

Green Simulation of Pandemic Disease Propagation

Spencer Wilson ^{1,†}, Abdullah Alabdulkarim ^{2,†}  and David Goldman ^{3,*,†}

¹ Industrial Engineering & Operations Research Department, University of California at Berkeley, Berkeley, CA 94720-1777, USA; spencerrowi@gmail.com

² Mechanical and Industrial Engineering Department, Majmaah University, Majmaah 11952, Saudi Arabia; a.alabdulkarim@mu.edu.sa

³ H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA

* Correspondence: sman@gatech.edu; Tel.: +1-404-822-8949

† All authors contributed equally to this work.

Received: 21 March 2019; Accepted: 18 April 2019; Published: 22 April 2019

Abstract: This paper is concerned with the efficient stochastic simulation of multiple scenarios of an infectious disease as it propagates through a population. In particular, we propose a simple “green” method to speed up the simulation of disease transmission as we vary the probability of infection of the disease from scenario to scenario. After running a baseline scenario, we incrementally increase the probability of infection, and use the common random numbers variance reduction technique to avoid re-simulating certain events in the new scenario that would not otherwise have changed from the previous scenario. A set of Monte Carlo experiments illustrates the effectiveness of the procedure. We also propose various extensions of the method, including its use to estimate the sensitivity of propagation characteristics in response to small changes in the infection probability.

Keywords: stochastic simulation; infectious disease; SIR model; variance reduction

1. Introduction

Successive global influenza pandemics have attracted a great deal of attention from society and public health officials. For instance, in the early 2000s, the dangerous avian influenza H5N1 strain jumped from poultry in Southeast Asia to humans and caused a serious pandemic outbreak (Abbott and Pearson [1], Ferguson et al. [2]). In 2009, the H1N1 influenza virus (“swine flu”) emerged as a worldwide outbreak (see, for instance, Andradóttir et al. [3]).

Disease outbreaks continue to be a major concern. For this reason, detailed study of influenza and other diseases is needed to help public policymakers with respect to the identification and control of such pandemics, and to evaluate pandemic preparedness plans. Various techniques have been developed over the years to do so. These techniques are often based on models involving differential equations or computer simulation.

In typical disease propagation models, as time progresses, individuals in a population make transitions from one health state to another according to a given rate or probability distribution. The most-basic disease spread model is what is known as the *SIR* model (Kermack and McKendrick [4]). At any time t , each individual must be either susceptible (*S*), infected (*I*), or removed (*R*), where “removed” means that the individual has either recovered or is dead, and will not be infected again; and it is only possible to move from *S* to *I* to *R*. If we let the quantities $S(t)$, $I(t)$, and $R(t)$ denote the numbers of individuals in states *S*, *I*, and *R*, respectively, at time t , then the instantaneous transition rates from one state to another are given by a set of simple differential equations:

$$\frac{dS(t)}{dt} = -\beta S(t)I(t), \quad \frac{dI(t)}{dt} = \beta S(t)I(t) - \delta I(t), \quad \frac{dR(t)}{dt} = \delta I(t),$$

where β and δ denote the infection and recovery rates. Thus, for instance, $S'(t)$ is the overall rate at which the population transitions from state S to state I. These classical equations can be used to model the number of individuals in each state at any time, as well as the probability that any individual will make a particular transition.

There is a rich literature of models based on differential equations, and these are widely accepted in the field. Some minor issues with these models concern the facts that they tend to perform poorly in small-population settings (where, notably, the numbers of individuals in each state must be integers); they usually do not take randomness into account; and they do not easily accommodate or study individual-level behavior.

A complementary simulation-based pandemic modeling literature has arisen to address some of the issues outlined above. Simulation is especially attractive due to the complex modeling issues that arise in the study of pandemics. For instance, what are the effects of the demographic composition of the population? How can intervention strategies such as vaccination or social distancing best be applied? Under what circumstances should we close a school as an intervention strategy? See, e.g., the papers Andradóttir et al. [3], Elveback et al. [5], Halloran et al. [6], Kelso et al. [7], Lee et al. [8], and Longini et al. [9,10] for additional background, motivation, and insights. Each of these papers uses a discrete-event / agent-based simulation approach, where individuals come into contact with each other on a daily basis at home, school, work, or outside of work (“compartments”). Potential criticisms of this approach are that (i) the models simulating daily life are often quite complicated (and are a bit of an art to develop); (ii) they produce “random” output; and (iii) the simulations may sometimes involve a tremendous amount of computation to complete.

In terms of potential drawback (i), it has come to pass that the use of simulation to model pandemic disease spread is now widely accepted, even if some of the modeling elements are sometimes regarded as ad hoc. To address this latter issue, there has been a great deal of work in the literature that incorporates sophisticated population modeling techniques—including population mixing and movement of individuals from location to location—along with a variety of mitigation strategies; see Andradóttir et al. [11] for additional motivation and references. With respect to (ii), one simply simulates any particular scenario over many independent replications and then conducts what can be regarded as standard, run-of-the-mill data analysis on the independent outputs. Our concern in this article is with simulation efficiency, item (iii).

With issue (iii) in mind, we discuss the simulation of a somewhat simplified version of the SIR model. Suppose that each susceptible individual may encounter each infected individual at most once a day. At such an encounter, the probability that the susceptible will become infected is p , and we assume that all encounters are independent so that each encounter is a Bernoulli trial. In some models, infected subjects first undergo a latent (noninfectious) period before actually become infectious; but for ease of exposition, the current paper will subsequently ignore that additional state. On any given day, individuals may move between contact groups (e.g., home, school, work, etc.), thereby encountering multiple cliques of people as the day goes on. A realization of the simulation usually proceeds until either no more susceptible individuals remain (only states I and R) or no more infected individuals remain (only states S and R), indicating that the disease has run its course. Moreover, the simulation is run over multiple independent replications, so as to obtain statistically sound estimates of system performance measures such as the expected number of individuals who are ultimately infected, the peak number of people infected over the course of the outbreak, and the expected length of the pandemic.

A critical driver of computational effort is the number of $\text{Bern}(p)$ trials that need to be carried out; this effort grows quickly as the population size increases and is an underlying concern of the current paper. Section 2 discusses how these trials are implemented, as well as some tricks that can be used to ameliorate the work effort. In particular, we propose the use of “green” simulation, in which we reuse many of the $\text{Bern}(p)$ trials between certain runs, thus avoiding recalculation effort (see, e.g., Feng and Staum [12] and Meterellioz et al. [13], the former of which coined the term). This is the main

contribution of our article. Section 3 presents several simple examples illustrating the method's efficacy. We also show how to use the methodology for purposes of sensitivity analysis—e.g., what is the derivative of the expected number of infected individuals as a function of p ? The latter application combines a simulation variance reduction technique (see Law [14]) with our green technology. Section 4 provides conclusions in addition to various potential augmentations of future interest.

2. Efficient Bernoulli Generation

A naive simulation of disease propagation would require us to generate a Bernoulli trial each time a susceptible individual (heretofore referred to as a “susceptible”) were to come in contact with an infected individual (an “infective”), for every replication of every experimental design point. Of course, a single $\text{Bern}(p)$ trial merely requires that the simulation generates a $\text{Unif}(0,1)$ pseudo-random number (PRN) U ; and if $p < U$, then the interpretation is that the susceptible becomes infected. However, the problem is that the entire exercise may require a tremendous number of $\text{Bern}(p)$ trials, a task which eventually becomes costly with respect to computer time.

What might one do to reduce the computational effort involving Bernoulli trials? There are a number of useful tricks, depending on the situation:

1. Within a particular simulation run, one could use a *binomial* random variable instead of a series of $\text{Bern}(p)$ random variables to determine the number of new infections. For instance, if an infective walks into a room with n susceptibles, and np is “small,” then it may be more efficient to (i) generate one $\text{Bin}(n, p)$ random variate X representing in one fell swoop the number of susceptibles who will become infected than to instead (ii) sample n individual $\text{Bern}(p)$ trials corresponding to the n susceptibles. If one were to take the $\text{Bin}(n, p)$ route, one could sample a single $\text{Unif}(0,1)$ PRN and use the discrete distribution version of the inverse-transform method to do a table look-up or instead use a normal distribution approximation to the binomial to generate the corresponding value for X (see, e.g., Law [14]). If $X = x \geq 1$, we would then have to randomly reassign x susceptibles to the infected state, which would take a little bit more work—though the expected number of assignments is only $E[X] = np$.
2. Within a particular simulation run, we could work with a single “super-infective” individual rather than multiple infectives. Consider the scenario above, except that k infectives walk into the room containing n susceptibles. Assuming that all of the infectives interact with all of the susceptibles in the form of $\text{Bern}(p)$ trials, an easy calculation shows that the probability that a particular susceptible will become infected that day is $1 - q^k$, where $q = 1 - p$. Then, instead of performing nk $\text{Bern}(p)$ trials with the individual infectives, we might equivalently perform only n $\text{Bern}(1 - q^k)$ trials using a “super-infective” entity—or even a single $\text{Bin}(n, 1 - q^k)$ trial; see Section 4 or references such as Tsai et al. [15].
3. Consider driving the simulation from different points of view. For example, is it more efficient to (i) generate each infected individual and see which susceptibles he infects as he proceeds through his day, or (ii) generate each susceptible individual and see if he gets infected as he proceeds through his day? See Shen et al. [16] and Tsai et al. [15] for insights and recommendations.
4. When re-running a simulation but with different input parameters, we might reuse from one simulation to the next as many of the random numbers as possible. This is the goal of the current paper. To illustrate, again consider the situation in which one infectious individual walks into a room with n susceptibles. Suppose that we perform n $\text{Bern}(p_1)$ trials to determine who, if anyone, becomes infected. Now suppose we run another simulation, but this time with a new probability of infection $p_2 > p_1$. We will show how to reuse many of the PRNs from the initial simulation to do the p_2 version of the job with less effort.

In what follows, we illustrate the workings of the “green” reuse of PRNs. We begin with a baseline example in which, on any given day, each infective has probability $p = 0.2$ of infecting any remaining susceptible. The simulation begins with $I(0) = 1$ infected person and $S(0) = 15$ susceptibles at the end of Day 0 / beginning of Day 1. The number of infectives at the beginning of Day t is $I(t - 1)$, and the probability that any particular susceptible will be infected by at least one of those infectives is $\tilde{p}(t) = 1 - q^{I(t-1)}$, as explained in Item 2 above. The numbers of individuals in states S, I, and R at the end of day t are $S(t)$, $I(t)$, and $R(t)$, respectively. Bern($\tilde{p}(t)$) trials are conducted until no infectives remain. Infectives stay infectious for exactly one day, whereupon they transit to the removed state.

Table 1 keeps track of one realization of the simulation. Day 1 starts with $S(0) = 15$ Bern(0.2) trials, since there is $I(0) = 1$ infective entering the room full of susceptibles. Any susceptible corresponding to a PRN < 0.2 becomes infected. For instance, the PRN for individual 6 is 0.082, so he transits to state I; and subsequent Bernoulli trials will no longer be necessary for him. At the end of Day 1, we have $S(1) = 12$ remaining susceptibles, $I(1) = 3$ newly infectious people, and $R(1) = 0$ removed people (the original infected “person 0” is no longer infected, but we will not count him in the $R(1)$ tally). On Day 2, we only need to take 12 Bern(0.488) trials, where $\tilde{p}(2) = 1 - q^{I(1)} = 1 - 0.8^3 = 0.488$. On Day 2, e.g., individual 3 becomes infected since his PRN = 0.027 < 0.488 . The exercise continues until Day 5, whereupon no infectives remain. In total, the pandemic lasted five days, 14 people were infected, and we used 37 PRNs.

Table 1. Baseline example with $S(0) = 15$ initial susceptibles, $p = 0.2$, $I(0) = 1$ initial infective, and length of infection of one day.

Day t $\tilde{p}(t) = 1 - q^{I(t-1)}$	1 0.2	2 0.488	3 0.67232	4 0.67232	5 0.2
Individual	PRN / State	PRN / State	PRN / State	PRN / State	PRN / State
1	0.637 S	0.886 S	0.855 S	0.975 S	0.828 S
2	0.992 S	0.986 S	0.688 S	0.335 I	
3	0.262 S	0.027 I			
4	0.244 S	0.877 S	0.145 I		
5	0.700 S	0.101 I			
6	0.082 I				
7	0.136 I				
8	0.857 S	0.372 I			
9	0.513 S	0.045 I			
10	0.333 S	0.496 S	0.039 I		
11	0.978 S	0.546 S	0.148 I		
12	0.980 S	0.567 S	0.147 I		
13	0.253 S	0.011 I			
14	0.253 S	0.965 S	0.658 I		
15	0.200 I				
$S(t)$	12	7	2	1	1
$I(t)$	3	5	5	1	0
$R(t)$	0	3	8	13	14
# PRNs	15	12	7	2	1

Table 2 illustrates the corresponding “green” realization of the simulation, using the same PRNs as those in Table 1, but with the new probability $p = 0.3$ that an infectious individual will infect any particular susceptible. Day 1 now starts with $S(0) = 15$ Bern(0.3) trials, since there is again $I(0) = 1$ infective entering the room full of susceptibles. Thus, in this case, any susceptible corresponding to a PRN < 0.3 becomes infected. In this riskier environment, seven Bern(0.3) trials—using the same PRNs as before—result in an infection, so subsequent Bernoulli trials will no longer be necessary for those individuals. At the end of Day 1, we have $S(1) = 8$ remaining susceptibles, $I(1) = 7$ newly infectious people, and $R(1) = 0$ removed people. On Day 2, we only need to take seven Bern(0.91765) trials, where $\tilde{p}(2) = 1 - q^{I(1)} = 1 - 0.7^7 = 0.91765$. Since the pandemic progresses more rapidly, we

now stop at Day 4, when no new state changes can occur. In total, the pandemic lasted four days, all 15 people were infected, and we used 24 PRNs (which themselves had already been generated for the previous Table 1 version of the example).

Table 2. “Green” example reusing PRNs from Table 1 with $S(0) = 15$ initial susceptibles, a new base probability of $p = 0.3$, $I(0) = 1$ initial infective, and a length of infection of one day.

Day t $\tilde{p}(t) = 1 - q^{I(t-1)}$	1 0.3	2 0.91765	3 0.91765	4 0.3
Individual	PRN / State	PRN / State	PRN / State	PRN / State
1	0.637 S	0.886 I		R
2	0.992 S	0.986 S	0.688 I	R
3	0.262 I		R	R
4	0.244 I		R	R
5	0.700 S	0.101 I		R
6	0.082 I		R	R
7	0.136 I		R	R
8	0.857 S	0.372 I		R
9	0.513 S	0.045 I		R
10	0.333 S	0.496 I		R
11	0.978 S	0.546 I		R
12	0.980 S	0.567 I		R
13	0.253 I		R	R
14	0.253 I		R	R
15	0.200 I		R	R
$S(t)$	8	1	0	0
$I(t)$	7	7	1	0
$R(t)$	0	7	14	15
# PRNs	15	8	1	0

Remark 1. A number of comments are in order here:

1. Generally speaking, the pandemic progresses quickly if p is very small (in which case the pandemic dies out rapidly) or very large (in which case the pandemic rapidly infects many people). Thus, for p very small or very large, we would expect to use only a small number of PRNs. Intermediate values of p might require more PRNs. See Section 3, Example 1.
2. How many PRNs will we need to generate in order to simulate multiple scenarios involving the same n susceptible individuals, with each scenario having a different parameter value, for instance, $p = 0.01, 0.02, \dots, 0.30$? With Remark 1.1 in mind, it might be prudent to generate an initial set of nT PRNs for possible use in any of the scenarios, where T is the maximum time horizon of interest. It is almost certain that many of the nT PRNs will never be used if, e.g., T is conservatively chosen to be so high as to guarantee that the pandemic will have run its course. However, if we are running a large number of scenarios, then the expected savings garnered by the use of green simulation would likely mitigate the conservativeness of T and the corresponding one-time charge of generating nT PRNs. For example, for the scenarios presented in Tables 1 and 2, the maximum possible number of days for the scenarios to run is 15, in which case $T = 15$ would be a conservative choice. See Section 4 for more insight on this matter.
3. It is possible (in rare cases) to obtain sample paths in which an individual becomes infectious on a certain day when simulating the Bernoulli trials for a particular probability parameter p_1 , but remains susceptible on that day when the simulation is re-run under a scenario with a larger probability parameter $p_2 > p_1$. For such an anomalous example, see the sample paths of Individual 3 in Tables 3 and 4 for the cases $p = 0.20$ and 0.25 , respectively; and see Section 4 for additional discussion.

Table 3. Remark 1.3 anomalous example with $S(0) = 6$ initial susceptibles, $p = 0.2$, $I(0) = 1$ initial infective, and length of infection of one day. For this sample path, Individual 3 is infectious at the end of Day 3.

Day t $\tilde{p}(t) = 1 - q^{I(t-1)}$	1 0.2	2 0.2	3 0.36
Individual	PRN / State	PRN / State	PRN / State
1	0.24 S	0.44 S	0.72 S
2	0.23 S	0.19 I	R
3	0.65 S	0.95 S	0.29 I
4	0.18 I	R	R
5	0.21 S	0.18 I	R
6	0.42 S	0.55 S	0.38 S

Table 4. Remark 1.3 anomalous example with $S(0) = 6$ initial susceptibles, $p = 0.25$, $I(0) = 1$ initial infective, and length of infection of one day. For this sample path, Individual 3 is susceptible at the end of Day 3.

Day t $\tilde{p}(t) = 1 - q^{I(t-1)}$	1 0.25	2 0.6836	3 0.25
Individual	PRN / State	PRN / State	PRN / State
1	0.24 I	R	R
2	0.23 I	R	R
3	0.65 S	0.95 S	0.29 S
4	0.18 I	R	R
5	0.21 I	R	R
6	0.42 S	0.55 I	R

3. Examples

With the mechanics of the methodology established, we put forth in this section a series of examples to assess the performance of the proposed green method, including its efficiency as compared to traditional (naive) simulation. Examples 1 and 2 in Section 3.1 give exact small-sample results that allow us to succinctly see how various system performance characteristics (e.g., the expected number of individuals infected and the expected length of the pandemic) are affected as the Bernoulli p parameter varies. Although these sorts of results are interesting in and of themselves, they also provide partial validation for the subsequent simulations that we will run. Such exact results might even allow us to avoid simulation altogether—but only for extremely small population sizes. Example 3 in Section 3.2 is concerned with a cohort size ($n = 100$ susceptibles) that is large enough so that we are forced to use simulation to assess system performance as we vary p . This example is useful for indicating what typical histograms might look like for the system performance characteristics of interest. Finally, Section 3.3 presents a number of results that directly address the contributions of our green simulation technology. Namely, Example 4 uses simulation to establish run-time savings on a larger, more-complicated scenario involving a cohort of size $n = 1000$; and Example 5 illustrates how we can use simulation variance reduction methods to conduct certain sensitivity analyses on the various performance characteristics.

3.1. Some Small-Sample Exact Results

First of all, it is natural to ask if it is possible to obtain exact analytical results for the problem at hand using only elementary combinatoric arguments. Indeed, one can sometimes carry out the necessary calculations, though these soon become too tedious even for very small problem instances.

Example 1. Suppose that $n = 3$ susceptibles are in a room, and they are exposed to one infective on Day 1. Moreover, suppose that each infective stays infectious for exactly one day (so that he only has one chance to infect his friends). Let D denote the length of the pandemic (i.e., how long the pandemic lasts until no infectives remain). In addition, let I denote the number of people who ultimately get infected (not counting person 0). Then, some slightly tedious calculations (algebra not shown) reveal that

$$E[D] = 1 + 3p + 3p^2 - 18p^3 + 18p^4 - 6p^5$$

and

$$E[I] = 3p + 6p^2 - 21p^4 + 21p^5 - 6p^6.$$

See Figures 1 and 2, which respectively illustrate the expected length of the outbreak and the expected number of individuals ultimately infected.

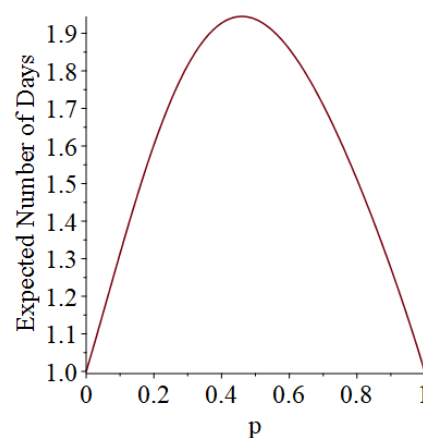


Figure 1. Exact expected length of pandemic (in days) for the case $n = 3$ with a one-day infectiousness period (Example 1).

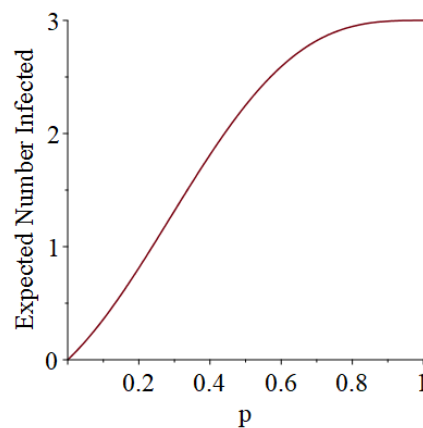


Figure 2. Exact expected number of infected individuals for the case $n = 3$ with a one-day infectiousness period (Example 1).

Let B denote the number of Bernoulli trial comparisons that would have to be made in a simulation of this scenario, where we use the “super-infective” trick described in Item 2 of Section 2. That is, B is the total number of days all of the susceptibles will be exposed to potential infection during the course of the pandemic. (The quantity B informally represents the simulation effort we would have to undertake.) Then, it turns out that

$$E[B] = 3 + 6p + 3p^2 - 15p^3 + 18p^4 - 6p^5. \quad \square$$

Example 2. Suppose that $n = 3$ susceptibles are in a room, and they are exposed to one infective on Day 1. However, now, each infective stays infectious for exactly two days (so that he gets two chances to infect his friends). Then, after a corvée of tedious calculations, we obtain

$$E[D] = 1 + 3q - 3q^2 + q^3 + 12q^4 - 15q^5 + 15q^6 - 12q^7 + 6q^8 - 18q^9 + 6q^{10} + 6q^{11}$$

and

$$E[I] = 3 - 6q^6 - 6q^8 + 15q^{10} - 6q^{12}. \quad \square$$

See Figures 3 and 4.

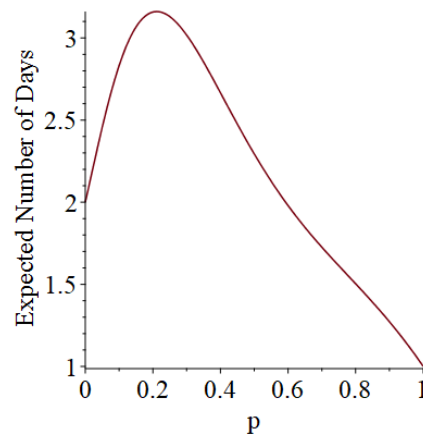


Figure 3. Exact expected length of pandemic (in days) for the case $n = 3$ with a two-day infectiousness period (Example 2).

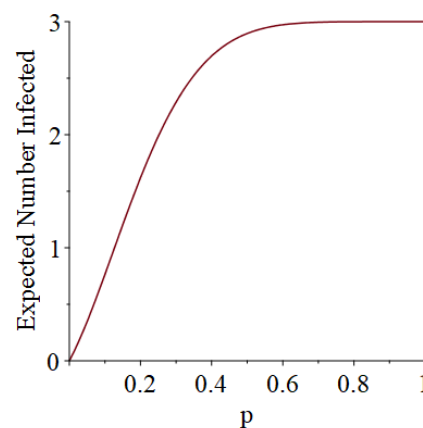


Figure 4. Exact expected number of infected individuals for the case $n = 3$ with a two-day infectiousness period (Example 2).

3.2. Simulation

It ought to be clear from Examples 1 and 2 in Section 3.1 that even small-sample exact calculations present combinatorial challenges. Luckily, it is easy to carry out simulations for more-substantive examples involving larger values of the cohort size n . In this section, we consider a simple example involving a non-trivial cohort size so as to illustrate what typical disease characteristic histograms might look like for moderate values of n , in particular, as we vary the infectiousness parameter p .

Example 3. Suppose that $n = 100$ susceptibles are in a facility, and all are exposed to one infective on Day 1; the probability that a susceptible becomes infected upon exposure to an infective is $p = 0.01$; and each infective stays infectious for exactly one day. Figures 5–7 respectively depict histograms of the length of the pandemic (in days), the total number of individuals who eventually become infected (out of 100), and the number of PRNs used in the simulation. These histograms are based on 1,000,000 independent replications of the simulated pandemic. Figures 8–10 depict the same outcome measures as the previous three, except that this latter set of figures is for the case $p = 0.05$. By design, replication i of the $p = 0.05$ runs used many of the same green PRNs as replication i of the $p = 0.01$ runs, $i = 1, 2, \dots, 10^6$.

Figure 5 shows that, for small $p = 0.01$, the sample probability mass function (p.m.f.) of the length of the pandemic dies off quickly. It is actually possible for the pandemic to end after only one day—with probability $(1 - 0.01)^{100} = 0.366 \approx 1/e$ —but the sample mean of the length of the pandemic is $\bar{D} = 3.88$ with a sample standard deviation of $S_D = 3.72$. Figure 6 reveals a similar p.m.f. for the number of individuals who become infected; here $\bar{I} = 6.67$ and $S_I = 10.93$, with the large standard deviation owing to a long right-hand tail. Figure 7 illustrates an interesting p.m.f. for the number B of PRNs (Bernoulli trials) used in the course of the simulations. That p.m.f. is characterized by distinct chunks of PRNs corresponding to the first few days of the pandemic; and we find that $\bar{B} = 355.7$ and $S_B = 310.7$.

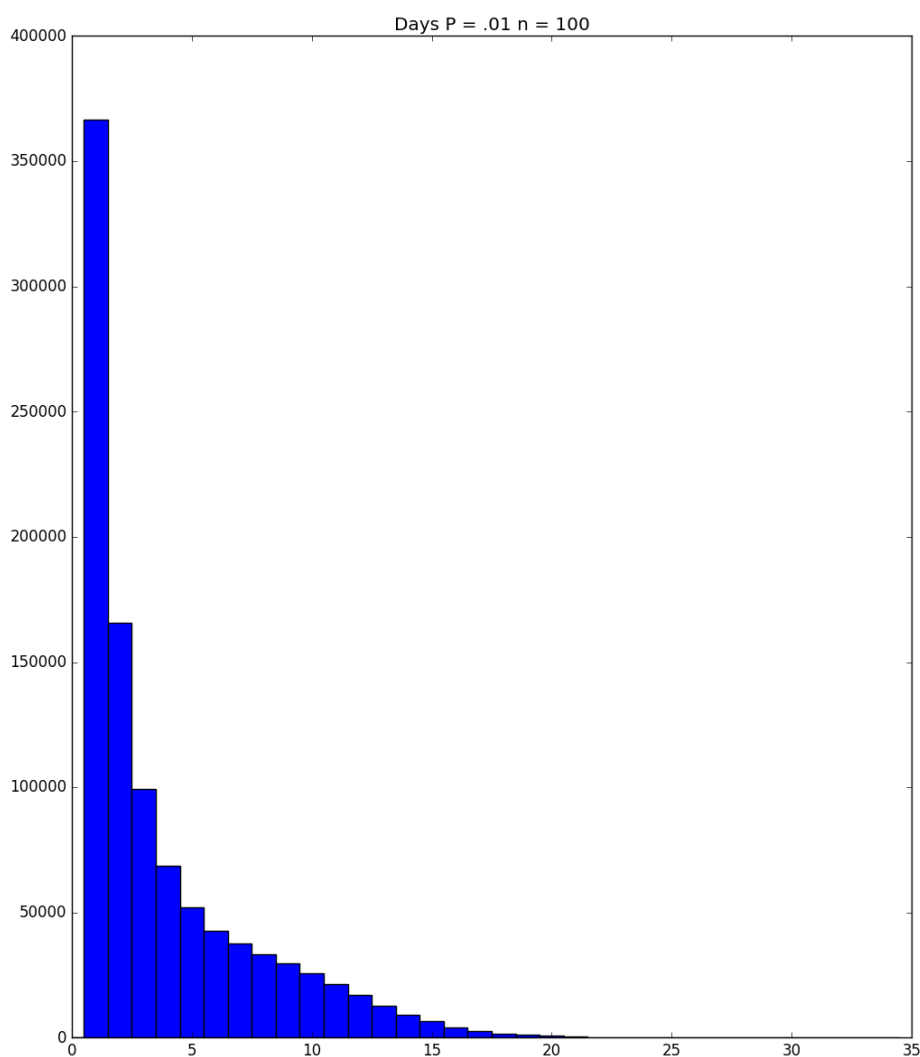


Figure 5. Histogram of length of pandemic (in days) for the case $n = 100$, $p = 0.01$, one initial infective, and a one-day infectiousness period (based on 1,000,000 replications).

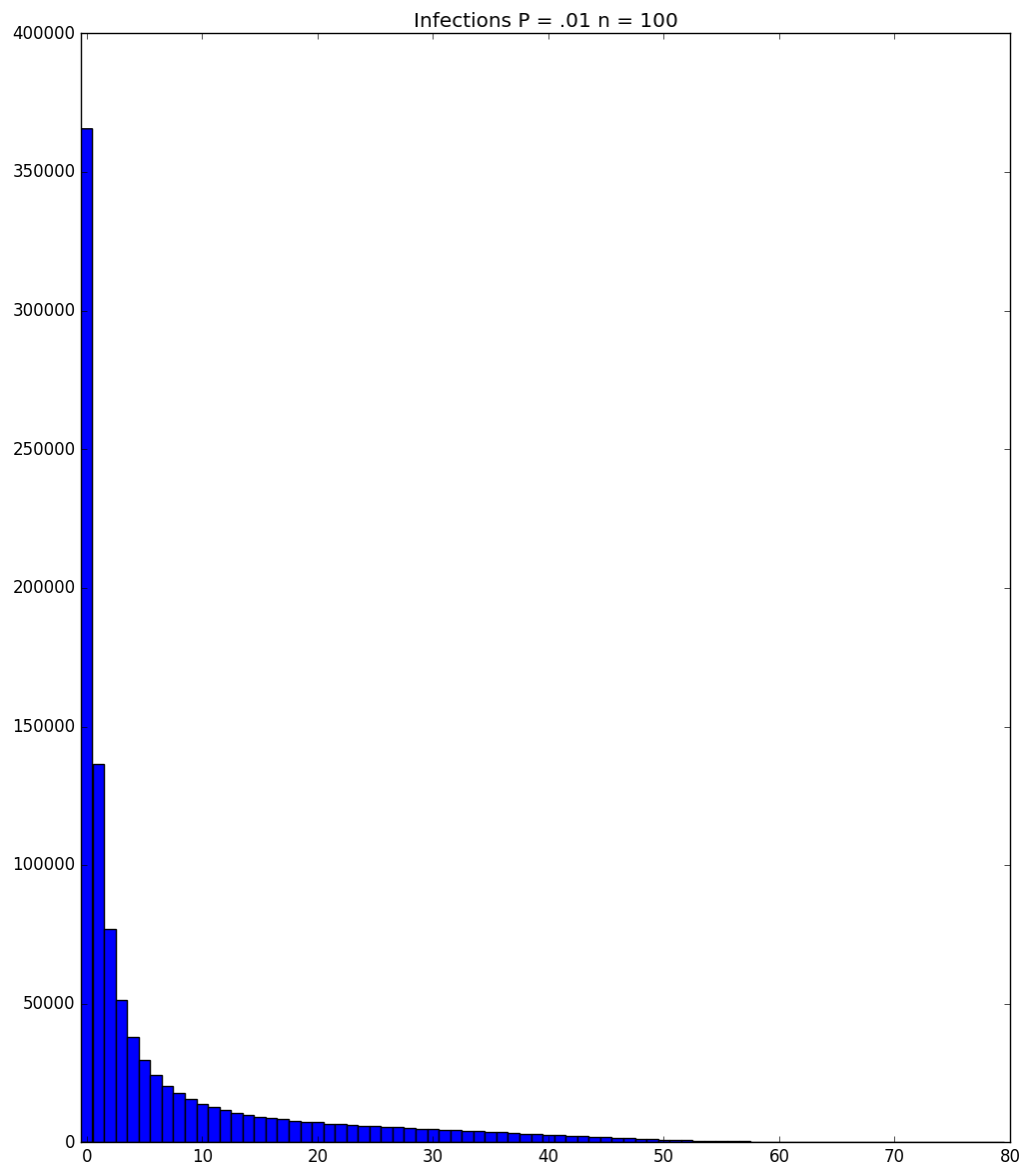


Figure 6. Histogram of number of individuals ultimately infected for the case $n = 100$, $p = 0.01$, one initial infective, and a one-day infectiousness period (based on 1,000,000 replications).

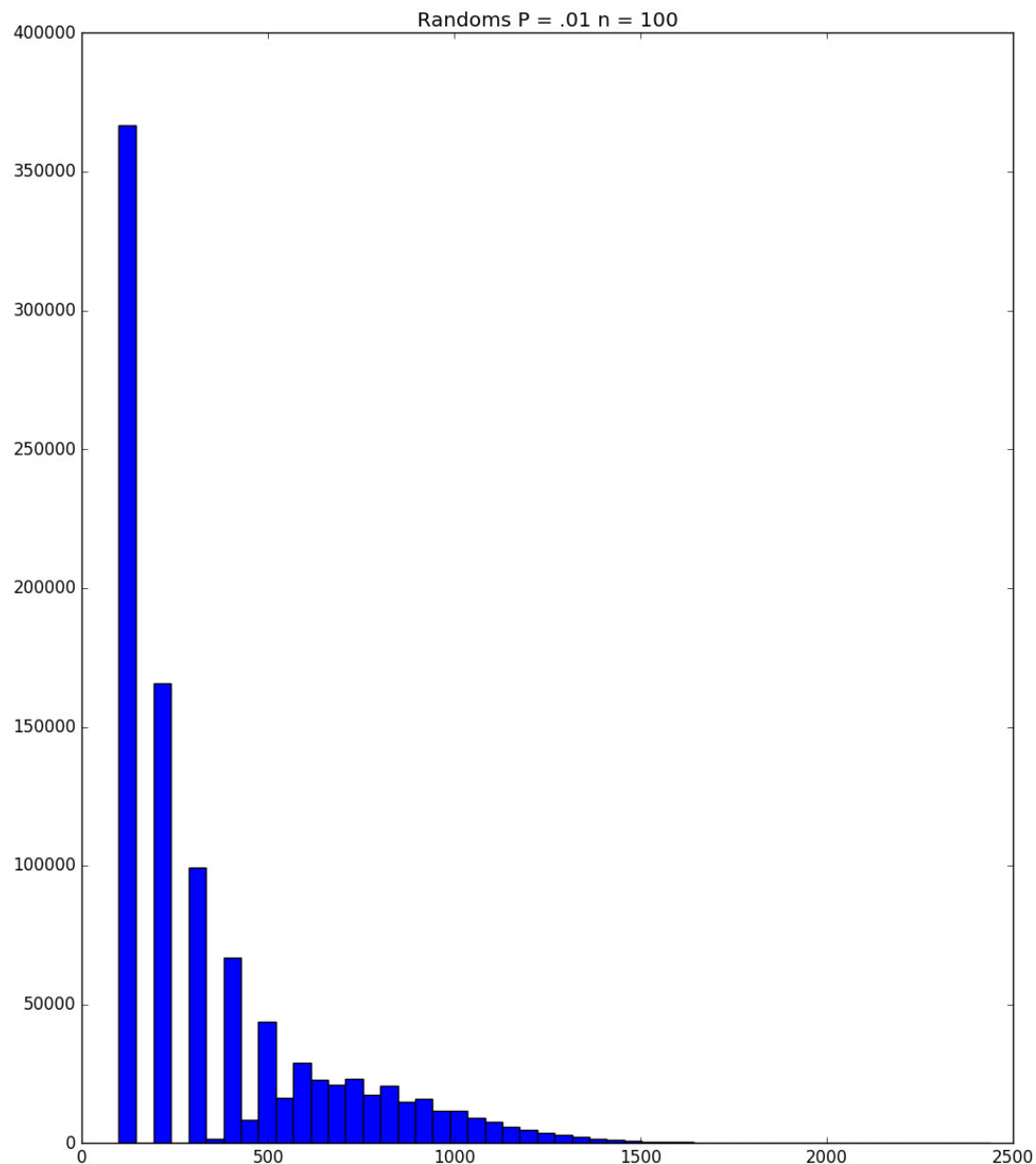


Figure 7. Histogram of the total number of PRNs used for the case $n = 100$, $p = 0.01$, one initial infective, and a one-day infectiousness period (based on 1,000,000 replications).

Figure 8 shows that, for the much higher $p = 0.05$, the *p.m.f.* of D is less variable, owing to the fact that the disease propagates throughout the population quite quickly, so that $\bar{D} = 3.25$ and $S_D = 0.88$. Figure 9 reveals that almost everyone eventually becomes infected, with $\bar{I} = 98.78$ and $S_I = 7.78$. Figure 10 illustrates a much tighter *p.m.f.* for the number of PRNs used in the $p = 0.05$ case (compared to the $p = 0.01$ case in Figure 7); here, we have $\bar{B} = 302.0$ and $S_B = 38.2$. \square

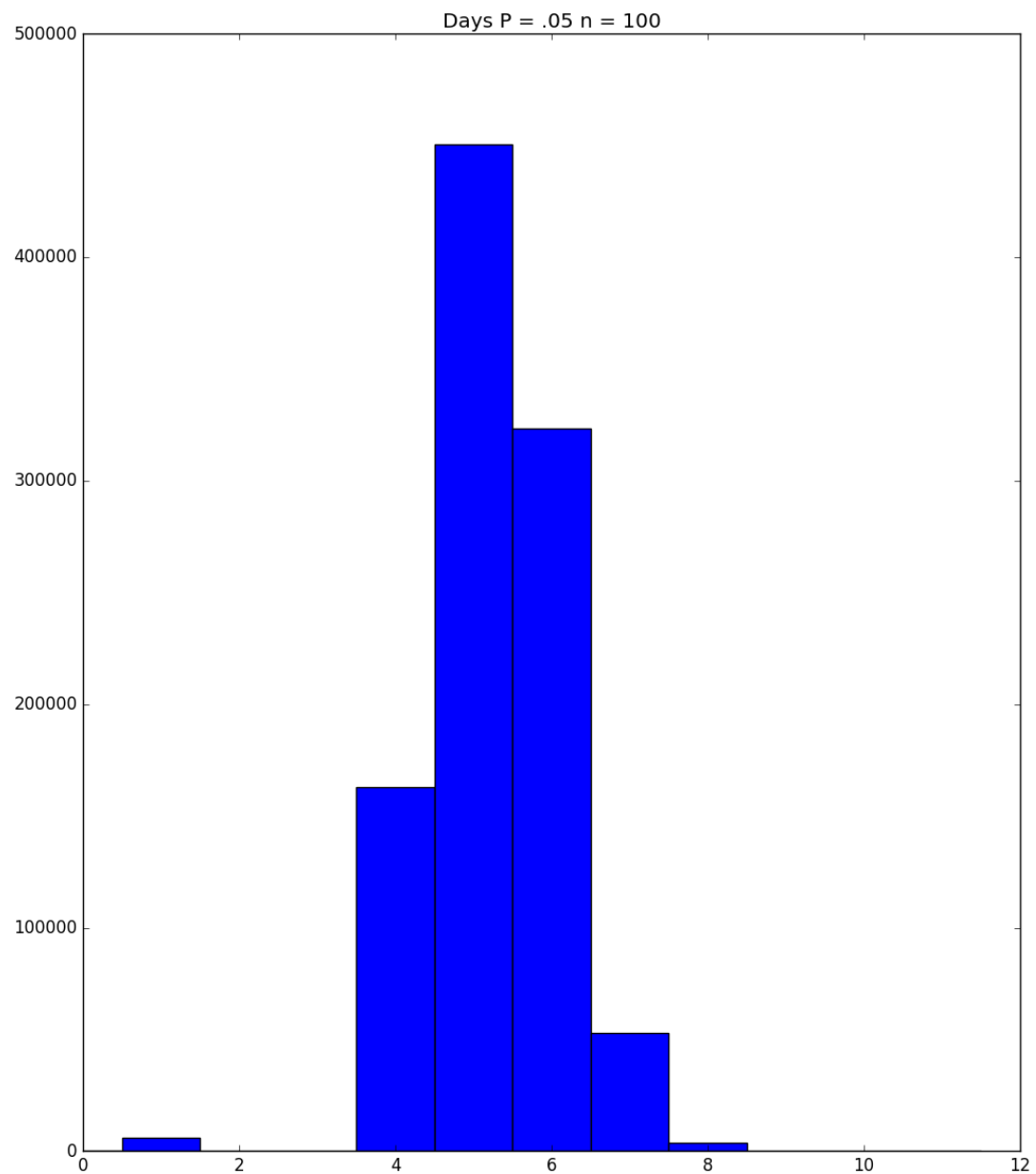


Figure 8. Histogram of length of pandemic (in days) for the case $n = 100$, $p = 0.05$, one initial infective, and a one-day infectiousness period (based on 1,000,000 replications).

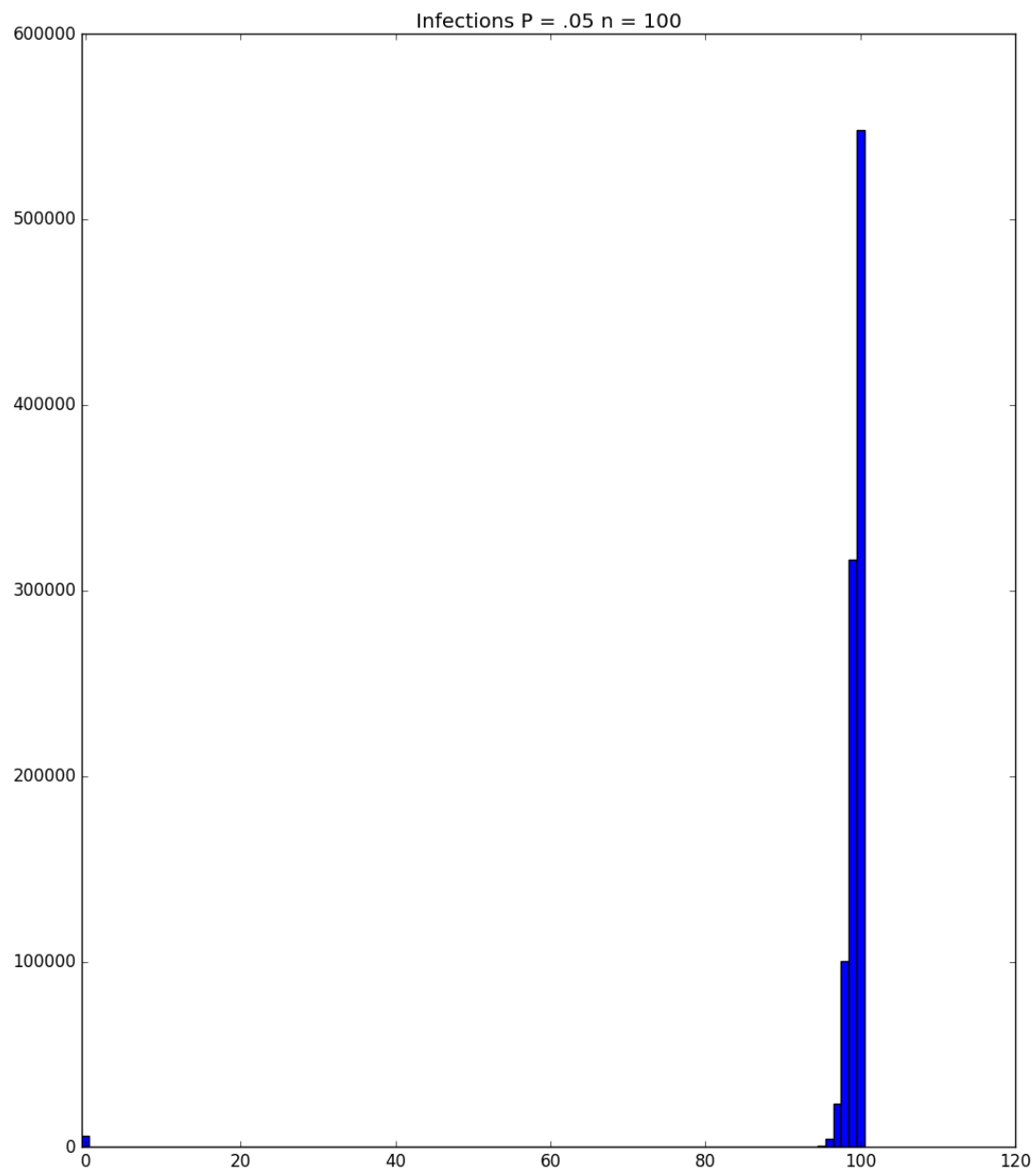


Figure 9. Histogram of number of individuals ultimately infected for the case $n = 100$, $p = 0.05$, one initial infective, and a one-day infectiousness period (based on 1,000,000 replications).

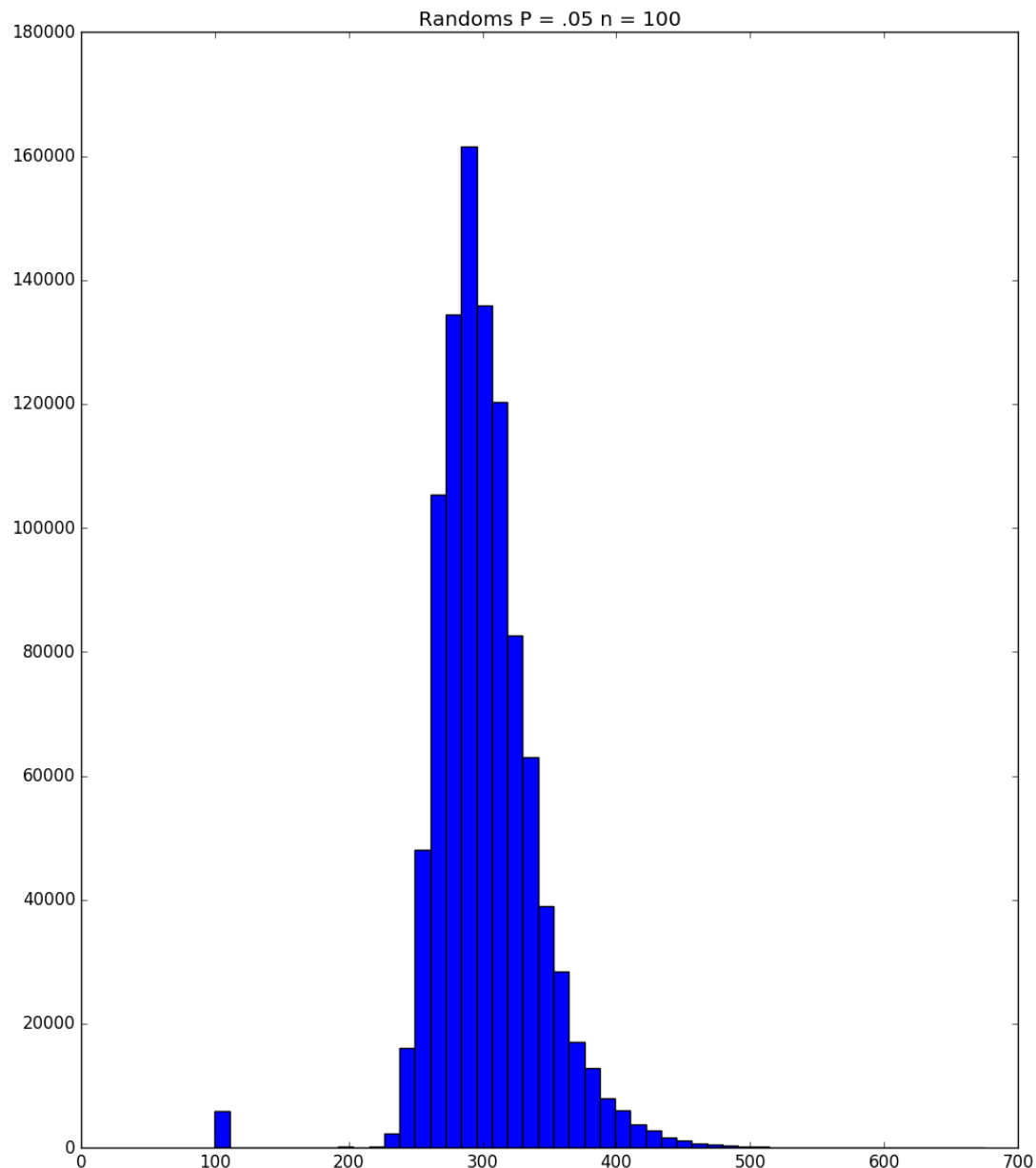


Figure 10. Histogram of the total number of PRNs used for the case $n = 100$, $p = 0.05$, one initial infective, and a one-day infectiousness period (based on 1,000,000 replications).

3.3. Run Time and Efficiency Considerations

The pandemic simulation effort requires more than simply generating Bernoulli trials. One also has to make decisions, update statistics, and access saved observations from temporary storage. How do these miscellaneous tasks affect the overall computation time with respect to green simulation? In this section, we present a number of results on the efficacy of our green simulation technology. Namely, Example 4 discusses simulation run-time savings on a more-realistic mixing example that what we have seen so far, and Example 5 illustrates how we can use variance reduction methods to carry out sensitivity analysis.

Example 4. Clock-Time Savings. Consider a school with 900 students and 100 staff members. The students are divided into 30 classrooms, each with 30 students; and each student remains in his/her assigned classroom all day. Each of the 100 staff members are randomly assigned 10 classrooms to visit during the day (and will visit those same assigned classrooms every day). Thus, each student interacts with the other 29 kids in his/her classroom as well as with the staff members who visit the classroom each day; and each staff member interacts with the other 99 staff members as well as with the 300 students whom the staff member visits in his/her assigned 10 classrooms. On Day 1, all 1000 people in the cohort are susceptibles. Unfortunately, the school principal (person 0) is infectious and visits all 1000 susceptibles on Day 1, any of whom subsequently become infected with probability p . The principal is so sick that he does not return after Day 1. However, anyone who gets infected becomes infectious the next day and remains infectious for a total of three days. We ran 200 replications of both the traditional (naïve) and green versions of the simulation for each of the five choices of $p = 0.020, 0.021, \dots, 0.024$.

To get an idea of what we are facing, Figures 11–13, respectively, depict histograms for the length of the pandemic, number of infections that occur, and the number of PRNs required for the baseline $p = 0.02$ scenario for this example. In any case, the $5 \times 200 = 1000$ traditional replications required about 118 min of real clock time on a standard Windows-based PC; and the corresponding 1000 green replications required about 103 min—i.e., a savings of about 15% over running the five sets of experiments individually. \square

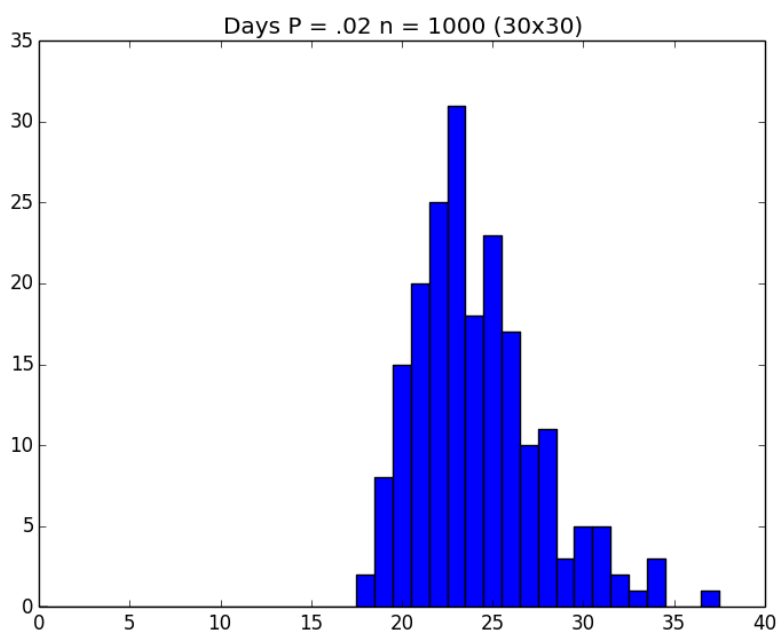


Figure 11. Histogram of length of pandemic (in days) for the case $n = 1000$, $p = 0.02$, one initial infective, and a three-day infectiousness period, as described in Example 4 (based on 200 replications).

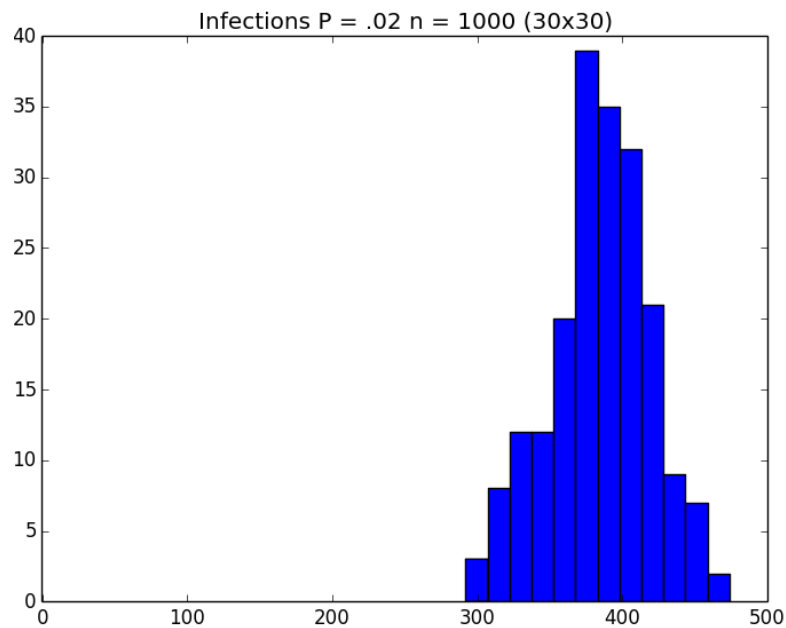


Figure 12. Histogram of number of individuals ultimately infected for the case $n = 1000$, $p = 0.02$, one initial infective, and a three-day infectiousness period, as described in Example 4 (based on 200 replications).

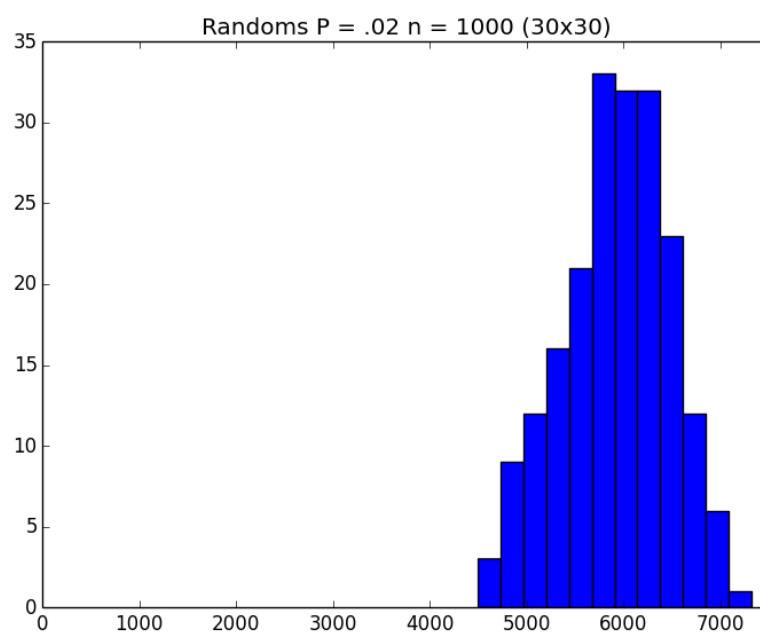


Figure 13. Histogram of the total number of PRNs used for the case $n = 1000$, $p = 0.02$, one initial infective, and a three-day infectiousness period, as described in Example 4 (based on 200 replications).

Example 5. Sensitivity and Variance Reduction. The fact that we incorporate many of the same PRNs between runs of the simulation that use different values for parameter p suggests that the outputs from the runs may be positively correlated. For instance, suppose that \bar{I}_p denotes the sample mean of the number of individuals infected during the course of the simulation when the baseline Bernoulli parameter is p . If $p_1 \approx p_2$ so that \bar{I}_{p_1} and \bar{I}_{p_2} are presumably positively correlated, then $\bar{I}_{p_1} - \bar{I}_{p_2}$ is an unbiased estimator for the true mean difference $E[\bar{I}_{p_1} - \bar{I}_{p_2}]$ and, notably,

$$\begin{aligned} \text{Var}(\bar{I}_{p_1} - \bar{I}_{p_2}) &= \text{Var}(\bar{I}_{p_1}) + \text{Var}(\bar{I}_{p_2}) - 2\text{Cov}(\bar{I}_{p_1}, \bar{I}_{p_2}) \\ &\leq \text{Var}(\bar{I}_{p_1}) + \text{Var}(\bar{I}_{p_2}), \end{aligned}$$

so that the estimator for the mean difference has a smaller variance than would be observed if the runs had been undertaken independently—a standard simulation variance reduction technique known as common random numbers (see Law [14]). As a bonus, we immediately obtain an estimator for the derivative dI_p/dp in the vicinity of $p_1 \approx p_2$,

$$\frac{\widehat{dI}_p}{dp} = \frac{\bar{I}_{p_1} - \bar{I}_{p_2}}{p_1 - p_2}.$$

Such derivatives are useful for sensitivity analysis, and are especially so if they are of reduced variance.

To illustrate this approach, suppose we simulate each of two scenarios $p_1 = 0.070$ and $p_2 = 0.071$, both involving $n = 30$ individuals, one initial infective, and a one-day infectiousness period. We simulated both of the scenarios with three sets of 100,000 replications to examine the variability of the outcomes under two sampling schemes: The two scenarios are carried out independently (Case A), and the scenarios are carried out in green fashion (Case B). Table 5 shows that the differences produced by the green simulations have less variability than those produced by Case A. \square

Table 5. Sensitivity example with $n = 30$ initial susceptibles, $p_1 = 0.070$, $p_2 = 0.071$, $I(0) = 1$ initial infective, and infectiousness period of one day. Three sets of 100,000 independent replications were carried out for Case A (independent scenarios) and Case B (green scenarios).

Set	Case A (Indep Scenarios)				Case B (Green Scenarios)			
	\bar{I}_{p_1}	\bar{I}_{p_2}	$\bar{I}_{p_1} - \bar{I}_{p_2}$	\widehat{dI}_p/dp	\bar{I}_{p_1}	\bar{I}_{p_2}	$\bar{I}_{p_1} - \bar{I}_{p_2}$	\widehat{dI}_p/dp
1	20.5945	21.0066	−0.4121	412	20.5593	20.9086	−0.3493	349
2	20.5325	20.8906	−0.3581	358	20.5773	20.9409	−0.3636	364
3	20.5611	20.8921	−0.3309	331	20.5604	20.9071	−0.3467	347

4. Conclusions and Future Work

We have shown how to conduct “green” simulations of the propagation of an infectious disease by reusing certain random variates under different input parameter settings. This reuse avoids some of the computation time that would be dedicated to generating a new set of random variates each time the parameters change. Implementation is straightforward, and computation time savings can be significant. Moreover, the green simulation can be regarded as a variance reduction technique when used for purposes of sensitivity analysis.

In terms of future research, there is still a great deal that we can do:

1. Item 1 of Section 2 discusses the use of a single $\text{Bin}(n, p)$ random variable instead of n $\text{Bern}(p)$ trials to determine the number of infections that occur on a given day. We can immediately implement efficient methods for generating the $\text{Bin}(n, p)$ random variables (Kachitvichyanukul and Schmeiser [17,18]), but we might also apply our green technology for $\text{Bin}(n, p)$ generation across scenarios involving different values of p .
2. Remark 1.2 in Section 2 suggests that one could pre-generate nT PRNs to be used on all scenarios of the simulation, but that nT may be a needlessly high number. In fact, one might instead generate each PRN only on a strict as-needed basis.
3. Remark 1.3 notes that it is possible (because of how we handle “super-infectives”) for an individual who is infected on a certain day under the p_1 scenario to at least temporarily escape infection for the $p_2 (> p_1)$ scenario. We are currently investigating ways to identify and take advantage of these (rare) events.
4. We are in the process of performing expanded analyses (i) on much larger population sizes, (ii) with a greater variety of parameter choices (including different probabilities of infection for different sets of infectives and susceptibles), and (iii) under more-general compartmental environments (e.g., individuals can move between home, school, and work on any given day).

5. The common random variates variance reduction technique is one of several that are popular in the literature. A worthy endeavor is that of incorporating other such methods in the context of disease propagation simulation.
6. Ultimately, we will use our efficient simulations to evaluate the effects of various mitigation strategies on outbreaks.

Author Contributions: The authors contributed in the following ways in the various phases of this research: conceptualization, S.W., A.A., and D.G.; methodology, S.W., A.A., and D.G.; software, S.W. and D.G.; validation, S.W., A.A., and D.G.; formal analysis, S.W., A.A., and D.G.; investigation, S.W., A.A., and D.G.; resources, S.W., A.A., and D.G.; data curation, N/A; writing—original draft preparation, S.W., A.A., and D.G.; writing—review and editing, S.W., A.A., and D.G.; visualization, S.W., A.A., and D.G.; supervision, A.A. and D.G.; project administration, A.A. and D.G.; funding acquisition, D.G.

Funding: This research was partially funded by National Science Foundation Grant No. CMMI-1233141.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Bern	Bernoulli random variable
Bin	Binomial random variable
p.m.f.	probability mass function
PRN	Pseudo-Random Number
SIR	Susceptible-Infected-Recovered
Unif	Uniform random variable

References

1. Abbott, A.; Pearson, H. Fear of human pandemic grows as bird flu sweeps through Asia. *Nature* **2004**, *427*, 472–473. [[CrossRef](#)] [[PubMed](#)]
2. Ferguson, N.M.; Cummings, D.A.; Cauchemez, S.; Fraser, C.; Riley, S.; Meeyai, A.; Iamsirithaworn, S.; Burke, D.S. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* **2015**, *437*, 209–214. [[CrossRef](#)] [[PubMed](#)]
3. Andradóttir, S.; Chiu, W.; Goldsman, D.; Lee, M.L.; Tsui, K.-L.; Sander, B.; Fisman, D.N.; Nizam, A. Reactive strategies for containing developing outbreaks of pandemic influenza. *BMC Public Health* **2011**, *11*, S1. [[CrossRef](#)] [[PubMed](#)]
4. Kermack, W.O.; McKendrick, A.G. Contributions of mathematical theory to epidemics. *Proc. R. Soc. Lond. Ser. A* **1927**, *115*, 700–721. [[CrossRef](#)]
5. Elveback, L.R.; Fox, J.P.; Ackerman, E.; Langworthy, A.; Boyd, M.; Gatewood, L. An influenza simulation model for immunization studies. *Am. J. Epidemiol.* **1976**, *103*, 152–165. [[CrossRef](#)] [[PubMed](#)]
6. Halloran, M.E.; Longini, I.M.; Nizam, A.; Yang, Y. Containing bioterrorist smallpox. *Science* **2002**, *128*, 1428–1432. [[CrossRef](#)] [[PubMed](#)]
7. Kelso, J.; Milne, G.; Kelly, H. Simulation suggests that rapid activation of social distancing can arrest epidemic development due to a novel strain of influenza. *BMC Public Health* **2009**, *9*, 117. [[CrossRef](#)] [[PubMed](#)]
8. Lee, B.Y.; Brown, S.T.; Cooley, P.C.; Zimmerman, R.K.; Wheaton, W.D.; Zimmer, S.M.; Grefenstette, J.J.; Assi, T.-M.; Furphy, T.J.; Wagener, D.K.; et al. A computer simulation of employee vaccination to mitigate an influenza epidemic. *Am. J. Prev. Med.* **2010**, *38*, 247–257. [[CrossRef](#)] [[PubMed](#)]
9. Longini, I.M.; Halloran, M.E.; Nizam, A.; Yang, Y. Containing pandemic influenza with antiviral agents. *Am. J. Epidemiol.* **2004**, *159*, 623–633. [[CrossRef](#)] [[PubMed](#)]
10. Longini, I.M.; Nizam, A.; Xu, S.; Ungchusak, K.; Hanshaoworakul, W.; Cummings, D.A.; Halloran, M.E. Containing pandemic influenza at the source. *Science* **2005**, *309*, 1083–1087. [[CrossRef](#)] [[PubMed](#)]
11. Andradóttir, S.; Chiu, W.; Goldsman, D.; Lee, M.L. Simulation of influenza propagation: Model development, parameter estimation, and mitigation strategies. *IIE Trans. Healthc. Syst. Eng.* **2014**, *4*, 27–48. [[CrossRef](#)]

12. Feng, M.; Staum, J. Green simulation designs for repeated experiments. In *Proceedings of the 2015 Winter Simulation Conference*; Yilmaz, L., Chan, W.K.V., Moon, I., Roeder, T.M.K., Macal, C., Rossetti, M.D., Eds.; Institute of Electrical and Electronics Engineers: Piscataway, NJ, USA, 2015; pp. 403–413.
13. Meterelliyoz, M.; Alexopoulos, C.; Goldsman, D. Folded overlapping variance estimators for simulation. *Eur. J. Oper. Res.* **2012**, *220*, 135–146. [[CrossRef](#)]
14. Law, A.M. *Simulation Modeling and Analysis*, 5th ed.; McGraw-Hill Education: New York, NY, USA, 2015.
15. Tsai, M.T.; Chern, T.C.; Chuang, J.H.; Hsueh, C.W.; Kuo, H.S.; Liao, C.J.; Riley, S.; Shen, B.J.; Shen, C.H.; Wang, D.W.; et al. Efficient simulation of the spatial transmission dynamics of influenza. *PLoS ONE* **2010**, *5*, e13292. [[CrossRef](#)] [[PubMed](#)]
16. Shen, X.; Wong, Z.S.-Y.; Ling, M.-H.; Goldsman, D.; Tsui, K.-L. Comparison of algorithms to simulate disease transmission. *J. Simul.* **2017**, *11*, 285–294. [[CrossRef](#)]
17. Kachitvichyanukul, V.; Schmeiser, B.W. Binomial random variate generation. *Commun. ACM* **1988**, *31*, 216–222. [[CrossRef](#)]
18. Kachitvichyanukul, V.; Schmeiser, B.W. Algorithm 678 BTPEC: Sampling from the binomial distribution. *ACM Trans. Math. Softw.* **1989**, *15*, 394–397. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).