STAT 371

Problems in discrete probability

Fall 2013

Byron Schmuland

Be discreet in all things, and so render it unnecessary to be mysterious.

Contents

1 Patterns

(a) HT

How many fair coin tosses on average are needed to see the pattern "HT"?

2	3	4	5	6	···
HT	HHT	HHHT	HHHHT	HHHHHT	···
	THT	THHT	THHHT	THHHHT	···
		TTHT	TTHHT	TTHHHT	···
			TTTHT	TTTHHT	···
				TTTTHT	···

There are $n-1$ outcomes of length n, all of the form

$$\underbrace{T\cdots T}_{t\geq 0}\underbrace{H\cdots H}_{h\geq 0}HT$$

with $t+h=n-2$.

An Aside. Since $\mathbb{P}(N_1=n)=(n-1)/2^n$ and the probabilities must add to 1, we conclude that $\sum_{n=2}^{\infty}\frac{n-1}{2^n}=1$.

Using the identity $n(n-1)=(n-2)(n-1)+2(n-1)$ we find that

$$\begin{aligned}\mathbb{E}(N_1) &= \sum_{n=2}^{\infty} n\,\mathbb{P}(N_1=n)=\sum_{n=2}^{\infty}\frac{n(n-1)}{2^n}\\ &= \sum_{n=2}^{\infty}\frac{(n-2)(n-1)}{2^n}+2\sum_{n=2}^{\infty}\frac{n-1}{2^n}\\ &= \frac{1}{2}\,\mathbb{E}(N_1)+2,\end{aligned}$$

which shows that $\mathbb{E}(N_1)=4$.

(b) HH

How many fair coin tosses on average are needed to see the pattern "HH"?

2	3	4	5	6	...
HH	THH	TTHH	TTTHH	TTTTHH	...
		HTHH	THTHH	TTHTHH	...
			HTTHH	THTTHH	...
				HTTTHH	...
				HTHTHH	...

The outcomes of length n are formed in two ways: by sticking a T in front of an outcome of length $n-1$ or sticking HT in front of an outcome of length $n-2$. This way we find that the number of outcomes of length n is $F(n-1)$ where F refers to the Fibonacci numbers.

Fibonacci numbers. They are defined by $F(0) = 0$, $F(1) = 1$, and $F(n) = F(n-1) + F(n-2)$ for $n \geq 2$. We have proved that $\sum_{n=2}^{\infty} \frac{F(n-1)}{2^n} = 1$.

The solution to our problem is

$$\mathbb{E}(N_2) = \sum_{n=2}^{\infty} \frac{n\,F(n-1)}{2^n} = 6.$$

(c) HH again

How many fair coin tosses on average are needed to see the pattern "HH"?

Conditioning on the first two tosses, there are three cases T, HT, or HH giving us

$$\mathbb{E}(N_2) = \frac{1}{2}[1 + \mathbb{E}(N_2)] + \frac{1}{4}[2 + \mathbb{E}(N_2)] + \frac{1}{4}[2],$$

which gives $\mathbb{E}(N_2) = 6$.

(d) HT again

How many fair coin tosses on average are needed to see the pattern "HT"? Conditioning on the first toss, there are two cases T and H giving us

$$\mathbb{E}(N_1) = \frac{1}{2}[1 + \mathbb{E}(N_1)] + \frac{1}{2}[1 + \mathbb{E}(Z)],$$

where Z is the number of further tosses needed to see HT, given an initial H.

Conditioning on the first toss, there are two cases T and H giving us

$$\mathbb{E}(Z) = \frac{1}{2}[1] + \frac{1}{2}[1 + \mathbb{E}(Z)],$$

which implies $\mathbb{E}(Z) = 2$. Plugging this back into the previous equation gives $\mathbb{E}(N_1) = 4$.

(e) Patterns with a head start

Example 1-1 Given two heads, how many further fair coin tosses on average are needed to see the pattern "HH" again?

1	3	4	5	6	⋯
HH|H	HH|THH	HH|TTHH	HH|TTTHH	HH|TTTTHH	⋯
			HH|THTHH	HH|TTHTHH	⋯
				HH|THTTHH	⋯

Conditioning on the first toss, there are two cases T and H giving us

$$\mathbb{E}(N_4) = \frac{1}{2}[1 + \mathbb{E}(N_2)] + \frac{1}{2}[1],$$

which gives $\mathbb{E}(N_4) = 4$.

Example 1-2 Given HT, how many further fair coin tosses on average are needed to see the pattern "HT" again?

2	3	4	5	6	$\cdots$
HT | HT	HT | HHT	HT | HHHT	HT | HHHHT	HT | HHHHHT	$\cdots$
	HT | T HT	HT | T HHT	HT | T HHHT	HT | T HHHHT	$\cdots$
		HT | T T HT	HT | T T HHT	HT | T T HHHT	$\cdots$
			HT | T T T HT	HT | T T T HHT	$\cdots$
				HT | T T T T HT	$\cdots$

In this case, the head start is useless and we get back the original problem. That is, $\mathbb{E}(N_3) = \mathbb{E}(N_1) = 4$.

Example 1-3 For the pattern HTHT things get a bit complicated. We solve it by introducing new notation and using linear algebra. For $j = 0, 1, 2, 3$, let e_j be the expected number of steps to see the pattern HTHT given a head start of its first j symbols.

Arguing as before, we get the following relations

$$\begin{aligned}
e_0 &= \frac{1}{2}(1 + e_1) + \frac{1}{2}(1 + e_0) \\
e_1 &= \frac{1}{2}(1 + e_1) + \frac{1}{2}(1 + e_2) \\
e_2 &= \frac{1}{2}(1 + e_3) + \frac{1}{2}(1 + e_0) \\
e_3 &= \frac{1}{2}(1 + e_1) + \frac{1}{2}(1 + 0).
\end{aligned}$$

These relations give the vector equation

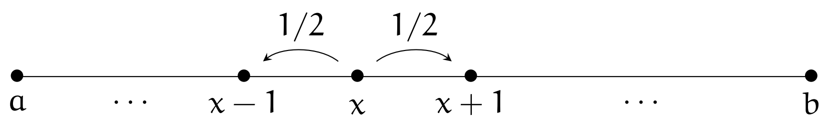
$$\begin{pmatrix} e_0 \\ e_1 \\ e_2 \\ e_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & 0 \end{pmatrix} \begin{pmatrix} e_0 \\ e_1 \\ e_2 \\ e_3 \end{pmatrix},$$

which can be solved to give

$$(e_0, e_1, e_2, e_3) = (20, 18, 16, 10).$$

2 First step analysis

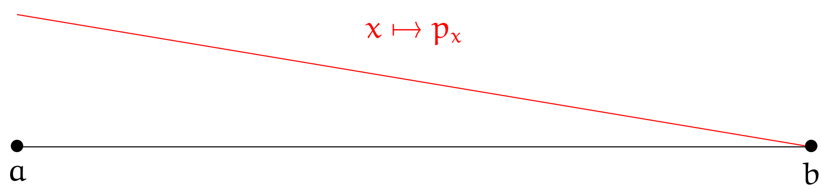
(a) Classical random walk



Place Let $a < b$ and consider the simple random walk with boundary points a and b. Define p_x as the probability that the random walk is absorbed at a starting at state x. These probabilities satisfy

$$p_a = 1, \; p_b = 0, \; p_x = \frac{p_{x-1}}{2} + \frac{p_{x+1}}{2} \; \text{ for } a < x < b.$$

This implies that $x \mapsto p_x$ is a straight line function; that is, $p_x = \frac{b-x}{b-a}$.



Details: Write the equation as

$$\frac{1}{2}(p_x - p_{x+1}) = \frac{1}{2}(p_{x-1} - p_x).$$

We have $p_x - p_{x+1} = c$ for some constant c, and since $p_b = 0$ this leads to

$$p_x = p_x - p_b = \sum_{j=x}^{b-1}(p_j - p_{j+1}) = (b-x)c.$$

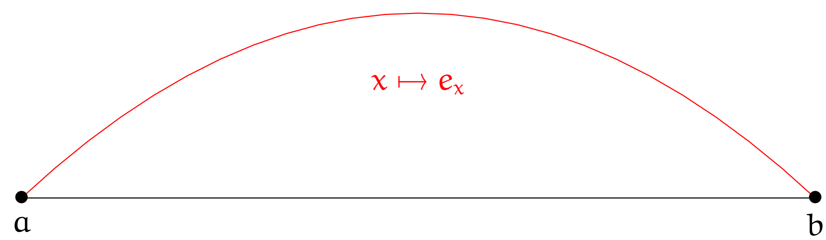
Plugging in $x = a$ gives $1 = p_a = (b-a)c$ so that $c = 1/(b-a)$, and hence

$$p_x = \frac{b-x}{b-a}.$$

Time Define e_x to be the expected number of steps until absorption starting at state x. These satisfy

$$e_a = e_b = 0,\ e_x = 1 + \frac{e_{x-1}}{2} + \frac{e_{x+1}}{2} \text{ for } a < x < b.$$

This forces $x \mapsto e_x$ to be a quadratic, $e_x = -x^2 + c_1 x + c_2$, and since we know the two roots a and b, we conclude that $e_x = (b-x)(x-a)$.



(b) Number of walks until no shoes

Someone lives in a house with two doors. He begins with n pairs of shoes at each door. Every day he randomly selects a door, puts on shoes, goes for a walk, and returns to a randomly chosen door. He leaves his shoes when he comes in the house.

What is the expected number of walks until he discovers no shoes available when he wants to take a walk?

For $0 \leq i \leq 2n$, let e_i be the expected number of completed walks starting with i pairs of shoes at the front door. For $1 \leq i \leq 2n-1$ considering the next walk, we get

$$e_i = 1 + \frac{e_{i-1}}{4} + \frac{e_i}{2} + \frac{e_{i+1}}{4}. \tag{1}$$

Some details: Equation (1) gives $\Delta^2 e_j = -4$ for $0 \leq j < 2n-1$. Summing gives us $\Delta e_i = \Delta e_0 - 4i$ for $0 \leq i < 2n$. Summing again gives

$$e_i = e_0 + [2 + \Delta e_0]i - 2i^2 \text{ for } 0 \leq i \leq 2n \tag{2}$$

For $i = 0$, a first step analysis gives

$$e_0 = \frac{1}{2} \cdot 0 + \frac{1}{4}[1 + e_1] + \frac{1}{4}[1 + e_0],$$

so that

$$e_0 = \frac{2 + e_1}{3}.$$

Plugging this into (2) gives us

$$e_i = e_0 + 2e_0 i - 2i^2. \tag{3}$$

Finally, by symmetry, we have $e_0 = e_{2n}$. Plugging this into (3) gives us the answer

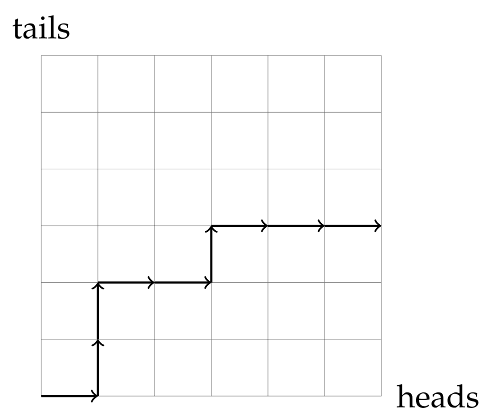
$$e_i = 2n + 4ni - 2i^2. \tag{4}$$

(c) Banach's matchbox problem

Professor Stefan Banach has, in each of his two pockets, a box with m matches. Now and then he randomly uses a match from a randomly chosen box. Find $\mathbb{E}(R)$ where R is the number of remaining matches when one box is found empty.

This problem is equivalent to tossing a fair coin until either $m+1$ heads, or $m+1$ tails comes up. If N is the number of coin tosses, then $N = (m+1) + (m-R)$. Thus, $\mathbb{E}(R) = 2m + 1 - \mathbb{E}(N)$.

This translates to a random walk problem where we start in the lower left hand corner, and take steps up or to the right until we hit a boundary.



Let $p_k = \mathbb{P}(N = k)$ for $m + 1 \leq k \leq 2m + 1$. We get $N = k$ by either m heads in the first $k - 1$ tosses followed by a head, or m tails in the first $k - 1$ tosses followed by a tail. By symmetry, we find

$$p_k = 2\binom{k-1}{m}\left(\frac{1}{2}\right)^{k-1}\left(\frac{1}{2}\right) = \binom{k-1}{m}\left(\frac{1}{2}\right)^{k-1}$$

An Aside.

$$\sum_{k\leq 2m} \binom{k}{m} \left(\frac{1}{2}\right)^k = 1$$

Time to calculate!

$$\begin{aligned}
\mathbb{E}(N) &= \sum_{k\leq 2m+1} k\binom{k-1}{m} \left(\frac{1}{2}\right)^{k-1} \\
&= \sum_{k\leq 2m+1} (m+1)\binom{k}{m+1} \left(\frac{1}{2}\right)^{k-1} \\
&= 2(m+1) \sum_{k\leq 2m+1} \binom{k}{m+1} \left(\frac{1}{2}\right)^{k} \\
&= 2(m+1) \left\{1 - \binom{2(m+1)}{m+1} \left(\frac{1}{2}\right)^{2(m+1)}\right\}.
\end{aligned}$$

Therefore

$$\mathbb{E}(R) = 2m+1-\mathbb{E}(N) = (2m+1)\binom{2m}{m}\left(\frac{1}{2}\right)^{2m} - 1.$$

(d) Three players

Huygens 1657 Three persons A, B, and C play a game. A bowl contains w white and b black marbles. The players successively draw a marble (with replacement) in the order $ABCABC\ldots$ until someone gets a white marble. He is the winner. Find the winning chances for the three players.

Solution: Letting $\alpha = \frac{w}{w+b}$ and $\beta = \frac{b}{w+b}$, first step analysis shows that $p_A = \alpha + \beta^3 p_A$. Similarly $p_B = \beta\alpha + \beta^3 p_B$, and $p_C = \beta^2\alpha + \beta^3 p_C$. Therefore

$$p_A = \frac{\alpha}{1-\beta^3}, \quad p_B = \frac{\beta\alpha}{1-\beta^3}, \quad p_C = \frac{\beta^2\alpha}{1-\beta^3}.$$

3 Exchangeability and symmetry

(a) Sampling with replacement

Pull 52 cards from an ordinary deck, replacing each card drawn and reshuffling thoroughly. For $1 \le i \le 52$ let A_i' be the event that the ith card drawn was red.

These events are independent with $\mathbb{P}(A_i') = 1/2$ for all i. In particular,

$$\mathbb{P}(A_{i_1}' A_{i_2}' \cdots A_{i_j}') = \left(\frac{1}{2}\right)^j. \tag{1}$$

The 2^{52} possible outcomes, e.g. $\underbrace{(B, R, R, \ldots, B)}_{52 \text{ places}}$, are equally likely.

(b) Sampling without replacement

Pull all 52 cards from an ordinary deck without replacement. For $1 \le i \le 52$ let A_i be the event that the ith card drawn was red.

The $\binom{52}{26}$ possible outcomes, e.g. $\underbrace{(B, R, R, \ldots, B)}_{26 \text{ Rs and } 26 \text{ Bs}}$, are equally likely.

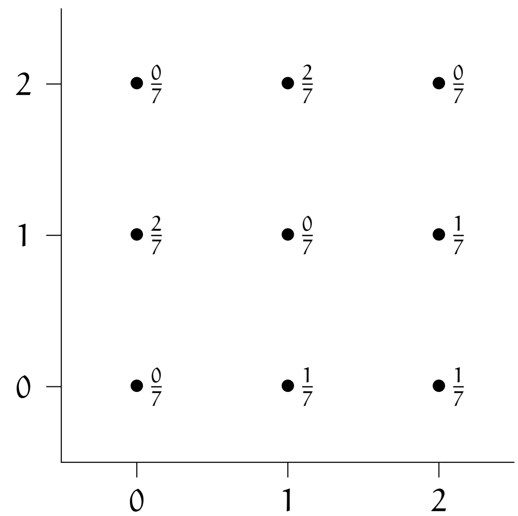
(c) Formal definition

Definition 3-1 A sequence $A_1, A_2, \ldots, A_n$ of events is called *exchangeable* if $\mathbb{P}(A_{i_1} A_{i_2} \cdots A_{i_j})$ is the same for any set of j distinct indices. Here $A_{i_1} A_{i_2} \cdots A_{i_j}$ means the intersection of these j events.

Similarly, a sequence $X_1, X_2, \ldots, X_n$ of random variables is called *exchangeable* if the vector $(X_{i_1}, X_{i_2}, \ldots, X_{i_j})$ has the same distribution for every sequence of j indices.

An alternative definition: The sequence $X_1, X_2, \ldots, X_n$ of random variables is exchangeable if and only if $(X_1, X_2, \ldots, X_n)$ and $(X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(n)})$ have the same distribution for any permutation π. That is, π is a one-to-one map from $\{1, 2, \ldots, n\}$ onto itself.

Example 3-1



Exchangeability implies being identically distributed, but not the reverse. An exchangeable sequence has extra symmetry.

For instance, the diagram above is the joint mass function of (X, Y), where X and Y take values in $\{0, 1, 2\}$. Checking the marginals shows that they are identically distributed, but

$$\mathbb{P}(X = 0, Y = 2) = 0/7 \neq \mathbb{P}(Y = 0, X = 2) = 1/7.$$

That is, the random variables X and Y have the same distribution, but the random vectors (X, Y) and (Y, X) don't.

(d) Sampling questions revisited

The independent events A'_i from example (a) are exchangeable, because of formula (1). For the events A_i from example (b), is the exchangeability "obvious for reasons of symmetry"?

Well, I'm not sure. Let's calculate the probability that the 3rd card drawn is red taking all cases into account:

$$\begin{aligned}
\mathbb{P}(A_3) &= \mathbb{P}(RRR) + \mathbb{P}(RBR) + \mathbb{P}(BRR) + \mathbb{P}(BBR) \\
&= \frac{26}{52}\cdot\frac{25}{51}\cdot\frac{24}{50} + \frac{26}{52}\cdot\frac{26}{51}\cdot\frac{25}{50} + \frac{26}{52}\cdot\frac{26}{51}\cdot\frac{25}{50} + \frac{26}{52}\cdot\frac{25}{51}\cdot\frac{26}{50} \\
&= \frac{1}{2}.
\end{aligned}$$

Details: There is a permutation of $\{1, 2, \ldots, 52\}$ that sends $(i_1, i_2, \ldots, i_j)$ to $(1, 2, \ldots, j)$.

Note that $\mathbb{P}(A_1) = 1/2$ and $\mathbb{P}(A_1 A_2) = (1/2) \cdot (25/51) = .245 \neq 1/4$ so these events are not independent.

(e) Random permutations

Randomly permute the numbers $\{1, 2, \ldots, n\}$ and let A_i be the event that number i is in the ith place.

(f) An urn model

An urn contains w white balls and b black balls. Draw the balls one at a time, without replacement, until the urn is empty. Let A_i be the event that the ith ball drawn is white.

Here is an exercise that is trivial if you use exchangeability:

$$\mathbb{P}(\text{all black balls are drawn before the last white ball}) = \frac{w}{w+b}.$$

(g) The German tank problem

Suppose that $\{x_1, x_2, \ldots, x_k\}$ is a random sample of size k, taken without replacement, from the set $\{1, 2, \ldots, N\}$. Putting $M = \max\{x_1, x_2, \ldots, x_k\}$, we want to calculate $\mathbb{E}(M)$.

Solution 1: For $k \le m \le N$, we have $\mathbb{P}(M = m) = \dfrac{\binom{m-1}{k-1}}{\binom{N}{k}}$.

An Aside. Since the probabilities must add to 1, a simple change of variables gives us the hockey stick formula:

$$\sum_{m \leq N} \binom{m}{k} = \binom{N+1}{k+1}.$$

The expected value $\mathbb{E}(M)$ is therefore

$$\sum_{m \leq N} m \frac{\binom{m-1}{k-1}}{\binom{N}{k}} = \frac{k}{\binom{N}{k}} \sum_{m \leq N} \binom{m}{k} = \frac{k}{\binom{N}{k}} \binom{N+1}{k+1} = \frac{k}{k+1}(N+1).$$

Solution 2: Here is an argument that uses the natural symmetry of the problem. For $1 \leq j \leq k$, define $x_{(j)}$ to be the jth order statistic of the set $\{x_1, x_2, \ldots, x_k\}$. To fill out these values, define $x_{(0)} = 0$ and $x_{(k+1)} = N + 1$.

For instance, if $N = 6$, $k = 3$, and $\{x_1, x_2, x_3\} = \{2, 3, 6\}$, then $x_{(1)} = 2$, $x_{(2)} = 3$, and $x_{(3)} = 6$. We also have $x_{(0)} = 0$ and $x_{(4)} = 7$.

The random variables $x_{(j+1)} - x_{(j)}$ for $j = 0, 1, \ldots, k$ are exchangeable, in particular, they are identically distributed. Now

$$\sum_{j=0}^{k} (x_{(j+1)} - x_{(j)}) = N + 1,$$

so we deduce that $\mathbb{E}(x_{(j+1)} - x_{(j)}) = \frac{N+1}{k+1}$. Writing $x_{(j)} = \sum_{i=0}^{j-1} (x_{(i+1)} - x_{(i)})$ we see that $\mathbb{E}(x_{(j)}) = \frac{j}{k+1}(N+1)$.

In particular, we have $M = x_{(k)}$ so $\mathbb{E}(M) = \frac{k}{k+1}(N+1)$.

(h) Drawing straws

The dirty dozen are a tough bunch of soldiers, and two of them have to go on a dangerous mission. In order to decide who goes, they draw straws. They begin with 10 long straws and two short straws, and draw successively until the short ones are chosen. The soldiers who get short straws have to go.

For $1 \leq i \leq 12$, let A_i be the event that soldier i draws the short straw. The argument in section (d) says that these events are exchangeable. In particular, $\mathbb{P}(A_i) = 2/12$ for all i; drawing straws is a completely fair way to make the decision.

Here is further twist. Suppose that you are able to take the first straw if you like, or to step in at any point and say "I'm next!". Can you reduce the risk of drawing a short straw? For instance, you may take your turn immediately after the first short straw. It turns out that it doesn't matter. Every strategy gives a 2/12 chance of getting a short straw, so you cannot reduce (or increase!) your risk.

4 Bijections

(a) Words, sets of sets, and perfect matchings

(C1) How many distinct "words" can be made with the letters $\{a, a, b, b, c, c\}$?

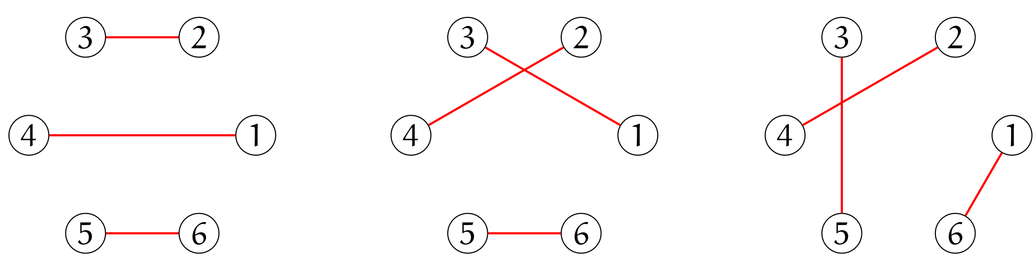
For example, $abbacc$, $babacc$, $cbabac$.

(C2) How many distinct sets of two disjoint 2-sets can be made from $\{1, 2, 3, 4, 5, 6\}$?

For example, $\{\{1, 4\}, \{2, 3\}\}$, $\{\{1, 3\}, \{2, 4\}\}$, $\{\{3, 5\}, \{2, 4\}\}$.

(C3) How many distinct ways can you match up six people into three pairs?

For example,



(b) Stars and bars

How many vectors $(x_1, x_2, \ldots, x_k)$ of positive integers are there whose sum is n? This problem has an easy answer once we set up a bijection:

$$(x_1, x_2, \ldots, x_k) \quad \longleftrightarrow \quad \underbrace{1\cdots 1}_{x_1} + \underbrace{111111\cdots 1}_{x_2} + \cdots + \underbrace{111\cdots 1}_{x_k}.$$

On the right, begin with a line of n ones and then insert $k-1$ plus signs into the $n-1$ spaces between the ones. The number of ways is $\binom{n-1}{k-1}$.

What if we allow zero, that is, $x_j \geq 0$? We could use the bijection

$$(x_1, x_2, \ldots, x_k) \quad \longleftrightarrow \quad (x_1+1, x_2+1, \ldots, x_k+1)$$

where $x_j + 1$'s are positive and add up to $n+k$. The number of different ways is therefore $\binom{n+k-1}{k-1}$.

(c) Non consecutive values

What's the chance of no consecutive values in the Lotto 6-49 draw?

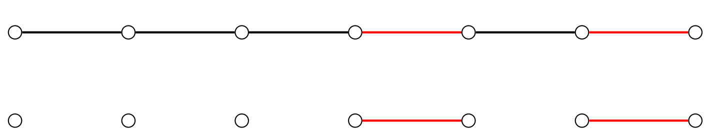
Solution: The total number of possibilities is $\binom{49}{6}$. How many of these consist of non-consecutive values? The answer, as we explain below is $\binom{44}{6}$ so the required probability is $\frac{\binom{44}{6}}{\binom{49}{6}} = .5048$.

In general, how many sets of k numbers x_j from $\{1, 2, \ldots, n\}$ with $x_1 < x_2 < \cdots < x_k$ have no consecutive values? We can use the bijection

$$(x_1, x_2, \ldots, x_k) \quad \longleftrightarrow \quad (x_1, x_2 - 1, \ldots, x_k - (k-1))$$

where $x_j - (j-1)$'s are distinct values from $\{1, 2, \ldots, n-(k-1)\}$. The number of different ways is therefore $\binom{n-k+1}{k}$.

Proof by picture: There are $\binom{5}{2}$ ways to choose two non-consecutive numbers from $\{1, 2, 3, 4, 5, 6\}$



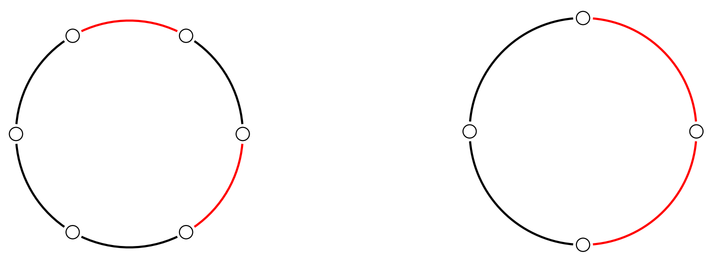
(d) Non consecutive values on a circle

How many sets of k numbers from $\{1, 2, \ldots, n\}$ arranged in a circle have no consecutive values?

Solution 1: We can take all linear arrangements without consecutive values, and then subtract out those that include both 1 and n:

$$\binom{n-k+1}{k} - \binom{n-k-1}{k-2}$$

Solution 2: It is possible to make a direct bijection, similar to the previous problem. The bijection is between patterns: on the right is an arbitrary pattern on the small circle; on the left is a pattern on the big circle without consecutive values. To go from the small circle to the big circle, move clockwise and add a black segment after every red segment.



For each pattern, the ratio of distinct numberings on the big circle compared to the small circle is $n/(n-k)$. From this we deduce that the number of arrangements is

$$\frac{n}{n-k}\binom{n-k}{k}.$$

5 The inclusion-exclusion principle

(a) One in a million

Choose one of the million numbers $000000, 000001, \ldots, 999999$ at random. Find the probability that at least one of the digits $0, 1, \ldots, 9$ appears exactly twice.

Solution 1:

Pattern	Choose values	Order them			
11112	$10 \cdot \binom{9}{4}$	$6!/2!$	1260×360	=	453,600
1122	$\binom{10}{2} \cdot \binom{8}{2}$	$6!/2!2!$	1260×180	=	226,800
123	$10 \cdot 9 \cdot 8$	$6!/3!2!$	720×60	=	43,200
222	$\binom{10}{3}$	$6!/2!2!2!$	120×90	=	10,800
24	$10 \cdot 9$	$6!/4!2!$	90×15	=	1,350
					735,750

This event has a probability of 0.73575.

Note: The values in the second and third columns of the table above may seem a bit *ad hoc*. The pattern becomes visible if you write them as *multinomial coefficients*. For instance, the two values in the first row can be written

$$\binom{10}{5, 4, 1} \quad \binom{6}{1, 1, 1, 1, 2}.$$

Under the 10 we have $5, 4, 1$ meaning that, for this pattern, out of the ten digits 5 of them will appear zero times, 4 of them will appear once, and 1 of them will appear twice. Under the six, we just put in the pattern with commas.

Similarly, the other four rows are

$$\binom{10}{6, 2, 2} \quad \binom{6}{1, 1, 2, 2}$$
$$\binom{10}{7, 1, 1, 1} \quad \binom{6}{1, 2, 3}$$
$$\binom{10}{7, 0, 3} \quad \binom{6}{2, 2, 2}$$
$$\binom{10}{8, 0, 1, 0, 1} \quad \binom{6}{2, 4}$$

Important formula. Let $A_1, A_2, \ldots, A_n$ be events. The probability that exactly k of them occur ($k \geq 0$) is

$$p_n(k) = \sum_{j \geq k} (-1)^{j-k} \binom{j}{k} S_j,$$

where $S_0 = 1$, $S_1 = \sum_i \mathbb{P}(A_i)$, $S_2 = \sum_{i<j} \mathbb{P}(A_i A_j)$, $S_3 = \sum_{i<j<k} \mathbb{P}(A_i A_j A_k)$, etc.

Solution 2: For $i = 0, 1, \ldots, 9$ define A_i as the event that the digit i appears exactly twice. Using binomial and multinomial probabilities we get

$$\begin{aligned}
P_1 &= \mathbb{P}(A_0) = \frac{6!}{2!4!}\left(\frac{1}{10}\right)^2\left(\frac{9}{10}\right)^4 = \frac{98{,}415}{1{,}000{,}000} \\
P_2 &= \mathbb{P}(A_0 A_1) = \frac{6!}{2!2!2!}\left(\frac{1}{10}\right)^2\left(\frac{1}{10}\right)^2\left(\frac{8}{10}\right)^2 = \frac{5{,}760}{1{,}000{,}000} \\
P_3 &= \mathbb{P}(A_0 A_1 A_2) = \frac{6!}{2!2!2!}\left(\frac{1}{10}\right)^2\left(\frac{1}{10}\right)^2\left(\frac{1}{10}\right)^2 = \frac{90}{1{,}000{,}000}
\end{aligned}$$

For exchangeable events: $S_j = \binom{n}{j} P_j.$

Plugging into the formula with $n = 10$, $k = 0$ gives the probability that no digit appears exactly twice to be

$$\binom{10}{0} P_0 - \binom{10}{1} P_1 + \binom{10}{2} P_2 - \binom{10}{3} P_3 = \frac{1{,}057}{4{,}000}.$$

The required probability is one minus this, i.e., $\frac{2{,}943}{4{,}000} = .73575$.

Solution 3: For every pair of distinct indices $i, j \in \{1, 2, \ldots, 6\}$ define $A_{i,j}$ to be the event that the ith and jth value are the same, and the rest are all different from this value.

There are $\binom{6}{2}$ such events with probability $\mathbb{P}(A_{i,j}) = (10 \cdot 9^4)/10^6$.

These events are not exchangeable, but their double intersections $A_{i,j}A_{k,\ell}$ comes in two types:

$$\mathbb{P}(A_{i,j}A_{k,\ell}) = \begin{cases} 0 & \text{if } \{i,j\} \cap \{k,\ell\} \neq \emptyset \\ (10 \cdot 9 \cdot 8^2)/10^6 & \text{if } \{i,j\} \cap \{k,\ell\} = \emptyset. \end{cases}$$

The triple intersection $A_{i,j}A_{k,\ell}A_{m,n}$ has non-zero probability only if the pairs $\{i,j\}$, $\{k,\ell\}$, and $\{m,n\}$ are disjoint, in which case

$$\mathbb{P}(A_{i,j}A_{k,\ell}A_{m,n}) = (10 \cdot 9 \cdot 8)/10^6.$$

Putting it all together, the probability of no pairs is

$$1 - \binom{6}{2}\frac{10 \cdot 9^4}{10^6} + \binom{6}{2}\binom{4}{2}\frac{1}{2!}\frac{10 \cdot 9 \cdot 8^2}{10^6} - \binom{6}{2}\binom{4}{2}\frac{1}{3!}\frac{10 \cdot 9 \cdot 8}{10^6} = \frac{1{,}057}{4{,}000}.$$

The required probability is one minus this, i.e., $\frac{2{,}943}{4{,}000} = .73575$.

(b) Another one in a million

Choose one of the million numbers $000000, 000001, \ldots 999999$ at random. Find the probability that at least two of the digits $0, 1, \ldots, 9$ appears exactly once.

Solution 1:

Pattern	Choose values	Order them			
111111	$\binom{10}{4,6}$	$\binom{6}{1,1,1,1,1,1}$	210×720	=	151,200
114	$\binom{10}{7,2,0,0,1}$	$\binom{6}{1,1,4}$	360×30	=	10,800
11112	$\binom{10}{5,4,1}$	$\binom{6}{1,1,1,1,2}$	1260×360	=	453,600
1113	$\binom{10}{6,3,0,1}$	$\binom{6}{1,1,1,3}$	840×120	=	100,800
1122	$\binom{10}{6,2,2}$	$\binom{6}{1,1,2,2}$	1260×180	=	226,800
					943,200

This event has a probability of 0.9432.

Another important formula. Let $A_1, A_2, \ldots, A_n$ be events. The probability that k or more of them occur ($k \geq 1$) is

$$P_n(k) = \sum_j (-1)^{j-k} \binom{j-1}{k-1} S_j.$$

Solution 2: For $i = 0, 1, \ldots, 9$ define A_i as the event that the digit i appears exactly once. Using binomial and multinomial probabilities we get

$$\begin{aligned}
P_2 &= \mathbb{P}(A_0A_1) = \frac{6!}{4!}\left(\frac{1}{10}\right)^2\left(\frac{8}{10}\right)^4 = \frac{122{,}880}{10^6} \\
P_3 &= \mathbb{P}(A_0A_1A_2) = \frac{6!}{3!}\left(\frac{1}{10}\right)^3\left(\frac{7}{10}\right)^3 = \frac{41{,}160}{10^6} \\
P_4 &= \mathbb{P}(A_0A_1A_2A_3) = \frac{6!}{2!}\left(\frac{1}{10}\right)^4\left(\frac{6}{10}\right)^2 = \frac{12{,}960}{10^6} \\
P_5 &= \mathbb{P}(A_0A_1A_2A_3A_4) = 6!\left(\frac{1}{10}\right)^5\left(\frac{5}{10}\right)^1 = \frac{3{,}600}{10^6} \\
P_6 &= \mathbb{P}(A_0A_1A_2A_3A_4A_5) = 6!\left(\frac{1}{10}\right)^6\left(\frac{4}{10}\right)^0 = \frac{720}{10^6}
\end{aligned}$$

Notice that $P_7 = P_8 = P_9 = P_{10} = 0$. Now using the formula with $k = 2$ and $n = 10$ we get the desired probability

$$\begin{aligned}
&\mathbb{P}(\text{at least two digits appear exactly once}) \\
= &\binom{1}{1}\binom{10}{2}P_2 - \binom{2}{1}\binom{10}{3}P_3 + \binom{3}{1}\binom{10}{4}P_4 - \binom{4}{1}\binom{10}{5}P_5 + \binom{5}{1}\binom{10}{6}P_6 \\
= &[5{,}529{,}600 - 9{,}878{,}400 + 8{,}164{,}800 - 3{,}628{,}800 + 756{,}000]/10^6 \\
= &943{,}200/10^6.
\end{aligned}$$

Solution 3: For $i = 1, \ldots, 6$ define A'_i as the event that the digit in position i appears exactly once. These are exchangeable events with

$$P'_2 = \mathbb{P}(A'_1A'_2) = \frac{10 \cdot 9 \cdot 8^4}{10^6} = \frac{368{,}640}{10^6}$$

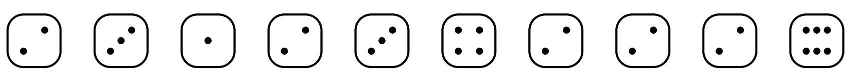
$$
\begin{aligned}
P'_3 &= \mathbb{P}(A'_1A'_2A'_3) = \frac{10\cdot 9\cdot 8\cdot 7^3}{10^6} = \frac{246{,}960}{10^6}\\
P'_4 &= \mathbb{P}(A'_1A'_2A'_3A'_4) = \frac{10\cdot 9\cdot 8\cdot 7\cdot 6^2}{10^6} = \frac{181{,}440}{10^6}\\
P'_5 &= \mathbb{P}(A'_1A'_2A'_3A'_4A'_5) = \frac{10\cdot 9\cdot 8\cdot 7\cdot 6\cdot 5}{10^6} = \frac{151{,}200}{10^6}\\
P'_6 &= \mathbb{P}(A'_1A'_2A'_3A'_4A'_5A'_6) = \frac{10\cdot 9\cdot 8\cdot 7\cdot 6\cdot 5}{10^6} = \frac{151{,}200}{10^6}
\end{aligned}
$$

Using the formula with $k = 2$ and $n = 6$ we get the desired probability

$$
\begin{aligned}
&\mathbb{P}(\text{at least two digits appear exactly once})\\
=\;& \binom{1}{1}\binom{6}{2}P'_2 - \binom{2}{1}\binom{6}{3}P'_3 + \binom{3}{1}\binom{6}{4}P'_4 - \binom{4}{1}\binom{6}{5}P'_5 + \binom{5}{1}\binom{6}{6}P'_6\\
=\;& [5{,}529{,}600 - 9{,}878{,}400 + 8{,}164{,}800 - 3{,}628{,}800 + 756{,}000]/10^6\\
=\;& 943{,}200/10^6.
\end{aligned}
$$

(c) Points when throwing several dice

Ten fair dice are thrown at random. What is the chance P that they add up to 27?



There are $6^{10} = 60{,}466{,}176$ equally likely outcomes. We need to count how many of these add to 27. We will solve this with inclusion-exclusion. Let's start with the set $\mathcal{S} = \{(y_1, y_2, \ldots, y_{10})\}$ where $y_k \geq 1$ and $y_1 + y_2 + \cdots + y_{10} = 27$.

By "stars and bars", the number of such strings is $\binom{27-1}{10-1} = 3{,}124{,}550$.

For $1 \leq i \leq 10$, let A_i be the subset of $\mathcal{S}$ such that $y_i \geq 7$. Subtracting 6 from the ith coordinate puts A_i in one-to-one correspondence with the set $\{(w_1, w_2, \ldots, w_{10})\}$ where $w_k \geq 1$ and $w_1 + w_2 + \cdots + w_{10} = 21$. By the argument above, there are $\binom{21-1}{10-1} = 167{,}960$ such strings.

For $i \neq j$, the set A_iA_j corresponds to strings $\{(v_1, v_2, \ldots, v_{10})\}$ where $v_k \geq 1$ and $v_1 + v_2 + \cdots + v_{10} = 15$. By the argument above, there are $\binom{15-1}{10-1} = 2{,}002$ such strings.

Triple intersections (and higher) are all empty, so the number of elements in $\mathcal{S}$ with all values less than or equal to 6 is

$$\begin{aligned} & |\mathcal{S}| - \sum_i |A_i| + \sum_{i<j} |A_i A_j| \\ = \; & \binom{26}{9} - \binom{10}{1}\binom{20}{9} + \binom{10}{2}\binom{14}{9} \\ = \; & 1{,}535{,}040. \end{aligned}$$

The required probability is $P = 1{,}535{,}040/6^{10} = 0.02538$.

(d) The checkerboard

A 3×3 checkerboard is randomly filled with zeros and ones, for example:

1	0	1
1	1	1
0	0	1

Find the probability that the resulting pattern contains a 2×2 square made of zeros.

Solution: Label the top left corner of the chequeboard as follows:

1	2	
3	4	

For $i = 1, 2, 3, 4$ define A_i to be the event that square i is the upper left corner of a 2×2 square full of zeros. These are not exchangeable events, in fact,

$$\mathbb{P}(A_{i_1} \cdots A_{i_j}) = \left(\frac{1}{2}\right)^N,$$

where N is the number of squares in the union $A_{i_1} \cup \cdots \cup A_{i_j}$.

The probability of one or more such squares is

$$\begin{aligned}
P_4(1) &= \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) + \mathbb{P}(A_4) \\
&- \mathbb{P}(A_1A_2) - \mathbb{P}(A_1A_3) - \mathbb{P}(A_1A_4) - \mathbb{P}(A_2A_3) - \mathbb{P}(A_2A_4) - \mathbb{P}(A_3A_4) \\
&+ \mathbb{P}(A_1A_2A_3) + \mathbb{P}(A_1A_2A_4) + \mathbb{P}(A_1A_3A_4) + \mathbb{P}(A_2A_3A_4) \\
&- \mathbb{P}(A_1A_2A_3A_4) \\
&= (1/2)^4 + (1/2)^4 + (1/2)^4 + (1/2)^4 \\
&- (1/2)^6 - (1/2)^6 - (1/2)^7 - (1/2)^7 - (1/2)^6 - (1/2)^6 \\
&+ (1/2)^8 + (1/2)^8 + (1/2)^8 + (1/2)^8 \\
&- (1/2)^9 \\
&= \frac{95}{512} = .18555.
\end{aligned}$$

(e) Coupling

A set of 200 people, consisting of 100 men and 100 women, is randomly divided into 100 pairs of 2 each. Find the probability that exactly 50 of these pairs will consist of a man and a woman.

Solution 1: For $1 \le i \le 100$, define A_i to be the event that woman i is paired with a man. These events are exchangeable, and

$$\mathbb{P}(A_1A_2\cdots A_i) = \frac{100}{199} \times \frac{99}{197} \times \cdots \times \frac{101-i}{201-2i}.$$

Therefore we get

$$\mathbb{P}(X = 50) = \sum_{i=50}^{100} (-1)^{i-50} \binom{i}{50} \binom{100}{i} \mathbb{P}(A_1A_2\cdots A_i),$$

which works out to

$$\begin{aligned}
\mathbb{P}(X = 50) &= \frac{26824906358673605656402521719558590657053327\,36}{16915402546484587405948993803982403353855810915} \\
&= .1585827253
\end{aligned}$$

Solution 2: Let's begin by counting the total number of possible pairings. Write out the 200 people in any order and underline them two by two: for example

$$\underline{78\,45}\ \underline{12\,105}\ \underline{137\,88}\ \cdots\ \underline{67\,7}.$$

Now, a lot of orders give the same pairings. I could write the pairs in any order, and within each pair I can swap the two partners without changing the pairing. For example,

$$\underline{105\,12}\ \underline{78\,45}\ \underline{137\,88}\ \cdots\ \underline{67\,7},$$

gives the same pairing as the other order. The total number of pairings is therefore $\frac{200!}{100!\,2^{100}}$. This value can be rewritten as

$$\frac{200!}{100!\,2^{100}} = 199 \cdot 197 \cdot 195 \cdots 3 \cdot 1 = 199!!,$$

where $n!!$ is called the *double factorial*.

To count how many pairings have exactly 50 man-woman couples, we first choose 50 men from 100, then 50 women from 100, then count how many man-woman couples we can get. The answer is $\binom{100}{50}\binom{100}{50}50!$.

Finally we must make pairings among the 50 remaining men, and then among the 50 remaining women. The number of ways this can be done is $49!! \cdot 49!!$, and so the final probability is

$$\mathbb{P}(X = 50) = \frac{\binom{100}{50}^2 50!\,(49!!)^2}{199!!} = .1585827253.$$

Solution 3: Write out the 200 men and women in any order and underline them two by two: for example

$$\underline{m\,m}\ \underline{m\,w}\ \underline{w\,m}\ \cdots\ \underline{m\,w}.$$

There are $\binom{200}{100}$ such orderings.

Let's call the number of the four different kinds of pairs $n(mm)$, $n(mw)$, $n(wm)$, and $n(ww)$. Of course $n(mm) + n(mw) + n(wm) + n(ww) = 100$, and you should check that $n(mm) = n(ww)$. The number of orderings with $n(wm) + n(mw) = 50$ is

$$\begin{aligned}
\mathbb{P}(X = 50) &= \frac{1}{\binom{200}{100}} \sum_{n(mw)=0}^{50} \frac{100!}{25!\,25!\,n(mw)!\,(50 - n(mw))!} \\
&= \frac{\binom{100}{50}\binom{50}{25}\sum_{n(mw)}\binom{50}{n(mw)}}{\binom{200}{100}}
\end{aligned}$$

$$= \frac{\binom{100}{50}\binom{50}{25}2^{50}}{\binom{200}{100}}$$
$$= .1585827253.$$

(f) Inclusion-exclusion approximation

The inclusion-exclusion formula for $P_n(1)$ says

$$P := \mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_n) = S_1 - S_2 + S_3 - \cdots + (-1)^{n-1} S_n.$$

It is worth knowing that the partial sums S_1, $S_1 - S_2$, $S_1 - S_2 + S_3$, etc. alternate between upper and lower bounds for P.

Example 5-1 If you roll a fair die ten times, what is the chance that there are at least three consecutive rolls with the same value? The sequence on page 20 is an example.

Solution: For $1 \leq i \leq 8$, define A_i to the event that rolls $i, i+1$, and $i+2$ have the same value.

Since $\mathbb{P}(A_i) = 1/6^2$ for all i, we have $S_1 = 8/36 = 2/9$.

For $\mathbb{P}(A_i A_j)$ it gets more complicated, since the probability depends on whether the overlap between $\{i, i+1, i+2\}$ and $\{j, j+1, j+2\}$ is zero, one, or two elements. For example, $A_1 \cap A_2$ means that rolls $1, 2$ and 3 are the same and also that rolls $2, 3$ and 4 are the same. This amounts to saying that rolls $1, 2, 3$, and 4 are all the same; the chance that this happens is $1/6^3$. On the other hand, since there is no overlap events A_1 and A_4 are independent, so $\mathbb{P}(A_1 \cap A_4) = (1/6^2)(1/6^2)$.

Accounting for all the different cases we calculate $S_2 = 7/144$ so that

$$.17361 = 2/9 - 7/144 \leq P \leq 2/9 = .22222.$$

With more patience, you can calculate further terms and get better approximations.

6 Zero-one random variables

(a) Moments and factorial moments

If $A_1, \ldots, A_n$ are events we introduce zero-one random variables $U_1, \ldots, U_n$ such that $U_i = 1$ if A_i occurs and $U_i = 0$ if A_i does not. The sum $X = U_1 + \cdots + U_n$ is the total number of events that occur.

We have

$$\mathbb{E}(U_i) = 1 \cdot \mathbb{P}(A_i) + 0 \cdot [1 - \mathbb{P}(A_i)] = \mathbb{P}(A_i),$$

so that

$$\mathbb{E}(X) = \sum_i \mathbb{P}(A_i) = S_1.$$

In a similar way, $\binom{X}{2} = \sum_{i<j} U_i U_j$ is the number of pairs of events that occur, so that $\mathbb{E}\left[\binom{X}{2}\right] = \sum_{i<j} \mathbb{P}(A_i A_j) = S_2$. In general, for $j \geq 0$,

$$\mathbb{E}\left[\binom{X}{j}\right] = S_j.$$

(b) Singles in a million

Randomly choose a 6-digit number $000000, 000001, \ldots, 999999$. What is the average number of digits that appear exactly once?

Solution: Let A_i be the event that the value i appears exactly once.

$$\mathbb{E}(X) = \sum_{i=0}^{9} \mathbb{P}(A_i) = 10\,\mathbb{P}(A_0) = 10\left[6\,\frac{1}{10}\left(\frac{9}{10}\right)^5\right] = 3.54294.$$

(c) Random ONEcards

In a class of n students, everyone's ONEcard is put in a pile and randomly given back to the students. What is the expected number of students that get their own cards back?

Solution: Let A_i be the event that student i gets his or her own card back. Then,

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{P}(A_i) = \sum_{i=1}^{n} \frac{1}{n} = 1.$$

(d) Where's the ace?

On average, where is the first ace in a well-shuffled deck?

Solution: For the 48 non-aces, define A_i, $1 \leq i \leq 48$ to be the event that card i appears before all the aces. The position of the first ace is $X = 1 + \sum_{i=1}^{48} U_i$, and since $\mathbb{P}(A_i) = 1/5$ we get

$$\text{average position of the first ace} = \mathbb{E}(X) = 1 + \sum_i \mathbb{P}(A_i) = 53/5 = 10.6.$$

(e) Numbers and colours

An urn contains n numbered balls and b blue balls. Balls are drawn and replaced until a blue ball is obtained, at which point the experiment stops. A repetition is noted each time a numbered ball is drawn that had been previously drawn. What is the expected number of repetitions?

Solution: Let N be the number of numbered balls drawn. When you add the terminating blue ball, the total number $N + 1$ of draws is a geometric random variable with $p = b/(b+n)$. Therefore, on average, we draw $\mathbb{E}(N+1) = 1/p = 1+n/b$ balls. Subtracting the blue ball, we get $\mathbb{E}(N) = n/b$.

For $1 \leq i \leq n$, let U_i be 1 if ball i appears before a blue ball and 0 otherwise. Then $\mathbb{E}(U_i) = 1/(b+1)$ and the average number of repetitions is

$$\mathbb{E}(N - \sum_{i=1}^{n} U_i) = n\left(\frac{1}{b} - \frac{1}{b+1}\right) = \frac{n}{b(b+1)}.$$

(f) Alphabet soup

Given a sequence $(a_1, a_2, \ldots, a_n)$ over the alphabet $\{1, 2, \ldots, m\}$ chosen uniformly at random among the m^n possibilities. What is the expected size of the *set* $\{a_1, a_2, \ldots, a_n\}$?

Solution: Let A_i be the event that the letter i is present in the random sequence. The size of the set is just $X = \sum_{i=1}^{m} U_i$ which has expected value

$$\mathbb{E}(X) = \sum_{i=1}^{m} \mathbb{P}(A_i) = m\left[1 - \left(1 - \frac{1}{m}\right)^n\right].$$

(g) Almost disjoint

Prove that if $\mathbb{P}(\cup_i A_i) = \sum_i \mathbb{P}(A_i)$, then $\mathbb{P}(A_i A_j) = 0$ for $i \neq j$.

Solution: Let $X_n = U_1 + \cdots + U_n$ as in section 6(a). Let X_0 be one on the set $\cup_i A_i$ and zero elsewhere. Then $X_n - X_0 \geq 0$, but the equation $\mathbb{P}(\cup_i A_i) = \sum_i \mathbb{P}(A_i)$ says that $\mathbb{E}(X_n - X_0) = 0$. Therefore $\mathbb{P}(X_n - X_0 = 0) = 1$, which implies $\mathbb{P}(X_n - X_0 \geq 1) = 0$.

Now for $i \neq j$, we have $A_i A_j \subseteq (X_n - X_0 \geq 1)$, so that $\mathbb{P}(A_i A_j) = 0$.

(h) Proof of the inclusion-exclusion formula

Let's prove the wonderful formula from page 19. We start by noting that for non-negative integers $x \geq k$ and any real number y we have

$$\sum_i \binom{i}{k}\binom{x}{i} y^{i-k} = \binom{x}{k} \sum_i \binom{x-k}{i-k} y^{i-k} = \binom{x}{k}(1+y)^{x-k}.$$

Substituting $y = -1$ gives us

$$\sum_i (-1)^{i-k} \binom{i}{k}\binom{x}{i} = \binom{x}{k} 0^{x-k}.$$

Since the sum on the left hand side is also zero when $x < k$, we get for any non-negative integers x, k:

$$\sum_i (-1)^{i-k} \binom{i}{k}\binom{x}{i} = 1_{[x=k]}.$$

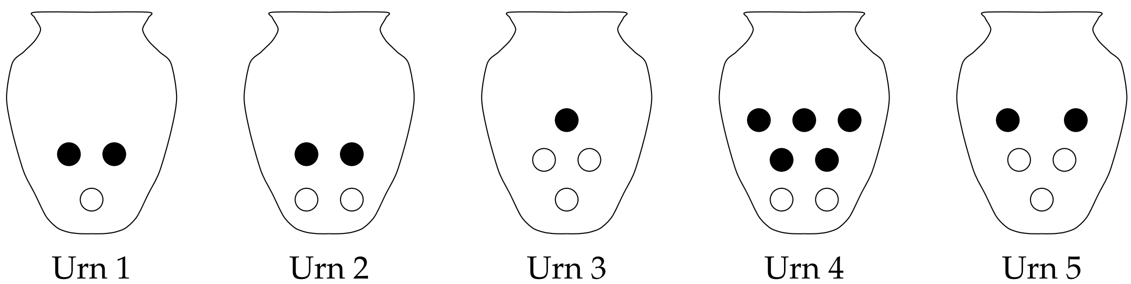
Substituting the random variable X for x and taking expectations gives

$$\boxed{\sum_i (-1)^{i-k} \binom{i}{k} \mathbb{E}\left[\binom{X}{i}\right] = \mathbb{P}(X = k).}$$

7 Generating functions

(a) Five urn problem

Suppose we have 5 urns with black and white balls as in the diagram.



We randomly select one ball from each urn. Let W be the total number of white balls drawn. What is $\mathbb{P}(W = 2)$?

We can just add the probabilities of the ten disjoint events

$$w_1 w_2 b_3 b_4 b_5, w_1 b_2 w_3 b_4 b_5, \ldots, b_1 b_2 b_3 w_4 w_5.$$

Each of these comes with a different probability, but this can be automatically taken into account by expanding the product

$$\left(\frac{1}{3}w_1 + \frac{2}{3}b_1\right)\left(\frac{1}{2}w_2 + \frac{1}{2}b_2\right)\left(\frac{3}{4}w_3 + \frac{1}{4}b_3\right)\left(\frac{2}{7}w_4 + \frac{5}{7}b_4\right)\left(\frac{3}{5}w_5 + \frac{2}{5}b_5\right),$$

and adding the coefficients of the ten events $w_1 w_2 b_3 b_4 b_5, \cdots$ etc.

In fact, since we are only interested in white balls, and only the number of them (and not what urns they came from) we may as well consider the polynomial

$$\begin{aligned} G(w) &= \left(\frac{1}{3}w + \frac{2}{3}\right)\left(\frac{1}{2}w + \frac{1}{2}\right)\left(\frac{3}{4}w + \frac{1}{4}\right)\left(\frac{2}{7}w + \frac{5}{7}\right)\left(\frac{3}{5}w + \frac{2}{5}\right) \\ &= \frac{3}{140}w^5 + \frac{39}{280}w^4 + \frac{137}{420}w^3 + \frac{283}{840}w^2 + \frac{16}{105}w + \frac{1}{42}. \end{aligned}$$

This function is called the *probability generating function* of W, and we can read off the answer $\mathbb{P}(W = 2) = 283/840 = .33690$.

Definition 7-1 For a positive integer valued random variable X the *probability generating function* of X is

$$G_X(s) = \mathbb{E}(s^X) = \sum_{k=0}^{\infty} s^k \, \mathbb{P}(X = k).$$

(b) Adding random numbers

Randomly choose one number from each of the sets $\{1,2\}$, $\{1,2,3\}$, and $\{1,2,3,4\}$. What is the chance that they add up to 7?

Solution: The probability generating function for the sum is

$$\begin{aligned} G_X(s) &= (s/2 + s^2/2)(s/3 + s^2/3 + s^3/3)(s/4 + s^2/4 + s^3/4 + s^4/4) \\ &= s^3/24 + s^4/8 + 5s^5/24 + s^6/4 + 5s^7/24 + s^8/8 + s^9/24. \end{aligned}$$

Taking the coefficient of s^7 we get $\mathbb{P}(X = 7) = 5/24$.

(c) Ten dice again

We revisit an old friend. Ten fair dice are thrown at random. What is the chance that they add up to 27?

Solution 1: The probability generating function for the sum of ten fair dice is

$$G_X(s) = [(s + s^2 + s^3 + s^4 + s^5 + s^6)/6]^{10}.$$

Expanding this with a computer and extracting the coefficient of s^{27} we get $\mathbb{P}(X = 27) = 2{,}665/104{,}976$.

Solution 2: We use the fact that $s + s^2 + s^3 + s^4 + s^5 + s^6 = s(1 - s^6)/(1 - s)$. Dividing by 6 and taking the tenth power gives

$$\begin{aligned} G_X(s) &= \frac{1}{6^{10}} s^{10} (1 - s^6)^{10} (1 - s)^{-10} \\ &= \frac{1}{6^{10}} s^{10} \sum_j \binom{10}{j} (-s^6)^j \sum_k \binom{-10}{k} (-s)^k. \end{aligned}$$

Thus the coefficient of s^{27} is

$$\begin{aligned} P &= \frac{1}{6^{10}}\left[-\binom{10}{0}\binom{-10}{17}+\binom{10}{1}\binom{-10}{11}-\binom{10}{2}\binom{-10}{5}\right] \\ &= \frac{1}{6^{10}}\left[\binom{26}{9}-10\binom{20}{9}+45\binom{14}{9}\right] \\ &= \frac{2{,}665}{104{,}976}. \end{aligned}$$

The general formula to get a sum of k when you throw n m-sided dice is

$$\boxed{\mathbb{P}(X=k)=\sum_{i=0}^{\lfloor (k-n)/m\rfloor}(-1)^i\binom{n}{i}\binom{k-im-1}{n-1}\frac{1}{m^n}}$$

(d) The pattern HH again

Find $p(k)=\mathbb{P}(N_2=k)$ where N_2 is the number of tosses of a fair coin needed to see the pattern HH.

Solution: We note that $p(0)=p(1)=0$, $p(2)=1/4$, and that for $k>2$ we have

$$p(k)=\frac{1}{2}p(k-1)+\frac{1}{4}p(k-2).$$

Thus the probability generating function satisfies

$$\begin{aligned} G(s) &= \sum_{k\geq 2}p(k)s^k \\ &= \frac{1}{4}s^2+\sum_{k>2}\frac{1}{2}p(k-1)s^k+\sum_{k>2}\frac{1}{4}p(k-2)s^k \\ &= \frac{1}{4}s^2+\frac{1}{2}sG(s)+\frac{1}{4}s^2G(s). \end{aligned}$$

Solving for $G(s)$ we get

$$G(s)=\frac{s^2}{4-2s-s^2}.$$

How can we extract coefficients from this expression? First we factor the quadratic into $4 - 2s - s^2 = -(r_1 - s)(r_2 - s)$ where $r_2 = -1 + \sqrt{5}$ and $r_1 = -1 - \sqrt{5}$. Secondly, we use the geometric series to write $1/(r - s) = (1/r)\sum_{k\geq 0}(s/r)^k$. Then we write

$$\frac{1}{-(r_1 - s)(r_2 - s)} = \frac{1}{r_2 - r_1}\left[\frac{1}{r_2 - s} - \frac{1}{r_1 - s}\right].$$

Therefore,

$$G(s) = \frac{1}{2\sqrt{5}}\sum_{k\geq 2}\left[\left(\frac{1}{r_2}\right)^{k-1} - \left(\frac{1}{r_1}\right)^{k-1}\right]s^k.$$

Extracting the probabilities, for $k \geq 2$ we get

$$p(k) = \frac{1}{2\sqrt{5}}\left[\left(\frac{1}{-1+\sqrt{5}}\right)^{k-1} - \left(\frac{1}{-1-\sqrt{5}}\right)^{k-1}\right].$$

(e) Crazy Dice

The sum of 2 normal dice has a triangular distribution over $\{2, 3, \ldots, 11, 12\}$ as seen using generating functions:

$$(s + s^2 + s^3 + s^4 + s^5 + s^6)^2 = s^2 + 2s^3 + 3s^4 + 4s^5 + 5s^6 + 6s^7 + 5s^8 + 4s^9 + 3s^{10} + 2s^{11} + s^{12}.$$

Question: can we put different positive numbers on a pair of dice so that the distribution of the sum stays the same? We will call these *crazy dice*.

Mathematically, we want positive integers a_i, b_i for $1 \leq i \leq 6$ so that

$$G_a(s)G_b(s) = (s + s^2 + s^3 + s^4 + s^5 + s^6)^2,$$

where $G_a(s) = \sum_{i=1}^{6} s^{a_j}$ and $G_b(s) = \sum_{i=1}^{6} s^{b_j}$.

We begin by factoring the right hand side:

$$s + s^2 + s^3 + s^4 + s^5 + s^6 = s(1 + s + s^2)(1 + s)(1 - s + s^2),$$

so that

$$(s + s^2 + s^3 + s^4 + s^5 + s^6)^2 = s^2(1 + s + s^2)^2(1 + s)^2(1 - s + s^2)^2.$$

This shows that G_a and G_b must consist of these factors. Since the a's and b's are strictly positive, each has one factor s. Both polynomials evaluate to 6 at $s = 1$, so that the factor

$(1+s+s^2)(1+s)$ also appears in both. If we give G_a and G_b one factor of $1-s+s^2$ we get back ordinary dice.

Alternatively, we could try to put both factors of $1-s+s^2$ into G_a. That gives

$$G_a(s) = s(1+s+s^2)(1+s)(1-s+s^2)^2 = s + s^3 + s^4 + s^5 + s^6 + s^8,$$

and

$$G_b(s) = s(1+s+s^2)(1+s) = s + 2s^2 + 2s^3 + s^4.$$

Translating back, the crazy dice are 1,3,4,5,6,8 and 1,2,2,3,3,4.

(f) Moments

Suppose X is a random variable taking non-negative integer values, and let $G_X(s) = \mathbb{E}(s^X)$. Differentiating once and setting $s = 1$ we get $G_X'(1) = \mathbb{E}(X)$.

For instance, the probability generating function for the pattern HH problem is $G(s) = s^2/(4-2s-s^2)$. Differentiating, we get $G'(s) = 2s(4-s)/(4-2s-s^2)^2$ and setting $s=1$ we get $\mathbb{E}(N_2) = 2\cdot 1\cdot(4-1)/1^2 = 6$.

In fact, for any non-negative integer i if we differentiate i times and set $s = 1$ we get $G_X^{(i)}(1) = \mathbb{E}[X(X-1)\cdots(X-i+1)]$. Consequently, we can expand G_X as a power series around the point $s=1$ to obtain

$$G_X(s) = \sum_i \mathbb{E}\left[\binom{X}{i}\right](s-1)^i.$$

Using the binomial theorem to expand the powers of $s-1$ we get

$$\begin{aligned} G_X(s) &= \sum_{i\geq 0} \mathbb{E}\left[\binom{X}{i}\right](s-1)^i \\ &= \sum_{i\geq 0} \mathbb{E}\left[\binom{X}{i}\right]\sum_{k\geq 0}\binom{i}{k}(-1)^{i-k}s^k \\ &= \sum_{k\geq 0} s^k \sum_{i\geq 0}(-1)^{i-k}\binom{i}{k}\mathbb{E}\left[\binom{X}{i}\right]. \end{aligned}$$

Extracting coefficients we get

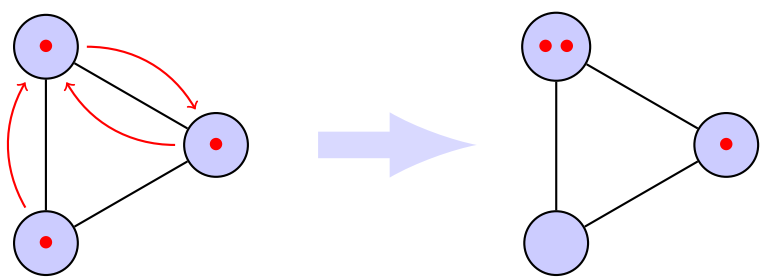
$$\mathbb{P}(X=k) = \sum_{i\geq 0}(-1)^{i-k}\binom{i}{k}\mathbb{E}\left[\binom{X}{i}\right],$$

another proof of the inclusion-exclusion formula.

(g) Triangle tango

Three independent random walkers start at different corners of a triangle. At every time point, each walker jumps either to the left or to the right at random. Eventually, all three walkers will meet at one corner and the game is finished. How long on average until this happens?

Solution: Define "state A" to be when all walkers are at different corners, "state B" when two walkers are at one corner, and the third walker is at a different corner, and "state C" when they are all together. The diagram below illustrates how we might move from state A to state B.



We are interested in M, the number of steps needed to finish the game, starting in state A. To solve this problem, we also need the random variable N, the number of steps needed to finish starting in state B.

To begin, imagine starting in state A. There are eight equally likely possibilities depending on how the walkers jump. In two of the cases we remain in state A, and in six cases we end up in state B. Hence

$$M = \begin{cases} 1 + M' & \text{with probability } 1/4 \\ 1 + N & \text{with probability } 3/4, \end{cases}$$

where M' has the same distribution as M.

Now imagine starting in state B. There are eight equally likely possibilities depending on how the walkers jump. In one case all the walkers go to the same corner and the game is finished. Otherwise, in five cases we stay in state B, and in two cases we go to state A.

Therefore

$$N = \begin{cases} 1 & \text{with probability } 1/8 \\ 1 + N' & \text{with probability } 5/8 \\ 1 + M & \text{with probability } 1/4, \end{cases}$$

where N' has the same distribution as N.

From these relationships, we find that the probability generating functions satisfy

$$\begin{aligned} G_M(s) &= \frac{s}{4}\, G_M(s) + \frac{3s}{4}\, G_N(s), \\ G_N(s) &= \frac{s}{8} + \frac{5s}{8}\, G_N(s) + \frac{s}{4}\, G_M(s). \end{aligned}$$

Solving this system of equations shows that

$$G_M(s) = \frac{3s^2}{32 - 28s - s^2} = \frac{3}{32}\, s^2 + \frac{21}{256}\, s^3 + \frac{153}{2048}\, s^4 + \frac{1113}{16384}\, s^5 + \cdots .$$

Differentiating G_M and setting $s = 1$ gives $\mathbb{E}(M) = G_M'(1) = 12$.

(h) Other generating functions

Definition 7-2 If (a_n) is a sequence of numbers, we call $H(s) = \sum_{n=0}^{\infty} a_n s^n$ the *generating function* of the sequence.

(h1) Even Odds

Let a_n be the probability that n Bernoulli trials result in an *even number of successes*. This occurs if an initial failure is followed by an even number of successes, or an initial success is followed by an odd number of successes. Therefore $a_0 = 1$ and for $n \geq 1$

$$a_n = q a_{n-1} + p(1 - a_{n-1}).$$

Multiplying by s^n and adding over n we see that the generating function satisfies

$$H(s) = 1 + qsH(s) + ps(1 - s)^{-1} - psH(s)$$

or

$$2H(s) = [1 - s]^{-1} + [1 - (q - p)s]^{-1}.$$

Expanding the right hand side using geometric series we find that the coefficients satisfy

$$a_n = \frac{1}{2} + \frac{(q-p)^n}{2}.$$

(h2) Non-consecutive values

Any set of k non-consecutive positive integers can be obtained by reading off the "b"s in a sequence of the form:

$$\mathbf{seq}(a)\left(b\,\mathbf{seq}_{\geq 1}(a)\right)^{k-1} b\,\mathbf{seq}(a).$$

The generating function for such sequences is

$$H(s) = \frac{1}{1-s}\left(s\,\frac{s}{1-s}\right)^{k-1} s\,\frac{1}{1-s} = \frac{s^{2k-1}}{(1-s)^{k+1}}.$$

In other words,

$$H(s) = s^{2k-1}\sum_{m=0}^{\infty}\binom{m+k}{k}s^m,$$

so the nth coefficient of H is $\binom{n-k+1}{k}$.

Similarly, we can find sets of k positive integers with a mimimum gap of size d using these sequences:

$$\mathbf{seq}(a)\left(b\,\mathbf{seq}_{\geq d}(a)\right)^{k-1} b\,\mathbf{seq}(a).$$

The generating function for such sequences is

$$H(s) = \frac{1}{1-s}\left(s\,\frac{s^d}{1-s}\right)^{k-1} s\,\frac{1}{1-s} = \frac{s^{d(k-1)+k}}{(1-s)^{k+1}}.$$

In other words,

$$H(s) = s^{d(k-1)+k}\sum_{m=0}^{\infty}\binom{m+k}{k}s^m,$$

so the nth coefficient of H is $\binom{n-d(k-1)}{k}$.

8 Random walks

(a) Definitions and notation

Let $X_0, X_1, \dots$ be a random walk taking values in some state space $\mathcal{S}$. For $x \in \mathcal{S}$, the notation $\mathbb{P}_x$ and $\mathbb{E}_x$ will refer to conditional probability and expectation given that we start at x, i.e., that $X_0 = x$. The transition probability from state x to y is defined as $p_{xy} := \mathbb{P}_x(X_1 = y)$.

If $f : \mathcal{S} \to \mathbb{R}$ is a real valued function on $\mathcal{S}$, we let $Pf : \mathcal{S} \to \mathbb{R}$ be the function that gives the average f value after taking one step, that is,

$$(Pf)(x) = \sum_{y \in \mathcal{S}} p_{xy}\, f(y).$$

If $\mathcal{B}$ is a subset of $\mathcal{S}$, we define the hitting time of $\mathcal{B}$ to be

$$T_{\mathcal{B}} = \inf\{n \geq 0 : X_n \in \mathcal{B}\}.$$

For future reference, let us also define the return time of $\mathcal{B}$ to be

$$R_{\mathcal{B}} = \inf\{n \geq 1 : X_n \in \mathcal{B}\}.$$

Notice the subtle, but important, difference.

(b) Hitting times

A lot of interesting questions can be answered using the following functions. Let $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{S}$ and define

$$f(x) = \mathbb{P}_x(T_{\mathcal{A}} = T_{\mathcal{B}}), \quad \text{and} \quad h(x) = \mathbb{E}_x(T_{\mathcal{B}}).$$

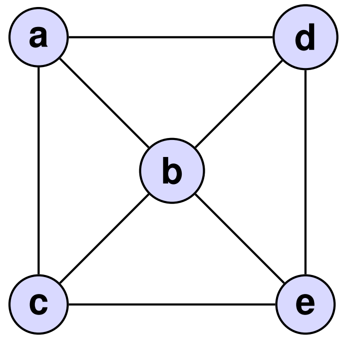
These two functions can often be calculated explicitly as they solve the following boundary value problems:

$$f(x) = \begin{cases} 0 & \text{if } x \in \mathcal{B} \setminus \mathcal{A} \\ 1 & \text{if } x \in \mathcal{A} \\ (Pf)(x) & \text{if } x \notin \mathcal{B}. \end{cases}$$

$$h(x) = \begin{cases} 0 & \text{if } x \in \mathcal{B} \\ 1 + (Ph)(x) & \text{if } x \notin \mathcal{B}. \end{cases}$$

These formulas are a consequence of first step analysis.

Example 8-1



Starting at state c, a random walk moves on this graph until it hits the boundary $\{d, e\}$. What is the chance that it ends up at state d?

Solution: We let $\mathcal{B} = \{d, e\}$, and $\mathcal{A} = \{d\}$. By the formula above the function f satisfies

$$\begin{aligned} f(a) &= \frac{1}{3}(1 + f(b) + f(c)) \\ f(b) &= \frac{1}{4}(f(a) + f(c) + 0 + 1) \\ f(c) &= \frac{1}{3}(0 + f(b) + f(a)). \end{aligned}$$

It's not too hard to solve this system using linear algebra, but we may as well exploit the symmetry in the problem by noting that $f(b) = 1/2$ and $f(a) + f(c) = 1$. Therefore the third equation can be rewritten as

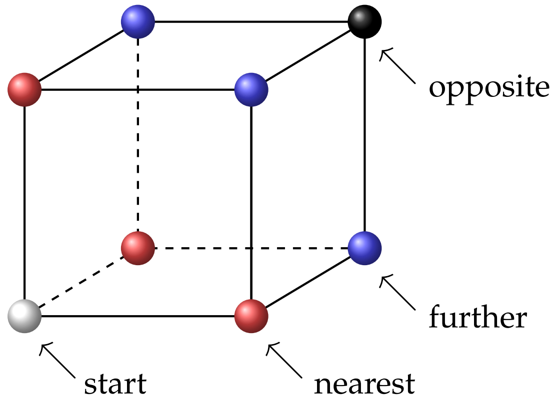
$$f(c) = \frac{1}{3}(0 + 1/2 + 1 - f(c)),$$

which gives $f(c) = 3/8$. ■

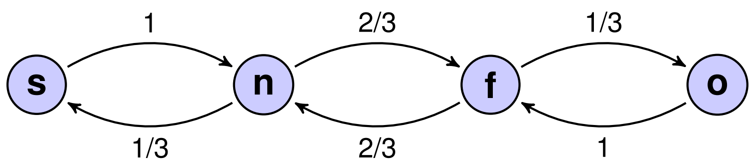
(c) Random walk on a cube

On average, how long does it take a random walk on the vertices of a cube to reach the opposite corner?

Solution:



Here we have a random walk with eight states, but by symmetry we can reduce this to four states: start, nearest neighbors, further neighbors, and opposite. In this case, the boundary is the single state $\{o\}$.

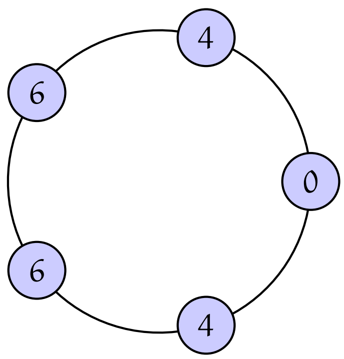


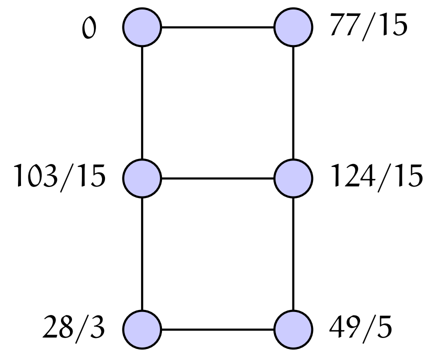
The formula for h says

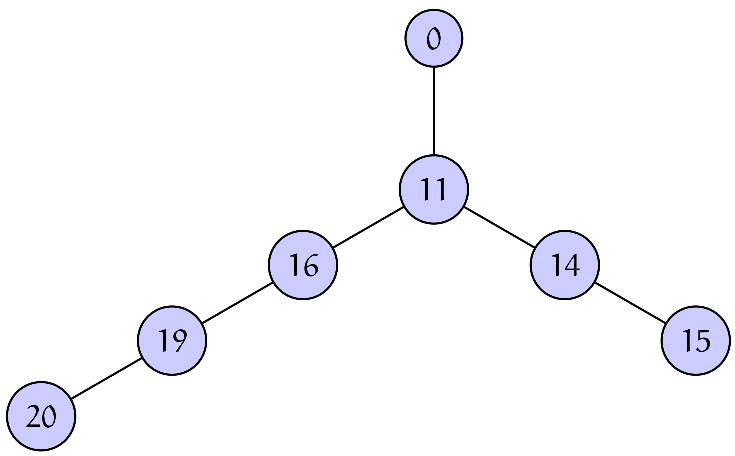
$$\begin{aligned} h(s) &= 1 + h(n) \\ h(n) &= 1 + \frac{1}{3}h(s) + \frac{2}{3}h(f) \\ h(f) &= 1 + \frac{1}{3}0 + \frac{2}{3}h(n). \end{aligned}$$

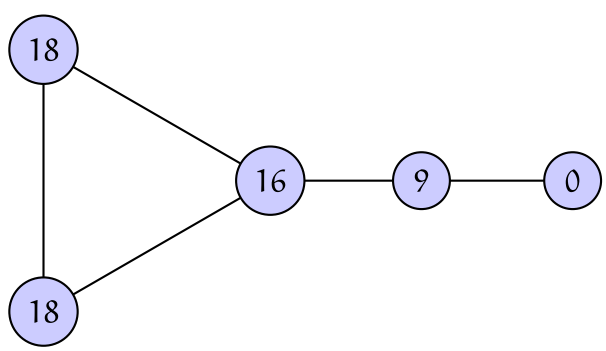
From here it is not too hard to work out that $h(s) = 10$. By the way, we also get $h(n) = 9$ and $h(f) = 7$. ■

(d) More random walk h functions





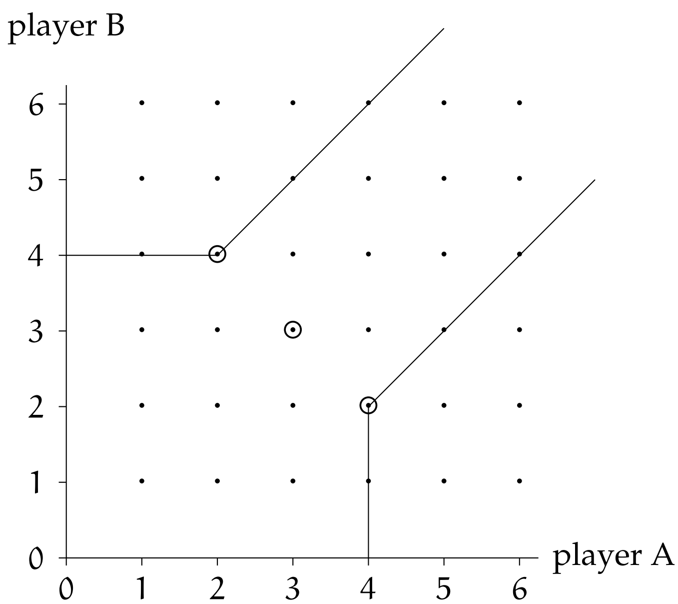




(e) Tennis

James Bernoulli 1713 Players A and B win a point in tennis with probabilities p and $q = 1 - p$, respectively. Find the winning probabilities of each player.

Solution: Here is a picture of the states of the process. We begin at $(0, 0)$ and move right with probability p or up with probability q. Player A wins if we first hit the lower boundary while player B wins if we first hit the upper boundary.



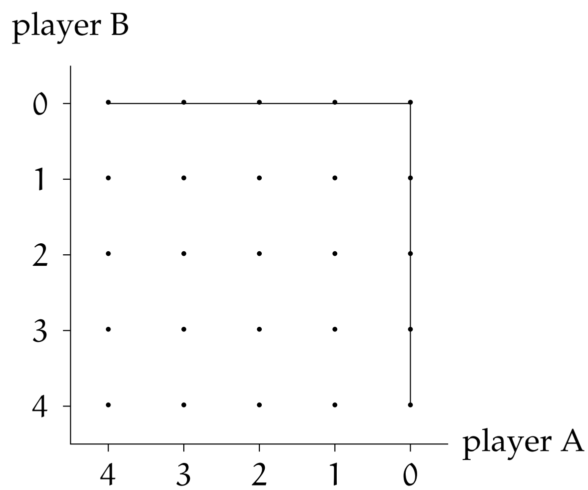
Let $f(i, j)$ be the probability that player A wins the game, starting at state (i, j). Start at the three circled states, where f is known and work your way back to state $(0, 0)$. You will find that

$$f(0,0) = \frac{p^4(3-2p)(4p^2-8p+5)}{2p^2-2p+1}.$$

(f) The problem of points

John Bernoulli 1710 Two persons participate in a series of games. In each game, A wins with probability p and B wins with probability $q = 1 - p$. The series ends when one of the players has won r games; he wins a pile of money. For some reason, the series is interrupted. At this moment, A has won $r - a$ games and B has won $r - b$ games. What is a fair way to divide the pile?

Solution 1: This the previous problem with a different boundary. Let's illustrate the case of a best-of-seven series, that is, $r = 4$. We will label the axes by the number of games needed to win the series.



If the series is tied 2-2, then the chance that player A will win the series is $3p^2 - 2p^3$.

Solution 2: Imagine playing all of the $a + b - 1$ remaining games. Player A wins the series if he wins at least a of these games, so

$$f(a, b) = \sum_{i=a}^{a+b-1} \binom{a+b-1}{i} p^i q^{(a+b-1)-i}.$$

Just to double check, let's plug in $a = b = 2$ and see what we get:

$$\sum_{i=2}^{3} \binom{3}{i} p^i q^{3-i} = 3p^2 q + p^3 = 3p^2 - 2p^3.$$

Solution 3: Imagine playing all of the $a + b - 1$ remaining games. Player A wins the series if his ath win occurs before player B wins b games, so

$$f(a, b) = \sum_{i=0}^{b-1} \binom{a+i-1}{i} p^a q^i.$$

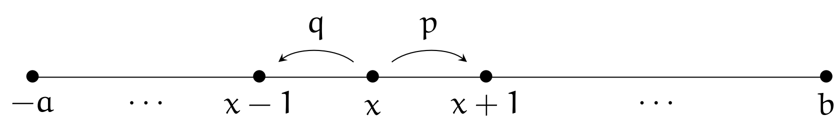
Just to double check, let's plug in $a = b = 2$ and see what we get:

$$\sum_{i=0}^{1} \binom{1+i}{i} p^2 q^i = p^2(1 + 2q) = 3p^2 - 2p^3.$$

(g) Gambler's ruin

James Bernoulli 1713 Players A and B begin with a and b counters to start. They play several rounds of a game where player A has probability p of winning. The winner of the round gets a counter from the other player, and the first player to get all the counters wins. Find the winning probabilities of each player.

Solution: Consider the random walk that tracks player A's gains. It moves forward with probability p and backward with probability $q = 1 - p$. The boundary consists of two points $\mathcal{B} = \{-a, b\}$. Set $\mathcal{A} = \{-a\}$ so that the probability that B wins the game is $f(0)$.



The function f satisfies $f(-a) = 1$, $f(b) = 0$, and in between $f(x) = qf(x-1) + pf(x+1)$. We rewrite this last equation as

$$q\,\Delta f(x-1) = p\,\Delta f(x).$$

It's not hard to show that

$$\Delta f(x) = \left(\frac{q}{p}\right)^{x+a} \Delta f(-a),$$

and then to use the boundary conditions to get

$$f(x) = \begin{cases} \dfrac{\left(\frac{q}{p}\right)^{b} - \left(\frac{q}{p}\right)^{x}}{\left(\frac{q}{p}\right)^{b} - \left(\frac{q}{p}\right)^{-a}} & \text{if } p \neq q \\[2ex] \dfrac{b - x}{b + a} & \text{if } p = q. \end{cases}$$

(h) One sided boundaries

Imagine a random walk on the integers that jumps to the left with probability q, or to the right with probability p, where $p + q = 1$.

We want to calculate $\mathbb{P}_0(T_{-1} < \infty)$ the probability that the random walk will ever hit the state to the left of its starting position.

Solution 1: Let $z = \mathbb{P}_0(T_{-1} < \infty)$, then a first step analysis gives

$$z = q + pz^2,$$

which has two roots 1 and q/p.

If $p \leq q$, we must have $z = 1$. Such a random walk is guaranteed to hit every state to the left of its start. If $p = q$, then the random walk will hit every state, in fact, it will hit every state infinitely often!

If $p > q$, it turns out that $z = q/p$. Accepting this for a moment, we combine both cases into one formula $z = 1 \wedge \frac{q}{p}$.

Solution 2: Let's try another approach to the problem. The hitting time T_{-1} is a random variable that takes values in $\{1, \ldots, \infty\}$. Define its probability generating function by $G(s) = \mathbb{E}_0(s^{T_{-1}})$ for $0 < s < 1$. First step analysis gives

$$G(s) = qs + ps\,G(s)^2,$$

which means that

$$G(s) = \frac{1 \pm \sqrt{1 - 4pqs^2}}{2ps}.$$

Since $|G(s)| \leq 1$, we must have

$$G(s) = \frac{1 - \sqrt{1 - 4pqs^2}}{2ps}.$$

Letting $s \to 1$, we get $\mathbb{P}_0(T_{-1} < \infty) = 1 \wedge \frac{q}{p}$.

Solution 3: We will start with the two boundary problem $-a < x < b$, and then let $b \to \infty$. When $p > q$,

$$\mathbb{P}_x(T_{-a} < \infty) = \lim_{b\to\infty} \mathbb{P}_x(T_{-a} < T_b) = \lim_{b\to\infty} \frac{\left(\frac{q}{p}\right)^b - \left(\frac{q}{p}\right)^x}{\left(\frac{q}{p}\right)^b - \left(\frac{q}{p}\right)^{-a}},$$

which equals $\left(\frac{q}{p}\right)^{x+a}$.

(i) Return to the start

The probability that the random walk returns to its starting position is

$$\mathbb{P}_x(\text{hit } x \text{ after time } 0) \quad = \quad q\left(1 \wedge \frac{p}{q}\right) + p\left(1 \wedge \frac{q}{p}\right)$$

$$
\begin{aligned}
&= q \wedge p + p \wedge q \\
&= 1 - |p - q|.
\end{aligned}
$$

9 Classical probability

(a) Coincidence or Rencontre

Montmort 1708 The cards on two identical, well-shuffled decks are turned over one at a time. What is the chance to see the same card simultaneously?

Solution: Let A_i be the event that card i is in the same position in both decks. If $i_1, i_2, \ldots, i_j$ are distinct indices, then

$$\mathbb{P}(A_{i_1} \cdots A_{i_j}) = \frac{(n-j)!}{n!}.$$

By inclusion-exclusion we get

$$\mathbb{P}(\text{no coincidence}) = p_n(0) = \sum_{j=0}^{n} (-1)^j \binom{n}{j} \frac{(n-j)!}{n!} = \sum_{j=0}^{n} \frac{(-1)^j}{j!}.$$

(b) Occupancy

de Moivre 1712 A symmetrical die with r faces is thrown n times. What is the chance that all faces appear at least once?

Solution: For $1 \leq i \leq r$, define A_i to be the event that the ith face does not occur. These are exchangeable events with $P_i = (1 - \frac{i}{r})^n$ and so the probability that there are no missing faces is

$$p_r(0) = \sum_{i=0}^{r} (-1)^i \binom{r}{i} \left(1 - \frac{i}{r}\right)^n.$$

■

The expected number of missing faces is $\mathbb{E}(X) = r\left(1 - \frac{1}{r}\right)^n$, and more generally $\mathbb{E}\left[\binom{X}{j}\right] =$

$\binom{r}{j}\left(1-\frac{j}{r}\right)^n$, and therefore

$$\mathbb{P}(X=k)=\sum_j(-1)^{j-k}\binom{j}{k}\binom{r}{j}\left(1-\frac{j}{r}\right)^n.$$

If we let $Y = r - X$ be the number of faces that occur, then the above formula can be rewritten as

$$\mathbb{P}(Y=k)=\binom{r}{k}\left(\frac{k}{r}\right)^n q_{kn},$$

where $q_{kn}=\sum_i(-1)^i\binom{k}{i}\left(1-\frac{i}{k}\right)^n$. Note that q_{kn} is the chance that all faces occur when you throw n dice with k sides.

Stirling numbers of the second kind Let's define $\left\{ {n \atop k} \right\}$ to be the number of ways to divide n objects into k non-empty groups. For instance, $\left\{ {4 \atop 2} \right\} = 7$ because of the seven ways shown here:

$$\{1,2,3\}\cup\{4\},\ \{1,2,4\}\cup\{3\},\ \{1,4,3\}\cup\{2\},\ \{4,2,3\}\cup\{1\}$$
$$\{1,2\}\cup\{3,4\},\ \{1,4\}\cup\{2,3\},\ \{1,3\}\cup\{2,4\}.$$

The Stirling numbers can be used to give an alternative solution to the occupancy problem. We can get all k faces in the following way: partition the index set $\{1,2,\ldots,n\}$ into k non-empty sets, then assign the values $1,2,\ldots,k$ to these sets.

For instance, let $n = 4$ and $k = 2$ (let's use a coin). We can construct an outcome where all k values occur by choosing $\{1,2,3\}\cup\{4\}$ and then assign H to the first set and T to the second. This results in the outcome HHHT. We had 7 ways to make the first choice and 2 ways to make the second choice. There are 14 outcomes where both values occur, so the probability is

$$\mathbb{P}(\text{both values occur})=\frac{14}{16}.$$

Generalizing this argument, we get

$$\mathbb{P}(\text{all } k \text{ values occur in } n \text{ throws})=q_{kn}=\frac{\left\{ {n \atop k} \right\}k!}{k^n}.$$

Example 9-1 *A Formula* Let Y be the number of distinct faces that occur in n throws of an r-sided die. Then

$$\mathbb{P}(Y=k)=\binom{r}{k}\left(\frac{k}{r}\right)^n\left\{ {n \atop k} \right\}\frac{k!}{k^n}.$$

Since these must add to 1, we deduce that

$$\sum_k \left\{ {n \atop k} \right\} r(r-1)\cdots(r-k+1) = r^n.$$

Example 9-2 *Coupon collecting* Let us sample with replacement from $\{1, 2, \ldots, k\}$, and let N be the number of trials needed to observe every value. By the occupancy problem, we see that

$$\mathbb{P}(N \le n) = k^{-n}\, k! \left\{ {n \atop k} \right\},$$

and

$$\mathbb{P}(N = n) = k^{-n}\, k! \left\{ {n-1 \atop k-1} \right\}.$$

For instance, if I throw a die until I see all six values, then the chance that I stop at the 10th throw is

$$\mathbb{P}(N = 10) = 6^{-10}\, 6! \left\{ {9 \atop 5} \right\} = .08277.$$

In the coupon collecting problem, we are usually more interested in the easier task of determining $\mathbb{E}(N)$. Write $N = Y_1 + Y_2 + \cdots + Y_k$ where Y_j is the number of trials needed to increase the collection from $j-1$ to j different values. The Y_js are geometric random variables with $p_j = 1 - \frac{j-1}{k}$. It follows that $\mathbb{E}(Y_j) = \frac{1}{p_j} = \frac{k}{k-j+1}$ and hence

$$\mathbb{E}(N) = k\left(\frac{1}{k} + \frac{1}{k-1} + \cdots + \frac{1}{2} + 1\right) \approx k(\log(k) + \gamma).$$

For the die throwing problem, we have

$$\mathbb{E}(N) = 6\left(\frac{1}{6} + \frac{1}{5} + \cdots + \frac{1}{2} + 1\right) = 14.7.$$

Example 9-3 *Birthdays* What is the chance that at least two out of n randomly selected people have the same birthday? We assume that there are r days on the calendar.

We *could* treat this as a missing faces problem, by plugging $k = r - n$ into the inclusion-exclusion formula. However, it is easier to attack the problem directly. If we consider checking the n people one at a time, we see that

$$p_n = \mathbb{P}(n \text{ distinct birthdays}) = \frac{r}{r}\,\frac{r-1}{r}\,\frac{r-2}{r}\cdots\frac{r-n+1}{r} = \binom{r}{n}\frac{n!}{r^n}.$$

For instance with $r = 365$ and $n = 23$, this gives a probability of $p_n = .4927$. The chance of a shared birthday is therefore $1 - p_n = .5073$.

We may also be interested in the number of people N that we need to check before we get the first shared birthday. The formula is

$$\mathbb{E}(N) = \sum_{n=0}^{\infty} \mathbb{P}(N > n) = \sum_{n=0}^{\infty} \binom{r}{n} \frac{n!}{r^n},$$

which unfortunately does not simplify. With $r = 365$ this gives $\mathbb{E}(N) = 24.617$. For large values of r, we have $\mathbb{E}(N) \approx \sqrt{\frac{r\pi}{2}} + \frac{2}{3}$.

Where'd that approximation come from?

$$\begin{aligned}
\sum_{n=0}^{\infty} p_n &= \sum_{n=0}^{\infty} \prod_{j=0}^{n-1} \left(1 - \frac{j}{r}\right) \approx \sum_{n=0}^{\infty} \prod_{j=0}^{n-1} e^{-j/r} = \sum_{n=0}^{\infty} e^{-\sum_{j=0}^{n-1} j/r} \\
&= \sum_{n=0}^{\infty} e^{-\binom{n}{2}/r} \approx \int_0^{\infty} e^{-n^2/2r}\, dn = \sqrt{\frac{r\pi}{2}}.
\end{aligned}$$

(c) Double Occupancy

Find the probability that every face appears twice or more when you roll 20 fair dice.

Solution 1: For $1 \le i \le r$, define A_i to be the event that face i shows up one or fewer times. These events are exchangeable and we will use inclusion-exclusion to find the chance that none of them occur.

Let's start with $\mathbb{P}(A_1)$. The number X_1 of "1"s that occur is a binomial random variable, so we have

$$\mathbb{P}(A_1) = \mathbb{P}(X_1 = 0) + \mathbb{P}(X_1 = 1) = \left(1 - \frac{1}{r}\right)^n + n\left(\frac{1}{r}\right)\left(1 - \frac{1}{r}\right)^{n-1}.$$

For $j > 1$ the vector $(X_1, X_2, \ldots, X_j, Z)$ records the number of "1"s, "2"s, ..., "j"s, plus the number of "others". This is a multinomial vector so

$$\mathbb{P}(A_1 \cdots A_j) = \sum_{0 \le x_1, \ldots, x_j \le 1} \binom{n}{x_1, \ldots, x_j, n - \sum x_k} \left(\frac{1}{r}\right)^{x_1} \cdots \left(\frac{1}{r}\right)^{x_j} \left(1 - \frac{j}{r}\right)^{n - \sum x_k}$$

$$= \sum_{s=0}^{j} \binom{j}{s} \binom{n}{0, \dots, 0, \underbrace{1, \dots, 1}_{s}, n-s} \left(\frac{1}{r}\right)^{s} \left(1 - \frac{j}{r}\right)^{n-s}$$

$$= \sum_{s=0}^{j} \binom{j}{s} \binom{n}{s} s! \left(\frac{1}{r}\right)^{s} \left(1 - \frac{j}{r}\right)^{n-s}.$$

The desired probability is

$$\mathbb{P}(\text{double occupancy}) = \sum_{j=0}^{r} (-1)^{j} \binom{r}{j} \mathbb{P}(A_1 \cdots A_j).$$

Using Stirling numbers, this simplifies to

$$\mathbb{P}(\text{double occupancy}) = \frac{r!}{r^n} \sum_{s=0}^{r} (-1)^{s} \binom{n}{s} \left\{ \begin{matrix} n-s \\ r-s \end{matrix} \right\}.$$

For $r = 6$ and $n = 20$ you find that the probability that every face appears twice or more when you roll 20 fair dice is $\frac{3102938914075}{8463329722368} = 0.3666333$.

Solution 2: Every vector of outcomes $(d_1, d_2, \dots, d_{20})$ has equal probability and there are 6^{20} such vectors. We need the number of such vectors that have at least two of each possible value. We follow the procedure used, for instance, in the example *"One in a million"* that starts the section on the Inclusion-Exclusion principle. The probability is equal to

$$\frac{1340469610880400}{6^{20}} = .3666333.$$

Choose values	Order them	Pattern			
$\binom{6}{5,1}$	$\binom{20}{10,2,2,2,2,2}$	22222 10	6×20951330400	=	125707982400
$\binom{6}{4,1,1}$	$\binom{20}{9,3,2,2,2,2}$	222239	30×69837768000	=	2095133040000
$\binom{6}{4,1,1}$	$\binom{20}{8,4,2,2,2,2}$	222248	30×157134978000	=	4714049340000
$\binom{6}{4,1,1}$	$\binom{20}{7,5,2,2,2,2}$	222257	30×251415964800	=	7542478944000
$\binom{6}{4,2}$	$\binom{20}{6,6,2,2,2,2}$	222266	15×293318625600	=	4399779384000
$\binom{6}{3,2,1}$	$\binom{20}{8,3,3,2,2,2}$	222338	60×209513304000	=	12570798240000
$\binom{6}{3,1,1,1}$	$\binom{20}{7,4,3,2,2,2}$	222347	120×419026608000	=	50283192960000
$\binom{6}{3,1,1,1}$	$\binom{20}{6,5,3,2,2,2}$	222356	120×586637251200	=	70396470144000
$\binom{6}{3,2,1}$	$\binom{20}{6,4,4,2,2,2}$	222446	60×733296564000	=	43997793840000
$\binom{6}{3,2,1}$	$\binom{20}{5,5,4,2,2,2}$	222455	60×879955876800	=	52797352608000
$\binom{6}{3,2,1}$	$\binom{20}{7,3,3,3,2,2}$	223337	60×558702144000	=	33522128640000
$\binom{6}{2,2,1,1}$	$\binom{20}{6,4,3,3,2,2}$	223346	180×977728752000	=	175991175360000
$\binom{6}{2,2,2}$	$\binom{20}{5,5,3,3,2,2}$	223355	90×1173274502400	=	105594705216000
$\binom{6}{2,2,1,1}$	$\binom{20}{5,4,4,3,2,2}$	223445	180×1466593128000	=	263986763040000
$\binom{6}{4,2}$	$\binom{20}{4,4,4,4,2,2}$	224444	15×1833241410000	=	27498621150000
$\binom{6}{4,1,1}$	$\binom{20}{6,3,3,3,3,2}$	233336	30×1303638336000	=	39109150080000
$\binom{6}{3,1,1,1}$	$\binom{20}{5,4,3,3,3,2}$	233345	120×1955457504000	=	234654900480000
$\binom{6}{3,2,1}$	$\binom{20}{4,4,4,3,3,2}$	233444	60×2444321880000	=	146659312800000
$\binom{6}{5,1}$	$\binom{20}{5,3,3,3,3,3}$	333335	6×2607276672000	=	15643660032000
$\binom{6}{4,2}$	$\binom{20}{4,4,3,3,3,3}$	333344	15×3259095840000	=	48886437600000
					1340469610880400

(d) Ménage

Touchard 1934 n couples are seated at a round table with men and women alternating. What is the probability that none of the men has his wife next to him?

Solution: We imagine the seats numbered from 1 to $2n$ in a circle. Let's put the ladies in the odd seats and the gents in the even seats. For $1 \le k \le 2n$ define H_k to be the event that a married couple occupies seats k and $k+1$. Because it's a circle, we consider seat $2n+1$ to be seat 1.

Note that $H_1 H_2$ is impossible, but $H_1 H_3$ is possible. In fact, for indices $i_1, \ldots, i_j$ between 1 and $2n$, we have

$$\mathbb{P}(H_{i_1} \cdots H_{i_j}) = \frac{(n-j)!}{n!},$$

if no two indices are consecutive, measured on the circle. Otherwise

$$\mathbb{P}(H_{i_1} \cdots H_{i_j}) = 0.$$

By the inclusion-exclusion formula, the required probability is

$$p_{2n}(0) = \sum_j (-1)^j \sum_{\{i_1,\ldots,i_j\}} \frac{(n-j)!}{n!},$$

where we only count sets of non-consecutive indices. So the question becomes: how many ways can we place j non-consecutive values on the circle? From section 4(d) we know that the answer is $\frac{2n}{2n-j}\binom{2n-j}{j}$. Plugging this in, gives the answer

$$p_{2n}(0) = \sum_j (-1)^j \frac{2n}{2n-j}\binom{2n-j}{j}\frac{(n-j)!}{n!}.$$

(e) Socks in the dryer

There are n distinct pairs of socks in the dryer, and we randomly pull them out one at a time. Let N be the number of draws needed to get a matching pair. Calculate $\mathbb{E}(N)$.

It is easy to see that $\mathbb{P}(N > k) = \frac{\binom{n}{k}2^k}{\binom{2n}{k}}$ so that

$$\mathbb{E}(N) = \sum_{k=0}^{n} \mathbb{P}(N > k) = \sum_{k=0}^{n} \frac{\binom{n}{k}2^k}{\binom{2n}{k}}.$$

Using the "aside" from Banach's matchbox problem, we simplify this sum

$$\begin{aligned}
\sum_{k=0}^{n} \frac{\binom{n}{k}2^k}{\binom{2n}{k}} &= \frac{1}{\binom{2n}{n}} \sum_{k=0}^{n} \binom{2n-k}{n-k} 2^k \\
&= \frac{1}{\binom{2n}{n}} \sum_{j=0}^{n} \binom{n+j}{j} 2^{n-j} \\
&= \frac{4^n}{\binom{2n}{n}} \sum_{j=0}^{n} \binom{n+j}{n} \left(\frac{1}{2}\right)^{n+j} \\
&= \frac{4^n}{\binom{2n}{n}}.
\end{aligned}$$

10 Random permutations

Suppose we scramble a set of symbols $\{A, B, B, A, \ldots, A\}$ and put them in a random order. There are a copies of A and b copies of B so the total number of permutations is $\binom{a+b}{a}$.

A *run* is a maximal subsequence of one symbol. For example, in $AABBA$ there is a run of two As, then a run of two B's, and finally a run of one A. Let X be the number of runs of As, Y be the the number of runs of Bs, and put $Z = X + Y$.

Define f_{ij} to be the number of permutations with $X = i$ and $Y = j$, and define f_k to be the number of permutations with $Z = k$.

Example 10-1 If $a = 3$ and $b = 2$, there are ten permutations:

[AAABB]	[AABAB]
[AABBA]	[ABAAB]
[ABABA]	[ABBAA]
[BAAAB]	[BAABA]
[BABAA]	[BBAAA]

The f_{ij}s and f_ks are given by the following two tables:

i/j	1	2	3
1	2	2	0
2	1	4	1

k	2	3	4	5
f_k	2	3	4	1

(a) Distribution of (X, Y)

How many ways can we get $(X, Y) = (i, j)$? First divide the A's into i groups; this can be done in $\binom{a-1}{i-1}$ ways. Also divide the B's into j groups; this can be done in $\binom{b-1}{j-1}$ ways.

Considering the number of ways of alternately placing these groups in a line, we get

$$f_{ij} = \begin{cases} \binom{a-1}{i-1}\binom{b-1}{j-1} & \text{if } |i-j| = 1 \\ 2\binom{a-1}{i-1}\binom{b-1}{j-1} & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Example 10-2 If $a = 3$ and $b = 2$ what is the chance that $(X, Y) = (2, 1)$?

Solution:

$$\frac{\binom{3-1}{2-1}\binom{2-1}{1-1}}{\binom{2+3}{2}} = \frac{2}{10}.$$

(b) Expected number of runs

Our random arrangement of symbols has length $a + b$. For $1 \le j \le a + b$ define s_j to be the symbol at position j, and for $1 \le j < a + b$ define

$$Z_j = \begin{cases} 1 & \text{if } s_j \neq s_{j+1} \\ 0 & \text{if } s_j = s_{j+1}. \end{cases}$$

In other words, the random variable Z_j indicates whether or not a new run begins at position $j+1$. The total number of runs is $R = 1 + \sum_{j=1}^{a+b-1} Z_j$, and so the expected number of runs is

$$\begin{aligned}\mathbb{E}(R) &= 1 + \sum_{j=1}^{a+b-1} \mathbb{E}(Z_j) \\ &= 1 + (a+b-1)\,\mathbb{P}(s_j \neq s_{j+1}) \\ &= 1 + \frac{2ab}{a+b}.\end{aligned}$$

Here we use the fact that

$$\mathbb{P}(s_j \neq s_{j+1}) = \frac{a}{a+b} \cdot \frac{b}{a+b-1} + \frac{b}{a+b} \cdot \frac{a}{a+b-1}.$$

(c) The Mississippi problem

If we randomly scramble the letters in the word "MISSISSIPPI", what is the chance that no consecutive letters are the same?

We will solve the problem in three steps:

1. First mix the four S's and four I's
2. Next blend in the two P's
3. Finally put in the M.

Let E_1, E_2, and E_3 be the number of consecutive letters of the same type after steps $1, 2, 3$ respectively. We are after $\mathbb{P}(E_3 = 0)$, and our strategy is to compute it as

$$\mathbb{P}(E_3 = 0) = \sum_{j=0}^{3} \sum_{k=0}^{1} \mathbb{P}(E_3 = 0 \,|\, E_1 = j, E_2 = k)\,\mathbb{P}(E_2 = k \,|\, E_1 = j)\,\mathbb{P}(E_1 = j).$$

Here is the joint distribution for the number of runs of S's and I's after step 1.

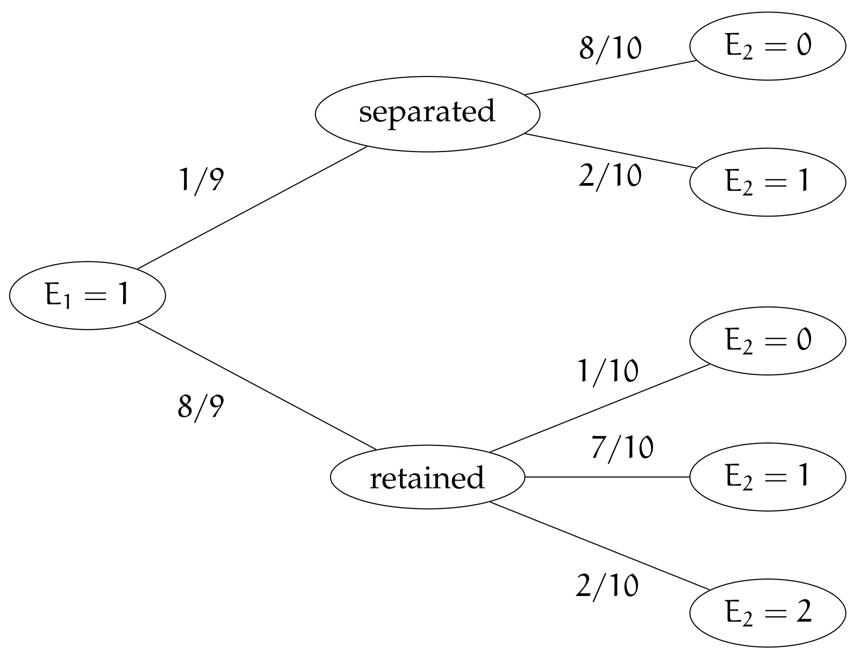
$i \backslash s$	1	2	3	4
1	$\frac{2}{70}$	$\frac{3}{70}$	0	0
2	$\frac{3}{70}$	$\frac{18}{70}$	$\frac{9}{70}$	0
3	0	$\frac{9}{70}$	$\frac{18}{70}$	$\frac{3}{70}$
4	0	0	$\frac{3}{70}$	$\frac{2}{70}$

From this we deduce that the distribution of E_1 is

j	0	1	2	3	4	5	6
$\mathbb{P}(E_1 = j)$	$\frac{2}{70}$	$\frac{6}{70}$	$\frac{18}{70}$	$\frac{18}{70}$	$\frac{18}{70}$	$\frac{6}{70}$	$\frac{2}{70}$

We only need the values $j = 0, 1, 2, 3$ from this table.

Using the following tree diagram, we can calculate the conditional probabilities $\mathbb{P}(E_2 = 0 \mid E_1 = 1) = \frac{16}{90}$ and $\mathbb{P}(E_2 = 1 \mid E_1 = 1) = \frac{58}{90}$. The branch points in the tree correspond to adding in a single letter P.



Similarly, we can analyze the cases $E_1 = 0, 2, 3$ and get the following conditional probabilities $\mathbb{P}(E_2 = k \mid E_1 = j)$:

$k \backslash j$	0	1	2	3
0	$\frac{72}{90}$	$\frac{16}{90}$	$\frac{2}{90}$	0
1	$\frac{18}{90}$	$\frac{58}{90}$	$\frac{28}{90}$	$\frac{6}{90}$

Now we note that $\mathbb{P}(E_3 = 0 \mid E_2 = 0) = 1$ and that $\mathbb{P}(E_3 = 0 \mid E_2 = 1) = 1/11$. We now have all the information we need to calculate that $\mathbb{P}(E_3 = 0) = \frac{16}{275}$. ∎

Example 10-3 Show that the probability that a random permutation of the word STATISTICS has no equal neighbors is 1/5.

Example 10-4 Show that the probability that a random permutation of the word EDMONTON has no equal neighbors is 4/7.

A sophisticated solution: These problems quickly get out of hand if the words are long and there are lots of multiple letters. Here is an alternative solution that uses ideas from *algebraic combinatorics*.

Define polynomials for $k \geq 1$ by $q_k(x) = \sum_{i=1}^{k} \frac{(-1)^{i-k}}{i!}\binom{k-1}{i-1}x^i$. Here are the first few polynomials:

$$q_1(x) = x, \quad q_2(x) = x^2/2 - x, \quad q_3(x) = x^3/6 - x^2 + x.$$

The number of permutations with no equal neighbors, using an alphabet with frequencies $k_1, k_2, \ldots$ is:

$$\boxed{\int_0^\infty \prod_j q_{k_j}(x)\, e^{-x}\, dx.}$$

Example 10-5 For the MISSISSIPPI problem, the product of the q functions is

$$\begin{aligned}
& q_4(x)^2\, q_2(x)\, q_1(x) \\
= \; & (x^4/24 - x^3/2 + 3x^2/2 - x)^2\,(x^2/2 - x)\,(x) \\
= \; & x^{11}/1152 - 13x^{10}/576 + 11x^9/48 - 7x^8/6 + 77x^7/24 - 19x^6/4 + 7x^5/2 - x^4.
\end{aligned}$$

Substituting $j!$ for x^j above, we get

$$\begin{aligned}
& 11!/1152 - 13 \cdot 10!/576 + 11 \cdot 9!/48 - 7 \cdot 8!/6 + 77 \cdot 7!/24 - 19 \cdot 6!/4 + 7 \cdot 5!/2 - 4! \\
= \; & 34650 - 81900 + 83160 - 47040 + 16170 - 3420 + 420 - 24 \\
= \; & 2016.
\end{aligned}$$

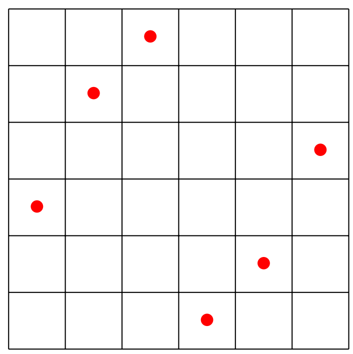
Dividing by the total number of permutations $\frac{11!}{4!4!2!1!} = 34650$, we get a probability of $\frac{2016}{34650} = \frac{16}{275}$. Amazing!

(d) Secret Santa

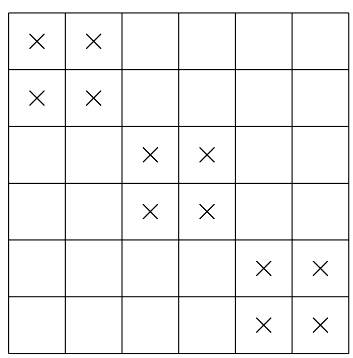
Imagine three couples who put their six names in a hat, and randomly draw one name each to buy a Christmas gift. We would like the probability that nobody gets their own name or the name of their partner.

In general, randomly permuting the values $1, 2, \ldots, N$ corresponds to randomly selecting N dots on a $N \times N$ chessboard, with one dot in each row and column.

For instance, the permutation $\langle 4, 2, 1, 6, 5, 3 \rangle$ corresponds to the diagram below.



Now suppose there are some forbidden positions



For each forbidden position i, let A_i be the event that a spot lands there. Then $S_j = \sum \mathbb{P}(A_{i_1} \cdots A_{i_j})$, where each $\mathbb{P}(A_{i_1} \cdots A_{i_j})$ is either equal to zero, or to $(N-j)!/N! = 1/(N)_j$. Here $(N)_j$ is defined to be $(N)(N-1)(N-2)\cdots(N-j+1)$. The probability is zero precisely when two of the positions $\{i_1, \ldots, i_j\}$ are in the same row or column.

Therefore we have $S_j = \frac{R_j}{(N)_j}$, where R_j is the number of ways to place j non-taking rooks in the forbidden zone. The probability that none of the spots lands in a forbidden place is

$$\mathbb{P}(X=0) = \sum_{j=0}^{N} (-1)^j \frac{R_j}{(N)_j}. \tag{1}$$

Example 10-1 When there are no forbidden places, the rook numbers are $R_j = 1_0(j)$ so plugging into (1) gives $\mathbb{P}(X=0) = (-1)^0 R_0/(N)_0 = 1$, as expected.

These "rook numbers" R_j are often calculated using the generating function $R(x) = \sum_j R_j x^j$ called the *rook polynomial*. Equation (1) can be rewritten in terms of R as

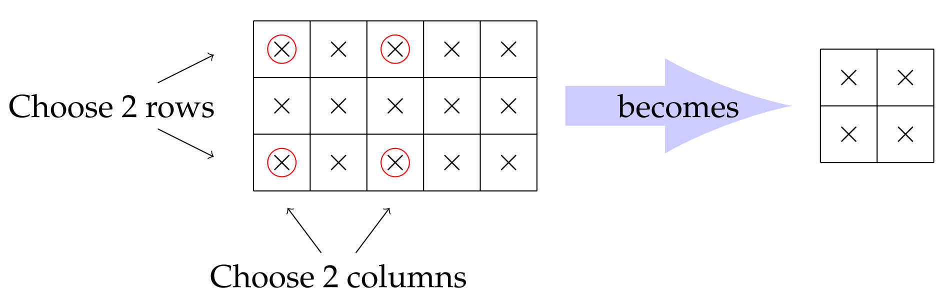
$$\mathbb{P}(X=0) = \frac{1}{N!} \int_0^\infty x^N \, R(-1/x) \, e^{-x} \, dx. \tag{2}$$

Equation (2) is useful for calculating $\mathbb{P}(X=0)$ using a computer.

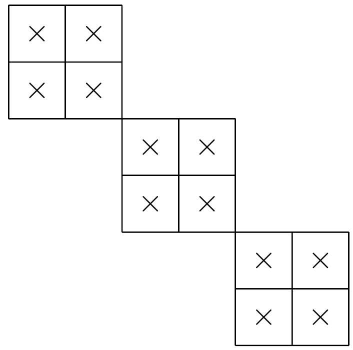
Example 10-2 The rook polynomial for a completely filled $m \times n$ rectangle is

$$R(x) = \sum_j \binom{m}{j} \binom{n}{j} j! \, x^j.$$

To see this, imagine placing j non-taking rooks in this rectangle. You can freely choose j rows and j columns to use. Now you have a $j \times j$ square, and there are $j!$ ways to place the rooks there.



Example 10-3 Consider the forbidden zone for the secret santa problem.



Each of the three blocks shown has its own rook polynomial, which, using the previous example, is $1 + 4x + 2x^2$. The overall rook polynomial for the forbidden zone illustrated above is

$$R(x) = (1 + 4x + 2x^2)^3 = 1 + 12x + 54x^2 + 112x^3 + 108x^4 + 48x^5 + 8x^6.$$

Substituting the rook numbers into (1) above, we get

$$\begin{aligned}\mathbb{P}(X = 0) &= 1 - \frac{12}{6} + \frac{54}{6 \cdot 5} - \frac{112}{6 \cdot 5 \cdot 4} + \frac{108}{6 \cdot 5 \cdot 4 \cdot 3} - \frac{48}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2} + \frac{8}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} \\ &= \frac{1}{9}.\end{aligned}$$

Of course, with only three couples it is much easier to calculate this by brute force.

For five couples, the following Maple code gives the answer immediately.

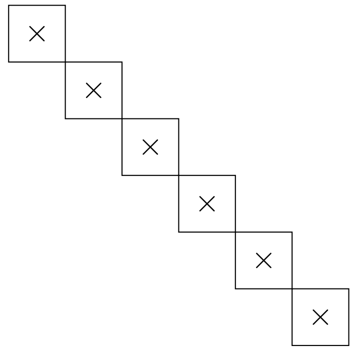
```
> R:=x->(1+4*x+2*x^2)^5:
> int(x^10*R(-1/x)*exp(-x),x=0..infinity)/10!;

                                3439
                                -----
                                28350

> evalf(%);

                             0.1213051146
```

Example 10-4 In the secret santa problem, if we only forbid people from drawing their own name then the forbidden zone consists of the main diagonal. In other words, this is the "rencontre" problem.



The rook polynomial for this zone is $R(x) = (1+x)^N$, which means that the rook numbers are $R_j = \binom{N}{j}$. Therefore,

$$\mathbb{P}(X=0) = \sum_{j=0}^{N} (-1)^j \frac{\binom{N}{j}}{(N)_j} = \sum_{j=0}^{N} \frac{(-1)^j}{j!}.$$

Example 10-5 Secret santa in a circle. Six people are seated at a round table, and they put their names in a hat for a gift draw. What is the chance that nobody draws their own name or the name of the person seated to their right?

Solution: For convenience I will mark the forbidden zone using numbers instead of $\times$. You can see that two rooks in this zone will be non-taking exactly when they land on non-consecutive numbers. Non-consecutive here means taken in a circle, so that 12 is beside both 11 and 1.

1					12
2	3				
	4	5			
		6	7		
			8	9	
				10	11

From section 4(d) we know that these rook numbers are

$$R_j = \frac{12}{12-j}\binom{12-j}{j},$$

so that

$$\mathbb{P}(X=0) = \sum_{j=0}^{6} (-1)^j \frac{12}{12-j} \binom{12-j}{j} \frac{(6-j)!}{6!} = \frac{1}{9}.$$

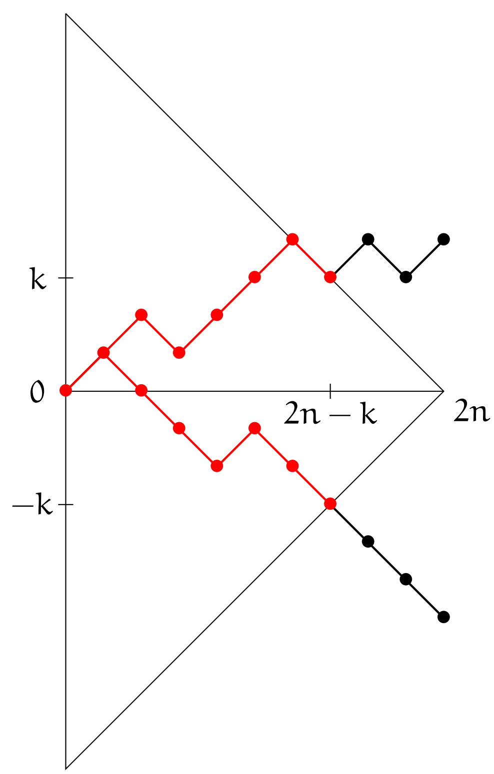
This formula is the same as "ménage" with 6 couples.

11 Counting paths

Example 11-1 We can prove the formula from "socks in the dryer"

$$\sum_{k=0}^{n} \binom{2n-k}{n-k} 2^k = 4^n,$$

by counting paths. Do you see why?



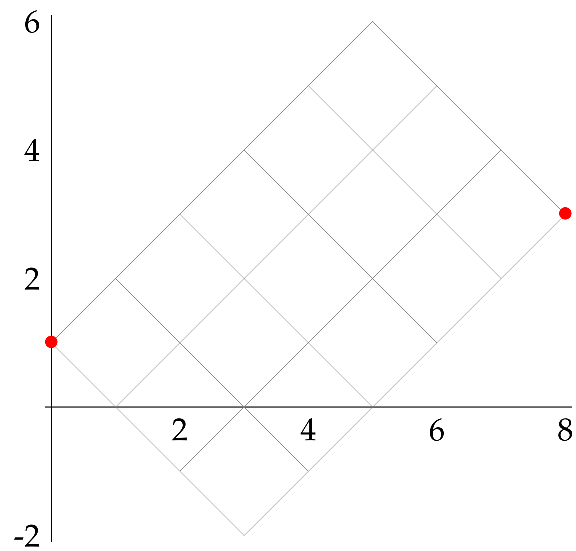
(a) The reflection principle

For the time being, we consider paths on the integers whose increments are either plus or minus one. Define $N_{m,n}(a, b)$ to be the number of such paths between (m, a) and (n, b). Each such path has α steps up and β steps down. Since $\alpha + \beta = |n - m|$ and $\alpha - \beta = b - a$ we deduce that

$$\alpha = \frac{1}{2}(|n - m| + b - a),$$

and therefore that

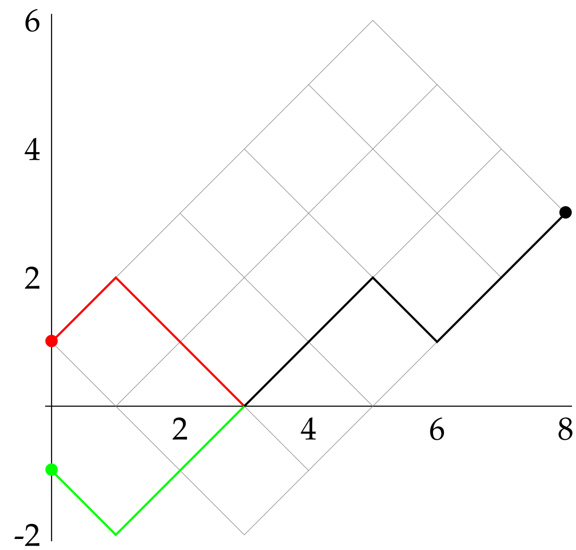
$$N_{m,n}(a, b) = \begin{cases} \binom{|n-m|}{\alpha} & \text{if } \alpha \text{ is an integer} \\ 0 & \text{otherwise.} \end{cases}$$



The number of paths from $(0, 1)$ to $(8, 3)$ is $\binom{|8-0|}{\frac{1}{2}[|8-0|+(3-1)]} = \binom{8}{5} = 56$.

We want to determine the number $N^0_{m,n}(a, b)$ of paths from (m, a) to (n, b) that do not touch the x-axis. We do this by considering the complement: let $N^1_{m,n}(a, b)$ count the number of paths from (m, a) to (n, b) that do touch the x-axis. The reflection principle says that, if $ab \geq 0$, then

$$\boxed{N^1_{m,n}(a, b) = N_{m,n}(-a, b).}$$



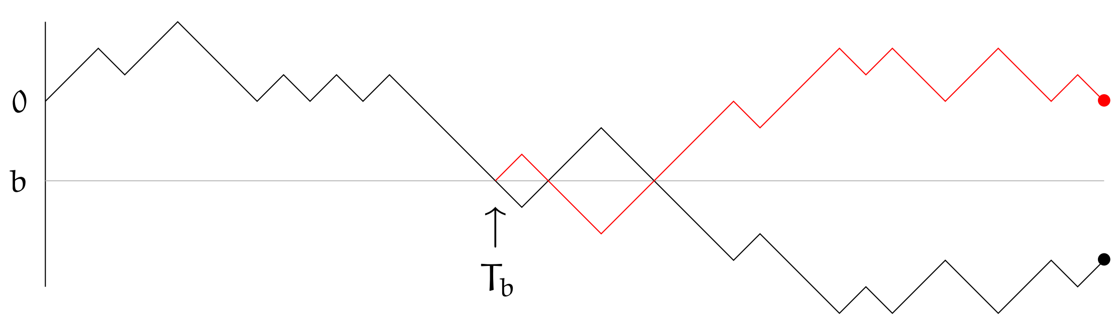
The number of paths from $(0, 1)$ to $(8, 3)$ that touch the x-axis
$=$ the number of paths from $(0, -1)$ to $(8, 3)$.

From this we get

$$N^0_{m,n}(a, b) = N_{m,n}(a, b) - N_{m,n}(-a, b).$$

Example 11-2 *Symmetric random walk* Let (S_n) be simple, symmetric random walk, and let b be a negative integer. By reflecting the path at the level b, where $b + n$ is odd, we find that

$$\mathbb{P}_0(S_n < b) = \frac{1}{2}\,\mathbb{P}_0(T_b < n).$$



Reflecting the path when it hits level b.

This can be rewritten as

$$\mathbb{P}_0(T_b \geq n) = \mathbb{P}_0(b \leq S_n \leq -b).$$

Plugging in $b = -1$ at time $2n$ gives

$$\mathbb{P}_0(T_{-1} \geq 2n) = \mathbb{P}_0(-1 \leq S_{2n} \leq 1).$$

In other words

$$\mathbb{P}_0(S_0 \geq 0, S_1 \geq 0, \ldots, S_{2n} \geq 0) = \mathbb{P}_0(S_{2n} = 0).$$

Example 11-3 Now let's look at the paths that don't hit zero from time 0 to time $2n$. We have

$$\begin{aligned}
\mathbb{P}_0(S_1 \neq 0, S_2 \neq 0, \ldots, S_{2n} \neq 0) &= \mathbb{P}_0(S_1 = \pm 1, R_0 > 2n) \\
&= 2\mathbb{P}_0(S_1 = 1, R_0 > 2n) \\
&= \mathbb{P}_1(T_0 \geq 2n) \\
&= \mathbb{P}_0(T_{-1} \geq 2n) \\
&= \mathbb{P}_0(S_{2n} = 0).
\end{aligned}$$

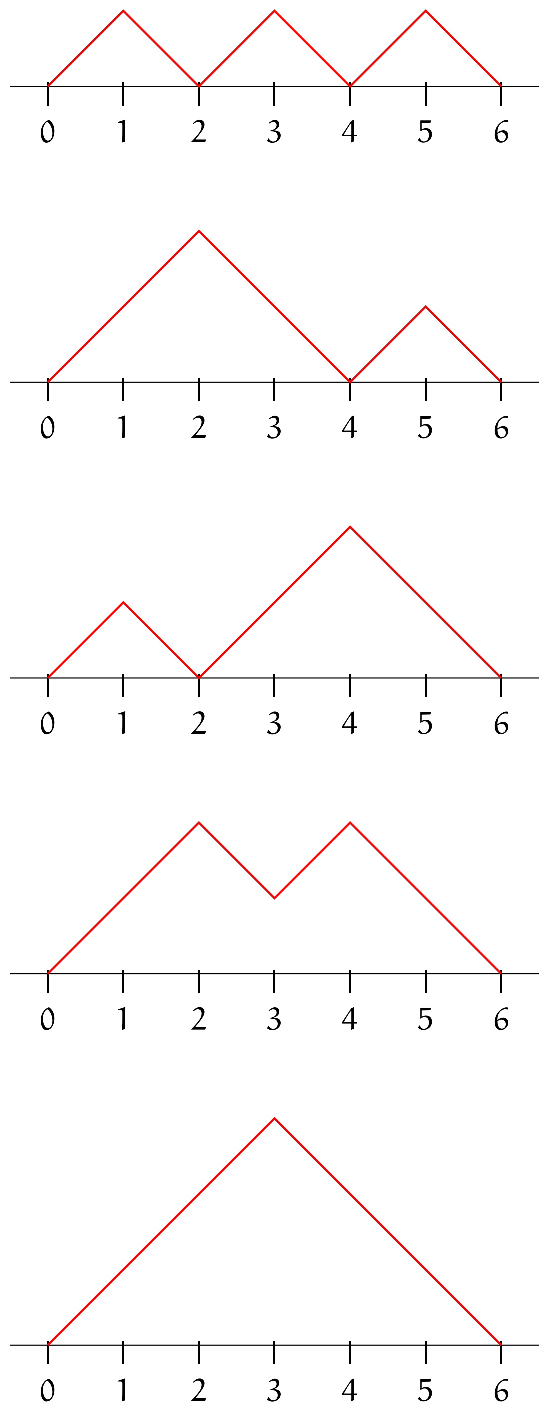
We can use these path counting arguments to derive a binomial identity. There are a total of 4^n paths up to time $2n$. If we let $2r$ stand for the last time that the path hits zero, then by considering cases and using the Example 11.2, we get

$$4^n = \sum_{r=0}^{n} \binom{2r}{r} \binom{2n-2r}{n-r}.$$

(b) Catalan numbers

The Catalan numbers C_n count the number of non-negative paths from $(0,0)$ to $(2n,0)$. The first few values are $C_0 = 1$, $C_1 = 1$, $C_2 = 2$, $C_3 = 5$, and $C_4 = 14$.

The five paths that demonstrate $C_3 = 5$



We will use the reflection principle to find a formula for the Catalan numbers. Lifting the path by one unit, we find that

$$\begin{aligned} C_n &= N^0_{0,2n}(1,1) \\ &= N_{0,2n}(1,1) - N_{0,2n}(-1,1) \\ &= \binom{2n}{n} - \binom{2n}{n+1} \\ &= \frac{1}{n+1}\binom{2n}{n}. \end{aligned}$$

Example 11-4 Draw cards one at a time from a well-shuffled deck and put them on the red pile or the black pile. What is the chance that the piles become equal for the first time, at the tenth draw?

Solution: The chance that the first ten cards have five red and five black is $\binom{26}{5}\binom{26}{5}/\binom{52}{10}$.

Randomly ordering these five red cards and five black cards, the chance that the piles are equal for the first time with the tenth card is $2C_4/\binom{10}{5}$.

Putting it together, the required probability is

$$\frac{2C_4\binom{26}{5}^2}{\binom{10}{5}\binom{52}{10}} = \frac{65780}{2164491} = 0.0304.$$

Example 11-5 What is the probability that a random walk will first return to zero at time $2n$?

Solution: We need to consider all paths from $(0,0)$ to $(2n,0)$ that do not touch the x-axis. The positive ones consist of an upstep, followed by a path from $(1,1)$ to $(2n-1,1)$ which doesn't hit the level zero, followed by a downstep. The total probability of this set of paths is

$$p \cdot C_{n-1}\, p^{n-1} q^{n-1} \cdot q.$$

Similarly, the total probability of the set of negative paths is

$$q \cdot C_{n-1}\, p^{n-1} q^{n-1} \cdot p,$$

and adding these together, we get the answer

$$\mathbb{P}_0(R_0 = 2n) = 2\, C_{n-1}\, p^n q^n.$$

(c) The ballot theorem

In an election, candidate A gets a votes and candidate B gets b votes, with $a > b$. What is the chance that A holds the lead throughout the election?

Solution: Count a vote for A as an upstep and a vote for B as a downstep. We want to count all paths from $(0,0)$ to $(a+b, a-b)$ that only touches the x-axis at $(0,0)$. The number of such paths is

$$\begin{aligned} N^0_{1,a+b}(1, a-b) &= N_{1,a+b}(1, a-b) - N_{1,a+b}(-1, a-b) \\ &= \binom{a+b-1}{a-1} - \binom{a+b-1}{a} \\ &= \frac{a-b}{a+b}\binom{a+b}{a}. \end{aligned}$$

Dividing by the total number of paths, we get $P = \frac{a-b}{a+b}$.

(d) Theater lineup

$f + t$ people are in line at a theatre; f with five dollar bills and t with ten dollar bills. The tickets cost five dollars each, and the till is empty to start. If each customer buys one ticket, what is the chance that nobody will have to wait for change?

Solution: If $t > f$, then the chance is zero, so let's assume that $t \leq f$.

Put the $f + t$ people in random order. Start at zero and take a step up for a five dollar bill and a step down for a ten dollar bill. The height of this sample path is the number of five dollar bills that the cashier has. We want this path to stay non-negative. This similar to the "ballot problem", except that we allow the path to touch the x-axis.

The good sample paths for our problem are in one-to-one correspondence with the good sample paths for the ballot problem with one extra five dollar bill. (Add the new person to the front of the line).

The ballot theorem says that the number of such paths is

$$\frac{f+1-t}{f+1+t}\binom{f+1+t}{t}.$$

Dividing by $\binom{f+t}{t}$ gives the desired probability $\dfrac{f+1-t}{f+1}$.

(e) The cycle lemma

In the next two sections, we will consider paths whose steps are integers less than or equal one. Let $\langle x_1, x_2, \ldots, x_n\rangle$ be a sequence of integers ≤ 1, and put $s_0 = 0$ and $s_j = \sum_{i=1}^{j} x_i$ for $1 \leq j \leq n$. We consider the path (j, s_j) as j runs from 0 to n.

A rotation of the path, is the path formed using a rotated sequence of increments

$$\langle x_r, x_{r+1}, \ldots, x_n, x_1, \ldots, x_{r-1}\rangle.$$

Cycle lemma. If $s_n > 0$, then there are exactly s_n rotations whose partial sums are strictly positive.

Corollary. If $s_n > 0$, then there are exactly s_n rotations so that $s_j < s_n$ for $0 \leq j < n$.

Proof: Apply the cycle lemma to the reversed sequence

$$x^{\mathrm{rev}} = \langle x_n, x_{n-1}, \ldots, x_1\rangle.$$

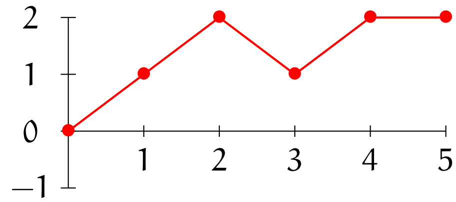
We have

$$s_n - s_j = s^{\mathrm{rev}}_{n-j}, \quad 0 \leq j < n.$$

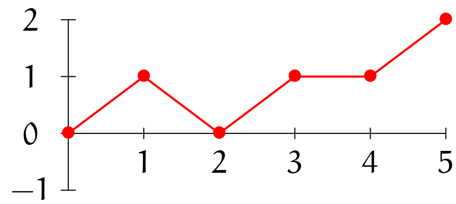
∎

Example 11-6

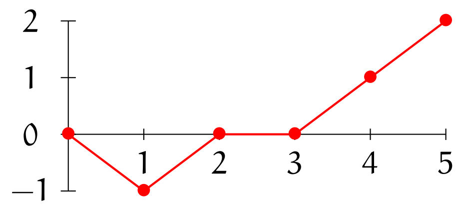
$x = \langle 1, 1, -1, 1, 0 \rangle$



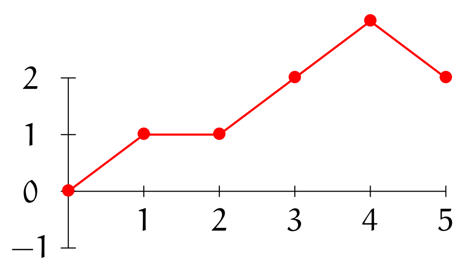
$x = \langle 1, -1, 1, 0, 1 \rangle$



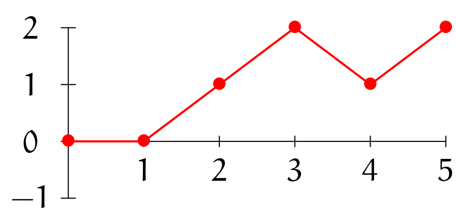
$x = \langle -1, 1, 0, 1, 1 \rangle$



$x = \langle 1, 0, 1, 1, -1 \rangle$



$x = \langle 0, 1, 1, -1, 1 \rangle$



(f) Applications

Example 11-7 *The ballot theorem revisited* Let candidate A get a votes and candidate B get b votes where $a \geq mb$ for some integer $m \geq 1$. The probability that A maintains more than m times as many votes as B throughout the election is $\frac{a-mb}{a+b}$.

Proof: For any arrangement of votes

$$AABAAAB \cdots AB,$$

replace each A with 1 and each B with $-m$. There are $a - mb$ rotations that maintain strictly positive partial sums. Since all $a + b$ rotations are equally likely we get the result. ∎

Let X_j be random variables taking values in $\{\ldots, -2, -1, 0, 1\}$ such that the distribution of $(X_1, X_2, \ldots, X_n)$ is invariant under rotations.

Example 11-8 For $k, n \geq 1$, the cycle lemma implies

$$\mathbb{P}_0(S_j > 0 \text{ for } 1 \leq j \leq n, S_n = k) = \frac{k}{n}\,\mathbb{P}_0(S_n = k).$$

Example 11-9 If all the steps are in $\{-1, 0, 1\}$, then we can generalize the above to all integers k:

$$\mathbb{P}_0(R_0 > n, S_n = |k|) = \frac{|k|}{n}\,\mathbb{P}_0(S_n = k).$$

Adding over k gives us $\mathbb{P}_0(R_0 > n) = \frac{1}{n}\,\mathbb{E}(|S_n|)$. In particular, letting $n \to \infty$ gives $\mathbb{P}_0(R_0 = \infty) = |\mu|$, where μ is the average step size.

Example 11-10 *The hitting time lemma* For $k \geq 1$, we have

$$\mathbb{P}_0(T_k = n) = \frac{k}{n}\,\mathbb{P}_0(S_n = k).$$

12 Formulas

(a) Binomial coefficients

For any integers n, k, we define

$$\binom{n}{k} = \begin{cases} \dfrac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots(1)} & \text{if } k > 0 \\ 1 & \text{if } k = 0 \\ 0 & \text{if } k < 0 \end{cases}$$

Basic identities:

$$\begin{aligned}
\binom{n}{k} &= \binom{n}{n-k} \qquad \text{if } n \geq 0 \\
\binom{n}{k} &= \binom{n-1}{k} + \binom{n-1}{k-1} \\
\binom{n}{k} &= (-1)^k \binom{k-n-1}{k} \\
\binom{n}{k}\binom{k}{m} &= \binom{n}{m}\binom{n-m}{k-m} \\
k\binom{n}{k} &= n\binom{n-1}{k-1}
\end{aligned}$$

Some Sums:

$$\begin{aligned}
\sum_{k \leq m} \binom{n+k}{k} &= \binom{n+m+1}{m} \\
\sum_{k} \binom{n}{k}\binom{m}{\ell-k} &= \binom{n+m}{\ell} \\
\sum_{k} \binom{n}{k} x^k y^{n-k} &= (x+y)^n \qquad \text{if } n \geq 0 \text{ or } |x/y| < 1
\end{aligned}$$

$$\sum_k \binom{n}{k}\binom{k}{m} x^k \;=\; x^m \binom{n}{m}(1+x)^{n-m} \qquad \text{if } n \geq 0 \text{ or } |x| < 1$$

Valid for $m, n \geq 0$**:**

$$\sum_{0 \leq k \leq m} \binom{k}{n} \;=\; \binom{m+1}{n+1}$$

$$\sum_k \binom{n+k}{k} x^k \;=\; \frac{1}{(1-x)^{n+1}} \qquad \text{if } |x| < 1$$

$$\sum_{k \leq m} \binom{n}{k}(-1)^k \;=\; (-1)^m \binom{n-1}{m}$$

(b) Stirling numbers

Recurrences:

$$\binom{n}{k} \;=\; \binom{n-1}{k} + \binom{n-1}{k-1}$$

$$\left\{ {n \atop k} \right\} \;=\; k\left\{ {n-1 \atop k} \right\} + \left\{ {n-1 \atop k-1} \right\}$$

Power conversion:

$$x^n \;=\; \sum_k \left\{ {n \atop k} \right\} x(x-1)\cdots(x-k+1)$$

(c) Finite calculus

If f is defined on the integers, then we let $\Delta f(x) = f(x+1) - f(x)$. The following telescoping sum is the analogue of the fundamental theorem of calculus:

$$f(x) = f(a) + \sum_{a \le y < x} \Delta f(y).$$

We can also take second differences:

$$\begin{aligned}
\Delta^2 f(x) &= \Delta f(x+1) - \Delta f(x) \\
&= [f(x+2) - f(x+1)] - [f(x+1) - f(x)] \\
&= f(x+2) - 2f(x+1) + f(x).
\end{aligned}$$

In fact, for any integer $n \ge 0$ we have

$$\Delta^n f(x) = \sum_k \binom{n}{k} (-1)^{n-k} f(x+k).$$

13 References

1. *Problems and Snapshots From the World of Probability*, Gunnar Blom, Lars Holst, and Dennis Sandell, Springer-Verlag (1994).

2. *Challenging Mathematical Problems with Elementary Solutions Volume 1: Combinatorial Analysis and Probability Theory*, Akiva M. Yaglom and Isaak M. Yaglom, Dover Publishing (1987).

3. *One Thousand Exercises in Probability*, Geoffrey Grimmett and David Stirzaker, Oxford University Press (2001).

4. *Fifty Challenging Problems in Probability*, Frederick Mosteller, Dover Publishing (1987).

5. *Concrete Mathematics (2nd edition)*, Ronald L. Graham, Donald E. Knuth, and Oren Patashnik, Addison-Wesley (1994).

6. *Problems column*, The American Mathematical Monthly, published ten times per year, `www.maa.org/pubs/monthly.html`.

7. *Stack Exchange – Mathematics*, `math.stackexchange.com`

8. *Analytic Combinatorics*, Philippe Flajolet and Robert Sedgewick, Cambridge University Press (2009).