

MA3227 Numerical Analysis II

Lecture 24: Stochastic SIR Model

Simon Etter



2019/2020

Stochastic SIR Model

Disclaimer

This lecture illustrates the theory developed in Lectures 22 and 23 by means of a concrete, real-world example.

The deterministic SIR model

Consider a population of N people which are exposed to a novel disease. We assume that at any given moment, the population can be split into three groups:

Susceptible, Infected, Recovered.

Let us denote by $S(t)$, $I(t)$ and $R(t)$ the number of people in each group, and let us assume that these quantities satisfy the system of ODEs

$$\dot{S} = -\frac{k}{N} SI, \quad \dot{I} = \frac{k}{N} SI - I, \quad \dot{R} = I \quad \text{for some } k > 0.$$

These equations are known as the (deterministic) SIR model.

We will see later why sometimes we need the “deterministic” qualifier to refer to the above model.

Stochastic SIR Model

The deterministic SIR model (continued)

SIR model copied from above:

$$\dot{S} = -\frac{k}{N} SI, \quad \dot{I} = \frac{k}{N} SI - I, \quad \dot{R} = I \quad \text{for some } k > 0.$$

The SIR model is a very crude approximation to a complicated real-world process, but it captures the correct qualitative behaviour. In particular:

- ▶ The $\pm \frac{k}{N} SI$ terms say that people move from the Susceptible to the Infected group as long as there are both susceptible and infected people in the population. Moreover, the disease spreads faster the more susceptible and infected people there are.
- ▶ The $\pm I$ terms say that the number of people recovering during a small time interval is proportional to the number of infected people.
- ▶ Once people reach the Recovered group, they are immune to the disease and cannot be infected anymore.

Stochastic SIR Model

Shortcomings of the deterministic SIR model

SIR model copied from above:

$$\dot{S} = -\frac{k}{N} SI, \quad \dot{I} = \frac{k}{N} SI - I, \quad \dot{R} = I \quad \text{for some } k > 0.$$

Some noticeable shortcomings of the SIR model are:

- ▶ The recovery terms $\pm I$ should include a rate constant analogous to the k in the infection terms $\pm \frac{k}{N} SI$. This would be easy to add, but I omitted this generality since from a qualitative point of view, the only thing which matters is whether the rate of infection is faster ($k > 1$) or slower ($k < 1$) than the rate of recovery.
- ▶ There is no lag between getting infected, infecting other people and recovering.
- ▶ S , I and R are all treated as real numbers when actually they should be integers. Implicitly, this assumes that S and I are sufficiently large that we can afford to ignore the less significant digits of S and I .
- ▶ The model does not take into account fluctuations in how fast people infect each other and recover.

Stochastic SIR Model

Numerical simulation of the deterministic SIR model

See `sir_ode()` and `simulate()`. Also, note that Julia has a function

The stochastic SIR model

The stochastic SIR model addresses the last two of the shortcomings of the deterministic model listed above.

The idea is to treat S , I and R as time-dependent random variables,

$$S, I, R : [0, T] \times \Omega \rightarrow \{0, \dots, N\},$$

and interpret the terms $\pm \frac{k}{N} SI$ and $\pm I$ as a probability density that a person gets infected or recovers in an infinitely small time interval.

The easiest way to present the details of the stochastic model is to present the algorithm which generates samples of the model.

This algorithm is known as the Gillespie algorithm and described on the next slide.

Stochastic SIR Model

The stochastic SIR model (continued)

Algorithm 1 Gillespie algorithm for the stochastic SIR model

- 1: Input: N, I_0, k, T
 - 2: Set $\ell = 0, \quad t_0 = 0, \quad S_0 = N - I_0, \quad R_0 = 0$
 - 3: **while** $t_\ell < T$ **do**
 - 4: Draw $\Delta t_\ell \sim \text{Exp}\left(\frac{k}{N} S_\ell I_\ell + I_\ell\right)$ and set $t_{\ell+1} = t_\ell + \Delta t_\ell$
 - 5: Draw $P_\ell \sim \text{Bernoulli}\left(\frac{\frac{k}{N} S_\ell I_\ell}{\frac{k}{N} S_\ell I_\ell + I_\ell}\right)$.
 - 6: If $P_\ell = 1$, set $S_{\ell+1} = S_\ell - 1, \quad I_{\ell+1} = I_\ell + 1$.
 - 7: If $P_\ell = 0$, set $I_{\ell+1} = I_\ell - 1, \quad R_{\ell+1} = R_\ell + 1$.
 - 8: Set $\ell = \ell + 1$.
 - 9: **end while**
-

Remarks:

- ▶ $\text{Exp}(\lambda)$ denotes the distribution with PDF $f(x) = \lambda \exp(-\lambda x)$.
- ▶ $X \sim \text{Bernoulli}(p)$ means $P(X = 1) = p, \quad P(X = 0) = 1 - p$.
- ▶ The next slide will discuss how to sample these distributions.

Stochastic SIR Model

Sampling the exponential distribution

The $\text{Exp}(\lambda)$ distribution can easily be sampled using transformation sampling. The CDF is given by

$$F(x) = \int_0^x \lambda \exp(-\lambda y) dy = 1 - \exp(-\lambda x);$$

hence if $U \sim \text{Uniform}[0, 1]$, then

$$X = F^{-1}(U) = -\frac{1}{\lambda} \log(1 - U) \sim \text{Exp}(\lambda).$$

See `exp_sampling()`. Also, note that Julia has a function `randexp()` in the `Random` package.

Sampling the Bernoulli distribution

Let $U \sim \text{Uniform}[0, 1]$ and set

$$X(U) = \begin{cases} 1 & \text{if } U < p, \\ 0 & \text{otherwise.} \end{cases}.$$

Then,

$$P(X = 1) = P(U < p) = p \quad \Longleftrightarrow \quad X \sim \text{Bernoulli}(p).$$

See `bernoulli_sampling()`.

Stochastic SIR Model

Numerical simulation of the stochastic SIR model

See `sir_gillespie()` and `simulate()`.

Discussion: scaling of the SIR model

It is easy to see that the deterministic SIR model is invariant under scaling of the population: if we fix $\alpha > 0$ and set

$$\tilde{N} = \alpha N, \quad \tilde{S}(t) = \alpha S(t), \quad \tilde{I}(t) = \alpha I(t), \quad \tilde{R}(t) = \alpha R(t),$$

then

$$\begin{aligned}\dot{\tilde{S}} &= \alpha \dot{S} = -\alpha \frac{k}{N} SI = -\frac{k}{\alpha N} (\alpha S)(\alpha I) = -\frac{k}{\tilde{N}} \tilde{S} \tilde{I}, \\ \dot{\tilde{I}} &= \alpha \dot{I} = \alpha \frac{k}{N} SI - \alpha I = \frac{k}{\alpha N} (\alpha S)(\alpha I) - \alpha I = \frac{k}{\tilde{N}} \tilde{S} \tilde{I} - \tilde{I}, \\ \dot{\tilde{R}} &= \alpha \dot{R} = \alpha I = \tilde{I},\end{aligned}$$

and

$$\begin{aligned}\tilde{S}(0) &= \alpha S(0) = \alpha (N - I_0) = \tilde{N} - \tilde{I}_0, \\ \tilde{I}(0) &= \alpha I(0) = \alpha I_0 = \tilde{I}_0, \\ \tilde{R}(0) &= \alpha R(0) = 0.\end{aligned}$$

Stochastic SIR Model

Discussion: scaling of the SIR models (continued)

The above means that $\tilde{S}, \tilde{I}, \tilde{R}$ satisfy the same ODEs as S, I, R but with scaled initial conditions. Since solutions to ODEs are unique (assuming the ODE is Lipschitz continuous, which is the case here), this conversely means that we can get the solution for scaled initial conditions simply by scaling the solution.

To see this numerically, replace $N = 1000$ in `simulate()` with e.g. $N = 10_000$ or $N = 100$. We observe that the solution to the deterministic model looks the same in each case on appropriately scaled axes.

The same is not true for the stochastic model: for large population sizes N , the stochastic solutions closely follow the deterministic solutions, but for small population sizes there is significant randomness in the stochastic solutions.

This confirms that using the deterministic model is justified if the population is large, but stochastic models should be used otherwise.

Stochastic SIR Model

Discussion: distribution of $R(\infty)$

To get a better idea of the variations in the stochastic model, we can look at the distribution of $R(\infty)$, which is the total number of people who have ever been infected by the virus. See `Rinf_distribution()`.

We observe that the distribution of $R(\infty)$ is fairly bimodal: in most runs, either $R(\infty) < 20$ or $R(\infty) > 60$.

Discussion: likelihood of pandemic

If you are a person in this population at time $t = 0$, you would probably be keen to know how likely it is that the disease will eventually infect a significant proportion of the population, i.e. you would like to estimate

$$P(R(\infty) \geq R_{\text{thres}}).$$

Given the distribution of $R(\infty)$, it further seems reasonable to choose $R_{\text{thres}} = 40$ (thres is short for threshold).

Stochastic SIR Model

Discussion: likelihood of pandemic

This question can easily be answered using Monte Carlo as follows.

Let us introduce the indicator random variable

$$X = \begin{cases} 1 & \text{if } R(\infty) \geq R_{\text{thres}}, \\ 0 & \text{otherwise.} \end{cases}$$

We then have

$$\begin{aligned} P(R(\infty) \geq R_{\text{thres}}) &= 0 \times P(R(\infty) < R_{\text{thres}}) + 1 \times P(R(\infty) \geq R_{\text{thres}}) \\ &= \mathbb{E}[X] \approx \tilde{\mathbb{E}}_N[X], \end{aligned}$$

i.e. the probability $P(R(\infty) \geq R_{\text{thres}})$ that we would like to compute can be expressed as the expectation of the random variable X , and the expectation in turn can be estimated using Monte Carlo sampling.

This is demonstrated in `p_epidemic()` (ignore the `var_p` and `error` parts for now).

Stochastic SIR Model

Error estimation in Monte Carlo algorithms

Of course, the number produced by `p_epidemic()` is only an estimate for the true value of $\mathbb{E}[X]$, and it would be useful if we could quantify how accurate this estimate is.

In Lecture 22, we have seen that the central limit theorem implies

$$\mathbb{E}\left[\left(\tilde{\mathbb{E}}_N[X] - \mathbb{E}[X]\right)^2\right] = \frac{1}{N} \text{Var}[X];$$

hence an estimate for the expected error can be easily computed once we have an estimate for $\text{Var}[X]$. Such an estimate can be obtained by applying Monte Carlo sampling to

$$\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2],$$

which yields

$$\text{Var}_N[X] = \tilde{\mathbb{E}}_N[(X - \tilde{\mathbb{E}}_N[X])^2].$$

This formula requires samples $X_1, \dots, X_N \stackrel{\text{iid}}{\sim} X$, but such samples have already been computed when evaluating $\tilde{\mathbb{E}}_N[X]$ and we can simply reuse these samples for $\text{Var}_N[X]$.

Stochastic SIR Model

Error estimation in Monte Carlo algorithms (continued)

We therefore conclude that the accuracy of a Monte Carlo estimate can be estimated simply by doing some extra post-processing of the samples X_1, \dots, X_N . See `p_epidemic()` for demonstration.

Remark: formulae for $\text{Var}_N[X]$

The last slide suggested to estimate the variance using

$$\text{Var}_N[X] = \tilde{\mathbb{E}}_N[(X - \tilde{\mathbb{E}}_N[X])^2].$$

Some sources propose to use $\frac{N}{N-1} \text{Var}_N[X]$ instead because it can be shown that $\frac{N}{N-1} \text{Var}_N[X]$ is an unbiased estimator for $\text{Var}[X]$.

Unbiasedness means that $\mathbb{E}[\frac{N}{N-1} \text{Var}_N[X]] = \text{Var}[X]$.

Unbiasedness is important for some applications in statistics, but it is not required for error estimation in Monte Carlo methods. Both $\frac{N}{N-1} \text{Var}_N[X]$ and $\text{Var}_N[X]$ are suitable estimators for $\text{Var}[X]$ for our purposes.

Stochastic SIR Model

Summary

- ▶ Error estimate for Monte Carlo sampling:

$$\mathbb{E} \left[\left(\tilde{\mathbb{E}}_N[X] - \mathbb{E}[X] \right)^2 \right] \approx \frac{1}{N} \text{Var}_N[X]$$

where

$$\text{Var}_N[X] = \tilde{\mathbb{E}}_N[(X - \tilde{\mathbb{E}}_N[X])^2].$$