exactly 5 consecutive heads somewhere in n tosses? There's a nice way to find this number. We simply find the probability of getting at least 5 consecutive blacks, and subtract off the probability of getting at least 6 consecutive blacks.

23.2 General Theory of Recurrence Relations

We've just seen the power of recurrence relations in probability. We used them to analyze the roulette problem, and found that what seemed like a sure-fire method is in fact fatally flawed. As there are many problems where recurrence relations pop-up, it's not a bad idea to know more about them. To help, we've collected some facts about them below.

23.2.1 Notation

Before developing the theory, we first set some notation. We'll study **linear recurrence relations**. A linear recurrence relation of depth k is a sequence of numbers $\{a_n\}_{n=0}^{\infty}$ where

$$a_{n+1} \;=\; c_1 a_n + c_2 a_{n-1} + \cdots + c_k a_{n-k+1} \tag{23.1}$$

for some fixed, given real numbers $c_1, c_2, \ldots, c_k$. If we specify the first k terms of the sequence, all remaining terms are uniquely determined. For example, for the Fibonacci numbers we have $k = 2$, $c_1 = c_2 = 1$, $F_0 = 0$ and $F_1 = 1$. Here the recurrence is $F_{n+1} = F_n + F_{n-1}$. The sequence starts off 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, and so on, where each term (from the third onward) is the sum of the previous two terms.

In some sense, we're done. Once we've specified the recurrence relation and the initial conditions, all subsequent terms are uniquely determined. As this is the case, why should we spend time developing an advanced theory? The main reason is efficiency. We saw in the roulette problem that we might only care about one specific term deep in the sequence; we'd love to be able to jump to it and not have to go through all the previous terms. Related to this, we might be interested in the general behavior of terms in the sequence. Is it possible to say something about their general behavior without computing exactly what they are? For these reasons, there is a real need to find a better approach than just computing term by term.

23.2.2 The Characteristic Equation

Let's see how to find a_n as a function of k, the c_i's and the initial conditions (the values for $a_0, a_1, \ldots, a_{k-1}$). We begin by guessing that $a_n = r^n$ for some constant r; this is the Method of Divine Inspiration we mentioned earlier (we could also use the methods of §19.2 to find the answer via generating functions). It turns out this will always give us a solution to Equation (23.1), though we'll have to do a little work to satisfy the initial conditions.

Plugging $a_n = r^n$ into Equation (23.1) gives

$$r^{n+1} \;=\; c_1 r^n + c_2 r^{n-1} + \cdots + c_k r^{n-k+1}. \tag{23.2}$$

Dividing both sides by r^{n-k+1}, Equation (23.2) becomes

$$r^k \;=\; c_1 r^{k-1} + c_2 r^{k-2} + \cdots + c_k. \tag{23.3}$$

We call Equation (23.3) the **characteristic polynomial** of the difference equation given by Equation (23.1). Subtracting $c_1 r^{k-1} + c_2 r^{k-2} + \cdots + c_k$ from both sides, we can rewrite Equation (23.3) as

$$r^k - c_1 r^{k-1} - c_2 r^{k-2} - \cdots - c_k \;=\; 0. \tag{23.4}$$

Equation (23.4) is a polynomial of degree k, and by the **Fundamental Theorem of Algebra** (see §20.8 for a review of this theorem) has k roots. We call these roots $r_1, r_2, \ldots, r_k$. Note these roots might not be distinct; in fact, if there are repeated roots the analysis is a little harder. For now, we'll assume the roots $r_1, r_2, \ldots, r_k$ are all distinct.

We know $a_n = r_i^n$ is a solution to Equation (23.1) for $1 \le i \le k$; each r_i solves the characteristic polynomial, and we created the characteristic polynomial by simple algebraic manipulation of Equation (23.2). Because we're solving a linear difference equation, once we know that each of $r_1^n, r_2^n, \ldots, r_k^n$ is a solution, we know that a linear combination of these solutions also satisfies Equation (23.1). That is, for constants $\gamma_1, \gamma_2, \ldots, \gamma_k$, we have

$$a_n \;=\; \gamma_1 r_1^n + \gamma_2 r_2^n + \cdots + \gamma_k r_k^n. \tag{23.5}$$

This fact depends on our original recurrence relation being linear. For example, if we had

$$a_{n+1} \;=\; n^2 a_n + e^n a_{n-1},$$

Equation (23.5) would not be valid.

Let's prove this in full gore for the Fibonacci numbers; the proof in general in similar. For the Fibonacci numbers, we get a characteristic equation of $r^2 - r - 1$, with roots $r_1 = (1 + \sqrt{5})/2$ and $r_2 = (1 - \sqrt{5})/2$. Knowing that each of these roots solves the characteristic equation, let's look at an arbitrary linear combination $\gamma_1 r_1^n + \gamma_2 r_2^n$ for F_n. We find

$$\begin{aligned}
&F_{n+1} - F_n - F_{n-1} \\
&= \left(\gamma_1 r_1^{n+1} + \gamma_2 r_2^{n+1}\right) - \left(\gamma_1 r_1^n + \gamma_2 r_2^n\right) - \left(\gamma_1 r_1^{n-1} + \gamma_2 r_2^{n-1}\right) \\
&= \gamma_1 \left(r_1^{n+1} - r_1^n - r_1^{n-1}\right) + \gamma_2 \left(r_2^{n+1} - r_2^n - r_2^{n-1}\right) \\
&= \gamma_1 r_1^{n-1} \left(r_1^2 - r_1 - 1\right) + \gamma_2 r_2^{n-1} \left(r_2^2 - r_2 - 1\right) \;=\; 0 + 0 \;=\; 0.
\end{aligned}$$

What makes the algebra work is the linearity: sums of solutions are solutions, and a multiple of a solution is a solution.

23.2.3 The Initial Conditions

We've made it about two-thirds of the way to finding a solution to Equation (23.1). We have Equation (23.5) as the general form for the a_n's. In addition, we solved the characteristic polynomial for the roots $r_1, r_2, \ldots, r_k$ (which we assume are distinct). Unfortunately, we're not done yet. We still need to determine the values of $\gamma_1, \gamma_2, \ldots, \gamma_k$ in order to find out what a_n is.

Using our initial conditions, which are the values for $a_0, a_1, \ldots, a_{k-1}$, and our assumption that $a_n = \gamma_1 r_1^n + \cdots + \gamma_k r_k^n$, we can set up the following system of equations:

$$\begin{aligned}
\gamma_1 + \gamma_2 + \cdots + \gamma_k &= a_0 \\
\gamma_1 r_1 + \gamma_2 r_2 + \cdots + \gamma_k r_k &= a_1 \\
\gamma_1 r_1^2 + \gamma_2 r_2^2 + \cdots + \gamma_k r_k^2 &= a_2 \\
\vdots &= \vdots \\
\gamma_1 r_1^{k-1} + \gamma_2 r_2^{k-1} + \cdots + \gamma_k r_k^{k-1} &= a_{k-1}.
\end{aligned}$$

From linear algebra, we know that we can rewrite this system of equations as the product of matrices:

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ r_1 & r_2 & \dots & r_k \\ r_1^2 & r_2^2 & \dots & r_k^2 \\ \vdots & & & \vdots \\ r_1^{k-1} & r_2^{k-1} & \dots & r_k^{k-1} \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_k \end{pmatrix} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{k-1} \end{pmatrix}. \tag{23.6}$$

It's a wonderful fact that if $r_1, r_2, \dots, r_k$ are distinct, then our $k \times k$ matrix is invertible. This is a non-trivial fact; for those who are really interested, a proof is given in §23.2.4. In this case, we can solve for the vector of $\gamma_1, \gamma_2, \dots, \gamma_k$ by multiplying both sides of Equation (23.6) to the left by the inverse of the $k \times k$ matrix:

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_k \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ r_1 & r_2 & \dots & r_k \\ r_1^2 & r_2^2 & \dots & r_k^2 \\ \vdots & & & \vdots \\ r_1^{k-1} & r_2^{k-1} & \dots & r_k^{k-1} \end{pmatrix}^{-1} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{k-1} \end{pmatrix}. \tag{23.7}$$

Then, Equation (23.7) gives us values for each of $\gamma_1, \gamma_2, \dots, \gamma_k$. We already solved for $r_1, r_2, \dots, r_k$ and, according to Equation (23.5), this is all the information we need to find a_n. That is, we substitute the r_i values that we found by solving the characteristic polynomial and the γ_i values we find by Equation (23.7) into Equation (23.5) to solve for a_n.

Let's end by applying this to the Fibonacci numbers. Remember $r_1 = (1+\sqrt{5})/2$ and $r_2 = (1-\sqrt{5})/2$, the initial conditions are $F_0 = 0$ and $F_1 = 1$, and $F_n = \gamma_1 r_1^n + \gamma_2 r_2^n$. Our system of equations becomes

$$\begin{pmatrix} 1 & 1 \\ r_1 & r_2 \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

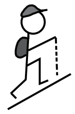
The determinant of the matrix is $r_2 - r_1 = -\sqrt{5}$; as this is non-zero, the matrix is invertible. We find

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \frac{-1}{\sqrt{5}} \begin{pmatrix} r_2 & -1 \\ -r_1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{5} \\ -1/\sqrt{5} \end{pmatrix}.$$

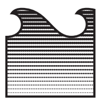
This leads to

$$a_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n,$$

and we recover **Binet's formula**. It's a spectacular formula. It allows us to jump to any Fibonacci number without having to compute the intermediate ones. It makes for very efficient computations.



We leave the rest of the roulette problem as an exercise for the interested reader. The difficulty is that the characteristic polynomial has degree 5, and there is no analogue of the quadratic formula. Sadly, this means we can't just write down the roots in terms of the coefficients of the polynomial, but instead have to approximate them. The five roots are approximately $-0.339175 \pm 0.229268i$, $0.0976883 \pm 0.424427i$, and 0.982974.



In analyzing the solutions to recurrence relations, the large n behavior is typically governed by the root whose absolute value is larger. This is because, as n grows, the powers of this root far exceed the powers of the other roots. The only time when it won't control the limiting behavior is if its corresponding coefficient happens to be zero (which only happen for very special, pathological choices of initial conditions).

23.2.4 Proof that distinct roots imply invertibility

To solve the recurrence relation, we needed the $k \times k$ matrix in Equation (23.7) to be invertible. We need to show that if

$$A = \begin{pmatrix} 1 & 1 & \dots & 1 \\ r_1 & r_2 & \dots & r_k \\ r_1^2 & r_2^2 & \dots & r_k^2 \\ \vdots & & & \vdots \\ r_1^{k-1} & r_2^{k-1} & \dots & r_k^{k-1} \end{pmatrix},$$

then A is invertible if and only if the roots are distinct. This is a very special type of matrix, called a **Vandermonde matrix**, and it turns out that a simple matching argument shows that it's invertible if the roots are distinct.

In linear algebra, you learned (or will learn) that a square matrix is invertible if and only if its determinant is non-zero. If two roots are the same, then two columns are the same and the matrix isn't invertible. We see we can therefore restrict ourselves to the case when all roots are distinct.

From linear algebra (basically expand by minors), we know that $\det(A)$ is a function of $r_1, r_2, \dots, r_k$. In addition, we know that, in calculating a determinant of a $k \times k$ matrix, we have $k!$ summands, with each summand a product of k terms. In the product, we always have exactly one element from each row, and exactly one element from each column. We're going to get a massive polynomial in $r_1, r_2, \dots, r_k$. The first question to ask is: what is its degree? Well, the first row is just all 1's, and thus contributes 0 to the degree. The second row gives us an r_i for some i, and this contributes 1 to the degree. For the third row, we get an r_j^2, which contributes 2 to the degree. And so on and so on until the last row, which gives us a factor like r_ℓ^{k-1}, and adds $k-1$ to the degree. Thus the degree of $\det(A)$ is

$$1 + 2 + \cdots + k - 1 \;=\; \frac{(k-1)k}{2}$$

(see Appendix A.2.1 for a proof of this sum). We know $\det(A)$ is a polynomial involving $r_1, \ldots, r_k$. We're going to show it's just $\prod_{1 \le i < j \le k} (r_j - r_i)$.

For a minute, let's go back and consider what happens if $r_i = r_j$ for some $i \neq j$. If this is the case, then $\det(A) = 0$ as two columns are equal. As i and j are arbitrary, we see $\det(A)$ must always be divisible by $r_i - r_j$, or $\prod_{1 \le i < j \le k} (r_j - r_i)$ divides $\det(A)$.

Now consider the degree of $\prod_{1 \le i < j \le k} (r_j - r_i)$, which we know to be a factor of $\det(A)$. We have $2 \le j \le k$ and $1 \le i \le j - 1$. Therefore, the degree of this polynomial is

$$\sum_{j=2}^{k} (j-1) \;=\; \sum_{j=1}^{k-1} j \;=\; \frac{k(k-1)}{2}.$$

We see that the degree of $\prod_{1 \le i < j \le k} (r_j - r_i)$, which we know to be a factor of $\det(A)$, is the same as the degree of $\det(A)$. This means that

$$\det(A) \;=\; \alpha \cdot \prod_{1 \le i < j \le k} (r_j - r_i) \tag{23.8}$$

for some constant α. Then, we see from Equation (23.8) that $\det(A)$ can be zero only if at least one of α or $\prod_{1 \le i < j \le k} (r_j - r_i)$ is zero. We know that $\prod_{1 \le i < j \le k} (r_j - r_i)$ is zero if $r_i = r_j$, but we've already shown that $\det(A)$ is zero in this case, and we're currently considering the situation in which we have k distinct roots. Thus, we assume that $\prod_{1 \le i < j \le k} (r_j - r_i) \neq 0$, and we must show only that $\alpha \neq 0$. What's really nice is that α is independent of the r_i's, so if we can determine α in one special case, we'll know it in every case.

Let's try $r_i = 10^{10^{i-1}}$. This sequence is growing rapidly. We have $r_1 = 1$, $r_2 = 10^{10}$, $r_3 = 10^{100}$ and so on. Clearly, r_k will be so large that the determinant cannot vanish (the determinant will be essentially $r_1^0 r_2^1 r_3^2 \cdots r_k^{k-1}$), and therefore we cannot have $\alpha = 0$. Consequently, we see that having k distinct roots $r_1, r_2, \ldots, r_k$ is enough to know that A will be invertible.

23.3 Markov Processes

We end with one final example of how difference equations can be applied to probability. We start with a completely deterministic, incredibly oversimplified situation; after we understand this problem, we'll make the model more reasonable and then explore related applications.