Theorem 5.3.1 *Fix an arbitrarily small positive number ϵ. If we flip a coin n times, the probability that the fraction of heads is between $0.5 - \epsilon$ and $0.5 + \epsilon$ tends to 1 as n tends to ∞.*

This theorem says, for example, that flipping a coin n times, the probability that the number of heads is between 49% and 51% is at least 0.99, if n is large enough. But how large must n be for this to hold? If $n = 49$ (which may sound pretty large) the number of heads can *never* be in this range; there are simply no integers between 49% of 49 (24.01) and 51% of 49 (24.99). How much larger does n have to be to assure that the number of heads is in this range for the majority of outcomes? This is an extremely important question in the statistical analysis of data: we want to know whether a deviation from the expected value is statistically significant.

Fortunately, much more precise formulations of the Law of Large Numbers can be made; one of these we can prove relatively easily, based on what we already know about Pascal's triangle. This proof will show that the Law of Large Numbers is not a mysterious force, but a simple consequence of the properties of binomial coefficients.

Theorem 5.3.2 *Let $0 \le t \le m$. Then the probability that out of $2m$ coin tosses, the number of heads is less than $m - t$ or larger than $m + t$, is at most $e^{-t^2/(m+t)}$.*

To illustrate the power of this theorem, let's go back to our earlier question: *How large should n be in order that the probability that the number of heads is between 49% and 51% is at least 0.99?* We want $m - t$ to be 49% of $n = 2m$, which means that $t = m/50$. The theorem says that the probability that the number of heads is not in this interval is at most $e^{-t^2/(m+t)}$. The exponent here is

$$-\frac{t^2}{m+t} = -\frac{\left(\frac{m}{50}\right)^2}{m + \frac{m}{50}} = -\frac{m}{2550}.$$

We want $e^{-m/2550} < 0.01$; taking the logarithm and solving for m, we get $m \ge 11744$ suffices. (This is pretty large, but, after all, we are talking about the "Law of Large Numbers.")

Observe that m is in the exponent, so that if m increases, the probability that the number of heads is outside the given interval drops very fast. For example, if $m = 1{,}000{,}000$, then this probability is less than 10^{-170}. Most likely, over the lifetime of the universe it never happens that out of a million coin tosses less than 49% or more than 51% are heads.

Normally, we don't need such a degree of certainty. Suppose that we want to make a claim about the number of heads with 95% certainty, but we would like to narrow the interval into which it falls as much as possible. In other words, we want to choose the smallest possible t so that the probability that the number of heads is less than $m-t$ or larger than $m+t$ less than 0.05. By Theorem 5.3.2, this will be the case if

$$e^{-t^2/(m+t)} < 0.05.$$

(This is only a sufficient condition; if this holds, then the number of heads will be between $m-t$ and $m+t$ with probability at least 0.95. Using more refined formulas, we would find a slightly smaller t that works.) Taking the logarithm, we get

$$-\frac{t^2}{m+t} < -2.996.$$

This leads to a quadratic inequality, which we could solve for t; but it should suffice for this discussion that $t = 2\sqrt{m}+2$ satisfies it (which is easy to check). So we get an interesting special case:

> *With probability at least* 0.95*, the number of heads among* $2m$ *coin tosses is between* $m-2\sqrt{m}-2$ *and* $m+2\sqrt{m}+2$.

If m is very large, then $2\sqrt{m}+2$ is much smaller than m, so we get that the number of heads is very close to m. For example, if $m = 1{,}000{,}000$ then $2\sqrt{m} = 2{,}002 \approx 0.002m$, and so it follows that with probability at least 0.95, the number of heads is within $\frac{1}{5}$ of a percent of $m = n/2$.

It is time now to turn to the proof of Theorem 5.3.2.

Proof. Let A_k denote the event that we toss exactly k heads. It is clear that the events A_k are mutually exclusive. It is also clear that for every outcome of the experiment, exactly one of the A_k occurs.

The number of outcomes for which A_k occurs is the number of sequences of length n consisting of k heads and $n-k$ tails. If we specify which of the n positions are heads, we are done. This can be done in $\binom{n}{k}$ ways, so the set A_k has $\binom{n}{k}$ elements. Since the total number of outcomes is 2^n, we get the following:

$$\mathsf{P}(A_k) = \frac{\binom{n}{k}}{2^n}.$$

What is the probability that the number of heads is far from the expected, which is $m = n/2$; say, it is less than $m-t$ or larger than $m+t$, where t is any positive integer not larger than m? Using Exercise 5.1.4, we see that the probability that this happens is

$$\frac{1}{2^{2m}}\left(\binom{2m}{0}+\binom{2m}{1}+\cdots+\binom{2m}{m-t-1}+\binom{2m}{m+t+1}+\cdots\right.$$
$$\left.+\binom{2m}{2m-1}+\binom{2m}{2m}\right).$$

Now we can use Lemma 3.8.2, with $k = m-t$, and get that

$$\binom{2m}{0}+\binom{2m}{1}+\cdots+\binom{2m}{m-t-1} < 2^{2m-1}\binom{2m}{m-t}\bigg/\binom{2m}{m}.$$

By (3.9), this can be bounded from above by

$$2^{2m-1}e^{-t^2/(m+t)}.$$

By the symmetry of Pascal's triangle, we also have

$$\binom{2m}{m+t+1} + \cdots + \binom{2m}{2m-1} + \binom{2m}{2m} < 2^{2m}e^{-t^2/(m+t)}.$$

Hence we get that the probability that we toss either fewer than $m-t$ or more than $m+t$ heads is less than $e^{-t^2/(m+t)}$. This proves the theorem. □

5.4 The Law of Small Numbers and the Law of Very Large Numbers

There are two further statistical "laws" (half serious): the *Law of Small Numbers* and the *Law of Very Large Numbers.*

The first one says that if you look at small examples, you can find many strange or interesting patterns that do not generalize to larger numbers. Small numbers exhibit only a small number of patterns, and looking at various properties of small numbers, we are bound to see coincidences. For example, "every odd number is a prime" is true for 3, 5 and 7 (and one may be tempted to say that it is also true for 1, which is even "simpler" than primes: instead of two divisors, it has only one). Of course, this fails for 9.

Primes are strange (as we'll see) and in their irregular sequence, many strange patterns can be observed, which than fail if we move on to larger numbers. A dramatic example is the formula n^2-n+41. This gives a prime for $n = 0, 1, \ldots, 40$, but for $n = 41$ we get $41^2 - 41 + 41 = 41^2$, which is not a prime.

Fibonacci numbers are not as strange as primes: We have seen many interesting properties of them, and derived an explicit formula in Chapter 4. Still, one can make observations for the beginning of the sequence that do not remain valid if we check them far enough. For example, Exercise 4.3.4 gave a (false) formula for the Fibonacci numbers, namely $\lceil e^{n/2-1} \rceil$, which was correct for the first 10 positive integers n. There are many formulas that give integer sequences, but these sequences can start only so many ways: we are bound to find different sequences that start out the same way.

So the moral of the "Law of Small Numbers" is that to make a mathematical statement, or even to set up a mathematical conjecture, it is not enough to observe some pattern or rule, because you can only observe small instances and there are many coincidences for these. There is nothing wrong with making conjectures in mathematics, generalizing facts observed in special cases, but even a conjecture needs some other justification (an imprecise