Chapter 1

Experiencing Statistical Regularity

1.1. A Simple Game of Chance

A good way to experience statistical regularity is to repeatedly play a game of chance. So let us consider a simple game of chance using a spinner. To attract attention, it helps to have interesting outcomes, such as falling into an alligator pit or winning a dash for cash (e.g., you receive the opportunity to run into a bank vault and drag out as many money bags as you can within thirty seconds). However, to focus on statistical regularity, rather than fear or greed, we consider repeated plays with a simple outcome.

In our game, the payoff in each of several repeated plays is determined by spinning the spinner. We pay a fee for each play of the game and then receive the payoff indicated by the spinner. Let the payoff on the spinner be uniformly distributed around the circle; i.e., if the angle after the spin is θ, then we receive $\theta/2\pi$ dollars. Thus our payoff on one play is U dollars, where U is a uniform random number taking values in the interval $[0, 1]$.

We have yet to specify the fee to play the game, but first let us simulate the game to see what cumulative payoffs we might receive, not counting the fees, if we play the game repeatedly. We perform the simulation using our favorite random number generator, by generating n uniform random numbers $U_1, \ldots, U_n$, each taking values in the interval $[0, 1]$, and then forming

associated partial sums by setting

$$S_k \equiv U_1 + \cdots + U_k, \quad 1 \le k \le n \ ,$$

and $S_0 \equiv 0$, where $\equiv$ denotes equality by definition. The n^{th} partial sum S_n is the total payoff after n plays of the game (not counting the fees to play the game). The successive partial sums form a *random walk*, with U_n being the n^{th} step and S_n being the position after n steps.

1.1.1. Plotting Random Walks

Now, using our favorite plotting routine, let us plot the random walk, i.e., the $n+1$ partial sums S_k, $0 \le k \le n$, for a range of n values, e.g., for $n = 10^j$ for several values of j. This simulation experiment is very easy to perform. For example, it can be performed almost instantaneously with the statistical package *S* (or *S-Plus*), see Becker, Chambers and Wilks (1988) or Venables and Ripley (1994), using the function

```
walk <- function(j) {
uniforms <- runif(10^j)            # generate random numbers
firstsums <- cumsum(uniforms)      # form the partial sums
sums <- c(0, firstsums)            # include a 0^th sum
index <- order(sums) -1            # adjust the index
plot(index, sums) }                # do the plotting
```

Plots of the random walk with $n = 10^j$ for $j = 1, \ldots, 4$ are shown in Figure 1.1. For small n, e.g., for $n = 10$, we see irregularly spaced (vertically) points increasing to the right, but as n increases, the spacing between the points becomes blurred and regularity emerges: The plots approach a straight line with slope equal to $1/2$, the mean of a single step U_k. If we look at the pictures in successive plots, ignoring the units on the axes, we see that the plots become independent of n as n increases. Looking at the plot for large n produces a macroscopic view of uncertainty.

The plotter automatically plots the random walk $\{S_k : 0 \le k \le n\}$ in the available space. Ignoring the units on the axes is equivalent to regarding the plot as a display in the unit square. By "unit square" we do not mean that the rectangle containing the plot is necessarily a square, but that new units can range from 0 to 1 on both axes, independent of the original units. The plotter automatically plots the random walk in the available space by

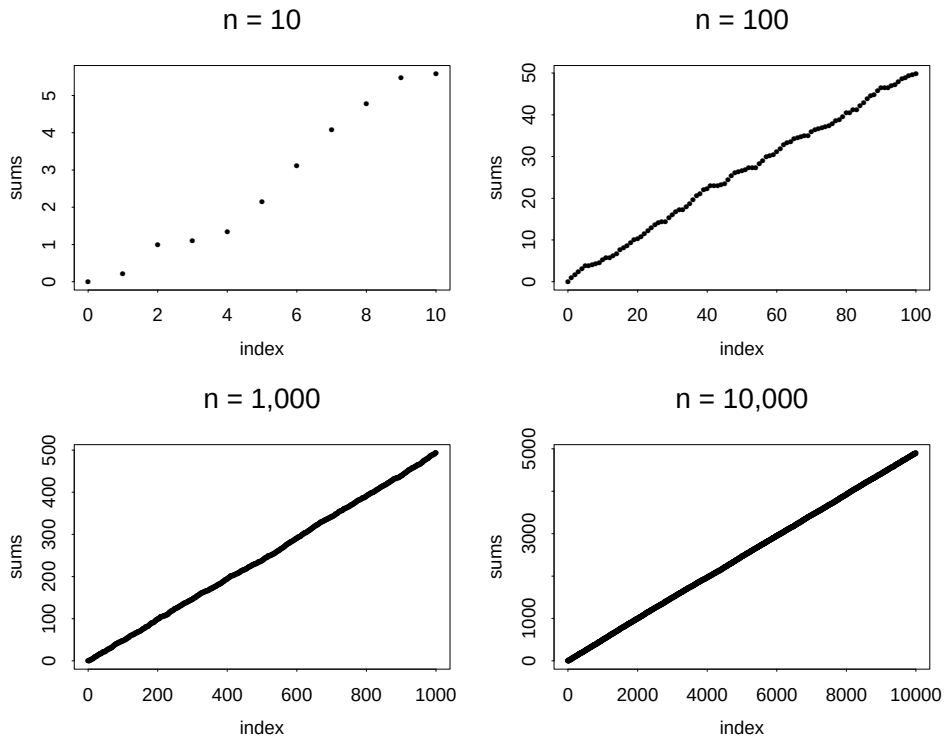

Figure 1.1: Possible realizations of the first 10^j steps of the random walk $\{S_k : k \geq 0\}$ with steps uniformly distributed in the interval $[0, 1]$ for $j = 1, \ldots, 4$.

scaling time and space (the horizontal and vertical dimensions). Time is scaled by placing the $n+1$ points $1/n$ apart horizontally. Space is scaled by subtracting the minimum and dividing by the range (assuming that the range is not zero); i.e., we interpret the plot as

$$plot(\{S_k : 0 \le k \le n\}) \equiv plot(\{(S_k - min)/range : 0 \le k \le n\}) \ ,$$

where

$$min \equiv min(\{S_k : 0 \le k \le n\})$$

and

$$range \equiv max(\{S_k : 0 \le k \le n\}) - min(\{S_k : 0 \le k \le n\}) \ .$$

Combining these two forms of scaling, the plotter displays the ordered pairs $(k/n, (S_k - min)/range)$ for $0 \le k \le n$. With that scaling, the ordered pairs do indeed fall in the unit square. Also note that $(S_k - min)/range$ must assume (approximately) the values 0 and 1 for at least one argument. That occurs because, without the rescaling, the plotting makes the units on the ordinate (y axis) range from the minimum value to the maximum value (approximately).

To confirm the regularity we see in Figure 1.1, we should repeat the experiment. When we repeat the experiment with different random number seeds (new uniform random numbers), the outcome for small n changes somewhat from experiment to experiment, but we always see essentially the same picture for large n. Thus the plots show regularity associated with both large n and repeated experiments.

1.1.2. When the Game is Fair

Now let us see what happens when the game is fair. Since the expected payoff is 1/2 dollar each play of the game, the game is fair if the fee to play is 1/2 dollar. To examine the consequences of making the game fair, we consider a minor modification of the simulation experiment above: We repeat the experiment after subtracting the mean 1/2 from each step of the random walk; i.e., we plot the *centered random walk* (i.e., the centered partial sums $S_k - k/2$ for $0 \le k \le n$) for the same values of n as before.

If we consider the case $n = 10^4$, it is natural to expect to see a horizontal line instead of the line with slope 1/2 in Figure 1.1. However, what we see is very different! Instead of a horizontal line, for $n = 10^4$ we see an irregular path, as shown in Figure 1.2.

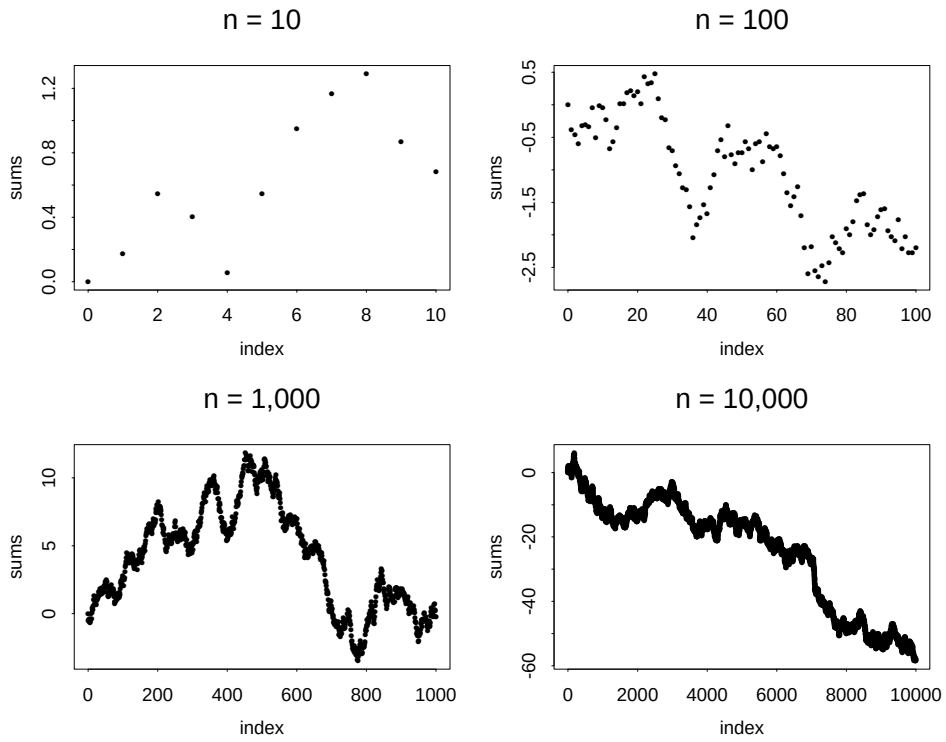

Figure 1.2: Possible realizations of the first 10^j steps of the centered random walk $\{S_k - k/2 : k \geq 0\}$ with steps uniformly distributed in the interval $[0, 1]$ for $j = 1, \ldots, 4$.

We do not see the horizontal line because *the data have been automatically rescaled by the plotter*. The centering has let the plotter *blow up the picture* to show extra detail not apparent from Figure 1.1.

After centering, the range of values (the maximum minus the minimum) for the partial sums decreases dramatically. The first 10^4 uncentered partial sums assume values approximately in the interval $[0, 5000]$, whereas the first 10^4 centered partial sums all fall in the interval $[-60, 5]$. Thus, the range has decreased from $5,000$ to less than 100.

At first glance, it may not be evident that there is any regularity for large n in Figure 1.2. We would hope to be able to predict what we will see if we repeat the experiment with new uniform random numbers. However, when we repeat the simulation experiment with different random number seeds, we obtain different irregular paths. To illustrate, six independent plots for $n = 10^4$ are shown in Figure 1.3. The six path samples look somewhat similar, but each is different from the others.

In Figure 1.3, just as in Figures 1.1 and 1.2, we let the plotter automatically do the scaling. Thus, the units on vertical axis change from plot to plot. We plot in this manner throughout this chapter, by design. We will show that these "automatic plots" reveal statistical regularity if we ignore the units and think of the plot as being on the unit square. But essentially the same conclusion can be drawn if we fix the units on the vertical axis. From Figure 1.3, after the fact, we can conclude that we could have fixed the units on the vertical axis, letting the values fall in the interval $[-100, 100]$. In either case, we are faced with the problem of understanding what we see.

We have arrived at a critical point, which may require us to adjust our thinking. To understand what we are seeing, we need to recognize that the irregular paths we see should be regarded as *random paths*. We then can understand that there actually is regularity underlying the six displayed paths in Figure 1.3, but it is *statistical regularity*.

We want to be able to predict what we will see when we increase n or perform additional experiments. For the uncentered random walks in Figure 1.1, we predict that the plot of $\{S_k : 0 \leq k \leq n\}$ will look like the diagonal line in the unit square for all n sufficiently large. However, for the centered random walks, the plots do not approach such a simple limit. What we should hope to predict when we repeat the experiment for the centered random walk (again ignoring the units on the axes) is the *probability distribution* of the random path. We should anticipate that the successive paths in repeated experiments will change from experiment to experiment, but we should look for a common probability distribution on the space of possible paths.

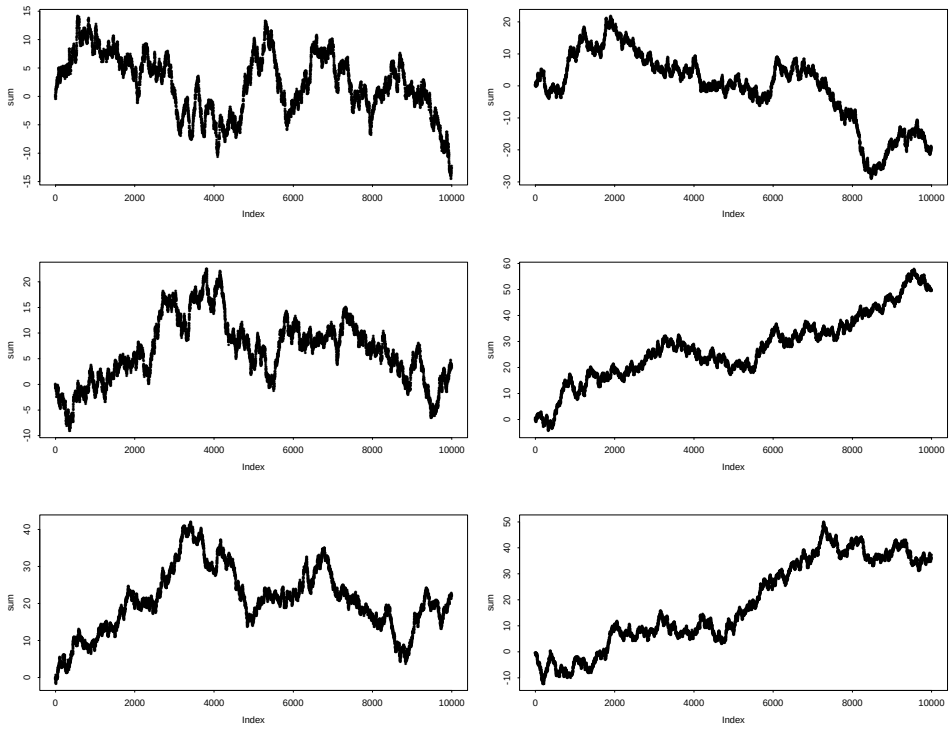

Figure 1.3: Six independent realizations of the first 10^4 steps of the centered random walk $\{S_k - k/2 : k \geq 0\}$ associated with steps uniformly distributed in the interval $[0, 1]$.

The simulation experiments suggest that, for all n sufficiently large, there tends to be a common probability distribution for the plotted random walk paths, where as before we ignore the units on the two axes or, equivalently, we regard the plot as being in the unit square. We can see part of the story when we generate new random walk paths for different values of n. For example, when we generate six centered random walk paths for $n = 10^5$ or $n = 10^6$, the plots look just like the plots in Figure 1.3. To make that clear, we plot six independent plots for the case $n = 10^6$ in Figure 1.4. As before, the units on the vertical axes change from plot to plot, but if we ignore the units on both axes, the plots in Figure 1.4 look just like the plots in Figure 1.3.

Looking at Figures 1.3 and 1.4, we should be confident about what we will see when $n = 10^8$ or $n = 10^{10}$. From Figure 1.4 and other similar plots, we see that, for n sufficiently large, the plots tend to be independent of n, provided that we ignore the units on the axes, and regard the plot as being in the unit square. Of course, as n increases, the units change on the two axes. And each new plot is a random path selected from the common probability distribution on the space of possible sample paths in the unit square.

As a consequence, we also see that the fluctuations in a smaller time scale are asymptotically negligible compared to the fluctuations in a larger time scale. Thus, for $j \geq 5$, the plots for 10^j are visually unchanged if we only keep the values at about 10^4 equally spaced indices. Indeed, such pruning of the data (reducing a data set of 10^j partial sums for $j \geq 5$ to 10^4 values) is useful to efficiently print the plots for large n.

The fact that the plots are independent of n for all n sufficiently large means that the plots tend to exhibit *self-similarity*. By self-similarity we mean that rescaled versions of the plot associated with increasing n tend to look like the original plot. More specifically, the probability distribution on the space of sample paths in the unit square tends to be unaffected by the scaling. Self-similarity will be a persistent theme; e.g., see Section 4.2.

When we consider rescaling, we can also decrease n. For instance, suppose that we consider the plot for $n = 10^7$ and select 10% of it from a subinterval of the plot. If we make a full plot of that 10% portion, then we obtain a plot for $n = 10^6$, which looks just like a random version of the original plot for $n = 10^7$. (By a "random version of the original plot" we mean that the probability distributions on the space of possible sample paths in the unit square tend to be the same.) Similarly, if we continue and select 10% of the new plot for $n = 10^6$ from any subinterval and plot it, then we obtain a plot for $n = 10^5$, which again looks like a random version of the

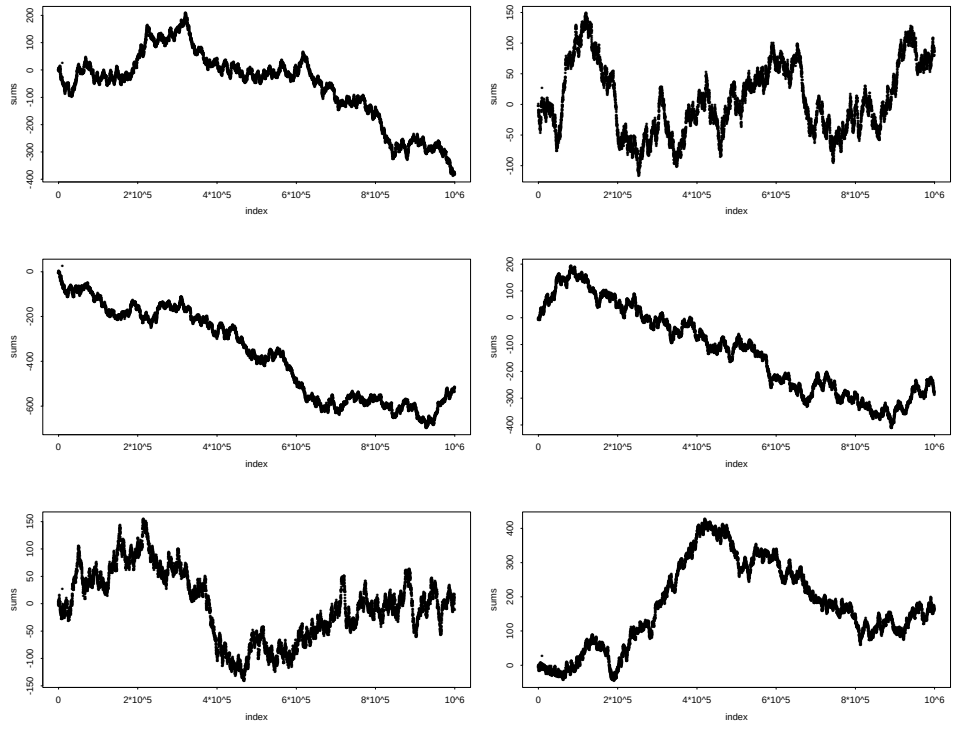

Figure 1.4: Six independent realizations of the first 10^6 steps of the centered random walk $\{S_k - k/2 : k \geq 0\}$ associated with steps uniformly distributed in the interval $[0, 1]$.

original plot for 10^7. Of course, Figure 1.2 shows that the self-similarity for the random walks associated with decreasing n breaks down when n is too small. It is interesting to contemplate a limiting continuous-time random path that permits self-similarity without end!

1.1.3. The Final Position

It is difficult to actually see the probability distribution of the entire random path, because the path is multidimensional, but we can easily look at any one position of the random walk. For instance, suppose that we focus on the final position of the centered random walk, i.e., the single centered partial sum $S_n - n/2$ for one fixed (large) value of n.

It is evident that the final position of the centered random walk, $S_n - n/2$, changes from experiment to experiment. We find statistical regularity when we perform many independent replications of the experiment and look at the distribution of the final positions. So, let us do that.

Remark 1.1.1. *The final position and the relative final position.* For simplicity, we now want to look at the final position of the centered random walk, $S_n - n/2$, independent of the rest of the random walk. If instead we looked at the final position in the unit square, ignoring the original units, we would be looking at the *relative final position*, which must assume a value between 0 and 1. Letting $M_n \equiv \max_{1 \le k \le n}\{S_k - k/2\}$ and $m_n \equiv \min_{1 \le k \le n}\{S_k - k/2\}$, the relative final position is

$$R_n \equiv \frac{S_n - n/2 - m_n}{M_n - m_n}, \quad n \ge 1 \ . \tag{1.1}$$

It turns out that there is statistical regularity associated with the relative final position, just as there is statistical regularity associated with the entire plot, but the relative final position is more complicated than the final position. Hence, now we focus on the final position. We discuss the relative final position in Remark 1.2.2 at the end of Section 1.2.4. ∎

Suppose that we consider the final position of the centered random walk with uniform random steps for $n = 1000$, and suppose that we perform 1000 replications of the experiment. We thus obtain 1000 independent samples of the centered sum $S_{1000} - 500$. We can estimate the probability density of this distribution using the nonparametric probability density estimator *density* from *S* (with the default parameter settings). The estimated probability density of the final position $S_{1000} - 500$ is plotted in Figure 1.5.

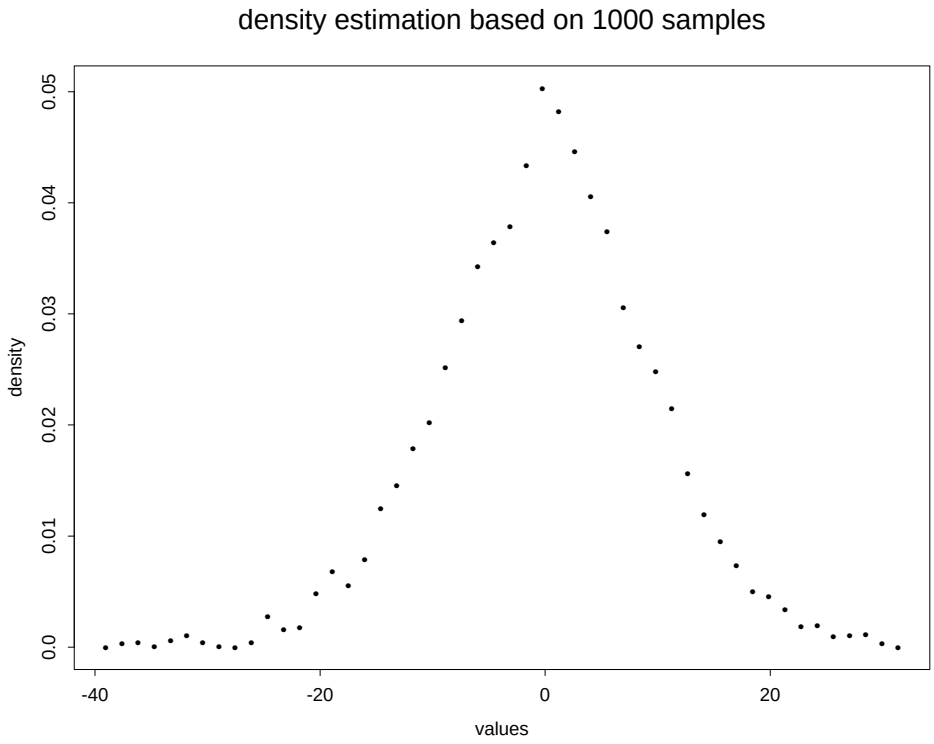

Figure 1.5: An estimate of the probability density of the final position of the random walk, obtained from 1000 independent samples of the centered partial sum $S_{1000} - 500$, where the steps U_k are uniformly distributed in the interval $[0, 1]$.

Figure 1.5 shows that nonparametric density estimation does not achieve high resolution with only a modest amount of data, but it suggests that the final position of the random walk after 1000 steps is approximately normally distributed with zero mean. That conclusion is more strongly supported by the QQ plot in Figure 1.6. The QQ plot compares the empirical distribution of the data to the normal distribution; e.g., see p. 122 of Venables and Ripley (1994). Specifically, the QQ plot compares the sorted data to the quantiles of the normal distribution. If there are n data points, then we consider the $n-1$ normal quantiles z_k, where

$$P(N(0,1) \leq z_k) = k/n, \quad 1 \leq k \leq n-1 \ ,$$

with $N(m, \sigma^2)$ denoting a random variable with a normal (or Gaussian) distribution having mean m and variance σ^2. When $n = 1,000$, the normal quantiles range from -3.1 to $+3.1$, with there being more quantiles near 0 than at the extremes. (Since we focus on the shape of the QQ plot, the QQ plot compares the distributions independent of location and scale; e.g., the shape of the QQ plot is independent of the mean and variance of the reference normal distribution.)

The near-linear plot in Figure 1.6 is approximately the same as the QQ plot for 1000 independent samples from a normal distribution. To make that clear, a QQ plot of a sample of 1000 observations from a normal distribution (with the same mean and variance) is also shown in Figure 1.6. Again the units are different in the two plots, because the range of values differs from sample to sample. The linearity that holds except for the tails strongly indicates that the final positions are indeed normally distributed.

But, in order to fairly draw that conclusion, we need more experience with QQ plots. We become more confident of the conclusion when we repeat these experiments a number of times; then we can observe the statistical variability in the QQ plots. We also gain confidence when we make QQ plots of various non-normal distributions; then we can see how departures from normality are reflected in the plots. When you think hard about the figures, they become invitations to perform additional experiments. Our main point here is that analysis with the QQ plots indicates that the final position of the centered random walk is indeed approximately normally distributed.

That conclusion is also supported by density estimates based on more data. To illustrate how the density estimates perform as a function of sample size, we display the estimates of the probability density of the same final position $S_{1000} - 500$ based on 10^j samples for $j = 2, \ldots, 5$ in Figure 1.7 (again using the nonparametric density estimator *density* from S with the default

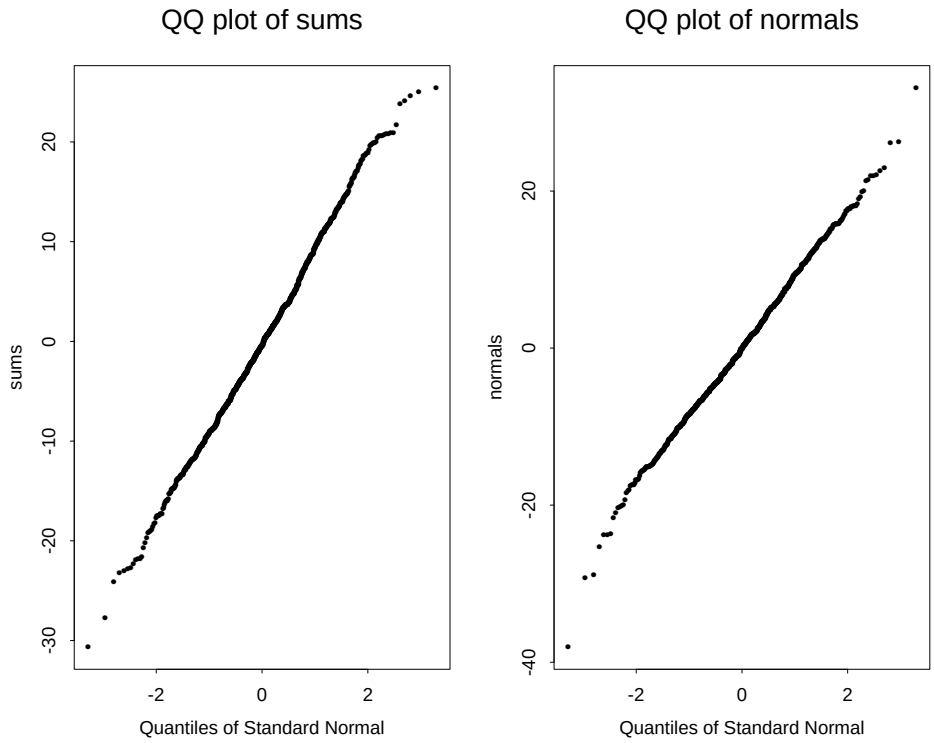

Figure 1.6: Two QQ plots of 1000 samples: the first for the sums, i.e., the final positions $S_{1000} - 500$ of the centered random walk, and the second for a normal distribution.

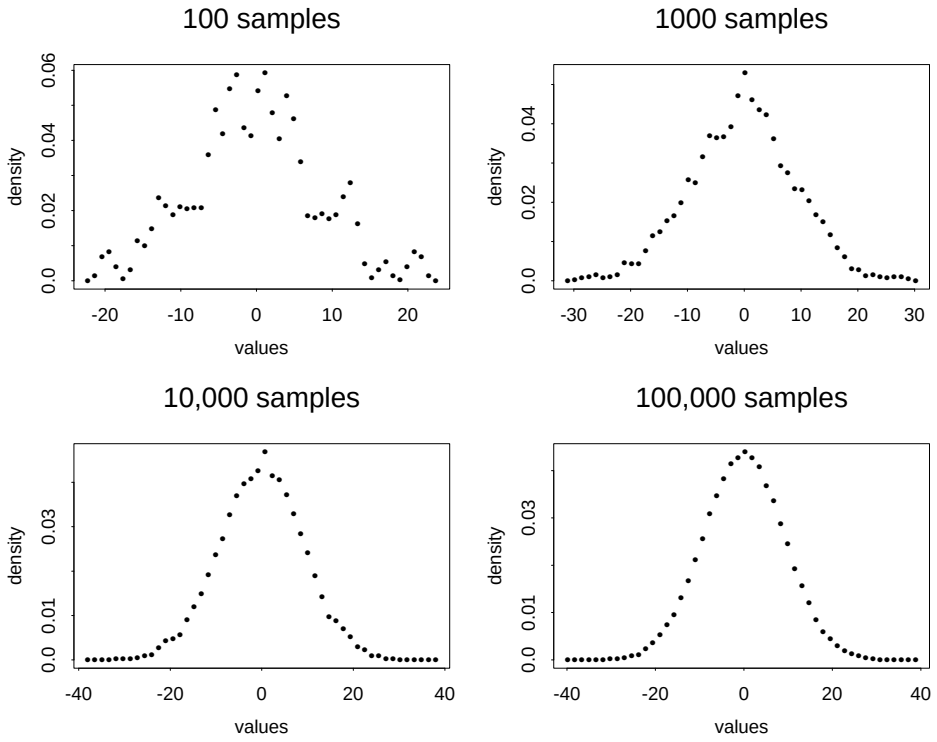

Figure 1.7: Estimates of the probability density of the final position of the random walk, obtained from 10^j independent samples of the centered partial sum $S_{1000} - 500$ for $j = 2, \ldots, 5$, for the case in which the steps U_k are uniformly distributed in the interval $[0, 1]$, based on the nonparametric density estimator *density* from *S*.

parameter settings). Essentially the same plots are obtained for independent samples from normal distributions. From Figure 1.7, it is evident that the density estimates converge to a normal pdf as $n \to \infty$. For more on density estimation, see Devroye (1987).

It is not our purpose to delve deeply into statistical issues, but it is worth remarking that we obtain new interesting plots, like the random walk plots, when we do. Our brief examination of the distribution of the final position of the random walk suggests looking for a more precise statistical test to determine whether or not the final position of the random walk is indeed approximately normally distributed. To evaluate whether some data can be regarded as an independent sample any specified probability distribution, it

is natural to carefully investigate how the empirical distribution of a sample from that probability distribution tends to differ from the underlying probability distribution itself.

Recall that the *cumulative distribution function* (cdf) F of a random variable X is the function

$$F(t) \equiv P(X \leq t) \quad \text{for} \quad t \in \mathbb{R} \ .$$

Similarly, the *empirical cdf* of a data set of size n is the proportion $F_n(t)$ of the n data points that are less than or equal to t, as a function of t.

The idea, then, is to look at the *difference* between a cdf and the empirical cdf obtained from an independent sample from that cdf. Moreover, it is natural to consider how that difference behaves as the sample size increases. Once we have made such a study, we can use the established behavior of samples from the specified probability distribution to *test* whether or not data from an unknown source can reasonably be regarded as a sample from the candidate probability distribution.

Example 1.1.1. *The empirical cdf of uniform random numbers.* To illustrate, we now consider the difference between the empirical cdf associated with n uniform random numbers on the interval $[0, 1]$ and the uniform cdf itself. Since the uniform cdf is $F(t) = t, 0 \leq t \leq 1$, we now want to plot $F_n(t) - t$ versus t for $0 \leq t \leq 1$. Since the function $F_n(t) - t, 0 \leq t \leq 1$, is a function of a continuous variable, the plotting is less routine than for the random walk. However, the empirical cdf F_n has special structure, making it possible to do the plotting quite easily. In particular, to do the plotting, let $U_k^{(n)}, 1 \leq k \leq n$, be the *order statistics* associated with the uniform random numbers $U_1, \ldots, U_n$, i.e., $U_k^{(n)}$ is the k^{th} smallest of the uniform random numbers. Note that

$$F_n(U_k^{(n)}) = k/n \quad \text{and} \quad F_n(U_k^{(n)}-) = (k-1)/n \ ,$$

$F_n(0) = 0$ and $F_n(1) = 1$, where $F_n(t-)$ is the left limit of the function F_n at t. Thus we can plot $F_n(t) - t$ versus t by plotting the points $(0, 0)$, $(1, 0)$, $(U_k^{(n)}, (k-1)/n - U_k^{(n)})$ and $(U_k^{(n)}, k/n - U_k^{(n)})$, $1 \leq k \leq n$, and connecting the points by lines (i.e., performing linear interpolation).

Plots for $n = 10^j$ for $j = 1, \ldots, 4$ are shown in Figure 1.8. The plots in Figure 1.8 look much like the plots of the uncentered random walks, but there is a subtle difference that can be confirmed by further replications of the experiment. Unlike before, here the final position is 0 just like the initial

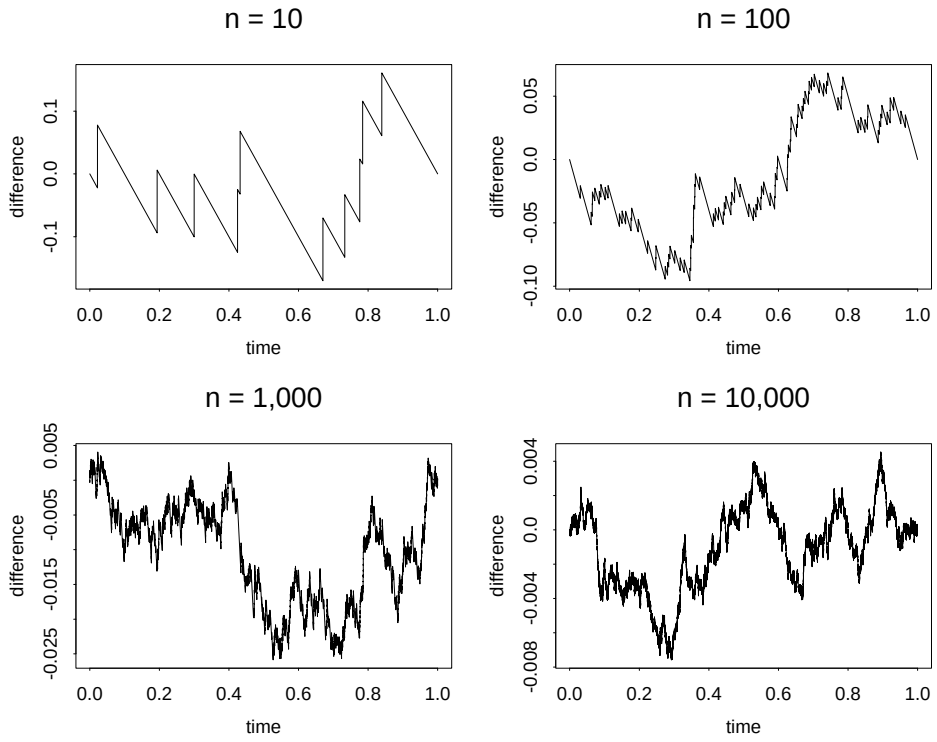

Figure 1.8: The difference between the empirical cdf and the actual cdf for samples of size 10^j from the uniform distribution over the interval $[0, 1]$ for $j = 1, \ldots, 4$.

position. That makes sense as well, because both the empirical cdf and the actual cdf must assume the common value 1 at the right endpoint.

It turns out that there is statistical regularity in the empirical cdf's just like there is in the random walks. As before, the plots look the same for all sufficiently large n. Moreover, except for having the final position be 0, the plots look just like the random-walk plots. More generally, this example illustrates that statistical analysis is an important source of motivation for stochastic-process limits. We discuss this example further in Section 2.2. There we show how to develop a statistical test applicable to any continuous cdf, including the normal cdf that is of interest for the final position of the random walk.

1.1.4. Making an Interesting Game

We have digressed from our original game of chance to consider the statistical regularity observed in the plots, which of course really is our main interest. But now let us return for a moment to the game of chance.

A gambling house cannot afford to make the game fair. The gambling house needs to charge a fee greater than the expected payoff in order to make a profit. What would be a good fee for the gambling house to charge?

From the perspective of the gambling house, one might think the larger the fee the better, but the players presumably have the choice of whether or not to play. If the gambling house charges too much, few players will want to play. The fee should be large enough for the gambling house to make money, but small enough so that potential players will want to play. We take that to mean that the individual players should have a good chance of winning.

One might think that those objectives are inconsistent, but they are not. The key to achieving those objectives is the realization that *the player and the gambling house experience the game in different time scales.* An individual player might contemplate playing the game 100 times on a single day, while the gambling house might offer the game to hundreds or thousands of players on each of many consecutive days.

Thus, the player might evaluate his experience by the possible outcomes from about 100 plays of the game, while the gambling house might evaluate its experience by the possible outcomes from something like $10^4 - 10^6$ plays of the game. What we need, then, is a fee close enough to \$0.50 that the player has a good chance of winning in 100 plays, while the gambling house receives a good reliable return over $10^4 - 10^6$ games.

A reasonable fee might be \$0.51, giving the gambling house a 1 cent or 2% advantage on each play. (Gambling houses actually tend to take more, which shows the appeal of gambling despite the odds.) To see how the \$0.51 fee works, let us consider the possible experiences of the player and the gambling house. In Figure 1.9 we plot six independent realizations of a player's position during 100 plays of the game when there is a fee of \$0.51 for each play. The game looks pretty interesting for the player from Figure 1.9. The player has a reasonable chance of winning. Indeed, the player wins in plots 3 and 5, and finishes about even in plot 2. How do things look for the gambling house?

To see how the gambling house fares, we should look at the net payoffs over a much larger number of games. Hence, in Figures 1.10 and 1.11 we plot

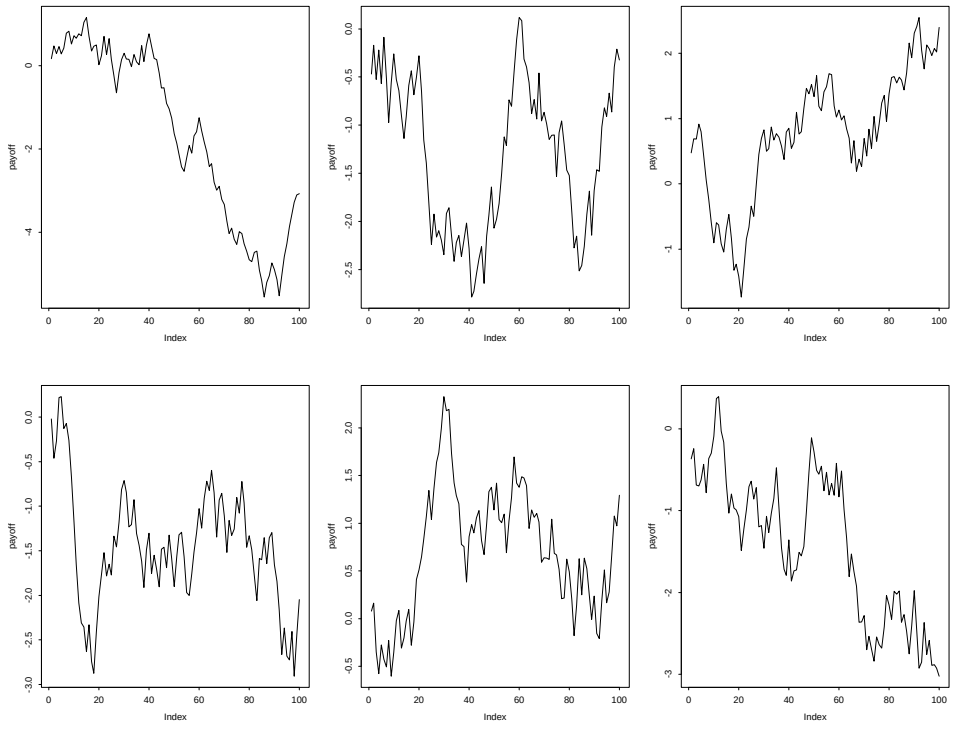

Figure 1.9: Six possible realizations of the first 100 net payoffs, positions of the random walk $\{S_k - 0.51k : k \geq 0\}$, with steps U_k uniformly distributed in the interval $[0, 1]$ and a fee of \$0.51.

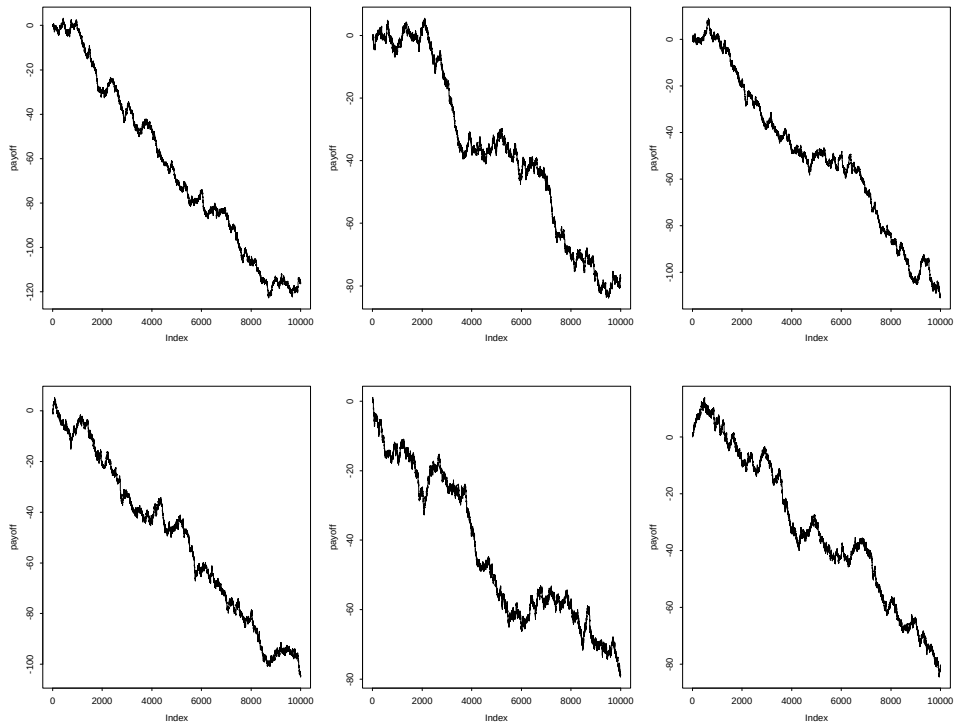

Figure 1.10: Possible realizations of the first 10^4 net payoffs (steps of the random walk $\{S_k - 0.51k : k \geq 0\}$ with steps U_k uniformly distributed in the interval $[0, 1]$.

six independent realizations of a player's position during 10^4 and 10^6 plays of the game. As before, we let the plotter automatically do the scaling, so that the units on the vertical axes change from plot to plot. But that does not alter the conclusions. In these larger time scales, we see that the player consistently loses money, so that a profit for the gambling house becomes essentially a sure thing. When we increase the number of plays to 10^6, there is little randomness left. That is shown in Figure 1.11. Further repetitions of the experiment confirm these observations. We again see the regularity associated with a macroscopic view of uncertainty.

Above we picked a candidate fee out of the air. We could instead be more systematic. For example, we might seek the largest fee such that the player satisfies some criteria indicating a good experience. Letting the fee for each game be f, we might want to constrain the probability p that a

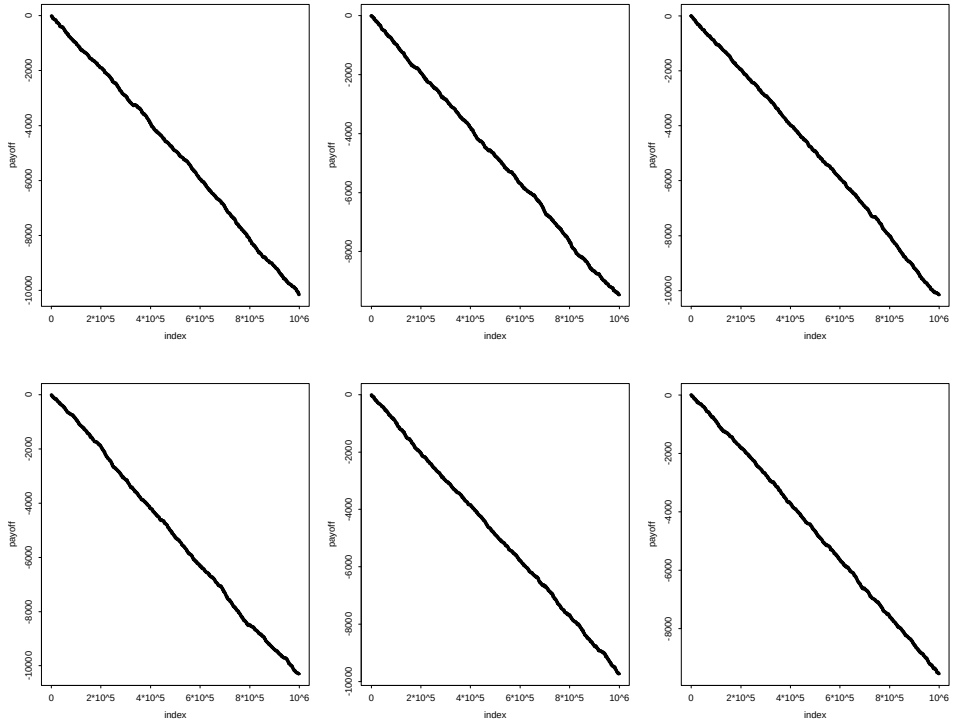

Figure 1.11: Possible realizations of the first 10^6 net payoffs (steps of the random walk $\{S_k - 0.51k : k \geq 0\}$ with steps U_k uniformly distributed in the interval $[0, 1]$.

player wins at least a certain amount w, i.e., by requiring that

$$P(S_{100} - f(100) \geq w) \geq p \ .$$

Given such a formulation, we can determine the optimal fee f, i.e., the maximum fee f such that the constraint is satisfied, which is attained when the probability just equals p.

As noted at the outset, when we consider making the game interesting, we might well conclude that a uniform payoff distribution for each play is boring. We might want to have the possibility of much larger positive and/or negative payoffs on one play. It is easy to devise more interesting games with different payoff distributions, but the statistical regularity associated with large numbers observed above tends to be the same. Readers are invited to make their own games and look at the net payoffs for 10^j plays for various values of j.

An extreme case that is often attractive is to have, like a lottery, some small chance of a very large payoff. However, with independent trials, as determined by successive spins of the spinner, the gambling house faces the danger of having to make too many large payoffs. Such large losses are avoided in lotteries by not letting the game be based on independent trials. In a lottery only a few prizes are awarded (and possibly shared) so that the people running the lottery are guaranteed a positive return. However, an insurance company cannot control the outcomes so tightly, so that careful analysis of the possible outcomes is necessary; e.g., see Embrechts, Klüppelberg and Mikosch (1997). We too will be interested in the possibility of exceptionally large values in random events.

1.2. Stochastic-Process Limits

The plots we have looked at indicate that there is statistical regularity associated with large n, i.e., with large sample sizes. We now want to understand *why* we see what we see, and what we will see in other related situations. For that purpose, we turn to probability theory; see Ross (1993) and Feller (1968) for introductions.

1.2.1. A Probability Model

We can use probability theory to explain what we have seen in the random walk plots. The first step is to introduce an appropriate mathematical model: Assuming that our random number generator is working properly

(an important issue, which we will not address, e.g., see p. 123 of Venables and Ripley (1994), L'Ecuyer(1998a,b) and references cited there), the observed values U_k, $1 \leq k \leq n$, should be distributed approximately as the first n values from a sequence of *independent and identically distributed* (IID) random variables uniformly distributed on $[0, 1]$ (defined on an underlying probability space). Indeed, the model fit is usually so good that there is a tendency to identify the mathematical model with the physical experiment (a mistake), but since the model fit is so good, we need not doubt that the mathematical conclusions are applicable.

Remark 1.2.1. *Mathematics and the physical world.* It is important to realize that a physical phenomenon, a mathematical model of that physical phenomenon and a simulation of that mathematical model are three different things. But, if the mathematical model is well chosen, the three may be closely related. In particular, a mathematical model, whether simulated or analyzed, may provide useful desciptions of the physical phenomenon.

We are interested in mathematical queueing models because of their ability to explain queueing phenomena, but we should not expect a perfect match. For example, mathematical models often succeed by exploiting the infinite, even though the physical phenomenon is finite. Random numbers generated on a computer are inherently finite, and yet simulations based on random numbers can be well described by mathematical models exploiting the infinite.

Here, we perform stochastic simulations to reveal statistical regularity, and we introduce and analyze mathematical models to explain that statistical regularity. We expect to capture key features, but we do not expect a perfect fit. We want the the mathematics to explain key features observed in the simulations, and we want the simulations to confirm key features predicted by the mathematics. ∎

With that attitude, let us consider the probability model consisting of a sequence of IID uniform random numbers. Within the context of that probability model, we want to formulate stochastic-process limits suggested by the plots. First, we see that as n increases the plotted random walk ceases to look discrete. For all sufficiently large n, the plotted random walk looks like a function of a continuous variable. Thus it is natural to seek a continuous-time representation of the original discrete-time random walk. We can do that by considering the associated continuous-time process $\{S_{\lfloor t \rfloor} : t \geq 0\}$, where $\lfloor \cdot \rfloor$ is the *floor function*, i.e., $\lfloor t \rfloor$ denotes the greatest integer less than or equal to t. If we also want to introduce centering,

then we do the centering first, and instead consider the centered process $\{S_{\lfloor t \rfloor} - m\lfloor t \rfloor : t \geq 0\}$ for appropriate centering constant m, which here is $1/2$. Thus the continuous-time representation of the random walk is a step function, which coincides with the random walk at integer arguments.

However, the step function is not the only possible continuous-time representation of the random walk. We could instead form a process with continuous sample paths by connecting the points by lines, i.e., by performing a *linear interpolation*. Then, instead of $S_{\lfloor t \rfloor}$, we consider

$$\tilde{S}(t) \equiv (t - \lfloor t \rfloor)S_{\lfloor t \rfloor + 1} + (1 + \lfloor t \rfloor - t)S_{\lfloor t \rfloor} \quad \text{for all} \quad t \geq 0 \ , \tag{2.1}$$

and similarly if we do centering. (With centering, we do the centering before doing the linear interpolation.) Possible initial segments of the two continuous-time processes associated with the discrete-time (uncentered) random walk for the case $n = 10$ are shown in Figure 1.12. (The vertical lines in the plot are not really part of the step function.) Even though the 10 random walk steps are the same for both continuous-time representations, the two initial segments of the continuous-time stochastic processes look very different in Figure 1.12. However, for large n, plots of the two continuous-time representations of the discrete-time random walk look virtually identical. To make that important point clear, we plot the two continuous-time representations of the same discrete-time centered random walk (same sample paths) for $n = 10^j$ for $j = 1, \ldots, 4$ in Figure 1.13. Figure 1.13 shows that the two alternative representations indeed look the same for all n sufficiently large. Thus, when we focus on the random-walk plots for large n, we regard the two alternatives as equivalent. For our remaining discussion here, though, we will only discuss the step functions.

We now want to scale time and space (the horizontal and vertical dimensions in the plots). Note that the plotter scales time by putting the $n+1$ random walk values in a region of fixed width. Thus, if we let 1 be the available width of the plot, then the $n + 1$ random walk values are spaced $1/n$ apart. Equivalently, time is scaled automatically by the plotting routine by multiplying time t by n, i.e., by replacing t with nt. Then, for each n, we only look at the process for t in the closed interval $[0, 1]$. The final position of the random walk for any n corresponds to $t = 1$.

We can also consider the space scaling in the same way. We can let 1 be the available height of the plot. Then the plotter automatically scales space by subtracting the minimum value and dividing by the range of the plotted values. Unfortunately, however, the range is random. Moreover, there is a complicated dependence between the path and its range. In formulating a

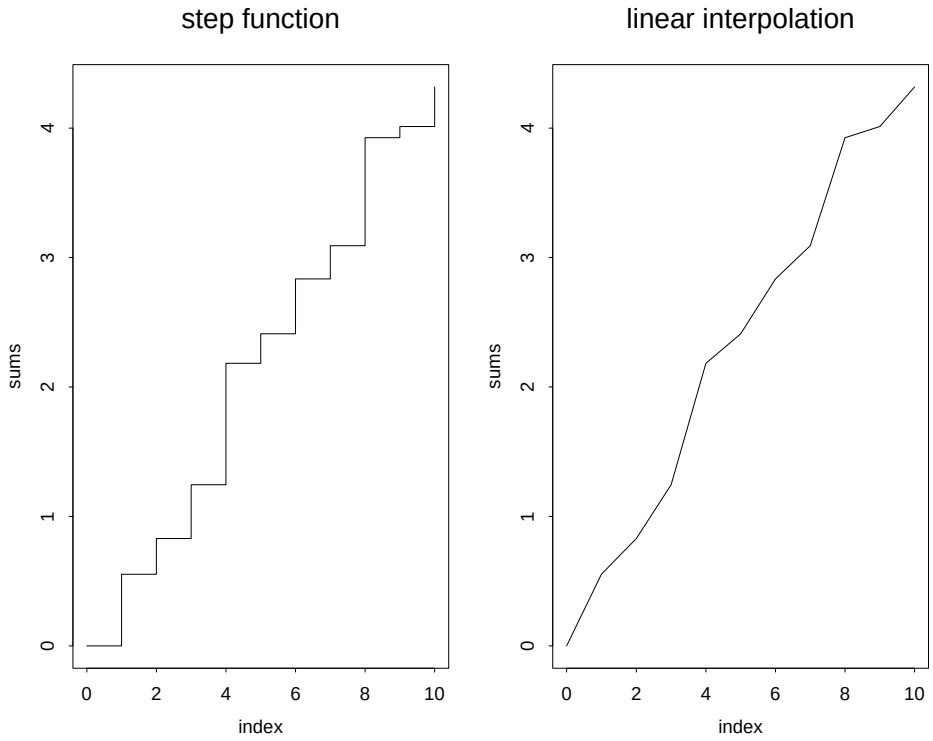

Figure 1.12: Possible initial segments of the two continuous-time stochastic processes constructed from one realization of an uncentered random walk with uniform steps for the case $n = 10$. The step-function representation appears on the left, while the linear-interpolation representation appears on the right.

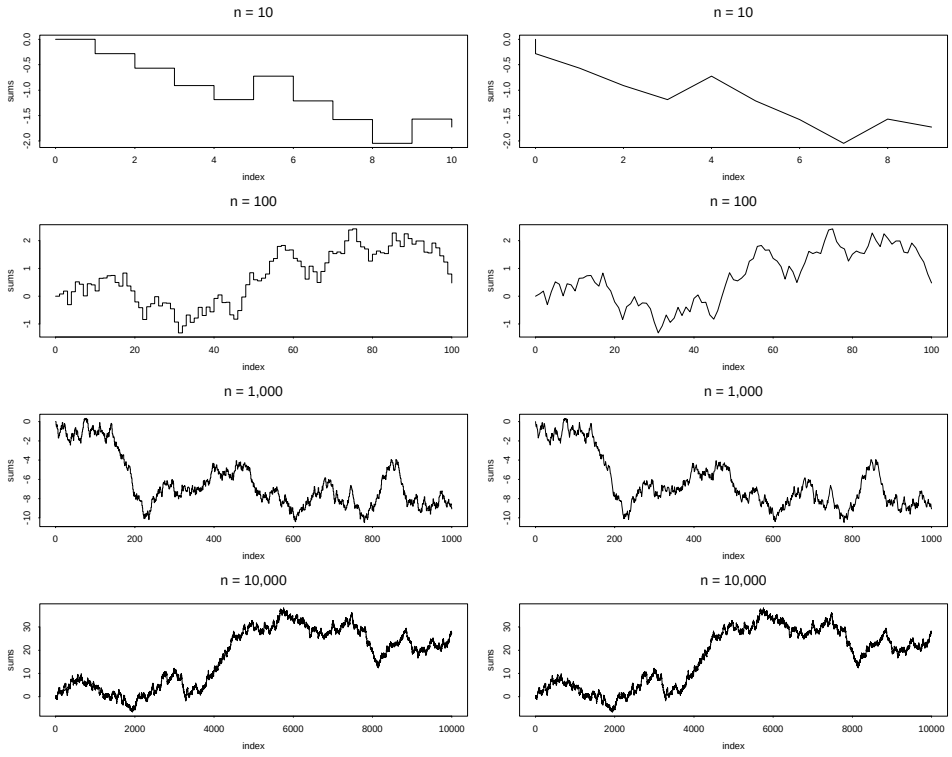

Figure 1.13: The two alternative continuous-time representations of a discrete-time process constructed from common realizations of a centered random walk with uniform steps for $n = 10^j$ with $j = 1, \ldots, 4$. The step-function representation appears on the left, while the linear-interpolation representation appears on the right.

stochastic-process limit, it is natural to try to perform the space scaling, like the time scaling, with a deterministic function of n. With such deterministic space scaling, we hope to achieve a nondegenerate limit as $n \to \infty$, but one for which the range is allowed to remain random. In the limit as $n \to \infty$, we will achieve essentially the same thing as the plots if the normalized range converges to a nondegenerate random limit.

What we do, then, is scale space by dividing by c_n, where $\{c_n : n \geq 1\}$ is a sequence of (deterministic) real numbers with $c_n \to \infty$ as $n \to \infty$. That is, for each n, we form the stochastic process

$$\mathbf{S}_n(t) \equiv c_n^{-1}(S_{\lfloor nt \rfloor} - m(\lfloor nt \rfloor), \quad 0 \leq t \leq 1 \ . \tag{2.2}$$

We then want to find an appropriate sequence $\{c_n : n \geq 1\}$ so that

$$\{\mathbf{S}_n(t) : 0 \leq t \leq 1\} \to \{\mathbf{S}(t) : 0 \leq t \leq 1\} \quad \text{as} \quad n \to \infty \ , \tag{2.3}$$

where $\mathbf{S} \equiv \{\mathbf{S}(t) : 0 \leq t \leq 1\}$ is an appropriate limit process with t ranging over the interval $[0, 1]$ and $\to$ in (2.3) is an appropriate mode of convergence. When we have a limit as in (2.3), we have a *stochastic-process limit*.

1.2.2. Classical Probability Limits

Classical probability limits help explain the statistical regularity we have seen. First, referring to the asymptotically linear plots in Figures 1.1 and 1.11, the *strong law of large numbers* (SLLN) implies that the scaled partial sums $n^{-1}S_n$ approach the mean m as $n \to \infty$ with probability 1 (w.p.1); e.g., see Chapter X of Feller (1968), Chapter VII of Feller (1971) and Chapter 5 of Chung (1974). (In Figure 1.1 the mean is 1/2; in Figure 1.11 the mean is -0.01.)

As an easy consequence of the SLLN, we can also conclude that

$$n^{-1}S_{\lfloor nt \rfloor} \to mt \quad \text{w.p.1} \quad \text{as} \quad n \to \infty$$

for each $t > 0$. Moreover, the pointwise convergence can actually be extended to uniform convergence over bounded intervals:

$$\{n^{-1}S_{\lfloor nt \rfloor} : 0 \leq t \leq 1\} \to \{mt : 0 \leq t \leq 1\} \quad \text{w.p.1} \quad \text{as} \quad n \to \infty \ ,$$

uniformly in t for t in the interval $[0, 1]$. In other words,

$$\sup\{|n^{-1}S_{\lfloor nt \rfloor} - mt| : 0 \leq t \leq 1\} \to 0 \quad w.p.1 \quad \text{as} \quad n \to \infty \ .$$

Thus, in the setting of Figure 1.1, the limit (2.3) holds without centering (with $m = 0$) for $c_n = n$ with the limit process $\mathbf{S}$ being the line with slope $1/2$ (the one-step mean) defined over the interval $[0, 1]$. In this case, the mode of convergence in (2.3) is convergence w.p.1 on a space of functions with the uniform distance

$$||x_1 - x_2|| \equiv \sup \{|x_1(t) - x_2(t)| : 0 \leq t \leq 1\} \ .$$

In this case, the stochastic-process limit is called a *functional strong law of large numbers* (FSLLN). Interestingly, the SLLN and the FSLLN are actually equivalent; see Theorem 3.2.1 in the Internet Supplement.

Next, turning to the plots of the centered random walks, with centering by the mean, in Figures 1.2, 1.3 and 1.4, we can appeal to the *central limit theorem* (CLT). The CLT implies that

$$(\sigma^2 n)^{-1/2}(S_n - nm) \Rightarrow N(0,1) \quad \text{as} \quad n \to \infty \ , \tag{2.4}$$

where $m \equiv EU_k = 1/2$ is the mean and $\sigma^2 \equiv Var\, U_k = 1/12$ is the variance of the uniform summand U_k, $\Rightarrow$ denotes convergence in distribution and the *standard* normal random variable $N(0,1)$ has cdf

$$\Phi(t) \equiv P(N(0,1) \leq x) \equiv \int_{-\infty}^{x} (2\pi)^{-1/2} e^{-u^2/2}\, du \ ; \tag{2.5}$$

e.g., see Section VIII.4 of Feller (1971) and Chapter 7 of Chung (1974).

It is useful to review what the limit (2.4) means: The convergence in distribution means that the cdf's converge, i.e.,

$$P(n^{-1/2}(S_n - mn) \leq x) \to P(N(0, \sigma^2) \leq x) \quad \text{as} \quad n \to \infty \tag{2.6}$$

for all x. More generally, given real-valued random variables Z_n, $n \geq 1$, and Z, there is convergence in distribution, by the standard definition, denoted by $Z_n \Rightarrow Z$, if the associated cdf's converge, i.e., if

$$F_n(x) \equiv P(Z_n \leq x) \to P(Z \leq x) \equiv F(x) \quad \text{as} \quad n \to \infty \tag{2.7}$$

for all x that are continuity points of the limiting cdf F, i.e., for which $P(Z = x) = 0$.

Since the normal distribution has a continuous cdf, the restriction to continuity points of the limiting cdf in (2.7) does not arise in (2.6). We need to allow non-convergence at discontinuity points in (2.7), because we want to say that we have convergence $Z_n \Rightarrow Z$ in situations such as the special

case in which $P(Z = z) = 1$ and $P(Z_n = z_n) = 1$ for all n and $z_n \to z$ as $n \to \infty$. If $z_n \to z$ with $z_n > z$ for all n, then $F_n(z) \equiv P(Z_n \le z) = 0$ for all n, while $F(z) \equiv P(Z \le z) = 1$. Since $F_n(x) \to F(x)$ for all x except $x = z$, we obtain the desired convergence $Z_n \Rightarrow Z$ if we require pointwise convergence of the cdf's everywhere except at discontinuity points of the limiting cdf F.

There also are other convenient equivalent characterizations of convergence in distribution. In particular, (2.7) holds if and only if

$$E[h(Z_n)] \to E[h(Z)] \quad \text{as} \quad n \to \infty \tag{2.8}$$

for every continuous bounded real-valued function h on $\mathbb{R}$, where E is the expectation operator. Moreover, (2.7) and (2.8) hold if and only if

$$g(Z_n) \Rightarrow g(Z) \quad \text{as} \quad n \to \infty \tag{2.9}$$

for every continuous function g on $\mathbb{R}$. The alternative characterizations (2.8) and (2.9) are useful because they generalize to random elements of more general spaces.

The CLT in (2.4) explains the statistical regularity associated with the final positions of the centered random walks: In agreement with Figures 1.5 – 1.7, the CLT tells us that the centered partial sums $S_n - mn$ should be approximately normally distributed with mean 0 for all n sufficiently large.

We can also apply the CLT to obtain a corresponding limit for the scaled random walk $\mathbf{S}_n$ in (2.2) at an arbitrary time t in the interval $[0, 1]$. More generally, we can consider an arbitrary $t \ge 0$. To do so, we set $c_n = \sqrt{n}$ and $m = 1/2$. In particular, it is an easy consequence of (2.4) that we must have

$$n^{-1/2}(S_{\lfloor nt \rfloor} - m\lfloor nt \rfloor) \Rightarrow \sigma N(0, t) \quad \text{in} \quad \mathbb{R} \quad \text{as} \quad n \to \infty \tag{2.10}$$

for each $t \ge 0$, where $m = 1/2$ and $\sigma^2 = 1/12$.

From (2.10) we clearly see that the space-scaling constants c_n in (2.2) must be asymptotically equivalent to $c\sqrt{n}$ for some constant c as $n \to \infty$. Moreover, the space scaling by $\sqrt{n}$ is consistent with the units on the axes in Figures 1.2–1.4. Indeed, if we instead scale by $c_n = n^p$ for $p > 1/2$, then the values converge to 0 as $n \to \infty$. Similarly, if we scale by $c_n = n^p$ for $p < 1/2$, then the values diverge as $n \to \infty$. (The absolute values diverge to infinity.) This property can be confirmed by further analysis of simulations, but we do not pursue it.

We now want to convert (2.10) into a stochastic-process limit of the form (2.3). Note that the left side of (2.10) coincides with $\mathbf{S}_n(t)$, but the right side of (2.10) is not a stochastic process evaluated at time t. What we need to do is identify the appropriate limit process $\mathbf{S}$ in (2.3).

1.2.3. Identifying the Limit Process

We should recognize that we have arrived at another critical point. Another important intellectual step is needed here. *We not only must identify the limit process; we need to realize that there indeed should be a limit process.*

The appropriate limit process turns out to be a *Brownian motion* (BM). Brownian motion stochastic processes can be characterized as the real-valued stochastic processes with stationary and independent increments having continuous sample paths. Brownian motion evaluated at time t turns out to be normally distributed with mean mt and variance $\sigma^2 t$ for some constants m and σ^2.

The special Brownian motion with parameters $m = 0$ and $\sigma^2 = 1$ is called *standard Brownian motion*; we shall refer to it by $\mathbf{B} \equiv \{\mathbf{B}(t) : t \geq 0\}$. It has marginal distributions

$$\mathbf{B}(t) \stackrel{\mathrm{d}}{=} N(0,t), \quad t \geq 0, \tag{2.11}$$

where $\stackrel{\mathrm{d}}{=}$ denotes equality in distribution.

An increment of Brownian motion is $\mathbf{B}(u) - \mathbf{B}(t)$ for $u > t$. By *stationary and independent increments*, we mean that the k-dimensional random vector

$$(\mathbf{B}(u_1 + h) - \mathbf{B}(t_1 + h), \ldots, \mathbf{B}(u_k + h) - \mathbf{B}(t_k + h))$$

has a distribution independent of h for all k, and that the k component random variables are independent, providing that $0 \leq t_1 \leq u_1 \leq t_2 \leq \cdots \leq u_k$.

Combining (2.10) and (2.11), we see that we can also express the limit (2.10) in terms of Brownian motion. In particular, after letting $c_n = \sqrt{n}$ in (2.2), we see that (2.10) is equivalent to

$$\mathbf{S}_n(t) \Rightarrow \sigma \mathbf{B}(t) \quad \text{in} \quad \mathbb{R} \quad \text{as} \quad n \to \infty \quad \text{for all} \quad t \geq 0, \tag{2.12}$$

where $\mathbf{B}$ is a standard Brownian motion,

$$\mathbf{S}_n(t) \equiv n^{-1/2}(S_{\lfloor nt \rfloor} - m(\lfloor nt \rfloor), \quad t \geq 0\ , \tag{2.13}$$

and $\sigma^2 = 1/12$ because the steps in the random walk are uniformly distributed over $[0,1]$. In equations (2.11), (2.12) and (2.13) we have let t range over the semi-infinite interval $[0,\infty)$, but we could also have restricted t to the closed interval $[0,1]$ to be consistent with the plots.

We can apply the limit in (2.12) to generate approximations for the terms of the original random walk. To generate approximations, we replace the convergence in distribution by approximate equality in distribution. From (2.12), we obtain the approximation

$$S_{\lfloor nt \rfloor} \approx m\lfloor nt \rfloor + n^{1/2}\sigma \mathbf{B}(t) \tag{2.14}$$

or

$$S_k \approx mk + n^{1/2}\sigma \mathbf{B}(k/n) \ , \tag{2.15}$$

where k is understood to be of order n and $\approx$ means approximately equal to in distribution. Note that the quality of the approximation for large n tends to depend more on the time scaling by n and the space scaling by $\sqrt{n}$ than the limit process $\sigma\mathbf{B}$.

The limit in (2.12) (with t ranging over the unit interval $[0, 1]$) can be regarded as the explanation for what we have seen in the random-walk plots. The limit in (2.12) is a *stochastic-process limit*, because it establishes convergence of the sequence of stochastic processes $\{\{\mathbf{S}_n(t) : 0 \leq t \leq 1\} : n \geq 1\}$ in (2.13) to the limiting stochastic process $\{\sigma\mathbf{B}(t) : 0 \leq t \leq 1\}$. However, we want to go beyond the limit as expressed via (2.12). We want to strengthen the form of convergence in order to be able to deduce convergence of related quantities of interest; in particular, we want to show that plots of the centered random walk converge to plots of standard Brownian motion as $n \to \infty$.

The probability law or distribution of a stochastic process is usually specified by the family of its finite-dimensional distributions (f.d.d.'s). Hence, a natural first step is to go beyond convergence of the one-dimensional marginal distributions, which is provided by (2.12), to convergence of the f.d.d.'s, i.e., the k-dimensional marginal distributions for all k. From the assumed independence among the random walk steps, it is not difficult to see that (2.12) can be extended to obtain

$$(\mathbf{S}_n(t_1), \ldots, \mathbf{S}_n(t_k)) \Rightarrow (\sigma\mathbf{B}(t_1), \ldots, \sigma\mathbf{B}(t_k)) \quad \text{in} \quad \mathbb{R}^k \tag{2.16}$$

as $n \to \infty$ for all positive integers k and all k time points $t_1, \ldots, t_k$ with $0 \leq t_1 < \cdots < t_k \leq 1$, where convergence in distribution of random elements of $\mathbb{R}^k$ is defined by the natural generalization of (2.7), (2.8) or (2.9). Because of the independence among the random walk steps in this example, there is little difference between (2.12) and (2.16), but in general (2.16) is a much stronger conclusion.

However, we want to go even further. We want to go beyond convergence of the f.d.d.'s in (2.16) to convergence of the plots. We want to establish

limits for more general functions of the stochastic processes. To do so, *we regard* $\mathbf{S}_n$ *and* $\mathbf{B}$ *as random elements of a function space containing all possible sample paths.* (A function space is a space of functions.)

For $\mathbf{B}$, we could consider the space $C \equiv C([0,1], \mathbb{R})$ of all continuous real-valued functions on the unit interval $[0,1]$, but to include $\mathbf{S}_n$, we need discontinuous functions. (We could work with the space C if we used linearly interpolated random walks, as in (2.1), but we are considering the step functions.) We could consider a space containing all continuous functions and the special step functions that capture the structure of $\mathbf{S}_n$, but with other applications in mind, we consider a larger set of functions. We let the function space be the set $D \equiv D([0,1], \mathbb{R})$ of all real-valued functions on $[0,1]$ that are right-continuous at all t in $[0,1)$ and have left limits everywhere in $(0,1]$, endowed with an appropriate topology (notion of convergence, see Chapter 3).

The desired generalization of (2.12) and (2.16) follows from *Donsker's theorem.* Donsker's theorem is a *functional central limit theorem* (FCLT), which implies here that

$$\mathbf{S}_n \Rightarrow \sigma\mathbf{B} \quad \text{in} \quad D, \tag{2.17}$$

where again $\mathbf{S}_n$ is the scaled random walk in (2.13), $\mathbf{B}$ is standard Brownian motion and the function space D is endowed with an appropriate topology. We discuss the topology on D and the precise meaning of (2.17) in Section 3.3.

Even though Brownian motion has a relatively simple characterization, it is a special stochastic process. For example, it has the self-similarity property observed in the plots (without limit). In particular, for all $c > 0$, the stochastic process $\{c^{-1/2}\mathbf{B}(ct) : 0 \le t \le 1\}$ has the same probability law on D; equivalently, it has the same finite-dimensional distributions, i.e., the random vector $(c^{-1/2}\mathbf{B}(ct_1), \ldots, c^{-1/2}\mathbf{B}(ct_k))$ has a distribution in $\mathbb{R}^k$ that is independent of c for any positive integer k and any k time points $t_i, 1 \le i \le k$, with $0 < t_1 < \cdots < t_k \le 1$.

Indeed, the self-similarity is a direct consequence of the stochastic-process limit in (2.17): First observe from (2.13) that, for any $c > 0$,

$$\mathbf{S}_{cn}(t) = c^{-1/2}\mathbf{S}_n(ct), \quad t \ge 0\ . \tag{2.18}$$

By taking limits on both sides of (2.18), we obtain

$$\{\mathbf{B}(t) : 0 \le t \le 1\} \stackrel{\mathrm{d}}{=} \{c^{-1/2}\mathbf{B}(ct) : 0 \le t \le 1\}\ . \tag{2.19}$$

For further discussion, see Section 4.2.

Even though we are postponing a detailed discussion of the meaning of the convergence in (2.17), we can state a convenient characterization, which explains the applied value of (2.17) compared to (2.12) and (2.16). Just as in (2.8), the limit (2.17) means that

$$E[h(\mathbf{S}_n)] \to E[h(\sigma\mathbf{B})] \quad \text{as} \quad n \to \infty \tag{2.20}$$

for every continuous bounded real-valued function h on D. The topology on D enters in by determining which functions h are continuous. Just as with (2.9), (2.20) holds if and only if

$$g(\mathbf{S}_n) \Rightarrow g(\sigma\mathbf{B}) \quad \text{in} \quad \mathbb{R} \tag{2.21}$$

for every continuous real-valued function g on D. (It is easy to see that (2.20) implies (2.21) because the composition function $h \circ g$ is a bounded continuous real-valued function whenever g is continuous and h is a bounded continuous real-valued function.) Interestingly, (2.21) is the way that Donsker (1951) originally expressed his FCLT. The convergence of the functionals (real-valued functions) in (2.21) explains why the limit in (2.17) is called a FCLT.

It turns out that we also obtain (2.21) for every continuous function g, regardless of the range. For example, the function g could map D into D. Then we can obtain new stochastic-process limits from any given one. That is an example of the continuous-mapping approach for obtaining stochastic-process limits; see Section 3.4. The representation (2.21) is appealing because it exposes the applied value of (2.17) as an extension of (2.12) and (2.16). We obtain many associated limits from (2.21).

1.2.4. Limits for the Plots

We illustrate the continuous-mapping approach by establishing a limit for the plotted random walks, where as before we regard the plot as being in the unit square $[0,1] \times [0,1]$.

To establish limits for the plotted random walks, we use the functions $sup : D \to \mathbb{R}$, $inf : D \to \mathbb{R}$, $range : D \to \mathbb{R}$ and $plot : D \to D$, defined for any $x \in D$ by

$$sup(x) \equiv \sup_{0 \le t \le 1} x(t),$$

$$inf(x) \equiv \inf_{0 \le t \le 1} x(t),$$

$$range(x) \equiv sup(x) - inf(x)$$

and

$$plot(x) \equiv (x - inf(x))/range(x) \ .$$

Note that $plot(x)$ is an element of D for each $x \in D$ such that $range(x) \neq 0$. Moreover, the function $plot$ is scale invariant, i.e., for each positive scalar c and $x \in D$ with $range(x) \neq 0$,

$$plot(cx) = plot(x) \ .$$

Fortunately, these functions turn out to preserve convergence in the topologies we consider. (The first three functions are continuous, while the final $plot$ function is continuous at all x for which $range(x) \neq 0$, which turns out to be sufficient.) Hence we obtain the initial limits

$$n^{-1/2} \max_{1 \leq k \leq n} \{S_k - mk\} = sup(\mathbf{S}_n) \Rightarrow sup(\sigma\mathbf{B}) \equiv \sup_{0 \leq t \leq 1} \{\sigma\mathbf{B}(t)\} \ ,$$

$$n^{-1/2} \min_{1 \leq k \leq n} \{S_k - mk\} = inf(\mathbf{S}_n) \Rightarrow inf(\sigma\mathbf{B}) \equiv \inf_{0 \leq t \leq 1} \{\sigma\mathbf{B}(t)\} \ ,$$

$$n^{-1/2} range(\{S_k - mk : 0 \leq k \leq n\}) \equiv range(\mathbf{S}_n) \Rightarrow range(\sigma\mathbf{B})$$

in $\mathbb{R}$ and the final desired limit

$$plot(\mathbf{S}_n) \Rightarrow plot(\sigma\mathbf{B}) = plot(\mathbf{B}) \quad \text{in} \quad D \ ,$$

where

$$plot(\{S_k - mk : 0 \leq k \leq n\}) = plot(\{c_n^{-1}(S_k - mk) : 0 \leq k \leq n\}) \equiv plot(\mathbf{S}_n) \ ,$$

from Donsker's theorem ((2.17) and (2.21)).

The limit $plot(\mathbf{S}_n) \Rightarrow plot(\mathbf{B})$ states that the plot of the scaled random walk converges to the plot of standard Brownian motion. Note that we use $plot$, not only as a function mapping D into D, but as a function mapping $\mathbb{R}^{n+1}$ into D taking the random walk segment into its plot.) Hence Donsker's theorem implies that the random walk plots can indeed be regarded as approximate plots of Brownian motion for all sufficiently large n. By using the FCLT refinement, we see that the stochastic-process limits do indeed explain the statistical regularity observed in the plots.

To highlight this important result, we state it formally as a theorem. Later chapters will provide a proof; specifically, we can apply Sections 3.4, 12.7 and 13.4.

Theorem 1.2.1. (convergence of plots to the plot of standard Brownian motion) *Consider an arbitrary stochastic sequence* $\{S_k : k \geq 0\}$. *Suppose that the limit in* (2.3) *holds in the space* D *with one of the Skorohod non-uniform topologies, where* $c_n = \sqrt{n}$ *and* $\mathbf{S} = \sigma\mathbf{B}$ *for some positive constant* σ, *with* $\mathbf{B}$ *being standard Brownian motion, as occurs in Donsker's theorem. Then*

$$plot(\{S_k - mk : 0 \leq k \leq n\}) \Rightarrow plot(\mathbf{B}) \ .$$

But an even more general result holds: *We have convergence of the plots for any space-scaling constants and almost any limit process.* We have the following more general theorem (proved in the same way as Theorem 1.2.1).

Theorem 1.2.2. (convergence of plots associated with any stochastic-process limit) *Consider an arbitrary stochastic sequence* $\{S_k : k \geq 0\}$. *Suppose that the limit in* (2.3) *holds in the space* D *with one of the Skorohod non-uniform topologies, where* c_n *and* $\mathbf{S}$ *are arbitrary. If*

$$P(range(\mathbf{S}) = 0) = 0 \ ,$$

then

$$plot(\{S_k - mk : 0 \leq k \leq n\}) \Rightarrow plot(\mathbf{S}) \ .$$

Note that the functions *sup*, *inf*, *range* and *plot* depend on more than one value $x(t)$ of the function x; they depend on the function over an initial segment. Thus, we exploit the strength of the limit in D in (2.17) as opposed to the limit in $\mathbb{R}$ in (2.12) or even the limit in $\mathbb{R}^k$ in (2.16). For the random walk we have considered (with IID uniform random steps), the three forms of convergence in (2.12), (2.16) and (2.17) all hold, but in general (2.16) is strictly stronger than (2.12) and (2.17) is strictly stronger than (2.16). Formulating the stochastic-process limits in D means that we can obtain many more limits for related quantities of interest, because many more quantities of interest can be represented as images of continuous functions on the space of stochastic-process sample paths.

Remark 1.2.2. *Limits for the relative final position.* As noted in Remark 1.1.1, if we look at the final position of the centered random walk in the plots, ignoring the units on the axes, then we actually see the relative final position of the centered random walk, as defined in (1.1). Statistical regularity for the relative final position also follows directly from Theorems 1.2.1 and 1.2.2, because the relative final position is just the plot evaluated at time 1, i.e.,

$plot(x)(1)$. Provided that 1 is almost surely a continuity point of the limit process $\mathbf{S}$, under the conditions of Theorem 1.2.2 we have

$$R_n \Rightarrow plot(\mathbf{S})(1) \quad \text{in} \quad \mathbb{R} \quad \text{as} \quad n \to \infty,$$

as a consequence of the continuous-mapping approach, using the projection map that maps $x \in D$ into $x(1)$. ∎

To summarize, the random-walk plots *reveal* remarkable statistical regularity associated with large n because the plotter automatically does the required scaling. In turn, the stochastic-process limits *explain* the statistical regularity observed in the plots. In particular, Donsker's FCLT implies that the random-walk plots converge in distribution to the plots of standard Brownian motion as $n \to \infty$.

1.3. Invariance Principles

The random walks we have considered so far are very special: the steps are IID with a uniform distribution in the interval $[0, 1]$. However, the great power of the SLLN, FSLLN, CLT and FCLT is that they hold much more generally. Essentially the same limits hold in many situations in which the step distribution is changed or the IID condition is relaxed, or both. Moreover, the limits each depend on only a single parameter of the random walk. The limits in the SLLN and the FSLLN only involve the single parameter m, which is the mean step size in the IID case. Similarly, after centering is done, the limits in the CLT and FCLT only involve the single parameter σ^2, which is the variance of the step size in the IID case. Thus these limit theorems are *invariance principles.*

Moreover, the plots have an even stronger invariance property, because the limiting plots have no parameters at all! (We are thinking of the plot being in the unit square $[0, 1] \times [0, 1]$ in every case, ignoring the units on the axes.) Assuming only that the mean is positive, the plots of the uncentered random walk (with arbitrary step-size distribution) approach the identity function $e \equiv e(t) \equiv t, \quad 0 \leq t \leq 1$. If instead the mean is negative, then the limiting plot is $-e$ over the interval $[0, 1]$. Similarly, the plots of the centered random walks approach the plot of standard Brownian motion over $[0, 1]$; i.e., the limiting plot does not depend on the variance σ^2. Thus, the random-walk plots reveal remarkable statistical regularity!

The power of the invariance principles is phenomenal. We will give some indication by giving a few examples and by indicating how they can be

applied. We recommend further experimentation to become a true believer. For example, the plots of the partial sums – centered and uncentered – should be contrasted with corresponding plots for the random-walk steps. Even for large n, plots of uniform random numbers and exponential (exponentially distributed) random numbers look very different, whereas the plots of the corresponding partial sums look the same (for all n sufficiently large).

1.3.1. The Range of Brownian Motion

We can apply the invariance property to help determine limiting probability distributions. For example, we can apply the invariance property to help determine the distribution of the limiting random variables $sup(\mathbf{B})$ and $range(\mathbf{B})$.

We first consider the supremum $sup(\mathbf{B})$. We can use combinatorial methods to calculate the distribution of $\max_{1\le k\le n}\{S_k - km\}$ for any given n for the special case of the *simple random walk*, with $P(X_1 = +1) = P(X_1 = -1) = 1/2$, as shown in Chapter III of Feller (1968) or Section 11 of Billingsley (1968). In that way, we obtain

$$P(sup(\mathbf{B}) > x) = 2P(N(0,1) > x) \equiv 2\Phi^c(x) \ , \tag{3.1}$$

where $\Phi^c(t) \equiv 1 - \Phi(t)$ for Φ in (2.5). Since $sup(\mathbf{B}) \stackrel{\mathrm{d}}{=} \mid N(0,1) \mid$,

$$E[sup(\mathbf{B})] = \sqrt{2/\pi} \approx 0.8 \tag{3.2}$$

and

$$E[sup(\mathbf{B})^2] = E[N(0,1)^2] = 1 \ .$$

These calculations are not entirely elementary; for details see 26.2.3, 26.2.41 and 26.2.46 in Abramowitz and Stegun (1972).

The limit $range(\sigma\mathbf{B})$ is more complicated, but it too can be characterized; see Section 11 of Billingsley (1968) and Borodin and Salminen (1996). There the combinatorial methods for the simple random walk are used again to determine the joint distribution of $inf(\mathbf{B})$ and $sup(\mathbf{B})$, yielding

$$P(a < inf(\mathbf{B}) < sup(\mathbf{B}) < b) = \sum_{k=-\infty}^{k=+\infty} (-1)^k [\Phi(b+k(b-a)) - \Phi(a+k(b-a))] \ ,$$

where Φ is again the standard normal cdf. From (3.2), we see that the mean of the range is

$$E[range(\mathbf{B})] = E[sup(\mathbf{B})] - E[inf(\mathbf{B})] = 2E[sup(\mathbf{B})] = 2\sqrt{2/\pi} \approx 1.6.$$

We can perform multiple replications of random-walk simulations to estimate the distribution of $range(\mathbf{B})$ and associated summary characteristics such as the variance. We show the estimate of the probability density function of $range(\mathbf{B})$ based on $10,000$ samples of the random walk with $10,000$ steps, each uniformly distributed on $[0,1]$, in Figure 1.14 (again obtained using the nonparametric density estimator *density* from S). The range of the centered random walk should be approximately $\sigma\sqrt{n}$ times the range $range(\mathbf{B})$, so we divide the observed ranges in this experiment by $\sqrt{n/12} = 28.8675$. The estimated mean and standard deviation of $range(\mathbf{B})$ were 1.58 and 0.474, respectively. The estimated 0.1, 0.25, 0.5, 0.75 and 0.9 quantiles were 1.05, 1.24, 1.50, 1.85 and 2.23, respectively. This characterization of the distribution of $range(\mathbf{B})$ helps us interpret what we see in the random-walk plots.

From the analysis above, we know approximately what the mean and standard deviation of the range should be in the random-walk plots. Since $E[range(\mathbf{B})] \approx 1.6$, the mean of the random walk range should be about $1.6\sigma\sqrt{n} \approx 0.46\sqrt{n}$. Similarly, since the standard deviation of $range(\mathbf{B})$ is approximately 0.47, the standard deviation of the range in the random-walk plot should be approximately $0.47\sigma\sqrt{n} \approx 0.14\sqrt{n}$. Hence the (mean, standard deviation) pairs in Figures 1.3 and 1.4 with $n = 10^4$ and $n = 10^6$ are, respectively, (46, 14) and (460, 140). Note that the six observed values in each case are consistent with these pairs.

Historically, the development of the limiting behavior of $sup(\mathbf{S}_n)$ played a key role in the development of the general theory; e.g. see the papers by Erdös and Kac (1946), Donsker (1951), Prohorov (1956) and Skorohod (1956). ∎

Remark 1.3.1. *Fixed space scaling.* In our plots, we have let the plotter automatically determine the units on the vertical axis. Theorems 1.2.1 and 1.2.2 show that there is striking statistical regularity associated with automatic plotting. However, for comparison, it is often desirable to have common units. Interestingly, Donsker's FCLT and the analysis of the range above shows how to determine appropriate units for the vertical axis for the centered random walk, before the simulations are run.

First, the CLT and FCLT tell us the range of values for the centered random walk should be of order $\sqrt{n}$ as the sample size n grows. The invariance principle tells us that, for suitably large n the scaling should depend on the random-walk-step distribution only through its variance σ^2.

The limit for the supremum $sup(\mathbf{S}_n) \equiv n^{-1/2} \max_{1 \le k \le n}\{S_k - mk\}$ tells us more precisely what fixed space scaling should be appropriate for the

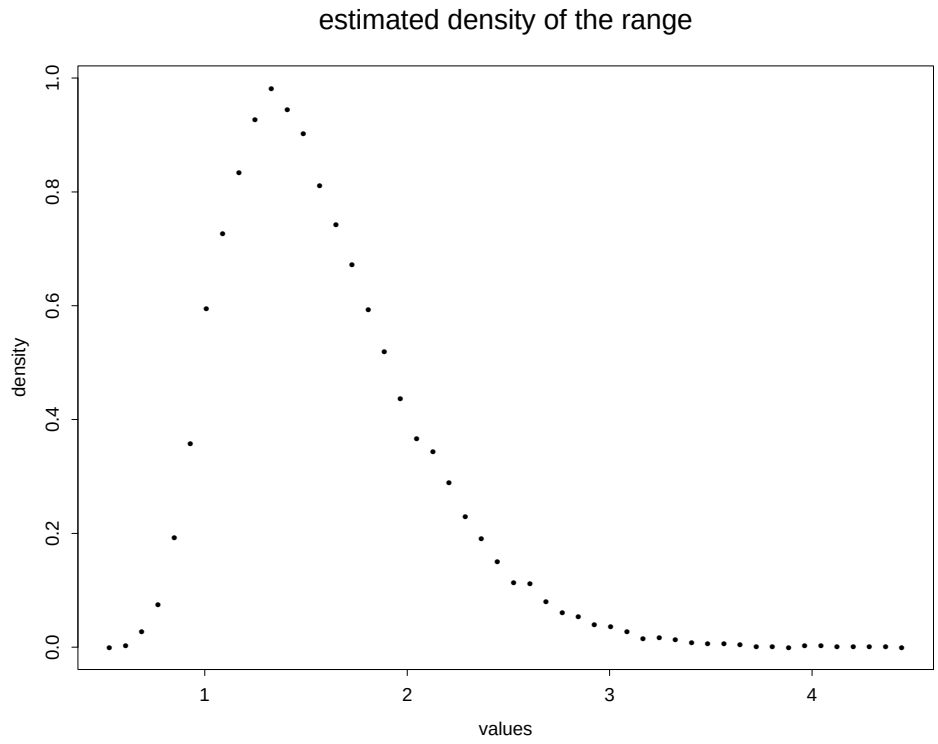

Figure 1.14: An estimate of the probability density of the range of Brownian motion over $[0, 1]$, obtained from $10,000$ independent samples of random walks with $10,000$ steps, each step being uniformly distributed in the interval $[0, 1]$.

plots. Since $2\,P(N(0,1) \geq 4)$ may be judged suitably small, from (3.1) we conclude that it should usually be appropriate to let the values on the vertical axis for a centered random walk fall in the interval $[-4\sigma\sqrt{n}, 4\sigma\sqrt{n}]$ as a function of n and σ^2. For example, we could use this space scaling to replot the six random-walk plots in Figure 1.4. Since $n = 10^6$ and $\sigma^2 = 1/12$ there, we would let the values on the vertical axies in Figure 1.4 fall in the interval $[-1155, 1155]$. Notice that the values for the six plots all fall in the interval $[-700, 450]$, so that this fixed space scaling would work in Figure 1.4. ∎

To gain a better appreciation of the invariance property, we perform some more simulations. First, we want to see that the IID conditions are *not* necessary.

1.3.2. Relaxing the IID Conditions

To illustrate how the IID conditions can be relaxed, we consider *exponential smoothing.*

Example 1.3.1. *Exponential smoothing.* We now consider a simple example of a random walk in which the steps are neither independent nor identically distributed. We let the steps be constructed by exponential smoothing. Equivalently, the steps are an autoregressive moving-average (ARMA) process of order (1,0); see Section 4.6.

In particular, suppose that we generate uniform random numbers U_k on the interval $[0, 1]$, $k \geq 1$, as before, but we now let the k^{th} step of the random walk be defined recursively by

$$X_k \equiv (1 - \gamma)X_{k-1} + \gamma U_k, \quad k \geq 1 \ , \tag{3.3}$$

where $X_0 = U_0$, where U_0 is another uniform random number on $[0, 1]$ and $0 < \gamma < 1$. Clearly, the new random variables X_k are neither independent nor identically distributed. Moreover, the distribution of X_k is no longer uniform. It is not difficult to see, though, that as k increases the distribution of X_k approaches a nondegenerate limit. More generally, the sequence $\{X_{n+k} : k \geq 0\}$ is asymptotically stationary as $n \to \infty$, but successive random variables remain dependent.

We now regard the random variables X_k as steps of a random walk; i.e., we let the successive positions of the random walk be

$$S_k \equiv X_1 + \cdots + X_k, \quad k \geq 1 \ , \tag{3.4}$$

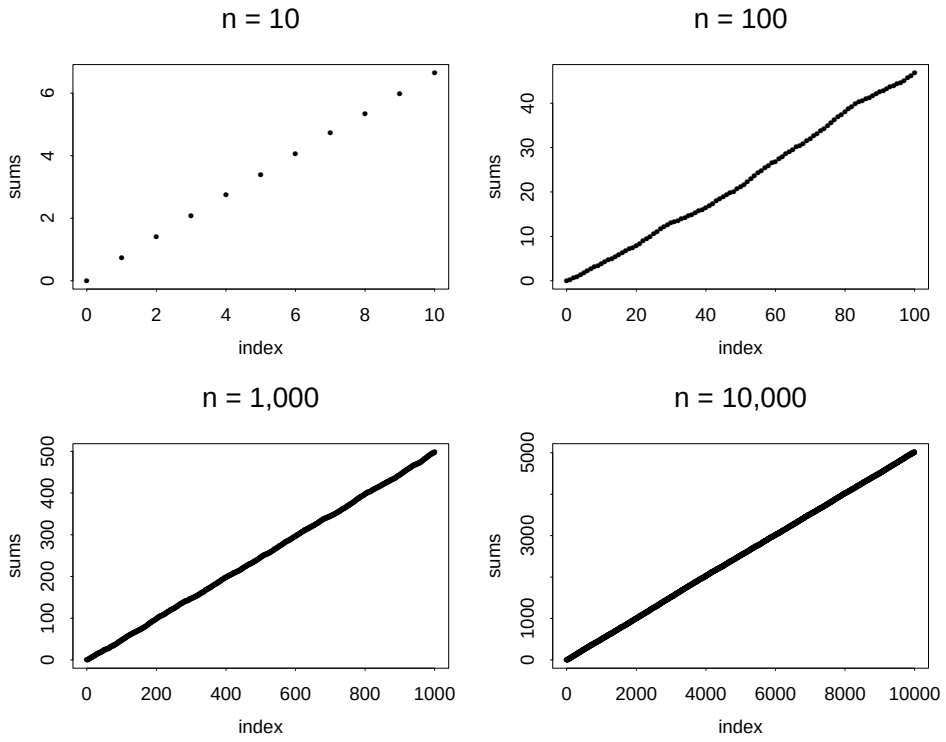

Figure 1.15: Possible realizations of the first 10^j steps of the uncentered random walk $\{S_k : k \geq 0\}$ with steps constructed by exponential smoothing, as in (3.3), for $j = 1, \ldots, 4$.

where $S_0 \equiv 0$. Next we repeat the experiments done before. We display plots of the uncentered and centered random walks with $\gamma = 0.2$ for $n = 10^j$ with $j = 1, \ldots, 4$ in Figures 1.15 and 1.16. To determine the appropriate centering constant (the steady-state mean of X_k), we solve the equation

$$E[X] = (1 - \gamma)E[X] + \gamma E[U]$$

to obtain $m \equiv E[X] = E[U] = 1/2$. Even though the distribution of X_k changes with k, the mean remains unchanged because of our choice of the initial condition.

Figures 1.15 and 1.16 look much like Figures 1.1 and 1.2 for the IID case. However, there is some significant difference for small n because the successive steps are positively correlated, causing the initial steps to be alike. However, the plots look like the previous plots for larger n. For the

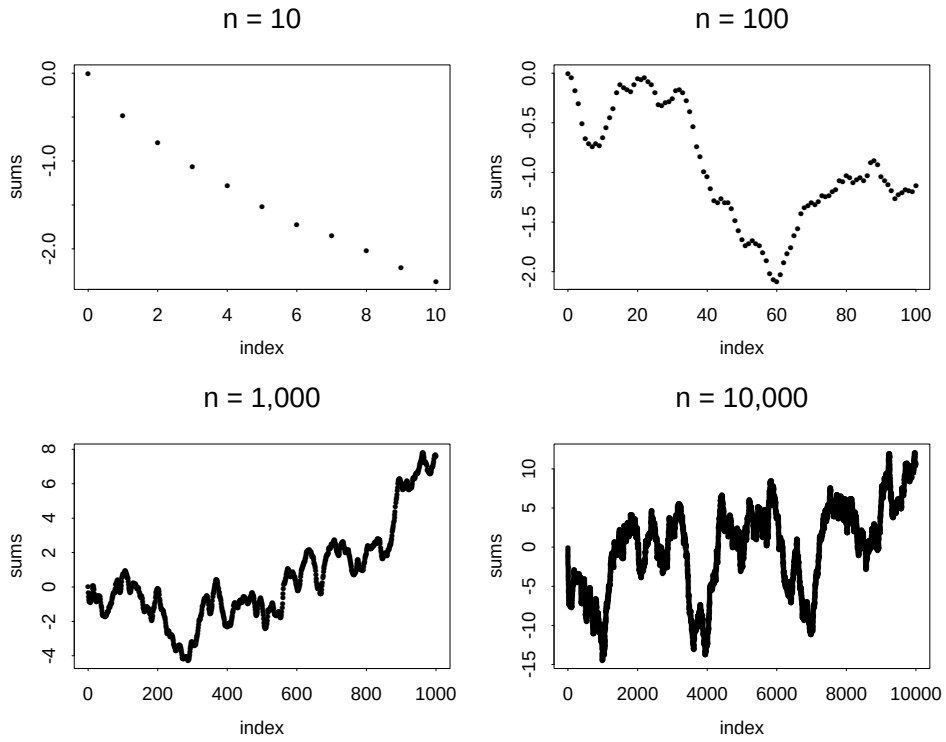

Figure 1.16: Possible realizations of the first 10^j steps of the centered random walk $\{S_k - k/2 : k \geq 0\}$ with steps constructed by exponential smoothing, as in (3.3), for $j = 1, \ldots, 4$.

centered random walks in Figure 1.16 with $n = 10^4$, what we see is again approximately a plot of Brownian motion. ∎

We can easily construct many other examples of random walks with dependent steps. For instance, we could consider a *random walk in a random environment*. A simple example has a two-state Markov-chain environment process with transition probabilities $P_{1,2} = 1 - P_{1,1} = p$ and $P_{2,1} = 1 - P_{2,2} = q$ for $0 < p < 1$ and $0 < q < 1$. We then let the k^{th} step X_k have one distribution if the Markov chain is in state 1 at the k^{th} step, and another distribution if the Markov chain is in state 2 then. We first run the Markov chain. Then, conditional on the realized states of the Markov chain, the random variables X_k are mutually independent with the appropriate distributions (depending upon the state of the Markov chain). If we consider a stationary version of the Markov chain, then the sequence $\{X_k : k \geq 1\}$ is stationary. Regardless of the initial conditions, we again see the same statistical regularity in the associated partial sums when n is sufficiently large. We invite the reader to consider such examples.

1.3.3. Different Step Distributions

Now let us return to random walks with IID steps and consider different possible step distributions. We now repeat the experiments above with various functions of the uniform random numbers, i.e., for $X_k \equiv f(U_k)$, $1 \leq k \leq n$, for different real-valued functions f. In particular, consider the following three cases:

$$\begin{aligned} &\text{(i)} \quad X_k \equiv -m\log(1-U_k) \quad \text{for} \quad m = 1,\ 10 \\ &\text{(ii)} \quad X_k \equiv U_k^p \quad \text{for} \quad p = 1/2,\ 3/2 \\ &\text{(iii)} \quad X_k \equiv U_k^{-1/p} \quad \text{for} \quad p = 1/2,\ 3/2\ . \end{aligned} \tag{3.5}$$

As before, we form partial sums associated with the new summands X_k, just as in (3.4).

Before actually considering the plots, we observe that what we are doing covers the general IID case. Given the sequence of IID random variables $\{U_k : k \geq 1\}$, by the method above we can create an associated sequence of IID random variables $\{X_k : k \geq 1\}$ where X_k has an arbitrary cdf F. Letting the left-continuous inverse of F be

$$F^{\leftarrow}(t) \equiv \inf\{s : F(s) \geq t\}, \quad 0 < t < 1\ ,$$

we can obtain the desired random variables X_k with cdf F by letting

$$X_k \equiv F^{\leftarrow}(U_k), \quad k \geq 1 \ . \tag{3.6}$$

Since

$$F^{\leftarrow}(s) \leq t \quad \text{if and only if} \quad F(t) \geq s \ , \tag{3.7}$$

we obtain

$$P(F^{\leftarrow}(U) \leq t) = P(U \leq F(t)) = F(t) \ ,$$

where U is a random variable uniformly distributed on $[0,1]$, which implies that $F^{\leftarrow}(U)$ has cdf F for any cdf F when U is uniformly distributed on $[0,1]$. For example, we see that X_k has an exponential distribution with mean m in case (i) of (3.5): If $F(t) = e^{-t/m}$, then $F^{\leftarrow}(t) = -m\log(1-t)$ and

$$P(X_k > t) = P(-m\log(1-U_k) > t) = P(1-U_k < e^{-t/m}) = e^{-t/m} \ .$$

Incidentally, we could also work with the right-continuous inverse of F, defined by

$$F^{-1}(t) \equiv \inf\{s : F(s) > t\} = F^{\leftarrow}(t+), \quad 0 < t < 1 \ ,$$

where $F^{\leftarrow}(t+)$ is the right limit at t, because

$$P(F^{-1}(U) = F^{\leftarrow}(U)) = 1 \ ,$$

since $F^{\leftarrow}$ and F^{-1} differ at, at most, countably many points.

Moreover, $F^{\leftarrow}(U_k)$, $k \geq 1$, are IID when U_k, $k \geq 1$, are IID. Of course, there also are other ways to generate IID random variables with specified distributions, but what we are doing is often a natural way.

So let us plot the uncentered and centered random walks with the step sizes in (3.5). When we do so for cases (i) and (ii), we see essentially the same pictures as before. For example, plots of the first 10^4 steps of the centered random walks in the four cases in (i) and (ii) of (3.5) are shown in Figure 1.17.

Again the plots look like plots of Brownian motion, indistinguishable from the plots for the uniform steps in Figure 1.3. Note that the units on the y axis change from plot to plot, but the plots themselves tend to have a common distribution.

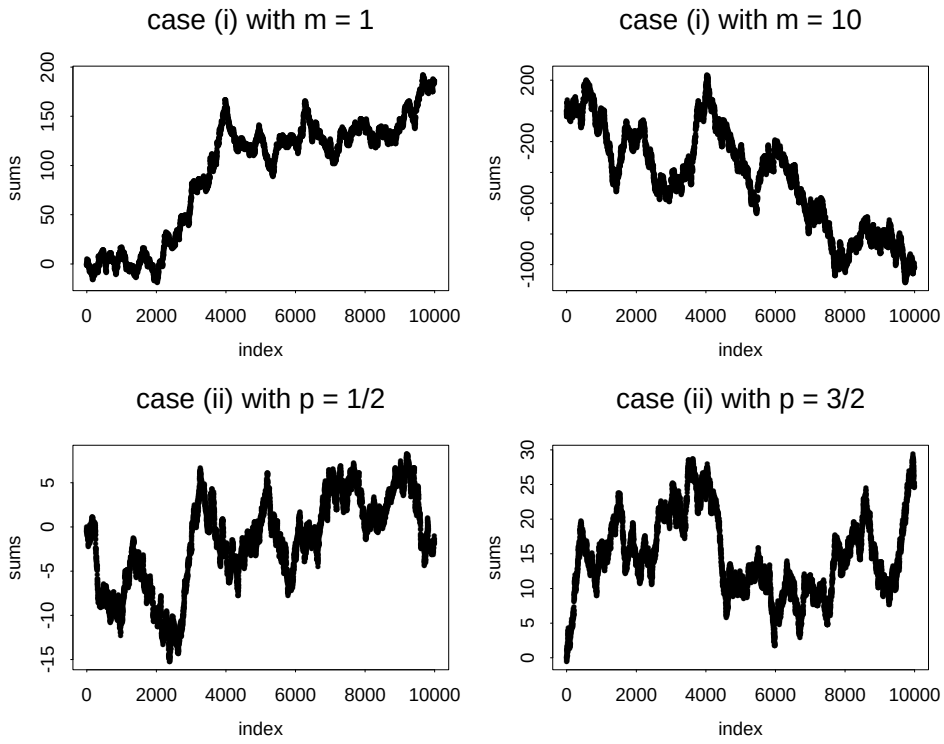

Figure 1.17: Possible realizations of the first 10^4 steps of the random walk $\{S_k - mk : k \geq 0\}$ with steps distributed as X_k in cases (i) and (ii) of (3.5).

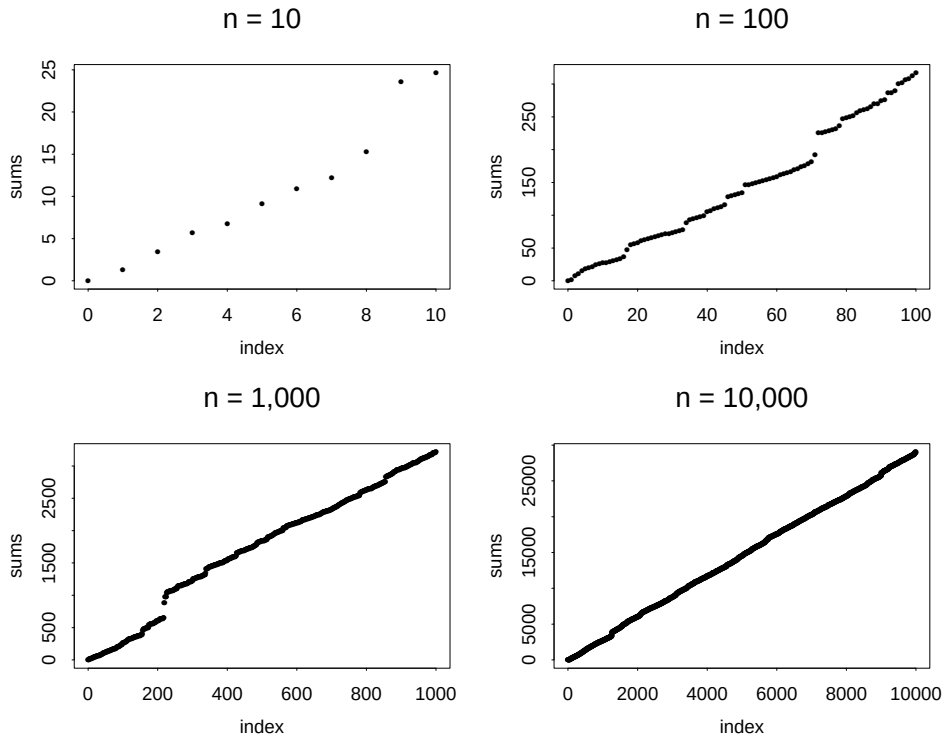

Figure 1.18: Possible realizations of the first 10^j steps of the uncentered random walk $\{S_k : k \geq 0\}$ with steps distributed as $U_k^{-1/p}$ in case (iii) of (3.5) for $p = 3/2$ and $j = 1, \ldots, 4$.

1.4. The Exception Makes the Rule

Just when boredom has begun to set in, after seeing the same thing in cases (i) and (ii) in (3.5), we should be ready to appreciate the startlingly different large-n pictures in case (iii). Plots of the uncentered random walks are plotted in Figures 1.18 and 1.19.

In the case $p = 3/2$ in Figure 1.18, the plot of the uncentered random walk is again approaching a line as $n \to \infty$, but not as rapidly as before. (Again we ignore the units on the axes when we look at the plots.) However, in the case $p = 1/2$ in Figure 1.19 we something radically different: For large n, the plots have *jumps*!

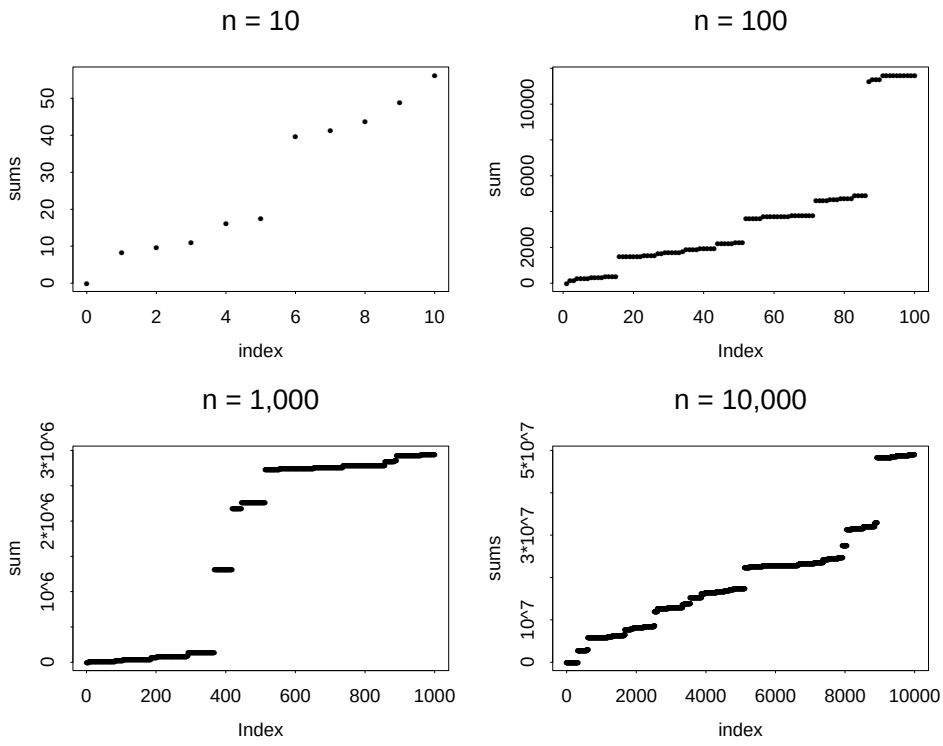

Figure 1.19: Possible realizations of the first 10^j steps of the uncentered random walk $\{S_k : k \geq 0\}$ with steps distributed as $U_k^{-1/p}$ in case (iii) of (3.5) for $p = 1/2$ and $j = 1, \ldots, 4$.

1.4.1. Explaining the Irregularity

Fortunately, probability theory again provides an explanation for the *irregularity* that we now see: The SLLN states, under the prevailing IID assumptions, that scaled partial sums $n^{-1}S_n$ will approach the mean EX_1 w.p.1 as $n \to \infty$, regardless of other properties of the probability distribution of X_1, *provided that a finite mean exists.* Knowing the SLLN, we should expect to see lines when $n = 10^4$ in all experiments except possibly in case (iii).

We might initially be fooled in case (iii), but we should anticipate occasional large steps because $U^{-1/p}$ involves *dividing* by very small values when U is small. Upon more careful examination, we see that $U^{-1/p}$ has a *Pareto distribution* with parameter p, which we refer to as Pareto(p), when U is uniformly distributed on $[0, 1]$, i.e.,

$$P(U^{-1/p} > t) = P(U < t^{-p}) = t^{-p}, \quad t \geq 1\,, \tag{4.1}$$

with mean

$$E(U^{-1/p}) = \int_0^\infty P(U^{-1/p} > t)dt = 1 + \int_1^\infty t^{-p}dt\ , \tag{4.2}$$

which is finite, and equal to $1 + (p-1)^{-1}$, if and only if $p > 1$; see Chapter 19 of Johnson and Kotz (1970) for background on the Pareto distribution and Lemma 1 on p. 150 of Feller (1971) for the integral representation of the mean.

Thus the SLLN tells us not to expect the same behavior observed in the previous experiments in case (iii) when $p \leq 1$. Thus, unlike all previous random walks considered, the conditions of the SLLN are *not satisfied* in case (iii) with $p = 1/2$.

Now let us consider the random walk with Pareto(p) steps for $p = 3/2$ in (3.5) (iii). Consistent with the SLLN, Figure 1.18 shows that the plots are approaching a straight line as $n \to \infty$ in this case. But what happens when we center?

1.4.2. The Centered Random Walk with $p = 3/2$

So now let us consider the centered random walk in case (iii) with $p = 3/2$. (Since the mean is infinite when $p = 1/2$, we cannot center when $p = 1/2$. We will return to the case $p = 1/2$ later.) We center by subtracting the mean, which in the case $p = 3/2$ is $1+(p-1)^{-1} = 3$. Plots of the centered random walk with $p = 3/2$ for $n = 10^j$ with $j = 1, 2, 3, 4$ are shown in Figure

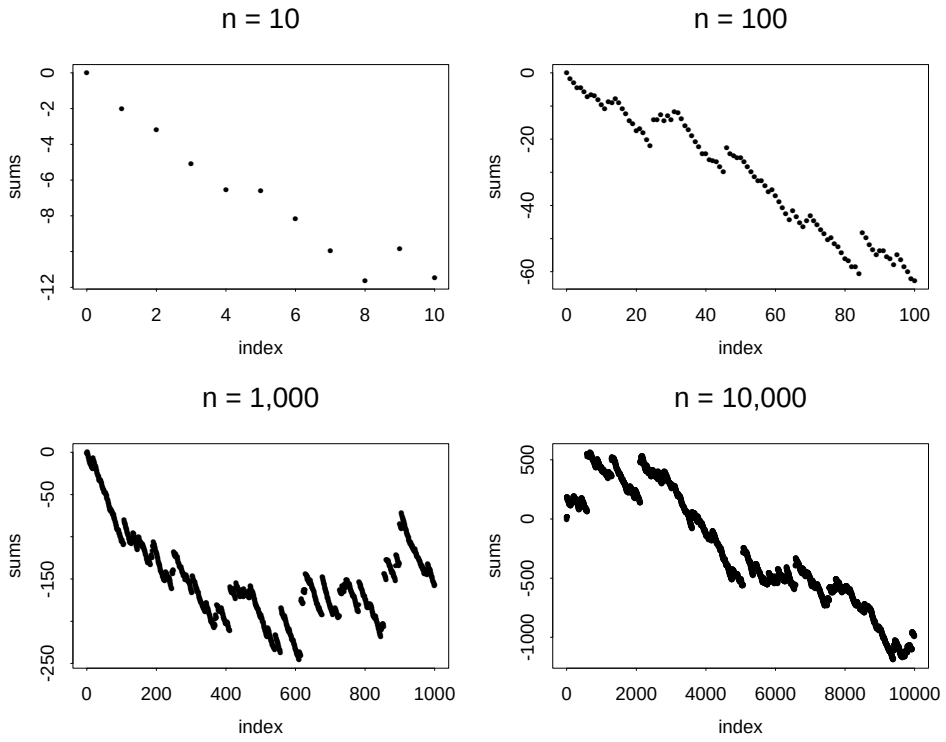

Figure 1.20: Possible realizations of the first 10^j steps of the centered random walk $\{S_k - 3k : k \geq 0\}$ associated with the Pareto steps $U_k^{-1/p}$ for $p = 3/2$, having mean 3 and infinite variance, for the cases $j = 1, \ldots, 4$.

1.20. As before, the centering causes the plotter to automatically blow up the picture. However, now the slight departures from linearity for large n in Figure 1.18 are magnified. Now, just as in Figure 1.19, we see jumps in the plot!

Once again, probability theory offers an explanation. Just as the SLLN ceases to apply when the IID summands have infinite mean, so does the (classical) CLT cease to apply when the IID summands have finite mean but infinite variance. Such a case occurs with the Pareto(p) summands in case (iii) in (3.5) when $1 < p \leq 2$. Thus, consistent with what we see in Figure 1.18, the SLLN holds, but the CLT does not, for the Pareto(p) random variable $U^{-1/p}$ in case (iii) when $p = 3/2$.

We have arrived at another critical point, where an important intellectual step is needed. We need to recognize that, *even though the sample paths are*

very different from the previous random-walk plots, which are approaching plots of Brownian motion, there may still be important statistical regularity in the new plots with jumps.

To see the statistical regularity, we need to repeat the experiment and consider larger values of n. Even though the plots look quite different from the previous random-walk plots, we can see statistical regularity in the plots (again ignoring the units on the axes). To confirm that observation, six possible realizations for $p = 3/2$ in the cases $n = 10^4$ and $n = 10^6$ are shown in Figures 1.21 and 1.22. Figures 1.21 and 1.22 show more irregular paths, but with their own distinct character, much like handwriting. (We might contemplate the probability of the path writing a word. With a suitable font for the script, we might see "Null" but not "Set".) Again, Figures 1.21 and 1.22 show that there is statistical regularity associated with the irregularity we see. The plots are independent of n for all n sufficiently large. Again we see self-similarity in the plots.

Even though the irregular paths in Figures 1.19 – 1.22 have jumps, as before we can look for statistical regularity through the distribution of these random paths. Again, to be able to see something, we can focus on the final positions. Focusing first on the case with $p = 3/2$, we plot the estimated density of the centered sums $S_n - 3n$ for $n = 1,000$. Once again, we obtain the density estimate by performing independent replications of the experiment. To have more data this time, we use 10,000 independent replications. We display the resulting density estimate in Figure 1.23.

When we look at the estimated density of the final position, we see that it is radically different from the previous density plots in Figures 1.5 and 1.7. Clearly, *the final position is no longer normally distributed!*

Nevertheless, there is statistical regularity. As before, when we repeat the experiment with different random number seeds, we obtain essentially the same result for all sufficiently large n. Examination shows that there is statistical regularity, just as before, but the approximating distribution of the final position is now different. In Figure 1.23, the peak of the density looks like a spike because the range of values is now much greater. In turn, the range of values is greater because the distribution of $S_n - 3n$ has a heavy tail.

The heavier tails are more clearly revealed when we plot the tail of the empirical cdf of the observed values. (By the tail of a cdf F, we mean the *complementary cdf or ccdf*, defined by $F^c(t) \equiv 1 - F(t)$.)

To focus on the tail of the cdf F, we plot the tail of the empirical cdf in $log - log$ scale in Figure 1.18; i.e., we plot $\log F^c(t)$ versus $\log t$. To use $log - log$ scale, we consider only those values greater than 1, of which there

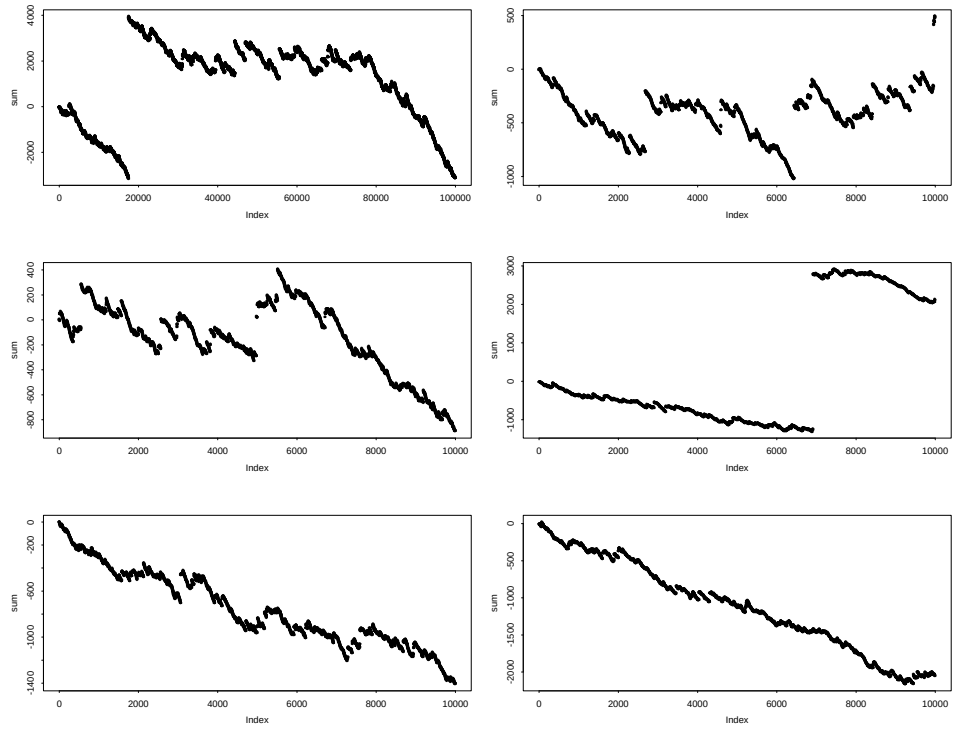

Figure 1.21: Six independent realizations of the first 10^4 steps of the centered random walk $\{S_k - 3k : k \geq 0\}$ associated with the Pareto steps $U_k^{-1/p}$ for $p = 3/2$, having mean 3 and infinite variance.

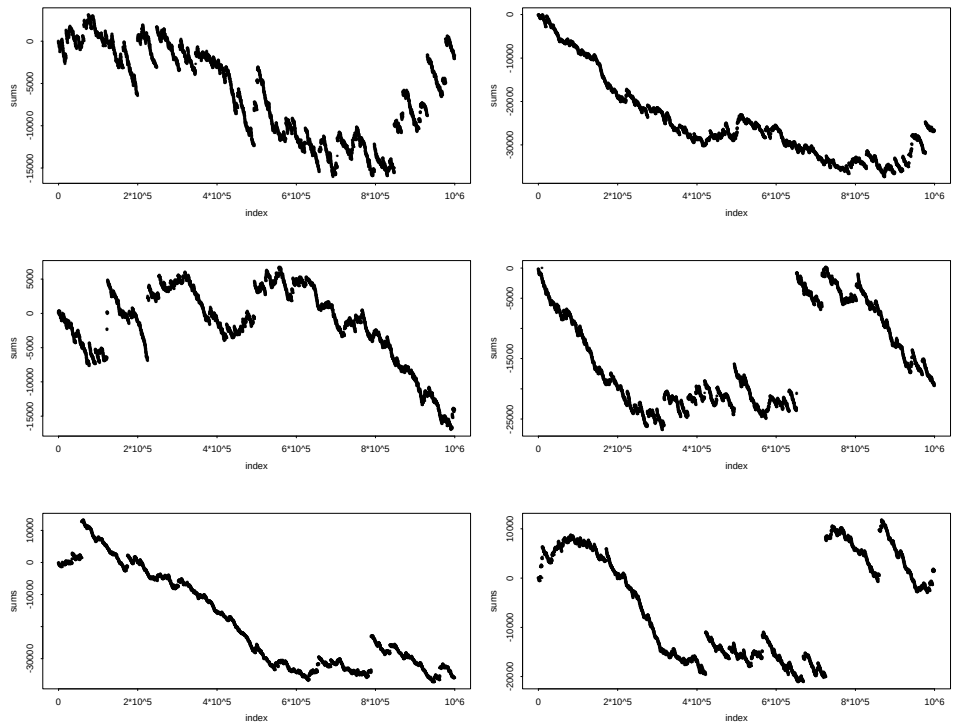

Figure 1.22: Six independent realizations of the first 10^6 steps of the centered random walk $\{S_k - 3k : k \geq 0\}$ associated with the Pareto steps $U_k^{-1/p}$ for $p = 3/2$, having mean 3 and infinite variance.

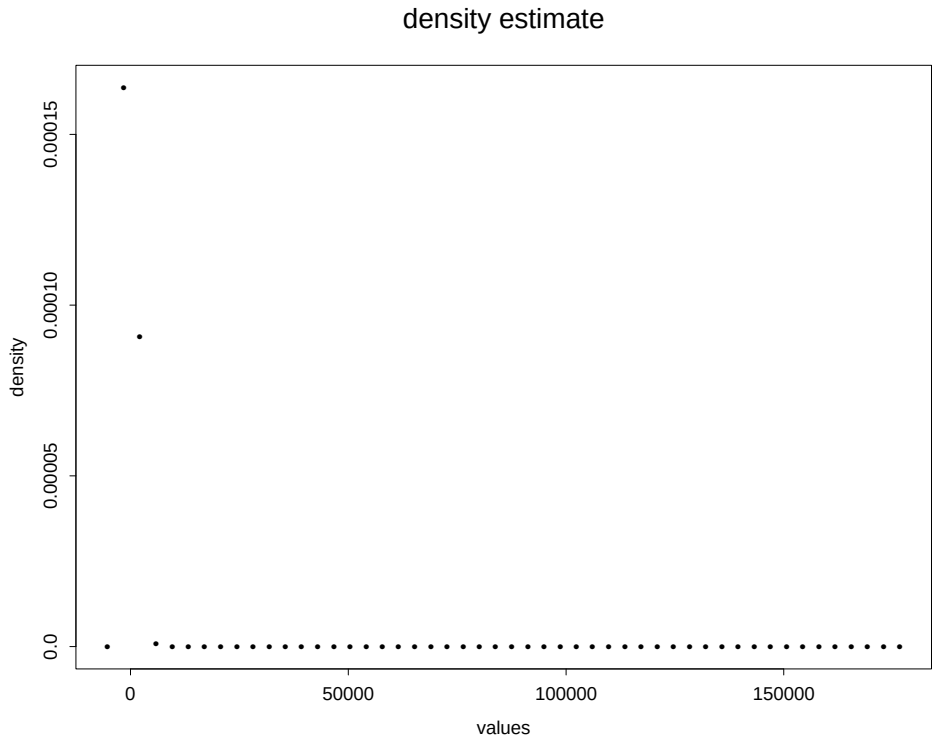

Figure 1.23: The density estimate obtained from $10,000$ independent samples of the final position of the centered random walk (i.e., the centered partial sum $S_{1000} - 3000$) associated with the Pareto steps $U_k^{-1/p}$ for $p = 3/2$.

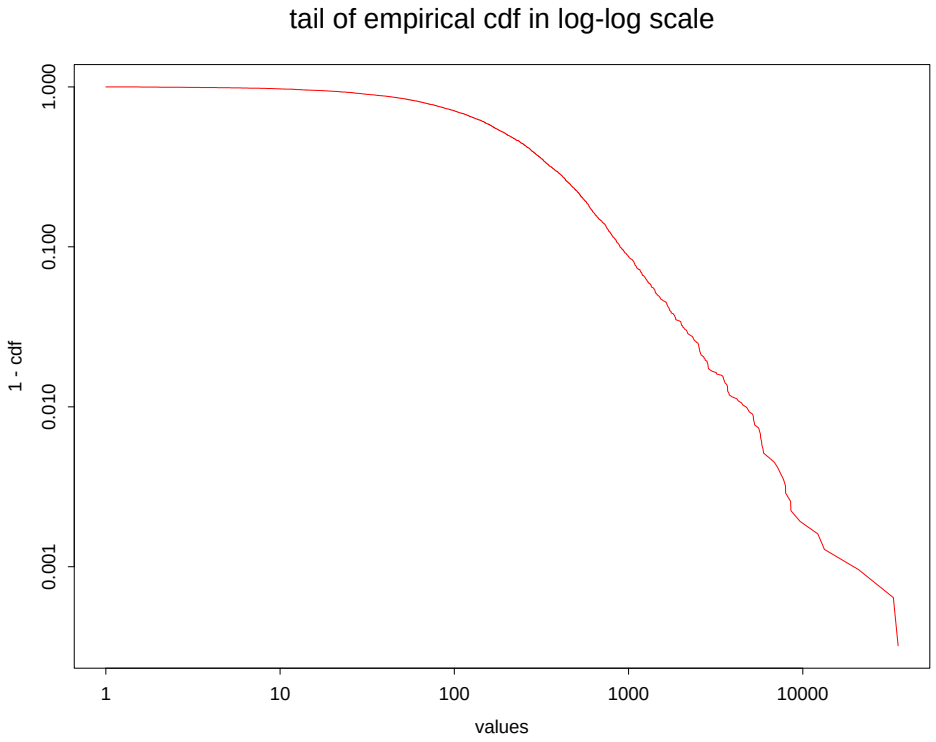

Figure 1.24: The tail of the empirical distribution function in $log - log$ scale obtained from $10,000$ independent samples of the final position of the centered random walk (i.e., the partial sum $S_{1000} - 3000$) associated with the Pareto steps $U_k^{-1/p}$ for $p = 3/2$ corresponding to the density in Figure 1.23. The results are based on the $3,121$ values greater than 1.

were 3,121 when $n = 10^4$.

From Figure 1.24, we see that for larger values of the argument t, the empirical ccdf has a linear slope in $log - log$ scale. That indicates a *power tail*. Indeed, if the ccdf is of the form

$$F^c(t) = \alpha t^{-\beta} \quad \text{for} \quad t \geq t_0 > 1 \ , \tag{4.3}$$

then

$$\log F^c(t) = -\beta \log t + \log \alpha \tag{4.4}$$

for $t > t_0$. Then the paremeters α and β in (4.3) can be seen as the intercept and slope in the $log - log$ plot.

Again there is supporting theory: A generalization of the CLT implies, under the IID assumptions and other regularity conditions (satisfied here), that properly scaled versions of the centered partial sums of Pareto(p) random steps converge in distribution, as in (2.7). In particular, when $1 < p < 2$,

$$n^{-1/p}(S_n - mn) \Rightarrow L \quad \text{in} \quad \mathbb{R} \,, \tag{4.5}$$

where $m = 1 + (p-1)^{-1}$ is the mean and the limiting random variable L has a *non-Gaussian stable law* (depending upon p); e.g., see Chapter XVII of Feller (1971). In our specific case of $p = 3/2$, we have space scaling by $n^{2/3}$.

Unlike the Pareto distribution, the limiting stable law is not a pure power, but it has a power tail; i.e., it is asymptotically equivalent to a power: for $1 < p < 2$,

$$P(L > t) \sim ct^{-p} \quad \text{as} \quad t \to \infty \tag{4.6}$$

for some positive constant c, where $f(t) \sim g(t) \quad \text{as} \quad t \to \infty$ means that f is *asymptotically equivalent* to g, i.e., $f(t)/g(t) \to 1 \quad \text{as} \quad t \to \infty$. Thus the tail of the limiting stable law has the same asymptotic decay rate as the Pareto distribution of a single step.

Unlike the standard CLT in (2.4), the space scaling in (4.5) involves $c_n = n^{1/p}$ for $1 < p < 2$ instead of $c_n = n^{1/2}$. Nevertheless, the generalized CLT shows that there is again remarkable statistical regularity in the centered partial sums when the mean is finite and the variance is infinite. We again obtain essentially the same probability distribution for all n. We also obtain essentially the same probability distribution for other nonnegative step distributions, provided that they are centered by subtracting the finite mean, and that the step-size ccdf $F^c(t)$ has the same asymptotic tail; i.e., we require that

$$F^c(t) \sim ct^{-p} \quad \text{as} \quad t \to \infty \tag{4.7}$$

for some positive constant c.

As before, there is also an associated stochastic-process limit. A generalization of Donsker's theorem (the FCLT) implies that the sequence of scaled random walks with Pareto(p) steps having $1 < p < 2$ converges in distribution to a *stable Lévy motion* as $n \to \infty$ in D. Now

$$\mathbf{S}_n \Rightarrow \mathbf{S} \quad \text{in} \quad D \,, \tag{4.8}$$

where

$$\mathbf{S}_n(t) \equiv n^{-1/p}(S_{\lfloor nt \rfloor} - m\lfloor nt \rfloor), \quad 0 \le t \le 1 \,, \tag{4.9}$$

for $n \geq 1$, m is the mean and $\mathbf{S}$ is a stable Lévy motion. That is, the stochastic-process limit (2.3) holds for $\mathbf{S}_n$ in (2.2), but now with $c_n = n^{1/p}$ and the limit process $\mathbf{S}$ being stable Lévy motion instead of Brownian motion. Moreover, a variant of the previous invariance property holds here as well. For nonnegative random variables (the step sizes) satisfying (4.7), the limit process depends on its distribution only through the decay rate p and the single parameter c appearing in (4.7). We discuss this FCLT further in Chapter 4.

Since the random walk steps are IID, it is evident that the limiting stable Lévy motion must have stationary and independent increments, just like Brownian motion. However, the marginal distributions in $\mathbb{R}$ or $\mathbb{R}^k$ are non-normal stable laws instead of the normal laws. Moreover, the stable Lévy motion has the self-similarity property, just like Brownian motion, but now with a different scaling. Now, for any $c > 0$, the stochastic process $\{c^{-1/p}\mathbf{S}(ct) : 0 \leq t \leq 1\}$ has a probability law on D, and thus finite-dimensional distributions, that are independent of c. Indeed, the proof is just like the proof for Brownian motion in (2.18).

It is significant that the space scaling to achieve statistical regularity is different now. In (4.9) above, we divide by $n^{1/p}$ for $1 < p < 2$ instead of by $n^{1/2}$. Similarly, in the self-similarity of the stable Lévy motion, we multiply by $c^{-1/p}$ instead of $c^{-1/2}$. The new scaling can be confirmed by looking at the values on the y-axis in the plots of Figures 1.20–1.22.

Figures 1.20–1.22 show that, unlike Brownian motion, stable Lévy motion must have *discontinuous sample paths*. Hence, *we have a stochastic-process limit in which the limit process has jumps*. The desire to consider such stochastic-process limits is a primary reason for this book.

1.4.3. Back to the Uncentered Random Walk with $p = 1/2$

Now let us return to the first Pareto(p) example with $p = 1/2$. The plots in Figure 1.19 are so irregular that we might not suspect that there is any statistical regularity there. However, after seeing the statistical regularity in the case $p = 3/2$, we might well think about reconsidering the case $p = 1/2$.

As before, we investigate by making some more plots. We have noted that we cannot center because the mean is ininite. So let us make more plots of the uncentered random walk with $p = 1/2$. Thus, in Figure 1.25 we plot six independent realizations of the uncentered random walk with 10^4 Pareto(0.5) steps. Now, even though these plots are highly irregular, with a single jump sometimes dominating the entire plot, we see remarkable statistical regularity.

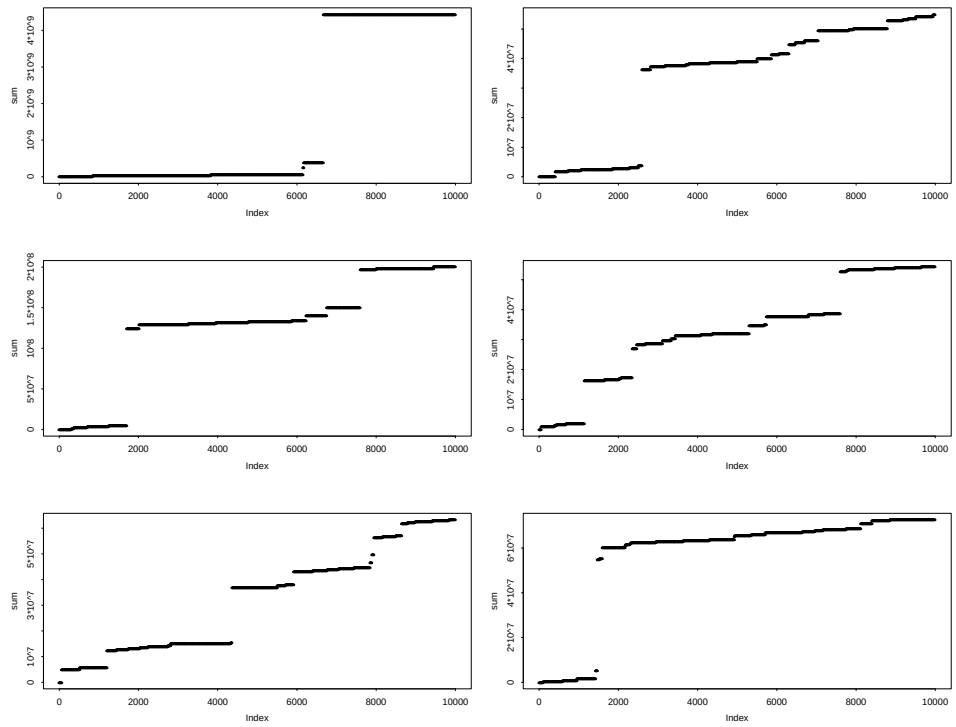

Figure 1.25: Six independent possible realizations of the first 10^4 steps of the uncentered random walk $\{S_k : k \geq 0\}$ with steps distributed as $U_k^{-1/p}$ in case (iii) of (3.5) for $p = 1/2$.

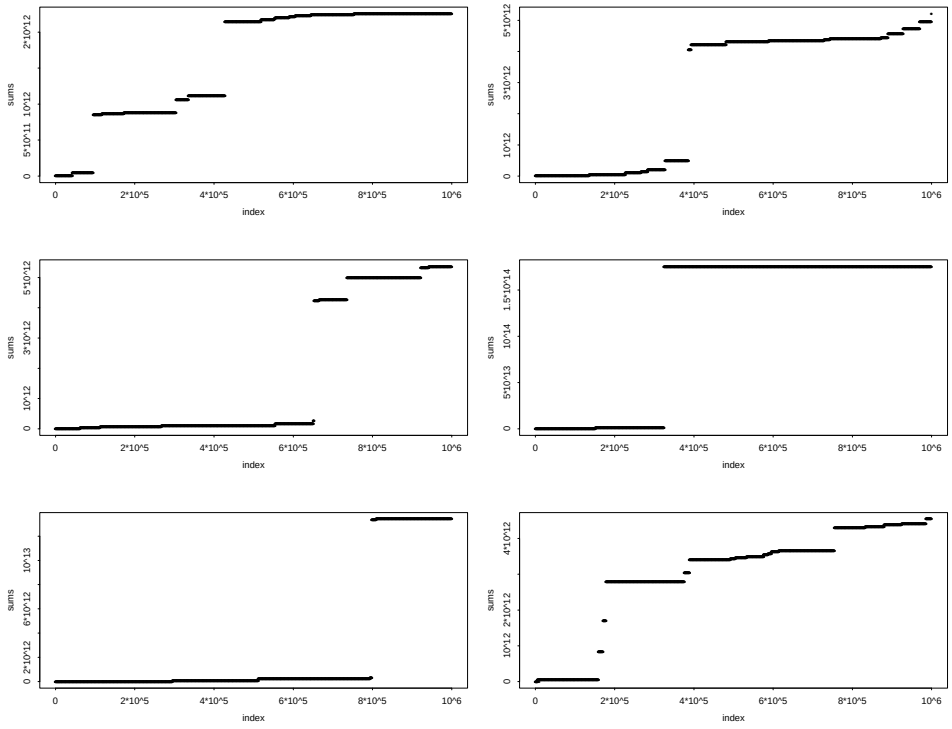

Figure 1.26: Six independent possible realizations of the first 10^6 steps of the uncentered random walk $\{S_k : k \geq 0\}$ with steps distributed as $U_k^{-1/p}$ in case (iii) of (3.5) for $p = 1/2$.

Paralleling Figures 1.4 and 1.22, we confirm what we see in Figure 1.25 by plotting six independent samples of the uncentered random walk in case (iii) with $p = 1/2$ for $n = 10^6$ in Figure 1.26. Even though the plots of the uncentered random walks with Pareto(0.5) steps in Figures 1.19 – 1.26 are radically different from the previous plots of centered and uncentered random walks, we see remarkable statistical regularity in the new plots. As before, the plots tend to be independent of n for all n sufficiently large, provided we ignore the units on the axes. Thus we see self-similarity, just as in the plots of the centered random walks before. *From the random-walk plots, we see that statistical regularity can occur in many different forms.*

Given what we have just done, it is natural to again look for statistical regularity in the final positions. Thus we consider the final positions S_n (without centering) for $n = 1000$ and perform 10,000 independent replica-

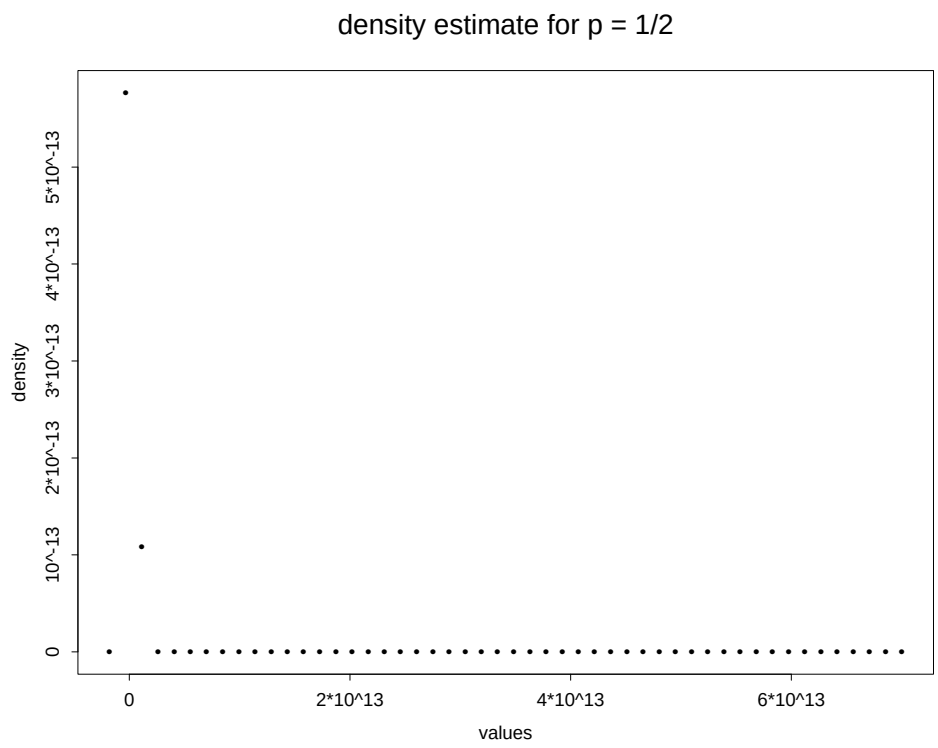

Figure 1.27: A density estimate obtained from 10,000 independent samples of the final position of the uncentered random walk (i.e., the partial sum S_{1000}) associated with the Pareto steps $U_k^{-1/p}$ in the case $p = 1/2$.

tions. Paralleling Figures 1.23 and 1.24 above, an estimate of the probability density and the tail of the empirical cdf are plotted in Figures 1.27 and 1.28 below.

Figures 1.27 and 1.28 are quite similar to Figures 1.23 and 1.24, but now the distribution has an even heavier tail. Again there is supporting theory: A generalization of the CLT states, under the IID assumptions and other regularity conditions (satisfied here), that for $0 < p < 1$ there is convergence in distribution of the *uncentered partial sums* to a non-Gaussian stable law if the partial sums are scaled appropriately, which requires that $c_n = n^{1/p}$. In particular, now with $p = 1/2$,

$$n^{-1/p} S_n \Rightarrow L \quad \text{in} \quad \mathbb{R}\,, \tag{4.10}$$

where the limiting random variable again has a a non-Gaussian stable law,

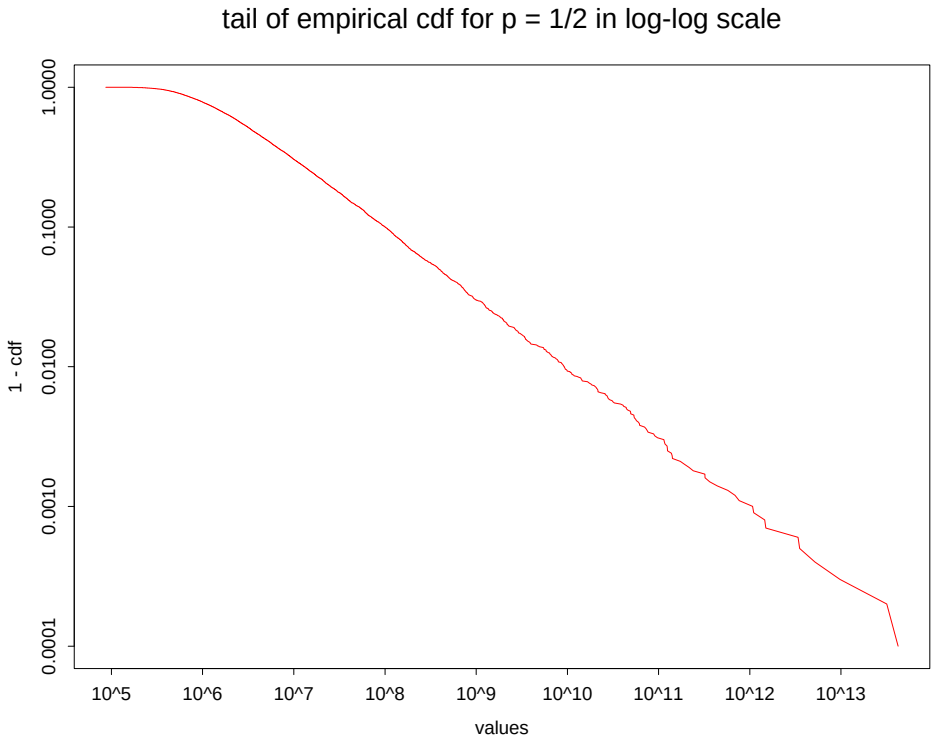

Figure 1.28: The tail of the empirical cumulative distribution function in $log-log$ scale obtained from $10,000$ independent samples of the final position of the uncentered random walk (i.e., the partial sum S_{1000}) associated with the Pareto steps $U_k^{-1/p}$ for $p = 1/2$ corresponding to the density in Figure 1.27.

which has an asymptotic power tail, i.e.,

$$P(L > t) \sim ct^{-p} \quad \text{as} \quad t \to \infty \tag{4.11}$$

for $p = 1/2$ and some positive constant c; again see Chapter XVII of Feller (1971). As before, the tail of the stable law has the same asymptotic decay rate as a single step of the random walk.

Moreover, there again is an associated stochastic-process limit. Another generalization of Donsker's FCLT implies that there is the stochastic-process limit (4.8), where

$$\mathbf{S}_n(t) \equiv n^{-1/p} S_{\lfloor nt \rfloor}, \quad 0 \le t \le 1 \ , \tag{4.12}$$

for $n \ge 1$, with the limit process $\mathbf{S}$ being another stable Lévy motion depending upon p.

Again there is an invariance property: Paralleling (4.7), we require that the random-walk step ccdf F^c satisfy

$$F^c(t) \sim ct^{-p} \quad \text{as} \quad t \to \infty \ , \tag{4.13}$$

where $p = 1/2$ and c is some positive constant. Any random walk with nonnegative (IID) steps having a ccdf satisfying (4.13) will satisfy the same FCLT, with the limit process depending on the step-size distribution only through the decay rate $p = 1/2$ and the constant c in (4.13).

As before, the plotter automatically does the proper scaling. However, the space scaling is different from both the previous two cases, now requiring division by $n^{1/p}$ for $p = 1/2$. Again, we can verify that the space scaling by $n^{1/p}$ is appropriate by looking at the values in the plots in Figures 1.19–1.26. Just as before, the stochastic-process limit in D implies that the limit process must be self-similar. Now, for any $c > 0$, the stochastic processes $\{c^{-1/p}\mathbf{S}(ct) : 0 \le t \le 1\}$ have probability laws in D that are independent of c.

Figures 1.19 and 1.25 show that the limiting stable Lévy motion for the case $p = 1/2$ must also have discontinuous sample paths. So we have yet another stochastic-process limit in which the limit process has jumps.

1.5. Summary

To summarize, in this chapter we have seen that there is remarkable statistical regularity associated with random walks as the number n of steps increases. That statistical regularity is directly revealed when we plot the

random walks. In great generality, as a consequence of Donsker's theorem, properly scaled versions of the centered random walks converge in distribution to Brownian motion as n increases. As a consequence, the random-walk plots converge to plots of standard Brownian motion.

The great generality of that result may make us forget that there are conditions for convergence to Brownian motion to hold. Through the exponential-smoothing example, we have seen that the conclusions of the classical limit theorems often still hold when the IID conditions are relaxed, but again there are limitations on the amount of dependence that can be allowed. That is easy to see by considering the extreme case in which *all the steps are identical*! Clearly, then the SLLN and the CLT break down. The classical limit theorems tend to remain valid when independence is replaced by *weak dependence*, but it is difficult to characterize the boundary exactly. We discuss FCLTs for weakly dependent sequences further in Chapter 4.

We also have seen for the case of IID steps that there are important situations in which the conditions of the FSLLN and Donsker's FCLT do not hold. We have seen that these fundamental theorems are not valid in the IID case when the step-size distribution has infinite mean (the FSLLN) or variance (the FCLT). Nevertheless, there often is remarkable statistical regularity associated with these heavy-tailed cases, but the limit process in the stochastic-process limit becomes a stable Lévy motion, which has jumps, i.e., it has discontinuous sample paths. We have thus seen examples of stochastic-process limits in which the limit process has jumps. We discuss such FCLTs further in Chapter 4.

If we allow greater dependence, which may well be appropriate in applications, then many more limit processes are possible, some of which will again have discontinuous sample paths. Again, see Chapter 4.