Chapter 5

Unexpected symmetry

The sampling problem in Chapter 4 made use of a symmetry property to simplify calculations of variances and covariances: if $X_1, X_2, \ldots$ identify the successive balls taken from an urn (with or without replacement) then each X_i has the same distribution, and each pair (X_i, X_j) with $i \neq j$ has the same distribution. Without the symmetry simplification, calculation of covariances for sampling without replacement would have been a fearsome task.

You should always look for symmetry properties before slogging your way through calculations with what might seen the obvious method. Symmetry, like a fairy godmother, can turn up in unexpected places.

<5.1> **Example.** Suppose an urn initially contains r red balls and b black balls. Suppose balls are sampled from the urn one at a time, but after each draw $k+1$ balls of the same color are returned to the urn (with thorough mixing between draws, blindfolds, and so on). If $k = 0$, the procedure is just sampling with replacement. The number of red balls in the first n draws would then have a Bin(n,r/(r+b)) distribution. If $k = -1$, the procedure is sampling without replacement. If $k \geq 1$ we will need a very big urn if we intend to sample for a long time; there will be $r + b + ki$ balls in the urn after the ith sampling.

•Polya urn

The return of multiple balls to the urn gives a crude model for contagion, whereby the occurrence of an event (such as selection of a red ball) makes the future occurrence of similar events more likely. The model is known as the POLYA URN scheme.

QUESTIONS (FOR GENERAL k):

(a) What is the distribution of the number of red balls in the first n draws?

(b) What is the probability that the ith ball drawn is red?

(c) What is the expected number of red balls in the first n draws?

To answer these questions we do not need to keep track of exactly which ball is selected at each draw; only its color matters. The questions involve only the events

$$R_i = \{i\text{th ball drawn from urn is red}\}$$

and their complements B_i, for $i = 1, 2, \ldots$. Clearly $\mathbb{P}R_1 = r/(r+b)$.

To get a feel for what is going on, start with some simple calculations for the first few draws, using straightforward conditioning.

$$\begin{aligned}
\mathbb{P}R_2 &= \mathbb{P}R_1R_2 + \mathbb{P}B_1R_2 \\
&= \mathbb{P}R_1\mathbb{P}(R_2 \mid R_1) + \mathbb{P}B_1\mathbb{P}(R_2 \mid B_1) \\
&= \frac{r}{r+b} \times \frac{r+k}{r+b+k} + \frac{b}{r+b} \times \frac{r}{r+b+k} \\
&= \frac{r(r+k)+rb}{(r+b)(r+k+b)} \\
&= \frac{r}{r+b}
\end{aligned}$$

Slightly harder:

$$
\begin{aligned}
\mathbb{P}R_3 &= \mathbb{P}(R_1R_2R_3) + \mathbb{P}(R_1B_2R_3) + \mathbb{P}(B_1R_2R_3) + \mathbb{P}(B_1B_2R_3) \\
&= \frac{r}{r+b} \times \frac{r+k}{r+b+k} \times \frac{r+2k}{r+b+2k} \\
&\quad + \frac{r}{r+b} \times \frac{b}{r+b+k} \times \frac{r+k}{r+b+2k} \\
&\quad + \frac{b}{r+b} \times \frac{r}{r+b+k} \times \frac{r+k}{r+b+2k} \\
&\quad + \frac{b}{r+b} \times \frac{b+k}{r+b+k} \times \frac{r}{r+b+2k}
\end{aligned}
$$

Each summand has the same denominator:

$$(r+b)(r+b+k)(r+b+2k),$$

corresponding to the fact that the number of balls in the urn increases by k after each draw. The sum of the numerators rearranges to

$$
\begin{aligned}
&\big(r(r+k)(r+2k) + r(r+k)b\big) + \big(rb(r+k) + rb(b+k)\big) \\
&= r(r+k)(r+2k+b) + rb(r+2k+b) \\
&= r(r+k+b)(r+2k+b)
\end{aligned}
$$

The last two factors, $r+k+b$ and $r+2k+b$, cancel with the same factors in the denominator, leaving $\mathbb{P}R_3 = r/(r+b)$.

REMARK:There is something wrong with the calculation of $\mathbb{P}R_3$ in the case $r = 1$ and $k = -1$ if we interpret each of the factors in a product like

$$\frac{r}{r+b} \times \frac{r+k}{r+b+k} \times \frac{r+2k}{r+b+2k}$$

as a conditional probability. The third factor would become $(-1)/(b-1)$, which is negative: the urn had run out of balls after the previous draw. Fortunately the second factor reduces to zero. The product of these factors is zero, which is the correct value for $\mathbb{P}(R_1R_2R_3)$ when $r = 1$ and $k = -1$. The oversight did not invalididate the final answer. *Moral: The value of a conditional probability needn't make sense if it is to be multiplied by zero.*

By now you probably suspect that the answer to question (b) is $r/(r+b)$, no matter what the value of k. A symmetry argument will prove your suspicions correct. Look for the pattern in probabilities like $\mathbb{P}(R_1R_2B_3\ldots)$ when expressed as a ratio of two products. The successive factors in the denominator correspond to the numbers of balls in the urn before each draw. The same factors will appear no matter what string of R_i's and B_i's is involved. In the numerator, the first appearance of an R_i contributes an r, the second appearance contributes an $r+k$, and so on. The B_i's contribute b, then $b+k$, then $b+2k$, and so on. For example,

$$\mathbb{P}(R_1R_2B_3B_4R_5B_6R_7) = \frac{r(r+k)(r+2k)(r+3k)b(b+k)(b+2k)}{(r+b)(r+b+k)(r+b+2k)(r+b+3k)\ldots(r+b+6k)}$$

You might like to rearrange the order of the factors in the numerator to make the representation as a product of conditional probabilities clearer.

In short, the probability of a particular string of R_i's and B_i's, corresponding to a particular sequence of draws from the urn, depends only on the number of R_i and B_i terms, and not on their ordering.

Answer to question (a)

For $i = 0, 1, \ldots, n$, we need to calculate the probability of getting exactly i red balls amongst the first n draws. There are $\binom{n}{i}$ different orderings for the first n draws that would give exactly i reds. (Think of the number of ways to choose the i positions for the red from the n available). The event $\{i$ reds in first n draws$\}$ is a disjoint union of $\binom{n}{i}$ equally likely events, whence

$$\begin{aligned}
&\mathbb{P}\{i \text{ reds in first } n \text{ draws}\}\\
&\quad= \binom{n}{i}\mathbb{P}R_1R_2\ldots R_iB_{i+1}B_{i+2}\ldots B_n\\
&\quad= \binom{n}{i}\frac{r(r+k)\ldots(r+k(i-1))b(b+k)\ldots(b+k(n-i-1))}{(r+b)(r+b+k)\ldots(r+b+k(n-1))}
\end{aligned}$$

As a quick check, notice that when k=0, the probability reduces to

$$\binom{n}{i}\left(\frac{r}{r+b}\right)^i\left(\frac{b}{r+b}\right)^{n-i},$$

as it should be for a $\text{Bin}(n, r/(r+b))$ distribution.

For the special case of sampling without replacement ($k=-1$), the probability becomes

$$\begin{aligned}
&\binom{n}{i}\frac{r(r-1)\ldots(r-i+1)b(b-1)\ldots(b-n+i+1)}{(r+b)(r+b-1)\ldots(r+b-n+1)}\\
&\quad= \frac{n!}{i!(n-i)!}\,\frac{r!}{(r-i)!}\,\frac{b!}{(b-n+i)!}\,\frac{(r+b-n)!}{(r+b)!}\\
&\quad= \frac{r!}{i!(r-i)!}\,\frac{b!}{(n-i)!(b-n+i)!}\,\frac{n!(r+b-n)!}{(r+b)!}\\
&\quad= \binom{r}{i}\binom{b}{n-i}\Big/\binom{r+b}{n} \qquad \text{<5.2>}
\end{aligned}$$

Notice that

$$\begin{aligned}
\binom{r}{i} &= \text{number of ways to choose } i \text{ from } r \text{ reds}\\
\binom{b}{n-i} &= \text{number of ways to choose } n-i \text{ from } b \text{ blacks}\\
\binom{r+b}{n} &= \text{number of ways to choose } n \text{ from } r+b \text{ in urn}
\end{aligned}$$

Compare <5.2> with the direct calculation based on a sample space where all possible subsets from the urn are given equal probability.

Unless you subscribe to tricky conventions about factorials or binomial coefficients, you might want to restrict the last calculation to values of i and n for which

$$\begin{aligned}
&0 \le i \le r\\
&0 \le n-i \le b\\
&1 \le n \le r+b
\end{aligned}$$

A random variable that takes on values of i in the range determined by these constraints, with the probabilities expressed by <5.2>, is said to have a HYPERGEOMETRIC distribution.

•hypergeometric

Answer to question (b)

The symmetry property that lets us ignore the ordering when calculating probabilities for particular sequences of draws also lets us eliminate much of the algebra we first used to find $\mathbb{P}R_3$. Reconsider that case. We broke the event R_3 into four disjoint pieces:

$$(R_1R_2R_3) \cup (R_1B_2R_3) \cup (B_1R_2R_3) \cup (B_1B_2R_3)\,.$$

Each triple ends with an R_3, with the first two positions giving all possible R and B combinations. The probability for each triple is unchanged if we permute the subscripts, because ordering does not matter. Thus

$$\mathbb{P}R_3 = \mathbb{P}(R_3R_2R_1) + \mathbb{P}(R_3B_2R_1) + \mathbb{P}(B_3R_2R_1) + \mathbb{P}(B_3B_2R_1)$$

Notice how the triple for each term now ends in an R_1 instead of an R_3. The last sum is just a decomposition for $\mathbb{P}R_1$ obtaining by splitting according to the outcome of the second and third draws. It follows that $\mathbb{P}R_3 = \mathbb{P}R_1$. Similarly,

$$\mathbb{P}\{i\text{th ball is red}\} = \mathbb{P}R_1 = r/(r+b) \qquad \text{for each } i.$$

Answer to question (c)

You should resist the urge to use the answer to question (a) in a direct attack on question (c). Instead, write the number of reds in n draws as $X_1 + \ldots + X_n$, where X_i denotes the indicator of the event R_i, that is,

$$X_i = \begin{cases} 1 & \text{if } i\text{th ball red} \\ 0 & \text{otherwise} \end{cases}$$

From the answer to question (b),

$$\mathbb{E}X_i = 1\mathbb{P}\{X_i = 1\} + 0\mathbb{P}\{X_i = 0\} = \mathbb{P}R_i = r/(r+b).$$

It follows that the expected number of reds in the sample of n is $nr/(r+b)$. This expected number does not depend on k; it is the same for $k = 0$ (sampling with replacement, draws independent) and $k \neq 0$ (draws are dependent), provided we exclude cases where the urn gets emptied out before the nth draw. □

The next Example illustrates a slightly different type of argument, where the symmetry enters conditionally.

<5.3> **Example.** A pack of cards consists of 26 reds and 26 blacks. I shuffle the cards, then deal them out one at a time, face up. You are given the chance to win a big prize by correctly predicting when the next card to be dealt will be red. You are allowed to make the prediction for only one card, and you must predict red, not black. What strategy should you adopt to maximize your probability of winning the prize?

First let us be clear on the rules. Your strategy will predict that card $\tau + 1$ is red, where τ is one of the values $0, 1, \ldots, 51$. That is, you observe the first τ cards then predict that the next one will be red. The value of τ is allowed to depend on the cards you observe. For example, a decision to choose $\tau = 3$ can be based on the observed colors of cards 0, 1, 2, and 3; but it cannot use information about cards $4, 5, \ldots, 52$. (In the probability jargon, τ is called a stopping rule, or stopping time, or several other terms that make sense in other contexts.)

Here are some simple-minded strategies: always choose the first card (probability 1/2 of winning); always choose the last card (probability 1/2 of winning). A more complicated strategy: if the first card is black choose card 2, otherwise choose card 52, which gives

$$\begin{aligned}\mathbb{P}\{\text{win}\} &= \mathbb{P}\{\text{first red, last red}\} + \mathbb{P}\{\text{first black, second red}\} \\ &= \frac{1}{2}\cdot\frac{25}{51} + \frac{1}{2}\cdot\frac{26}{51} \\ &= \frac{1}{2}.\end{aligned}$$

Notice the hidden appeal to (conditional) symmetry to calculate

$$\mathbb{P}\{\text{last red} \mid \text{first red}\} = \mathbb{P}\{\text{second red} \mid \text{first red}\} = \frac{25}{51}.$$

All three stategies give the same probability of a win.

We have to be a bit more cunning. How about: wait until the proportion of reds in the remaining cards is high enough and then go for the next card. As you will soon see, the extra cunning gets us nowhere, because all strategies have the same probability, 1/2, of winning. Amazing!

Consider first an analogous problem for a pack of 3 red and 3 black cards. Why doesn't the following strategy improve one's chances of winning?

> WAIT UNTIL
> NUMBER OF REDS OBSERVED IS $<$ NUMBER OF BLACKS OBSERVED,
> THEN CHOOSE THE NEXT CARD.

With such a small deck we are able to list all possible ways that the cards might appear, calculate τ for each outcome, then calculate the probability of a win. There are $\binom{6}{3} = 20$ possible orderings of 3 reds and 3 blacks, each equally likely. (Here I am treating all red cards as equivalent. You could construct a more detailed sample space, with 6! orderings for the 6 cards, but the calculations would end up with the same conclusion.) With r denoting a red card, and b a black card, the outcomes are:

pattern	**value of τ**	**win?**
b$\underline{\text{b}}$brrr	1	
b$\underline{\text{b}}$rbrr	1	
b$\underline{\text{b}}$rrbr	1	
b$\underline{\text{b}}$rrrb	1	
b$\underline{\text{r}}$bbrr	1	✓
b$\underline{\text{r}}$brbr	1	✓
b$\underline{\text{r}}$brrb	1	✓
b$\underline{\text{r}}$rbbr	1	✓
b$\underline{\text{r}}$rbrb	1	✓
b$\underline{\text{r}}$rrbb	1	✓
rbb$\underline{\text{b}}$rr	3	
rbb$\underline{\text{r}}$br	3	✓
rbb$\underline{\text{r}}$rb	3	✓
rbrbb$\underline{\text{r}}$	5	✓
rbrbrb	?	
rbrrbb	?	
rrbbb$\underline{\text{r}}$	5	✓
rrbbrb	?	
rrbrbb	?	
rrrbbb	?	

Where possible I have underlined the card that the strategy would predict to be red. Even though the game ends after the card is predicted, I have written out the whole string, to make calculation of probabilities a mere matter of counting up equally probable events. Notice that in 5 cases (rbrbrb,…,rrrbbb) the strategy fails to predict. We could modify the strategy by adding

…, BUT IF ONLY ONE CARD REMAINS, CHOOSE IT.

Notice that the addendum has no effect on the probability of a win. There are still only 10 of the 20 equally likely cases that lead to win. The strategy again has probability 1/2 of winning.

The enumeration of outcomes gives a clue to why we keep coming back to 1/2. Look, for example, at the block of ten outcomes beginning b?????. Each of them gives $\tau = 1$. There are only ten possible continuations, each having conditional probability 1/10. The strategy τ has conditional probability 6/10 of leading to a win; six of the ten possible continuations have an r where τ wants it. By symmetry, six of the ten possible continuations have an r in the last position. Thus

$$\mathbb{P}\{\tau \text{ wins} \mid b?????\} = \mathbb{P}\{br???? \mid b?????\} = \mathbb{P}\{b????r \mid b?????\}.$$

It follows that τ has the same conditional probability for a win as the strategy for which $\tau \equiv 5$.

Now try the same idea on the original problem. Consider a string $x_1, x_2, \dots, x_{52}$ of 26 reds and 26 blacks in some order such that a strategy τ would choose card i. The strategy must be using information from only the first i cards. We must have $\tau = i$ for all strings

$$x_1, x_2, \dots, x_i, ?\dots?$$

with the same i cards at the start. Conditioning on this starting fragment, which triggered the choice $\tau = i$, we get

$$\begin{aligned}\mathbb{P}\{\tau \text{ wins} \mid x_1, x_2, \dots, x_i, ?\dots? \} &= \mathbb{P}\{x_1, x_2, \dots, x_i, r, ?\dots? \mid x_1, x_2, \dots, x_i, ?\dots? \} \\ &= \mathbb{P}\{x_1, x_2, \dots, x_i, ?\dots?r \mid x_1, x_2, \dots, x_i, ?\dots? \}.\end{aligned}$$

If we write LAST for the strategy of always choosing the 52nd card, the equality becomes

$$\mathbb{P}\{\tau \text{ wins} \mid x_1, x_2, \dots, x_i, ?\dots? \} = \mathbb{P}\{\text{LAST wins} \mid x_1, x_2, \dots, x_i, ?\dots? \}$$

Multiply both sides by $\mathbb{P}\{x_1, x_2, \dots, x_i, ?\dots?\}$ then sum over all possible starting fragments that trigger a choice for τ to deduce that

$$\mathbb{P}\{\tau \text{ wins}\} = \mathbb{P}\{\text{LAST wins}\} = 1/2.$$

Maybe the LAST strategy is not so simple-minded after all. □

✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠

The strategy of waiting for the the proportion of red cards left in the deck to exceed 1/2, then betting on the next red, works except when the proprtion of reds never gets above 1/2. How likely is that? The answer can be deduced from a result known as the BALLOT THEOREM. According to that Theorem (see the next Example), if a deck contains $n+1$ red cards and n black cards then

•Ballot Theorem

$$\mathbb{P}\{\#\text{reds sampled} > \#\text{blacks sampled, always}\} = \frac{1}{2n+1}.$$

If we condition on the first card being red, then we get

$$\frac{1}{2n+1} = \frac{n+1}{2n+1}\mathbb{P}\{\text{subsequent \#reds} \geq \text{\#blacks} \mid \text{first card red}\},$$

where the conditional probability is the same as the probability, for a deck of n red cards and n black cards, that the number of black cards dealt never strictly exceeds the number of red cards dealt. Solving for that probability, we find that the strategy of waiting for a higher proportion of reds in the deck will fail with probability $1/(n+1)$ for a deck of n red and n black cards. The probability might not seem very large, but apparaently it is just large enough to offset the slight advantage gained when the strategy works.

<5.4> **Example.** Suppose an urn contains r red balls and b black balls, with $r > b$. As balls are sampled without replacement from the urn, keep track of the total number of red balls removed and the total number of black balls removed after each draw. Show that the probability of 'the number of reds removed always strictly exceeds the number of blacks removed' is equal to $(r-b)/(r+b)$.

For simplicity, I will refer to the event whose probability we seek as "red always leads".

The sampling scheme should be understood to imply that all $(r+b)!$ orderings of the balls (treating balls of the same color as distinguishable for the moment) are equally likely.

There is a sneaky way to generate a random permutation, which will lead to an elegant solution to the problem.

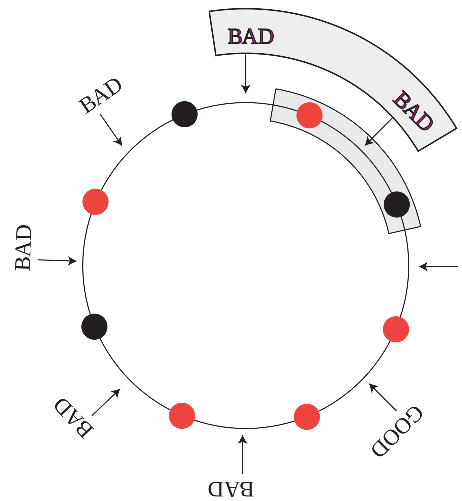
Imagine that the balls are placed into a circular track as they are removed, without any special marker to indicate the position of the first ball. After all the balls are placed in the track, choose a starting position at random, with each of the $r+b$ possible choices equally likely, then select the balls in order moving clockwise from the starting position.

To calculate $\mathbb{P}\{\text{red always leads}\}$, condition on the "circle", the ordering of the balls around the cirular track. I will show that

<5.5> $$\mathbb{P}\{\text{red always leads} \mid \text{circle}\} = \frac{r-b}{r+b},$$

for every circle configuration. Regardless of the probabilities of the various circle configuration, the weighted avarage of these conditional probabilities must give the asserted result.

The calculation of the conditional probability in <5.5> reduces to a simple matter of counting: How many of the $r+b$ possible starting positions generate a "GOOD" permutation where red always leads?



Imagine the $r+b$ positions labelled as GOOD or BAD, as in the picture. Somewhere around the circle there must exist a pair red-black, with the black ball immediately following the red ball in the clockwise ordering.

Two of the positions—the one between the red-black pair, and the one just before the inital red—are obviously bad. (Look at the first few balls in the resulting permutation.)

Consider the effect on the total number (not probability) of GOOD starting positions if the red-black pair is removed from the track. Two BAD starting positions are eliminated immediately. It is less obvious, but true, that removal of the pair has no effect on any of the other starting positions: a GOOD starting position stays GOOD, and a BAD starting position stays BAD. (Consider the effect on the successive red and black counts.) The total number of GOOD starting positions is unchanged.

Repeat the argument with the new circle configuration of $r+b-2$ balls, eliminating one more red-black pair but leaving the total GOOD count unchanged. And so on.

After removal of b red-black pairs all $r-b$ remaining balls are red, and all $r-b$ starting positions are GOOD. Initally, therefore, there must also have been $r-b$ of the GOOD positions out of the $r+b$ available. The assertion <5.5>, and thence the main assertion, follow. $\square$