6

Simulation

Sometimes probabilistic models are so complex that the tools of mathematical analysis are not sufficient to answer all relevant questions about them. Stochastic simulation is an alternative approach: values are generated for the random variables and inserted into the model, thus mimicking outcomes for the whole system. It is shown in this chapter how one can use uniform random number generators to mimic random variables. Also two larger simulation examples are presented.

6.1 What is simulation?

In many areas of science, technology, government, and business, models are used to gain understanding of some part of reality (the portion of interest is often referred to as "the system"). Sometimes these are physical models, such as a scale model of an airplane in a wind tunnel or a scale model of a chemical plant. Other models are abstract, such as macroeconomic models consisting of equations relating things like interest rates, unemployment, and inflation or partial differential equations describing global weather patterns.

In *simulation*, one uses a model to create specific situations in order to study the response of the model to them and then interprets this in terms of what would happen to the system "in the real world." In this way, one can carry out experiments that are impossible, too dangerous, or too expensive to do in the real world—addressing questions like: What happens to the average temperature if we reduce the greenhouse gas emissions globally by 50%? Can the plane still fly if engines 3 and 4 stop in midair? What happens to the distribution of wealth if we halve the tax rate?

More specifically, we focus on situations and problems where randomness or uncertainty or both play a significant or dominant role and should be modeled explicitly. Models for such systems involve random variables, and we speak of *probabilistic* or *stochastic models*. Simulating them is *stochastic simulation*. In the preceding chapters we have encountered some of the tools of probability theory, and we will encounter others in the chapters to come. With these tools we can compute quantities of interest explicitly for many models. Stochastic simulation of a system means generating values for all the random variables in the model, according to their specified distributions, and recording and analyzing what happens. We refer to the generated values as *realizations* of the random variables.

For us, there are two reasons to learn about stochastic simulation. The first is that for complex systems, simulation can be an alternative to mathematical analysis, sometimes the only one. The second reason is that through simulation, we can get more feeling for random variables, and this is why we study stochastic simulation at this point in the book. We start by asking how we can generate a realization of a random variable.

6.2 Generating realizations of random variables

Simulations are almost always done using computers, which usually have one or more so-called *(pseudo) random number generators*. A call to the random number generator returns a random number between 0 and 1, which mimics a realization of a $U(0,1)$ variable. With this source of uniform (pseudo) randomness we can construct any random variable we want by transforming the outcome, as we shall see.

QUICK EXERCISE 6.1 Describe how you can simulate a coin toss when instead of a coin you have a die. Any ideas on how to simulate a roll of a die if you only have a coin?

Bernoulli random variables

Suppose U has a $U(0,1)$ distribution. To construct a $Ber(p)$ random variable for some $0 < p < 1$, we define

$$X = \begin{cases} 1 & \text{if } U < p, \\ 0 & \text{if } U \geq p \end{cases}$$

so that

$$\begin{aligned} \mathrm{P}(X = 1) &= \mathrm{P}(U < p) = p, \\ \mathrm{P}(X = 0) &= \mathrm{P}(U \geq p) = 1 - p. \end{aligned}$$

This random variable X has a Bernoulli distribution with parameter p.

QUICK EXERCISE 6.2 A random variable Y has outcomes 1, 3, and 4 with the following probabilities: $\mathrm{P}(Y = 1) = 3/5$, $\mathrm{P}(Y = 3) = 1/5$, and $\mathrm{P}(Y = 4) = 1/5$. Describe how to construct Y from a $U(0,1)$ random variable.

Continuous random variables

Suppose we have the distribution function F of a continuous random variable and we wish to construct a random variable with this distribution. We show how to do this if F is strictly increasing from 0 to 1 on an interval. In that case F has an inverse function F^{inv}. Figure 6.1 shows an example: F is strictly increasing on the interval $[2, 10]$; the inverse F^{inv} is a function from the interval $[0, 1]$ to the interval $[2, 10]$.

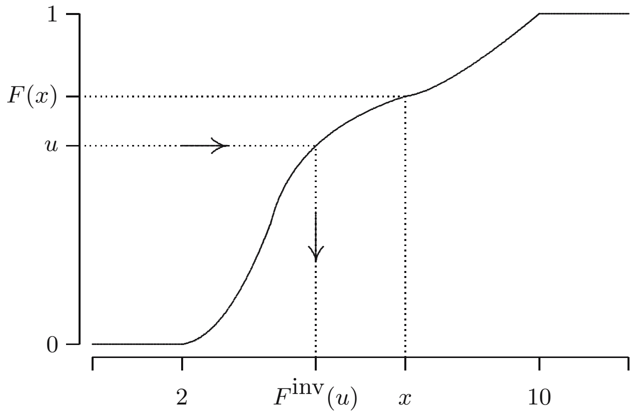


Fig. 6.1. Simulating a continuous random variable using the distribution function.

Note how u relates to $F^{\mathrm{inv}}(u)$ as $F(x)$ relates to x. We see that $u \leq F(x)$ is equivalent with $F^{\mathrm{inv}}(u) \leq x$. If instead of a real number u we consider a $U(0, 1)$ random variable U, we obtain that the corresponding events are the same:

$$\{U \leq F(x)\} = \{F^{\mathrm{inv}}(U) \leq x\}. \tag{6.1}$$

We know about the $U(0, 1)$ random variable U that $\mathrm{P}(U \leq b) = b$ for any number $0 \leq b \leq 1$. Substituting $b = F(x)$ we see

$$\mathrm{P}(U \leq F(x)) = F(x).$$

From equality (6.1), therefore,

$$\mathrm{P}\big(F^{\mathrm{inv}}(U) \leq x\big) = F(x);$$

in other words, the random variable $F^{\mathrm{inv}}(U)$ has distribution function F.

What remains is to find the function F^{inv}. From Figure 6.1 we see

$$F(x) = u \quad \Leftrightarrow \quad x = F^{\mathrm{inv}}(u),$$

so if we solve the equation $F(x) = u$ for x, we obtain the expression for $F^{\mathrm{inv}}(u)$.

Exponential random variables

We apply this method to the exponential distribution. On the interval $[0, \infty)$, the *Exp*(λ) distribution function is strictly increasing and given by

$$F(x) = 1 - \mathrm{e}^{-\lambda x}.$$

To find F^{inv}, we solve the equation $F(x) = u$:

$$\begin{aligned} F(x) = u \quad &\Leftrightarrow \quad 1 - \mathrm{e}^{-\lambda x} = u \\ &\Leftrightarrow \quad \mathrm{e}^{-\lambda x} = 1 - u \\ &\Leftrightarrow \quad -\lambda x = \ln(1-u) \\ &\Leftrightarrow \quad x = -\frac{1}{\lambda}\ln(1-u), \end{aligned}$$

so $F^{\mathrm{inv}}(u) = -\frac{1}{\lambda}\ln(1-u)$ and if U has a $U(0,1)$ distribution, then the random variable X defined by

$$X = F^{\mathrm{inv}}(U) = -\frac{1}{\lambda}\ln(1-U)$$

has an *Exp*(λ) distribution.

In practice, one replaces $1-U$ with U, because both have a $U(0,1)$ distribution (see Exercise 6.3). Leaving out the subtraction leads to more efficient computer code. So instead of X we may use

$$Y = -\frac{1}{\lambda}\ln(U),$$

which also has an *Exp*(λ) distribution.

Quick exercise 6.3 A distribution function F is 0 for $x < 1$ and 1 for $x > 3$, and $F(x) = \frac{1}{4}(x-1)^2$ if $1 \leq x \leq 3$. Let U be a $U(0,1)$ random variable. Construct a random variable with distribution F from U.

Remark 6.1 (The general case). The restriction we imposed earlier, that the distribution function should be strictly increasing, is not really necessary. Furthermore, a distribution function with jumps or a flat section somewhere in the middle is not a problem either. We illustrate this with an example in Figure 6.2.

This F has a jump at 4 and so for a corresponding X we should have $\mathrm{P}(X = 4) = 0.2$, the size of the jump. We see that whenever U is in the interval $[0.3, 0.5]$, it is mapped to 4 by our method, and that this happens with exactly the right probability!

The flat section of F between 7 and 8 seems to pose a problem: the equation $F(a) = 0.85$ has as its solution any a between 7 and 8, and we cannot define a unique inverse. This, however, does not really matter, because $\mathrm{P}(U = 0.85) = 0$, and we can define the inverse $F^{\mathrm{inv}}(0.85)$ in any way we want. Taking the left endpoint, here the number 7, agrees best with the definition of quantiles (see page 66).

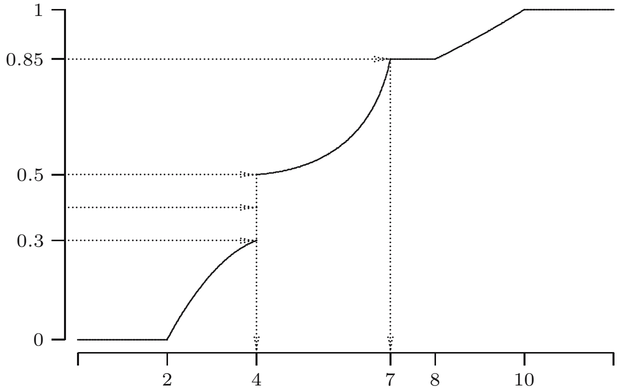


Fig. 6.2. A distribution function with a jump and a flat section.

Remark 6.2 (Existence of random variables). The previous remark supplies a sketchy argument for the fact that any nondecreasing, rightcontinuous function F, with $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$, is the distribution of some random variable.

Generating sequences

For simulations we often want to generate realizations for a large number of random variables. Random number generators have been designed with this purpose in mind: each new call mimics a new $U(0,1)$ random variable. The sequence of numbers thus generated is considered as a realization of a sequence of $U(0,1)$ random variables U_1, U_2, U_3,... with the special property that the events $\{U_i \leq a_i\}$ are independent[1] for every choice of the a_i.

6.3 Comparing two jury rules

At the Olympic Games there are several sports events that are judged by a jury, including gymnastics, figure skating, and ice dancing. During the 2002 winter games a dispute arose concerning the gold medal in ice dancing: there were allegations that the Russian team had bribed a French jury member, thereby causing the Russian pair to win just ahead of the Canadians. We look into operating rules for juries, although we leave the effects of bribery to the exercises (Exercise 6.11).

Suppose we have a jury of seven members, and for each performance each juror assigns a grade. The seven grades are to be transformed into a final score. Two rules to do this are under consideration, and we want to choose

[1] In Chapter 9 we return to the question of independence between random variables.

the better one. For the first one, the highest and lowest scores are removed and the final score is the average of the remaining five. For the second rule, the scores are put in ascending order and the middle one is assigned as final score. Before you continue reading, consider which rule is better and how you can verify this.

A probabilistic model

For our investigation we assume that the scores the jurors assign deviate by some random amount from the true or deserved score. We model the score that juror i assigns when the performance deserves a score g by

$$Y_i = g + Z_i \qquad \text{for } i = 1, \ldots, 7, \tag{6.2}$$

where $Z_1, \ldots, Z_7$ are random variables with values around zero. Let h_1 and h_2 be functions implementing the two rules:

$$\begin{aligned} h_1(y_1, \ldots, y_7) &= \text{average of the middle five of } y_1, \ldots, y_7, \\ h_2(y_1, \ldots, y_7) &= \text{middle value of } y_1, \ldots, y_7. \end{aligned}$$

We are interested in deviations from the deserved score g:

$$\begin{aligned} T &= h_1(Y_1, \ldots, Y_7) - g, \\ M &= h_2(Y_1, \ldots, Y_7) - g. \end{aligned} \tag{6.3}$$

The distributions of T and M depend on the individual jury grades, and through those, on the juror-deviations $Z_1, Z_2, \ldots, Z_7$, which we model as $U(-0.5, 0.5)$ variables. This more or less finishes the modeling phase: we have given a stochastic model that mimics the workings of a jury and have defined, in terms of the variables in the model, the random variables T and M that represent the errors that result after application of the jury rules.

In any serious application, the model should be *validated*. This means that one tries to gather evidence to convince oneself and others that the model adequately reflects the workings of the real system. In this chapter we are more interested in showing what you can do with simulation once you have a model, so we skip the validation.

The next phase is analysis: which of the deviations is closer to zero? Because T and M are random variables, we would have to clarify what we mean by that, and answering the question certainly involves computing probabilities about T and M. We cannot do this with what we have learned so far, but we know how to simulate, so this is what we do.

Simulation

To generate a realization of a $U(-0.5, 0.5)$ random variable, we only need to subtract 0.5 from the result we obtain from a call to the random generator.

We do this 7 times and insert the resulting values in (6.2) as jury deviations $Z_1, \ldots, Z_7$, and substitute them in equations (6.3) to obtain T and M (the value of g is irrelevant: it drops out of the calculation):

$$\begin{aligned} T &= \text{average of the middle five of } Z_1, \ldots, Z_7, \\ M &= \text{middle value of } Z_1, \ldots, Z_7. \end{aligned} \tag{6.4}$$

In simulation terminology, this is called a *run*: we have gone through the whole procedure once, inserting realizations for the random variables. If we repeat the whole procedure, we have a second run; see Table 6.1 for the results of five runs.

Table 6.1. Simulation results for the two jury rules.

Run	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	T	M
1	−0.45	−0.08	−0.38	0.11	−0.42	0.48	0.02	−0.15	−0.08
2	−0.37	−0.18	0.05	−0.10	0.01	0.28	0.31	0.01	0.01
3	0.08	0.07	0.47	−0.21	−0.33	−0.22	−0.48	−0.12	−0.21
4	0.24	0.08	−0.11	0.19	−0.03	0.02	0.44	0.10	0.08
5	0.10	0.18	−0.39	−0.24	−0.36	−0.25	0.20	−0.11	−0.24

QUICK EXERCISE 6.4 The next realizations for $Z_1, \ldots, Z_7$ are: −0.05, 0.26, 0.25, 0.39, 0.22, 0.23, 0.13. Determine the corresponding realizations of T and M.

Table 6.1 can be used to check some computations. We also see that the realization of T was closest to zero in runs 3 and 5, the realization of M was closest to zero in runs 1 and 4, and they were (about) the same in run 2. There is no clear conclusion from this, and even if there was, one could wonder whether the next five runs would yield the same picture. Because the whole process mimics randomness, one has to expect some variation—or perhaps a lot. In later chapters we will get a better understanding of this variation; for the moment we just say that judgment based on a large number of runs is better. We do one thousand runs and exchange the table for pictures. Figure 6.3 depicts, for juror 1, a histogram of all the deviations from the true score g. For each interval of length 0.05 we have counted the number of runs for which the deviation of juror 1 fell in that interval. These numbers vary from about 40 to about 60.

This is just to get an idea about the results for an individual juror. In Figure 6.4 we see histograms for the final scores. Comparing the histograms, it seems that the realizations of T are more concentrated near zero than those of M.

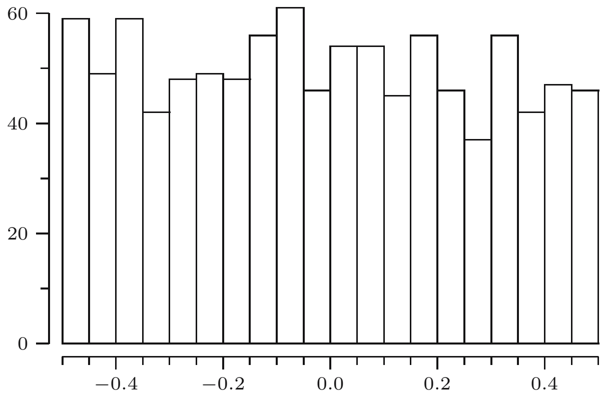


Fig. 6.3. Deviations of juror 1 from the deserved score, one thousand runs.

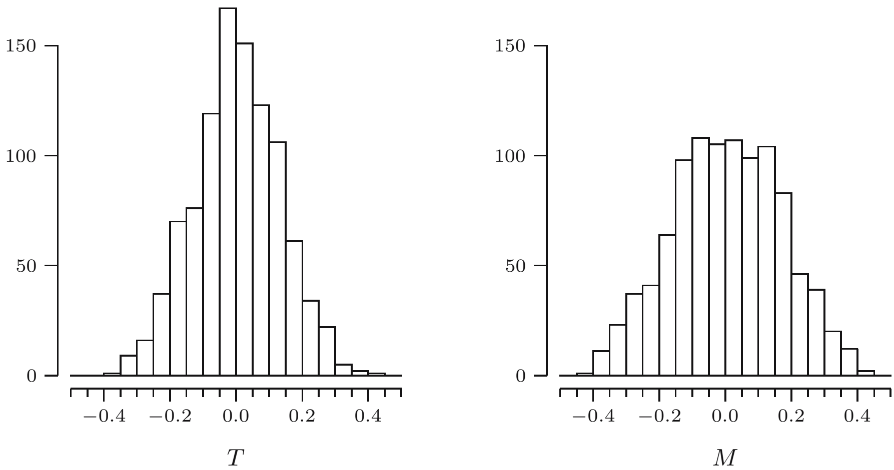


Fig. 6.4. One thousand realizations of T and M.

However, the two histograms do not tell us anything about the relation between T and M, so we plot the realizations of pairs (T, M) for all one thousand runs (Figure 6.5). From this plot we see that in most cases M and T go in the same direction: if T is positive, then usually M is also positive, and the same goes for negative values. In terms of the final scores, both rules generally overvalue and undervalue the performance simultaneously. On closer examination, with help of the line drawn from $(-0.5, -0.5)$ to $(0.5, 0.5)$, we see that the T values tend to be a little closer to zero than the M values.

This suggests that we make a histogram that shows the difference of the absolute deviations from true score. For rule 1 this absolute deviation is $|T|$, for rule 2 it is $|M|$. If the difference $|M| - |T|$ is positive, then T is closer to zero than M, and the difference tells us by how much. A negative difference

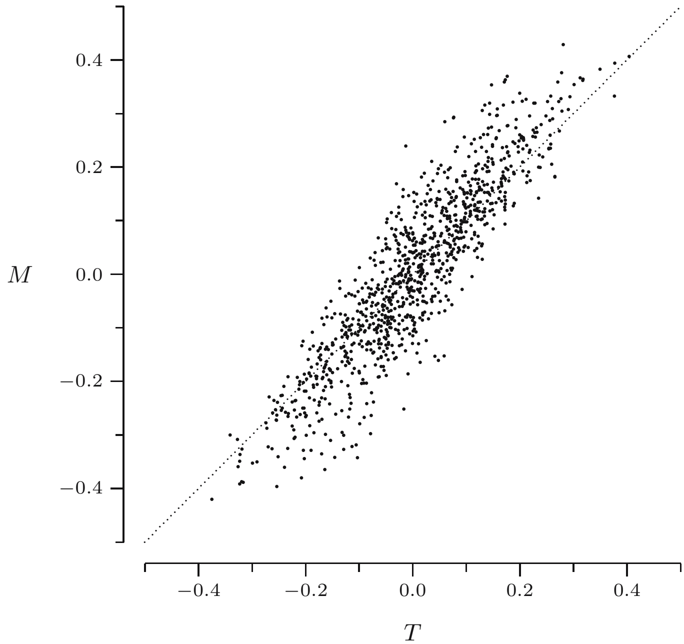


Fig. 6.5. Plot of the points (T, M), one thousand runs.

means that M was closer. In Figure 6.6 all the differences are shown in a histogram. The bars to the right of zero represent 696 runs. So, in about 70% of the runs, rule 1 resulted in a final score that is closer to the true score than rule 2. In about 30% of the cases, rule 2 was better, but generally by a smaller amount, as we see from the histogram.

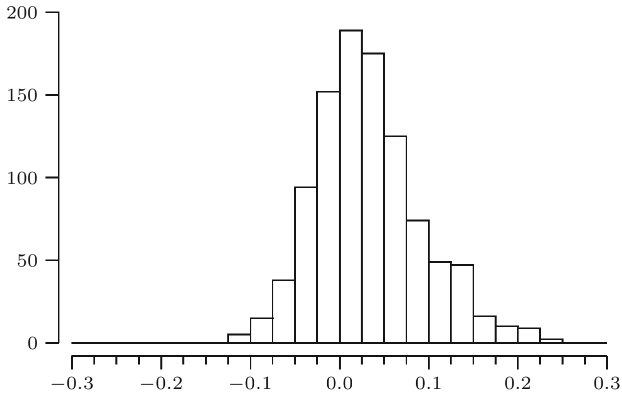


Fig. 6.6. Differences $|M| - |T|$ for one thousand runs.

6.4 The single-server queue

There are many situations in life where you stand in a line waiting for some service: when you want to withdraw money from a cash dispenser, borrow books at the library, be admitted to the emergency room at the hospital, or pump gas at the gas station. Many other queueing situations are hidden: an email message you send might be queued at the local server until it has sent all messages that were submitted ahead of yours; searching the Internet, your browser sends and receives packets of information that are queued at various stages and locations; in assembly lines, partly finished products move from station to station, each time waiting for the next component to be added.

We are going to study one simple queueing model, the so-called single-server queue: it has one *server* or service mechanism, and the arriving customers await their turn in order of their arrival. For definiteness, think of an oasis with one big water well. People arrive at the well with bottles, jerry cans, and other types of containers, to pump water. The supply of water is large, but the pump capacity is limited. The pump is about to be replaced, and while it is clear that a larger pump capacity will result in shorter waiting times, more powerful pumps are also more expensive. Therefore, to prepare a decision that balances costs and benefits, we wish to investigate the relationship between pump capacity and system performance.

Modeling the system

A stochastic model is in order: some general characteristics are known, such as how many people arrive per day and how much water they take on average, but the individual arrival times and amounts are unpredictable. We introduce random variables to describe them: let T_1 be the time between the start at time zero and the arrival of the first customer, T_2 the time between the arrivals of the first and the second customer, T_3 the time between the second and the third, etc.; these are called the *interarrival times*. Let S_i be the length of time that customer i needs to use the pump; in standard terminology this is called the *service time*. This is our description so far:

Arrivals at:	T_1	$T_1 + T_2$	$T_1 + T_2 + T_3$	etc.
Service times:	S_1	S_2	S_3	etc.

The pump capacity v (liters per minute) is not a random variable but a model parameter or decision variable, whose "best" value we wish to determine. So if customer i requires R_i liters of water, then her service time is

$$S_i = \frac{R_i}{v}.$$

To complete the model description, we need to specify the distribution of the random variables T_i and R_i:

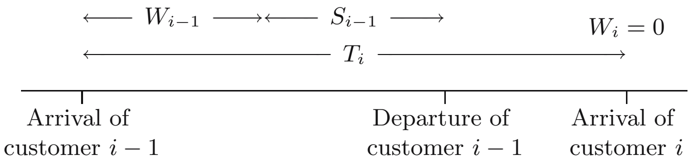
Interarrival times: every T_i has an $Exp(0.5)$ distribution (minutes);
Service requirement: every R_i has a $U(2,5)$ distribution (liters).

This particular choice of distributions would have to be supported by evidence that they are suited for the system at hand: a validation step as suggested for the jury model is appropriate here as well. For many arrival type processes, however, the exponential distribution is reasonable as a model for the interarrival times (see Chapter 12). The particular uniform distribution chosen for the required amount of water says that all amounts between 2 and 5 liters are equally likely. So there is no sheik who owns a 5000-liter water truck in "our" oasis.

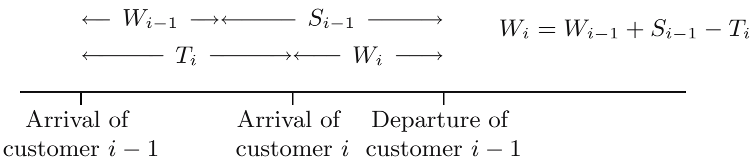
To evaluate system performance, we want to extract from the model the waiting times of the customers and how busy it is at the pump.

Waiting times

Let W_i denote the waiting time of customer i. The first customer is lucky; the system starts empty, and so $W_1 = 0$. For customer i the waiting time depends on how long customer $i-1$ spent in the system compared to the time between their respective arrivals. We see that if the interarrival time T_i is long, relatively speaking, then customer i arrives *after* the departure of customer $i-1$, and so $W_i = 0$:



On the other hand, if customer i arrives *before* the departure, the waiting time W_i equals whatever remains of $W_{i-1} + S_{i-1}$:



Summarizing the two cases, we see obtain:

$$W_i = \max\{W_{i-1} + S_{i-1} - T_i, 0\}. \tag{6.5}$$

To carry out a simulation, we start at time zero and generate realizations of the interarrival times (the T_i) and service requirements (the R_i) for as long as we want, computing the other quantities that follow from the model on the way. Table 6.2 shows the values generated this way, for two pump capacities ($v = 2$ and 3) for the first six customers. Note that in both cases we use the same realizations of T_i and R_i.

Table 6.2. Results of a short simulation.

	Input realizations			$v = 2$		$v = 3$	
i	T_i	Arr.time	R_i	S_i	W_i	S_i	W_i
1	0.24	0.24	4.39	2.20	0	1.46	0
2	1.97	2.21	4.00	2.00	0.23	1.33	0
3	1.73	3.94	2.33	1.17	0.50	0.78	0
4	2.82	6.76	4.03	2.01	0	1.34	0
5	1.01	7.77	4.17	2.09	1.00	1.39	0.33
6	1.09	8.86	4.24	2.12	1.99	1.41	0.63

QUICK EXERCISE 6.5 The next four realizations are T_7: 1.86; R_7: 4.79; T_8: 1.08; and R_8: 2.33. Complete the corresponding rows of the table.

Longer simulations produce so many numbers that we will drown in them unless we think of something. First, we summarize the waiting times of the first n customers with their average:

$$\bar{W}_n = \frac{W_1 + W_2 + \cdots + W_n}{n}. \tag{6.6}$$

Then, instead of giving a table, we plot the pairs $(n, \bar{W}_n)$, for $n = 1, 2, \ldots$ until the end of the simulation. In Figure 6.7 we see that both lines bounce up and down a bit. Toward the end, the average waiting time for pump capacity 3 is about 0.5 and for $v = 2$ about 2. In a longer simulation we would see each of the averages converge to a limiting value (a consequence of the so-called law of large numbers, the topic of Chapter 13).

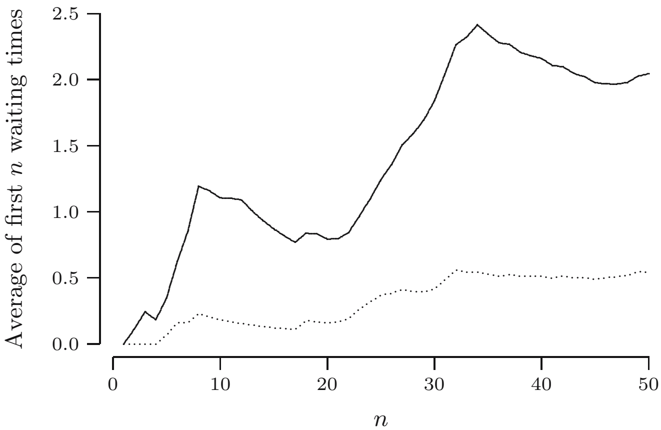


Fig. 6.7. Averaged waiting times at the well, for pump capacity 2 and 3.

Work-in-system

To show how busy it is at the pump one could record how many customers are waiting in the queue and plot this quantity against time. A slightly different approach is to record at every moment how much work there is in the system, that is, how much time it would take to serve everyone present at that moment. For example, if I am halfway through filling my 4-liter jerry can and three persons are waiting who require 2, 3, and 5 liters, respectively, then there are 12 liters to go; at $v = 2$, there is 6 minutes of work in the system, and at $v = 3$ just 4.

The amount of work in the system just before a customer arrives equals the waiting time of that customer, because it is exactly the time it takes to finish the work for everybody ahead of her. The work-in-system at time t tells us how long the wait would be *if* somebody were to arrive at t. For this reason, this quantity is also called the *virtual waiting time*.

Figure 6.8 shows the work-in-system as a function of time for the first 15 minutes, using the same realizations that were the basis for Table 6.2. In the top graph, corresponding to $v = 2$, the work in the system jumps to 2.20 (which is the realization of $R_1/2$) at $t = 0.24$, when the first customer arrives. So at $t = 2.21$, which is 1.97 later, there is $2.20 - 1.97 = 0.23$ minute of work left; this is the waiting time for customer 2, who brings an amount of work of 2.00 minutes, so the peak at 1.97 is $0.23 + 2.00 = 2.23$, etc. In the bottom graph we see the work-in-system reach zero more often, because the individual (work) amounts are 2/3 of what they are when $v = 2$. More often, arriving

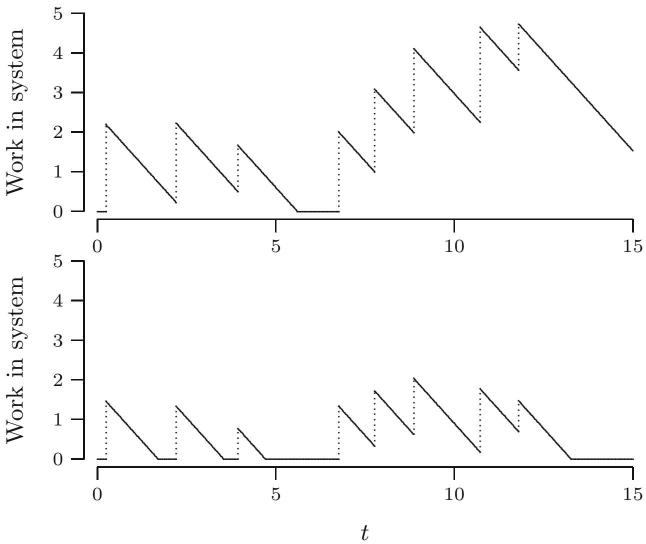


Fig. 6.8. Work in system: top, $v = 2$; bottom, $v = 3$.

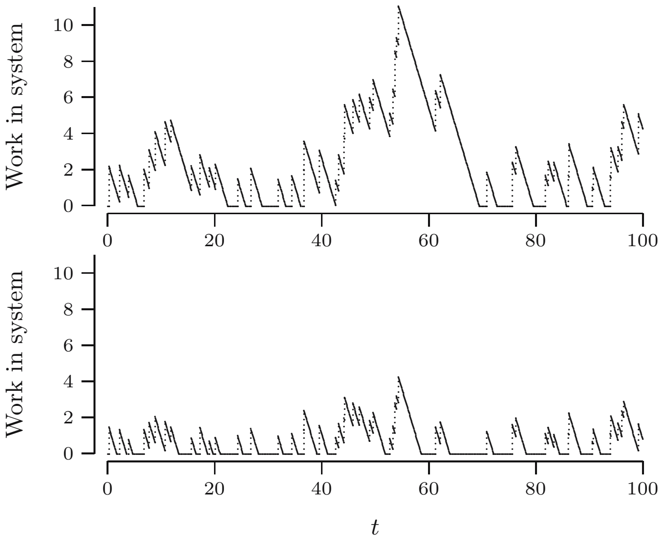


Fig. 6.9. Work in system: top, $v = 2$; bottom, $v = 3$.

customers find the queue empty and the pump not in use; they do not have to wait.

In Figure 6.9 the work-in-system is depicted as a function of time for the first 100 minutes of our run. At pump capacity 2 the virtual waiting time peaks at close to 11 minutes after about 55 minutes, whereas with $v = 3$ the corresponding peak is only about 4 minutes. There also is a marked difference in the proportion of time the system is empty.

6.5 Solutions to the quick exercises

6.1 To simulate the coin, choose any three of the six possible outcomes of the die, report heads if one of these three outcomes turns up, and report tails otherwise. For example, heads if the outcome is odd, tails if it is even.

To simulate the die using a coin is more difficult; one solution is as follows. Toss the coin three times and use the following conversion table to map the result:

Coins	HHH	HHT	HTH	HTT	THH	THT
Die	1	2	3	4	5	6

Repeat the coin tosses if you get TTH or TTT.

6.2 Let the $U(0,1)$ variable be U and set:

$$Y = \begin{cases} 1 & \text{if } U < \frac{3}{5}, \\ 3 & \text{if } \frac{3}{5} \leq U < \frac{4}{5}, \\ 4 & \text{if } U \geq \frac{4}{5}. \end{cases}$$

So, for example, $\mathrm{P}(Y = 3) = \mathrm{P}\left(\frac{3}{5} \leq U < \frac{4}{5}\right) = \frac{1}{5}$.

6.3 The given distribution function F is strictly increasing between 1 and 3, so we use the method with F^{inv}. Solve the equation $F(x) = \frac{1}{4}(x-1)^2 = u$ for x. This yields $x = 1 + 2\sqrt{u}$, so we can set $X = 1 + 2\sqrt{U}$. If you need to be convinced, determine F_X.

6.4 In ascending order the values are $-0.05, 0.13, 0.22, 0.23, 0.25, 0.26, 0.39$, so for M we find 0.23, and for T $(0.13 + 0.22 + 0.23 + 0.25 + 0.26)/5 = 0.22$.

6.5 We find:

Input realizations				$v=2$		$v=3$	
i	T_i	Arr.time	R_i	S_i	W_i	S_i	W_i
7	1.86	10.72	4.79	2.39	2.25	1.60	0.18
8	1.08	11.80	2.33	1.16	3.57	0.78	0.70

6.6 Exercises

6.1 Let U have a $U(0,1)$ distribution.

a. Describe how to simulate the outcome of a roll with a die using U.

b. Define Y as follows: round $6U + 1$ down to the nearest integer. What are the possible outcomes of Y and their probabilities?

6.2 ⊡ We simulate the random variable $X = 1 + 2\sqrt{U}$ constructed in Quick exercise 6.3. As realization for U we obtain from the pseudo random generator the number $u = 0.3782739$.

a. What is the corresponding realization x of the random variable X?

b. If the next call to the random generator yields $u = 0.3$, will the corresponding realization for X be larger or smaller than the value you found in **a**?

c. What is the probability the next draw will be smaller than the value you found in **a**?

6.3 Let U have a $U(0,1)$ distribution. Show that $Z = 1 - U$ has a $U(0,1)$ distribution by deriving the probability density function or the distribution function.

6.4 Let F be the distribution function as given in Quick exercise 6.3: $F(x)$ is 0 for $x < 1$ and 1 for $x > 3$, and $F(x) = \frac{1}{4}(x-1)^2$ if $1 \leq x \leq 3$. In the answer it is claimed that $X = 1 + 2\sqrt{U}$ has distribution function F, where U is a $U(0,1)$ random variable. Verify this by computing $\mathrm{P}(X \leq a)$ and checking that this equals $F(a)$, for any a.

6.5 ⊞ We have seen that if U has a $U(0,1)$ distribution, then $X = -\ln U$ has an $Exp(1)$ distribution. Check this by verifying that $\mathrm{P}(X \leq a) = 1 - \mathrm{e}^{-a}$ for $a \geq 0$.

6.6 ⊡ Somebody messed up the random number generator in your computer: instead of uniform random numbers it generates numbers with an $Exp(2)$ distribution. Describe how to construct a $U(0,1)$ random variable U from an $Exp(2)$ distributed X.
Hint: look at how you obtain an $Exp(2)$ random variable from a $U(0,1)$ random variable.

6.7 ⊞ In models for the lifetimes of mechanical components one sometimes uses random variables with distribution functions from the so-called Weibull family. Here is an example: $F(x) = 0$ for $x < 0$, and

$$F(x) = 1 - \mathrm{e}^{-5x^2} \qquad \text{for } x \geq 0.$$

Construct a random variable Z with this distribution from a $U(0,1)$ variable.

6.8 A random variable X has a $Par(3)$ distribution, so with distribution function F with $F(x) = 0$ for $x < 1$, and $F(x) = 1 - x^{-3}$ for $x \geq 1$. For details on the Pareto distribution see Section 5.4. Describe how to construct X from a $U(0,1)$ random variable.

6.9 ⊡ In Quick exercise 6.1 we simulated a die by tossing three coins. Recall that we might need several attempts before succeeding.

a. What is the probability that we succeed on the first try?

b. Let N be the number of tries that we need. Determine the distribution of N.

6.10 ⊞ There is usually more than one way to simulate a particular random variable. In this exercise we consider two ways to generate geometric random variables.

a. We give you a sequence of independent $U(0,1)$ random variables U_1, U_2, $\ldots$. From this sequence, construct a sequence of Bernoulli random variables. From the sequence of Bernoulli random variables, construct a (single) $Geo(p)$ random variable.

b. It is possible to generate a $Geo(p)$ random variable using just *one* $U(0,1)$ random variable. If calls to the random number generator take a lot of CPU time, this would lead to faster simulation programs. Set $\lambda = -\ln(1-p)$ and let Y have a $Exp(\lambda)$ distribution. We obtain Z from Y by rounding to the nearest integer *greater* than Y. Note that Z is a discrete random variable, whereas Y is a continuous one. Show that, nevertheless, the event $\{Z > n\}$ is the same as $\{Y > n\}$. Use this to compute $\mathrm{P}(Z > n)$ from the distribution of Y. What is the distribution of Z? (See Quick exercise 4.6.)

6.11 Reconsider the jury example (see Section 6.3). Suppose the first jury member is bribed to vote in favor of the present candidate.

a. How should you now model Y_1? Describe how you can investigate which of the two rules is less sensitive to the effect of the bribery.

b. The International Skating Union decided to adopt a rule similar to the following: *randomly* discard two of the jury scores, then average the remaining scores. Describe how to investigate this rule. Do you expect this rule to be more sensitive to the bribery than the two rules already discussed, or less sensitive?

6.12 ⊞ **A tiny financial model.** To investigate investment strategies, consider the following:

You can choose to invest your money in one particular stock or put it in a savings account. Your initial capital is € 1000. The interest rate r is 0.5% per month and does not change. The initial stock price is € 100. Your stochastic model for the stock price is as follows: next month the price is the same as this month with probability 1/2, with probability 1/4 it is 5% lower, and with probability 1/4 it is 5% higher. This principle applies for every new month. There are no transaction costs when you buy or sell stock.

Your investment strategy for the next 5 years is: convert all your money to stock when the price drops below € 95, and sell all stock and put the money in the bank when the stock price exceeds € 110.

Describe how to simulate the results of this strategy for the model given.

6.13 We give you an unfair coin and you do not know $\mathrm{P}(H)$ for this coin. Can you simulate a fair coin, and how many tosses do you need for each fair coin toss?