克牌中的）*A*，发球得分.

15. (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?

Ans: For any $0 \le i \le N$, let $p(i)$ be the probability that the gambler will end up with N dollars, given that he starts with i dollars. If $i = 0$, i.e., he starts with 0 dollars, then according to the rules, he will stop immediately and can never accumulate N dollars. Hence $p(0) = 0$. Similarly, we have $p(N) = 1$.

Now assume that $0 < i < N$. We condition the probability $p(i)$ on the result of the first game. If he wins the first game (with a probability of p), then we can consider the game as restarting, with the player's initial capital being $i + 1$; if he loses the first game (with a probability of q), then we can consider the game as restarting, with the player's initial capital being $i - 1$. By the law of total probability, we have

$$p(i) = p \times p(i + 1) + q \times p(i - 1).$$

Hence the sequence $\{p(i)\}$ is a homogenous linear recurrence sequence. Its characteristic equation is $x = p \cdot x^2 + q$, or equivalently, $px^2 - x + q = 0$. The discriminant of this quadratic polynomial is $\Delta = 1 - 4pq = 1 - 4p(1 - p) = (2p - 1)^2$.

1. If $p = q = \frac{1}{2}$, then the characteristic equation has a double root $x = 1$. According to the theory of linear recurrence sequences, the general solution to the recurrence relation is

$$p(i) = (A + B \cdot i) \cdot 1^i = A + B \cdot i,$$

where A and B are constants. Since $p(0) = 0$ and $p(N) = 1$, we have $A = 0$ and $B = \frac{1}{N}$. Hence $p(i) = \frac{i}{N}$.

2. If $p \neq \frac{1}{2}$, then the characteristic equation has two distinct roots, i.e., $x_1 = 1$ and $x_2 = \frac{q}{p}$. According to the theory of linear recurrence sequences, the general solution to the recurrence relation is

$$p(i) = A \cdot 1^i + B \cdot \left(\frac{q}{p}\right)^i = A + B \cdot \left(\frac{q}{p}\right)^i,$$

where A and B are constants. Since $p(0) = 0$ and $p(N) = 1$, we have $A = 0$ and $B = \frac{1}{1-\left(\frac{q}{p}\right)^N}$. Hence $p(i) = \frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{1-p}{p}\right)^N}$.

Remark 29 *gambler /ˈgæmblər/, n.* 赌徒.

Remark 30 *ruin /ruːɪn/, n.* 毁灭，崩溃，破产.

Remark 31 *accumulate /əˈkjuːmjəleɪt/, v.* 积累，积聚.

16. (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 40-th throw, the probability that she will score in the 41-th throw is $\frac{23}{40}$. After 100 throws (including the first and the second), what is the probability that she scores exactly 50 baskets?

Ans: For any integers n and k such that $n \geq 2$ and $0 \leq k \leq n$, let $A_{n,k}$ be the event that the player scores exactly k baskets after n throws and let $P_{n,k} = P(A_{n,k})$. According to the rules, we have $P_{n,0} = P_{n,n} = 0$ for any $n \geq 2$. Moreover, we have $P_{2,1} = 1$.

Let's condition $P_{n,k}$ ($k \geq 1$) on the result of the n throw. If the n-th throw is a basket, then the player has scored exactly $k-1$ baskets after $n-1$ throws. According to the rules, in this case, the probability that she will score in the n-th throw is $\frac{k-1}{n-1}$. If the n-th throw is a miss, then the player has scored exactly k baskets after $n-1$ throws. According to the rules, the probability that she will miss in the n-th throw is $\frac{n-1-k}{n-1}$. By the law of total probability, we have

$$P_{n,k} = \frac{k-1}{n-1}P_{n-1,k-1} + \frac{n-1-k}{n-1}P_{n-1,k}.$$

We have
$$P_{3,1} = \frac{0}{2} \cdot P_{2,0} + \frac{2-1}{2} \cdot P_{2,1} = \frac{1}{2},$$
$$P_{3,2} = \frac{1}{2} \cdot P_{2,1} + \frac{2-2}{2} P_{2,2} = \frac{1}{2},$$
$$P_{4,1} = \frac{0}{3} \cdot P_{3,0} + \frac{3-1}{3} \cdot P_{3,1} = \frac{1}{3},$$
$$P_{4,2} = \frac{1}{3} \cdot P_{3,1} + \frac{3-2}{3} \cdot P_{3,2} = \frac{1}{6} + \frac{1}{6} = \frac{1}{3},$$
and
$$P_{4,3} = \frac{2}{3} \cdot P_{3,2} + \frac{3-3}{3} \cdot P_{3,3} = \frac{1}{3}.$$
Hence we may conjecture that $P_{n,k} = \frac{1}{n-1}$ for any $n \geq 2$ and $1 \leq k \leq n-1$.

We can prove the conjecture by induction. Assume that $P_{n,k} = \frac{1}{n-1}$ for any $n \geq 2$ and $1 \leq k \leq n-1$. Then for any $1 \leq k \leq n$, using the recurrence relation, we have
$$P_{n+1,k} = \frac{k-1}{n} \cdot P_{n,k-1} + \frac{n-k}{n} \cdot P_{n,k}.$$
If $k = 1$, then
$$P_{n+1,1} = 0 \cdot P_{n,0} + \frac{n-1}{n} \cdot P_{n,1} = \frac{1}{n}.$$
If $k = n$, then
$$P_{n+1,n} = \frac{n-1}{n} \cdot P_{n,n-1} + 0 \cdot P_{n,n} = \frac{1}{n}.$$
If $1 < k < n$, then
$$P_{n+1,k} = \frac{k-1}{n} \cdot \frac{1}{n-1} + \frac{n-k}{n} \cdot \frac{1}{n-1} = \frac{1}{n}.$$
Therefore, by induction, we have $P_{n,k} = \frac{1}{n-1}$ for any $n \geq 2$ and $1 \leq k \leq n-1$. In particular, we have $P_{100,50} = \frac{1}{99}$.

Remark 32 *hoop /huːp/, n.* 籃筐.

17. (Cars on Road) If the probability of observing at least one car on a highway during any 20-minute time interval is $\frac{609}{625}$, 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.

Ans: We can break down the 20-minute interval into a sequence of 4 non-overlapping 5-minute intervals. Because of the constant probability of observing a car at any given moment, the probability of observing a car during any 5-minute interval is constant. Let's denote this probability by p. Then the probability of not observing a car during any 5-minute interval is $1-p$. Hence the probability of not observing a car in all four of such independent 5-minute intervals is $(1-p)^4$. By assumption, we have $(1-p)^4 = 1 - \frac{609}{625} = \frac{16}{625}$, which implies that $p = \frac{3}{5}$.

Remark 33 *highway* /ˈhaɪweɪ/, *n.* 高速公路.

1.3 Discrete and Continuous Distributions

1.3.1 Important Discrete and Continuous Distributions

The following are the most commonly encountered discrete random variables.

1. Binomial random variable represents the number of successes in a sequence of n experiments when each trial is independently a success with probability p. The probability mass function of a binomial random variable X is given by

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \cdots, n.$$

The expected value of X is $\mathbb{E}(X) = np$ and the variance of X is $\text{Var}(X) = npq$, where $q = 1 - p$.

2. Poisson (/pwəˈsoʊn/) random variable represents the number of events occurring in a fixed time interval, given that events happen at a known average rate λ and independently of the time since the previous event, resulting in an expected count of λt, where t is the length of the time interval. The probability mass function of a Poisson random variable X is given by

$$p(x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, \quad x = 0, 1, 2, \cdots.$$

The expected value of X is $\mathbb{E}(X) = \lambda t$ and the variance of X is $\text{Var}(E) = \lambda t$.

3. Geometric random variable represents the number (n) of trials to get the first success when each trial is independently a success with probability p. The probability mass function of a geometric random variable X is given by

$$p(x) = p(1-p)^{x-1}, \quad x = 1, 2, 3, \cdots.$$

The expected value of X is $\mathbb{E}(X) = \frac{1}{p}$ and the variance of X is $\text{Var}(X) = \frac{q}{p^2}$, where $q = 1 - p$.

4. Negative binomial random variable represents the number of trials to get the r-th success when each trial is independently a success with probability p. The probability mass function of a negative binomial random variable X is given by

$$p(x) = \binom{n-1}{r-1} p^r (1-p)^{x-r}, \quad x = r, r+1, r+2, \cdots.$$

The expected value of X is $\mathbb{E}(X) = \frac{r}{p}$ and the variance of X is $\mathrm{Var}(X) = \frac{rq}{p^2}$, where $q = 1 - p$.

The following are the most commonly encountered continuous random variables.

1. Uniform distribution describes a random variable uniformly distributed over a finite interval $[a, b]$. The probability density function of a uniform random variable X is given by

$$p(x) = \begin{cases} 0 & x < a, \\ \frac{1}{b-a} & a \leq x \leq b, \\ 0 & x > b. \end{cases}$$

The expected value of X is $\mathbb{E}(X) = \frac{a+b}{2}$ and the variance of X is $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$.

2. Normal/Gaussian (/ˈgaʊsiən/) distribution is by far the most popular continuous distribution. The probability density function of a normal random variable is given by

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

The expected value of X is $\mathbb{E}(X) = \mu$ and the variance of X is $\mathrm{Var}(X) = \sigma^2$.

3. Exponential distribution models the arrival time of an event if it has a constant arrival rate λ. The probability density function of an exponential random variable T is given by

$$p(t) = \begin{cases} 0 & t < 0, \\ \lambda e^{-\lambda x} & t \geq 0. \end{cases}$$

The expected value of T is $\mathbb{E}(T) = \frac{1}{\lambda}$ and the variance of T is $\mathrm{Var}(T) = \frac{1}{\lambda^2}$.

4. Gamma distribution models the amount of time one has to wait until a total of n events occur. The probability density function of a gamma random variable X is given by

$$p(x) = \begin{cases} 0 & x < 0, \\ \frac{\lambda e^{-\lambda x} (\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & x \geq 0, \end{cases}$$

where $\Gamma(a) = \int_0^\infty e^{-y} y^{a-1} \, \mathrm{d}y$. The expected value of X is $\mathbb{E}(X) = \frac{\alpha}{\lambda}$ and the variance of X is $\mathrm{Var}(X) = \frac{\alpha}{\lambda^2}$.

5. Beta distributions are used to model events that are constrained within a defined interval. By adjusting the shape parameters α and β, it can model different shapes of probability density functions. The probability density function of a beta random variable X is given by

$$p(x) = \begin{cases} 0 & x < 0, \\ \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & 0 \leq x \leq 1, \\ 0 & x > 1. \end{cases}$$

The expected value of X is $\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}$ and the variance of X is $\mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

1.3.2 Memorylessness of Exponential Distribution

Exponential distribution models the arrival time of an event if it has a constant arrival rate λ. Since the event has a constant arrival rate λ, the expected arrival time should be $\frac{1}{\lambda}$.

An important property of exponential distribution is memorylessness: if a random variable T follows an exponential distribution with parameter λ, then

$$P(T > s + t | T > s) = P(T > t), \quad \forall\ s, t \geq 0; \tag{5}$$

that is, if we have waited for s time units, the extra waiting time follows the same distribution as the waiting time when we start at time 0. Indeed, the probability density function of T is

$$p(t) = \begin{cases} 0 & t < 0, \\ \lambda e^{-\lambda t} & t \geq 0, \end{cases}$$

which implies that the cumulative distribution function of T is

$$F(t) = P(T < t) = \begin{cases} 0 & t < 0, \\ 1 - e^{-\lambda t} & t \geq 0. \end{cases}$$

It follows that

$$\begin{aligned} P(T > s + t | T > s) &= \frac{P(T > s + t, T > s)}{P(T > s)} \\ &= \frac{P(T > s + t)}{P(T > s)} \\ &= \frac{1 - (1 - e^{-\lambda(s+t)})}{1 - (1 - e^{-\lambda s})} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} \\ &= 1 - (1 - e^{-\lambda t}) = P(T > t) \end{aligned}$$

for any $s, t \geq 0$.

Conversely, assume that T is a continuous random variable with expected value $\frac{1}{\lambda}$ whose probability density function vanishes on $(-\infty, 0)$. If T satisfies the memoryless property (5), then T must follow the exponential distribution with parameter λ. Indeed, let F be the cumulative distribution function of T and let $G = 1 - F$. Then G is a continuous function. The memoryless property can be reformulated as

$$G(s + t) = G(t)G(s)$$

for any $s, t \geq 0$. This is the well-known Cauchy equation, whose general solution has the form $G(t) = e^{ct}$, where c is a constant. Since $\lim_{t \to \infty} G(t) = 0$, we have $c < 0$. Then probability density function of T is $p(t) = -G'(t) = -ce^{ct}$ ($t \geq 0$) and the expected value of T is

$$\mathbb{E}(T) = \int_0^\infty -tce^{ct}\, \mathrm{d}t = \frac{1}{c}e^{ct}\Big|_0^\infty - e^{ct}t\Big|_0^\infty.$$

Since $c < 0$, we have $\mathbb{E}(T) = -\frac{1}{c}$. By assumption, we have $c = -\lambda$ and thus the density function of T is

$$p(t) = \begin{cases} 0 & t < 0, \\ \lambda e^{-\lambda t} & t \geq 0. \end{cases}$$

This tells us that the memoryless property is a characteristic property of the exponential distribution.

1.4   Poisson Process

When the arrivals of a series of events each independently follow an exponential distribution with arrival rate λ, the number of arrivals between time 0 and t can be modeled as a Poisson process

$$P(N(t) = x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, \quad x = 0, 1, 2, \cdots.$$

Let's give a rigorous derivation. Let $T_1, T_2, \cdots$ be i.i.d. (independent and identically distributed) exponential random variables with parameter λ. We interpret T_i as the waiting time between the $(i-1)$-th and i-th arrival. For any $n \geq 1$, define the arrival time

$$S_n = T_1 + T_2 + \cdots + T_n,$$

i.e., S_n is the time that the n-th event occurs. For any $t \geq 0$, put

$$N(t) = \max\{n : \ S_n \leq t\},$$

i.e., the total number of arrivals up to time t. We want to show that $N(t) \sim \text{Poisson}(\lambda t)$. For any $k \geq 0$, put $p_k(t) = P\big(N(t) = k\big)$.

First, we compute

$$p_0(t) = P\big(N(t) = 0\big) = P(\text{no arrivals occur between time } 0 \text{ and time } t).$$

This is exactly the event that the first arrival time T_1 is greater than t. Since $T_1 \sim \text{Exp}(\lambda)$, we have

$$p_0(t) = P(T_1 > t) = e^{-\lambda t}.$$

Next, assume $k \geq 1$. We will construct a recurrence relation using the memeoryless property of exponential distribution. Indeed, we can set up an integral equation by conditioning on when the first arrival occurs.

- Suppose that the first arrival time is $s \in (0, t)$.

- Then, for there to be k arrivals in $[0, t]$, we need exactly $k - 1$ additional arrivals in the remaining time $(s, t]$.

- By the memoryless property of the exponential distribution, once the first arrival happens at time s, the number of arrivals in $(s, t]$ has the same distribution as $N(t - s)$.

Hence we have

$$p_k(t) = \int_0^t \Big(P\big(\text{first arrival at time } [s, s + ds)\big) \times P\big(\text{there are } k - 1 \text{ more arrivals in } (s, t]\big) \Big).$$

Since $T_1 \sim \text{Exp}(\lambda)$, the density for the first arrival at s is $\lambda e^{-\lambda s}$. And the conditional probability of getting $k - 1$ arrivals thereafter is $p_{k-1}(t - s)$. Hence we have

$$p_k(t) = \int_0^t \lambda e^{-\lambda s} p_{k-1}(t - s) \, ds$$

for any $k \geq 1$. Along with

$$p_0(t) = e^{-\lambda t},$$

we obtain a convolution-type recurrence relation for $p_k(t)$. We can prove by induction that

$$p_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

for any $k \geq 0$.

1.4.1 Moment Generating Function

Let X be a discrete/continuous random variable with probability mass/density function $p(x)$. The moment generating function of X is defined as

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

We may assume that X is continuous. Then

$$M'(t) = \int_{-\infty}^{\infty} xe^{tx}p(x)\,\mathrm{d}x = \mathbb{E}[Xe^{tX}],$$

which implies that $M'(0) = \mathbb{E}[X]$. Similarly, we have

$$M''(t) = \int_{-\infty}^{\infty} x^2 e^{tx}p(x)\,\mathrm{d}x = \mathbb{E}[X^2 e^{tX}],$$

which implies that $M''(0) = \mathbb{E}[X^2]$. In general, we have

$$M^{(n)}(0) = \mathbb{E}[X^n].$$

1.4.2 Exercises

1. (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?

Ans: Assume that banker A arrives X minutes after 5:00 am and banker B arrives Y minutes after 5:00 am. Then both X and Y are uniformly distributed over the interval $[0, 60]$, and they are independent. Since both only stay exactly five minutes, A and B meet if and only if $|X - Y| \leq 5$.

Hence the probability that A and B will meet is simply the area of the shadowed region divided by the area of the whole square. The area of the shadowed region can be found by subtracting the areas of the two small triangles from the area of the whole square. We have

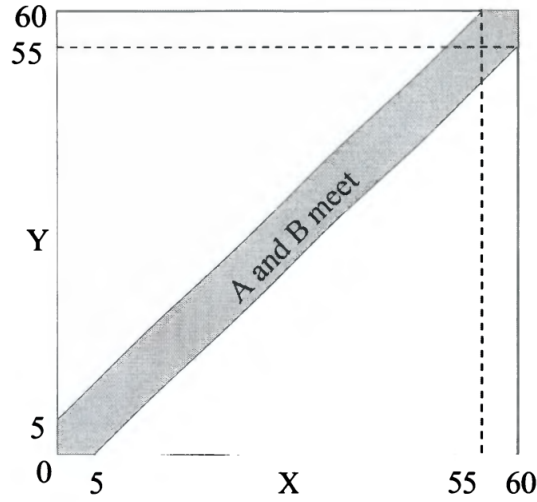$$P(|X - Y| \leq 5) = \frac{60^2 - 55^2}{60^2} = \frac{(60 + 55) \times (60 - 55)}{60^2} = \frac{23}{144}.$$

Figure 1: Distributions of Banker A's and Banker B's arrival times

Remark 34 *subtract /səbˈtrækt/, v.* 减去，扣除.

2. (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?

Ans: With loss of generality, we may assume that the length of the stick is 1. Let X and Y denote the coordinates of the cut points of first and second cuts, respectively. Then both X and Y are uniformly distributed over the interval $[0, 1]$, and they are independent. Let A be the event that the 3 segments can form a triangle. Then we have

$$P(A) = P(A, \ X < Y) + P(A, \ X > Y).$$

If $X < Y$, then the three segments are X, $Y - X$ and $1 - Y$. They can form a triangle if and only if the following conditions are satisfied:

$$\begin{cases} X + Y - X > 1 - Y, \\ Y - X + 1 - Y > X, \\ X + 1 - Y > Y - X, \end{cases} \Leftrightarrow \begin{cases} Y > \frac{1}{2}, \\ X < \frac{1}{2}, \\ 2X - 2Y + 1 > 0, \end{cases}$$

which implies that

$$P(A, X < Y) = P(Y > \frac{1}{2}, X < \frac{1}{2}, 2X - 2Y + 1 > 0).$$

Page 27

Hence the probability $P(A, X < Y)$ is simply the area of the shadowed region divided by the area of the whole square.
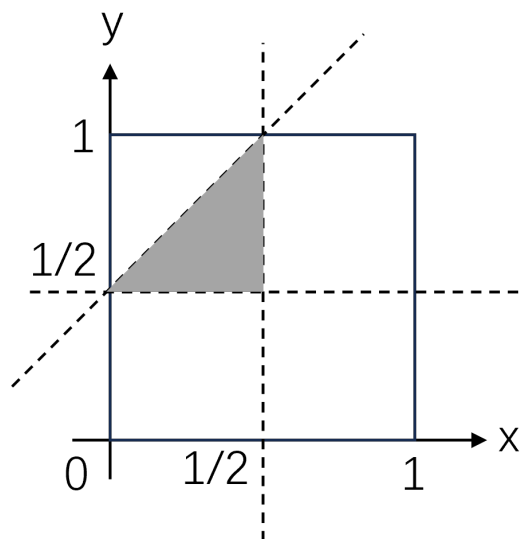

Figure 2: Distributions of X and Y

We have
$$P(A, X < Y) = \frac{1}{8}.$$

By symmetry, we have
$$P(A, Y < X) = P(A, X < Y) = \frac{1}{8}$$

and thus
$$P(A) = \frac{1}{4}.$$

3. (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 (A = 0.1/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?

Ans: Since the buses form a Poisson process with an average inter-arrival time of 10 minutes, the waiting-time for the next bus is exponentially distributed with mean 10. By the memoryless property of the exponential distribution, if we arrive at a random time, our expected waiting time for the next bus is still 10 minutes. If we look back in time, the memoryless property still applies. So on average, the last bus arrived 10 minutes ago as well.

4. (Moments of Normal Distribution) If X follows the standard normal distribution ($X \sim N(0, 1)$), what is $\mathbb{E}[X^n]$ for $n = 1, 2, 3, 4$?

Ans: The first moment of a distribution is always the expected value. Hence we have $\mathbb{E}[X] = 0$. Since the $\mathbb{E}[X] = 0$, the second moment of the stand normal distribution is the variance. Hence $\mathbb{E}[X^n] = 1$. Note that

$$\mathbb{E}[X^n] = \int_{-\infty}^{\infty} x^n p(x),$$

where $p(x)$ is the probability density function of X. Since $p(x)$ is symmetric about 0, we know that $\mathbb{E}[X^n] = 0$ for any odd positive number n.

For $\mathbb{E}[X^4]$, we can use integration by parts. However, to solve $\mathbb{E}[X^n]$ for any integer n, the approach using moment generating function may be a better choice. The moment generating function of X is

$$M(t) = \mathbb{E}[e^{tX}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} \cdot e^{-\frac{x^2}{2}} \, \mathrm{d}x = e^{\frac{t^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2}} \, \mathrm{d}x = e^{\frac{t^2}{2}}.$$

It follows that

$$M'(t) = te^{\frac{t^2}{2}},$$
$$M''(t) = e^{\frac{t^2}{2}} + t^2 e^{\frac{t^2}{2}},$$
$$M'''(t) = 3te^{\frac{t^2}{2}} + t^3 e^{\frac{t^2}{2}},$$

and

$$M^{(4)}(t) = 3e^{\frac{t^2}{2}} + 6t^2 e^{\frac{t^2}{2}} + t^4 e^{\frac{t^2}{2}},$$

which implies that

$$\mathbb{E}[X^4] = M^{(4)}(0) = 3 + 6 \cdot 0^2 + 0^4 = 3.$$

We can also use the Taylor expansion of $e^{t^2/2}$ at 0 to compute $\mathbb{E}[X^n]$. We have

$$M(t) = e^{\frac{t^2}{2}} = \sum_{n=0}^{\infty} \frac{(2n)!}{2^n n!} \frac{t^{2n}}{(2n)!},$$

which implies that

$$\mathbb{E}[X^n] = M^{(n)}(0) = \begin{cases} 0 & \text{if } n \text{ is odd,} \\ \frac{(n)!}{2^{\frac{n}{2}}(\frac{n}{2})!} & \text{if } n \text{ is even.} \end{cases}$$

In particular, we have $\mathbb{E}[x^4] = \frac{4!}{2^2 \cdot 2!} = 3$.

1.5 Expected Value, Variance and Covariance

1. If X and Y are two independent random variables, then

$$E(g(X)h(Y)) = E[g(X)]E(h(Y))$$

for any functions g and h.

2. For any random variables X and Y, the covariance of them is defined as

$$\text{Cov}(X,Y) = \mathbb{E}\Big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\Big] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

It is clear that
$$\text{Cov}(X,X) = \text{Var}(X).$$

Moreover, if X and Y are independent, then we have $\text{Cov}(X,Y) = 0$. However, the converse is not true. Generally, if $\text{Cov}(X,Y) = 0$, then X and Y are said to be uncorrelated.

The correlation coefficient of X and Y are defined as

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

Using the Cauchy-Schwarz (/ʃwɑːz/) inequality, we have $|\rho(X,Y)| \leq 1$.

3. Let $X_1, X_2, \cdots, X_n$ be n random variables. Then

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

In particular, if $X_1, X_2, \cdots, X_n$ are pairwise uncorrelated/independent, then

$$\text{Var}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \text{Var}(X_i).$$

4. Let X and Y be two random variables and let g be a function. Then the conditional expected value of $g(X)$ given $Y = y$ is defined by

$$\mathbb{E}[g(X)|Y = y] = \begin{cases} \sum_x g(x)p_{X|Y}(x|y) & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} g(x)p_{X|Y}(x|y)\,\mathrm{d}x & \text{if } X \text{ is continuous.} \end{cases}$$

5. (Law of total expectation) Let X and Y be two random variables. Then

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]] = \begin{cases} \sum_y \mathbb{E}[X|Y = y]p(Y = y) & \text{if } Y \text{ is discrete,} \\ \int_{-\infty}^{\infty} \mathbb{E}[X|Y = y]p_Y(y)\,\mathrm{d}y & \text{if } Y \text{ is continuous.} \end{cases}$$

Remark 35 *correlate /'kɔːrəleɪt/, v.* 相互关联.

1.5.1 Exercises

1. (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.

Ans: We consider the general case where there are n ($n \geq 1$) noodles. Let X_n be the random variable representing the number of circles formed by the n noodles and let $p_n(k) = P(X_n = k)$ for any $0 \leq k \leq n$.

It is clear that $p_n(0) = 0$. Moreover, the n noodles will ultimately form n loops if and only if, at each step, the two selected ends come from the same noodle. In the first selection, there are a total of $\binom{2n}{2}$ ways to choose two ends, out of which n ways correspond to selecting both ends from the same noodle. In the second selection, there are a total of $\binom{2n-2}{2}$ ways to choose two ends, out of which $n - 1$ ways correspond to selecting both ends from the same noodle. Proceeding in this way, we have

$$P_n(n) = \frac{n(n-1)\cdots 1}{\binom{2n}{2}\binom{2n-2}{2}\cdots\binom{2}{2}} = \frac{2^n n!}{(2n)!}.$$

Now assume that $n \geq 2$ and $1 \leq k \leq n-1$. Let Y be the event that the first two ends selected come from the same noodle. Then we have $P(Y) = \frac{n}{\binom{2n}{2}} = \frac{1}{2n-1}$. If Y occurs, a cycle is formed and the remaining $n - 1$ noodles remain unchanged. It follows that

$$P(X_n = k|Y) = P(X_{n-1} = k - 1) = p_{n-1}(k - 1).$$

If Y does not occur, then no cycle is formed any two noodles are connected together to form a longer noodle. We may view this longer noodle as a single noodle. It follows that

$$P(X_n = k|Y^c) = P(X_{n-1} = k) = p_{n-1}(k).$$

Applying the law of total probability, we have

$$
\begin{aligned}
p_n(k) &= P(X_n = k) \\
&= P(X_n = k|Y)P(Y) + P(X_n = k|Y^c)P(Y^c) \\
&= \frac{1}{2n-1}p_{n-1}(k-1) + \frac{2n-2}{2n-1}p_{n-1}(k).
\end{aligned}
$$

It follows that for any $n \geq 2$, we have

$$
\begin{aligned}
\mathbb{E}[X_n] &= \sum_{k=1}^{n} kp_n(k) \\
&= \sum_{k=1}^{n-1} \left(\frac{1}{2n-1}kp_{n-1}(k-1) + \frac{2n-2}{2n-1}kp_{n-1}(k) \right) + \frac{2^n n! \cdot n}{(2n)!} \\
&= \frac{1}{2n-1}\sum_{k=1}^{n-1} kp_{n-1}(k-1) + \frac{2n-2}{2n-1}\sum_{k=1}^{n-1} kp_{n-1}(k) + \frac{2^n n! \cdot n}{(2n)!} \\
&= \frac{1}{2n-1}\sum_{k=1}^{n} kp_{n-1}(k-1) + \frac{2n-2}{2n-1}\sum_{k=1}^{n-1} kp_{n-1}(k) + \frac{2^n n! \cdot n}{(2n)!} - \frac{n}{2n-1}p_{n-1}(n-1) \\
&= \frac{1}{2n-1}\sum_{k=1}^{n} kp_{n-1}(k-1) + \frac{2n-2}{2n-1}\sum_{k=1}^{n-1} kp_{n-1}(k) \\
&= \frac{1}{2n-1}\sum_{k=1}^{n} (k-1)p_{n-1}(k-1) + \frac{1}{2n-1}\sum_{k=1}^{n} p_{n-1}(k-1) + \frac{2n-2}{2n-1}\sum_{k=1}^{n-1} kp_{n-1}(k) \\
&= \frac{1}{2n-1}\mathbb{E}(X_{n-1}) + \frac{1}{2n-1} + \frac{2n-2}{2n-1}\mathbb{E}(X_{n-1}) \\
&= \mathbb{E}(X_{n-1}) + \frac{1}{2n-1}.
\end{aligned}
$$

Using $\mathbb{E}(X_1) = 1$ and the above recursion formula, we can compute $\mathbb{E}(X_{100})$.

It can be seen that the process of deriving the recursive formula for $\mathbb{E}[X_n]$ above is quite complex. Next, we will take a different approach. Consider a random variable R_n following a Bernoulli (/bərˈnuːli/) distribution with parameter $p = \frac{1}{2n-1}$ (that is, R_n takes the value 1 with probability $\frac{1}{2n-1}$ and the value 0 with

probability $\frac{2n-2}{2n-1}$), which is independent of X_{n-1}. Put $Z_n = R_n + X_{n-1}$. Then we can see that

$$
\begin{aligned}
P(Z_n = n) &= P(R_n = 1)P(X_{n-1} = n-1) \\
&= \frac{1}{2n-1} \cdot \frac{2^{n-1}(n-1)!}{(2n-2)!} \\
&= \frac{2^n n!}{(2n)!} = p_n(n)
\end{aligned}
$$

and

$$
\begin{aligned}
P(Z_n = k) &= P(R_n = 1)P(X_{n-1} = k-1) + P(R_n = 0)P(X_{n-1} = k) \\
&\frac{1}{2n-1}p_{n-1}(k-1) + \frac{2n-2}{2n-1}p_{n-1}(k) = p_n(k) \\
&= p_n(k)
\end{aligned}
$$

for any $1 \le k \le n-1$. Hence Z_n and X_n follow the same distribution and thus

$$
\mathbb{E}[X_n] = \mathbb{E}[Z_n] = \mathbb{E}[R_n] + \mathbb{E}[X_{n-1}] = \frac{1}{2n-1} + \mathbb{E}[X_{n-1}].
$$

Remark 36 *blindfold* /ˈblaɪndfəʊld/ *n.* 眼罩. *v.* 蒙住眼睛.

2. (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 σ_A^2; the variance of B's return is σ_B^2; their correlation coefficient is ρ.

Ans: Let r_A and r_B be the random variables representing the returns of stock A and stock B, respectively. Suppose that we short h shares of B. Then the variance of the hedged position is

$$
\begin{aligned}
\operatorname{Var}(r_A - hr_B) &= \operatorname{Var}(r_A) - 2h\operatorname{Cov}(r_A, r_B) + h^2 \operatorname{Val}(r_B) \\
&= \sigma_B^2 h^2 - 2h\rho\sigma_A\sigma_B + \sigma_A^2.
\end{aligned}
$$

This is a quadratic function of h. When h takes the coordinate of the axis of symmetry of its graph, i.e., $-\frac{-2h\rho\sigma_A\sigma_B}{2\sigma_B^2} = \rho\frac{\sigma_A}{\sigma_B}$, the variance of the hedged position is minimized. Hence the optimal hedge ratio is $\rho\frac{\sigma_A}{\sigma_B}$.

Remark 37 *hedge /hedʒ/ v.* 用树篱围住；限制；对冲，采取保值措施以规避风险. *n.* 树篱；防范措施；对冲.

Remark 38 *short v.* 卖空，做空.

Remark 39 *stock /stɑːk/ n.* 股票.

Remark 40 *position n.* 持仓量，投资头寸.

Remark 41 *hedged position* 对冲后的持仓.

Remark 42 *ratio /ˈreɪʃiəʊ/ n.* 比率.

3. (Dice Game) Suppose that you roll a dice. For each roll, you are paid the face value. Ifa 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?

Ans: Let X be the random variable representing the payoff of this game and let Y be the random variable representing the result of the first roll. For any $i \in \{1, 2, 3, 4, 5, 6\}$, we have $P(Y = i) = \frac{1}{6}$. If $i \in \{1, 2, 3\}$, then the distribution of X given $Y = i$ is the degenerate distribution with $X = i$. If $i \in \{4, 5, 6\}$, then the distribution of the conditional random variable $X|(Y = i)$ coincides with that of $i + X$. In particular, we have $\mathbb{E}[X|Y = i] = i + \mathbb{E}[X]$. Applying the law of total expectation, we have

$$\mathbb{E}[X] = \mathbb{E}\Big[\mathbb{E}[X|Y]\Big]$$
$$= \sum_{i=1}^{6} \frac{1}{6} \mathbb{E}[X|Y = i]$$
$$= \frac{1 + 2 + 3}{6} + \frac{4 + \mathbb{E}[X] + 5 + \mathbb{E}[X] + 6 + \mathbb{E}[X]}{6}$$
$$= \frac{1}{2} \cdot \mathbb{E}[X] + \frac{7}{2},$$

which implies that $\mathbb{E}[X] = 7$.

---

4. (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?

Ans: We consider the general case where there are n ($n \geq 4$) cards in total, which contains 4 aces. Let X_n be the random variable representing the number of cards that need to be turned over to see the first ace and let A be the event that the first card turned over is an ace. It is clear that $P(A) = \frac{4}{n}$, $P(X_4 = 1) = 1$ and $\mathbb{E}[X_4] = 1$.

Now assume that $n \geq 5$. If A occurs, then the random variable $X_n | A$ is the degenerate distribution with $X_n | A \equiv 1$, which implies that $\mathbb{E}[X_n | A] = 1$. If A does not occur, then it is easy to see that the distribution of $X_n | A^c$ coincides with that of $1 + X_{n-1}$. In particular, we have $\mathbb{E}[X_n | A^c] = 1 + \mathbb{E}[X_{n-1}]$. Applying the law of total expectation, we have

$$
\begin{aligned}
\mathbb{E}[X_n] &= P(A) \times \mathbb{E}[X_n | A] + P(A^c) \times \mathbb{E}[X_n | A^c] \\
&= \frac{4}{n} + \frac{n-4}{n}\big(1 + \mathbb{E}[X_{n-1}]\big) \\
&= 1 + \frac{n-4}{n} \cdot \mathbb{E}[X_{n-1}].
\end{aligned}
$$

We have $E[X_5] = 1 + \frac{1}{5} \cdot \mathbb{E}[X_4] = \frac{6}{5}$, $\mathbb{E}[X_6] = 1 + \frac{2}{6} \cdot \mathbb{E}[X_5] = \frac{7}{5}$ and $\mathbb{E}(X_7) = 1 + \frac{3}{7} \cdot \mathbb{E}[X_6] = \frac{8}{5}$. By induction, we can show that $\mathbb{E}[X_n] = \frac{n+1}{5}$ for any $n \geq 4$. In particular, we have $\mathbb{E}[X_{52}] = \frac{53}{5}$.

Let's give a different approach.

Ans: There are 4 aces and 48 other cards. Let's label the other 48 cards as $1, 2, \cdots, 48$. For any $1 \leq i \leq 48$, let X_i be the random variable given by

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

Then the total number of cards that need to turned over in order to see the first ace is

$$
X = 1 + \sum_{i=1}^{48} X_i,
$$

which implies that

$$
\mathbb{E}[X] = 1 + \sum_{i=1}^{48} \mathbb{E}[X_i].
$$

By symmetry, these $\mathbb{E}[X_i]$'s are all the same. As shown in the following sequence, card 1 is equally likely to be in one of the five regions separated by 4 aces:

$$
1 \text{ A } 2 \text{ A } 3 \text{ A } 4 \text{ A } 5
$$

Hence the probability that card 1 appears before all 4 aces is $\frac{1}{5}$, which implies that $\mathbb{E}[X_1] = \frac{1}{5}$. It follows that

$$E[X] = 1 + \sum_{i=1}^{48} \mathbb{E}[X_i] = 1 + \frac{48}{5} = \frac{53}{5}.$$

Remark 43 *This is just a special case for random ordering of m ordinary cards and n special cards. The expected position of the first special card is*

$$1 + \sum_{i=1}^{m} \mathbb{E}[X_i] = 1 + \frac{m}{n+1}.$$

Remark 44 *If the above explanation of the fact $P(X_1 = 1) = \frac{1}{5}$ is not clear to you, you can consider the following more rigorous proof. We randomly arrange the 52 cards in a row, and we want card 1 to be positioned before all 4 aces. There are 52! ways to arrange the 52 cards. To ensure that card 1 is placed before all four aces, we first determine the position of card 1. Since we need to leave at least four spaces for the aces at the end, there are 48 possible positions for card 1. If card 1 is placed in the i-th position, then there are $52 - i$ positions remaining after it. We then choose four of these positions to place the four aces, which can be done in A_{52-i}^4 ways. After placing card 1 and the four aces, the remaining 47 cards can be arranged freely in the remaining 47 positions. Hence the total number of ways to arrange the 52 cards such that card 1 is placed before all four aces equals*

$$47! \cdot \sum_{i=1}^{48} A_{52-i}^4 = 47! \cdot \sum_{i=4}^{51} A_i^4 = 47! \cdot 4! \cdot \sum_{i=4}^{51} \binom{i}{4}.$$

Applying the Hockey Stick Identity, we have

$$\sum_{i=4}^{51} \binom{i}{4} = \binom{52}{4},$$

which implies that the total number of ways to arrange the 52 cards such that card 1 is placed before all four aces equals

$$47! \cdot 4! \cdot \binom{52}{4} = \frac{52!}{5}.$$

Hence the probability that card 1 is placed before all four aces is $\frac{1}{5}$.

Remark 45 *The Hockey Stick Identity states that for any $n \geq k \geq 0$, we have*

$$\sum_{i=k}^{n} \binom{i}{k} = \binom{n+1}{k+1}.$$

Here is a combinatorial interpretation: imagine that you want to choose $k+1$ objects from a set of $n+1$ items. The identity shows that you can count this in another way: consider the largest item selected and suppose it is an position $i+1$ (where i ranges from k to n). Then you need to choose the remaining k objects from the first i items, which can be done in $\binom{i}{k}$ ways. Summing over all possible positions i gives the total number of ways, which is exactly $\binom{n+1}{k+1}$.

Remark 46 *We can simplify the above argument as follows. First, we choose 5 positions from the 52 available positions, which will be occupied by these 5 key cards. The number of ways to do this is $\binom{52}{5}$. Among these 5 positions, card 1 must be placed before all four aces, meaning it must occupy the leftmost (smallest index) position among these 5 chosen positions. The four aces can be arranged freely in the remaining 4 positions, and the number of ways to arrange them is $4!$. The remaining 47 cards can be freely arranged in the remaining 47 positions, and the number of ways to do this is $47!$. Thus, the total number of favorable arrangements is*

$$\binom{52}{5} \cdot 4! \cdot 47! = \frac{52!}{5},$$

Remark 47 *We have another much simpler interpretation of the fact $P(X_1 = 1) = \frac{1}{5}$. Still, we want to randomly arrange the 52 cards. We can first arrange the remaining 47 cards in a sequence. Then, we arrange card 1 and the four aces separately. Next, we insert these five cards into the sequence of 47 cards while preserving their original order. Clearly, any arrangement of the 52 cards can be uniquely obtained through this method. Note that the first step of arranging the 47 cards and the final step of inserting the five cards while preserving their order do not alter the relative order of these five cards. Therefore, we know that in all possible arrangements of the 52 cards, the proportion of arrangements where card 1 appears before the four aces is equal to the proportion of arrangements where card 1 appears before the four aces in the separate arrangement of the five cards. The latter is clearly $\frac{1}{5}$.*

Remark 48 *proportion /prəˈpɔːʃən/ n. 份额，比例，重要性.*

5. (Sum of Random Variables) Assume that X_1, X_2, $\cdots$, X_n are n independent and identically distributed random variables with uniform distribution between 0 and 1. What is the probability that $S_n = X_1 + X_2 + \cdots + X_n \leq 1$?

Ans: It is clear that $P(S_n \leq 1)$ equals the n-dimensional volume of the region

$$\Omega_n = \{(x_1, \cdots, x_n) \in \mathbb{R}^n : 0 \leq x_i \leq 1, \ x_1 + x_2 + \cdots + x_n \leq 1\}.$$

Note that if $x_i \geq 0$ for any $1 \leq i \leq n$, then the condition that $x_i \leq 1$ in the definition of Ω_n is redundant. It is clear that $V(\Omega_1) = 1$. Now assume that $n \geq 2$. We have

$$V(\Omega_n) = \int_{\Omega_n} \mathrm{d}x_1 \cdots \mathrm{d}x_n$$

$$= \int_0^1 \left(\int_{\substack{x_2, \cdots, x_n \geq 0 \\ x_2 + \cdots + x_n \leq 1 - x_1}} \mathrm{d}x_2 \cdots \mathrm{d}x_n \right) \mathrm{d}x_1.$$

By making the following variable substitution

$$x_2 = (1 - x_1)y_2, \quad x_3 = (1 - x_1)y_3, \quad \cdots, \quad x_n = (1 - x_1)y_n,$$

we obtain that

$$V(\Omega_n) = \int_0^1 \left(\int_{\substack{y_2, \cdots, y_n \geq 0 \\ y_2 + \cdots + y_n \leq 1}} (1 - x_1)^{n-1} \, \mathrm{d}y_2 \cdots \mathrm{d}y_n \right) \mathrm{d}x_1$$

$$= \int_0^1 (1 - x_1)^{n-1} \left(\int_{\substack{y_2, \cdots, y_n \geq 0 \\ y_2 + \cdots + y_n \leq 1}} \mathrm{d}y_2 \cdots \mathrm{d}y_n \right) \mathrm{d}x_1$$

$$= \int_0^1 (1 - x_1)^{n-1} \cdot V(\Omega_{n-1}) \, \mathrm{d}x_1$$

$$= V(\Omega_{n-1}) \int_0^1 (1 - x_1)^{n-1} \, \mathrm{d}x_1$$

$$= V(\Omega_{n-1}) \times \frac{-1}{n} \cdot (1 - x)^n \Big|_0^1$$

$$= V(\Omega_{n-1}) \times \frac{1}{n}.$$

It follows that $V(\Omega_n) = \frac{1}{n!}$ for any $n \geq 1$. Hence the probability that $S_n \leq 1$ is $\frac{1}{n!}$.

6. (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 coupons?

Ans:

(a) For any $1 \leq i \leq N$, let X_i be the number of additional coupons needed to obtain the i-th type after $(i-1)$ distinct types have been collected. Hence the total number of coupons needed is

$$X = X_1 + X_2 + \cdots + X_N.$$

For any $1 \leq i \leq N$, if $i-1$ distinct types of coupons have already been collected, then a new coupon is of a different type with probability

$$\frac{1-(i-1)}{N} = \frac{N-i+1}{N}.$$

It is not difficult to see that the random variable X_i follows a geometric distribution with $p = \frac{N-i+1}{N}$, which implies that

$$\mathbb{E}[X_i] = \frac{N}{N-i+1}.$$

Hence

$$\mathbb{E}[X] = \sum_{i=1}^{N} \mathbb{E}[X_i] = \sum_{i=1}^{N} \frac{N}{N-i+1} = N(1 + \frac{1}{2} + \cdots + \frac{1}{N}).$$

(b) For any $1 \leq i \leq N$, let Y_i be the random variable given by

$$Y_i = \begin{cases} 1 & \text{if the } i\text{-th type of coupon has been collected,} \\ 0 & \text{otherwise.} \end{cases}$$

Then the total number of distinct coupons collected is

$$Y = Y_1 + Y_2 + \cdots + Y_N,$$

By symmetry, we have $\mathbb{E}[Y_1] = \cdots = \mathbb{E}[Y_N]$. For each collected coupon, the probability that it is not of the i-th type is $\frac{N-1}{N}$. Since all the n coupons are independent, the probability that none of the n coupons is of the i-th type is $(\frac{N-1}{N})^n$. It follows that

$$\mathbb{E}[Y_1] = P(Y_1 = 1) = 1 - P(Y_1 = 0) = 1 - (\frac{N-1}{N})^n.$$

Hence we have

$$\mathbb{E}[Y] = \sum_{i=1}^{N} \mathbb{E}[Y_i] = N \cdot \left(1 - (\frac{N-1}{N})^n\right).$$

Remark 49 *coupon /ˈkuːpɑn/ n.* 赠券，优惠券.

Remark 50 *cereal /ˈsɪriəl/ n.* 谷类食品，谷物植物. *adj.* 谷类的，谷物的.

7. (Join 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?

> **Ans:** Let X and Y be the events that A and B default, respectively. Then $P(X) = 50\%$ and $P(Y) = 30\%$. Since probability is a monotonic increasing function of events, we have
>
> $$P(X \cup Y) \geq \max\big(P(X), P(Y)\big) = 50\%.$$
>
> We can decompose $X \cup Y$ into two following two disjoint events: X and $Y \setminus X$. By the additivity of probability, we have
>
> $$P(X \cup Y) = P(X) + P(Y \setminus X) \leq P(X) + P(Y) = 80\%.$$
>
> Hence the range of probability that at least one bond defaults is $[50\%, 80\%]$.
>
> Let 1_X and 1_Y be the indicator random variables of X and Y, respectively. Then we have $\mathbb{E}[1_X] = 0.5$, $\mathbb{E}[1_B] = 0.3$, $\mathrm{Var}(1_A) = P(X) \times \big(1 - P(X)\big) = 0.25$ and $\mathrm{var}(1_B) = P(Y) \times \big(1 - P(Y)\big) = 0.21$. The correlation coefficient of X and Y is
>
> $$\rho(1_X, 1_Y) = \frac{\mathrm{Cov}(1_X, 1_Y)}{\sqrt{\mathrm{Var}(1_X)\mathrm{Var}(1_Y)}} = \frac{\mathbb{E}[1_X 1_Y] - \mathbb{E}[1_X]\mathbb{E}[1_Y]}{\sqrt{\mathrm{Var}(1_X)\mathrm{Var}(1_Y)}}.$$
>
> By definition, we have $\mathbb{E}[1_X 1_Y] = P(X \cap Y)$. The range of $P(X \cap Y)$ is $\Big[0, \min\big(P(X), P(Y)\big)\Big]$, i.e., $[0, 0.3]$. Hence the range of $\rho(1_X, 1_Y)$ is $[-\sqrt{\frac{3}{7}}, \sqrt{\frac{3}{7}}]$.

Remark 51 *bond /bɑːnd/ n.* 债券.

Remark 52 *default /dɪˈfɔːlt/ v.* 违约.

1.6 Order Statistics

1.6.1 Key Concept Review

Let X_1, X_2, $\cdots$, X_n be n independent and identically distributed random variables with cumulative distribution function $F(x)$. Put $Y = \min(X_1, X_2, \cdots, X_n)$ and $Z = \max(X_1, X_2, \cdots, X_n)$.

For any $x \in \mathbb{R}$, we have

$$\begin{aligned}
P(Y \geq x) &= P(X_1 \geq x, X_2 \geq x, \cdots, X_n \geq x) \\
&= P(X_1 \geq x)P(X_2 \geq x) \cdots P(X_n \geq x) \\
&= \left(1 - F(x)\right)^n,
\end{aligned}$$

which implies that cumulative distribution function of Y is

$$1 - \left(1 - F(x)\right)^n.$$

If F is differentiable, then the probability density function of Y is

$$n\left(1 - F(x)\right)^{n-1} F'(x).$$

Similarly, for any $x \in \mathbb{R}$, we have

$$\begin{aligned}
P(Z < x) &= P(X_1 < x, X_2 < x, \cdots, X_n < x) \\
&= P(X_1 < x)P(X_2 < x) \cdots P(X_n < x) \\
&= F(x)^n,
\end{aligned}$$

which implies that the cumulative distribution function of Z is

$$F(x)^n.$$

If F is differentiable, then the probability density function of Z is

$$nF(x)^{n-1} F'(x).$$

1.6.2 Exercises

1. (Expected Value of Max and Min) Let $X_1, X_2, \cdots, X_n$ 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 $Z_n = \max(X_1, \cdots, X_n)$? What are the cumulative distribution function, the probability density function and expected value of $Y_n = \min(X_1, \cdots, X_n)$?

Ans: Let $F(x)$, $p(x)$ denote the cumulative distribution function and probability density function, respectively, of a uniform distribution over the interval $[0, 1]$. Then

$$p(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

---

and
$$F(x) = \begin{cases} 0 & \text{if } x < 0, \\ x & \text{if } 0 \le x \le 1, \\ 1 & \text{if } x > 1. \end{cases}$$

Then the cumulative distribution function of Z_n is
$$F_Z(x) = F(x)^n = \begin{cases} 0 & \text{if } x < 0, \\ x^n & \text{if } 0 \le x \le 1, \\ 1 & \text{if } x > 1, \end{cases}$$

and the probability density function of Z_n is
$$p_Z(x) = nF(x)^{n-1}p(x) = \begin{cases} 0 & \text{if } x < 0, \\ nx^{n-1} & \text{if } 0 \le x \le 1, \\ 0 & \text{if } x > 1. \end{cases}$$

Hence the expected value of Z_n is
$$\mathbb{E}(Z_n) = \int_0^1 xp_Z(x)\,\mathrm{d}x = \int_0^1 nx^n\,\mathrm{d}x = \frac{n}{n+1}.$$

Moreover, the cumulative distribution function of Y_n is
$$F_Y(x) = 1 - \big(1 - F(x)\big)^n = \begin{cases} 0 & \text{if } x < 0, \\ 1 - (1-x)^n & \text{if } 0 \le x \le 1, \\ 1 & \text{if } x > 1, \end{cases}$$

and the probability density function of Y_n is
$$p_Y(x) = \begin{cases} 0 & \text{if } x < 0, \\ n(1-x)^{n-1} & \text{if } 0 \le x \le 1, \\ 0 & \text{if } x > 1. \end{cases}$$

Hence the expected value of Y_n is
$$\mathbb{E}(Y_n) = \int_0^1 xp_Y(x)\,\mathrm{d}x = \int_0^1 nx(1-x)^{n-1}\,\mathrm{d}x = \frac{1}{n+1}.$$

2. (Correlation of Max and Min) Let X_1 and X_2 be IID random variables with uniform

distributions 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 coefficient of Y and Z?

Ans: Let $F(x)$ and $p(x)$ be the cumulative distribution function and probability density function, respectively, of a uniform distribution on $[0, 1]$. Then we have

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

and

$$p(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

For any $y, z \in [0, 1]$, we have

$$\begin{aligned} P(Y \geq y | z \leq z) &= \frac{P(Y \geq y, Z \leq z)}{P(Z \leq z)} \\ &= \frac{P(X_1 \geq y, X_2 \geq y, X_1 \leq z, X_2 \leq z)}{P(X_1 \leq z, X_2 \leq z)} \\ &= \frac{P(y \leq X_1 \leq z, y \leq X_2 \leq z)}{P(X_1 \leq z, X_2 \leq z)} \\ &= \frac{P(y \leq X_1 \leq z)P(y \leq X_2 \leq z)}{P(X_1 \leq z)P(X_2 \leq z)} \\ &= \begin{cases} 0 & \text{if } z \leq y, \\ \frac{(z-y)^2}{z^2} & \text{if } z > y. \end{cases} \end{aligned}$$

It follows that for any $y, z \in [0, 1]$, we have

$$\begin{aligned} P(Y \leq y, Z \leq z) &= P(Z \leq z)P(Y \leq y | Z \leq z) \\ &= P(Z \leq z)\left(1 - P(Y \geq y | Z \leq z)\right) \\ &= \begin{cases} z^2 & \text{if } z \leq y, \\ (2z - y)y & \text{if } z > y. \end{cases} \end{aligned}$$

It follows that the joint density function of Y and Z is

$$p(y, z) = \frac{\partial^2 P(Y \leq y, Z \leq z)}{\partial z \partial y} = \begin{cases} 2 & \text{if } 0 \leq y < z \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Page 43

The cumulative distribution function of Y is

$$F_Y(y) = 1 - \big(1 - F(y)\big)^2 = \begin{cases} 0 & \text{if } y < 0, \\ 1 - (1 - y)^2 & \text{if } 0 \le y \le 1, \\ 1 & \text{if } y > 1, \end{cases}$$

and the probability density function of Y is

$$p_Y(y) = 2(1 - y)p(y) = \begin{cases} 0 & \text{if } y < 0, \\ 2(1 - y) & \text{if } 0 \le y \le 1, \\ 0 & \text{if } y > 1. \end{cases}$$

Then the cumulative distribution function of Z is

$$F_Z(z) = F(z)^2 = \begin{cases} 0 & \text{if } z < 0, \\ z^2 & \text{if } 0 \le z \le 1, \\ 1 & \text{if } z > 1, \end{cases}$$

and the probability density function of Z is

$$p_Z(z) = 2zp(z) = \begin{cases} 0 & \text{if } z < 0, \\ 2z & \text{if } 0 \le z \le 1, \\ 0 & \text{if } z > 1. \end{cases}$$

Hence the expected value of Y is

$$\mathbb{E}(Y) = \int_0^1 y p_Y(y) \, \mathrm{d}y = \int_0^1 2y(1 - y) \, \mathrm{d}y = \frac{1}{3},$$

the expected value of Z is

$$\mathbb{E}(Z) = \int_0^1 z p_Z(z) \, \mathrm{d}z = \int_0^1 2z^2 \, \mathrm{d}z = \frac{2}{3},$$

the variance of Y is

$$\mathrm{Var}(Y) = \int_0^1 y^2 p_Y(y) \, \mathrm{d}y - \mathbb{E}(Y)^2 = \int_0^1 2y^2(1 - y) \, \mathrm{d}y - \frac{1}{9} = \frac{1}{18},$$

and

$$\mathrm{Var}(Z) = \int_0^1 z^2 p_Z(z) \, \mathrm{d}z - \mathbb{E}(Z)^2 = \int_0^1 2z^3 \, \mathrm{d}z - \frac{4}{9} = \frac{1}{18}.$$

Finally, we need to compute the covariance of Y and Z. Since

$$\mathrm{Cov}(X, Y) = \mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z],$$

it suffices to compute $\mathbb{E}[YZ]$. We have

$$\begin{aligned}
\mathbb{E}[YZ] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yz\, p(y, z)\, \mathrm{d}y\, \mathrm{d}z \\
&= \int_0^1 \left(\int_0^z yz\, p(y, z)\, \mathrm{d}y \right) \mathrm{d}z \\
&= 2 \int_0^1 \left(\int_z^1 yz\, \mathrm{d}y \right) \mathrm{d}z = \frac{1}{4},
\end{aligned}$$

which implies that

$$\mathrm{Cov}(X, Y) = \frac{1}{4} - \frac{1}{3} \cdot \frac{2}{3} = \frac{1}{36}$$

and

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{1}{2}.$$

Remark 53 *There is an alternative and simpler approach to calculate $\mathbb{E}[YZ]$. Indeed, we always have $YZ = X_1 X_2$. Hence YZ and $X_1 X_2$ follow the same distribution. It follows that*

$$\mathbb{E}[YZ] = \mathbb{E}[X_1 X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2] = \frac{1}{4}$$

since X_1 and X_2 are independent.

Remark 54 *There is another way to calculate the joint density function of Y and Z. Indeed, for any $y, z \in \mathbb{R}$, we have*

$$P(Y \geq y, Z \leq z) = P(Y \geq y, Z \leq z, X_1 < X_2) + P(Y \geq y, Z \leq z, X_1 > X_2).$$

By symmetry, we must have

$$P(Y \geq y, Z \leq z, X_1 < X_2) = P(Y \geq y, Z \leq z, X_1 > X_2),$$

which implies that

$$\begin{aligned}
P(Y \geq y, Z \leq z) &= 2 \cdot P(Y \geq y, Z \leq z, X_1 < X_2) \\
&= 2 \cdot P(X_1 \geq y, X_2 \leq z, X_1 < X_2) \\
&= 2 \cdot P(y \leq X_1 < X_2 \leq z).
\end{aligned}$$

Remark 55 *That Y and Z have positive correlation coefficient makes sense, since when Y becomes large, Z tends to becomes as well (note that $Z \geq Y$).*

3. (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?

Ans: The difficulty of this problem lies in the fact that when two ants collide, they need to change their directions. However, we can use a clever transformation to avoid this difficulty.

Please refer to the above diagram. When an ant A collides with another ant B, both of them switch direction. But if we exchange the ants' labels, it is as if the collision never happened: A continues to move to the right and B continues to move to the left. Moreover, after making this equivalent transformation, the total time required remains unchanged, since we are considering the time until the last ant falls off. Hence we can remove the collision rule and consider the ants moving independently.

Now let T_1 denote the time for the first ant to fall off the string, let D denote its initial direction and let x_0 be the coordinate of its initial position. Then for any $t \in \mathbb{R}$,

1. if $t \leq 0$, then $P(T_1 < t) = 0$;

2. if $t \geq 1$, then $P(T_1 < t) = 1$;

3. if $0 < t < 1$, then

$$
\begin{aligned}
P(T_1 < t) &= P(D = \text{left}) \cdot P(T_1 < t | D = \text{left}) \\
&\quad + P(D = \text{right}) \cdot P(T_1 < t | D = \text{right}) \\
&= \frac{1}{2} \cdot P(x_0 < t) + \frac{1}{2} \cdot P(x_0 > 1 - t) \\
&= \frac{t}{2} + \frac{t}{2} = t.
\end{aligned}
$$

Hence T_1 also follows that uniform distribution on the interval $[0, 1]$ and the problem reduces to calculating the expected value of the maximum of 500 independent

---

and identically distributed random variables with uniform distribution on $[0, 1]$. We know that the probability density function of the maximum is

$$p(x) = 500x^{499}.$$

Hence the expected value of the maximum equals

$$\int_0^1 xp(x)\,\mathrm{d}x = \int_0^1 x \cdot 500x^{499}\,\mathrm{d}x = \frac{500}{501}.$$

Remark 56 *foot n. 英尺.*

Remark 57 *collide /kəˈlaɪd/ v. 碰撞.*

Remark 58 *collision /kəˈlɪʒn/ n. 碰撞.*

2 Stochastic Process and Stochastic Calculus

2.1 Markov Chain

2.1.1 Key Concept Review

1. A **Markov** (/ˈmɑːrkɔːv/) **chain** with state space S is a sequence of infinitely many random variables $X = (X_n)_{n \geq 0}$ taking values in S with the Markov property that given the present state, the future states and the past states are independent:

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) = P(X_{n+1} = j | X_n = i) \quad (6)$$

for any $n \in \mathbb{N}$ and any $i_0, i_1, \cdots, i_{n-1}, i, j \in S$.

2. If the probability in equation (6) does not depend on n, then the Markov chain is said to be **homogeneous**. For a homogeneous Markov chain, the **transition matrix** P is defined by $(p_{ij})_{i,j \in S}$, where $p_{ij} = P(X_{n+1} = j | X_n = i)$. It is clear that a homogeneous Markov chain is completely determined by its **initial distribution** π_0 (the distribution of X_0) and its transition matrix P.

3. The transition matrix P satisfies the following properties:

   (a) $p_{ij} \geq 0$ for any $i, j \in S$;

   (b) $\sum_{j \in S} p_{ij} = 1$ for any $i \in S$, which implies that the rows of P sum to 1.

   Any matrix satisfying these two properties is called a **stochastic matrix**.

4. Applying the Markov property, we can calculate **the probability of any path**:

$$P(X_1 = i_1, X_2 = i_2, \cdots, X_n = i_n | X_0 = i_0) = p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}$$

for any $i_0, i_1, i_2, \cdots, i_n \in S$.

5. For any $n \geq 0$ and $i, j \in S$, put $p_{ij}(n) = P(X_n = j | X_0 = i)$. Then we have $P^n = (p_{ij}(n))$.

6. For any stochastic matrix, we can use the **transition graph** to express it graphically. The vertices of the graph are the states of the Markov chain and there is a directed edge from state i to state j if $p_{ij} > 0$. The weight of the edge from i to j is p_{ij}.

7. A state $j \in S$ is called **accessible** from state i if there is a direction path in the transition graph from i to j, i.e., if there exists $n \geq 0$ such that $p_{ij}(n) > 0$. Two states $i, , j \in S$ called said to **communicate** if they are accessible from each other. It is clear that communication is an equivalence relation on S. If any two states in S communicate, then the Markov chain is said to be **irreducible**.

8. A state $i \in S$ is called **recurrent** if every state j that is accessible from i, i is also accessible from j. A state $i \in S$ is called **transient** (/ˈtrænzɪənt/, adj. 转瞬即逝的，短暂的；暂住的.) if it is not recurrent. It is clear that recurrence and transience are communication classes.

9. For any state $s \in S$, consider the probability of ever returning to s:

$$f_{ss} = P(X_n = s \text{ for some } n \geq 1 | X_0 = s).$$

Then

$$f_{ss} \begin{cases} = 1 & \text{if } s \text{ is recurrent,} \\ < 1 & \text{if } s \text{ is transient.} \end{cases}$$

Indeed, assume that s is transient, then there exists $j \in S$ such that $p_{sj}(n_0) > 0$ for some $n_0 \in \mathbb{N}_+$ but $p_{js}(n) = 0$ for any $n \in \mathbb{N}_+$. We may assume that n_0 is the smallest such integer and claim that for any $n \geq 1$, $P(X_n = s, X_{n_0} = j | X_0 = s) = 0$. Otherwise, there exists $n_1 \geq 1$ such that $P(X_{n_1} = s, X_{n_0} = j | X_0 = s) > 0$. If $n_1 = n_0$, then $P(X_{n_0} = j | X_{n_0} = s) > 0$, which is impossible. If $n_1 < n_0$, then $P(X_{n_0} = j | X_{n_1} = s) > 0$. Using the homogeneity, we have $P_{sj}(n_0 - n_1) > 0$, which contradicts the minimality of n_0. If $n_1 > n_0$, then $P(X_{n_1} = s | X_{n_0} = j) > 0$. Using the homogeneity, we have $p_{js}(n_1 - n_0) > 0$, which contradicts the assumption on j. Hence for any $n \geq 1$, $P(X_n = s, X_{n_0} = j | X_0 = s) < 1$, which implies that

$$P(X_n = s \text{ for some } n \geq 1, X_{n_0} = j | X_0 = s) = 0.$$

However, by assumption, we have $P(X_{n_0} = j|X_0 = s) = p_{sj}(n_0) > 0$, which implies that
$$P(X_n \neq s \text{ for any } n \geq 1, X_{n_0} = j|X_0 = s) > 0.$$
It follows that $P(X_n \neq s \text{ for any } n \geq 1|X_0 = s) > 0$, which implies that
$$f_{ss} = P(X_n = s \text{ for some } n \geq 1|X_0 = s) < 1.$$

Conversely, assume that
$$f_{ss} = P(X_n = s \text{ for some } n \geq 1|X_0 = s) < 1.$$

Then
$$P(X_n \neq s \text{ for any } n \geq 1|X_0 = s) > 0.$$

We have
$$\{X_n \neq s \text{ for any } n \geq 1\} = \bigcup_{j \in S \setminus \{s\}} \{X_n \neq s \text{ for any } n \geq 1, \ X_1 = j\}.$$

It follows that there exists $j \in S \setminus \{s\}$ such that
$$P(X_n \neq s \text{ for any } n \geq 1, \ X_1 = j|X_0 = s) > 0.$$

In particular, we have $p_{sj} = P(X_1 = j|X_0 = s) > 0$.

10. A state $i \in S$ is called **absorbing** (/əbˈzɔːbɪŋ/, v. 吸收的. adj. 引人入胜的.) if it is possible to leave this state, i.e., $p_{ii} = 1$ and $p_{ij} = 0$ for any $j \neq i$.

11. (**Equations for absorption probability**) Assume that the state space is $S = \{1, \cdots, M\}$. For any state $i \in S$ that is absorbing or transient, let a_i denote the probability of reaching an absorbing state s starting from state i. Then (a_i) is the unique solution to the following system of equations:
$$\begin{cases} a_s = 1, \\ a_i = 0 \text{ for any absorbing state } i \neq s, \\ a_i = \sum_{j=1}^{M} a_j p_{ij} \text{ for any transient state } i. \end{cases}$$

Why we don't consider the probability of absorption from recurrent states? Because a recurrent state will never visit an absorbing state except itself.

12. (Equations for the expected time to absorption) For any state $i \in S$ that is absorbing or transient, let μ_i be the expected time to absorb from state i. Then (μ_i) is the unique solution to the following system of equations:
$$\begin{cases} \mu_i = 0, \text{ for any absorbing state } i, \\ \mu_i = 1 + \sum_{j=1}^{M} \mu_j p_{ij}, \text{ for any transient state } i. \end{cases} \tag{7}$$

2.1.2 Exercises

1. (Gamblers' Ruin Game) Player M has \$1 and player N has \$2. Each game gives the winner \$1 from the other. As a better player, M wins $\frac{2}{3}$ of the games. They play until one of them is bankrupt. What is the probability that M wins?

Ans: We consider the general case where player M begins with m dollars and player N begins with n dollars. Let $p(m, n)$ denote the probability that player M wins given that player M has m dollars and player N has n dollars. Then it is clear that $p(m, 0) = 1$ and $p(0, n) = 0$ for any $m, n \in \mathbb{N}_+$.

Now assume that $m, n \geq 1$. Consider the first round. Player M wins with probability $\frac{2}{3}$ and loses with probability $\frac{1}{3}$. If player M wins the first round, then player M has $m + 1$ dollars and player N has $n - 1$ dollars. different rounds are independent., we may regard the game as restarting with M having $m + 1$ dollars and N having $n - 1$ dollars. Hence given that M wins the first round, the probability that M wins at last is $p(m + 1, n - 1)$. Similarly, given that M loses the first round, the probability that M wins at last is $p(m - 1, n + 1)$. Applying the law of total probability, we have

$$p(m, n) = \frac{2}{3} p(m + 1, n - 1) + \frac{1}{3} p(m - 1, n + 1).$$

It follows that

$$p(1, 1) = \frac{2}{3} p(2, 0) + \frac{1}{3} p(0, 2) = \frac{2}{3},$$

and

$$\begin{cases} p(2, 1) = \frac{2}{3} p(3, 0) + \frac{1}{3} p(1, 2) = \frac{2}{3} + \frac{1}{3} p(1, 2), \\ p(1, 2) = \frac{2}{3} p(2, 1) + \frac{1}{3} p(0, 3) = \frac{2}{3} p(2, 1), \end{cases}$$

Solving the above system of equations, we have $p(2, 1) = \frac{6}{7}$ and $p(1, 2) = \frac{4}{7}$. Hence the probability that M wins is $\frac{4}{7}$.

Remark 59 *Note that the sum of money of the two players is always* 3. *Hence we only need to consider the number of dollars that player M has. In this way, this problem is essentially a one-dimensional random walk with constraints* 0 *and* 3.

Remark 60 *bankrupt /ˈbæŋkrʌpt/ adj. 破产的.*

We give another approach using the Markov chain, which is essentially the same. The most difficult part of Markov chain problems ofter lies in how to choose the right state space and define the transition probabilities p_{ij}'s.

Ans: Define the state space as the possible pairs of dollars that player M and player N have. Then it is clear that the state space is $S = \{(3,0),(2,1),(1,2),(0,3)\}$. Since the sum of the dollars that player M and player N have is always 3, we only need to consider the number of dollars that player M has. Hence the state space can be simplified as $S = \{0,1,2,3\}$. The initial state is $X_0 = 1$, the transition matrix is

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

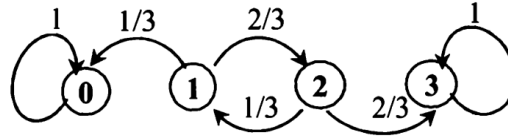and the transition graph is as follows:

Figure 3: Transition graph for Gamblers' ruin problem

We can see that state 3 and state 0 are absorbing states. To calculate the probability that M reaches absorbing state 3, we can apply the absorption probability equations. For any $i \in S$, let a_i be the absorption probability of reaching absorbing state 3 starting from state i. Then we have

$$\begin{cases} a_3 = 1, \\ a_0 = 0, \\ a_1 = \sum_{j=0}^{3} p_{1j}a_j, \\ a_2 = \sum_{j=0}^{3} p_{2,j}a_j, \end{cases} \Leftrightarrow \begin{cases} a_3 = 1, \\ a_0 = 0, \\ a_1 = \frac{1}{3}a_0 + \frac{2}{3}a_2, \\ a_2 = \frac{1}{3}a_1 + \frac{2}{3}a_3. \end{cases}$$

Solving the above system of equations yields $a_1 = \frac{4}{7}$ and $a_2 = \frac{6}{7}$. Hence the probability that M wins is $\frac{4}{7}$.

2. (Dice Question) Two players bet on roll(s) of the total of two standard six-face dice. Player A bets that a sum of 12 will occur first. Player B bets that two consecutive 7s will occur first. The players keep rolling the dice and record the sums until one player wins. What is the probability that A will win?

Ans: Let X be the event that A wins. Conditioning $P(X)$ on the first throw's sum F, we have

$$P(X) = P(X|F = 12)P(F = 12) + P(X|F = 7)P(F = 7)$$
$$+ P\left(X|F \notin \{7, 12\}\right)P\left(F \notin \{7, 12\}\right).$$

Let (a, b) be the result of the first throw. Then

$$P(F = 12) = P(a = 6, b = 6) = \frac{1}{36},$$

$$P(F = 7) = P(a = 1, b = 6) + P(a = 6, b = 1)$$
$$+ P(a = 2, b = 5) + P(a = 5, b = 2)$$
$$+ P(a = 3, b = 4) + P(a = 4, b = 3) = \frac{6}{36},$$

and thus

$$P(F \notin \{7, 12\}) = 1 - \frac{1}{36} - \frac{2}{36} = \frac{29}{36}.$$

Therefore, we have

$$P(X) = \frac{1}{36}P(X|F = 12) + \frac{6}{36}P(X|F = 7) + \frac{29}{36}P\left(X|F \notin \{7, 12\}\right). \quad (8)$$

1. If $F = 12$, then A wins, and thus $P(X|F = 12) = 1$.

2. If $F \notin \{7, 12\}$, then we may view the game as restarting. Hence

$$P\left(X|F \notin \{7, 12\}\right) = P(X).$$

3. If $F = 7$, then we condition $P(X|F = 7)$ on the second throw's sum G. We have $P(G = 12) = P(F = 12) = \frac{1}{36}$, $P(G = 7) = P(F = 7) = \frac{6}{36}$, and

$$P(X|F = 7) = P(X|F = 7, G = 12)P(G = 12)$$
$$+ P(X|F = 7, G = 7)P(G = 7)$$
$$+ P\left(X|F = 7, G \notin \{7, 12\}\right)P\left(G \notin \{7, 12\}\right)$$
$$= \frac{1}{36}P\left(X|F = 7, G = 12\right) + \frac{6}{36}P(X|F = 7, G = 7)$$
$$+ \frac{29}{36}P(X).$$

(a) If $F = 7$ and $G = 12$, then A wins, and thus $P(X|F = 7, G = 12) = 1$.

(b) If $F = 7$ and $G = 7$, then B wins, and thus $P(X|F = 7, G = 7) = 0$.

(c) If $F = 7$ and $G \notin \{7, 12\}$, then we may view the game as restarting. Hence
$$P\Big(X|F = 7, G \notin \{7, 12\}\Big) = P(X).$$

It follows that
$$P(X|F = 7) = \frac{1}{36} + \frac{29}{36}P(X).$$

Substituting all the conditional probabilities into equation (8), we have
$$P(X) = \frac{1}{36} + \frac{6}{36} \cdot \Big(\frac{1}{36} + \frac{29}{36}P(X)\Big) + \frac{29}{36}P(X),$$

which implies that $P(X) = \frac{7}{13}$.

Remark 61 *consecutive /kənˈsekjətɪv/ adj.* 连续的.

We again provide a Markov chain approach. The key part is still to choose the right state space and define the transition probabilities.

Ans: It is clear that we have the following states: 12 (A wins) and $7 - 7$ (B wins), S (starting state) and 7 (one 7 occurs, yet no 12 or $7 - 7$ occurred). The transition matrix (the row and column indices are sequentially: 12, $7 - 7$, S and 7) is
$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{1}{36} & 0 & \frac{29}{36} & \frac{6}{36} \\ \frac{1}{36} & \frac{6}{36} & \frac{29}{36} & 0 \end{pmatrix},$$
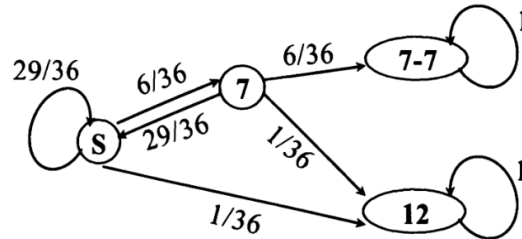and the transition graph is as follows:

Figure 4: Transition graph for Dice question

It is clear that 12 and $7-7$ are absorbing states, S and 7 are transient states. To calculate the probability that A wins, we can apply the absorption probability equations. For any state i, let a_i be the absorption probability of reaching state 12 starting from state i. Then we have

$$\begin{cases} a_{12} = 1, \\ a_{7-7} = 0, \\ a_S = \sum_j p_{Sj} a_j = \frac{1}{36} a_{12} + \frac{29}{36} a_s + \frac{6}{36} a_7, \\ a_7 = \sum_j p_{7j} a_j = \frac{1}{36} a_{12} + \frac{6}{36} a_{7-7} + \frac{29}{36} a_S. \end{cases}$$

Solving the above system of equations yields that $a_S = \frac{7}{13}$.

3. (Coin Triplets)

   (a) If you keep on tossing a fair coin, what is the expected number of tosses such that you can have HHH (heads heads heads) in a row? What is the expected number of tosses to have THH (tails heads heads) in a row?

   (b) Keep flipping a fair coin util either HHH or THH occurs in the sequence. What is the probability that you get an HHH subsequence before THH?

   (c) Let's add more fun to the triplet game. Instead of fixed triplets for the two players, the new game allows both to choose their own triplets. Player 1 chooses a triplet first and announces it; then player 2 chooses a different triplet. The players again toss the coins until one of the two triplet sequences appears. The player whose chosen triplet appears first wins the game. If both player 1 and player 2 are perfectly rational and both want to maximize their probability of winning, would you go first (as player 1)? If you go second, what is your probability of winning?

   Ans:

   (a) We use the Markov chain approach. We first consider the question on the expected number of tosses to have HHH in a row. The state space is
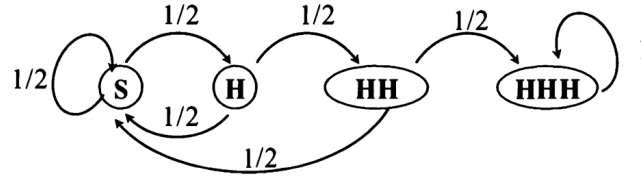
   $$\Omega = \{S, H, HH, HHH\},$$

   where S is the starting state, H is the state that one head occurs, HH is the state that only two consecutive heads occur, and HHH is the state that three consecutive heads occur.

If we are in the starting state S, the tossing a coin will result in heads with probability $\frac{1}{2}$ and tails with probability $\frac{1}{2}$. The first outcome corresponds to state H while the second outcome corresponds to state S since we can regard the game as restarting. Following a similar analysis, we can obtain the transition matrix, which is (the row and columns indices are sequentially: S, H, HH and HHH)

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The corresponding transition graph is as follows:

For any state $i \in \Omega$, let μ_i denote the expected number of tosses to reach state HHH starting from state i. Applying the equations for the expected time to absorption, we have

$$\begin{cases} \mu_{HHH} = 0, \\ \mu_{HH} = 1 + \frac{1}{2}\mu_S + \frac{1}{2}\mu_{HHH}, \\ \mu_H = 1 + \frac{1}{2}\mu_S + \frac{1}{2}\mu_{HH}, \\ \mu_S = 1 + \frac{1}{2}\mu_S + \frac{1}{2}\mu_H. \end{cases}$$

Solving the above equation yields $\mu_S = 14$. Hence the expected number of tosses to have HHH in a row is 14.
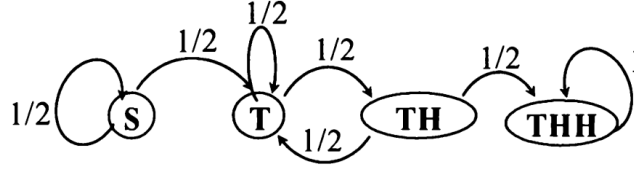
Next, we consider the question on the expected number of tosses to have THH in a row. The state space is

$$\Omega = \{S, T, TH, THH\},$$

and the transition matrix is (the row and column indices are sequentially: S, T, TH and THH)

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The transition graph is as follows:

It is clear that THH is the absorbing state. For any $i \in \Omega$, let μ_i be the expected number of tosses to reach state THH starting from state i. Applying the equations for the expected time to absorption, we have
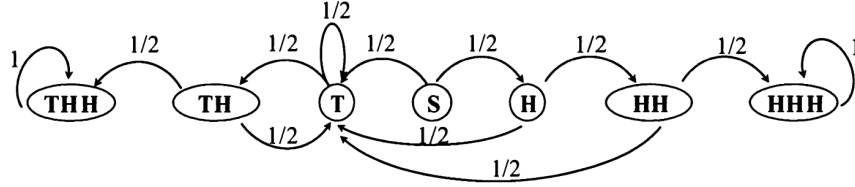
$$\begin{cases} \mu_{THH} = 0, \\ \mu_{TH} = 1 + \frac{1}{2}\mu_T + \frac{1}{2}\mu_{THH}, \\ \mu_T = 1 + \frac{1}{2}\mu_T + \frac{1}{2}\mu_{TH}, \\ \mu_S = 1 + \frac{1}{2}\mu_S + \frac{1}{2}\mu_T. \end{cases}$$

Solving the above system of equations yields $\mu_S = 8$. Hence the expected number of tosses to have THH is 8.

(b) The state space is much complicated now, which is

$$\Omega = \{S, H, HH, HHH, T, TH, THH\}$$

and the transition graph is as follows:

For any state $i \in \Omega$, let a_i be the probability of reaching state HHH before reaching state THH starting from state i. Applying the law of total probability, we have

$$\begin{cases} a_{HHH} = 1, \\ a_{THH} = 0, \\ a_S = \frac{1}{2}a_H + \frac{1}{2}a_T, \\ a_H = \frac{1}{2}a_T + \frac{1}{2}a_{HH}, \\ a_{HH} = \frac{1}{2}a_T + \frac{1}{2}a_{HHH}, \\ a_T = \frac{1}{2}a_T + \frac{1}{2}a_{TH}, \\ a_{TH} = \frac{1}{2}a_T + \frac{1}{2}a_{THH}. \end{cases}$$

Solving the above system of equations yields

$$\begin{cases} a_S = \frac{1}{8}, \\ a_H = \frac{1}{4}, \\ a_T = 0, \\ a_{HH} = \frac{1}{2}, \\ a_{TH} = 0. \end{cases}$$

Hence the probability that we get an HHH subsequence before THH is $\frac{1}{8}$.

The computational cost of solving this system of equations is relatively high. However, we should have noticed that $a_T = 0$. Indeed, once a tail occurs, we will always get THH before HHH, since we need two H to have THH but three H to have HHH. A similar argument applies to a_{TH}.

(c) A common misconception is that there is always a best sequence that beats other sequences. This misconception is often founded on a wrong assumption that these sequences are transitive: if sequence A has a higher probability occurring before sequence B and sequence B has a higher probability occurring before sequence C, then sequence A has a higher probability occurring before C. In reality, such transitivity does not exist for this game.

Let's explain why. Let T_A, T_B, T_C be the random variables representing the number of tosses to have sequence A, B and C for the first time, respectively, and assume that $P(T_A < T_B) > \frac{1}{2}$ and $P(T_B < T_C) > \frac{1}{2}$. Then we cannot conclude that $P(T_A < T_C) > \frac{1}{2}$.

The key point to solve this problem is that no matter what sequence player 1 chooses, player 2 can always choose another sequence with a probability of winning higher than $\frac{1}{2}$. Please refer to the following table:

2's winning Probability		Player 1							
		HHH	THH	HTH	HHT	TTH	THT	HTT	TTT
Player 2	HHH	/	1/8	2/5	1/2	3/10	5/12	2/5	1/2
	THH	**7/8**	/	1/2	**3/4**	1/3	1/2	1/2	3/5
	HTH	3/5	1/2	/	1/3	3/8	1/2	1/2	7/12
	HHT	1/2	1/4	**2/3**	/	1/2	5/8	**2/3**	7/10
	TTH	7/10	**2/3**	5/8	1/2	/	**2/3**	1/4	1/2
	THT	7/12	1/2	1/2	3/8	1/3	/	1/2	3/5
	HTT	3/5	1/2	1/2	1/3	**3/4**	1/2	/	**7/8**
	TTT	1/2	2/5	5/12	3/10	1/2	2/5	1/8	/

Figure 5: Player 2's winning probability with different coin sequence pairs

As shown in the above table, no matter what triplet player 1 chooses, player 2 can always choose another sequence to have better odds of winning. The best sequences that player 2 can choose in response to player 1's choices are highlighted in bold. In order the maximize his odds of winning, player 1 should choose among HTH, HTT, THH and THT. Even in these cases, player 2 wins with a probability of $\frac{2}{3}$.

Remark 62 *triplet /ˈtrɪplət/ n.* 三胞胎.