The Gillespie Algorithm: Simulating Stochasticity

Hwai-Ray Tung

September 2024

1 Gillespie Use Cases

The Gillespie algorithm is a generic useful algorithm that's great for simulating Markov Process where events happen at exponential rate. Examples of common applications include chemical reactions when there are few particles, epidemics in initial stages, and other times when there aren't very many particles in play. We'll run through a few examples for today's exercise.

2 Introduction Example: Branching process

We define a branching process as

- each particle births at rate b
- each particle dies at rate d

Suppose there are currently n particles. To understand the system, we should understand

1. when the next event (birth or death) happens
2. what the next event is (birth or death)

Once we have these two pieces of information, it is straightforward to ask the same two questions, but with population $n-1$ or $n+1$ and repeat until you don't want to simulate anymore.

2.1 When the next event happens

Let T be the time of the next event. We could (but computationally shouldn't) generate a separate exponential random variable (rv) $T_{k,b}$ or $T_{k,d}$ where k is the kth particle and b and d stand for birth or death. Then

$$T = \min(T_{1:n,b}, T_{1:n,d})$$

Nothing that by the cdf of the exponential rv,

$$\begin{aligned} P(T_{k,b} > t) &= e^{-bt} \\ P(T_{k,d} > t) &= e^{-dt} \end{aligned}$$

we can then write down the probability distribution of T as

$$\begin{aligned} P(T > t) &= \prod_{k=1}^{n} P(T_{k,b} > t) P(T_{k,d} > t) \\ &= \prod_{k=1}^{n} e^{-bt} e^{-dt} \\ &= \exp(-n(b+d)t) \end{aligned}$$

This means that T has the distribution of a exponential rv with rate $n(b+d)$, so we only need to generate one exponential rv to find the time of the next event! **In general, the next event occurs at time distributed from an exponential rv with rate equal to the sum of all the rates involved.**

2.2 What the next event is

Let T_b be the time of the next birth and T_d be the time of the next death. By the prior section, both are exponential rvs with rate nb and nd, respectively. Then the probability the next event is birth is

$$P(T_b < T_d) = \int_0^\infty f_{T_b}(t) P(T_d > t) dt = \int_0^\infty nbe^{-nbt} e^{-ndt} dt = \frac{nb}{nb + nd} = \frac{b}{b + d}$$

It is straightforward to show this does not change when the time of the next event is known.

We see in this case that the probability the next event is birth is the sum of all the rates that result in a cell being born (nb) divided by the sum of all the rates ($nb + nd$). **In general, the probability that the next event is some event A is the sum of all rates that generate event A divided by the sum of all the rates.**

2.3 Exercises

In your favorite language with a buddy, code up a branching process with birth rate 2 and death rate 1 and initial population 1. Use this to answer or do the following

1. Graph the population over time until time 10
2. Estimate the probability the branching process ever reaches 0.
3. If the process does not go extinct, graph the amount of time before the population reaches 100.
4. Try running with a large initial size.
   - How long does it take for the population to increase tenfold? Compare with the amount of computational time for a population of 10 to reach 100.
   - Consider an alternative method for simulating stochasticity when the population is large. Hint: Poisson might help.

3 Slightly more complicated: Epidemic Model

Suppose we are simulating an SIR model. For any pair of infected and susceptible people, the infected infects the uninfected with a rate β. The infected recovers with rate γ. There is a constant population of n individuals.

Quick check: What is the infection rate when there is 1 infected, 20 recovered, and 79 susceptible?

Let's do a coding example where $n = 1000$, 1 initial infected, and everyone is susceptible. $\beta = 0.002$ and $\gamma = 1$.

1. Graph a time series of the populations of S, I, and R.
2. Make a graph of the final size (ie total number of people that got sick in the epidemic). What is the probability the epidemic never really takes off?
3. What system of ODEs is associated with this simulation?

4 Epidemic on a network

Suppose we are given a network where each node represents a person. The edges represent contact, their weights represent contact rates (the βs in the previous problem), and an infected person recovers with rate γ. The previous example can be seen as a special case of this example, where the network is a complete graph and the weights are all identical and equal to β.

For this coding example, consider the following network structure and each time an event happens, print out the state of the network and how long it took to happen.