Active Infections and Disease Extinction in the Stochastic SIR Model

Gugan Thoppe (IISc)

Collaborators: Gal Dalal (nVidia) and Balazs Szorenyi (Yahoo)
Event: Bengaluru Probability Seminar
Date: 12 Vaishakh 1943

Cheat Sheet

One Slide Summary

- SIR models are useful for studying the spread of infections.

- Our work concerns a discrete-time SIR model and focuses on:

  1. *Active Infections: Confirmed cases minus recoveries.*
  2. *Disease Extinction Time: Active infections become zero.*

- Our analysis via novel stopping times is fundamentally different.

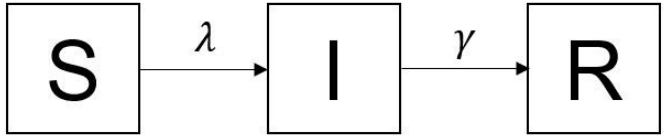
- Our results include:

  1. *Bounds for active infections and disease extinction time.*
  2. *Estimate for the expected value of the largest epidemic size.*

Background

SIR Models: Preliminaries

- The SIR model is an example of a compartment model.
- A fixed size population is divided into different compartments.



- Arrows indicate how the status of an individual may change.
- λ and γ are parameters that control the rate of evolution.
- Two categories of SIR models: deterministic and stochastic.

Deterministic SIR Models

- First proposed by Kermack and McKendriek in 1927.
- Good if community is homogeneous and people mix uniformly.
- Described by ordinary differential equations:

$$\dot{S}(t) = -\lambda \frac{SI}{n}, \qquad \dot{I}(t) = \lambda \frac{SI}{n} - \gamma I, \qquad \dot{R}(t) = \gamma I,$$

  where n is the fixed population size.

- Two major issues:
  1. No analytical estimates exist for active infections, i.e., $I(t)$.
  2. Early disease termination not possible as $I(t) \neq 0$ for finite t.

When are deterministic models insufficient?

- Inherent uncertainty in epidemic
  - Community is small, e.g., school.

- Epidemic fails to start
  - Large community, but outbreak started by few individuals.

- Noisy data
  - Disease outbreak data has standard measurement errors.

- Early disease extinction
  - Such questions are of interest.

Stochastic SIR Models

- Described via discrete/continuous time Markov chains or stochastic differential equations.

- Major issue with Markov chain based models:

  Complicated analysis via multiple approximations,
  e.g., branching process for early and final stages, ODE for middle.

Our Model

Continuous-time Stochastic SIR Model

- Introduced by Barlett in 1949.
- Fixed population size: n
- There is an independent $\exp(\lambda/(n-1))$ clock for each pair. Each time a clock ticks, the corresponding pair meet.
- Mean waiting before individual i meets another person is $1/\lambda$.
- An infected person recovers in $\exp(\gamma)$ time, independently.
- Basic reproduction number $\mathcal{R}_0 = \dfrac{\lambda}{\gamma} = \dfrac{\text{mean recovery time}}{\text{mean waiting time}}$.

Discrete-time Stochastic SIR Model

- At jump $t \geq 0$,

S_t : Susceptibles	I_t : Infected	R_t : Recovered

- $I_t = 0$ implies $I_{t+1} = 0$, $S_{t+1} = S_t$, and $R_{t+1} = R_t$.

- Suppose $I_t > 0$. Then, conditioned on the value of (S_t, I_t, R_t),

$$(S_t, I_t, R_t) \to \begin{cases} (S_t - 1, I_t + 1, R_t) & \text{w.p.} \quad \dfrac{\lambda S_t}{\lambda S_t + \gamma(n-1)} \\ (S_t, I_t - 1, R_t + 1) & \text{w.p.} \quad \dfrac{\gamma(n-1)}{\lambda S_t + \gamma(n-1)} \end{cases}$$

Our Approach

Key Idea

- We do not try to directly estimate the value of I_t for all t.
- Instead, propose new stopping times (T_i) and focus on I_{T_i}.

Stopping Times T_i and $T_{\max}$

- $T_{\max} = \min\{t \geq 0 : I_t = 0\}$.
- $T_0 = 0$.
- $T_i = \min\{t \geq 0 : S_t = S_0 - i\}$.
- $\tau_i = \min\{T_i, T_{\max}\}$.
- Since $T_{\max} < \infty$, we have $\tau_i < \infty$ a.s.

Usefulness of τ_i

- Suppose $I_{\tau_i} > 0$. Then, $\tau_i = T_i$ and $S_{\tau_i} = S_0 - i$.

- Between τ_i and τ_{i+1}, one and only one of the following could occur:

  1. there are a bunch of recoveries followed by an infection.
  2. all I_{τ_i} infected people recover.

- At each recovery, the value of S_t does not change from $S_0 - i$.

- Recoveries between τ_i and τ_{i+1} is a truncated geometric random variable with parameter $1/q_i$, where $q_i = 1 + \frac{1}{\mathcal{R}_0}\left(\frac{n-1}{n-I_0-i}\right)$.

- Truncation is needed since recoveries cannot exceed I_{τ_i}.

Our Results

Expected Number of Active Infections

Theorem

For any $t \in \{0, \ldots, S_0\}$,

$$\mathbb{E}[I_{\tau_t} - I_0] = \sum_{i=0}^{t-1} \left[1 - \frac{1}{\mathcal{R}_0}\left(\frac{n-1}{n-I_0-i}\right)\right] \mathbb{P}\left[T_{i+1} \leq T_{\mathsf{max}}\right].$$

Early Termination Bounds ($I_0 = 1$)

Theorem

- *Suppose* $0 \leq i \leq (n-3)/2$ *and* $\zeta = 1/(\mathcal{R}_0 + 1)$.
- $\mathbb{P}\{T_i \geq T_{\mathsf{max}}\} \geq \zeta$.
- *If* $\gamma/(\gamma+\lambda) \leq 0.0654$, *then*

$$\mathbb{P}\{T_i \geq T_{\mathsf{max}}\} \leq 2.9\zeta.$$

- *If* $\gamma/(\gamma+\lambda) \leq 0.101$, *then*

$$\limsup_{n\to\infty} \mathbb{P}\{T_i \geq T_{\mathsf{max}}\} \leq 1.38\zeta.$$

Remarks

- If mean recovery period is 14 days, then $\gamma = 1/14 \approx 0.071$.
- If the mean time for one to meet others is 0.5, then $\lambda = 2$.
- $\zeta \approx 0.034$.
- The bound does not depend on i.

Early Termination Bounds ($l_0 \geq 1$)

Theorem

Suppose $\mathcal{R}_0 \geq 4$ *and* $0 \leq i + l_0 \leq (n-1)/2$. *Then,*

$$\mathbb{P}\{T_i \geq T_{\max}\} \leq c_1 e^{-c_2 l_0},$$

where $c_1 = e^{1/30}/(1 - e^{-1/30})$ *and* $c_2 = 0.2/3$.

Remarks

- The bound $c_1 e^{-c_2 l_0} \leq 1$ for $l_0 \geq 52$.
- The bound does not depend on i.

Bounds for Active Infections ($I_0 \geq 1$)

Theorem

- *Suppose* $\mathcal{R}_0 \geq 4$ *and* $0 \leq i + I_0 \leq (n-1)/2$.

- *For* $q_j = 1 + \frac{1}{\mathcal{R}_0}\left(\frac{n-1}{n-I_0-j}\right)$, *let* $\mu_i = \sum_{j=0}^{i} q_j = O(i)$.

- *Let* $\epsilon \in (0,1)$ *be such that*

$$\mathcal{E}_i := \Big(2[i+1] - [1+\epsilon]\mu_i,\ 2[i+1] - [1-\epsilon]\mu_i\Big) \subseteq [2 - I_0, \infty).$$

- *Then, for* c_1, c_2 *as before and* $c_3 = \frac{2}{3}\left(1 + \frac{\gamma}{\lambda}\ln(2)\right)$,

$$\mathbb{P}\{(I_{\tau_{i+1}} - I_{\tau_0}) \notin \mathcal{E}_i\} \leq c_1 e^{-c_2 I_0} + e^{-c_3 i\epsilon\left(1-\frac{1}{\sqrt{1+\epsilon}}\right)} + e^{-c_3 i\epsilon\left(\frac{1}{\sqrt{1-\epsilon}}-1\right)}.$$

Remarks

- $\mathcal{E}_i \subseteq [2 - l_0, \infty)$ implies that $l_{\tau_{i+1}} > 0$.

- $l_{\tau_{i+1}} > 0$ implies $l_{\tau_j} > 0$ for all $j \in \{0, \ldots, j\}$.

Proof Ideas

Early Termination Bounds

- Use coupling to show that, for $k \geq 0$ and $\mathcal{E}_0, \ldots, \mathcal{E}_k \subset [2 - l_0, \infty)$,

$$\mathbb{P}\left[\bigcap_{i=0}^{k} \left\{ \left(l_{\tau_{i+1}} - l_{\tau_0}\right) \in \mathcal{E}_i \right\}\right] = \mathbb{P}\left[\bigcap_{i=0}^{k} \left\{ \left((i+1) - \sum_{j=0}^{i} H_j \right) \in C_i \right\}\right],$$

  where $H_j \sim Geom(1/q_j)$ are independent random variables.

- $\mathbb{P}\{T_{k+1} < T_{\max}\} = \mathbb{P}\left[\bigcap_{i=0}^{k} \{l_{\tau_{i+1}} \geq 2\}\right] \geq 1 - \sum_{i=0}^{k} \mathbb{P}\left\{ \sum_{j=0}^{i} H_j \geq i + l_0 \right\}$

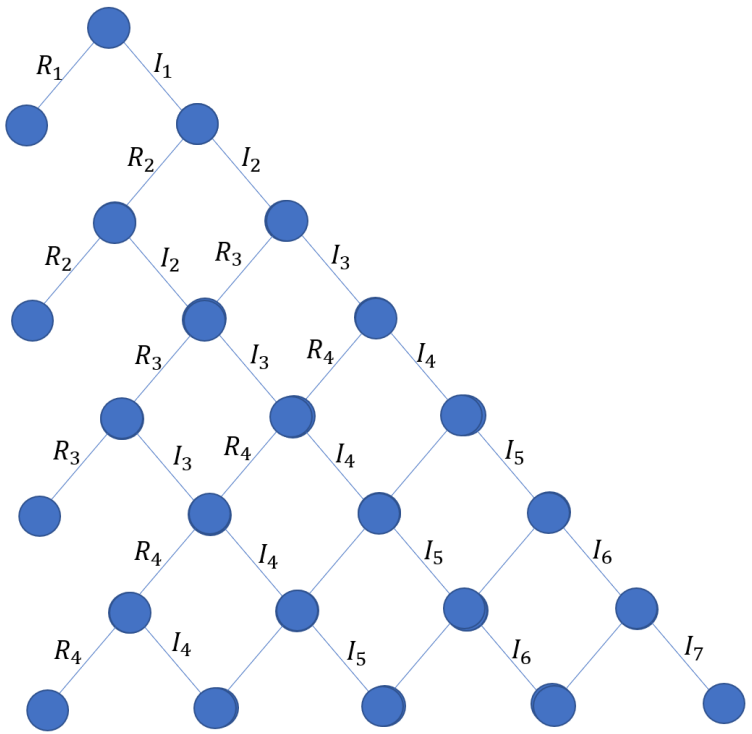
- Use bounds for sums of geometric variables from [Janson '14].

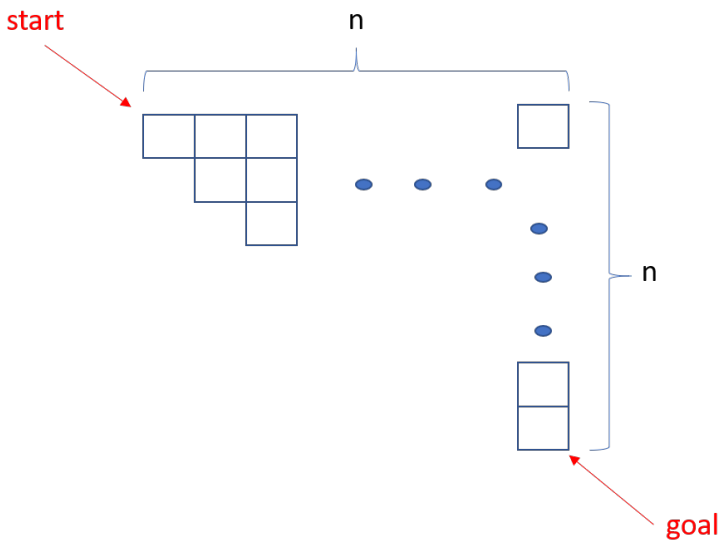
Bounds for Active Infections

For $\mathcal{E}_i \subseteq [2 - I_0, \infty)$,

$$\mathbb{P}\{(I_{\tau_{i+1}} - I_{\tau_0}) \in \mathcal{E}_i\} = \mathbb{P}\left[\left\{T_i < T_{\max}\right\} \cap \left\{\left(i + 1 - \sum_{j=0}^{i} H_j\right) \in \mathcal{E}_i\right\}\right].$$

Early Termination Bounds ($I_0 = 1$)





Early Termination Bounds ($l_0 = 1$)

- Use direct counting to obtain a bound for $\mathbb{P}\{T_{\max} = k\}$.
- $l_{t+1} - l_t = 2(S_t - S_{t+1}) - 1$.
- $l_t - l_0 = 2(S_0 - S_t) - t$.
- $\{T_{\max} < T_i\} = \{S_{T_{\max}} > S_0 - i\} = \{T_{\max} < 2i + l_0\}$.

Expected Number of Active Infections

- Let $X_t \equiv (S_t, I_t, R_t)$.

- Suppose $I_{\tau_i} > 0$. Then, for any $T_i \leq t < T_{i+1}$,

$$\mathbb{E}[I_{t+1} - I_t | X_t] = \left(1 - \frac{1}{\mathcal{R}_0}\left[\frac{n-1}{n - I_0 - i}\right]\right)\mathbb{E}[S_t - S_{t+1}|X_t].$$

- $\mathbb{E}[I_{\tau_{i+1}} - I_{\tau_i}] = \left(1 - \frac{1}{\mathcal{R}_0}\left[\frac{n-1}{n-I_0-i}\right]\right)\mathbb{E}[S_{\tau_i} - S_{\tau_{i+1}}]$.

- $S_{\tau_i} - S_{\tau_{i+1}} = \mathbb{I}[T_{i+1} \leq T_{\mathsf{max}}]$.

Future Directions

- Extend the first early termination result to cover the case $I_0 = k$.
- Obtain bounds for $\mathbb{E}[I_{\tau_i} - I_{\tau_0}]$ using estimates for $\mathbb{P}\{T_i \leq T_{\max}\}$.
- Obtain estimates for expected time for disease extinction.