

# Stochastic Epidemic Models

Monday, December 01, 2008  
12:09 PM

HW 4 posted, submit between Friday, December 5 and Wednesday, December 10 at 5 PM (final deadline). Will grade within about a day except anything submitted after Monday, December 8 at 5 PM will be graded on Wednesday evening.

Final Exam on Thursday, December 11 at 3 PM.

Thursday's class: guest lecturer Daniel Forger from University of Michigan will speak on "Stochastic simulation of molecular networks." That will last until about 1 PM and then after that we'll do course evaluations.

Danny Forger will also give a colloquium talk on how noise affects circadian clock regulation on Tuesday, December 2 at 4 PM in Ricketts 211.

## Deterministic models for epidemics (Hethcote, *SIAM Review*, 2000)

Divide the population into three subgroups

S is the number of susceptible individuals

I is the number of infective individuals

R is the number of recovered/immune (or dead!) individuals

$$N = S + I + R$$

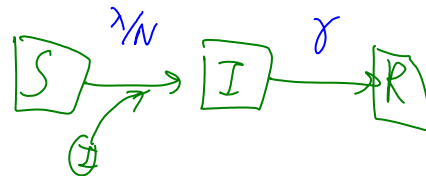
A simple way to model the dynamics of the spread of the disease through this population is by the

**Kermack-McKendrick model:**

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

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

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



$\lambda$  describes the per capita transmissivity rate (including contact rate and probability of infection upon contact)

$\gamma$  describes the rate of recovery (death).

This deterministic model has more sophisticated variations, and something like it is often used in practice for epidemiological modeling.

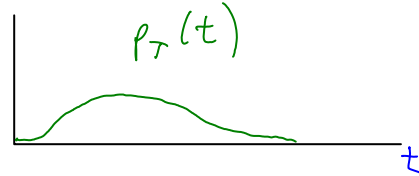
But one possible shortcoming of the model is that really it is premised on all the subpopulations being large so that in essence their interactions are modeled by just some mean field behavior. In practice, often one is concerned with how a small outbreak of the disease will progress (ebola, SARS). Here the disease may, under different scenarios, either die off without too much damage or create a large epidemic, even for situations where the governing parameters may be similar. That is to say, chance events early on may govern whether the disease spreads or is contained. For this reason, people are interested now in quantifying the risk factors in the spread of the disease, meaning not just what the mean number of infected people is, but what is the probability of a large outbreak?

To address these kinds of questions concerning risk, people turn to stochastic versions of these epidemic models.

Here is one way to make the Kermack-McKendrick model incorporate statistical effects.

The disease progresses as follows: Each infective person communicates his disease to each susceptible person according to a Poisson process with rate  $\lambda/N$ . All these infection events are independent of each other.

Each infective person remains infective for a random time  $T$  which is governed by an arbitrary probability density function  $p_T(t)$ , which need not be exponential. Moreover the duration of each disease is independent for each individual.



Special cases: If  $T$  is deterministic

$$p_T(t) = \delta(t - t_I)$$

then one can formulate a discrete-time Markov chain model (Reed-Frost model) with each epoch corresponding to the length of time of the disease  $t_I$ . (Andersson & Britton Sec. 1.2)

If  $p_T(t)$  is an exponential distribution:  $p_T(t) = \delta e^{-\delta t}$   $\delta = \frac{1}{\langle T \rangle}$

then one can write down a continuous-time Markov chain model for the epidemic with transition rates:

$$\begin{aligned} (s, i) &\rightarrow (s-1, i+1) & \text{rate } \frac{\lambda s i}{N} \\ (s, i) &\rightarrow (s, i-1) & \text{rate } \delta i \end{aligned}$$

Unfortunately exponential distribution is not very realistic for the duration of an illness; illnesses tend not to behave in quite a Markovian way. So we generally will need to work in the non-Markovian framework where the recovery time is not exponentially distributed. Then the process of the epidemic is not a Markov process, but it is built from a Markov process (for the infection) and a sequence of independent random variables (for the recovery) so progress may still be possible through appropriate tools.

Simulating this stochastic epidemic model is not that difficult (homework problem). What about analyzing it?

Suppose we start the population with  $m$  infective people and  $n$  susceptible individuals with  $N=n+m$ . Can we make some analytical statements about the probability distribution for the total number of people that ever get infected?

One key tool is the **Sellke construction**

Reformulate the model in the following terms.

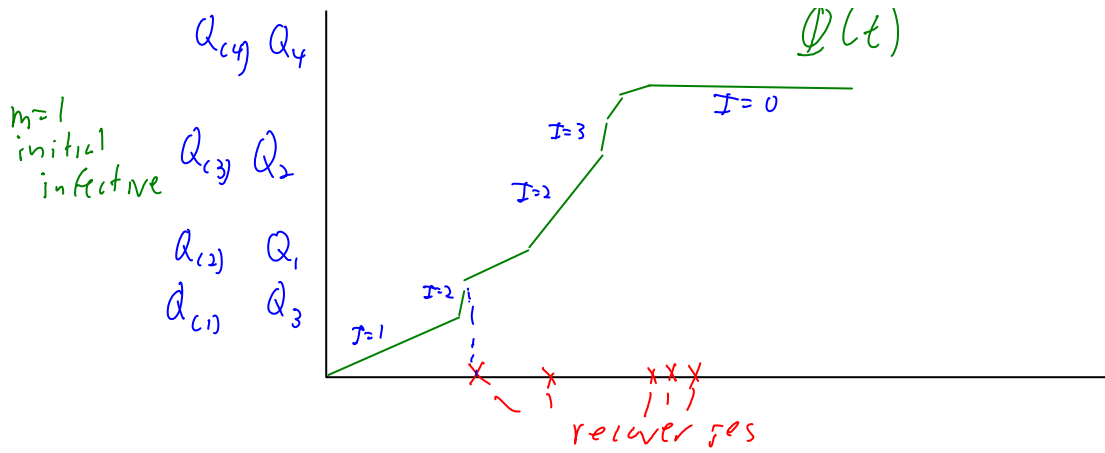
Define **infection pressure**  $\Phi(t) = \frac{\lambda}{N} \int_0^t I(s) ds$

Associate to each initially susceptible person (index them by  $j=1,2,\dots,n$ ) an **infection threshold**  $Q_j$  which are independently, identically distributed random variables with exponential distribution  $p_Q(q) = e^{-q}$

Susceptible person  $j$  becomes infected at the time  $t$  such that  $\Phi(t) = Q_j$ .

Recovery proceeds as before.





Why is the Sellke construction equivalent to the original stochastic epidemic model?

Recovery process is the same, so we just need to check that the infection process proceeds in statistically the same way. Specifically we should check that a given individual susceptible at time  $t$  has a probability  $\frac{\lambda}{N} I(t) \Delta t + o(\Delta t)$

to be infected during the time interval  $[t, t + \Delta t]$  independently of what's happened in the past.

So under the Sellke construction

$$\begin{aligned}
 & P(\text{individual } \#j \text{ infected during } [t, t + \Delta t] \mid \#j \text{ susceptible at } t) \\
 &= P(\Phi(t + \Delta t) \geq Q_j \mid \Phi(t) < Q_j) \\
 &= 1 - P(\Phi(t + \Delta t) < Q_j \mid \Phi(t) < Q_j) \\
 &= 1 - P(Q_j > \Phi(t + \Delta t) \mid Q_j > \Phi(t)) \\
 &= 1 - P(Q_j > \Phi(t + \Delta t) - \Phi(t)) \\
 &\quad \text{because } Q \text{ is exp. dist.} \\
 &\quad P(Q > b \mid Q > a) = P(Q > b - a) \text{ for } b > a \\
 &= 1 - e^{-(\Phi(t + \Delta t) - \Phi(t))} \\
 &\quad \text{Taylor expand} \\
 &= 1 - e^{-\Delta t \Phi'(t) + o(\Delta t)} \\
 &= 1 - (1 - (\Delta t \Phi'(t) + o(\Delta t)) + o((\Delta t)^2)) \\
 &= \Delta t \Phi'(t) + o(\Delta t) \\
 &= \Delta t \frac{\lambda I(t)}{N} + o(\Delta t) \quad \checkmark
 \end{aligned}$$

Extra knowledge about the past doesn't affect the calculation. This shows the statistical equivalence of the Sellke construction and the original stochastic epidemic model.

We now mathematically formulate the question of the total number of people that ever get infected by the disease.

$Z$  will denote the number of initially susceptible individuals that ever become infected. infection pressure from  $(-m-1), \dots, (i)$

$$Z = \min \left\{ i : Q_{(i+1)} > \frac{\lambda}{N} \sum_{j=-m-1}^i T_j \right\}$$

$Q_{(j)}$  = ordered values of the infection thresholds  $Q_{(1)} \leq Q_{(2)} \leq \dots \leq Q_{(n)}$

$T_j$  = infective period of  $j$ th individual, were they to be infected

Individuals  $(-m-1), \dots, 0$  are initial infectives.

To calculate the statistics of  $Z$  we will use a recursion argument based on a similar arrangement but with a smaller number  $n'$  of initial susceptibles.

$Z^{n'}$ ,  $I^{n'}(t)$  denote the total number of initial susceptibles that becomes infected and infection pressure for a population with  $n'$  initial infectives but

always an infection rate  $\frac{\lambda}{n+m}$

$$I^{n'}(t) = \frac{\lambda}{n+m} \int_0^t I^{n'}(s) ds$$

$$I^{n'}(0) = n'$$

$$I^{n'}(0) = m$$

$$P_k^{n'} = p(Z^{n'} = k)$$

$$\tilde{P}_k^{n'} = p(I_1 I_2 \dots I_k S S \dots S_{n'})$$

$$P_k^{n'} = \binom{n'}{k} \tilde{P}_k^{n'}$$

Recursion: Choose  $k \leq l \leq n'$

Then 
$$\tilde{P}_k^{n'} = \tilde{P}_k^{n-l} \mathbb{E} \left( e^{-Q^{(n-l)}(i-l)} \mid Z^l = k \right)$$

Then 
$$\tilde{p}_k^{n'} = \tilde{p}_k^l \mathbb{E} \left( e^{-Q^l(\infty)(n-l)} \mid Z^l = k \right)$$

Proof: 
$$\begin{aligned} \tilde{p}_k^{n'} &= P \left( \underset{1}{I} \underset{1}{I} \dots \underset{k}{I} \underset{k+1}{S} \dots \underset{n'}{S} \right) \\ &= P \left( \underset{1}{I} \underset{1}{I} \dots \underset{k}{I} \underset{k+1}{S} \dots \underset{l}{S} \right) P \left( \underset{k+1}{S} \dots \underset{n'}{S} \mid \underset{1}{I} \dots \underset{k}{I} \underset{k+1}{S} \dots \underset{l}{S} \right) \end{aligned}$$

To develop the recursion scheme, need to evaluate this conditional expectation.

$$Q^l(\infty) = \frac{\lambda}{n+m} \sum_{j=-m}^{Z^l} T_j$$

where we recall

$$Z^l = \min \left\{ i: Q_{(i+1)} \right\} = \frac{\lambda}{n+m} \sum_{j=-m}^i T_j$$

The total infection pressure looks like a random sum. The infection times  $T_j$  are all independent of each other, but they are not independent of the number of terms in the sum  $Z^l$ .

So we can't quite use the tools for summing up a random number of independent random variables when these are all independent of each other (like we did for branching processes).

However, our random sum does have the property that  $Z^l$

behaves like a Markov time with respect to the sequence of random variables  $\{T_1, T_2, \dots, T_n\}$

Combine this observation with the Wald identity (which more or less came from optional stopping theorem for martingales) allows the conditional expectation to be computed exactly.

Once you do that, then you can develop a deterministic recursion scheme...see the text.