ELEMENTARY EXAMPLES OF PROBABILITY

CHI-YU CHENG

"It stood there with secrets and compartments, with possible treasuries and traps; it might have a great deal to give, but would probably ask for equal services in return, and would collect this debt to the last shilling."

— Henry James, *Preface to The Wings of the Dove*

Contents

1. Introduction

The preparation of this note was initiated back in March 2024, when the author was transitioning as a math educator at the Citadel in the USA to the financial industry. It is the author's intention to invest the time to maintain a stable, organized, and reliable record of the solutions to some of the problems he thought about.

The note does not require measure theory. The references of this note are the Society of Actuary Probability Exam at [oA], the quantitative finance interview book [Zho20], and the textbook [Ros14]. The note is not intended to replace any textbook on probability. Rather, it provides an organized review of some of the interesting problems from the three sources cited.

The topics discussed include the probability and expectation from poker, tournament matches, coupon collections, coin toss sequences, gene pairing, gambler's ruin, and the order statistics of random variables.

1.1. **The road map of this note.** We start by setting up the notations and recalling the basics in Section 2. This section also serves as a dictionary for the entire note. We then recall the definitions of some random variables in Section 3, where the expectations and variances are included as well. Nevertheless, many of the verifications are pushed to Section 5.2. Specifically, we discuss

(1) the binomial random variable in Example 3.1,
(2) the negative binomial random variable in Example 3.3,
(3) the hypergeometric random variable in Example 3.4,
(4) the negative hypergeometric random variable in Example 3.5,
(5) the Poisson random variable in Example 3.6,
(6) the uniform random variable in Example 3.9,
(7) the normal random variable in Example 3.12,
(8) the exponential random variable in Example 3.14,
(9) the gamma random variable in Example 3.17,
(10) the beta random variable in Example 3.20, and
(11) the relation between the Poisson and the exponential random variables in Section 3.3.

In Section 4 we delve directly into examples of probability. The types of problems are categorized into the following subsections:

(1) Poker in Section 4.1;
(2) Matching in Section 4.2;
(3) Coin toss sequence in Section 4.3;
(4) Gene pairings in Section 4.4;
(5) Monty Hall and the Russian roulette problems in Section 4.5;
(6) Candies in a jar in Section 4.6;
(7) Gambler's ruin in Section 4.7.

In Section 5, we take a serious look at the techniques to calculate expectations and variances. In Section 5.2, we look at the expectations and variances of the following random variables:

(1) The geometric random variable in Example 5.11;
(2) The negative binomial random variable in Example 5.12;
(3) The hypergeometric random variable in Example 5.13;
(4) The negative hypergeometric random variable in Example 5.16.

We then expand to following topics:

(1) Coupon collections in Section 5.3;
(2) Matches in random pairings in Section 5.4;
(3) Coin toss sequences in Section 5.5.

The exposition culminates at Section 5.6, where several trickier expectation problems are thoroughly discussed.

In Section 6, we discuss the techniques to deal with the ordering of random variables. Finally, we provide a few counterexamples in Section 7.

2. Preliminaries

In this section, we set up some of the notations used throughout the note. We also recall some basic definitions and properties of probability such as independence of events (Definition 2.3), the multiplication principle (Equation (2.1), Equation (2.2)), and the inclusion-exclusion principle (Equation (2.3)).

2.1. **Probability of events.** Let E be an event. We denote by $\Pr(E)$ the probability that event E occurs. The complement of E is written as E^c.

Given events $E_1, E_2, \ldots, E_r$, the probability that all r events occur is written as $\Pr(E_1, E_2, \ldots, E_r)$ or more simply as $\Pr(E_1 E_2 \cdots E_r)$. The conditional probability that E occurs given that A occurred is

$$\Pr(E|A) := \frac{\Pr(EA)}{\Pr(A)}.$$

It is straightforward to verify the multiplication principle:

$$\Pr(E_1E_2\cdots E_r) = \Pr(E_1)\Pr(E_2|E_1)\Pr(E_3|E_1E_2)\cdots\Pr(E_r|E_1\cdots E_{r-1}). \tag{2.1}$$

More generally, for any event A,

$$\begin{aligned}\Pr(E_1E_2\cdots E_r|A) = &\Pr(E_1|A)\times\Pr(E_2|E_1A)\times\Pr(E_3|E_1E_2A)\times\cdots \\ &\times\Pr(E_r|E_1\cdots E_{r-1}A).\end{aligned} \tag{2.2}$$

Definition 2.1. We say event A is independent of event B if $\Pr(B) = 0$ or $\Pr(A|B) = \Pr(A)$.

We note that event A is independent of event B if and only if $\Pr(AB) = \Pr(A)\Pr(B)$. Therefore, independence between two events is symmetric (but not transitive: Example 7.1). We note without proof of the

Proposition 2.2. *Event A is independent of event B if and only if event A is independent of event B^c.*

More generally, we may define independence among multiple events.

Definition 2.3. We say events $E_1, \ldots, E_r$ are jointly independent if for all $1 \le i_1 < i_2 < \cdots < i_k \le r$, we have

$$\Pr(E_{i_1}E_{i_2}\cdots E_{i_k}) = \Pr(E_{i_1})\Pr(E_{i_2})\cdots\Pr(E_{i_k}).$$

We will see in Example 7.4 that all subsets are required in Definition 2.3. We now recall the inclusion-exclusion principle. If one requires one of $E_1, \ldots, E_r$ to occur, the probability is written as $\Pr(\cup_{i=1}^r E_i)$. The inclusion-exclusion principle states that

$$\begin{aligned}\Pr(\cup_{i=1}^r E_i) = &\sum_{i=1}^r \Pr(E_i) - \sum_{i_1<i_2}\Pr(E_{i_1}E_{i_2}) + \cdots \\ &+ (-1)^{k+1}\sum_{i_1<i_2<\cdots<i_k}\Pr(E_{i_1}E_{i_2}\cdots E_{i_k}) \\ &+(-1)^{r+1}\Pr(E_1E_2\cdots E_r).\end{aligned} \tag{2.3}$$

In the simplest form we have $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(AB)$, but sometimes this equivalent form is used:

$$\Pr(AB) = \Pr(A) + \Pr(B) - \Pr(A \cup B).$$

2.2. **Basic definitions for distributions.** Here we discuss the definitions of (cumulative) probability density functions for a single random variable, for multiple random variables, and for conditional distributions in Section 2.2.1, Section 2.2.2, and Section 2.2.3 respectively. For multiple random variables, we also define their independence (Definition 2.11), their covariance and correlation (Definition 2.15).

2.2.1. *Single distributions.*

Definition 2.4. Let X be a random variable. If X is discrete, the *probability mass function of* X is the real valued function

$$x \mapsto \Pr(X = x)$$

defined on the sample space of X. If X is continuous, the *probability density function of* X is the real valued function $f(x)$ defined on the sample space of X such that

$$\Pr(X \in B) = \int_B f(x)\, dx.$$

The *cumulative distribution function of* X regardless of X being discrete or continuous, is the function

$$x \mapsto \Pr(X \leq x)$$

defined on the sample space of X.

Definition 2.5. Let X be a random variable, either continuous with probability density function $f(x)$, or discrete. The *expectation of* X is

$$E[X] := \begin{cases} \int_{\mathbf{R}} x f(x)\, dx \text{ if } X \text{ is continuous,} \\ \\ \sum_{x \in \mathbf{Z}} x \cdot \Pr(X = x) \text{ if } X \text{ is discrete.} \end{cases}$$

The *variance of* X is

$$\mathrm{Var}(X) := E[(X - E[X])^2] = E[X^2] - (E[X])^2.$$

The *standard deviation of* X is defined as

$$\sigma_x := \sqrt{\mathrm{Var}(X)}.$$

2.2.2. *Joint distributions.*

Definition 2.6. Given two random variables X, Y, the *joint cumulative distribution function of* X and Y is defined as

$$F(a, b) = \Pr(X \leq a, Y \leq b).$$

The *marginal cumulative distribution function of* X is

$$\begin{aligned} F_X(a) &:= \Pr(X \leq a) \\ &= \Pr(X \leq a, Y \leq \infty) \\ &= \lim_{b \to \infty} F(a, b). \end{aligned}$$

Similarly, the *marginal cumulative distribution function of* Y is

$$F_Y(b) := \Pr(Y \leq b) = \lim_{a \to \infty} F(a, b).$$

Definition 2.7. In the case that both X and Y are discrete, we define the *joint probability mass function of* X and Y as

$$p(x, y) = \Pr(X = x, Y = y).$$

In the case that both X and Y are continuous, we define the *joint probability density function of* X and Y as the function $f(x, y)$ such that

$$\Pr((x, y) \in C) = \iint_{(x,y)\in C} f(x, y)\, dA$$

for any subset $C \subset \mathbf{R}^2$.

Definition 2.8. If both X and Y are discrete with joint probability mass function $p(x, y)$, the *marginal probability mass function of* X is

$$p_X(x) := \sum_{y\in\mathbf{Z}} p(x, y).$$

Similarly the *marginal probability mass function of* Y is

$$p_Y(y) := \sum_{x\in\mathbf{Z}} p(x, y).$$

If both X and Y are continuous with joint probability density function $f(x, y)$, the *marginal probability density function of* X is

$$f_X(x) := \int_{y\in\mathbf{R}} f(x, y)\, dy.$$

Similarly, the *marginal probability density function of* Y is

$$f_Y(y) := \int_{x\in\mathbf{R}} f(x, y)\, dx.$$

Remark 2.9. More generally, we say $F(x_1, x_2, \ldots, x_r)$ is the joint cumulative distribution function of the random variables $X_1, X_2, \ldots, X_r$ if

$$F(x_1, x_2, \ldots, x_r) = \Pr(X_1 \le x_1, X_2 \le x_2, \ldots, X_r \le x_r).$$

In the case that $X_1, X_2, \ldots, X_r$ are continuous, their joint probability density function $f(x_1, x_2, \ldots, x_r)$ can be obtained via the fundamental theorem of calculus

$$f(x_1, x_2, \ldots, x_r) = \frac{\partial^r F}{\partial x_1 \partial x_2 \cdots \partial x_r}(x_1, x_2, \ldots, x_r).$$

The marginal density function of X_i can be obtained by integrating $f(x_1, x_2, \ldots, x_r)$ over all variables except x_i.

Example 2.10. Suppose 3 balls are chosen without replacements from an urn containing 5 white balls and 8 red balls. Suppose the white balls are numbered. Let X_i be equal 1 if the i-th white ball is selected. Otherwise, let X_i be equal to 0. Find the joint probability mass function for X_1, X_2, X_3. Repeat the problem, except now the balls are chosen with replacements.

solution. When the balls are chosen without replacements, we get

$$\begin{aligned}
&\Pr(X_1 = X_2 = X_3 = 0) = \frac{\binom{10}{3}}{\binom{13}{3}},\\
&\Pr(X_1 = 1, X_2 = X_3 = 0) = \Pr(X_2 = 1, X_1 = X_3 = 0)\\
&= \Pr(X_3 = 1, X_1 = X_2 = 0) = \frac{\binom{10}{2}}{\binom{13}{3}},\\
&\Pr(X_1 = 0, X_2 = X_3 = 1) = \Pr(X_2 = 0, X_1 = X_3 = 1)\\
&= \Pr(X_3 = 0, X_1 = X_2 = 1) = \frac{\binom{10}{1}}{\binom{13}{3}},\\
&\Pr(X_1 = X_2 = X_3 = 1) = \frac{\binom{10}{0}}{\binom{13}{3}}.
\end{aligned}$$

When the balls are chosen with replacements, a ball can be selected multiple times. Therefore, we now have

$$\begin{aligned}
&\Pr(X_1 = X_2 = X_3 = 0) = \frac{10^3}{13^3},\\
&\Pr(X_1 = 1, X_2 = X_3 = 0) = \Pr(X_2 = 1, X_1 = X_3 = 0)\\
&= \Pr(X_3 = 1, X_1 = X_2 = 0) = \frac{\binom{3}{1}\cdot 10^2 + \binom{3}{2}\cdot 10 + \binom{3}{3}}{13^3},\\
&\Pr(X_1 = 0, X_2 = X_3 = 1) = \Pr(X_2 = 0, X_1 = X_3 = 1)\\
&= \Pr(X_3 = 0, X_1 = X_2 = 1) = \frac{3\cdot 2\cdot 10 + 2\cdot 3}{13^3},\\
&\Pr(X_1 = X_2 = X_3 = 1) = \frac{3!}{13^3}.
\end{aligned}$$

□

Similar to independence among events (Definition 2.3), we can define independence among random variables.

Definition 2.11. We say the r random variables $X_1, X_2, \dots, X_r$ are independent if for any subsets $B_1, \dots, B_r$ in the sample space of $X_1, \dots, X_r$ respectively, we have

$$\begin{aligned}
&\Pr(X_1 \in B_1, X_2 \in B_2, \dots, X_r \in B_r) =\\
&\Pr(X_1 \in B_1) \times \Pr(X_2 \in B_2) \times \cdots \times \Pr(X_r \in B_r).
\end{aligned}$$

Remark 2.12. If $X_1, \dots, X_r$ are independent random variables, then any subset of $\{X_1, X_2, \dots, X_r\}$ consists of independent random variables.

Proposition 2.13. *Let $X_1, X_2, \dots, X_r$ be r continuous random variables. The following statements are equivalent.*

(1) $X_1, X_2, \dots, X_r$ are independent.
(2) The joint cumulative distribution function is the product of the marginal cumulative density functions.
(3) The joint probability density function is the product of the probability density functions.

Similar statements hold if $X_1, X_2, \dots, X_r$ are discrete random variables.

Example 2.14 (Sample mean and variance)**.** Let $X_1, X_2, \ldots, X_n$ be identically distributed and independent random variables with common mean μ and variance σ^2. The *sample mean* is

$$\overline{X} = \frac{1}{n} \cdot \sum_i X_i.$$

The *sample variance* is

$$S^2 = \sum_i \frac{(X_i - \overline{X})^2}{n-1}.$$

In fact, we have

$$\begin{aligned} \mathrm{Var}(\overline{X}) &= \frac{\sigma^2}{n}, \\ E[S^2] &= \sigma^2. \end{aligned}$$

Finally, we define the correlation between two random variables.

Definition 2.15. Let X, Y be two random variables. The *covariance between* X and Y is

$$\mathrm{Cov}(X,Y) := E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y].$$

The *correlation between* X and Y is

$$\rho(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{\mathrm{Cov}(X,Y)}{\sigma_x \sigma_y}.$$

In Corollary 5.3 and Remark 5.4, we will see that if X, Y are continuous random variables with joint probability density function $f(x,y)$, then given any function $g(x,y)$, we have

$$E[g(X,Y)] = \iint g(x,y) f(x,y)\, dA.$$

In particular, if X, Y are independent, then

$$E[XY] = \iint xy f_X(x) f_Y(y)\, dA = \int_{x \in \mathbf{R}} x f_X(x)\, dx \int_{y \in \mathbf{R}} y f_Y(y)\, dy = E[X]E[Y].$$

Therefore, $\mathrm{Cov}(X,Y) = 0$ if X, Y are independent (true for any two independent random variables, continuous or not). The converse however, is not true. A counterexample is presented at Example 7.6.

Proposition 2.16. *Let X, Y be two random variables. We then have*

(1) $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$. *In general, for any collection of random variables $X_1, \ldots, X_n$, we have*

$$\mathrm{Var}(\sum_i X_i) = \sum_i \mathrm{Var}(X_i) + 2 \sum_{i<j} \mathrm{Cov}(X_i, X_j).$$

In the case that $X_1, \ldots, X_n$ are independent, the variance of the sum is the sum of the variances.

(2) The correlation between X and Y satisfies

$$-1 \leq \rho(X,Y) \leq 1,$$

where the equality holds if and only if X and Y are related linearly. More precisely, we have

$$\rho(X,Y) = -1 \Leftrightarrow Y = -\frac{\sigma_y}{\sigma_x}X + b \text{ for some constant } b. \text{ Similarly,}$$
$$\rho(X,Y) = 1 \Leftrightarrow Y = \frac{\sigma_y}{\sigma_x}X + b \text{ for some constant } b.$$

Proof. Property (1) is standard textbook material. For (2), it is easy to verify first that if there are constants $a \neq 0$ and b so that $Y = aX + b$, then $\rho(X,Y) = \pm 1$ depending on the sign of a. To show the converse, since the variance is always non-negative, we have

$$0 \leq \operatorname{Var}\left(\frac{X}{\sigma_x} + \frac{Y}{\sigma_y}\right) = 2 + 2\operatorname{Cov}\left(\frac{X}{\sigma_x}, \frac{Y}{\sigma_y}\right) = 2(1 + \rho(X,Y)).$$

Therefore, we get $\rho(X,Y) \geq -1$.

When $\rho(X,Y) = -1$, we get $\operatorname{Var}\left(\frac{X}{\sigma_x} + \frac{Y}{\sigma_y}\right) = 0$, which is only possible if $\frac{X}{\sigma_x} + \frac{Y}{\sigma_y} = b$ for some constant b. The other part is proved by considering $\operatorname{Var}\left(\frac{X}{\sigma_x} - \frac{Y}{\sigma_y}\right)$. □

2.2.3. *Conditional distributions.*

Definition 2.17. Let X and Y be two discrete random variables with joint probability mass function $p(x,y)$. The *conditional probability mass function of X given that $Y = y$* is

$$p_{X|Y}(x|y) := \Pr(X = x|Y = y) = \frac{\Pr(X = x, Y = y)}{\Pr(Y = y)} = \frac{p(x,y)}{p_Y(y)}.$$

Definition 2.18. Let X and Y be two continuous random variables with joint probability density function $f(x,y)$. The *conditional probability density function of X given that $Y = y$* is

$$f_{X|Y}(x|y) := \frac{f(x,y)}{f_Y(y)}.$$

Sometimes a discrete random variable is conditioned by a continuous random variable. For example, suppose the probability X of a coin landing on its head is a uniform distribution on the interval $(0,1)$ (see Example 3.9). The number N of heads in a sequence of m tosses given that $X = x$ has the probability mass function

$$\Pr(N = k|X = x) = \binom{m}{k} x^k (1-x)^{m-k}.$$

Given the conditional probability $\Pr(N = n|X = x)$ where N is discrete and X is continuous, we can derive the conditional distribution of X given $N = n$ via the

Proposition 2.19. *Let X be a continuous random variable with probability density function $f(x)$ and let N be a discrete random variable. The* conditional probability density function of X given that $N = n$ *is*

$$f_{X|N}(x|n) := \frac{\Pr(N = n|X = x)}{\Pr(N = n)} f(x).$$

Remark 2.20. We note first that one may compute probabilities by conditioning. Namely, given random variables X and Y, we have

$$\Pr(X \in B) = \begin{cases} \displaystyle\sum_{y \in \mathbf{Z}} \Pr(X \in B | Y = y) \cdot \Pr(Y = y) \text{ if } Y \text{ is discrete,} \\ \\ \displaystyle\int_{y \in \mathbf{R}} \Pr(X \in B | Y = y) \cdot f_Y(y)\, dy \text{ if } Y \text{ is continuous} \end{cases}$$

for any subset B of the sample space of X.

Second, conditional distributions are probability distributions in the following sense: For any y in the sample space of Y,

$$1 = \begin{cases} \displaystyle\sum_{x \in \mathbf{Z}} f_{X|Y}(x|y) \text{ if } X \text{ is discrete,} \\ \\ \displaystyle\int_{x \in \mathbf{R}} f_{X|Y}(x|y)\, dx \text{ if } X \text{ is continuous.} \end{cases}$$

Example 2.21. The probability X of a coin landing on its head is first selected from a uniform distribution on $(0, 1)$. Let N be the number of heads in a $n + m$ toss sequence. What is the conditional distribution of the success probability X, given that $N = n$?

solution. We have $f_{N|X}(n|x) = \binom{n+m}{n} x^n (1 - x)^m$. Therefore by Proposition 2.19, we have

$$f_{X|N}(x|n) = \frac{\binom{n+m}{n} x^n (1 - x)^m}{\Pr(N = n)}, 0 < x < 1.$$

Therefore $f_{X|N}(x|n) = cx^n(1 - x)^m$ for some constant c. As we have seen in Remark 2.20, $f_{X|N}(x|n)$ is a probability density function. Hence it must be the probability density function of the beta distribution (Example 3.20). □

3. Some important random variables

We introduce some commonly seen random variables, together with their expectations and variances. More precisely, we present

(1) the binomial random variable in Example 3.1,
(2) the negative binomial random variable in Example 3.3,
(3) the hypergeometric random variable in Example 3.4,
(4) the negative hypergeometric random variable in Example 3.5,
(5) the Poisson random variable in Example 3.6,
(6) the uniform random variable in Example 3.9,
(7) the normal random variable in Example 3.12,
(8) the exponential random variable in Example 3.14,
(9) the gamma random variable in Example 3.17, and
(10) the beta random variable in Example 3.20.

We also consider sums of independent random variables. Specifically, we discuss the sum of independent

(1) binomial random variables in Proposition 3.2,
(2) Poisson random variables in Proposition 3.7,
(3) normal random variables in Proposition 3.13, and
(4) gamma random variables in Proposition 3.18.

Finally, we relate the exponential random variable and the Poisson random variable in Section 3.3.

3.1. Discrete random variables.

Example 3.1 (The binomial random variable)**.** The *binomial random variable with parameters* (n, p) is the number of successes X in n independent trials, where each trial has probability p of success. The probability mass function of X is

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^k.$$

In this case, we have

$$E[X] = np,$$
$$\operatorname{Var}(X) = np(1-p).$$

Proposition 3.2. *Let X and Y be two independent binomial random variables (Definition 2.11) with parameters (n, p) and (m, p) respectively. The sum $X + Y$ is a binomial random variable with parameters $(m + n, p)$.*

Example 3.3 (The negative binomial random variable)**.** Suppose that independent trials, each having probability p of being a success are performed. The *negative binomial random variable with parameters* (r, p) is the number of trials X required to obtain a total of r successes. The probability mass function of X is

$$\Pr(X = n) = \binom{n-1}{r-1} (1-p)^{n-r} p^r.$$

When $r = 1$, the random variable is also known as the *geometric random variable.*

We will see in Example 5.11 and Example 5.12 that

$$E[X] = \frac{r}{p},$$
$$\operatorname{Var}(X) = \frac{r(1-p)}{p^2}.$$

Example 3.4 (The hypergeometric random variable)**.** In a collection of N balls, m are special, and the other $N - m$ are ordinary. The *hypergeometric random variable* in this case is the number of special balls X obtained when n balls are selected without replacement. The probability mass function of X is

$$\Pr(X = k) = \frac{\binom{m}{k} \cdot \binom{N-m}{n-k}}{\binom{N}{n}}.$$

We will see in Example 5.13 that

$$E[X] = \frac{mn}{N},$$
$$\operatorname{Var}(X) = \frac{mn(m-1)(n-1)}{N(N-1)} + \frac{mn}{N} - \left(\frac{mn}{N}\right)^2.$$

Example 3.5 (The negative hypergeometric random variable)**.** In a collection of N balls, m are special, and the other $N - m$ are ordinary. Selections without replacements are made until r special balls are obtained. In this case the *negative hypergeometric random variable* is the number X of selections made in order to obtain r special balls. The probability mass function of X is

$$\Pr(X = k) = \frac{m}{N} \cdot \frac{\binom{m-1}{r-1} \cdot \binom{N-m}{k-r}}{\binom{N-1}{k-1}} = \frac{\binom{m}{r-1} \cdot \binom{N-m}{k-r}}{\binom{N}{k-1}} \cdot \frac{m-r+1}{N-k+1}.$$

We will see in Example 5.16 that

$$E[X] = (N-m) \cdot \frac{r}{m+1} + r.$$

The variance is also calculated in Example 5.16, but we leave it out here due to its complexity.

Caution! The denominator of the probability mass function in Example 3.5 is not $\binom{N}{k}$. Why not?

Example 3.6 (The Poisson random variable)**.** The Poisson random variable X with parameter λ has the probability mass function

$$\Pr(X = k) = e^{-\lambda} \cdot \left(\frac{\lambda^k}{k!}\right), k \geq 0.$$

In this case, we have

$$E[X] = \mathrm{Var}(X) = \lambda.$$

Proposition 3.7. *Let X and Y be two independent Poisson random variables (Definition 2.11) with parameters λ_x and λ_y. The sum $X+Y$ is the Poisson random variable with parameter $\lambda_x + \lambda_y$.*

Example 3.8. Suppose the number of people entering a drug store in an hour is a Poisson random variable with mean $\lambda = 10$. Assume each customer is equally likely to be a man or a woman, independent of the number of customers in the store.

Given that 10 women entered the drug store in an hour, what is the probability that at most three men entered the store in that hour?

solution. Let W (resp. M) be the number of women (resp. men) who entered the drug store in a given hour. Let $X = W + M$ be the number of people who entered the drug store in that hour. By assumption,

$$\begin{aligned}
\Pr(W = a, M = b) =& \Pr(X = a+b) \cdot \Pr(W = a, M = b | X = a+b) \\
=& \Pr(X = a+b) \cdot \binom{a+b}{a} \cdot \left(\frac{1}{2}\right)^{a+b} \\
=& e^{-\lambda} \cdot \frac{\lambda^{a+b}}{(a+b)!} \cdot \binom{a+b}{a} \cdot \left(\frac{1}{2}\right)^{a+b} \\
=& e^{-\lambda/2} \frac{(\lambda/2)^a}{a!} \cdot e^{-\lambda/2} \frac{(\lambda/2)^b}{b!}.
\end{aligned}$$

Therefore, W and M are independent (Definition 2.11) Poisson random variables with the common parameter $\lambda/2$. It follows that

$$\Pr(M \leq 3 | W = 10) = \Pr(M \leq 3) = e^{-5} \cdot \left(1 + \frac{5}{1} + \frac{5^2}{2} + \frac{5^3}{3!}\right).$$

□

3.2. **Continuous random variables.**

Example 3.9 (The uniform random variable)**.** The probability density function of the *uniform random variable on the interval* (a, b) is

$$f(x) = \begin{cases} \dfrac{1}{b-a} \text{ if } a < x < b, \\ 0 \text{ otherwise.} \end{cases}$$

In this case, we have

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

Example 3.10. A man and a woman agree to meet at a certain location. Suppose the man arrives at a time uniformly distributed between 12:15 and 12:45, and the woman independently arrives at a time uniformly distributed between 12:00 and 1:00 pm. Each will depart after five minutes of arrival. What is the probability that they will meet?

solution. Let W (resp. M) be the woman's (resp. the man's) arrival time. The joint distribution of their arrival times is

$$\frac{1}{30 \cdot 60}, (m, w) \in (15, 45) \times (0, 60).$$

For the man and the woman to meet, it is equivalent to require that

$$|W - M| < 5.$$

One can easily visualize this region as a parallelogram having height 30 and base 10. This parallelogram has area 300. The sample space for (m, w) is a rectangle with area 1800. Therefore, the probability that they meet is $1/6$. Formally, one integrates over this parallelogram to get

$$\int_{15}^{45} \int_{m-5}^{m+5} \frac{1}{1800} \, dw \, dm = \frac{1}{6}.$$

□

Example 3.11. Repeat the previous example, except now assume the man is willing to wait for 10 minutes.

solutions setup. We present solutions in the perspective of Remark 2.20, and relate them to direct solutions using double integrals. For this, we let W (resp. M) be the arrival time for the woman (resp. the man). We write the joint density of M, W as

$$f(m, w) = \frac{1}{30 \cdot 60}, (m, w) \in (15, 45) \times (0, 60).$$

The marginal density functions are written as f_W and f_M respectively. Let E be the event that the two meet.

solution 1. Given that $M = m$ where $15 \leq m \leq 45$, if the woman arrives earlier than the man, since she can wait for up to five minutes, she must arrive no earlier than $m - 5$ minutes to meet the man. On the other hand if she arrives later than the man, since the man can wait for up to 10 minutes, she must arrive no later than $m + 10$ minutes in order to meet the man.

Summing up, the woman must arrive within the interval $(m-5, m+10)$. The interval has length 15, so

$$\Pr(E|M=m) = \frac{15}{60} = \frac{1}{4}.$$

By Remark 2.20,

$$\Pr(E) = \int_{15}^{45} \Pr(E|M=m) \cdot f_M(m)\, dm = \int_{15}^{45} \frac{1}{4} \cdot \frac{1}{30}\, dm = \frac{1}{4}.$$

For the sake of practice, let us also condition on the arrival time W of the woman.

solution 2. By the same argument from above, given $W = w$, the man should arrive in the interval $(w-10, w+5)$. Nevertheless, $15 \le m \le 45$, so for some w the interval $(w-10, w+5)$ might not fit in the interval $(15, 45)$. Instead, we have a more complicated description

$$\Pr(E|W=w) = \begin{cases} \frac{w+5-15}{45} = \frac{w-10}{45} \text{ if } 10 < w \le 25, \\ \\ \frac{w+5-(w-10)}{45} = \frac{1}{3} \text{ if } 25 < w < 40, \\ \\ \frac{45-(w-10)}{45} = \frac{55-w}{45} \text{ if } 40 \le w < 55, \\ 0 \text{ otherwise.} \end{cases}$$

Using Remark 2.20, we have

$$\begin{aligned} \Pr(E) &= \int_0^{60} \Pr(E|W=w) f_W(w)\, dw \\ &= \frac{1}{60} \cdot \left(\int_{10}^{25} \frac{w-10}{45}\, dw + \int_{25}^{40} \frac{1}{3}\, dw + \int_{40}^{55} \frac{55-w}{45}\, dw \right) \\ &= \frac{1}{4}. \end{aligned}$$

If one were to solve the problem with double integrals, the region over which one integrates is the parallelogram with vertices $(15, 10)$, $(15, 25)$, $(45, 40)$, $(45, 55)$ in the mw-plane. It is easier to integrate with respect to w first over this region. In fact,

$$\begin{aligned} \Pr(E) &= \int_{15}^{45} \int_{m-5}^{m+10} \frac{1}{30 \cdot 60}\, dw\, dm \\ &= \int_{15}^{45} \frac{1}{4} \cdot \frac{1}{30}\, dm = \int_{15}^{45} \Pr(E|M=m) f_M(m)\, dm, \end{aligned}$$

which is precisely the integral we used in the first solution. If one integrates with respect to m first, the integral one gets is precisely the integral we used in the second solution. □

Example 3.12 (The normal random variable)**.** The probability density function of the *normal random variable with parameters* μ *and* σ^2 is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-u)^2}{2\sigma^2} \right), x \in \mathbf{R}.$$

In this case, we have

$$E[X] = \mu,$$
$$\mathrm{Var}(X) = \sigma^2.$$

A *standard normal random variable* is a normal random variable with $\mu = 0$ and $\sigma = 1$.

Proposition 3.13. *If X_i, $i = 1, \ldots, n$ are independent normal random variables (Definition 2.11) with parameters μ_i, σ_i^2 respectively, then the sum $\sum_{i=1}^{n} X_i$ is normally distributed with parameter $\sum_i \mu_i$ and $\sum_i \sigma_i^2$.*

Example 3.14 (The exponential random variable)**.** The probability density function of the *exponential distribution with parameter* λ is

$$f(x) = \begin{cases} \lambda e^{-\lambda x} \text{ if } x \geq 0, \\ 0 \text{ otherwise.} \end{cases}$$

In this case, we have

$$E[X] = \frac{1}{\lambda},$$
$$\mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

Remark 3.15. The exponential distribution is memoryless in the following sense. Suppose the time T a computer breaks down is exponentially distributed with parameter λ. We have

$$\Pr(T > t) = 1 - \Pr(T < t) = 1 - \int_0^t \lambda e^{-\lambda u}\, du = e^{-\lambda t}.$$

Therefore,

$$\Pr(T > t + s | T > t) = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = \Pr(T > s).$$

This means if the computer has not broken down after t units of time, the chance that it breaks down in the next s units of time is the same as the chance that it does in s units of time from the beginning. We can try to reconcile this somewhat surprising phenomenon by saying that if the computer has not broken down, it is probably a good computer.

Example 3.16. The number of accidents that a person has in a given year is a Poisson random variable with mean λ. However, λ changes from person to person. Assume that the proportion of the population having a value $\lambda \leq x$ is $1 - e^{-x}$. Assume accidents are independent from year to year. Find the probability that a person will have 3 accidents, given that they have 0 accident in the previous year.

solution. The probability density function of λ for any person is $-e^{-x}$, $x > 0$ obtained by differentiating $1 - e^{-x}$. Therefore, λ is exponentially distributed with parameter 1.

By Remark 2.20, the probability that the person has 0 accident is

$$\int_0^\infty e^{-x} \cdot \Pr(N = 0 | \lambda = x)\, dx = \int_0^\infty e^{-x} \cdot e^{-x}\, dx = \frac{1}{2}.$$

Next, the probability that the person has 0 accident and then 3 accidents is given by

$$\begin{aligned}\int_0^\infty e^{-x} \cdot \Pr(N = 0 \text{ then } N = 3|\lambda = x)\, dx &= \int_0^\infty e^{-x} \cdot e^{-x} \cdot e^{-x} \cdot \frac{x^3}{3!}\, dx \\ &= \frac{1}{81}.\end{aligned}$$

Therefore, the conditional probability in question is $2/81$. □

Example 3.17 (The gamma random variable)**.** Define the gamma function

$$\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1}, \alpha > 0.$$

The probability density function of the *gamma random variable with parameters* $(\alpha, \lambda), \lambda > 0, \alpha > 0$ is

$$f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & \text{if } x \geq 0, \\ 0 \text{ otherwise.} \end{cases}$$

In this case, we have

$$\begin{aligned} E[X] &= \frac{\alpha}{\lambda}, \\ \operatorname{Var}(X) &= \frac{\alpha}{\lambda^2}. \end{aligned}$$

We note that the exponential random variable is a special case of the gamma random variable by letting $\alpha = 1$.

Proposition 3.18. *If X, Y are independent gamma random variables with parameters (s, λ) and (t, λ), then the sum $X + Y$ is again gamma with parameters $(s + t, \lambda)$.*

Here is a relation between the normal random variables and the gamma random variables.

Proposition 3.19. *Let Z be a standard normal random variable. The square Z^2 is a gamma random variable with parameters $(1/2, 1/2)$. If $Z_1, \ldots, Z_n$ are independent (Definition 2.11) standard normal random variables, then the sum $\sum_i Z_i^2$ is the gamma random variable with parameters $(n/2, 1/2)$ and is called the* chi-squared distribution.

Example 3.20 (The beta random variable)**.** Define the function

$$B(a, b) = \int_0^1 x^{a-1}(1 - x)^{b-1}\, dx.$$

The probability density function of the *beta random variable with parameters* a, b is

$$f(x) = \begin{cases} \dfrac{1}{B(a, b)} x^{a-1}(1 - x)^{b-1} & \text{if } 0 < x < 1, \\ 0 \text{ otherwise.} \end{cases}$$

In this case, we have

$$E[X] = \frac{a}{a+b},$$
$$\mathrm{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}.$$

3.3. **Relating the Poisson and the exponential random variables.**

Definition 3.21. We say the occurrence of an event follows the Poisson process at the rate λ if the following holds:

(1) The frequency of occurrences of the event is λ. More precisely, the probability that exactly 1 event occurs in a given time interval of length h is $\lambda h + o(h)$. where $o(h)$ is some function $f(h)$ with $\lim_{h \to 0} \frac{f(h)}{h} = 0$.
(2) The probability that 2 or more events occur in an interval of length h is equal to $o(h)$.
(3) For any integers $j_1, j_2, \ldots, j_n$ and any n non-overlapping time intervals, if we define E_i as the event that exactly j_i of the events under consideration occur in the ith of these intervals, then the events $E_1, E_2, \ldots, E_n$ are jointly independent(Definition 2.3).

For any time interval of length t, the event occurs randomly at any point. Therefore, it is reasonable to sub-divide the interval into infinitesimal ones. Namely, we divide it into n equally spaced sub-intervals, each having length $\frac{t}{n}$ for sufficiently large n.

Condition (1) then implies the probability that the event occurs in a particular sub-interval is $\lambda \frac{t}{n} + o(\frac{t}{n})$. Conditions (2) and (3) imply that the number of occurrences of the event in a time interval of length t is approximately the binomial distribution with parameters $(n, \frac{\lambda}{t} n)$. Interested readers can read chapter 4 of [Ros14] for the details.

Let $N(t)$ denote the number of times the event occurs in the time interval. We then have

$$\begin{aligned}
\Pr(N(t) = k) \approx & \binom{n}{k} \cdot \left(\frac{\lambda t}{n}\right)^k \cdot \left(1 - \frac{\lambda t}{n}\right)^{n-k} \\
= & \frac{n \cdots (n-k+1)}{k! n^k} \cdot (\lambda t)^k \cdot \left(1 - \frac{\lambda t}{n}\right)^{n-k} \\
\to & \frac{(\lambda t)^k}{k!} \cdot e^{-\lambda t} \text{ as } n \to \infty.
\end{aligned}$$

Therefore, for any $t > 0$, the random variable $N(t)$ is the Poisson random variable with parameter λt. On the other hand, the time T before the first occurrence of the event satisfies

$$\begin{aligned}
\Pr(T \le t) &= 1 - \Pr(T > t) \\
&= 1 - \Pr(N(t) = 0) = 1 - e^{-\lambda t},
\end{aligned}$$

which is the cumulative density function of the exponential distribution with parameter λ. We have derived the following:

Proposition 3.22. *If an event follows the Poisson process at the rate λ, its number of occurrences on a time interval of length t follows the Poisson distribution with parameter λt. The time before the first occurrence of the event follows the exponential distribution with parameter λ.*

4. Examples of probability

In this section we categorize the examples into

(1) poker in Section 4.1,
(2) matching in Section 4.2,
(3) coin toss sequence in Section 4.3,
(4) gene pairings in Section 4.4,
(5) decision makings in Section 4.5,
(6) candies in a jar in Section 4.6, and
(7) gambler's ruin in Section 4.7.

Specifically speaking in Section 4.2 we look at how people are grouped or paired together in a tournament match or in a social gathering. In Section 4.5 we envision ourselves as the players in the Monty Hall and the Russian roulette games, and strategize for better chances of winning. In Section 4.6, we discuss the order of the types of candies that are depleted in a jar.

4.1. **Poker.**

Example 4.1. We consider the probability that a set of 5 cards out of 52 in a poker game is

- straight. Namely, there are 5 consecutive values but not all of the same suit;
- a full house;
- two pairs;
- four of a kind.

Here if we only allow consecutive values to range from $A, 2, 3, 4, 5$ to $10, J, Q, K, A$, the probability to get a straight set would be

$$\frac{10(4^5-4)}{\binom{52}{5}}.$$

To get a full house, we need two different values, one for 3 cards and one for 2 cards. Therefore, the probability is

$$\frac{13\cdot\binom{4}{3}\cdot 12\cdot\binom{4}{2}}{\binom{52}{5}}.$$

To get two pairs, we first need two values, one for each pair. The 5-th card has to be a different value, so there are only $52-4\cdot 2=44$ choices. One needs to be careful that the two pairs have the same number of cards. Instead of using $13\cdot 12$ when choosing values, we use $\binom{13}{2}$. The probability to get two pairs is

$$\frac{\binom{13}{2}\cdot\binom{4}{2}\cdot\binom{4}{2}\cdot 44}{\binom{52}{5}}.$$

Finally, to get four of a kind, we need a value for four cards. The fifth card is any other card. The probability to get four of a kind is then

$$\frac{13\cdot(52-4)}{\binom{52}{5}}.$$

Example 4.2. Cards from a deck of 52 cards are turned up one at a time. Each card has the same probability to be the one immediately after the first ace (except the first ace itself). To see this, one first excludes the card to be immediately after the first ace, then arranges the remaining 51 cards first. For each of the 51!

arrangements, there is precisely one place for the excluded card to be the successor of the first ace. The probability is therefore always

$$\frac{51!}{52!} = \frac{1}{52}.$$

Example 4.3. Fifty two poker cards are split into two heaps, each having 26 cards. A card was drawn from the first heap. It is an ace. The ace is then inserted into the second heap, which is shuffled immediately. A card is then drawn from the second heap. What is the probability that the card drawn is an ace?

solution. Let A_1 be the ace drawn from the first heap. Let A_2, A_3, A_4 be the remaining aces. Let E be the event that a card drawn from the second heap is an ace, and let E_i be the event that the card drawn from the second heap is the ace A_i.

Note that there are 27 cards in the second heap. So $\Pr(E_1) = 1/27$. For other aces to be drawn, they have to be in the second heap in the first place. Nevertheless, A_1 was originally excluded from the second heap. Therefore, the second heap has only 51 cards to choose from. It follows that

$$\Pr(E_i) = \frac{26}{51} \cdot \frac{1}{27}, i = 2, 3, 4.$$

Since the events $E_1, \dots, E_4$ are mutually exclusive and exhaustive, we get

$$\Pr(E) = \sum_i \Pr(E_i) = \frac{43}{459}.$$

□

Example 4.4. A friend randomly chooses two cards, without replacement, from an ordinary deck of 52 cards. Determine the probability that both cards are aces given the following scenarios:

(1) The first card is an ace.
(2) The second card is an ace.
(3) One of the cards is an ace.

solution. Let F (resp. S) be the event that the first (resp. second) card is an ace. The probability for (1) is

$$\Pr(FS|F) = \frac{4 \times 3}{4 \times 51} = \frac{1}{17}.$$

Similarly, the probability for (2) is

$$\Pr(FS|S) = \frac{4 \times 3}{51 \times 4} = \frac{1}{17}.$$

For (3), the probability is

$$\Pr(FS|F \cup S) = \frac{\Pr(FS)}{\Pr(F \cup S)} = \frac{4 \times 3}{4 \times 51 + 4 \times 51 - 4 \times 3} = \frac{1}{33}.$$

□

4.2. **Matching problems.**

Example 4.5. There are n men, each wearing a hat. The hats are thrown into a pile and then each man picks up a hat at random. Find the probability that

(1) a particular man gets his own hat;
(2) each man in a particular k-men group gets his own hat;
(3) none of the men gets his own hat;
(4) exactly k men get their own hats.

Compare Example 5.28.

solution. Each man has n hats to choose from. Therefore, the probability for a man to get his own hat is $1/n$. For (2)-(4), let E_i be the event that the i-th man gets his own hat.

For (2), suppose we are looking at men $i_1, i_2, \ldots, i_k$. We then have

$$\begin{aligned}\Pr(E_{i_1}E_{i_2}\cdots E_{i_k}) =& \Pr(E_{i_1})\Pr(E_{i_2}|E_{i_1})\cdots\Pr(E_{i_k}|E_{i_1}\cdots E_{i_{k-1}})\\ =&\frac{1}{n}\cdot\frac{1}{n-1}\cdot\dots\cdot\frac{1}{n-k+1}=\frac{(n-k)!}{n!}.\end{aligned}$$

The result does not depend on the k-men chosen, we therefore let the probability in (2) be written as r_k.

For (3), the probability that there is at least one man getting his own hat is

$$p_n := \Pr(\cup_{i=1}^n E_i).$$

With the inclusion-exclusion principle, Equation (2.3), we get

$$\begin{aligned}\Pr(\cup_{i=1}^n E_i) =& \sum_{k=1}^n(-1)^{k+1}\sum_{i_1<i_2<\cdots<i_k}\Pr(E_{i_1}E_{i_2}\cdots E_{i_k})\\ =&\sum_{k=1}^n(-1)^{k+1}\binom{n}{k}\frac{(n-k)!}{n!}=\sum_{k=1}^n(-1)^{k+1}\frac{1}{k!}.\end{aligned}$$

Therefore, the probability that no one gets his own hat is

$$q_n := 1-p_n = \sum_{k=0}^n(-1)^k\frac{1}{k!}.$$

For (4), given that each man in a particular k-men group gets their own hat, the probability that they are the only matches is q_{n-k}. Therefore, the probability for (4) is

$$s_k := \binom{n}{k}r_k q_{n-k} = \frac{q_{n-k}}{k!} = \frac{1}{k!}\sum_{i=0}^{n-k}\frac{(-1)^i}{i!}.$$

It is interesting to note that $q_{n-k}\to e^{-1}$ as $n\to\infty$. Therefore,

$$s_k \to \frac{e^{-1}}{k!} \text{ as } n\to\infty.$$

□

Example 4.6. Suppose there are n married couples, and they are regrouped randomly into n pairs. Find the probability that

(1) a particular couple is paired together;
(2) k particular couples are paired together;
(3) none of the couples are paired together;

(4) exactly k couples are paired together.

Also compare Example 5.33.

solution. There are a total of $\frac{(2n)!}{2^n}$ ways to group $2n$ people into the first pair, the second pair, and so on till the n-th pair. Therefore, there are a total of

$$\frac{(2n)!}{2^n n!}$$

ways to group $2n$ people into n pairs. The probability that a particular couple is paired together is then

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

For (2)-(4), let E_i be the event that the i-th couple is paired together.

For (2), suppose we are looking at couples $i_1, i_2, \ldots, i_k$. We then have

$$\begin{aligned}\Pr(E_{i_1}E_{i_2}\cdots E_{i_k}) =& \Pr(E_{i_1})\Pr(E_{i_2}|E_{i_1})\cdots\Pr(E_{i_k}|E_{i_1}\cdots E_{i_{k-1}})\\ =& \frac{1}{2n-1}\cdot\frac{1}{2(n-1)-1}\cdots\cdot\frac{1}{2(n-k+1)-1}.\end{aligned}$$

The probability does not depend on which k couples are chosen. We therefore give this probability in (2) the symbol r_k.

For (3), the probability that there is at least one couple that is paired together is

$$p_n := \Pr(\cup_{i=1}^n E_i).$$

Using the inclusion-exclusion principle, Equation (2.3), we get

$$\begin{aligned}\Pr(\cup_{i=1}^n E_i) =& \sum_{k=1}^n (-1)^{k+1} \sum_{i_1<i_2<\cdots<i_k} \Pr(E_{i_1}E_{i_2}\cdots E_{i_k})\\ =& \sum_{k=1}^n (-1)^{k+1}\binom{n}{k} r_k.\end{aligned}$$

Hence, the probability for (3) is

$$q_n := 1 - p_n.$$

Finally, given that k particular couples are grouped together, the probability that they are the only couples paired together would be q_{n-k}. Therefore, the probability for (4) is

$$\binom{n}{k} r_k q_{n-k}.$$

□

Example 4.7. Suppose there are $2n$ defensive players and $2n$ offensive players in a sports team. The $4n$ players are paired at random. Find the probability that there are exactly $2i$ pairs consisting of one offensive player and one defensive player, where $i = 0, 1, \ldots, n$. Compare Example 5.33.

solution. As we have seen in Example 4.6, there are a total of

$$M := \frac{(4n)!}{2^{2n}(2n)!}$$

possible pairings. To have $2i$ pairs of one offensive and one defensive player, we can choose $2i$ offensive players and $2i$ defensive players. These players then have $(2i)!$ possible pairings. These amount to $\binom{2n}{2i}^2(2i)!$ pairs.

We now need to group the remaining $2n-2i$ offensive players into $n-i$ pairs (similarly for the remaining $2n-2i$ defensive players). There are precisely

$$\left(\frac{(2n-2i)!}{2^{n-i}(n-i)!}\right)^2$$

possible pairs. By the multiplication principle, there are

$$N := \binom{2n}{2i}^2 (2i)! \left(\frac{(2n-2i)!}{2^{n-i}(n-i)!}\right)^2$$

pairs meeting the condition in question. The probability for this example is then N/M. □

Example 4.8. There are 2^n players in a chess tournament with skills $1 > 2 > \cdots > 2^n$. In each round, the player with the higher skill always wins. Suppose two players are drawn at random, except in the last round. Find the probability that the top two players meet in the final.

solution. Let E_i be the event that the top two players do not meet in the i-th round, where $1 \le i \le n-1$. The probability in question is $\Pr(E_1E_2\cdots E_{n-1})$.

As we have seen in problem (1) of Example 4.6, the probability that the top two players meet in round 1 is $\frac{1}{2^n-1}$. Therefore,

$$\Pr(E_1) = 1 - \frac{1}{2^n-1} = 2 \times \frac{2^{n-1}-1}{2^n-1}.$$

It is then clear that

$$\begin{aligned}\Pr(E_1E_2\cdots E_{n-1}) =& \Pr(E_1)\Pr(E_2|E_1)\cdots\Pr(E_{n-1}|E_1\cdots E_{n-2})\\ =&\frac{2(2^{n-1}-1)}{2^n-1} \times \frac{2(2^{n-2}-1)}{2^{n-1}-1} \times \cdots \times \frac{2(2^1-1)}{2^2-1}\\ =&\frac{2^{n-1}}{2^n-1}.\end{aligned}$$

□

Example 4.9. There are 2^n players in a chess tournament with equal skills. In each round, each player has $1/2$ chance of winning. Suppose players are drawn at random, except in the last round. Find the probability that two specific players A and B never play each other.

solution. Let E_i be the event that the two players meet in the i-th round, where $1 \le i \le n$. The probability in question is

$$1 - \Pr(\cup_{i=1}^n E_i) = 1 - \sum_{i=1}^n \Pr(E_i).$$

To meet in round i, the two players must never meet in the first $i-1$ rounds. In addition, they must also win in the first $i-1$ rounds. As we have seen in Example 4.8, the probability that the two players do not meet in round 1 through round $i-1$ is

$$\frac{2(2^{n-1}-1)}{2^n-1} \times \frac{2(2^{n-2}-1)}{2^{n-1}-1} \times \cdots \times \frac{2(2^{n-i+1}-1)}{2^{n-i+2}-1} = \frac{2^{i-1}(2^{n-i+1}-1)}{2^n-1}.$$

Next, after $i-1$ rounds, there are 2^{n-i+1} players left. The probability that two particular players are paired in round i is then $\frac{1}{2^{n-i+1}-1}$. Therefore,

$$\begin{aligned}\Pr(E_i) =&\frac{2^{i-1}(2^{n-i+1}-1)}{2^n-1} \times \left(\frac{1}{2^{i-1}}\right)^2 \times \frac{1}{2^{n-i+1}-1}\\ =&\frac{1}{2^{i-1}(2^n-1)}.\end{aligned}$$

It now remains to verify the geometric series $\sum_{i=1}^{n} \Pr(E_i) = \frac{1}{2^{n-1}}$. The probability in question is then

$$1-\frac{1}{2^{n-1}}.$$

□

Example 4.10. Suppose there are n married couples lined up in a row. Find the probability that

(1) there are k particular husbands next to their wives;
(2) no husband is next to his wife.

Do the same when these couples are arranged randomly at a round table. Also compare Example 5.31.

Proof. Let E_i be the event that the i-th husband stands next to his wife in the row. For k particular couples to not be separated, we can first treat these k couples as k individuals, as if we are lining up $2n-k$ people in a row. Each of the k couples has two possible arrangements. Therefore, for any $i_1 < i_2 < \cdots < i_k, 1 \le k \le n$, we have

$$r_k := P(E_{i_1}E_{i_2}\cdots E_{i_k}) = \frac{(2n-k)!2^k}{(2n)!}.$$

By the inclusion-exclusion principle Equation (2.3), the probability that there is at least one husband next to his wife is

$$\begin{aligned}p_n =&\Pr(\cup_{i=1}^n E_i) = \sum_{k=1}^{n}(-1)^{k+1} \sum_{i_1<i_2<\cdots<i_k} \Pr(E_{i_1}E_{i_2}\cdots E_{i_k})\\ =&\sum_{k=1}^{n}(-1)^{k+1}\binom{n}{k}r_k.\end{aligned}$$

The probability that no husband is next to his wife is then $1-p_n$.

When sitting at a round table, it is easier to mod out the rotation equivalence. Using the same notation E_i for the same event, we have

$$r_k := P(E_{i_1}E_{i_2}\cdots E_{i_k}) = \frac{\frac{(2n-k)!2^k}{2n-k}}{\frac{(2n)!}{2n}} = \frac{(2n-k-1)!2^k}{(2n-1)!}.$$

The rest can be solved using the inclusion-exclusion principle as before. □

4.3. **Coin toss sequence probabilities.**

Example 4.11 (Cumulating results)**.** A coin lands on its head with probability p. Find the probability that n heads are accumulated before m tails are. Compare Example 5.34.

solution. For n heads to be accumulated before m tails, it is necessary and sufficient that there are at least n heads in the first $n+m-1$ flips. The probability in question is then

$$\sum_{k=n}^{n+m-1} \binom{m+n-1}{k} p^k (1-p)^{n+m-k-1}.$$

□

Example 4.12 (Successive results)**.** A coin lands on its head with probability p. Find the probability that a successive string of n heads occurs before a string of m tails does. Compare Example 5.35.

solution. Let E be the event that a string of n consecutive heads occur before a string of m tails does. Also let H (resp. T) be the event that the first flip results in a head (resp. tail). Conditioning on the first flip, we get

$$\Pr(E) = p\Pr(E|H) + q\Pr(E|T),$$

where $q = 1-p$.

Next, given H, any tail in the next $n-1$ flips resets the state to T. Conditioning on the position of the first tail in the next $n-1$ flips, we get

$$\begin{aligned}\Pr(E|H) =& q\Pr(E|T) + pq\Pr(E|T) + \cdots + p^{n-2}q\Pr(E|T) + p^{n-1}\\ =&(1-p^{n-1})\Pr(E|T) + p^{n-1}.\end{aligned}$$

Similarly, given T, any head in the next $m-1$ flips resets the state to H. Therefore,

$$\begin{aligned}\Pr(E|T) =& p\Pr(E|H) + qp\Pr(E|H) + \cdots + q^{m-2}p\Pr(E|H)\\ =&(1-q^{m-1})\Pr(E|H).\end{aligned}$$

One then easily solves

$$\begin{aligned}\Pr(E|H) =& \frac{p^{n-1}}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}},\\ \Pr(E|T) =& \frac{(1-q^{m-1})p^{n-1}}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}},\\ \Pr(E) =& \frac{p^{n-1}(1-q^m)}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}}.\end{aligned} \tag{4.1}$$

□

Example 4.13 (Longest run)**.** A coin that lands on its head with probability p is to be flipped n times. Find the probability that there is a string of k consecutive heads.

solution. We first define the following: A run of k consecutive heads from the n flips terminates at $i = k+1, k+2, \ldots, n$ if the string switches to a tail at the i-th flip. Let E_i be the event that a run of k heads terminates at i, and let F be the event that the last k flips of the n flips all result in heads. The probability in question is then

$$\Pr\left((\cup_{i=k+1}^n E_i) \cup F\right).$$

We will give this probability a precise formula using the inclusion-exclusion principle, Equation (2.3).

For starters, $\Pr(E_i) = p^k q$ and $\Pr(F) = p^k$, where $q = 1-p$. Next, to enumerate the concurrency among the $E_i's$, we associate each E_i with the length $(k+1)$-string, where we also include the tail at the i'th position. For the events $E_{i_1}, E_{i_2}, \ldots, E_{i_r}$ to concur, their associated strings must be disjoint. Therefore, to enumerate how many r of the E_i's can concur, we can shrink these length $(k+1)$-strings to a singleton. It now comes down to choose r positions out of $n - r(k+1) + r = n - rk$ positions. It follows that there are $\binom{n-rk}{r}$ many.

Similarly, for $E_{i_1}, \ldots, E_{i_{r-1}}, F$ to concur, their associated strings must be disjoint. This comes down to choose $r-1$ positions out of $n-k-(r-1)k = n-rk$. It follows that there are $\binom{n-rk}{r-1}$ many $r-1$ of the $E_i's$ that can concur with F.

With the information we just obtained, we get

$$\sum_{i_1<i_2<\cdots<i_r} \Pr(E_{i_1}\cdots E_{i_r}) = \binom{n-rk}{r}(p^k q)^r,$$

$$\sum_{i_1<i_2<\cdots<i_{r-1}} \Pr(E_{i_1}\cdots E_{i_{r-1}}F) = \binom{n-rk}{r-1}p^{rk}q^{r-1}.$$

Using the inclusion-exclusion principle, we obtain

$$\Pr\left((\cup_i E_i)\cup F\right) = \sum_{r=1}^{n-k+1}(-1)^{r+1}\left(\binom{n-rk}{r} + \frac{1}{q}\binom{n-rk}{r-1}\right)(p^k q)^r,$$

where we set $\binom{m}{j} = 0$ if $m < j$. $\square$

Remark 4.14. The above example has the following interpretation. Let L_n denote the length of the longest run of heads in the n flips. The event that there is a string of k consecutive heads is equivalent to the event that $L_n \geq k$. Therefore, the probability that the longest run has length k is

$$\Pr(L_n \geq k) - \Pr(L_n \geq k+1).$$

From this the expected value of L_n can also be derived.

Example 4.15. Independent flips of a coin that lands on heads with probability p are made. What is the probability that the sequence T, H, H, H occurs before the sequence H, H, H, H?

solution. Simply note that the pattern T, H, H, H occurs first if and only if there is a T in the first four flips. Therefore, the probability in question is $1 - p^4$. $\square$

Example 4.16. Two players A and B compete in a series of games. Each game is independently won by A with probability p and by B with probability $1-p$. The game stops when the total wins of a player is two greater than that of the other. The player with the greater number of wins is declared the winner of the series. Find the probability that A wins.

solution. Let E be the event that the player A wins. Let W (resp. L) be the event that player A wins (resp. loses) in the first round. We then have

$$\Pr(E) = p\Pr(E|W) + q\Pr(E|L),$$

where $q = 1-p$. Moreover, conditioning on the result of the second round, we get

$$\begin{aligned}\Pr(E|W) =& p + (1-p)\Pr(E),\\ \Pr(E|L) =& p\Pr(E).\end{aligned}$$

Putting these back to the equation for $\Pr(E)$, one easily solves that

$$\Pr(E) = \frac{p^2}{1-2pq}.$$

□

4.4. **Boys and girls.**

Example 4.17. A new couple, known to have two children, just moved into town. Suppose that one of the children is a girl, and that the couple gives birth to a boy and a girl with equal likelihood, independent of the previous births. What is the probability that both children are girls?

solution. Let G_1 (resp. G_2) be the event that the eldest child is a girl (resp. that the second child is a girl). Let B_1 and B_2 be likewise defined for boys.

Since one of the children is a girl, the sample space consists of G_1G_2, G_1B_2, and B_1G_2. Therefore, the probability in question is $1/3$. □

Example 4.18. A new couple, known to have two children, just moved into town. Suppose that the mother is seen walking with a girl, and that the couple gives birth to a boy and a girl with equal likelihood, independent of the previous births. What is the probability that both children are girls?

solution. Let G_1 (resp. G_2) be the event that the eldest child is a girl (resp. that the second child is a girl). Let B_1 and B_2 be likewise defined for boys. Let G be the event that one of the children seen walking with the mother is a girl. The probability in question is $\Pr(G_1G_2|G)$.

A common mistake is to neglect that there is a choice from the mother involved. If we assume that the mother is equally likely to choose either child to walk with, irrespective to their age and sex, we then have

$$\begin{aligned}\Pr(G) =& \Pr(G_1G_2)\Pr(G|G_1G_2)\\ &+\Pr(G_1B_2)\Pr(G|G_1B_2) + \Pr(B_1G_2)\Pr(G|B_1G_2)\\ =& \frac{1}{4}\cdot 1 + \frac{1}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{1}{2} = \frac{1}{2}\end{aligned}$$

so that

$$\Pr(G_1G_2|G) = \frac{1/4}{1/2} = \frac{1}{2}.$$

Now assume the mother chooses to walk with a girl with probability p, we get

$$\Pr(G) = \frac{1}{4} + \frac{1}{2}p$$

so that

$$\Pr(G_1G_2|G) = \frac{1}{1+2p}.$$

In particular, if the mother always chooses to walk with a daughter, the probability that both children are girls is $1/3$, which is the same answer to Example 4.17. Note that this is because under this condition, seeing the mother walking with her daughter is equivalent to the event that one of the children is a girl. □

Example 4.19. Each of 2 balls is painted either black or gold and then placed in an urn. Suppose each ball is colored black with probability $1/2$, and the paintings of the two balls are independent. Compute the probabilities that both balls are painted gold, given the scenarios:

(1) You know the gold paint is used.
(2) A ball is randomly selected from the urn. The ball is gold.

solution. These are just Example 4.17 and Example 4.18 stated in a different way. The condition in problem (1) means one of the balls is painted gold. □

Example 4.20. The color of a person's eyes is determined by a pair of genes. If a person has two blue-eyed (resp. brown-eyed genes), the color of their eyes is blue (resp. brown). The brown-eyed gene is dominant over the blue-eyed gene in the sense that if a person has one blue-eyed gene and one brown-eyed gene, then the color of their eyes are brown.

Assume Smith and his parents all have brown eyes, but his sister has blue eyes. Also assume a newborn child independently receives one eye gene from each of their parents, and the gene they receive from a parent is equally likely to be either of the two eye genes of that parent.

(1) What is the probability that Smith possesses a blue-eyed gene?
(2) Suppose that Smith's wife has blue eyes. What is the probability that his first born child has blue eyes?
(3) Suppose Smith's first born child has brown eyes. What is the probability that his second child also has brown eyes?

solution. For the following discussions, we let BL (resp. BR) be the blue-eyed gene (resp. brown-eyed gene). We also denote the pair of eye genes a child receives from their parents by (α, β), where the first entry α is the gene from their father and the second entry β is the gene from their mother. For example, the tuple (BR, BL) means the child receives a brown-eyed gene from their father, and a blue-eyed gene from their mother.

The colors of Smith's parents' and his sister's eyes dictate that each of the parents must have a pair of a blue-eyed gene and a brown-eyed gene. Let A_1 be the event that Smith has two brown-eyed genes and let A_2 be the event that Smith has one blue-eyed gene and one brown-eyed gene. Since Smith has brown eyes he can only have the pair (BL, BR), (BR, BL), or (BR, BR). Since each pair is equally likely, we have

$$\Pr(A_1) = \frac{1}{3} \text{ and } \Pr(A_2) = \frac{2}{3}.$$

The probability for (1) is precisely $\Pr(A_2) = \frac{2}{3}$.

For (2) and (3), let B_i be the event that the i-th child of Smith's has brown eyes. The probability for (2) is $\Pr(B_1^c)$ and that for (3) is $\Pr(B_2|B_1)$. We now compute $\Pr(B_1)$. For this, we have

$$\begin{aligned}\Pr(B_1) =& \Pr(A_1)\Pr(B_1|A_1) + \Pr(A_2)\Pr(B_2|A_2)\\ =& \frac{1}{3}\cdot 1 + \frac{2}{3}\cdot\frac{1}{2} = \frac{2}{3}.\end{aligned}$$

Hence the answer for (2) is 1/3.

For (3), we have

$$\begin{aligned}\Pr(B_1B_2) =& \Pr(A_1)\Pr(B_1B_2|A_1) + \Pr(A_2)\Pr(B_1B_2|A_2)\\ =& \frac{1}{3}\cdot 1^2 + \frac{2}{3}\cdot\left(\frac{1}{2}\right)^2 = \frac{1}{2}.\end{aligned}$$

Therefore,

$$\Pr(B_2|B_1) = \frac{1/2}{2/3} = \frac{3}{4}.$$

□

Example 4.21. Genes relating to albinism are denoted by A and a. Only those people who receive the gene a from both parents will have albinism. Persons having the pair (A, a) will be normal but, because they can pass on the trait to their offspring, are called carriers. Suppose a normal couple has two children, exactly one of which is normal. Suppose this normal child mates with another carrier.

(1) What is the probability that the first child is an albino?
(2) Given that the firstborn is normal, what is the probability that the second offspring is an albino?

solution. For the following discussions, we denote the pair of eye genes a child receives from their parents by (α, β), where the first entry α is the gene from their father and the second entry β is the gene from their mother. For example, the tuple (A, a) means the child receives the gene A from their father, and the gene a from their mother.

The conditions stated in the problem dictate that both parents are carriers. Let A_1 be the event that the normal child receives the gene A from both parents and let A_2 be the event that the normal child receives one gene A, and one gene a from the parents. The normal child can only have the pair (A, A), (A, a), or (a, A). We therefore get

$$\Pr(A_1) = \frac{1}{3} \text{ and } \Pr(A_2) = \frac{2}{3}.$$

Now let B_i be the event that the i-th offspring of the normal child is normal. The probability for (1) is then $\Pr(B_1^c)$. For minor convenience for (2), we calculate instead the complement

$$\begin{aligned}\Pr(B_1) =& \Pr(A_1)\Pr(B_1|A_1) + \Pr(A_2)\Pr(B_1|A_2) \\ =& \frac{1}{3}\cdot 1 + \frac{2}{3}\cdot\frac{3}{4} = \frac{5}{6}.\end{aligned}$$

For (2), we first need

$$\begin{aligned}\Pr(B_2^c B_1) =& \Pr(A_1)\Pr(B_2^c B_1|A_1) + \Pr(A_2)\Pr(B_2^c B_1|A_2) \\ =& \frac{1}{3}\cdot 1\cdot 0 + \frac{2}{3}\cdot\frac{3}{4}\cdot\frac{1}{4} \\ =& \frac{1}{8}.\end{aligned}$$

Therefore, the probability for (2) is

$$\Pr(B_2^c|B_1) = \frac{1/8}{5/6} = \frac{3}{20}.$$

□

Example 4.22. Assume that each child in a village is, independently, equally likely to be a boy or a girl.

(1) If every family in the village has two children, what proportion of all sons are the eldest sons?

(2) If every family in the village has three children, what proportion of all sons are the eldest sons?

solution. For (1), we use an ordered tuple (α, β) to describe the children of a family, where the age of α is bigger than that of β. We let B be a boy (resp. G be a girl). For example, (B, G) means the family has an older brother and a younger sister.

To be able to pick out the eldest son, the family we choose from has to have children (B, B), (G, B), or (B, G). The son picked out from families having (B, B) has the chance of $1/2$ being the eldest. Nevertheless, when a family has precisely a girl and a boy, the boy is always the eldest. Therefore, picking the eldest son from such families comes down to picking the boy. The probability that the child picked out from the village is the eldest son, is then

$$\frac{1}{4} \cdot \left(\frac{1}{2} + \binom{2}{1} \cdot \frac{1}{2} \right) = \frac{3}{8}.$$

Since there is obviously fifty percent of sons in the village, the probability that a son is the eldest is

$$\frac{3/8}{1/2} = \frac{3}{4}.$$

For (2), we use an ordered triple (α, β, γ) to express the children in each family, where α is the eldest child. Using the same reasoning from (1), the probability that a child is the eldest son is

$$\frac{1}{8} \cdot \left(\frac{1}{3} + \binom{3}{2} \cdot \frac{1}{3} + \binom{3}{1} \cdot \frac{1}{3} \right) = \frac{7}{24},$$

where

- the term $1/3$ comes from the triple (B, B, B);
- the term $\binom{3}{2} \cdot \frac{1}{3}$ comes from the triples (G, B, B), (B, G, B), and (B, B, G);
- the term $\binom{3}{1} \cdot \frac{1}{3}$ comes from the triples (G, G, B), (B, G, G), and (G, B, G).

Therefore, the answer to (2) is $\frac{7}{12}$. □

4.5. **To be or not to be.**

Example 4.23 (Monty Hall problem)**.** In the show the guest is presented with three doors. Behind one door is a car. Behind the other two are goats. After the guest chooses a door, the host Monty opens a door that he knows has a goat. The guest is then allowed to switch doors. In terms of the probability of winning the car, should the guest switch?

solution. If the guest does not switch doors at all, the probability of winning the car stays $1/3$. If the guest switches, they always win if they missed the door that has the car behind it. Therefore, the probability of winning by switching is $2/3$. □

Example 4.24 (Russian Roulette)**.** Assume the revolver has n empty chambers. A bullet is randomly inserted into one of the chambers. Two players take turns to pull the trigger, aiming at their own heads. The one who gets shot in the head dies and loses the game.

(1) Find the probability of winning for the first and for the second player.
(2) Assume now we spin the barrel after every trigger pull. Will you choose to be the first or the second player?

(3) Assume two bullets are inserted randomly into the chambers. Given that your opponent pulled the trigger first and survived, should you spin the barrel?

(4) Assume $n = 6$ and two bullets are inserted in consecutive positions. If your opponent survived the first round, should you spin the barrel?

solution. For (1) and (2), let F (resp. S) be the event that the first (resp. second) player wins. Label the chamber containing the bullet as 0, and let the barrel spin in the direction of the $0, 1, 2, \ldots$th chamber.

For (1), the first player wins if and only if the spin starts at positions $2, 4, \ldots, (n-1)$ if n is odd. If n is even, the first player wins if and only if the spin starts at odd positions. Therefore,

$$\Pr(F) = \begin{cases} \frac{n-1}{2n} \text{ if } \text{ is odd,} \\ \frac{1}{2} \text{ if } n \text{ is even.} \end{cases}$$

In sum, the first player has a lower chance of winning than the second when n is odd. When n is even, both players have an equal chance of winning.

For (2), once the barrel spins after the first trigger pull and the first player survives, it is as if the second player is now the first player. The second player must lose for the first to win. Therefore,

$$\Pr(F) = \frac{n-1}{n} \cdot (1 - \Pr(F)),$$

resulting in

$$\Pr(F) = \frac{n-1}{2n-1} < 1 - \Pr(F) = \Pr(S).$$

In terms of probability you should be the second player.

For (3), since the opponent survived, the two bullets are randomly distributed among the remaining $n-1$ chambers. Since either bullet has the chance of $1/(n-1)$ being in the next chamber, the probability of you getting shot in the head without spinning is $2/(n-1)$. If we spin the barrel, the chance of you getting shot is $2/n$. Therefore, by spinning the barrel you get a higher chance of winning.

For (4), label the position of the two bullets as $0, 1$, and assume the barrel spins in the direction of increasing labels. Given that the opponent survived in the first round, you get shot in the next round if and only if the barrel started at position 5. Therefore, the probability of losing by not spinning is $1/4$. If we spin the barrel, the chance of getting shot is $2/6$. In terms of probability it is safer not to spin the barrel. □

4.6. **The candy lineup.** Here we demonstrate the technique of reducing the sample space when we care about the arrangements of only a subset of objects. This technique is particularly useful in dealing with expectations. See Example 5.16 and Example 5.17. The following example alludes to the sitcom series Seinfeld.

Example 4.25. Suppose there are 10 Twix, 20 Fifth Avenue bars, and 30 A Hundred Grand bars in a jar. A bar is taken out of the jar one at a time.

(1) What is the probability that A Hundred Grand bars run out before the Twix bars?

(2) What is the probability that the bars are depleted in the order of A Hundred Grand bars, The Fifth Avenue bars, and Twix?

(3) What is the probability that there are at least one Twix bar and one Fifth Avenue bar when all A hundred Grand bars are taken out?

solution. Since all orderings of the bars that are taken are equally likely, we can imagine these bars are lined up in a row. For (1), the positions of the Fifth Avenue bars are irrelevant. We only need to require that the last position is one of the Twix when permuting 10 Twix against 30 A Hundred Grand bars. The probability for (1) is therefore $10/40 = 1/4$.

For (2), we first require the last position is a Twix in the lineup among all bars. This has probability $10/60 = 1/6$. Next, given that a Twix is in the last position, the probability that A Hundred Grand Bars run out before the Fifth Avenue bars is $20/50 = 2/5$. Here the sample space is the 59 bars, among which we only care about how A Hundred Grand bars are permuted against the Fifth Avenue bars. Therefore, the probability for (2) is $1/6 \cdot 2/5 = 1/15$.

For (3), it is equivalent to require that the A Hundred Grand is the first to be depleted among the three types of bars. In this case, the order of depletion is either in the order in (2), or A Hundred Grand bars, Twix, and then the Fifth Avenue bars. The probability of the later is $2/6 \cdot 1/4 = 1/12$. Therefore, the probability of (3) is

$$1/12 + 1/15.$$

□

4.7. **Gambler's ruin.**

Example 4.26 (Gambler's ruin)**.** Two gamblers A and B, bet on the outcomes of successive flips of a coin. If the coin comes up with the head, then A receives 1 unit from B. Otherwise, B receives 1 unit from A. The coin flip continues until one of the gamblers runs out of money. The one who gets all the money is then declared as the victor. Assume the flips are independent, and the coin lands on head with probability p. What is the probability that A is the winner if A starts with i units of money and B starts with $n - i$ units?

solution. Let P_i be the probability that A wins if A starts with i units of money. We then immediately have

$$P_0 = 0, P_n = 1.$$

Conditioning on the outcome of the coin flip, we have

$$P_i = pP_{i+1} + qP_{i-1},$$

where $q = 1 - p$. Rewriting the left hand side as $pP_i + qP_i$, we get

$$p(P_{i+1} - P_i) = q(P_i - P_{i-1}).$$

This is equivalent to

$$P_{i+1} - P_i = \frac{q}{p}(P_i - P_{i-1}).$$

It is more convenient to work with the equations

$$P_i - P_{i-1} = \frac{q}{p}(P_{i-1} - P_{i-2}), i = 2, \ldots, n.$$

For each i, one gets

$$\begin{aligned}
P_i - P_{i-1} =& \frac{q}{p} \cdot (P_{i-1} - P_{i-2}) \\
=& \left(\frac{q}{p}\right)^2 \cdot (P_{i-2} - P_{i-3}) \\
=& \cdots \\
=& \left(\frac{q}{p}\right)^{i-1} \cdot (P_1 - P_0) = \left(\frac{q}{p}\right)^{i-1} \cdot P_1.
\end{aligned}$$

Summing up the above equations from 2 to i, we get

$$P_i - P_1 = \left(\frac{q}{p} + \cdots + \left(\frac{q}{p}\right)^{i-1}\right) \cdot P_1.$$

Therefore,

$$P_i = \left(1 + \frac{q}{p} + \cdots + \left(\frac{q}{p}\right)^{i-1}\right) \cdot P_1 = \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \frac{q}{p}} \cdot P_1.$$

Using $P_N = 1$, we get

$$P_1 = \frac{1 - \frac{q}{p}}{1 - \left(\frac{q}{p}\right)^N},$$

and that

$$P_i = \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^N}.$$

We note that when $p = q$, we get $P_i - P_{i-1} = P_{i-1} - P_{i-2}$, from which one easily derives

$$P_i = \frac{i}{n}.$$

In sum, we have

$$P_i = \begin{cases} \dfrac{1 - \left(\dfrac{q}{p}\right)^i}{1 - \left(\dfrac{q}{p}\right)^N} \text{ if } p \neq q, \\ \\ \dfrac{i}{n} \text{ if } p = q. \end{cases}$$

□

Example 4.27. Players 1,2,..., r compete in a contest, with player i starting at n_i units. In each stage, two players are randomly chosen. The winner receives one unit from the loser. Any player whose fortune drops to zero is eliminated. The contest continues until one player has all $n = \sum_i n_i$ units of money. Assume each player has a 1/2 chance of winning in each stage. What is the probability that the i-th player is the final victor?

solution. For player i to be the victor, they have to receive $n - n_i$ units from other players, irrespective of how others played. Therefore, we only need to consider two players, where player i starts at n_i and another virtual player starts at $n - n_i$. The probability that player i is the victor is then $\frac{n_i}{n}$. □

Example 4.28. An investor owns shares in a stock whose present value is 25. She decides that she must sell her stock if the value either drops down to 10 or rises above 40. If each change of price is either up 1 point with probability 0.55, or down 1 point with probability 0.45, and the successive changes are independent, what is the probability that she retires a winner?

solution. Let P_i be the probability that the investor retires a winner if her stock's present value is i, where $10 \leq i \leq 40$. We then have

$$P_{40} = 1, P_{10} = 0.$$

Letting $p = 0.55$ and $q = 0.45$, we get

$$P_i = pP_{i+1} + qP_{i-1}, i = 11 \ldots, 39.$$

Using the strategies we have seen in Example 4.26, we get

$$P_i = \frac{1 - \left(\frac{q}{p}\right)^{i-10}}{1 - \left(\frac{q}{p}\right)^{30}}.$$

In particular,

$$P_{25} = \frac{1 - \left(\frac{q}{p}\right)^{15}}{1 - \left(\frac{q}{p}\right)^{30}} \approx 0.953.$$

□

5. Examples of expectation

In this section we introduce properties of expectations, and apply these properties to numerous problems. While expectations and variances were introduced back in Definition 2.5, we develop further tools to compute them.

First, we recall how to compute the expectation of a function of a random variable in Corollary 5.3. We then recall conditional expectations and conditional variances, and that expectations and variances can be computed by conditioning in Proposition 5.5. Before delving into problems, we also cover a method to compute the moments of a random variable in Example 5.7. This technique is widely used throughout the rest of the note.

The problems presented in this section are categorized into coupon collections (Section 5.3), matches in random pairings (Section 5.4), and coin toss sequences (Section 5.5). Some of the expectations and variances presented in Section 3 are verified in Section 5.2. Finally, we apply all the techniques to trickier problems in Section 5.6.

5.1. Properties of expectation.

We begin with the following lemma, whose derivation is omitted as it comes down to exchanging the order of integration or summation:

Lemma 5.1. *Let X be a random variable either continuous with probability density function $f(x)$, or discrete. We have*

$$\begin{cases} \int_0^\infty xf(x)\,dx = \int_0^\infty \Pr(X > t)\,dt \text{ if } X \text{ is continuous,} \\ \\ \sum_{x>0} x\Pr(X = x) = \sum_{t=0}^\infty \Pr(X > t) \text{ if } X \text{ is discrete.} \end{cases}$$

Similarly,

$$\begin{cases} \int_{-\infty}^0 xf(x)\,dx = -\int_0^\infty \Pr(X < -t)\,dt \text{ if } X \text{ is continuous,} \\ \\ \sum_{x<0} x\Pr(X = x) = -\sum_{t=0}^\infty \Pr(X < -t) \text{ if } X \text{ is discrete.} \end{cases}$$

With Definition 2.5 and Lemma 5.1 the following is immediate.

Proposition 5.2. *Let X be a random variable, either continuous or discrete. We have*

$$E[X] = \begin{cases} \int_0^\infty \Pr(X > t)\,dt - \int_0^\infty \Pr(X < -t)\,dt \text{ if } X \text{ is continuous,} \\ \\ \sum_{t=0}^\infty \Pr(X > t) - \sum_{t=0}^\infty \Pr(X < -t) \text{ if } X \text{ is discrete.} \end{cases}$$

We now prove the following tool to compute the expectation of a function $g(X)$ of a random variable X. By Definition 2.5, we need to know the probability mass function or the probability density function of $g(X)$. Such function for $g(X)$ may be difficult to describe. A workaround is the

Corollary 5.3. *Let X be a random variable, either continuous with probability density function $f(x)$ or discrete. Then*

$$E[g(X)] = \begin{cases} \int_{\mathbf{R}} g(x)f(x)\,dx \text{ if } X \text{ is continuous,} \\ \\ \sum_{x\in\mathbf{Z}} g(x)\Pr(X = x) \text{ otherwise.} \end{cases}$$

Proof. For the continuous case,

$$E[g(X)] = \int_0^\infty \Pr(g(X) > t)\,dt - \int_0^\infty \Pr(g(X) < -t)\,dt$$

by Proposition 5.2. However,

$$\begin{aligned} \int_0^\infty \Pr(g(X) > t)\,dt &= \int_0^\infty \int_{g(x)>t} f(x)\,dx\,dt = \int_{g(x)>0} \int_0^{g(x)} f(x)\,dt\,dx \\ &= \int_{g(x)>0} g(x)f(x)\,dx. \end{aligned}$$

Similarly, we have $-\displaystyle\int_{-\infty}^0 \Pr(g(X) < -t)\,dt = \int_{g(x)<0} g(x)f(x)\,dx$. Hence

$$E[g(X)] = \int_{g(x)>0} g(x)f(x)\,dx + \int_{g(x)<0} g(x)f(x)\,dx = \int_{\mathbf{R}} g(x)f(x)\,dx.$$

The discrete case can be proved similarly. $\square$

Remark 5.4. The above proof can be extended to joint distributions (Section 2.2.2). For example, if X, Y are two continuous random variables with the joint probability density function $f(x, y)$, then

$$E[g(X,Y)] = \iint_{\mathbf{R}^2} g(x,y) f(x,y)\, dA.$$

In particular,

$$\begin{aligned}
E[X+Y] &= \iint_{\mathbf{R}^2} (x+y) f(x,y)\, dA \\
&= \iint_{\mathbf{R}^2} x f(x,y)\, dA + \iint_{\mathbf{R}^2} y f(x,y)\, dA \\
&= \int_{\mathbf{R}} \int_{\mathbf{R}} x f(x,y)\, dy\, dx + \int_{\mathbf{R}} \int_{\mathbf{R}} y f(x,y)\, dx\, dy \\
&= \int_{\mathbf{R}} x \int_{\mathbf{R}} f(x,y)\, dy\, dx + \int_{\mathbf{R}} y \int_{\mathbf{R}} f(x,y)\, dx\, dy \\
&= \int_{\mathbf{R}} x f_X(x)\, dx + \int_{\mathbf{R}} y f_Y(y)\, dy \\
&= E[X] + E[Y].
\end{aligned}$$

Here f_X (resp. f_Y) denotes the marginal density function of X (resp. Y).

We now recall how to compute the expectation and the variance of a random variable by conditioning.

Proposition 5.5. *Let X and Y be two random variables. For simplicity, assume they are either both discrete or continuous. Define $E[X|Y]$ as the function of the random variable Y, where*

$$\begin{aligned}
E[X|Y](y) :=& E[X|Y=y] \\
:=& \begin{cases} \sum_x x \cdot \Pr(X=x|Y=y) \ \textit{if } X, Y \textit{ are discrete,} \\ \\ \int_{\mathbf{R}} x \cdot f_{X|Y}(x,y)\, dx \ \textit{if } X, Y \textit{ are continuous.} \end{cases}
\end{aligned}$$

If needed, $f_{X|Y}$ is introduced back in Definition 2.18.

We then define

$$\operatorname{Var}(X|Y) := E[X^2|Y] - (E[X|Y])^2.$$

The expectation and variance satisfy the following properties:

(1) $E[X+Y] = E[X] + E[Y]$ *(true regardless of dependency between X and Y).*
(2) $E[E[X|Y]] = E[X]$.
(3) $\operatorname{Var}(X) = E[\operatorname{Var}(X|Y)] + \operatorname{Var}(E[X|Y])$.

Proof. Property (1) is established in Remark 5.4. Property (2) is due to the fact that probabilities can be computed by conditioning (Remark 2.20). Let us prove (3).

For this, note that by definition

$$\operatorname{Var}(X|Y) = E[X^2|Y] - (E[X|Y])^2.$$

Therefore,

$$E[\operatorname{Var}(X|Y)] = E[E[X^2|Y]] - E[(E[X|Y])^2] = E[X^2] - E[(E[X|Y])^2].$$

On the other hand,

$$\operatorname{Var}(E[X|Y]) = E[(E[X|Y])^2] - (E[E[X|Y]])^2 = E[(E[X|Y])^2] - E([X])^2.$$

Adding up, we get the desired result. $\square$

Remark 5.6. The properties of expectation we have seen in Remark 5.4 also apply to conditional expectations. For example, if X, Y are discrete with conditional probability mass function $p_{X|Y}$, then

$$E[g(X)|Y = y] = \sum_{x \in \mathbf{Z}} g(x) p_{X|Y}(x, y).$$

Similarly, if X, Y are continuous with conditional probability density function $f_{X|Y}$, then

$$E[g(X)|Y = y] = \int_{\mathbf{R}} g(x) f_{X|Y}(x, y)\ dx.$$

We are now ready to apply the properties of expectations to problems. We begin with a technique that is frequently used.

Example 5.7 (A technique to compute moments)**.** In many of the examples that follow, we calculate the expectation and variance of a random variable by expressing it as a sum of indicators. Let $X_1, X_2, \ldots, X_n$ be n indicators for n the events $E_1, E_2 \ldots, E_n$ respectively. Namely,

$$X_i = \begin{cases} 1 \text{ if } E_i \text{ occurs,} \\ 0 \text{ otherwise.} \end{cases}$$

The sum $X = X_1 + X_2 + \cdots + X_n$ is then the number of these events that occur. By (1) from Proposition 5.5, we have

$$E[X] = \sum_i E[X_i] = \sum_i \Pr(X_i = 1) = \sum_i \Pr(E_i).$$

To calculate the second moment $E[X^2]$, we consider the number of pairs of the n events that occur. For each $i < j$, we define the indicator $X_{ij} = X_i X_j$, where

$$X_{ij} = \begin{cases} 1 \text{ if } E_i, E_j \text{ occur,} \\ 0 \text{ otherwise.} \end{cases}$$

The sum $\sum_{i<j} X_{ij}$ is then the number of pairs of events that occur. On the other hand, this in terms of X is precisely $\binom{X}{2}$. Therefore, we have

$$\begin{aligned} E\left[\binom{X}{2}\right] =& E\left[\sum_{i<j} X_{ij}\right] \\ =& \sum_{i<j} E[X_{ij}] = \sum_{i<j} \Pr(E_i E_j). \end{aligned}$$

One can obtain $E[X^2]$ via the following. First, compute for each pair $i < j$ the probability $\Pr(E_i E_j)$ and take the sum $\sum_{i<j} \Pr(E_i E_j)$. Next, use

$$E\left[\binom{X}{2}\right] = E\left[\frac{1}{2} \cdot \left(X^2 - X\right)\right] = \frac{1}{2} \cdot \left(E[X^2] - E[X]\right)$$

to solve for

$$\begin{aligned} E[X^2] =& 2 \cdot E\left[\binom{X}{2}\right] + E[X] \\ =& 2 \cdot \sum_{i<j} \Pr(E_i E_j) + \sum_i \Pr(E_i). \end{aligned}$$

From here one obtains also the variance $\mathrm{Var}(X) = E[X^2] - (E[X])^2$. One may obtain higher moments of X by considering triplets of the events and so on.

We now provide a quick example for the technique we just discussed.

Example 5.8. An algorithm selects the maximum from a list of distinct values in the following way: The algorithm reads the list linearly. The algorithm keeps the value until a bigger value is read. When the whole list has been read, the value the algorithm holds is the maximum. Find the expected value and the variance of the number of times a new value is assigned in the algorithm, where the reading of the first entry in the list is counted as an assignment as well.

solution. For each $i = 1, \ldots, 100$, define the random variable X_i via the following. We declare $X_i = 1$ if the i-th value in the list is the biggest among the first i values. Otherwise, we set $X_i = 0$. For example, if the first three values of the list are 13, 9, 22, we then have $X_1 = 1$, $X_2 = 0$, $X_3 = 1$.

The sum $X = \sum_{i=1}^{100} X_i$ is then the number of assignments of new values in the algorithm. The expectation in question is therefore

$$E[X] = \sum_{i=1}^{100} E[X_i] = \sum_{i=1}^{100} \Pr(X_i = 1).$$

We now simply note that in a permutation of i values, the probability that the maximum is in the last position is $1/i$. Therefore, $\Pr(X_i = 1) = 1/i$ so that

$$E[X] = \sum_{i=1}^{100} \frac{1}{i}.$$

To calculate the variance, let us consider $\Pr(X_i, X_j = 1)$ for any $i < j$. For this, we use $\Pr(X_i, X_j = 1) = \Pr(X_j = 1)\Pr(X_i = 1|X_j = 1)$. As was discussed, $\Pr(X_j = 1) = 1/j$. Given this, we have now a list of $j - 1$ values. For the i-th value to be the biggest among the first i values, we have $\Pr(X_i = 1|X_j = 1) = 1/i$. Therefore,

$$\Pr(X_i, X_j = 1) = \frac{1}{ij} = \Pr(X_i = 1)\Pr(X_j = 1).$$

This shows that X_i and X_j are independent. By Proposition 2.16,

$$\begin{aligned} \mathrm{Var}(X) =& \sum_{i=1}^{100} \mathrm{Var}(X_i) = \sum_{i=1}^{100} \Pr(X_i = 1) - \Pr(X_i = 1)^2 \\ =& \sum_{i=1}^{100} \left(\frac{1}{i} - \frac{1}{i^2}\right). \end{aligned}$$

□

In the above example, while it was straightforward to compute $\Pr(X_i = 1|X_j = 1)$, let us also try to compute $\Pr(X_j = 1|X_i = 1)$ directly, without the knowledge that $\Pr(X_i, X_j = 1) = 1/(ij)$. Let S be the value that position i holds.

For $X_i = 1$, we must have $i \leq S \leq j$. When $S \neq j$ and $X_i = 1$, the value j can only be in positions $i+1, i+2, \ldots, j$. Therefore,

$$\Pr(X_j = 1 | X_i = 1, S = k) = \frac{1}{j-i} \text{ for } k < j.$$

When $S = j$, we get $X_j = 0$. On the other hand, we get $X_i = 1$ so that

$$\Pr(S = j | X_i = 1) = \frac{\Pr(S = j)}{\Pr(X_i = 1)} = \frac{i}{j}.$$

With the information we have obtained so far, we have

$$\begin{aligned}
\Pr(X_j = 1 | X_i = 1) &= \sum_{k=i}^{j} \Pr(S = k | X_i = 1) \Pr(X_j = 1 | X_i = 1, S = k) \\
&= \sum_{k=i}^{j-1} \Pr(S = k | X_i = 1) \Pr(X_j = 1 | X_i = 1, S = k) \\
&= \sum_{k=i}^{j-1} \Pr(S = k | X_i = 1) \frac{1}{j-i} \\
&= (1 - \Pr(S = j | X_i = 1)) \cdot \frac{1}{j-i} \\
&= (1 - \frac{i}{j}) \cdot \frac{1}{j-i} = \frac{1}{j}.
\end{aligned}$$

Hence,

$$\Pr(X_i, X_i = 1) = \Pr(X_i = 1) \Pr(X_j = 1 | X_i = 1) = \frac{1}{ij},$$

agreeing with what we had.

Here is a classic example of finding the expectation by conditioning (property (2) of Proposition 5.5).

Example 5.9. A miner is trapped in a mine containing 3 doors. The first door leads to a tunnel that will take him to safety after 3 hours of travel. The second door leads to a tunnel that will take him back to the mine after 5 hours of travel. The third door leads to a tunnel that loops back to the mine after 7 hours of travel. Due to darkness, the miner could only choose the doors randomly. Find the expected value and the variance of the length of time until he reaches safety.

solution. Let X be the length of time until the miner reaches safety. Let us condition on the door chosen. Specifically, we define $N = i$ if the i-th door is picked by the miner. Property (2) of Proposition 5.5 then implies that

$$\begin{aligned}
E[X] &= E[E[X|N]] \\
&= \frac{1}{3} \cdot (E[X|N=1] + E[X|N=2] + E[X|N=3]) \\
&= \frac{1}{3} \cdot (3 + 5 + E[X] + 7 + E[X]).
\end{aligned}$$

One then easily solves that

$$E[X] = 15.$$

For the variance, we can compute $E[X^2]$ also by conditioning. Since each selection is independent of the previous ones, we have

$$\Pr(X = x | N = 2) = \Pr(X = x - 5).$$

Similarly,

$$\Pr(X = x|N = 3) = \Pr(X = x - 7).$$

Therefore,

$$\begin{aligned} E[X^2|N = 2] =& \sum_x x^2 \Pr(X = x|N = 2) = \sum_x x^2 \Pr(X = x - 5) \\ =& \sum_x x^2 \Pr(X + 5 = x) = E[(X + 5)^2] \\ =& E[X^2] + 10E[X] + 25. \end{aligned}$$

Similarly,

$$E[X^2|N = 3] = E[X^2] + 14E[X] + 49$$

so that

$$\begin{aligned} E[X^2] =& E[E[X^2|N]] \\ =& \frac{1}{3}(3^2 + E[X^2] + 10E[X] + 25 + E[X^2] + 14E[X] + 49). \end{aligned}$$

From here $E[X^2]$ and therefore $\mathrm{Var}(X)$ can be solved. The answer is 218. □

We also apply property (3) of Proposition 5.5 to a problem from the Society of Actuary probability exam:

Example 5.10. A driver made three errors. Each error independently results in an accident with probability 0.25. Each accident results independently a loss exponentially distributed with mean 0.8. Suppose the loss is independent from the number of accidents caused by these errors. Find the variance of the loss.

solution. Let N be the number of accidents occurred. Then N is binomial with $n = 3$ and $p = 0.25$. Therefore,

$$E[N] = 3 \cdot 0.25$$

and

$$\mathrm{Var}(N) = 3 \cdot 0.25 \cdot 0.75.$$

Let L_i, $i = 1, \ldots, N$ be the loss of each accident. We want to find $\mathrm{Var}(\sum_{i=1}^N L_i)$.

By property (4) of Proposition 5.5,

$$\mathrm{Var}\left(\sum_{i=1}^N L_i\right) = E\left[\mathrm{Var}\left(\sum_{i=1}^N L_i|N\right)\right] + \mathrm{Var}\left(E\left[\sum_{i=1}^N L_i|N\right]\right),$$

where

$$\begin{aligned} E\left[\sum_{i=1}^N L_i|N = n\right] =& E\left[\sum_{i=1}^n L_i|N = n\right] \\ =& \sum_{i=1}^n E\left[L_i|N = n\right] = \sum_{i=1}^n E[L_i] \end{aligned}$$

by independence of L_i and N. Since each L_i is exponential with mean 0.8, we get $\sum_{i=1}^n E[L_i] = 0.8n$, implying

$$E\left[\sum_i^N L_i|N\right] = 0.8N.$$

Hence

$$\operatorname{Var}\left(E\left[\sum_{i=1}^{N} L_i|N\right]\right) = 0.64 \cdot \operatorname{Var}(N) = 0.64 \cdot 3 \cdot 0.25 \cdot 0.75.$$

By independence among L_i and independence of each L_i from N, we have

$$\begin{aligned}\operatorname{Var}\left(\sum_{i=1}^{N} L_i \middle| N = n\right) =& \operatorname{Var}\left(\sum_{i=1}^{n} L_i|N = n\right) = \sum_{i=1}^{n} \operatorname{Var}(L_i)\\ =&0.64n.\end{aligned}$$

This implies $\operatorname{Var}\left(\sum_{i=1}^{N} L_i|N\right) = 0.64N$ so that

$$E\left[\operatorname{Var}\left(\sum_{i=1}^{N} L_i|N\right)\right] = 0.64 \cdot E[N] = 0.64 \cdot 3 \cdot 0.25.$$

Summing these two up, we get

$$\operatorname{Var}\left(\sum_{i=1}^{N} L_i\right) = 0.64 \cdot 3 \cdot 0.25 \cdot 1.75.$$

□

5.2. **The expectations and variances of common random variables.** In this section, we compute the expectations and the variances of the following random variables:

(1) The geometric random variable in Example 5.11;
(2) The negative binomial random variable in Example 5.12;
(3) The hypergeometric random variable in Example 5.13;
(4) The negative hypergeometric random variable in Example 5.16.

Example 5.11 (Geometric random variable)**.** Independent trials, each resulting in a success with probability p, are successively performed. Let N be the time of the first success. Find $E[N]$ and $\operatorname{Var}(N)$.

solution. Define the random variable

$$X = \begin{cases} 1 \text{ if the first trial is a success,} \\ 0 \text{ otherwise.} \end{cases}$$

We will compute $E[N]$ and $E[N^2]$ using (2) from Proposition 5.5, where we condition on X. More generally, it is clear that $E[N^k|X = 1] = 1$ for any k. For $E[N^k|X = 0]$, we have

$$E[N^k|X = 0] = \sum_i i^k \cdot \Pr(N = i|X = 0)$$

by Remark 5.6. By independence of each trial, we also have

$$\Pr(N = i|X = 0) = \Pr(N = i - 1).$$

Therefore,

$$\begin{aligned}E[N^k|X = 0] =& \sum_i i^k \cdot \Pr(N = i - 1) = \sum_i i^k \cdot \Pr(N + 1 = i)\\ =&E[(N + 1)^k].\end{aligned} \tag{5.1}$$

In particular,

$$E[N|X=0]=E[N+1]=E[N]+1.$$

Intuitively, if the first trial fails, we are back to the initial state, plus an additional step used. We now get

$$\begin{aligned}E[N] =&E[E[N|X]]\\ =&p(E[N|X=1])+(1-p)E([N|X=0])\\ =&p+(1-p)(1+E[N]).\end{aligned}$$

Solving the above equation for $E[N]$, we get

$$E[N]=\frac{1}{p}.$$

For $\mathrm{Var}(N)$, first note

$$\begin{aligned}E[N^2] =&E[E[N^2|X]]\\ =&pE[N^2|X=1]+(1-p)E[N^2|X=0]\\ =&p+(1-p)E[(N+1)^2]\\ =&p+(1-p)(E[N^2]+2E[N]+1)=p+(1-p)(E[N^2]+\frac{2}{p}+1).\end{aligned}$$

We then get

$$E[N^2]=\frac{2-p}{p^2}.$$

Therefore,

$$\mathrm{Var}(N)=E[N^2]-(E[N])^2=\frac{1-p}{p^2}.$$

□

Example 5.12 (Negative binomial random variable)**.** Independent trials, each resulting in a success with probability p, are successively performed. Let N be the time of the first r successes. Find $E[N]$ and $\mathrm{Var}(N)$.

solution. Let X_i be the number of trials to obtain the i-th success after the $(i-1)$-th success, where $1\le i\le r$. Each X_i is then a geometric random variable with probability p. Since $N=\sum_i X_i$,

$$E[N]=\sum_i E[X_i]=\frac{r}{p}.$$

Since $X_1,\dots,X_r$ are independent, we get

$$\mathrm{Var}(N)=\sum_i \mathrm{Var}(X_i)=r\cdot\frac{1-p}{p^2}.$$

□

Example 5.13 (Hypergeometric random variable)**.** In a collection of N balls, m are special, and the other $N-m$ are ordinary. Let X be the number of special balls selected when a selection of n balls without replacement is made. Find the expectation and the variance of X.

solution. Define the random variable X_i, one for each special ball, where $X_i = 1$ if the i-th special ball is selected. Otherwise, $X_i = 0$. We then have $X = X_1 + \cdots + X_m$. Since each special ball has a probability of n/N of being selected, the expectation of X is mn/N.

The variance can be found by considering pairs of special balls that are selected (see Example 5.7). For any $i < j$, the probability that i, j-th balls are selected is

$$\frac{n}{N} \cdot \frac{n-1}{N-1}.$$

Hence, the expected number of pairs of special balls selected is

$$E\left[\binom{X}{2}\right] = \binom{m}{2} \frac{n(n-1)}{N(N-1)}.$$

This implies

$$E[X^2 - X] = E[X^2] - E[X] = \frac{mn(m-1)(n-1)}{N(N-1)}.$$

Therefore,

$$\begin{aligned}\mathrm{Var}(X) =& E[X^2] - (E[X])^2 \\ =& \frac{mn(m-1)(n-1)}{N(N-1)} + \frac{mn}{N} - \left(\frac{mn}{N}\right)^2.\end{aligned}$$

□

Example 5.14. Let $\mathbf{X} = \langle X_1, X_2, \ldots, X_n\rangle$ and $\mathbf{Y} = \langle Y_1, Y_2, \ldots, Y_n\rangle$ be two independent orderings of k ones and $n-k$ zeros. Let M be the number of i's such that $X_i = 1$ but $Y_i = 0$. Describe the distribution of M, and find its expected value.

solution. We fix $\mathbf{X}$ and arrange $\mathbf{Y}$ accordingly for $M = m$ to hold. Note that to arrange $\mathbf{Y}$, it is sufficient to choose which components to insert the k ones.

We first disperse $k - m$ ones of $\mathbf{Y}$ to the components where $\mathbf{X}$ is one. We then disperse the remaining m ones to the components where $\mathbf{X}$ is zero. Therefore,

$$\Pr(M = m) = \frac{\binom{k}{k-m}\binom{n-k}{m}}{\binom{n}{k}}.$$

It follows that M is the hypergeometric distribution, where there are $n-k$ special and k ordinary balls. The probability that $M = m$ is precisely the probability that m special balls are chosen when k selections are made.

To further make sense of this, given a vector $\mathbf{X}$, a choice to insert a one for $\mathbf{Y}$ among the $n-k$ positions where $X_i = 0$ results in an increment of M. Therefore, the $n-k$ positions where $X_i = 0$ are special. On the other hand, inserting ones in the k positions where $X_i = 1$ does not change the value of M. Therefore, those k positions can be thought of as ordinary. By Example 5.13,

$$E[M] = \frac{(n-k)k}{n}.$$

□

Remark 5.15. In the above example, one does not need to interpret M as a hypergeometric random variable to find its expected value. The probability that

$X_i = 1$ but $Y_i = 0$ is just $\frac{k}{n} \cdot \frac{n-k}{n}$. Since there are n positions, the expected value of M is

$$n \cdot \frac{k}{n} \cdot \frac{n-k}{n} = \frac{k(n-k)}{n}.$$

Moreover, if N is the number of positions where $X_i \neq Y_i$, then $N = 2M$. Therefore, $E[N] = 2E[M]$.

Example 5.16 (Negative hypergeometric random variable)**.** A ball is selected one at a time without replacement from a collection of N balls. Among them m balls are special, and the other $N - m$ are ordinary. Let X be the number of selections needed in order to obtain r special balls. Find the expectation and the variance of X.

solution. Define a random variable X_i for each ordinary ball, where we declare $X_i = 1$ if the ball is selected before r special balls. We then have $X = \sum_i X_i + r$ so that

$$E[X] = \sum_i E[X_i] + r.$$

Now for a particular ordinary ball to be selected before r special balls, it is sufficient to require this ordinary ball to be in the first r positions when lined up with all special balls. Therefore, $E[X_i] = \frac{r}{m+1}$ so that

$$E[X] = (N - m)\frac{r}{m+1} + r.$$

For the variance, first note that if $Z = \sum_i X_i$, then $\mathrm{Var}(X) = \mathrm{Var}(Z)$. Using Example 5.7, we consider the pairs of ordinary balls selected before r special balls. For any two balls to be selected before r special balls, it is sufficient to require they be in the first $r + 1$ positions when lined up against all special balls. Therefore, the probability is

$$\frac{(r+1)r}{(m+2)(m+1)}.$$

We then get

$$E\left[\binom{Z}{2}\right] = \binom{N-m}{2}\frac{(r+1)r}{(m+2)(m+1)}.$$

It follows that

$$E[Z^2] - E[Z] = (N-m)(N-m-1)\frac{(r+1)r}{(m+2)(m+1)}.$$

From here, the variance of Z (and therefore the variance of X) can be derived. □

Example 5.17. What is the expected number of turnovers needed in a regular 52-card deck in order to obtain the first ace?

solution. The number X of turnovers to see the first ace is the negative hypergeometric random variable, where $m = 4$ (4 ace cards are special), $n = 52$, and $r = 1$. We can also solve this problem by directly defining X_i for each ordinary card.

We declare $X_i = 1$ if the card is turned over prior to all aces. We then have

$$\Pr(X_i = 1) = \frac{1}{5}.$$

The number of turnovers in question is then

$$X = 1 + \sum_i X_i$$

so that

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

□

Example 5.18. Let X be the minimum of k-distinct values selected from $1, 2, \ldots, n$. Find the expected value of X.

solution. There is an interpretation of X as a negative hypergeometric random variable. First, i is the minimum if and only if it is selected, and other $k-1$ values are selected among $i+1, i+2, \ldots, n$. Therefore,

$$\begin{aligned}
\Pr(X = i) =& \frac{\binom{n-i}{k-1}}{\binom{n}{k}} = \frac{\frac{(n-i)!}{(k-1)!(n-i-k+1)!}}{\frac{n!}{k!(n-k)!}} \\
=& k \cdot \frac{\frac{(n-k)!}{(n-i-k+1)!}}{\frac{n!}{(n-i)!}} = \frac{k}{i} \cdot \frac{\binom{n-k}{i-1}}{\binom{n}{i}} \\
=& \frac{k}{n-i+1} \cdot \frac{\binom{n-k}{i-1}}{\binom{n}{i-1}}.
\end{aligned}$$

This is the probability to get a special ball using i selections without replacements, where there are k special balls among n balls. The expected value of X is therefore

$$1 + \frac{n-k}{k+1} = \frac{n+1}{k+1}$$

by Example 5.16.

If the reader finds it difficult to relate X to the negative hypergeometric random variable we just described, let us reinterpret the selection process. Imagine your friend inspecting what numbers you chose. Your friend reads out one number at a time from 1 to n. Each time your friend asks you if you took this number. The process continues until k numbers are confirmed (or $n-k$ numbers are rejected). The first number you say yes to is the smallest among the k values. □

Example 5.19 (Expected length of initial runs)**.** Find the expected length of initial runs when n ones and m zeros are lined up at random. For example, if the first 5 digits are $1, 1, 1, 0, 0$, the length of the initial run is 3. If the first five digits are $0, 0, 1, 0, 1$, the initial run has length 2.

solution. Let X be the length of the initial run. We can condition $E[X]$ on the first digit. If the first digit is 0, then the initial run is the number of selections needed to obtain 1. This is negative hypergeometric in the sense that there are $n+m-1$ balls in an urn, among which n are special. We want the number of selections without replacements to get the first special ball. The expected number of selections needed is

$$1 + \frac{m-1}{n+1} = \frac{m+n}{n+1}$$

by Example 5.16. One can easily verify that

$$E[X] = \frac{m}{m+n} \cdot \frac{m+n}{n+1} + \frac{n}{m+n} \cdot \frac{m+n}{m+1} = \frac{m}{n+1} + \frac{n}{m+1}.$$

□

5.3. **Coupon collections.** Many properties of expectation are best demonstrated by the coupon collection problems.

Example 5.20. Suppose there are n types of coupons. Assume each selection is independent of the previous ones, and each selection is equally likely to be any of the n types. Find the expected value and the variance of the number of selections needed to amass all types of coupons.

solution. Let X_i be the number of selections needed to obtain a new i-th type after $i-1$ many types have been collected. Then $X = X_1 + X_2 + \cdots X_n$ is the number of selections until all types of coupons have been collected. Since each X_i is a geometric random variable with probability $\frac{n-i+1}{n}$, we get $E[X_i] = \dfrac{n}{n-i+1}$ so that

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

For $\mathrm{Var}(X)$, simply note that X_i and X_j are independent for all $i < j$. Therefore,

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i),$$

where

$$\mathrm{Var}(X_i) = \frac{n(i-1)}{(n-i+1)^2}$$

by Example 5.11. □

Example 5.21 (Singletons in coupon collections)**.** Suppose there are n types of coupons. Assume each selection is independent of the previous ones, and each selection is equally likely to be any of the n types. Let the collection continue until all n types have been collected. Find the expected value and the variance of the number of types that are singletons in the collection.

solution. Let X_i be the indicator for the i-th type of coupon selected in the process. We declare $X_i = 1$ if that type is a singleton and we declare $X_i = 0$ otherwise. We need to find $\Pr(X_i = 1)$ by Example 5.7.

For this, we let E^i_j, $i \le j \le n-1$ be the event that type i stays a singleton as the number of types selected reaches from j to $j+1$. Event E^i_j then occurs when the first selection after collecting j types is a new $(j+1)$-th type. Or, a coupon from types $1, 2, \ldots, i-1, i+1, \ldots, j$ is first selected, then a new type and so on.

Therefore,

$$\Pr(E^i_j) = \frac{n-j}{n} + \frac{j-1}{n} \cdot \frac{n-j}{n} + (\frac{j-1}{n})^2 \cdot \frac{n-j}{n} + \cdots .$$

This is a geometric series with common ratio $\frac{j-1}{n}$ so

$$\Pr(E^i_j) = \frac{n-j}{n-j+1}.$$

For type i to remain as a singleton, we require a type i coupon is never selected again as the number of types selected goes from i to n. By independence of each selection, we get

$$\Pr(X_i = 1) = \prod_{j=i}^{n-1} \Pr(E^i_j) = \frac{1}{n-i+1}.$$

It follows that

$$E[X] = \sum_{i=1}^{n} \frac{1}{n-i+1} = \sum_{i=1}^{n} \frac{1}{i}.$$

Next, to find the variance, we first find $\Pr(X_i = X_j = 1)$ for all $i < j$ by Example 5.7. The probability can be obtained by requiring type i stays a singleton till j types are selected, and from there both type i and type j stay singletons till all types have been collected. A similar calculation, albeit tedious, shows that

$$\Pr(X_i = X_j = 1) = \frac{1}{n-i+1} \cdot \frac{2}{n-j+2}.$$

This yields

$$E[X^2] - E[X] = \sum_{i<j} \frac{1}{n-i+1} \cdot \frac{4}{n-j+2}.$$

From here the variance can be easily deduced. □

We now deal with types of coupons whose probabilities to be selected are possibly different. The following is going to be useful.

Proposition 5.22 (The maximum-minimum identity)**.** *For arbitrary numbers $x_1, x_2, \ldots, x_n$,*

$$\max_i x_i = \sum_i x_i - \sum_{i<j} \min(x_i, x_j) + \sum_{i<j<k} \min(x_i, x_j, x_k) + \cdots + (-1)^{n+1} \min(x_1, \ldots, x_n).$$

Proof. See Proposition 2.2, chapter 7 from [Ros14]. □

Remark 5.23. We note that there is the minimum-maximum identity

$$\min_i x_i = \sum_i x_i - \sum_{i<j} \max(x_i, x_j) + \cdots + (-1)^{n+1} \max(x_1, \ldots, x_n).$$

As an application, if there are random variables $X_1, \ldots, X_n$, then $X = \max(X_1, \ldots, X_n)$ satisfies

$$E[X] = \sum_i E[X_i] - \sum_{i<j} E[\min(X_i, X_j)] + \cdots + (-1)^{n+1} E[\min(X_1, \ldots, X_n)]. \tag{5.2}$$

Example 5.24. Suppose there are N types of coupon and that, independently of past types selected, each new one obtained is type j with probability p_j, $\sum_{j=1}^{N} p_j = 1$. Find the expected number of selections needed to amass all types.

solution. Let X_i be the number of selections needed to obtain the i-th type. In this case,

$$X = \max(X_1, \ldots, X_N)$$

is the number of selections to amass all types.

Each X_i geometric with probability p_i. Hence $E[X_i] = 1/p_i$. Moreover, $\min(X_{i_1}, \ldots, X_{i_j})$ is the number of selections needed to obtain either a type $i_1, i_2, \ldots i_j$ coupon. Therefore, $\min(X_{i_1}, \ldots, X_{i_j})$ is geometric with probability $p_{i_1} + \cdots + p_{i_j}$. By Equation (5.2),

$$E[X] = \sum_i \frac{1}{p_i} - \sum_{i<j} \frac{1}{p_i + p_j} + \sum_{i<j<k} \frac{1}{p_i + p_j + p_k} + \cdots + (-1)^{N+1} \frac{1}{p_1 + \cdots + p_N}.$$

□

Example 5.25. Suppose there are N types of coupon and that, independently of past types selected, each new one obtained is type j with probability p_j, $\sum_{j=1}^{N} p_j = 1$. Find the expected value and variance of the number of types of coupons selected among the first n selections.

solution. It is cleaner to work with the opposite indicator. We declare $X_i = 1$ if type i is not among the first n selections, and we declare $X_i = 0$ otherwise. Then

$$Y = N - \sum_{i=1}^{N} X_i$$

is the number of types selected.

Since

$$E[X_i] = \Pr(X_i = 1) = (1 - p_i)^n,$$

we get

$$E[Y] = N - \sum_{i=1}^{N} (1 - p_i)^n.$$

To find $\mathrm{Var}(Y) = \mathrm{Var}(X)$, we use from Example 5.7 that

$$E[X^2] - E[X] = 2 \sum_{i<j} \Pr(X_i = X_j = 1).$$

Therefore,

$$\begin{aligned}\mathrm{Var}(X) =& 2 \sum_{i<j} \Pr(X_i = X_j = 1) + E[X] - (E[X])^2 \\ =& 2 \sum_{i<j} (1 - p_i - p_j)^n + \sum_{i=1}^{N} (1 - p_i)^n - \big(\sum_{i=1}^{N} (1 - p_i)^n \big)^2.\end{aligned}$$

□

Example 5.26. There are 4 different types of coupons, the first 2 comprise one group and the second 2 another group. Each new coupon obtained is type i with probability p_i, where $p_1 = p_2 = 1/8$, $p_3 = p_4 = 3/8$. Find the expected number of coupons one must obtain to have at least one of

(1) all 4 types;
(2) all types of the first group;
(3) all types of the second group;
(4) all types of either group.

solution. Let X_i, $i = 1, \ldots, 4$ be the number of selections needed to obtain a type i coupon. Each X_i is then geometric with probability p_i. By Example 5.24, the answer to (1) is

$$2 \times 8 + 2 \times 8/3 - (4 + 2 \times 4 + 4/3) + 8/5 \times 2 + 8/7 \times 2 - 1 = 437/35.$$

(2) and (3) are also direct consequences of Example 5.24. The answers for (2), (3) respectively are 12 and 4. For (4), we let Y (resp. Z) be the number of selections needed to obtain all types in group 1 (resp. group 2). Then $\min(Y, Z)$ is the number of selections needed to obtain all types in either group. By the minimum-maximum identity,

$$E[\min(Y, Z)] = E[Y] + E[Z] - E[\max(Y, Z)].$$

Therefore, the answer to (4) is simply $12+4-\frac{437}{35}$. □

There is another solution to (4) using Proposition 5.2.

solution 2 to (4). Let N (resp. M) be the number of selections needed to obtain all types in either group (resp. both groups). Using the same Y and Z from above, we have

- Both $Y, Z > t$ if and only if $N > t$;
- Either $Y > t$ or $Z > t$ if and only if $M > t$.

By Proposition 5.2, we get

$$\begin{aligned}
E[N] =& \sum_{t\geq 0} \Pr(N > t) \\
=& \sum_{t\geq 0} \Pr(Y \& Z > t) \\
=& \sum_{t\geq 0} \Pr(Y > t) + \Pr(Z > t) - \Pr(Y \text{ or } Z > t) \\
=& \sum_{t\geq 0} \Pr(Y > t) + \Pr(Z > t) - \Pr(M > t) \\
=& E[Y] + E[Z] - E[M].
\end{aligned}$$

□

Example 5.27. Suppose there are N types of coupon and that, independently of past types selected, each new one obtained is type j with probability p_j, $\sum_{j=1}^{N} p_j = 1$. Let the selection continue until all types of coupons have been selected. Find the expected number of type i coupons selected. In the case that all $p_i = 1/N$, find the variance of the number of type i coupons selected.

solution. This is another good place to apply Proposition 5.5. We let M be the number of selections needed to amass all types of coupons, and we let X_i be the number of type i coupons selected. We can find $E[X_i]$ and $\operatorname{Var}(X_i)$ by conditioning on M.

First, when $M = m$, X_i is binomial with parameters m, p_i. Hence $E[X_i|M = m] = mp_i$ and $\operatorname{Var}(X_i|M = m) = mp_i(1-p_i)$. Therefore,

$$E[X_i] = E[E[X_i|M] = p_i E[M],$$

$$E[\operatorname{Var}(X_i|M)] = p_i(1-p_i)E[M],$$

and

$$\operatorname{Var}(E[X_i|M]) = p_i^2 \operatorname{Var}(M).$$

Therefore $E[X_i]$ can be derived from Example 5.24, and $\operatorname{Var}(X_i)$ can be derived from Example 5.20. □

5.4. **Expected matches.**

Example 5.28. There are n men, each carries a hat unique to himself. Hats are shuffled and redistributed randomly to the men. Find the expected value and the variance of the number of men who have their own hats. Compare Example 4.5.

solution. Let $X_1, \dots, X_n$ be the indicators, one for each man. We declare $X_i = 1$ if the i-th man gets his hat. Otherwise, $X_i = 0$. The sum $X = \sum_i X_i$ is the number of men getting their own hats.

Obviously $\Pr(X_i) = \frac{1}{n}$, and $\Pr(X_i = X_j = 1) = \frac{1}{n} \cdot \frac{1}{n-1}$. By Example 5.7 one deduces that

- $E[X] = 1$;
- $E[\binom{X}{2}] = \binom{n}{2} \cdot \frac{1}{n} \cdot \frac{1}{n-1} = \frac{1}{2}$.

It follows that $\mathrm{Var}(X) = 1$ also. □

Example 5.29. A set of 1000 cards numbered 1 through 1000 are distributed randomly to 1000 people. Find the expected number of persons whose age matches the number on the card.

solution. Each person has the probability of 1/1000 to get the card whose number matches their age. Therefore, the expected value is 1. □

Example 5.30. A group of n men and n women are lined up at random.

(1) Find the expected number of men who have a woman next to them.
(2) Repeat (1), but now assuming they are seated at a round table.

Proof. Let $X_1, \dots, X_n$ be the indicators, one for each man. We declare $X_i = 1$ if there is a woman next to the i-th man. Otherwise, $X_i = 0$. The sum $X = \sum_i X_i$ is the number of men having a woman next to them.

We need to find $\Pr(X_i = 1)$ by Example 5.7. For both problems, it is easier to find $\Pr(X_i = 0)$ first. For (1), we can condition on whether the i-th man is in the middle, at the beginning, or at the end of the line. If the man is in the middle, then we require the person in front of him and the person behind him to be men. If the man is not in the middle, we only need to require the person next to him to be a man. These lead to

$$\begin{aligned}\Pr(X_i = 0) =& \frac{2}{2n} \cdot \frac{n-1}{2n-1} + (1 - \frac{1}{n}) \cdot \frac{n-1}{2n-1} \cdot \frac{n-2}{2n-2} \\ =& \frac{n-1}{4n-2}.\end{aligned}$$

Therefore,

$$\Pr(X_i = 1) = 1 - \Pr(X_i = 0) = \frac{3n-1}{4n-2},$$

leading to

$$E[X] = \frac{3n^2 - n}{4n-2}.$$

For (2), there are always two people next to a man. So

$$\Pr(X_i = 0) = \frac{n-1}{2n-1} \cdot \frac{n-2}{2n-2} = \frac{n-2}{4n-2}.$$

Therefore,

$$\Pr(X_i = 1) = \frac{3n}{4n-2},$$

leading to

$$E[X] = \frac{3n^2}{4n-2}.$$

□

Example 5.31. If n couples are seated randomly at a round table, find the expected value and the variance of the number of couples sitting together. Also compare Example 4.10.

solution. Let $X_1, \ldots, X_n$ be the indicators, one for each couple. We declare $X_i = 1$ if the i-th couple sits together. Otherwise, $X_i = 0$. The sum $X = \sum_i X_i$ is the number of couples sitting together. For the expected value of X, we need to find $\Pr(X_i = 1)$ by Example 5.7.

We can treat the i-th couple as a single entity, so the probability that they sit together would be

$$\Pr(X_i = 1) = \frac{2(2n-2)!}{(2n-1)!} = \frac{2}{2n-1}.$$

In fact, the probability that a particular group of r couples sitting together is

$$\Pr(X_{i_1}, \ldots, X_{i_r}) = \frac{2^r(2n-2r+r-1)!}{(2n-1)!}. \tag{5.3}$$

Therefore,

$$E[X] = \frac{2n}{2n-1}.$$

For the variance, we need $\Pr(X_i = X_j = 1)$ by Example 5.7. By Equation (5.3),

$$\Pr(X_i = X_j = 1) = \frac{2^2(2n-3)!}{(2n-1)!} = \frac{2}{(2n-1)(n-1)}.$$

From here it can be calculated that

$$E\left[\binom{X}{2}\right] = \frac{n}{2n-1}$$

so that

$$\mathrm{Var}(X) = \frac{4n^2-4n}{(2n-1)^2}.$$

□

Remark 5.32. In the above example, it can be calculated that

$$\begin{aligned}\mathrm{Cov}(X_i, X_j) =& \Pr(X_i = X_j = 1) - \Pr(X_i = 1)\cdot\Pr(X_j = 1)\\ =& \frac{2}{(2n-1)^2(n-1)} > 0.\end{aligned}$$

This means having a couple that sits together increases the likelihood that another couple sits together.

Example 5.33. Suppose there are $2n$ people, among which n are men and n are women. These people are paired randomly into n pairs. Find the expected value and the variance of the number of pairs consisting of a man and a woman.

Similarly, suppose there are n couples that are re-paired randomly. Find the expected value and the variance of the number of couples that are still paired together. Also compare Example 4.6 and Example 4.7.

Proof. For the first problem, we define $X_1, \ldots, X_n$, one for each pair. We declare $X_i = 1$ if the i-th pair consists of a man and woman. Otherwise, $X_i = 0$. The sum $\sum_i X_i$ is the number of pairs consisting of a man and a woman. By Example 5.7, we only need to calculate $\Pr(X_i = 1)$, and $\Pr(X_i = X_j = 1)$.

Since there are n women and n men possible for the i-th pair, we find

$$\Pr(X_i) = \frac{n^2}{\binom{2n}{2}} = \frac{n}{2n-1}.$$

Moreover,

$$\Pr(X_i = X_j = 1) = \Pr(X_i = 1) \cdot \Pr(X_j = 1 | X_i = 1) = \frac{n}{2n-1} \cdot \frac{n-1}{2n-3}.$$

For the second problem, we define $X_1, \dots, X_n$, one for each couple. We declare $X_i = 1$ if the i-th couple are grouped together. Otherwise, $X_i = 0$. The sum $\sum_i X_i$ is the number of couples that are still paired together. By Example 5.7, we only need to calculate $\Pr(X_i = 1)$, and $\Pr(X_i = X_j = 1)$.

Since the couple can be grouped in any of the n pairs, we have

- $\Pr(X_i = 1) = \frac{n}{\binom{2n}{2}} = \frac{1}{2n-1}$;
- $\Pr(X_i = X_j = 1) = \frac{1}{2n-1} \cdot \frac{1}{2n-3}$.

□

5.5. **Coin toss sequence expectations.** In this section, the main topics are the two:

- Expected coin flips to obtain either (both) m heads or (and) n tails in total.
- Expected coin flips to obtain either (both) m successive heads or (and) n successive tails.

The maximum-minimum identity (Proposition 5.22) is frequently used.

Example 5.34 (Cumulating results)**.** A coin lands on its head with probability p. Find the expected flips to obtain

(1) both n heads and m tails;
(2) either n heads or m tails.

solution. Let M (resp. N) be the number of flips to obtain m tails (resp. n heads). Let X be the number of flips to accumulate both n heads and m tails and let Y be that of the flips to accumulate either n heads or m tails. We then have

$$X = \max(M, N) \text{ and } Y = \min(M, N).$$

We will first obtain $E[X]$ by conditioning on the number of heads H in the first $n + m - 1$ flips.

Note that when $H = k \leq n - 1$ (resp. $H = k > n - 1$), m tails (resp. n heads) have been accumulated, and there remain $n - k$ heads (resp. $k + 1 - n$ tails) to be accumulated.

Therefore,

$$E[X|H = k] = \begin{cases} n + m - 1 + \frac{n-k}{p} & \text{if } k \leq n - 1, \\ n + m - 1 + \frac{k+1-n}{1-p} & \text{otherwise.} \end{cases}$$

Next,

$$\Pr(H = k) = \binom{n + m - 1}{k} p^k (1 - p)^{m+n-1-k}$$

so that

$$E[X] = \sum_{k=0}^{n+m-1} \Pr(H=k)\cdot E[X|H=k]$$
$$= n+m-1+\sum_{k=0}^{n-1}\binom{n+m-1}{k}p^k(1-p)^{m+n-1-k}\cdot\frac{n-k}{p}$$
$$+\sum_{k=n}^{n+m-1}\binom{n+m-1}{k}p^k(1-p)^{n+m-1-k}\cdot\frac{k+1-n}{1-p}.$$

This finishes (1). For (2), we apply Equation (5.2) to get

$$E[Y] = E[M]+E[N]-E[X] = \frac{m}{1-p}+\frac{n}{p}-E[X].$$

□

For the upcoming example, it is useful to invoke the following

$$\sum_{i=1}^{n} ip^{i-1}(1-p) = \frac{1-p^n}{1-p} - np^n \text{ for any } 0<p<1. \tag{5.4}$$

Note that the above equation is the n-th partial sum of the expected number of flips to obtain a tail:

$$\sum_{i=1}^{\infty} ip^{i-1}(1-p) = \frac{1}{1-p}.$$

Therefore, to derive Equation (5.4), it is sufficient to derive

$$\sum_{k=1}^{\infty}(n+k)p^{n+k-1}(1-p) = p^n\sum_{k=1}^{\infty}(n+k)p^{k-1}(1-p)$$
$$= np^n\sum_{k=1}^{\infty}p^{k-1}(1-p)+p^n\sum_{k=1}^{\infty}k(1-p)p^{k-1}$$
$$= np^n+\frac{p^n}{1-p}.$$

It follows that

$$\sum_{i=1}^{n} ip^{i-1}(1-p) = \sum_{i=1}^{\infty} ip^{i-1}(1-p) - \sum_{k=1}^{\infty}(n+k)p^{n+k-1}(1-p)$$
$$= \frac{1-p^n}{1-p} - np^n.$$

Example 5.35 (To amass consecutive heads)**.** Suppose a coin lands on its head with probability p. Find the expected number of flips to obtain n successive heads.

solution. Let E be the expected number in question. We condition on the position of the first tail in the first n flips. Whenever a tail occurs, the state resets to the initial state. Based on these, we get

$$\begin{aligned}E =&(1-p)(1+E)+p(1-p)(2+E)+\cdots p^{n-1}(1-p)(n+E)+np^n \\ =&\sum_{i=1}^{n} ip^{i-1}(1-p)+\sum_{i=1}^{n} p^{i-1}(1-p)E+np^n \\ \overset{(5.4)}{=}&\frac{1-p^n}{1-p}+(1-p^n)E.\end{aligned}$$

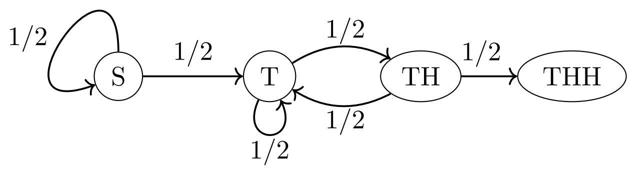
Solving for E, we get

$$E=\frac{1-p^n}{p^n(1-p)}.$$

When $p=\frac{1}{2}$, the expected value is $2^{n+1}-2$. □

Example 5.36. A coin lands on its head with probability 1/2. Find the expected number of flips to obtain the triplet THH. Do the same for the triplet HHH.

solution. The following diagram summarizes the states from the beginning S to achieving the triplet THH:



.

For example, an initial H resets the state back to S. After obtaining a first T, any additional T obtained before the triplet THH resets the state to an initial T.

For any state α, we let μ_α be the expected number of flips starting at α, until we reach the triplet THH. For example, μ_T is the expected number of flips to obtain THH starting with an initial T. Similarly μ_{TH} is the expected number of flips to obtain THH starting with an initial TH. Our goal is to find precisely μ_S.

Looking at the S node in the above figure, there are two possible outcomes. An H leading back to S, and a T. Conditioning on these two outcomes (outgoing arrows), we obtain

$$\mu_s=\frac{1}{2}(1+\mu_S)+\frac{1}{2}(1+\mu_T)=1+\frac{1}{2}\mu_S+\frac{1}{2}\mu_T.$$

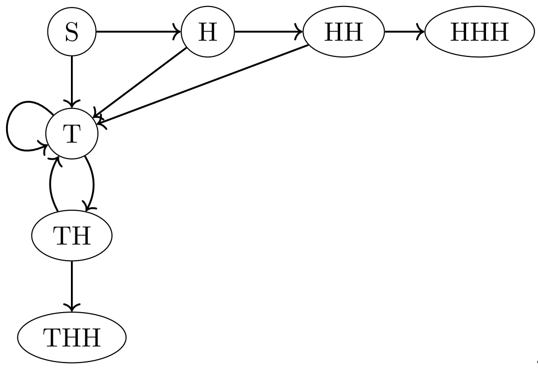
Iterating over all nodes, we obtain in addition

- $\mu_T=1+\frac{1}{2}\mu_T+\frac{1}{2}\mu_{TH}$;
- $\mu_{TH}=1+\frac{1}{2}\mu_T$.

We thus have obtained three linear equations in μ_S, μ_T and μ_{THH}. One easily verifies that $\mu_{TH}=4$, $\mu_T=6$, and $\mu_S=8$. The expected number of flips to obtain HHH can be obtained either by the similar technique we just presented, or Example 5.35. The value is 14. □

Example 5.37. As a continuation of Example 5.36, find the expected number of flips to obtain either HHH, or THH. Do the same for the number of flips to obtain both THH and HHH.

solution. The state diagram to obtain either triplet would be



,

where each arrow has probability $1/2$.

This diagram demonstrates that if any T is obtained before HHH, the triplet THH is the only final state possible. That is why nodes T and TH give the same equations as before. Therefore, we still have $\mu_T = 6$ and $\mu_{TH} = 4$. Iterating through the rest of the nodes, we get $\mu_{HH} = 4$, $\mu_H = 6$, and $\mu_S = 7$.

To get the expected number of flips to obtain both triplets, we use Equation (5.2). More precisely, suppose the flip continues indefinitely, and M (resp. N) are the number of flips to obtain THH (resp. HHH). The value $\max(M, N)$ (resp. $\min(M, N)$) is then the number of flips to obtain both (resp. either) of the triplets. Therefore,

$$E[\max(M, N)] = E[M] + E[N] - E[\min(M, N)] = 8 + 14 - 7 = 15.$$

□

Example 5.38 (Cumulating successive results)**.** Instead of dealing with the most generality, we consider the expected number of flips to obtain 3 successive heads or 2 successive tails, and then use Equation (5.2) to calculate the expected number of flips to obtain both. The strategy described in this example generalizes easily.

Let the coin land on its head with probability $1/2$. Let M (resp. N) be the number of flips to obtain 2 tails (resp. 3 heads). The value $X = \min(M, N)$ is the number of flips to obtain either 3 successive heads or 2 successive tails. In addition, we let μ_H (resp. μ_T) be the expected number of flips to achieve the goal, starting with an initial head (resp. tail).

Conditioning on the first flip, we get

$$\begin{aligned} E[X] =& \frac{1}{2} \cdot E[X|H] + \frac{1}{2} \cdot E[X|T] \\ =& \frac{1}{2} \cdot (1 + \mu_H) + \frac{1}{2} \cdot (1 + \mu_T) \\ =& 1 + \frac{1}{2} \cdot (\mu_H + \mu_T)\,. \end{aligned} \tag{5.5}$$

Given H, any tail before 3 successive heads resets the state to an initial tail. Conditioning on the first position of the tail in the next 2 flips, we get

$$\begin{aligned} \mu_H =& \sum_{i=1}^{3-1} \left(\frac{1}{2}\right)^{i-1} \cdot \frac{1}{2} \cdot (i + \mu_T) + \frac{1}{4} \cdot 2 \\ =& \frac{3}{2} + \frac{3}{4}\mu_T. \end{aligned}$$

Similarly, given T, any head before 2 successive heads resets the state to an initial head. Conditioning on the first position of the head in the next flip, we get

$$\mu_T = \frac{1}{2}(1+\mu_H) + \frac{1}{2} = 1 + \frac{1}{2}\mu_H.$$

We can then solve $\mu_H = \frac{18}{5}$ and $\mu_T = \frac{14}{5}$, yielding

$$E[X] = \frac{21}{5}.$$

Finally, the value $Y = \max(M, N)$ is the number of flips to obtain both 3 successive heads and 2 successive tails. Equation (5.2) and Example 5.35 implies

$$E[Y] = E[M] + E[N] - E[X] = 2^3 - 2 + 2^4 - 2 - \frac{21}{5} = \frac{79}{5}.$$

One can also obtain $E[Y]$ by conditioning. Everything is the same as before, except that when either successive results are obtained, we need to obtain the other. The equations for μ_H and μ_T becomes

$$\begin{aligned}
\mu_H =& \sum_{i=1}^{3-1} \left(\frac{1}{2}\right)^{i-1} \frac{1}{2} \cdot (i + \mu_T) + \frac{1}{4} \cdot (2 + E[M]) \\
=& 3 + \frac{3}{4}\mu_T, \\
\mu_T =& \frac{1}{2} \cdot (1 + E[N]) + \frac{1}{2}(1 + \mu_H) \\
=& 8 + \frac{1}{2}\mu_H.
\end{aligned}$$

We then solve $\mu_H = \frac{72}{5}$, and $\mu_T = \frac{76}{5}$, resulting in

$$E[Y] = 1 + \frac{1}{2} \cdot (\mu_H + \mu_T) = \frac{79}{5}.$$

In the most generality where the coin lands on head with probability p, the equation for the expected number of flips X to obtain either n consecutive heads or m consecutive tails is

$$E[X] = pE[X|H] + qE[X|T] = 1 + p\mu_H + q\mu_T,$$

where $q = 1 - p$. Moreover,

$$\begin{aligned}
\mu_H =& \sum_{i=1}^{n-1} p^{i-1} q (i + \mu_T) + (n-1)p^{n-1} \\
=& \frac{1 - p^{n-1}}{q} + (1 - p^{n-1})\mu_T, \\
\mu_T =& \sum_{i=1}^{m-1} q^{i-1} p (i + \mu_H) + (m-1)q^{m-1} \\
=& \frac{1 - q^{m-1}}{p} + (1 - q^{m-1})\mu_H.
\end{aligned}$$

As for the number of flips Y to obtain both consecutive n heads and m tails, we let M (resp. N) be the number of flips to obtain m consecutive tails (resp. n

consecutive heads). The equations are

$$\begin{aligned}
E[Y] =& 1 + p\mu_H + q\mu_T, \\
\mu_H =& \sum_{i=1}^{n-1} p^{i-1} q(i + \mu_T) + p^{n-1}(n - 1 + E[M]) \\
=& \frac{1 - p^{n-1}}{q} + p^{n-1}E[M] + (1 - p^{n-1})\mu_T, \\
\mu_T =& \sum_{i=1}^{m-1} q^{i-1} p(i + \mu_H) + q^{m-1}(m - 1 + E[N]) \\
=& \frac{1 - q^{m-1}}{p} + q^{m-1}E[N] + (1 - q^{m-1})\mu_H,
\end{aligned}$$

where $E[M]$ and $E[N]$ were discussed in Example 5.35.

5.6. **Some trickier expectations.**

Example 5.39 (A pill's puzzle)**.** Suppose there are m half pills and n large pills in a jar. Each pill is equally likely to be selected. If a half pill is selected, it is consumed. If a large pill is taken, it is split into two half pills. One of the half pill is consumed, and the other half is put back into the jar. Find the expected number of half pills in the jar when all large pills are consumed.

solution. Let X be an indicator for a particular half pill and let Y be an indicator for a particular large pill. We declare $X = 1$ when this particular half pill remains after all large pills are consumed. Otherwise, we set $X = 0$. Similarly, declare $Y = 1$ if one of the half pill split from the particular large pill remains after all large pills are consumed. Declare $Y = 0$ otherwise. We need to find $\Pr_{m,n}(X = 1)$ and $\Pr_{m,n}(Y = 1)$.

It is easier to find $\Pr_{m,n}(X = 1)$. Intuitively, $X = 1$ when all large pills are selected prior to this particular half pill. Therefore,

$$\Pr_{m,n}(X = 1) = \frac{1}{n+1}.$$

One can show this by induction on m and n, where the base case being

$$\Pr_{i,0}(X = 1) = 1 \text{ for all } i.$$

Assume we have shown that $\Pr_{i,j}(X = 1) = \dfrac{1}{j+1}$ for all $j < n$ and for all i. To establish the identity for $\Pr_{m,n}(X = 1)$, we must start at $\Pr_{1,n}(X = 1)$.

Since there is only one half pill, for $X = 1$ we must only select the large pills. We therefore get

$$\begin{aligned}
\Pr_{1,n}(X = 1) =& \frac{n}{n+1} \cdot \Pr_{2,n-1}(X = 1) = \frac{n}{n+1} \cdot \frac{1}{n} \\
=& \frac{1}{n+1},
\end{aligned}$$

giving the desired identity for $\Pr_{1,n}(X = 1)$.

Next, assume we have shown $\Pr_{m-1,n}(X=1) = \frac{1}{n+1}$. Conditioning on whether half pills or large pills are selected, we get

$$\begin{aligned}\Pr_{m,n}(X=1) =& \frac{m-1}{n+m} \cdot \Pr_{m-1,n}(X=1) + \frac{n}{n+m} \cdot \Pr_{m+1,n-1}(X=1)\\ =& \frac{m-1}{n+m} \cdot \frac{1}{n+1} + \frac{n}{n+m} \cdot \frac{1}{n}\\ =& \frac{1}{n+1}.\end{aligned}$$

This establishes the desired identity for $\Pr_{m,n}(X=1)$.

For $\Pr_{m,n}(Y=1)$, we claim that

$$\Pr_{m,n}(Y=1) = \frac{1}{n} \cdot (1 + \frac{1}{2} + \cdots + \frac{1}{n}).$$

This comes from these observations. First, it is obvious that $\Pr_{i,1}(Y=1) = 1$ for all i. It is also not difficult to show $\Pr_{0,2}(Y=1) = \frac{3}{4}$. We now fix the number of large pills at 2, and consider $\Pr_{1,2}(Y=1)$ by conditioning on whether the half pill, the particular large pill, or the other large pill is selected. This gives

$$\begin{aligned}\Pr_{1,2}(Y=1) =& \frac{1}{3} \cdot \Pr_{0,2}(Y=1) + \frac{1}{3} \cdot \Pr_{2,1}(X=1) + \frac{1}{3} \cdot \Pr_{2,1}(Y=1)\\ =& \frac{1}{3} \cdot \frac{3}{4} + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3}\\ =& \frac{3}{4} = \Pr_{0,2}(Y=1),\end{aligned}$$

where X is the indicator for the half pill split from the Y-pill. This suggests that like $\Pr_{m,n}(X=1)$, $\Pr_{m,n}(Y=1)$ does not depend on m. Assuming this, and using X as the indicator for the small pill split from the Y-pill, we get

$$\begin{aligned}\Pr_{0,n}(Y=1) =& \frac{1}{n} \cdot \Pr_{1,n-1}(X=1) + \frac{n-1}{n} \cdot \Pr_{1,n-1}(Y=1)\\ =& \frac{1}{n} \cdot \frac{1}{n} + \frac{n-1}{n} \cdot \Pr_{0,n-1}(Y=1)\\ =& \cdots\\ =& \frac{1}{n} \cdot (\frac{1}{n} + \frac{1}{n-1} + \cdots + 1 \cdot \Pr_{0,1}(Y=1))\\ =& \frac{1}{n} \cdot (\frac{1}{n} + \frac{1}{n-1} + \cdots + 1).\end{aligned}$$

The above observations suggest that

$$\Pr_{m,n}(Y=1) = \frac{1}{n} \cdot (1 + \frac{1}{2} + \cdots + \frac{1}{n}),$$

which can also be proved by induction.

For example, assume

$$\Pr_{i,j}(Y=1) = \frac{1}{j} \cdot (1 + \cdots + \frac{1}{j})$$

for all $j < n$ and for all i, we get

$$\begin{aligned}\Pr_{0,n}(Y=1) =& \frac{1}{n}\Pr_{1,n-1}(X=1) + \frac{n-1}{n}\Pr_{1,n-1}(Y=1)\\ =& \frac{1}{n}\cdot\frac{1}{n} + \frac{n-1}{n}\cdot\frac{1}{n-1}\cdot(1+\cdots+\frac{1}{n-1})\\ =& \frac{1}{n}\cdot(1+\cdots+\frac{1}{n}).\end{aligned}$$

Next, assume we have established the identity for $\Pr_{m-1,n}(Y=1)$. By conditioning on the pills taken, we get

$$\begin{aligned}\Pr_{m,n}(Y=1) =& \frac{m}{m+n}\cdot\Pr_{m-1,n}(Y=1) + \frac{1}{m+n}\cdot\Pr_{m+1,n-1}(X=1)\\ &+ \frac{n-1}{m+n}\cdot\Pr_{m+1,n-1}(Y=1)\\ =& \frac{m}{m+n}\cdot\Pr_{m-1,n}(Y=1) + \frac{1}{m+n}\cdot(1+\cdots+\frac{1}{n})\\ =& \frac{1}{n}\cdot(1+\cdots+\frac{1}{n}).\end{aligned}$$

We have established the desired identities for $\Pr_{m,n}(X=1)$ and $\Pr_{m,n}(Y=1)$.

The expected number of half pills in the jar after all large pills are consumed is

$$\begin{aligned}& m\cdot\Pr_{m,n}(X=1) + n\cdot\Pr_{m,n}(Y=1)\\ =& \frac{m}{n+1} + \frac{1}{n} + \frac{1}{n-1} + \cdots + 1.\end{aligned}$$

□

Remark 5.40. In the above example, we can also find the expectation of the number of pills Z consumed until all large pills are taken. For this, let us relate Z and the number X of half pills left in the jar.

For starters, n pills need to be consumed to exhaust the large pills. This results in $m+n$ half pills, among which $Z-n$ are taken. It follows that

$$X + (Z-n) = m+n.$$

The expected value of Z is then

$$E[Z] = m + 2n - E[X].$$

Example 5.41 (Gambler's ruin)**.** Suppose there are 2 players. Let the first player have i dollars and the second player have $n-i$ dollars. The winner of a game receives one dollar from the loser. The victor is the one who gets all n dollars. Find the expected number of games to be played until a victor arises. For simplicity, assume in each game, each player is equally likely to win.

Proof. Let $X_i, i = 0, \ldots, n$ be the random variable of the number of games to be played until a victor occurs when the first player has i dollars. Clearly we have $X_n = X_0 = 0$. We will calculate $E[X_i]$ for $0 < i < n$.

Conditioning on the result of the first game, we have

$$\begin{aligned}E[X_i] =& \frac{1}{2}(1+E[X_{i-1}]) + \frac{1}{2}(1+E[X_{i+1}])\\ =& 1 + \frac{1}{2}(E[X_{i-1}] + E[X_{i+1}])\end{aligned}$$

for $i = 1, \dots, n-1$. A little bit of algebra shows that

$$E[X_i] - E[X_{i+1}] = 2 + (E[X_{i-1}] - E[X_i]),\ i = 1, \dots, n-1.$$

From this it follows that

$$E[X_i] - E[X_{i+1}] = 2i - E[X_1],\ i = 1, \dots, n-1. \tag{5.6}$$

Adding the equations (5.6) up, we get

$$\begin{aligned} E[X_1] - E[X_i] =& 2(1 + \cdots + (i-1)) - (i-1)E[X_1] \\ =& i(i-1) - (i-1)E[X_1]. \end{aligned} \tag{5.7}$$

This yields

$$-E[X_i] = i(i - 1 - E[X_1]).$$

In particular,

$$-E[X_n] = 0 = n(n - 1 - E[X_1]),$$

implying

$$E[X_1] = n - 1.$$

Putting this back to 5.7, we obtain $-E[X_i] = i(i-n)$ so that

$$E[X_i] = i(n-i).$$

Note that we have the symmetry

$$E[X_i] = E[X_{n-i}]$$

as both players are equally likely to win. □

Example 5.42 (Amplification of gambler's ruin)**.** Suppose there are r players, each having an initial amount of n_i dollars. In each game two players are chosen. The winner of the game receives one dollar from the loser. A player is eliminated when they do not have any money left. The victor is the one who gets all $\sum_i n_i = n$ dollars. Find the expected number of games to be played until a victor occurs. For simplicity, assume each player has the same chance of winning in each game.

solution. Define $X_1, \dots, X_r$ as the random variables so that X_i is the number of games played by the i-th participant. Since each game is played by two participants, the total number of games is

$$X = \frac{1}{2} \cdot \sum_i X_i.$$

Therefore,

$$E[X] = \frac{1}{2} \cdot \sum_i E[X_i].$$

It remains to find $E[X_i]$.

For participant i, they either win the total n dollars, or lose n_i dollars. Therefore,

$$E[X_i] = n_i(n - n_i)$$

by Example 5.41. □

Example 5.43 (To win gambler's ruin)**.** This is a follow up for Example 5.41. Recall X_i is the number of games played when the first player has i dollars. Also define for each i the random variable

$$W_i = \begin{cases} 1 \text{ if the first player wins starting } i \text{ dollars,} \\ 0 \text{ if the first player loses starting at } i \text{ dollars.} \end{cases}$$

Let us find the expected number of games played given that the first play wins. That is, find

$$E[X_i|W_i = 1].$$

From now on we shall abbreviate this conditional expectation as E_i.

As before, we can find E_i by conditioning on the outcome S of the first game. Let

$$p_i = \Pr(W_i = 1) = \frac{i}{n}.$$

Let $S = w$ (resp. $S = l$) denote the outcome that the first player wins (resp. loses) in the first game. We then note that

$$\begin{aligned} E[X_i|W_i = 1, S = w] &= \sum_k (k+1) \cdot \frac{\Pr(X_i = k+1, S = w, W_i = 1)}{\Pr(S = w, W_i = 1)} \\ &= \sum_k (k+1) \cdot \frac{\Pr(S = w) \cdot \Pr(X_{i+1} = k, W_{i+1} = 1)}{\Pr(S = w) \cdot \Pr(W_{i+1} = 1)} \\ &= 1 + E_{i+1}. \end{aligned}$$

Similarly,

$$E[X_i|W_i = 1, S = l] = 1 + E_{i-1}.$$

Conditioning on S, we get

$$E_i = 1 + \Pr(S = w|W_i = 1)E_{i+1} + \Pr(S = l|W_i = 1)E_{i-1}.$$

It is not easy to derive a solution for E_i directly. Instead, let

$$R_i = \Pr(W_i = 1) \cdot E_i = \frac{i}{n} E_i.$$

We then get the adjusted relations

$$R_i = \frac{i}{n} + \frac{1}{2} R_{i+1} + \frac{1}{2} R_{i-1},$$

which is equivalent to

$$R_i - R_{i+1} = \frac{2i}{n} + (R_{i-1} - R_i).$$

From here it follows that

$$R_{i-1} - R_i = \frac{i(i-1)}{n} + (R_0 - R_1) = \frac{i(i-1)}{n} - R_1.$$

Summing these up, we get

$$\begin{aligned} R_1 - R_i &= \frac{1}{n} \cdot \sum_{k=1}^{i-1} k(k+1) - (i-1)R_1 \\ &= \frac{i(i-1)(i+1)}{3n} - (i-1)R_1. \end{aligned}$$

Hence

$$R_i = iR_1 - \frac{i(i-1)(i+1)}{3n}.$$

In particular

$$R_n = nR_1 - \frac{(n-1)(n+1)}{3} = 0.$$

This gives

$$R_1 = \frac{n^2 - 1}{3n} \text{ and } R_i = \frac{i}{3n}(n^2 - i^2).$$

Finally, we obtain

$$E_i = \frac{n^2 - i^2}{3}.$$

Remark 5.44. By symmetry, we should also get that

$$E[X_i|W_i = 0] = \frac{n^2 - (n-i)^2}{3}$$

since $W_i = 0$ means the second player wins starting at $n - i$ dollars. This can be checked by

$$\begin{aligned} i(n-i) =& E[X_i] \\ =& \Pr(W_1 = 1) \cdot E[X_i|W_i = 1] + \Pr(W_0 = 1) \cdot E[X_i|W_i = 0] \\ =& \frac{i}{n} \cdot \frac{n^2 - i^2}{3} + \frac{n-i}{n} E[X_i|W_i = 0]. \end{aligned}$$

Upon division by $n - i$, we arrive at

$$i = \frac{i}{n} \cdot \frac{n+i}{3} + \frac{1}{n} E[X_i|W_i = 0],$$

giving

$$E[X_i|W_i = 0] = \frac{2ni - i^2}{3} = \frac{n^2 - (n-i)^2}{3}$$

as desired.

Example 5.45 (Color balls)**.** Before delving into this example, one should be familiar with Example 4.26, Example 5.41, and Example 5.43.

Assume there are n balls in a jar in n distinct colors. Each time two balls are selected from the jar. The second ball will then be painted in the same color as the first ball. The process continues until all balls are of the same color. Find the expected time until the process stops.

solution. One may in the first attempt conditions on the number of colors in the jar, where in the initial state there are n colors. In the final state there is one color of balls left in the jar.

The issue is that the probability to go from j colors to $j+1, j$, or $j-1$ colors depends on the number of balls in each color. One would then further condition on the number of balls in each color, resulting in a very complicated system. The trick, is to condition on the number of a specific color.

First, note that each of the n colors are equally likely to be the lone color at the end of the process. Namely, if we let F_j be the event that the j-th color remains at the end, we have $\Pr(F_j) = 1/n$.

Therefore, if X is the time the process stops, we have

$$E[X] = \sum_{j=1}^{n} \frac{1}{n} \cdot E[X|F_j].$$

We shall see not surprisingly that $E[X|F_j]$ does not depend on j so that $E[X] = E[X|F_j]$ for any j.

For the ease of notations let us compute $E[X|F_1]$. We will condition on the number i of balls that are of the first color, where $1 \le i \le n$. Note that since color one remains, $i \neq 0$.

We now also define F_1^i as the event that color one balls remain, starting with i of them in the jar. Our original F_1 is now F_1^1. We additionally define $P_{i,j}$ as the probability that the number of color one balls changes from i to j in a single round, given F_1^i. We then obtain the system of equations

$$E[X|F_1^i] = 1 + P_{i,i-1}E[X|F_1^{i-1}] + P_{i,i}E[X|F_1^i] + P_{i,i+1}E[X|F_1^{i+1}], 2 \le i \le n-1.$$

When $i = 1$, the equation is

$$E[X|F_1^1] = 1 + P_{1,1}E[X|F_1^1] + P_{1,2}E[X|F_1^2].$$

The boundary case is

$$E[X|F_1^n] = 0.$$

Let us calculate $P_{i,i}$, $P_{i,i-1}$ and $P_{i,i+1}$.

For these, we need $\Pr(F_1^i)$ first. Condition on the outcome of the selection, we get the system

$$\Pr(F_1^i) = \Pr(i \to i-1)\Pr(F_1^{i-1}) + \Pr(i \to i)\Pr(F_1^1) + \Pr(i \to i+1)\Pr(F_1^{i+1}),$$

where $1 \le i \le n-1$, $F_1^0 = 0$ and $F_1^n = 1$. To nail down $\Pr(i \to j)$, where $j = i-1, i$ or $i+1$, simply note that balls that are not of color one are all equivalent. Therefore,

$$\Pr(i \to j) = \begin{cases} \frac{(n-i)i}{n(n-1)} \text{ if } j = i-1, \\ \\ \frac{i(i-1)+(n-i)(n-i-1)}{n(n-1)} \text{ if j=i}, \\ \\ \frac{i(n-i)}{n(n-1)} \text{ if } j = i+1. \end{cases}$$

Note that the above probabilities add up to 1. Putting these back to the equation for $\Pr(F_1^i)$, we get

$$\Pr(F_1^i) - \Pr(F_1^{i+1}) = \Pr(F_1^{i-1}) - \Pr(F_1^i), 1 \le i \le n-1.$$

One easily verifies that

$$\Pr(F_n^i) = \frac{i}{n},$$

which now gives

$$P_{i,j} = \frac{\Pr(i \to j) \cdot \Pr(F_1^j)}{\Pr(F_1^i)} = \begin{cases} \frac{(n-i)(i-1)}{n(n-1)} \text{ if } j = i-1, \\ \\ \Pr(i \to i) \text{ if } j = i, \\ \\ \frac{(n-i)(i+1)}{n(n-1)} \text{ if } j = i+1. \end{cases}$$

Note that $P_{i,i-1} < \Pr(i \to i-1)$ and $P_{i,i+1} > \Pr(i \to i+1)$. These should come as no surprises. For example, the probability that the number of color one balls decreases, given that color one remains, is less likely than when there is no condition on color one balls.

The system of equations for $E[X|F_1^i]$ can now be rearranged as

$$E[X|F_1^i] - E[X|F_1^{i+1}] = \frac{n(n-1)}{(n-i)(i+1)} + \frac{i-1}{i+1}(E[X|F_1^{i-1}] - E[X|F_1^i]) \tag{5.8}$$

for $2 \le i \le n-1$, and

$$E[X|F_1^1] - E[X|F_1^2] = \frac{2}{n}.$$

For the ease of notations we let

- $w_i = E[X|F_1^i] - E[X|F_1^{i+1}]$;

- $a_i = \frac{n(n-1)}{(n-i)(i+1)}$;
- $c_i = \frac{i-1}{i+1}$.

Note that

$$E[X|F_1^1] = E[X|F_1^1] - E[X|F_1^n] = \sum_{i=1}^{n-1} w_i.$$

Unpacking Equation (5.8) repeatedly, we get

$$\begin{aligned} w_i =& a_i + c_i w_{i-1} = a_i + c_i(a_{i-1} + c_{i-1} w_{i-2}) \\ =& a_i + c_i a_{i-1} + c_i c_{i-1} w_{i-2} \\ =& \cdots \\ =& a_i + c_i a_{i-1} + c_i c_{i-1} a_{i-2} + \cdots + c_i c_{i-1} \cdots c_2 a_1 \\ =& \sum_{j=1}^{i} \frac{j(j+1)}{i(i+1)} a_j = \frac{n(n-1)}{i(i+1)} \cdot \sum_{j=1}^{i} \frac{j}{n-j}. \end{aligned}$$

Here one uses the facts

$$c_i c_{i-1} \cdots c_k = \frac{(k-1)k}{i(i+1)}, \text{ and } w_1 = a_1.$$

Finally,

$$\begin{aligned} \sum_{i=1}^{n-1} w_i =& n(n-1) \cdot \sum_{i=1}^{n-1} \sum_{j=1}^{i} \frac{1}{i(i+1)} \cdot \frac{j}{n-j} \\ =& n(n-1) \cdot \sum_{j=1}^{n-1} \frac{j}{n-j} \sum_{i=j}^{n-1} \frac{1}{i(i+1)} \\ =& n(n-1) \cdot \sum_{j=1}^{n-1} \frac{j}{n-j} \sum_{i=j}^{n-1} (\frac{1}{i} - \frac{1}{i+1}) \\ =& n(n-1) \cdot \sum_{j=1}^{n-1} (\frac{j}{n-j} (\frac{1}{j} - \frac{1}{n})) \\ =& n(n-1) \cdot \sum_{j=1}^{n-1} \frac{j}{n-j} \frac{n-j}{nj} = n(n-1)\frac{n-1}{n} = (n-1)^2. \end{aligned}$$

Since all the computations we have done do not depend on the color which one conditions on, we conclude that

$$E[X] = E[X|F_1] = (n-1)^2.$$

□

Example 5.46. In Example 5.45, if one only chooses two balls of distinct colors, find the expected time until the process stops.

solution. Since each of the two balls selected can be either the first or the second with equal probability, this set up is just the generalized gambler's ruin (Example 5.42), where there are n players, each having 1 dollar. The expected time is then

$$\frac{n(n-1)}{2}.$$

□

Example 5.47. Two dice are rolled. If the sum of the dice is 2, 3, or 12, the player loses. If the sum is 7 or 11, the player wins. If the sum is any other number i, the player continues to roll the dice until the sum is 7 or i. If the sum is 7, the player loses. If it is i, the player wins. Let X be the number of rolls to complete the game. Find the conditional expectation $E[X|\text{player wins}]$.

solution. Let p_i be the probability that the sum is i. Let us determine Pr(win) first. For this, we can condition on the sum S of the first roll:

$$\begin{aligned}\Pr(\text{win}) =& p_7 + p_{11} + \sum_i p_i \cdot \Pr(\text{win}|S = i)\\ =& p_7 + p_{11} + \sum_i p_i \cdot \Pr(i \text{ occurs before } 7)\\ =& p_7 + p_{11} + \sum_i p_i \cdot \frac{p_i}{p_i + p_7},\end{aligned}$$

where the sum is over i not equal to 2, 3, 12, 7, 11.

Let us condition the expectation on S. We have

$$E[X|\text{win}] = \Pr(S = 7, 11|\text{win}) + \sum_i \Pr(S = i|\text{win})E[X|\text{win}, S = i],$$

where the sum is over i not equal to 2, 3, 12, 7, 11.

For each such i, we have

$$\begin{aligned}E[X|\text{win}, S = i] =& \sum_{k=1}^{\infty}(k + 1) \cdot \Pr(X = k + 1|\text{win}, S = i)\\ =& \sum_{k=1}^{\infty}(k + 1) \cdot \frac{p_i^2(1 - p_i - p_7)^{k-1}}{p_i^2/(p_i + p_7)}\\ =& \sum_{k=1}^{\infty}(k + 1) \cdot (1 - p_i - p_7)^{k-1} \cdot (p_i + p_7)\\ =& 1 + \sum_{k=1}^{\infty} k \cdot (1 - p_i - p_7)^{k-1} \cdot (p_i + p_7).\end{aligned}$$

Now note that $\sum_{k=1}^{\infty} k \cdot (1 - p_i - p_7)^{k-1} \cdot (p_i + p_7)$ is the expected number of rolls to get a sum of either i or 7. Therefore,

$$E[X|\text{win}, S = i] = 1 + \frac{1}{p_i + p_7}(= E[X|S = i]).$$

We then have

$$\begin{aligned}E[X|\text{win}] =& \Pr(S = 7, 11|\text{win}) + \sum_i \Pr(S = i|\text{win})(1 + \frac{1}{p_i + p_7})\\ =& 1 + \sum_i \Pr(S = i|\text{win}) \cdot \frac{1}{p_i + p_7}\\ =& 1 + \sum_i \frac{p_i^2}{\Pr(\text{win})(p_i + p_7)^2}.\end{aligned}$$

□

Remark 5.48. In the above example,

- $E[X]$ is much easier than $E[X|\text{win}]$. $E[X]$ can be obtained by conditioning on S. One does not have to obtain $E[X]$ via $E[X|\text{win}]$ and $E[X|\text{lose}]$.
- Although the calculation says so, it is not intuitively clear at this point that $E[X|\text{win}, S = i] = E[X|S = i]$.

Example 5.49. Let X (resp. Y) be the number of rolls of a dice to obtain the first 6 (resp. 5). Find the conditional expectation $E[X|Y = 5]$.

solution. We can condition the expectation on whether $X < 5$ or not. Namely,

$$\begin{aligned} E[X|Y = 5] =& \Pr(X < 5|Y = 5)E[X|Y = 5, X < 5] \\ &+ \Pr(X > 5|Y = 5)E[X|Y = 5, X > 5] \end{aligned}$$

First,

$$\begin{aligned} &\Pr(X < 5|Y = 5)E[X|Y = 5, X < 5] \\ =& \sum_{k=1}^{4} k \cdot \Pr(X < 5|Y = 5) \cdot \frac{\Pr(X = k, Y = 5)}{\Pr(X < 5, Y = 5)} \\ =& \sum_{k=1}^{4} k \cdot \Pr(X = k|Y = 5) \\ =& \sum_{k=1}^{4} k \cdot (\frac{4}{5})^{k-1} \cdot \frac{1}{5}. \end{aligned}$$

Next,

$$\begin{aligned} &\Pr(X > 5|Y = 5)E[X|Y = 5, X > 5] \\ =& \Pr(X > 5|Y = 5) \cdot \sum_{k=1}^{\infty} (k + 5) \cdot \frac{\Pr(X = k + 5, Y = 5)}{\Pr(X > 5, Y = 5)} \\ =& \Pr(X > 5|Y = 5) \cdot \sum_{k=1}^{\infty} (k + 5) \cdot \Pr(X = k) \\ =& (\frac{4}{5})^4 \cdot (5 + E[X]) = (\frac{4}{5})^4 \cdot (5 + 6). \end{aligned}$$

□

Example 5.50. There are two biased coins. The probabilities for landing on heads are 0.4 and 0.7 respectively. One coin is randomly selected and is flipped 10 times. Given that there are 2 heads in the first 3 flips, what is the expected number of heads in the 10 flips?

solution. Let F be the event that there are two heads in the first 3 flips. Also let A (resp. B) be the event that the coin whose probability to land on head is 0.4 (resp. 0.7) is flipped. Finally, let X be the number of heads in the 10 flips. We will find $E[X|F]$ by conditioning on which coin is flipped. Namely,

$$E[X|F] = \Pr(A|F) \cdot E[X|F, A] + \Pr(B|F) \cdot E[X|F, B]$$

The trick is that whenever we condition on a specific coin, the flips become independent. The number of heads given a specific coin is binomial. Therefore,

$$E[X|F, A] = 2 + 7 \cdot 0.4 \text{ and } E[X|F, B] = 2 + 7 \cdot 0.7,$$

where we were already given 2 heads in the first 3 flips. It follows that

$$E[X|F] = 2 + \Pr(A|F) \cdot 7 \cdot 0.4 + \Pr(B|F) \cdot 7 \cdot 0.7.$$

It remains to find $\Pr(A|F)$ and $\Pr(B|F)$. They are given as

$$\begin{aligned}\Pr(A|F) =& \frac{\Pr(A) \cdot \Pr(F|A)}{\Pr(A) \cdot \Pr(F|A) + \Pr(B) \cdot \Pr(F|B)} \\ =& \frac{\frac{1}{2} \cdot \binom{3}{2} \cdot 0.4^2 \cdot 0.6}{\frac{1}{2} \cdot \binom{3}{2} \cdot 0.4^2 \cdot 0.6 + \frac{1}{2} \cdot \binom{3}{2} \cdot 0.7^2 \cdot 0.3}, \\ \Pr(B|F) =& \frac{\Pr(B) \cdot \Pr(F|B)}{\Pr(A) \cdot \Pr(F|A) + \Pr(B) \cdot \Pr(F|B)} \\ =& \frac{\frac{1}{2} \cdot \binom{3}{2} \cdot 0.7^2 \cdot 0.3}{\frac{1}{2} \cdot \binom{3}{2} \cdot 0.4^2 \cdot 0.6 + \frac{1}{2} \cdot \binom{3}{2} \cdot 0.7^2 \cdot 0.3}.\end{aligned}$$

□

6. Order statistics

This section deals with the ordering of independent identically distributed random variables. The most relevant topics are probably computing the probability density functions of the maximum, the minimum, and the range of a sample of uniform random variables. These are dealt with in Example 6.4. We begin by formalizing the concept of ordering in probability.

Let $X_1, X_2, \dots, X_n$ be n independent and identically distributed continuous random variables. Define the reordering of the n random variables

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}.$$

Namely, $X_{(1)}, \dots, X_{(n)}$ are the ordered values of $X_1, \dots, X_n$. The values $X_{(1)}, \dots, X_{(n)}$ are known as the *order statistics corresponding to* $X_1, \dots, X_n$.

Suppose f (resp. F) is the common probability density function (resp. cumulative density function). The joint probability density function of $X_{(1)}, \dots, X_{(n)}$ is then given by

$$n! f(x_1) f(x_2) \cdots f(x_n), x_1 < x_2 < \cdots < x_n.$$

The n factorial accounts for the fact that there are $n!$ orderings possible among $X_1 \dots, X_n$, and each ordering is equally likely.

One can then obtain the probability density function of each $X_{(i)}$ by integrating the joint probability density function. One can integrate with respect to $x_1, x_2, \dots x_{i-1}$, and then integrate with respect to $x_{i+1}, \dots, x_n$. The range for dx_1 would be $(-\infty, x_2)$. The range for dx_k for $2 \leq k \leq i-1$ would be $(-\infty, x_{k+1})$. Finally, the range for dx_k for $n > k > i$ would be $[x_i, x_{k+1}]$, and the range for dx_n would be $[x_i, \infty)$. The result would be

$$\frac{n!}{(i-1)! \cdot 1! \cdot (n-i)!} F(x_i)^{i-1} f(x_i) (1 - F(x_i))^{n-i}. \tag{6.1}$$

Equation (6.1) can be interpreted as the following: For $X_{(i)}$ to be close to x_i, one requires

- $(i-1)$ values among $X_1, X_2, \dots, X_n$ to be smaller than x_i,
- 1 value from the remaining of $X_1, X_2, \dots, X_n$ to be close to x_i, and
- the last $n-i$ values to be bigger than x_i.

We can also obtain the joint probability density function of $X_{(i)}, X_{(j)}$ for $i < j$ by integration. The result would be

$$\begin{aligned}\frac{n!}{(i-1)!\cdot 1!\cdot (j-i-1)!\cdot 1!\cdot (n-j)!}F(x_i)^{i-1}f(x_i)\\ \times (F(x_j)-F(x_i))^{j-i-1}f(x_j)[1-F(x_j)]^{n-j}.\end{aligned} \tag{6.2}$$

We can interpret Equation (6.2) as the following: For $X_{(i)}$ and $X_{(j)}$ to be close to x_i and x_j respectively, we divide $X_1, \ldots, X_n$ into 5 groups, where

- the first group has $(i-1)$ values, all of which are smaller than x_i,
- the second group has 1 value close to x_i,
- the third group has $(j-i-1)$ values, all of which are between x_i and x_j,
- the fourth group has 1 value close to x_j,
- the fifth group has $(n-j)$ values, all of which are bigger than x_j.

Example 6.1. If three trucks break down at points randomly distributed on a road of length L, find the probability that no 2 of the trucks are within a distance of d when $d \le L/2$.

solution. Let the position of the three trucks be ordered as

$$X_{(1)} < X_{(2)} < X_{(3)}.$$

Their joint distribution is

$$f(x_1, x_2, x_3) = \frac{3!}{L^3}, x_1 < x_2 < x_3.$$

We then require $X_{(1)} < X_{(2)} - d$ and $X_{(2)} < X_{(3)} - d$. The probability in question is obtained by the integral

$$\begin{aligned}\int_{2d}^{L}\int_{d}^{x_3-d}\int_{0}^{x_2-d}\frac{3!}{L^3}\,dx_1\,dx_2\,dx_3 &= \frac{3!}{L^3}\int_{2d}^{L}\int_{d}^{x_3-d}(x_2-d)\,dx_2\,dx_3\\ &=\frac{3!}{L^3}\int_{2d}^{L}\frac{1}{2}(x_3-2d)^2\,dx_3 = (1-\frac{2d}{L})^3.\end{aligned}$$

□

Example 6.2. If X_1, X_2, X_3 are three independent uniform distributions on $(0, 1)$, find the probability that the maximum is greater than the sum of the other two.

Proof. The joint distribution of the order statistics $X_{(1)}, X_{(2)}, X_{(3)}$ is given by

$$f(x_1, x_2, x_3) = 3!, x_1 < x_2 < x_3.$$

We require $x_3 > x_1 + x_2$ in the domain of $f(x_1, x_2, x_3)$. The projection of the corresponding region onto the x_1x_2-plane is the triangle

$$x_1 < x_2 < 1 - x_1$$

in the first quadrant. Therefore, the probability in question is the integral

$$\begin{aligned}3!\int_{0}^{1/2}\int_{x_1}^{1-x_1}\int_{x_1+x_2}^{1}dx_3\,dx_2\,dx_1 &= 3!\int_{0}^{1/2}\int_{x_1}^{1-x_1}(1-x_1-x_2)\,dx_2\,dx_1\\ &=3!\int_{0}^{1/2}\frac{-1}{2}(1-2x_1)^2\,dx_1 = \frac{1}{2}.\end{aligned}$$

□

Example 6.3. If a sample of size 3 from a uniform distribution on $(0,1)$ is observed, find the probability that the median is between $1/4$ and $3/4$.

solution. Let $X_{(1)} < X_{(2)} < X_{(3)}$ be the order statistics. The median $X_{(2)}$ has probability density function

$$\frac{3!}{1!1!1!}x_2(1-x_2), 0 < x_2 < 1.$$

Upon integration we get the answer $11/16$. □

Example 6.4. Let $X_1, X_2, \ldots, X_n$ be n independent uniform distributions on $(0,1)$. Find the probability density function for $Y = \min(X_1, \ldots, X_n)$, $Z = \max(X_1, \ldots, X_n)$, and their joint distribution. Finally, find the distribution of the range of sample $R = Z - Y$.

solution. We can directly apply Equation (6.1) and Equation (6.2), or argue as the following:

$$\begin{aligned}
\Pr(Z < z) =& \Pr(X_1, \ldots, X_n < z) = z^n \\
\Pr(Y < y) =& 1 - \Pr(Y \geq y) = 1 - (1-y)^n \\
\Pr(Y < y, Z < z) =& \Pr(Z < z) - \Pr(Z < z, Y \geq y) \\
=& \Pr(Z < z) - \Pr(y \leq X_1, X_2, \ldots, X_n < z) \\
=& z^n - (z-y)^n.
\end{aligned}$$

Therefore,

- The probability density function of Z is
$$\frac{d}{dz}\Pr(Z < z) = nz^{n-1}, 0 < 1 < z;$$
- The probability density function of Y is
$$\frac{d}{dy}\Pr(Y < y) = n(1-y)^{n-1}, 0 < y < 1;$$
- The joint probability density function is
$$\frac{\partial^2}{\partial y \partial z}\Pr(Y < y, Z < z) = n(n-1)(z-y)^{n-2}, 0 < y < z < 1.$$

To derive the probability density function of the range R, we note that

$$\begin{aligned}
\Pr(R < r) =& 1 - \Pr(R \geq r) = 1 - \Pr(Z \geq Y + r) \\
=& 1 - \int_0^{1-r}\int_{y+r}^1 n(n-1)(z-y)^{n-2}\,dz\,dy \\
=& n(1-r)r^{n-1} + r^n.
\end{aligned}$$

Whence the probability density function of R is

$$\frac{d}{dr}\Pr(R < r) = n(n-1)r^{n-2}(1-r), 0 < r < 1.$$

□

Example 6.5. Let $X_1, X_2, \ldots, X_n$ be n independent uniform distributions on $(0,1)$. Find the expected value of $Y = \min(X_1, \ldots, X_n)$, $Z = \max(X_1, \ldots, X_n)$. In the case that $n = 2$, find the variances of Y and Z, and their correlation.

solution. By Example 6.4,

$$E[Y] = \int_0^1 ny \cdot (1-y)^{n-1}\,dy = \frac{1}{n+1}$$
$$E[Z] = \int_0^1 nz \cdot z^{n-1}\,dz = \frac{n}{n+1}.$$

When $n = 2$, we get $E[Y] = \frac{1}{3}$ and $E[Z] = \frac{2}{3}$. To find the correlation, we need

$$E[Y^2] = \int_0^1 2y^2(1-y)\,dy = \frac{1}{6},$$
$$E[Z^2] = \int_0^1 2z^3\,dz = \frac{1}{2},$$
$$E[YZ] = \int_0^1 \int_y^1 2yz\,dz\,dy = \frac{1}{4}.$$

Hence, $\mathrm{Var}(Y) = \frac{1}{18}, \mathrm{Var}(Z) = \frac{1}{18}, \mathrm{Cov}(Y,Z) = \frac{1}{36}$. These yield the correlation

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

□

Example 6.6 (Random ants)**.** Five hundred ants are placed randomly and simultaneously on a string of 1 unit length. Each ant randomly moves towards either end at a constant speed of 1 unit per minute until it reaches the end and falls off. Whenever two ants meet, they change to the opposite directions. Find the expected time for all ants to fall off the string.

solution. This seemingly intimidating problem becomes easy once we imagine that the ants exchange heads whenever they meet. Namely, the other ant carries on the rest of the journey. Now let X_i for $i = 1, 2, \ldots, 500$ be the time before the i-th ant falls off the string, provided that it never changes directions once it starts. These random variables are independent uniform random variables on the interval $(0, 1)$. The expected time in question is precisely

$$\max(X_1, X_2, \ldots, X_{500}) = \frac{500}{501}$$

by Example 6.4. □

Example 6.7. Let $X_1, X_2, \ldots, X_6$ be 6 continuous independent and identically distributed random variables. Find

$$\Pr(X_6 > X_2 | X_1 = \max(X_1, X_2, \ldots, X_5)).$$

solution. It is immediate that

$$\Pr(X_1 = \max(X_1, X_2, \ldots, X_5)) = \frac{1}{5}.$$

It remains to find

$$\Pr(X_6 > X_2, X_1 = \max(X_1, \ldots, X_5)).$$

For this, we condition on whether $X_6 > X_1$ or not.

$$\begin{aligned}
&\Pr(X_6 > X_2, X_1 = \max(X_1, \ldots, X_5)) \\
=&\Pr(X_6 > X_1 = \max(X_1, \ldots, X_5)) + \Pr(X_1 = \max(X_1, \ldots, X_6), X_6 > X_2) \\
=&\frac{1}{6} \cdot \frac{1}{5} + \frac{1}{6} \cdot \frac{1}{2} = \frac{7}{60}.
\end{aligned}$$

Therefore, the conditional probability in question is $7/12$. □

7. Counterexamples

Example 7.1. If events A, B are independent and events B, C are independent, are events A, C independent?

solution. Consider the three subsets $A = \{2, 5\}$, $B = \{2, 4, 6\}$, $C = \{4, 5\}$ of the outcomes of rolling a fair six-sided dice. We then have $\Pr(AB) = \Pr(A)\Pr(B)$ and $\Pr(BC) = \Pr(B)\Pr(C)$. Nevertheless,

$$\Pr(AC) = 1/6 \neq \Pr(A)\Pr(C).$$

Therefore, independence is not transitive. □

Example 7.2. If events B, C are independent and events B, A are independent, are events $B, (A \cup C)$ independent?

solution. Use the subsets from Example 7.1 to demonstrate the negative answer to the problem posed. □

Example 7.3. If events A, C are independent, is it true that

$$\Pr(AC|B) = \Pr(A|B)\Pr(C|B) \text{ for any event } B?$$

solution. Use the same subsets from Example 7.1 to demonstrate the negative answer to the problem posed. □

Example 7.4. If events A, B, events B, C, and events A, C are independent, is it necessarily true that A, B, C are jointly independent? Namely, does

$$\Pr(ABC) = \Pr(A)\Pr(B)\Pr(C)$$

also hold?

solution. Consider the subsets $A = \{1, 2\}$, $B = \{2, 3\}$, and $C = \{1, 3\}$ of the outcomes of rolling a fair four-sided dice. One checks easily that the conditions stated in the problem are all satisfied, yet

$$\Pr(ABC) = 0 \neq \Pr(A)\Pr(B)\Pr(C).$$

□

Example 7.5. Let X be a random variable with mean μ. Is it true that

$$\Pr(X \leq \mu) = \frac{1}{2}?$$

solution. While this is true for symmetric distributions such as the normal distributions and the uniform distributions, this is not true in general. Consider the exponential random variable X with mean $\frac{1}{\lambda}$. The cumulative density function is $F(t) = 1 - e^{-\lambda t}$. We see that

$$F\left(\frac{1}{\lambda}\right) = 1 - e^{-1} \neq \frac{1}{2}.$$

□

Example 7.6. A deck of 52 cards is shuffled and a bridge of 13 cards is dealt out. Let X and Y denote respectively the number of aces and spades in the hand. The two random variables X, Y are not independent, but uncorrelated.

solution. That X, Y are not independent can be established by

$$\Pr(X = Y = 0) \neq \Pr(X = 0) \cdot \Pr(Y = 0).$$

To show that X, Y are uncorrelated, we define X_1, X_2, X_3, X_4 and $Y_1, Y_2 \ldots, Y_{13}$ as indicators for the aces and the spades respectively. Note that there is a spade ace. For simplicity, we let X_1 and Y_1 be the spade ace. We then have

$$\begin{aligned} E[XY] =& 51 \cdot \Pr(X_i = Y_j = 1) + \Pr(X_1 = Y_1 = 1) \text{ for any } i, j \neq 1 \\ =& 51 \cdot \frac{13 \cdot 12}{52 \cdot 51} + \frac{13}{52} = \frac{13}{4}. \end{aligned}$$

On the other hand,

$$E[X] \cdot E[Y] = 4 \cdot \frac{13}{52} \cdot 13 \cdot \frac{13}{52} = \frac{13}{4}.$$

Hence $E[XY] = E[X]E[Y]$ as desired. □

References

[oA] Society of Actuary. Society of actuary probability exams. https://www.soa.org/education/exam-req/edu-exam-p-detail/.

[Ros14] Sheldon Ross. *A first course in probability*. Pearson, nine edition, 2014.

[Zho20] Xinfeng Zhou. *A practical guide to quantitative finance interviews*. Two edition, 2020.