Module 11: Lesson 1 Lecture Notes

Callum Arnold

7/19/2021

Contents

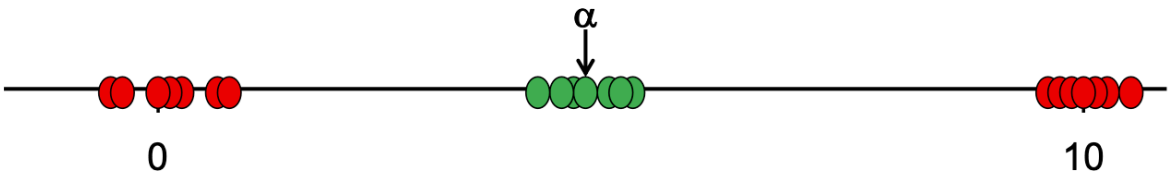
Simulation

Introducing simulation

- Suppose we have a stochastic epidemic model (e.g. SIR)
- Simulation = producing a realisation of the model (i.e. possible outcome)
  - Producing an outcome according to the correct distribution of all possible outcomes
- For the SIR model
  - producing a set of infection and removal times according to the correct distribution inherent in the model
- Why is it useful?
  - Helps us understand model behaviour
  - It is useful for testing our inference procedure (finding or estimating model parameters) e.g. validate our method against data from the simulation where we know the "truth"
- For example:
  - Perform N (=1000) simulations
  - Have model with parameter vector θ fixed at θ_T
  - For each simulation, estimate the model parameters to get $\theta_1, ..., \theta_N$
  - The average of the model estimates should be close to the true value of θ_T
- Also useful for model checking
  - Let's say we estimate model parameter α

  - For each value of α, we perform a large number of simulations and see whether the output of each is similar to the actual data

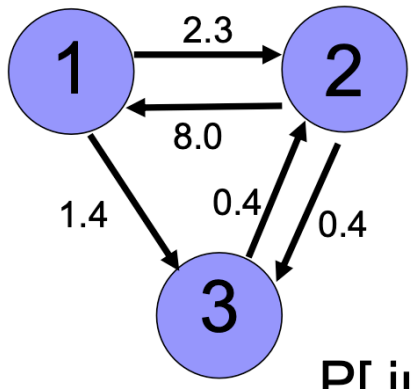
- Suppose we have a mixture model, e.g. looking at seroprevalence data where some people infected and others aren't, and x_k is the IgG value
  - $x_k \sim N(0,1)$ with probability 0.5, and $x_k \sim N(10,1)$, if we just fit a single parameter, we will be way off and we can check our predictions against the data to show that our model guess will never look like the data



- Simulations can be used for prediction
  - e.g. estimate epidemic model parameters up to time T and simulate forward using these estimates

Simulating Markov models

- Let the state space be denoted $S = \{1, 2, 3, ..., n\}$
  - S is the set of states the MC can visit and each state can be multidimensional
- We care about the tendency of the chain to move from $i \to j$
  - $\Pr(X(t+dt) = j | X(t) = i) = q_{ij}dt + o(dt)$
- The chain stays in state i for time T_i
  - $T_i \sim \text{Exp}(\sum_{j \neq i} q_{ij})$
  - $P(T_i > t) = \exp(-\sum_{j \neq i} q_{ij})$
- When it leaves state i, the chain jumps to state j with probability $q_{ij} / \sum_{j \neq i} q_{ij}$
- The time spent in state i and the choice of where to jump to **are independent**, and they are also independent of the same quantities in other states and at other times



Here $q_{12} = 2.3$, $q_{21} = 8.0$ etc

If chain enters state 1, time spent there is

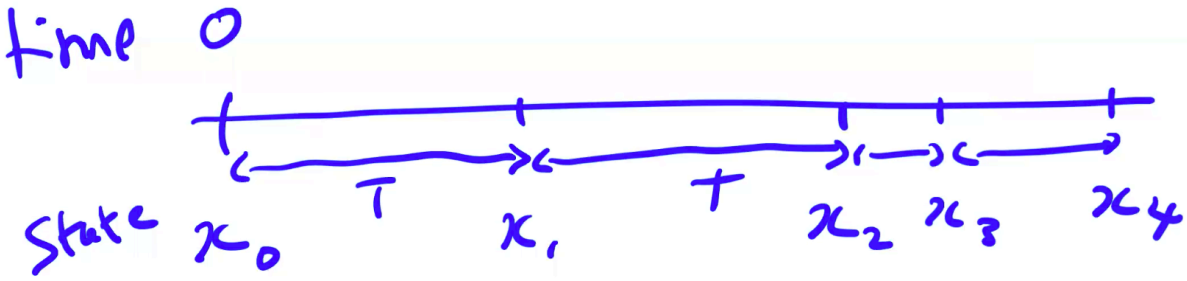
$T_1 \sim \text{Exp}(2.3+1.4) = \text{Exp}(3.7)$

$$P[\text{jump to state 2 next}] = \frac{2.3}{2.3+1.4}$$

Gillespie algorithm (next event simulation)

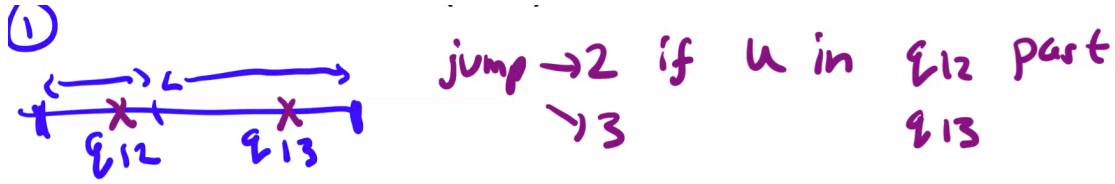
- All that is needed is to generate the time spent in a state, and the next state that is visited

Algorithm:



- Initialise (t=0, X(0) = x0)
- for state i:
  - Calculate L (sum of the jump rates out of state i)
    - $L = \sum_{j \neq i} q_{ij}$
  - Generate time spent in state i
    - $T_i \sim \text{Exp}(L)$
  - Sample from a uniform distribution ($u = \text{U}[0, 1]$)
  - If $u < \frac{\sum_{j \leq 1, j \neq i} q_{ij}}{L}$
    - $k = 1$
    - Divide by L to normalize so it's between 0 and 1
  - Else if $u < \frac{\sum_{j \leq 2, j \neq i} q_{ij}}{L}$
    - $k = 2$

- Continue until k determined



  - Update current time
  - Record t and k
    - Time of next event $= t + T_i$
    - State jumped to is k
- The algorithm outputs a sequence of times (t_k) and a corresponding sequence of states (x_k)

Example: general epidemic model (SIR)

- Due to the Poisson process infection mechanism and exponentially distributed infectious period
  - $\{[S(t), I(t)] : t \geq 0\}$
- If the chain is currently at (s, i), then it can jump to:
  - $(s-1, i+1)$ (infection) at rate $\beta si/N$
  - $(s, i-1)$ (removal) at rate γi
- Therefore, the time spent in (s, i)
  - $T_{(s,i)} \sim \text{Exp}([\beta si/N] + \gamma i)$
- Once the chain leaves (s, i)
  - Probability of infection $(s-1, i+1)$: $\frac{\beta s}{\beta s + N\gamma}$
  - Probability of recovery $(s, i-1)$: $\frac{N\gamma}{\beta s + N\gamma}$
  - Calculate probabilities by dividing rate of interest by sum of the rates
- Apply the algorithm iteratively
- Can sometimes be useful to keep track of the type of each event e.g. infection or recovery

Simulating non-Markov models

- Same idea as before
  - Generating time until next event
  - But independence properties of Markov chain are lost so we need to explicitly generate the times of future events as the algorithm evolves
- In the Markov SIR model:
  - Infections occur according to a Poisson process of rate $\beta S_t I_t/N$
  - Each infective remains so for a period of time $T_I \sim \text{Exp}(\gamma)$
- In the non-Markov SIR model
  - A common generalisation is to let the infectious period distribution T_I be non-exponential e.g. constant, Gamma

  - Infectious period T_I drawn from specified distribution with parameter vector θ
  - Now two model parameters β and θ
- To simulate the epidemic
  - Generate removal time of each individual as they become infected
    - The time of next removal is known, as is the identity of the removed individual
  - Generate possible time-to-next infection
    - $T \sim \text{Exp}(\beta si/N)$
    - If $T <$ time of next removal, next event is an infection
      - Otherwise, next removal occurs
      - If removal occurs first, then i in $T \sim \text{Exp}(\beta si/N)$ changes, so we no longer have the correct distribution for time to next infection and it needs to be updated

Example: non-Markov SIR model

- Lets assume we have a fixed infectious period
  - $T_I = c$
- We need $\vec{r}$ which contains the removal times of all the current infectives
- Initialize:
  - $S = N - 1$
  - $I = 1$
  - $t = 0$
  - $\vec{r} = (c)$
    - We only have one infective, so $\vec{r}$ only contains 1 removal time $(0 + c)$
- While $i > 0$:
  - $T \sim \text{Exp}(\beta SI/N)$
    - Draw from the infectious period distribution
  - $R = \min(r)$
    - Time of the next removal
  - if $t + T < R$ (current time + potential time to next infection $<$ time of next removal)
    - $S = S - 1, I = I + 1$
    - Add new removal time $r \leftarrow (t + T + c)$
    - Update current time $t = t + T$

Discussion

When infections and removals happen at the same time, do we consider to be Markov or non-Markov

Things are actually happening instantaneously, it's just that we make them discrete when we aggregate data.

In the non-Markov model example, the time to infection is exponential, so why is it not a Markov model

- The joint distribution of the number of susceptibles and infected at each time
  - Markov model is a markov chain
  - Not a markov chain in non-markov model as infectious period distribution is non-exponential
-