Chapter 8

Solutions to Exercises

Winfried Just[1], Hannah Callender[2], M. Drew LaMar[3] and Natalia Toporikova[4]
[1]*Department of Mathematics, Ohio University, Athens, OH, USA,* [2]*University of Portland, Portland, OR, USA,*
[3]*The College of William and Mary, Williamsburg, VA, USA,* [4]*Washington and Lee University, Lexington, VA, USA*

Exercise 8.1. It might appear that the number of reported cases of new infections that are filed for a given day or week would give a good estimate of the incidence during that day or week. However, this will not always be the case. Why not?

Sample Solution: Because these reports are filed by health care providers, they only cover patients who actually seek treatment. These will generally comprise only a fraction of all infected hosts. Especially for diseases with relatively mild symptoms, we should expect significant *underreporting*. On the other hand, if suspected rather than confirmed cases are reported, then we may get an overestimate of nI.

Also, the reporting system may not work perfectly, both in terms of filing the reports and correctly diagnosing illnesses in the first place.

Moreover, there will be a time lag between when people get infected and when they seek treatment, so that the number of new cases reported for a given day will generally represent new infections that occurred sometime earlier.

For example, Lyme disease is a bacterial infection that is transferred to humans by tick bites. Shortly after the bite, a characteristic skin rash develops, but it soon disappears and often goes unnoticed. After the rash subsides, symptoms remain nonspecific and relatively mild for a number of years. But several years later an infected person may develop severe health problems [1]. Patients often seek treatment at this late stage of symptoms and frequently get misdiagnosed. Thus in the case of Lyme disease, the epidemic curve cannot be reliably inferred from the number of reported cases. □

Exercise 8.2. Why did we need to assume that Δt was "sufficiently short"? More generally, why can we have only approximate equality in (8.1)?

Sample Solution: If the time interval is too long, there may be multiple contacts between i and j. But $v_{i,j}$ as we have defined it above stands only for the probability of transmission during one contact, so we could not simply multiply the probabilities. Moreover, over a long time interval, j may already have ceased to be infectious, or i may have already become infected by another host before a contact with j occurred. If the length Δt of the time interval is very short, then these complicating effects become negligible, but they do not disappear entirely. For this reason we have only approximate equality in (8.1). □

Exercise 8.3. There are a lot of other things that we may find interesting about real people: whether they are shy or assertive, if they wear fashionable clothes, whether they have a standoffish or emotionally expressive demeanor, how much they pay for their car insurance, whether they ever received an organ transplant, or whether both of their grandfathers died at the age of 97. Could any of these personal characteristics actually *matter* in terms of $b_{i,j}$?

Algebraic and Discrete Mathematical Methods for Modern Biology. http://dx.doi.org/10.1016/B978-0-12-801213-0.00028-9

Sample Solution: Perhaps you were inclined to answer with a resounding "Of course not!" If so, congratulations! This shows that you have a knack for focusing on the essentials, which is key to mathematical modeling.

But let us think again: assertive people tend to have social interactions with more people than shy ones do. Thus, assertiveness increases the probability $c_{i,j}$ of making a contact, and hence increases $b_{i,j}$ as well.

Similarly, people who dress well will have more social opportunities than people who do not; at least this is what the fashion industry wants us to believe.

It may not be obvious whether or not emotionally expressive people will have *more* social contacts; this will depend to some extent on the cultural norms of the given population. But touchy-feely people will on average make literally closer contacts than standoffish ones, which may increase the probability $v_{i,j}$ of transmission during any given contact, and may increase $b_{i,j}$ as well.

Similarly, car insurance companies demand higher premiums of clients whom they perceive as prone to risky behavior, and such proclivity may lead to an increased probability of disease transmission per contact.

Recipients of organ transplants need to take medication that suppresses their immune system. This will increase their likelihood $v_{i,j}$ of contracting the disease from any given contact.

The strength of an individual's immune system also depends on genetic factors that are heritable. Long-lived grandparents very likely had strong immune systems, and may have passed this trait to their grandchildren. □

Exercise 8.4. Can you think of situations when even the probabilities $b_{i,j}$ would vary over time, that is, depend on t_{curr} for a fixed Δt? When would one want to incorporate this dependence on time into our models?

Sample Solution: Actually, in most outbreaks in real populations there will be variability. For example, human hosts may move to another house, change employers, or even change spouses. This alters $c_{i,j}$ and hence $b_{i,j}$. However, this kind of background variability is usually something that one can ignore.

More interesting is the case when one wants to study the effectiveness of behavior modification after the start of an outbreak. This would usually translate into lowering the probabilities $v_{i,j}$ and hence $b_{i,j}$ over time t. □

Exercise 8.5. To which of the times $T_i^E, T_i^Y, T_i^I, T_i^C$ that we introduced above are the data summarized in the top panel of Figure 8.1 most directly related?

Sample Solution: The data tell us approximately for how many hosts i the times T_i^Y of onset of symptoms were reported on a given day. We cannot automatically assume that the data give us the corresponding counts for the times T_i^I, unless $T_i^I \approx T_i^Y$. The curve does not give us any direct information about T_i^E or T_i^C. □

Exercise 8.6. Suppose the current time is early morning on Friday the 13^{th} and time is measured in days. How would you translate "within the next 10 days, I will start coughing and sneezing" and "eventually I will start coughing and sneezing" into the language of probabilities and r.v.s T_i^E, T_i^I, T_i^R?

Sample Solution: Strictly speaking, these probabilities cannot be expressed in terms of T_i^E, T_i^I, T_i^R. The verbal descriptions would translate into probabilities $P(13 \leq T_7^Y < 23)$ and $P(13 \leq T_7^Y < \infty)$.

However, because for the flu the onset of symptoms roughly coincides with the onset of infectiousness, these probabilities are close to $P(13 \leq T_7^I < 23)$ and $P(13 \leq T_7^I < \infty)$. □

Exercise 8.7. The ellipsis at the end of (8.2) suggest that the list goes on. Can you predict what will happen at future times?

Sample Solution: Yes, you can. At time $t = 3.9$, both boxes **I** and **E** become empty. There are no preinfectious or infectious hosts left in the population. There are susceptible hosts in box **S**, but there is nobody left in the population who could infect them in the future. The outbreak has run its course, and (8.2) gives us in fact the complete list. □

Exercise 8.8. Based on (8.2) and your simulation, derive each of the following:

(a) The final size of the outbreak. Try to figure this out based only on the state of the population at time $t = 3.9$, as represented by the contents of your boxes at the end of the simulation.

(b) A graph that depicts the prevalence function $|\mathbf{I}(t)|$ by showing how the proportion of infectious hosts in the population changes over time.

(c) Your best guess at the epidemic curve, assuming that time is measured in weeks.

(d) A rough estimate of the mean duration of infectiousness $\langle \tau^I \rangle$.

(e) A rough estimate of the mean duration of the latent period $\langle \tau^E \rangle$.

Sample Solution:

(a) The final size was defined above as the proportion of hosts who experience infection at any time during the outbreak. The most direct calculation would involve redoing the simulation and keeping a tally of all hosts whose slips visited box **I** at some time. But here is a simpler method: all these slips, and only these slips, must have ended up in box **R** at the end of the simulation. This box will contain slips for hosts $1, 2, 4, 7$ at the end, so the final size will be $\frac{4}{7}$. Note, however, that this shortcut works only if we start with an initial state where box **R** is empty and if all infectious hosts eventually move to box **R** and stay there permanently.

(b) See Figure S8.1.

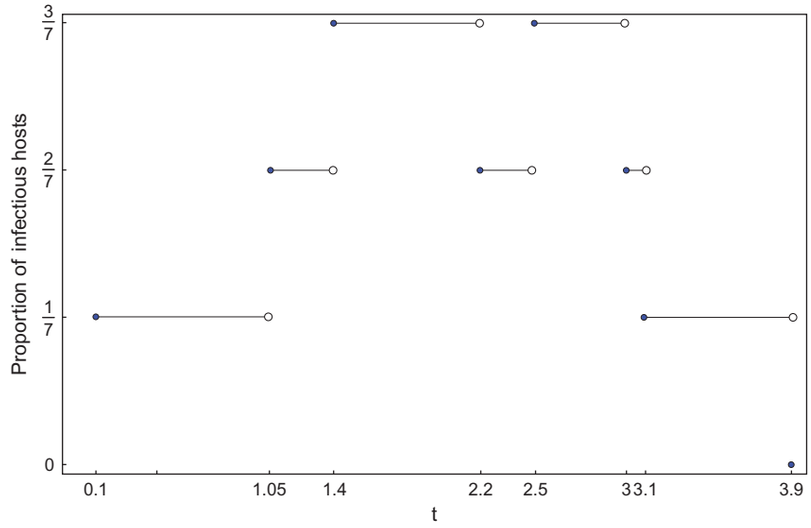

FIGURE S8.1 Proportion of infectious hosts as a function of time for Exercise 8.8.

(c) To construct this curve, we add up all new infections in the time interval $[0, 1)$ to find the height of the curve from $t = 0$ to $t = 1$. Then along the next interval, from $t = 1$ to $t = 2$, we add up all new infections in this time interval. Note that in the fourth week there are no new infections, so the curve height is 0 along this time interval.

We might base the constructions either on the transition times T_i^E or the transition times T_i^I. Figure S8.2 is based on the first interpretation, while Figure S8.3 is based on the second. So which solution is correct? The answer is that it really depends. There is ambiguity in the wording of the definition of the epidemic curve. The interpretation in terms of times T_i^E is closer to the definition given in the text, while the other interpretation will be closer to the kind of data that actually can be observed. So there is value in both interpretations.

(d) We only have data for hosts $1, 2, 4, 7$. The durations of infectiousness for these hosts are $\tau_1^I = T_1^R - T_1^I = 2.2 - 0.1 = 2.1$, $\tau_2^I = T_2^R - T_2^I = 3.9 - 2.5 = 1.4$, $\tau_4^I = T_4^R - T_4^I = 3.1 - 1.05 = 2.05$, $\tau_7^I = T_7^R - T_7^I = 3 - 1.4 = 1.6$, respectively. The mean can be calculated as $\frac{2.1+1.4+2.05+1.6}{4} = 1.7875$ and gives an estimate

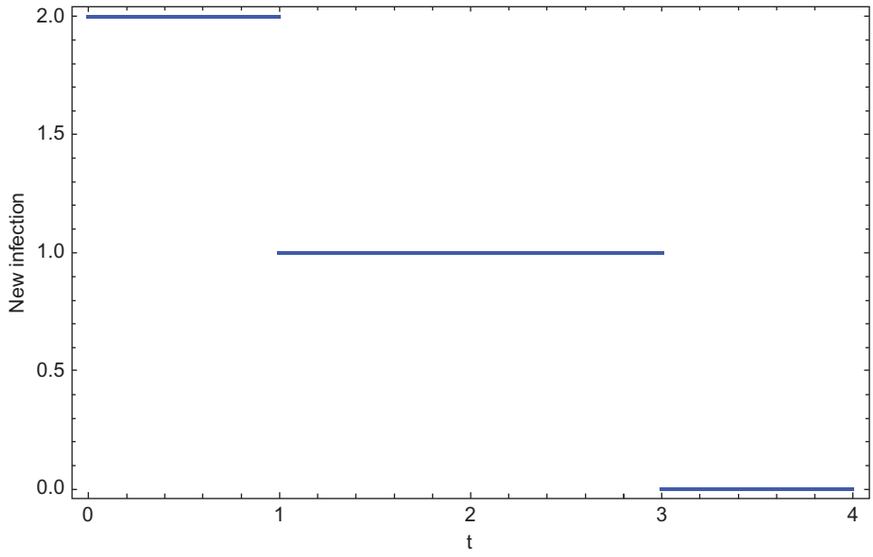

FIGURE S8.2 Epidemic curve for Exercise 8.8c based on times of exposure.

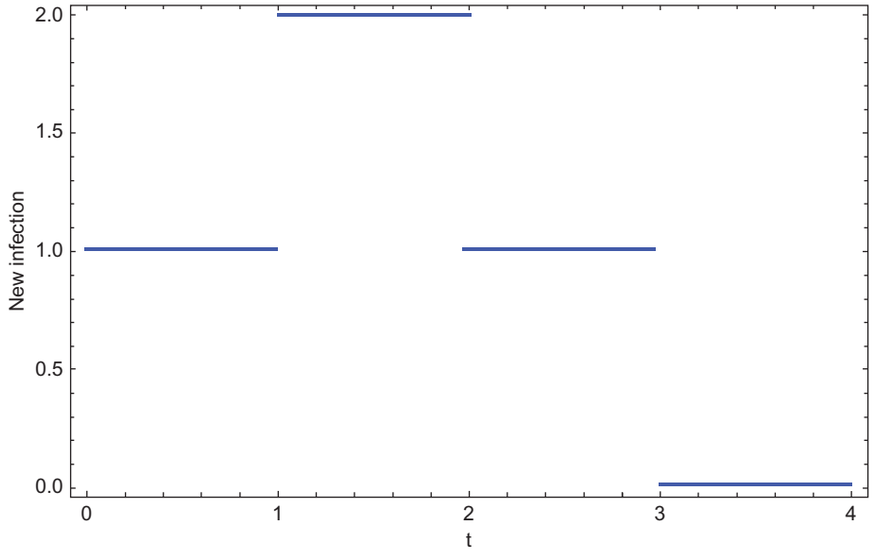

FIGURE S8.3 Epidemic curve for Exercise 8.8c based on onset of infectiousness.

of $\langle \tau^I \rangle$, the overall mean for this particular disease. This estimate is based on very few data points, though. We should not place any confidence in the last few digits; at best we might conclude that the mean duration of infectiousness appears to be ≈ 1.8.

(e) We only have data for hosts $1, 2, 4, 7$. The duration of the latency periods for these hosts are $\tau_1^E = T_1^I - T_1^E = 0.1 - 0 = 0.1$, $\tau_2^E = T_2^I - T_2^E = 2.5 - 2.4 = 0.1$, $\tau_4^E = T_4^I - T_4^E = 1.05 - 0.9 = 0.15$, $\tau_7^E = T_7^I - T_7^E = 1.4 - 1.3 = 0.1$, respectively. The mean can be calculated as $\frac{0.1+0.1+0.15+0.1}{4} = 0.1125$ and gives an estimate of $\langle \tau^E \rangle$, the overall mean for this particular disease. Again, this estimate is based on very few data points and we should not place any confidence in the last few digits. At best we might conclude that the mean duration of latency appears to be ≈ 0.1.

We will see later that numbers such us $\langle \tau^I \rangle$ or $\langle \tau^E \rangle$ will enter our models as parameters. The values of such parameters must be estimated from data on past outbreaks. Points (d) and (e) exemplify such calculations. If we have very few or not very reliable data, then we cannot be very confident in the values of these estimates. □

Exercise 8.9. Complete the list of (8.3) for the whole simulation.
Sample Solution:

$$(ESSSSSS), (ISSSSSS), (ISSESSS), (ISSISSS), (ISSISSS),$$
$$(ISSISSE), (ISSISSI), (RSSISSI), (RESISSI), (RISISSI),$$
$$(RISISSR), (RISRSSR), (RRSRSSR). \tag{S8.1}$$

□

Exercise 8.10. Probabilities must be hiding behind some mask. Can you spot their disguise?
Sample Solution: Given in the text. □

Exercise 8.11. Why? What strikes you as fishy about $\tau_i^I = \infty$?
Sample Solution: All hosts will eventually die. At this point they will cease to be infectious. Where do we move them at this time when there is no **R**-compartment?

This actually tells you that there is something suspicious about Figure 8.4c itself. The trick of setting $T_i^I = \infty = \tau_i^I$ only made us aware of the problem. We will return to this issue shortly. □

Exercise 8.12. Our definition of transition events allows for the possibility that, for example, $T_6^I = T_{77}^I = T_8^R = t$ for some t. In other words, each of these state transitions would happen exactly at the same time. Why can we ignore this possibility in continuous-time models?
Sample Solution: Think about a situation when $t = \pi$. Then onset of immunity for hosts 6 and 77 and recovery of host 8 would all need to happen simultaneously *exactly* at time $t = 3.141592653589793\ldots$. The probability of this possibility is 0, so we can ignore it. □

Exercise 8.13. Consider the state transition sequence (8.2). Find the state transition events for times $t = 1, 2, 3$ in the corresponding discrete-time model.
Sample Solution: $T_1^I = T_4^E = 1$, $T_4^I = T_7^E = T_7^I = 2$, $T_1^R = T_2^E = T_2^I = T_7^R = 3$. □

Exercise 8.14. Compare the predictions of four *SIR*-models of the same disease. In one model, choose the continuous-time option; in the second and third models, choose the discrete-time approximation with time steps $\Delta t = 0.02$ and $\Delta t = 1$, respectively. In the fourth model, set the probability of removal a to 1 so as to obtain a next-generation model. Detailed instructions for the exercise are given in Section 8.4 of [2].
Sample Solution: The average final size both in the continuous-time model and the discrete-time model with time step $\Delta t = 0.02$ should be close to 0.6, and the mean duration of the outbreak for these two models should be a little less than 10. It is not immediately clear whether the small differences are attributable to random fluctuations or differences in the models. It certainly appears that for a very small time step, the discrete model gives very good approximations to the continuous-time model.

For a large time step Δt we observed large differences between the continuous and discrete models, with average final size around 0.9 and mean duration larger than 11. Because we ran as many as 500 simulations in each experiment, these cannot be attributed to random fluctuations alone. For a large Δt, discrete-time models are no longer good approximations to continuous-time models.

The results for next-generation models fall in between. They appear to give overestimates of the final size by about 10% for our parameter settings, and still give mean durations near 10 that are close to the ones predicted for continuous-time models.

In Section 8.6 of the Online Appendix [2], we will return to these observations and examine how they can be explained in terms of mathematical models. We will also discuss implications in terms of usefulness of the three types of models in terms of predicting the course of actual outbreaks. □

Exercise 8.15. How would you modify the above pseudocode for simulating an *SIS*-model instead of an *SIR*-model?

Sample Solution: Simply replace line 4 by:
If $t_{\text{next}} = T_i^R$ **then** change the state of host i to S. □

Exercise 8.16. What would happen in the simulation if we did not generate a new t_{next} and went straight to line 2 again? How can we fix this problem?

Sample Solution: Suppose the simulator would simply move on to line 2. Because the old values of t_{next} and for the next transition event have not been erased, the simulator would see the previous values of the corresponding variables and carry out the specified assignment instructions. These would not change any values, but still take up CPU time. After finishing with line 4, the terminating condition would still not hold. Thus, the simulator would run through an infinite loop while the state of the simulation would remain unchanged (as it actually should). The user meanwhile may watch in bewilderment a frozen screen while the simulation supposedly continues until she has the good sense to terminate the simulation herself.

This can be easily fixed by "loading the die" so that if there are no more infectious hosts, line 1 will give us a value $t_{\text{next}} > T_{\max}$. □

Exercise 8.17. What would happen if we left the order of lines as in the previous pseudocode? Why did the previous arrangement of the code work just fine for the model with continuous time?

Sample Solution: If we first change the states of all hosts in *New* to I and then execute line 3n, the newly infectious hosts would also be moved to the **R**-compartment and we would have no infectious hosts left at time $t_{\text{curr}} + 1$. This does not present a problem in the pseudocode for the continuous-time model where only one of the **If**-conditions in lines 3 and 4 can be satisfied, and our simulator will never need to execute both lines in the same step of the simulation. □

Exercise 8.18. Note that p_i represents the probability that host i will be included in the set *New*, that is, that host i will be infectious at the next time step. What do these probabilities depend on? Will they be fixed throughout the simulation? Will p_i be the same for each i?

Sample Solution: Given in the text. □

Exercise 8.19. Using your results from Experiment 1 and the Central Limit Theorem, calculate the number of simulations needed to obtain 95% confidence that the observed mean number of removed hosts is within an error tolerance of 0.5 from the true mean.

Sample Solution: The section on normal distributions of the review of basic concepts of probability theory on our web page [3] gives details of these calculations. For our experiment, we obtained a sample standard deviation of 4.219. Using an error tolerance of 0.5, we therefore found that in order to obtain 95% confidence, we would need

$$\frac{4(4.219)^2}{.5^2} = 285 \text{ trials.}$$

□

Exercise 8.20. Using ideas similar to our discussion in the previous paragraphs, explain how our simulator would use these parameters in the case of an *SEIR*-model for a population of $N = 1$ hosts if at time t_{curr} host 1 is in the exposed state.

Sample Solution: If we have $N = 1$ host who is exposed at time t_{curr}, then the next transition event must be transition of host 1 at time $T_1^I = t_{\text{next}} > t_{\text{curr}}$ from exposed to infectious. So, just as in the case removal in an *SIR*-model with $N = 1$ host, our simulator only needs to determine t_{next}, or, equivalently, the value of the random variable $t_{\text{next}} - t_{\text{curr}}$. Again, our simulator suffers from amnesia. Thus $t_{\text{next}} - t_{\text{curr}}$ must be a memoryless continuous r.v. and hence must have an exponential distribution with probability density function $g(x) = \gamma e^{-\gamma x}$ for $x \geq 0$ and some parameter $\gamma > 0$. Then, in analogy with (8.8), we will have

$P(t_{\text{next}} - t_{\text{curr}} \leq \Delta t) = 1 - \mathrm{e}^{-\gamma \Delta t} \approx \gamma \Delta t$, where the approximation on the right is valid as long as Δt is sufficiently small. Here the parameter γ of the exponential distribution of $t_{\text{next}} - t_{\text{curr}}$ is the rate γ at which exposed hosts become infectious. □

Exercise 8.21. The next transition event presumably should be one from this list. How could our simulator choose t_{next} and the particular state transition that will occur at this time from the list?

Sample Solution: The time t_{next} would be the smallest value on this list, and the next state transition would be the scenario for which this value was generated. □

Exercise 8.22. Suppose you have built a model and estimated parameters α and β for transition rates per day. How would you need to rescale these parameters if you were to reinterpret time in units of hours instead? Does this make sense in view of how the mean duration of infectiousness $\langle \tau^I \rangle$ is calculated from these parameters?

Sample Solution: If α, β are transition rates per day, then the corresponding rates per hour would be $\alpha_h = \frac{\alpha}{24}, \beta_h = \frac{\beta}{24}$.

For example, a daily rate $\alpha = 0.5$ translates into a mean duration of infectiousness $\langle \tau^I \rangle = \frac{1}{\alpha} = \frac{1}{0.5} = 2$ days or $\langle \tau^I \rangle = \frac{1}{\alpha_h} = \frac{24}{0.5} = 48$ h, which makes sense. □

Exercise 8.23.

(a) Suppose host 1 is infectious at time $t_{\text{curr}} > 1$ of the discrete model. How can you calculate the probability that host 1 will no longer be infectious at the next time step of this model?

(b) How would you calculate the probability $b_{i,j}$ of at least one effective contact between hosts i and j over a time interval of length Δt from the rate $\beta_{i,j}$ for an SI-model?

Sample Solution:

(a) Recall that our simulator has amnesia. Thus the probability of removal by time $t_{\text{curr}} + 1$ does not depend on the value of t_{curr} and will also be a.

(b) This part is discussed in the text. □

Exercise 8.24. How many hospital beds were needed to cope with the outbreak? Express your answer in terms of the incidence and prevalence functions, and critically evaluate it.

Sample Solution: Under the scenario outlined above, the number of hospital beds that were needed is equal to the maximum of the prevalence function $|\mathbf{I}(t)|$.

We can argue as follows. Because each patient was infectious for exactly 2 weeks, the number $|\mathbf{I}(t)|$ for week number t should be equal to the sum $nI(t-1) + nI(t)$ of new infections that occurred in week t and the previous week. Because $nI(0) = 0 = nI(6)$ in our example, this would give us the following estimate of the prevalence function:

$$|\mathbf{I}(1)| = 6, \quad |\mathbf{I}(2)| = 24, \quad |\mathbf{I}(3)| = 54, \quad |\mathbf{I}(4)| = 55, \quad |\mathbf{I}(5)| = 26, \quad |\mathbf{I}(6)| = 7. \tag{S8.2}$$

It appears that 55 hospital beds would have been needed. In fact, this is the *minimum* number that we can deduce from the given information, but the actual number may have been even higher. For example, more hospital beds may be needed if there were more new cases near the end of week 3 than near the beginning of week 3 and/or more new cases near the beginning of week 5 than near its end. The magnitude of overlap between hospital stays of patients that were admitted during week 3 and during week 5 cannot be deduced from week-by-week reports. □

Exercise 8.25. What would be the most plausible explanation? How does this relate to independence?

Sample Solution: Sally probably stayed home this weekend. We did not tell you that Sally actually attended this week's party, only that she attends about half of the parties.

Now substitute any of Steve's friends other than $\{j_1, j_2, \ldots, j_9\}$ for Sally. If you want to model effective contacts in terms of tossing 9 coins 90 times, you would need to assume that about half of the time *all* coins come up tails, and half of the time there is a high probability for each coin to come up heads. Mathematicians would say that the outcomes of the coins in each of the 90 rounds of tossing are *highly correlated*, which is the opposite of independence. In this case, the correlation will be positive.

Even if we were to tell you that Sally did attend the party, you could not automatically assume independence. The partial data we gave you could also indicate that this may have been a somewhat unusual Wednesday for her; perhaps she did not drink that night, or behaved in an unusually restrained way for a different reason. We really do not know what the particular cause was, but in each case we would want to revise our estimate of $P(F_6)$ down from $b_{i,6}$ based on the information that none of the events F_1, F_2, F_3, F_4 happened. □

Exercise 8.26. Redo the previous exercise for this new scenario.

Sample Solution: At a round-table dinner, there is a much higher chance of disease transmission from the two persons seated next to Sally than of transmission from the other attendees. In the absence of any additional information, our description of Steve's quiet Sunday night round-table dinners would make it rather likely that Sally sat next to Pam, Patty, Peter, or Paul. The information that none of the events F_1, F_2, F_3, F_4 occurred this Sunday makes this *less* likely, though. But Sally must have been seated between *some* neighbors among the 12 attendees. Thus the additional information makes it *more* likely that Ian was seated next to her, and we should revise our estimate of $P(F_6)$ *up* from $b_{i,6}$ instead of down. If we want to think about this scenario in terms of coin tosses, then the outcomes of the coins in each toss would be negatively correlated.

Such considerations do not apply in the story about the wwW parties, because we told you that "everybody hugs everybody else and people constantly mingle." □

Exercise 8.27. Should we revise our estimate $b_{i,6}$ of the probability that Sally made effective contact with Ian at *this* week's wwW party?

Sample Solution: Which way would we revise our probability—up or down from $b_{i,6}$? We wrote that "each" of Steve's friends attends about half of all parties, so these 99 friends are assumed not to differ in terms of their frequency of attendance. If Sally did indeed miss the party of last year for which we gave you data, this would not make it any more or less likely that she attended this week's party. If she did attend, the information about last year's party may indicate that Sally in general drinks responsibly, or shows more restraint in her social interactions than most attendees do. But such characteristics would already go into estimating $b_{i,6}$, which we can think of as a long-term average. Sally may have changed in some important ways since last year, but we have no way to infer from the data in which direction. In this case, an assumption of independence seems warranted. □

REFERENCES

[1] Centers for Disease Control and Prevention. Lyme disease. Available from: http://www.cdc.gov/lyme/. Accessed: 2014-08-14.

[2] Just W, Callender H, LaMar MD, Toporikova N. Online Appendix: Transmission of infectious diseases: data, models, and simulations. In: Robeva, R, editor. Algebraic and discrete mathematical methods for modern biology. New York: Academic Press; 2015.

[3] Just W, Callender H, LaMar MD. Exploring transmission of infectious diseases on networks with NetLogo. Available from: https://qubeshub.org/iontw and also http://www.ohio.edu/people/just/IONTW/.