Stochastic Epidemic Models

Thursday, December 10, 2015
2:05 PM

Reading: Andersson and Britton Ch. 2

Homework 4 due Wednesday, December 16 at 5 PM; strict deadline. No late submissions. Turn it in to my office (Amos Eaton 310) or better yet my mailbox in Amos Eaton 301.

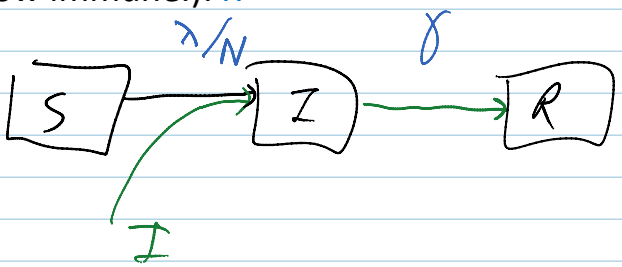
Optional lecture on Wednesday, December 16 at 1 PM in Lally 102 (this room) on Renewal Processes.

Will have office hour on Tuesday, December 15 -- time TBA (see website).

First, a standard deterministic model for disease spread.
(Hethcote, *SIAM Review*, 2000).

Assume a fixed population of N individuals, and divide them into three categories:

- Susceptible (healthy, never infected so far): S
- Infective (currently have and spread disease): I
- Recovered (had disease, but now immune.): R



λ is the per capita transmittivity rate of the disease (includes both exposure rate and likelihood of transmission upon exposure)

γ is the rate of recovery (the inverse of the mean time of the infectious period of the disease)

Law of mass action (assume all subpopulations are large enough) gives the standard Kermack-McKendrick model or SIR model:

$$\frac{dS}{dt} = -\lambda \frac{SI}{N}$$

$$\frac{dS}{dt} = -\frac{\lambda}{N} SI$$

$$\frac{dI}{dt} = \frac{\lambda}{N} SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

Some comments:

- the factor $\frac{\lambda}{N}$ is supposed to be proportional to the rate at which a given individual meets another given individual in the population. Keeping λ constant and dividing by N corresponds to assuming that as a population becomes larger, a given individual has a certain fixed number of contacts per day, so those contacts are diluted by the size the population. If dilution of contacts does not occur with population growth, then just replace the factor by λ.
- Note the SIR model has no real regard for individual welfare; it only is concerned with the process of infection. The model can be augmented to keep track of infectious vs. infected vs. exposed and to include births, deaths, immigration, etc.

These DE models, like any mass action model, only works well when all the populations involved are large. It's not so appropriate when a novel disease is introduced into a community, where one is more interested in characterizing risks of epidemic spread than just mean behavior.

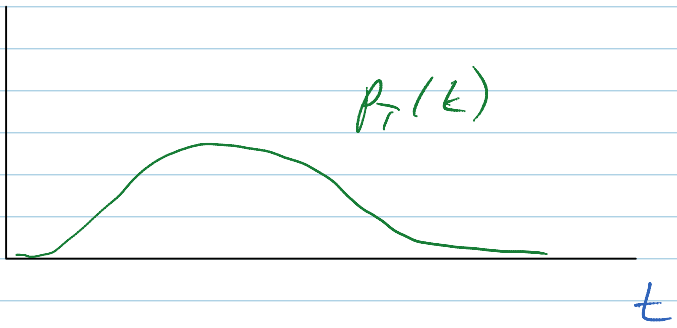
We'll illustrate how to construct a basic stochastic version of the Kermack-McKendrick model. The first attempt would be simply to make a CTMC version of the model. We can take the state space to be $\{0,1,2,\dots,N\} \times \{0,1,2\dots,N\}$, which we parameterize by (S, I), the number of susceptible, respectively infectious individuals. Note that we can recover $R = N - S - I$.

We would have two reaction channels:

- Infection: $(s, i) \to (s-1, i+1)$, rate $\frac{\lambda}{N} si$

- Recovery: $(s,i) \to (s,i-1)$, rate γi

One particular shortcoming of a CTMC model is that recovery process is poorly modeled as memoryless, and exponentially distributed. Rather, the recovery time PDF typically looks like:



So one important variation to the CTMC model is to break out of the CTMC framework, and instead explicitly incorporate some prescribed recovery time distribution $p_T(t)$ which can be fairly arbitrary.

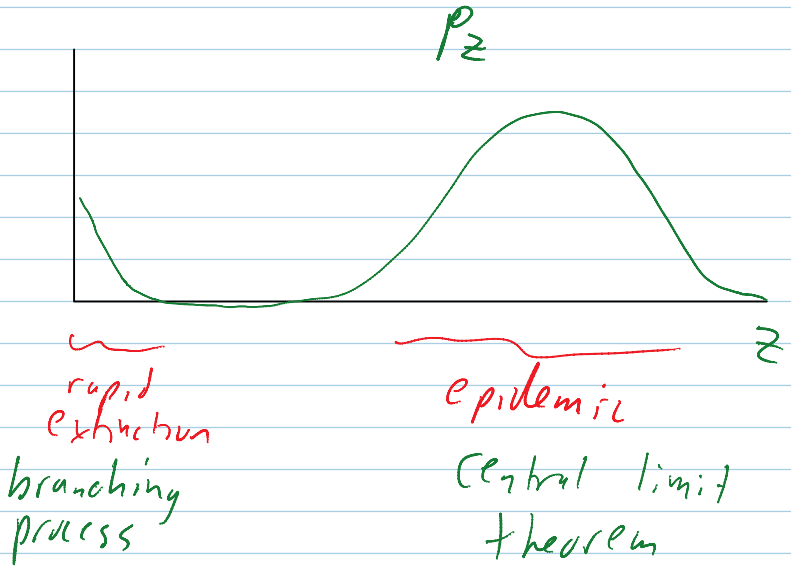
So in this stochastic epidemic model, the dynamics proceed as follows:

- infections still occur as in the CTMC model, which means when the system has s susceptible individuals and i infective individuals, then the probability for an infection to occur over the next small time interval of size Δt is $\frac{\lambda}{N} si\Delta t + o(\Delta t)$, independently of any other current or past information.
- But...you need to keep track of the times at which the currently infective subpopulation started their infectious period, and generate for each infective person an iid random variable with PDF $p_T(t)$ for how long that infective person will remain infective.

This model is not Markovian in time (at the level of (S,I) variables), but has the flavor of a hybrid Markov/renewal process.

Simulating this model by extending event-based approach of CTMCs is not difficult, but how do we analyze such a model since all CTMC tools are unavailable? We will develop a special purpose approach that can calculate the full probability distribution of the number of individuals who will ever be infected via a deterministic algorithm.

If we define $Z = \lim_{t\to\infty} R(t) = \lim_{t\to\infty} N - S(t) - I(t) = \lim_{t\to\infty} N - S(t)$ as the total epidemic size, then the goal is to compute the probability distribution for Z:



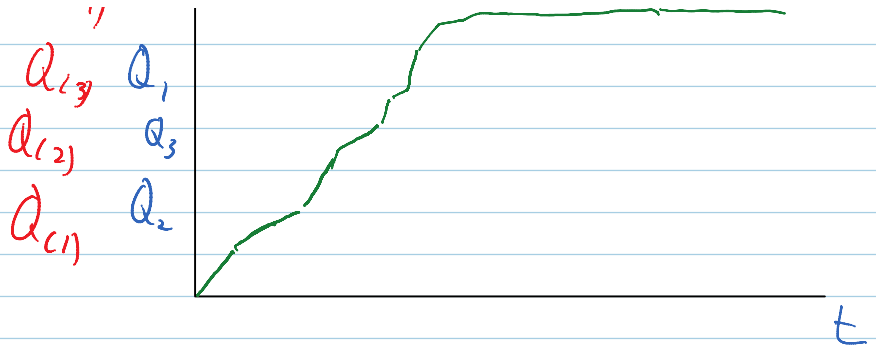
The key to analyzing our stochastic epidemic model is the Sellke construction:

Formulate the following model in terms of stochastic integer-valued process $I(t)$ representing the number of infective individuals at time t:

Define infection pressure: $\Phi(t) = \frac{\lambda}{N}\int_0^t I(t')dt'$ (essentially the total time contributed by all infectious individuals up to time t, multiplied by rate of transmission)

Define infection thresholds for each individual as iid random variables $\{Q_j\}$ which are exponentially distributed with mean 1.

Then individual j becomes infectious at the moment (if ever) when $\Phi(t) \geq Q_j$. When an individual becomes infectious, then they stay infectious for a random time which is drawn iid from the PDF $p_T(t)$, and their infection increases $I(t)$ by 1. Recoveries decrease $I(t)$ by 1.



Why is this model equivalent to the original formulation? Clearly the recovery process is the same. Just need to check that the infection process is the same.

We'll show that the Sellke construction will replicate the following property of the original model:

Given that we have $I(t)$ infected individuals at time t, and individual number j is not infected at time t, then the probability that he will be infected over the time interval $[t, t+\Delta t]$ will be $\frac{\lambda}{N}\Delta t + o(\Delta t)$, independent of anything else. A

$$P\left(A \mid j \text{ not infected at time } t, \mathcal{F}_t\right)$$

$$= P\left(\underline{\phi}(t+\Delta t) > Q_j \mid \underline{\phi}(t) < Q_j, \mathcal{F}_t\right)$$

$$= 1 - P\left(\underline{\phi}(t+\Delta t) < Q_j \mid \underline{\phi}(t) < Q_j, \mathcal{F}_t\right)$$

$$= 1 - P\left(Q_j > \underline{\phi}(t+\Delta t) \mid Q_j > \underline{\phi}(t), \mathcal{F}_t\right)$$

$$= 1 - e^{-(\underline{\phi}(t+\Delta t) - \underline{\phi}(t))}$$

because Q is exponential + therefore memoryless

$$= 1 - e^{-\phi'(t)\,\delta t + o(\delta t)}$$

$$= 1 - (1 - \phi'(t)\,\delta t + o(\delta t) + o((\delta t)^2))$$

$$= \phi'(t)\,\delta t + o(\delta t)$$

$$= \frac{\lambda}{N} I(t)\,\delta t + o(\delta t) \quad \checkmark$$

Now we formulate the analysis for the final total size of the disease spread in terms of the Sellke construction as follows:

Organize the population by writing it in terms of n initial susceptibles and m initial infectives, with $n + m = N$. Give positive indices to the initial susceptibles $\{1,2,\dots,n\}$, and negative indices to the initial infectives $\{-(m-1),\dots,0\}$.

Define $\{Q_{(j)}\}_{j=1}^{n}$ to be the order statistics of the $\{Q_j\}_{j=1}^{n}$, meaning reorder these values so that $Q_{(1)} \le Q_{(2)} \le \cdots Q_{(n)}$.

Assign to each susceptible $(1), (2), \dots$ the infectious period $T_1, T_2,$ etc. i.e., T_1 is the time that the individual with the lowest infectious threshold would remain infectious if they become infected. Same as the times at which the first newly infected individual remains infectious, etc.

$$Z = \min\left\{ i \in \{0, 1, \dots, n\} \;\middle|\; Q_{(i+1)} > \frac{\lambda}{N} \sum_{j=-(m-1)}^{i} T_j \right\}$$

total infection pressure

contributed by individuals $-(n-1)\dots,1$

To set up a deterministic method to calculate the probability distribution for Z, we will build a recursion/induction based on the initial number of susceptibles n.

To set up the induction process, define $Z^{n'}, \Phi^{n'}(t)$ to be the corresponding total infection size and the time-dependent infection pressure for a model with the same parameters but $n' \leq n$ initial susceptibles.

$$\Phi^{n'}(t) = \frac{\lambda}{\eta+m}\int_0^t I^{n'}(t')\,dt'$$

$$S^{n'}(t=0) = n'$$

$$I^{n'}(t=0) = m$$

$$p_k^{n'} = P(Z^{n'} = k)$$

is the PMF for $Z^{n'}$.

$$\tilde{p}_k^{n'} = P(\underbrace{I\,I\,I\dots I}_{k}\ \underbrace{S\,S\,S\dots S}_{n'-k})$$

$$p_k^{n} = \binom{n}{k}\tilde{p}_k^{n'}$$

Recursion is based on the following observation:

Choose $k \le \ell \le n'$ then:

$$\tilde{p}^{n'}_k = \tilde{p}^{\ell}_k \, \mathbb{E}\left[\exp\left(-\bar{Q}^{(\ell)}(\infty)(n'-\ell) \,\middle|\, Z^\ell = k \right) \right]$$

Proof: $\tilde{p}^{n'}_k = P(\underbrace{I\,I\,I\cdots I}_{1\ldots k}\,\underbrace{S\,S\,S\,S\,S\,S}_{k+1\ldots n'})$

$= P(\underbrace{I\,I\cdots I}_{1\ldots k}\,\underbrace{S\,S\,S}_{k+1\ldots \ell})$

$\times Pr(\underbrace{S\,S\,S\cdots S}_{\ell+1,\ldots,n'} \,|\, \underbrace{I\,I\cdots I}_{1\ldots k}\,\underbrace{S\,S\cdots S}_{k+1\ldots \ell})$

The main challenge in developing the recursion is computing the conditional expectation

$$\mathbb{E}\left[\exp\left(-(n-\ell)\bar{Q}^{\ell}(\infty) \,\middle|\, Z^\ell = k \right) \right]$$

$$= \mathbb{E}\left[\exp\left(-\frac{\lambda(n-\ell)}{n+m} \sum_{j=-(m-1)}^{Z^\ell} T_j \right) \,\middle|\, Z^\ell = k \right]$$

where $Z^k = \min\left\{ i \in \{k+1,\ldots,\ell\} : Q_{(i,1)} > \frac{\lambda}{n+m} \sum_{j=-(m-1)}^{i} T_j \right\}$

It looks like we are calculating the conditional expectation of a random sum (i.e., a sum over a random number of iid rvs), but the number of terms in the sum is not independent of the terms in the sum. So the standard generating function approach doesn't work.

But….Z^{ℓ} behaves as a Markov time w.r.t. the filtration generated by the infection times $\{T_1, T_2, \ldots, T_n\}$.

Next one can construct an exponential (or Wald) martingale associated to the desired random variable whose conditional expectation is being taken. And then one applies the optional stopping/sampling theorem to compute the conditional expectation of the martingale evaluated at the Markov time Z^{ℓ}; this particular implementation is what's known as Wald's identity (Karlin and Taylor Sec. 6.4).