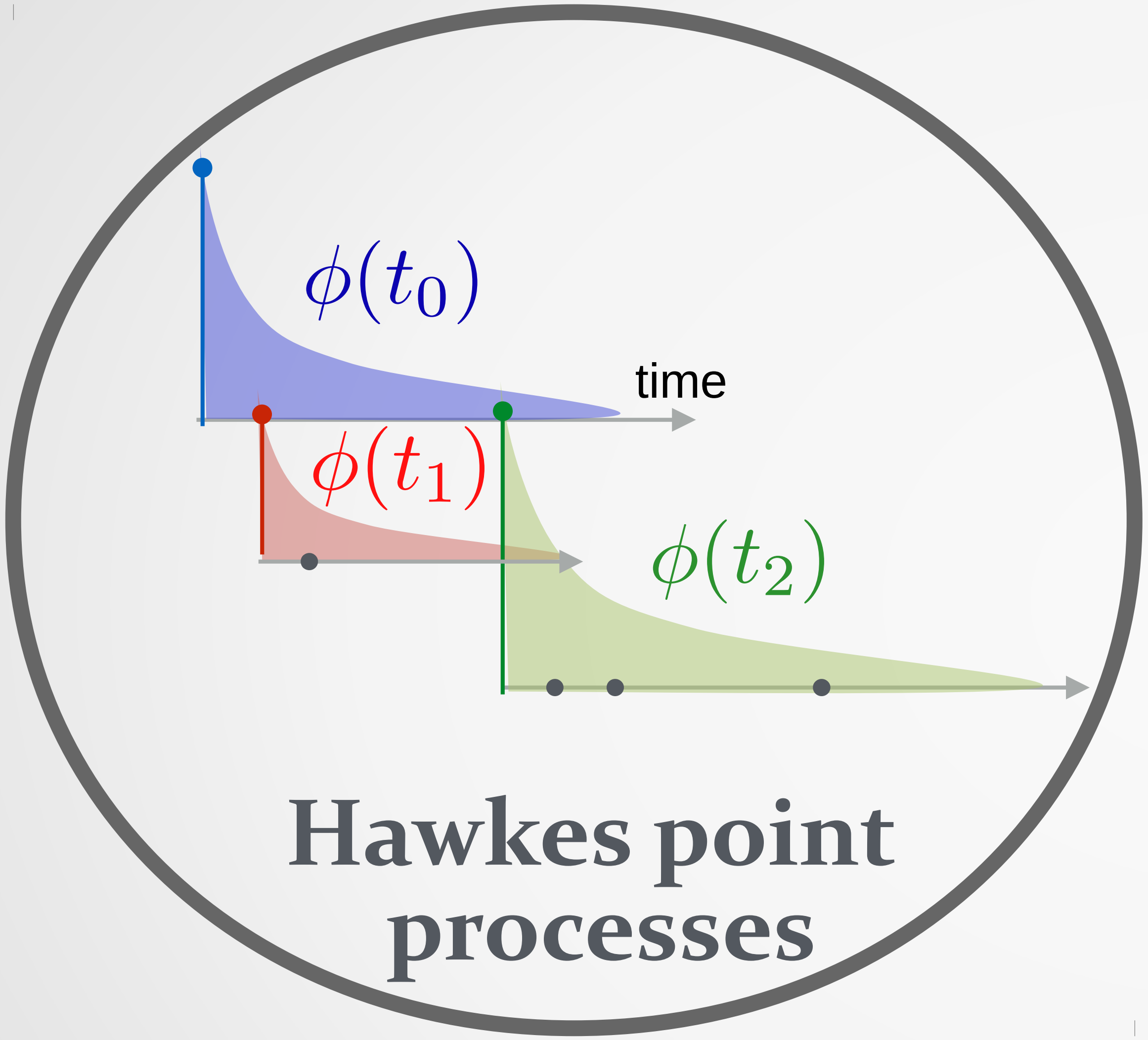
SIR-Hawkes:
Linking Epidemic Models and Hawkes Processes

Marian-Andrei Rizoiu **Swapnil Mishra** **Quyu Kong**

Mark Carman **Lexing Xie**

ComputationalMedia @ANU: http://cm.cecs.anu.edu.au

Divided we model



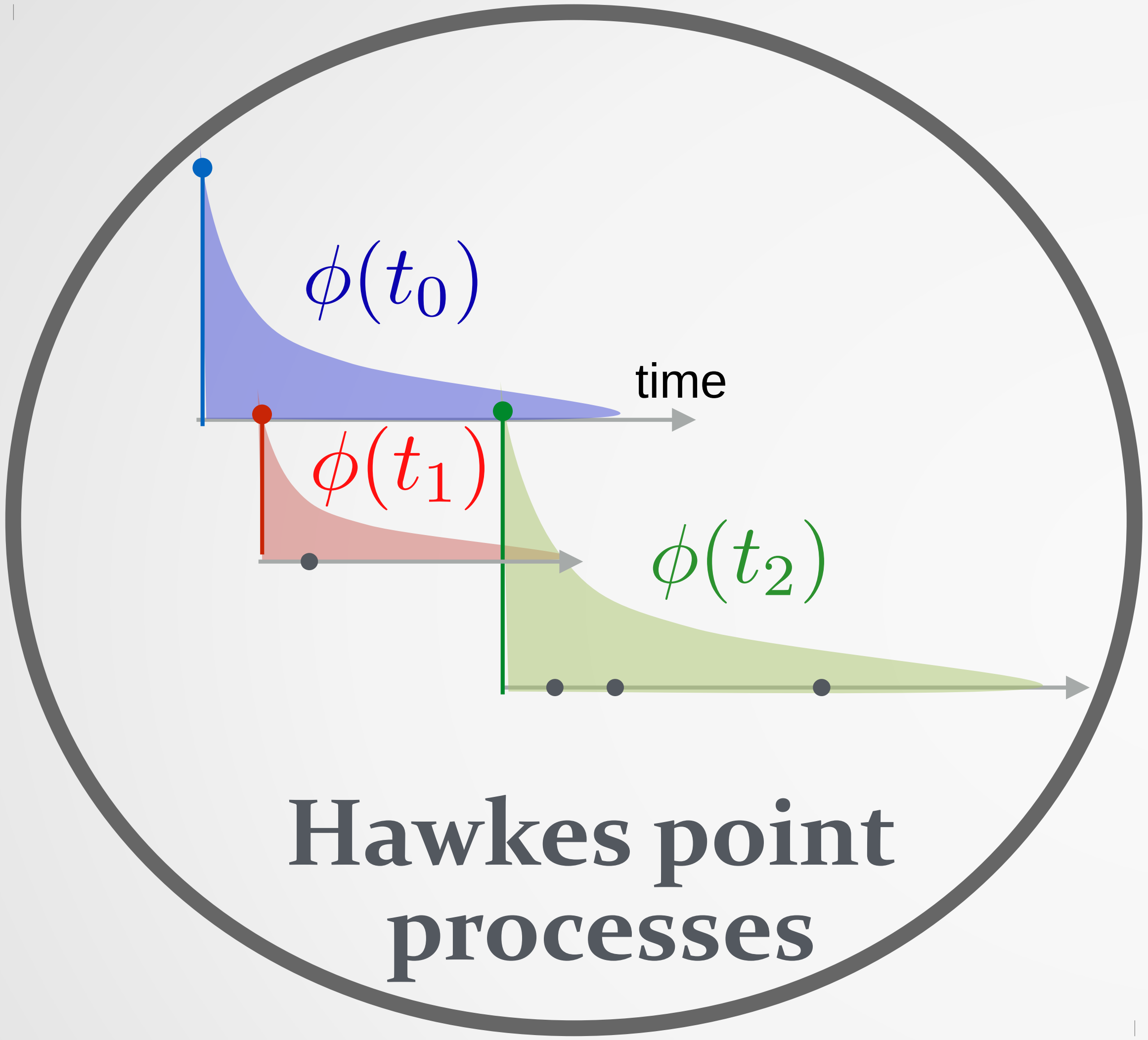
[Zhao et al KDD'15]

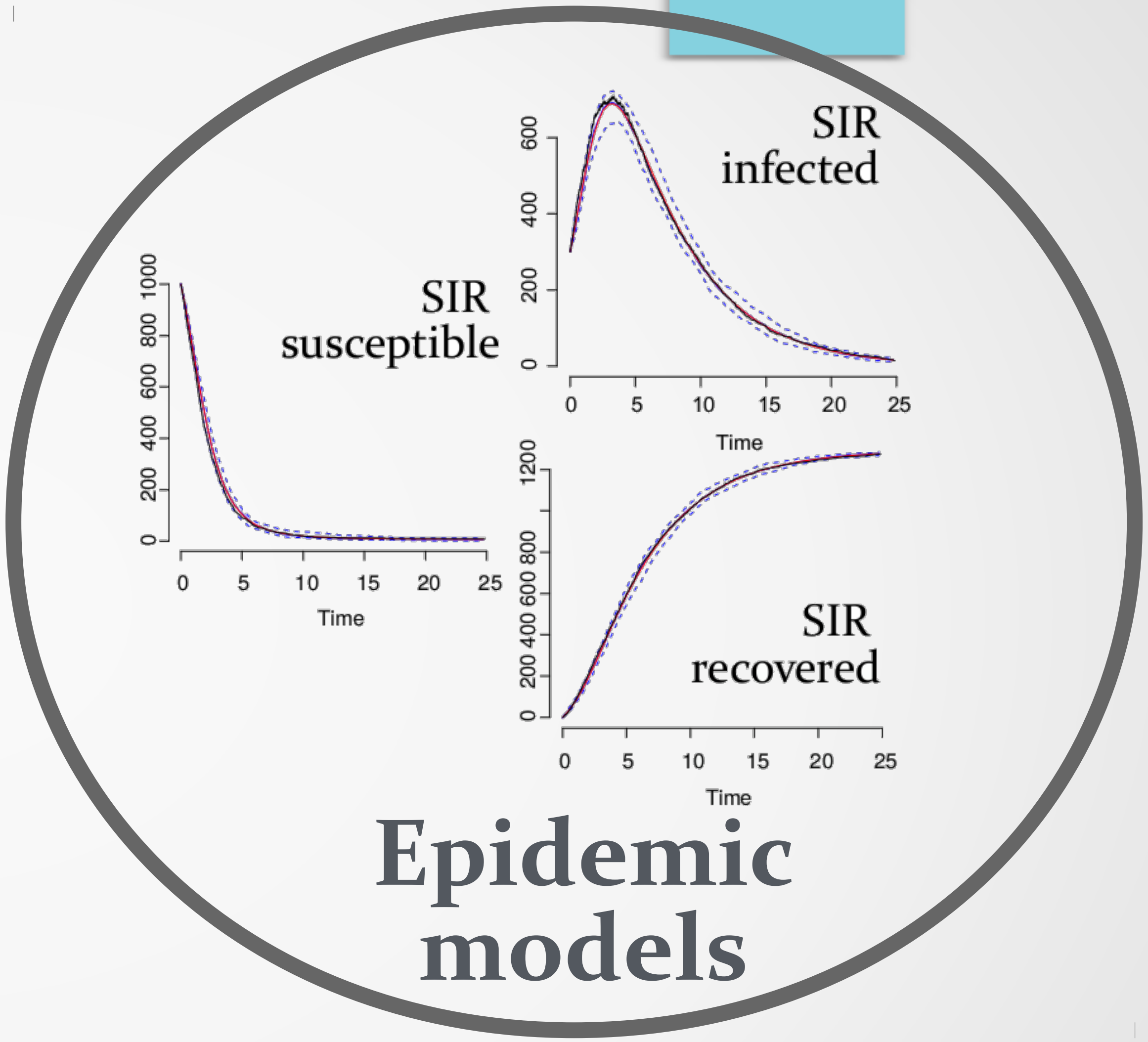
[Mishra et al CIKM'16]

[Farajtabar et al NIPS'15]

[Shen et al AAAI'14]

Divided we model

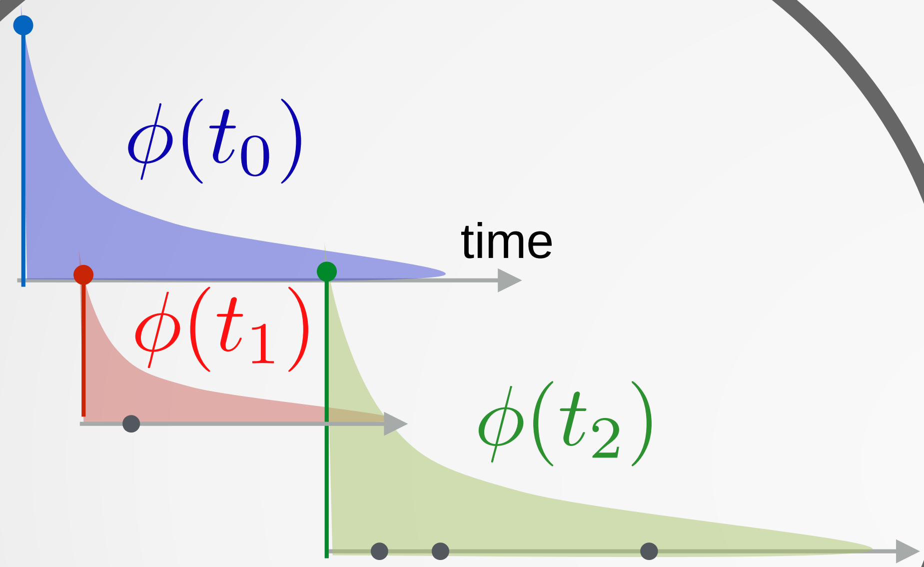




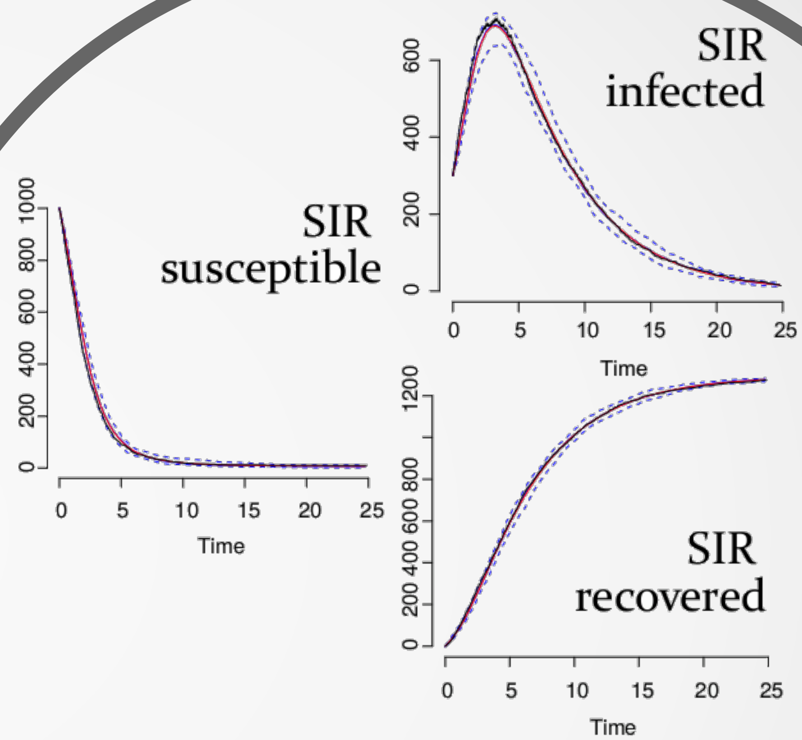
[Zhao et al KDD‘15]
[Mishra et al CIKM‘16]
[Farajtabar et al NIPS‘15]
[Shen et al AAAI‘14]

[Martin et al WWW‘16]
[Wu and Chen Springer+‘16]
[Bauckhage et al ICWSM‘15]
[Goel et al Manag.Sci.‘15]

Divided we model



Hawkes point processes

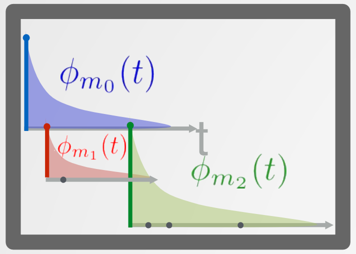


Epidemic models

[Zhao et al KDD‘15]
[Mishra et al CIKM‘16]
[Farajtabar et al NIPS‘15]
[Shen et al AAAI‘14]

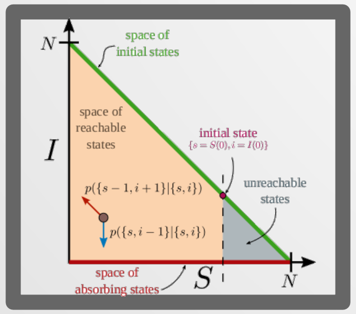
[Martin et al WWW‘16]
[Wu and Chen Springer+‘16]
[Bauckhage et al ICWSM‘15]
[Goel et al Manag.Sci.‘15]

Presentation outline



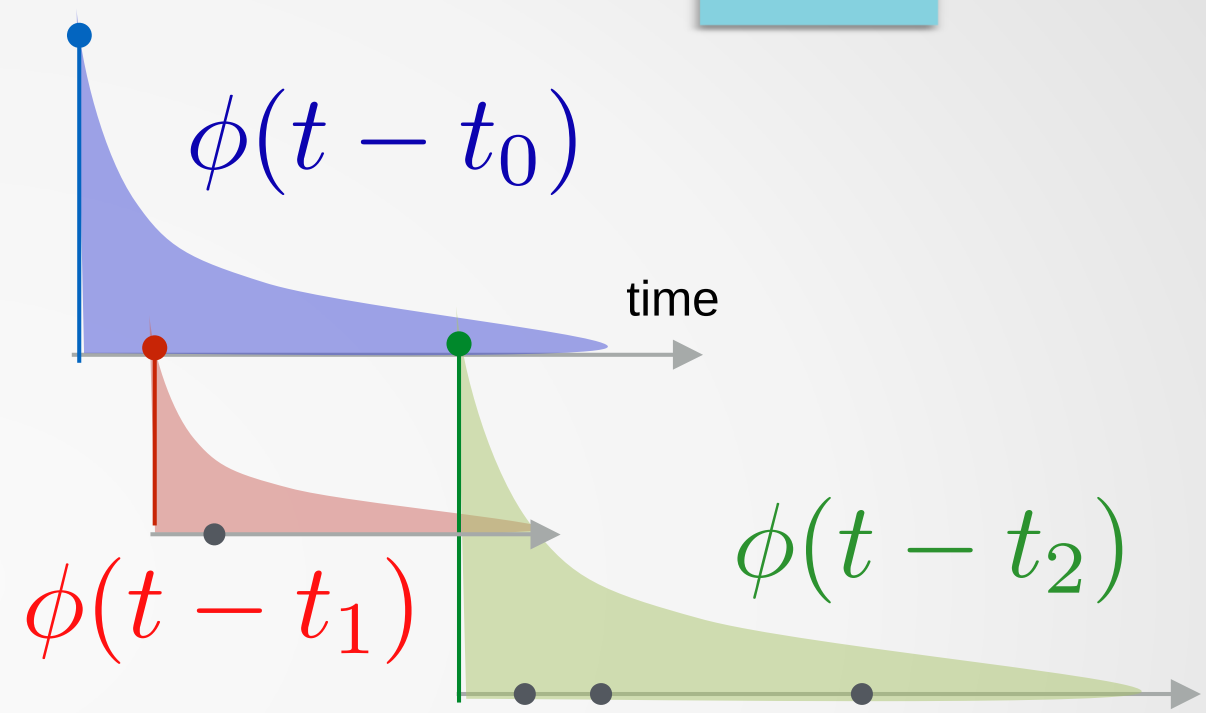
Prerequisites: Hawkes point processes and SIR infectious models

Linking SIR and the Hawkes processes



Computing the distribution of diffusion size

The Hawkes Process [Hawkes '71]



background event rate

$$\lambda(t) = \mu + \sum_{t_j < t} \phi(t - t_j)$$

event intensity

self-excitation

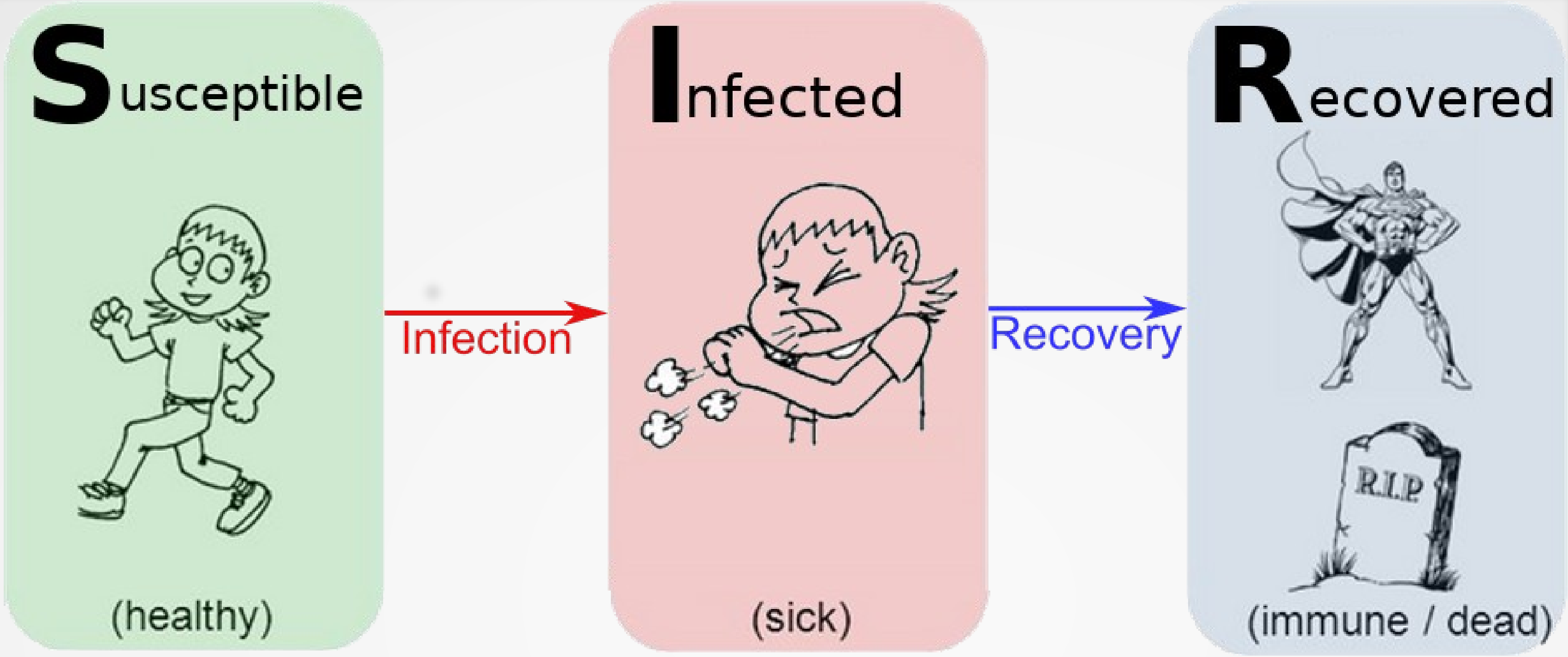
the rate of 'daughter' events

content virality

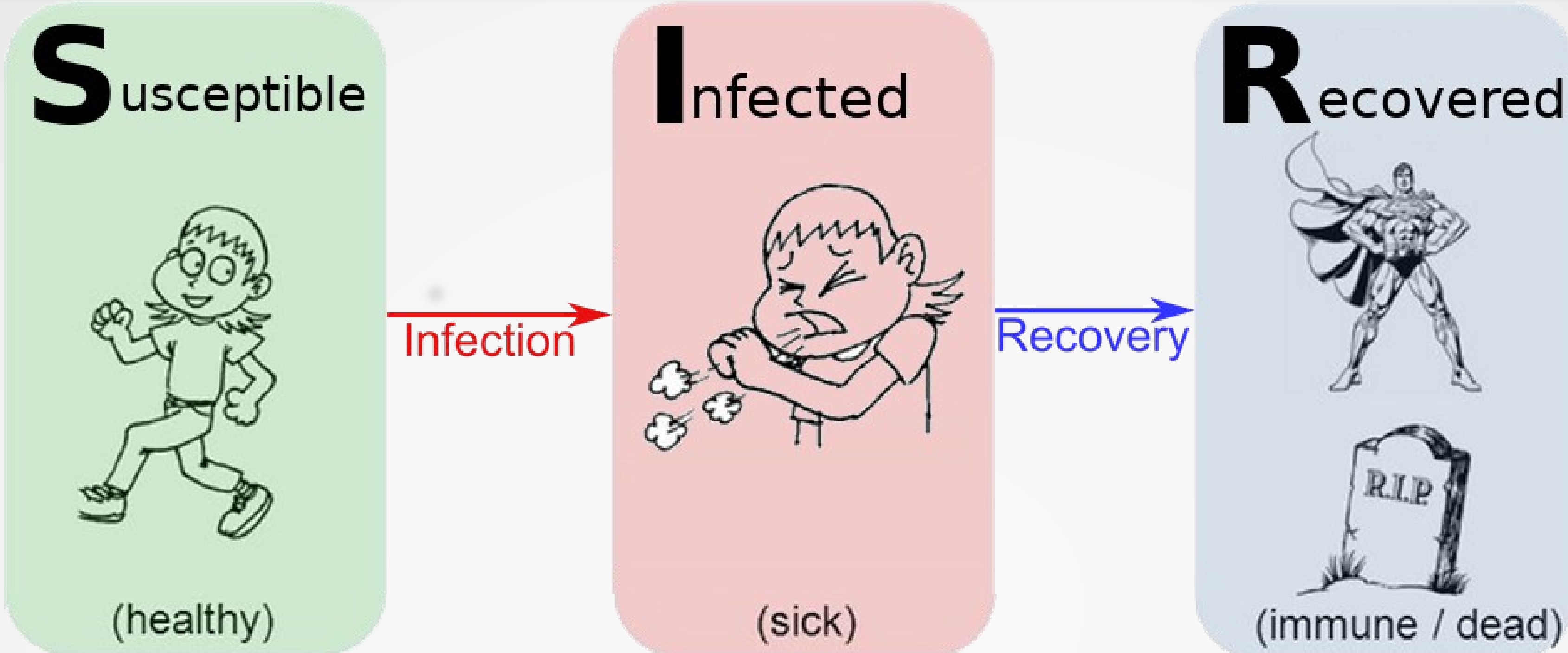
memory decay

$$\phi(\tau) = \kappa \, \theta e^{-\theta\tau}$$

The SIR epidemic model



The SIR epidemic model



$$\frac{dS(t)}{dt} = -\beta \frac{S(t)}{N} I(t)$$

$$\frac{dI(t)}{dt} = \beta \frac{S(t)}{N} I(t) - \gamma I(t)$$

$$\frac{dR(t)}{dt} = \gamma I(t)$$

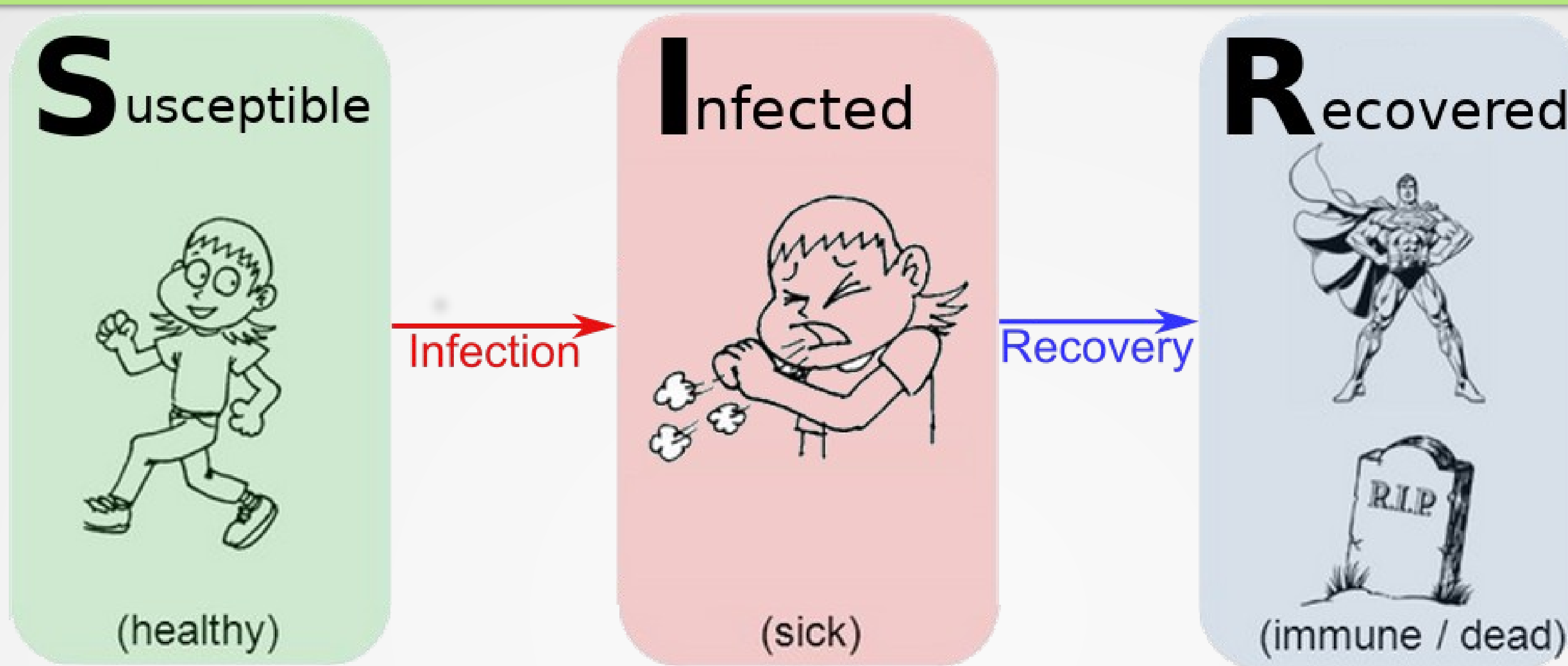
infection rate

Population size (known and fixed)

recovery rate

Deterministic SIR

The SIR epidemic model



$$\frac{dS(t)}{dt} = -\beta \frac{S(t)}{N} I(t)$$

$$\frac{dI(t)}{dt} = \beta \frac{S(t)}{N} I(t) - \gamma I(t)$$

$$\frac{dR(t)}{dt} = \gamma I(t)$$

infection rate

Population size (known and fixed)

recovery rate

$$\lambda^I(t) = \beta \frac{S_t}{N} I_t$$

$$\lambda^R(t) = \gamma I_t$$

Deterministic SIR

Stochastic SIR

SIR as a bivariate point process

Infection process C_t

Recovery process R_t

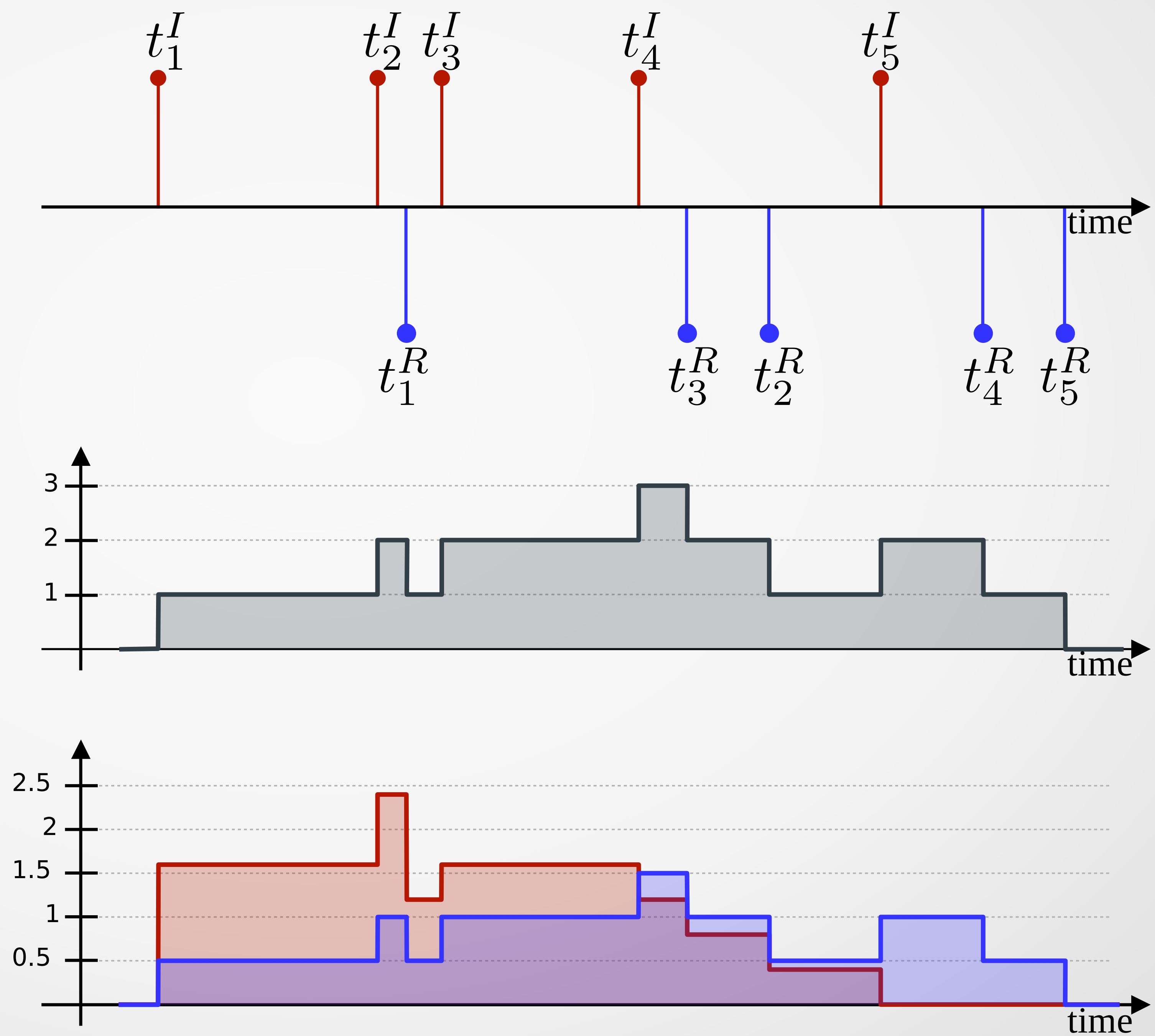
Number of infected I_t

New infection rate

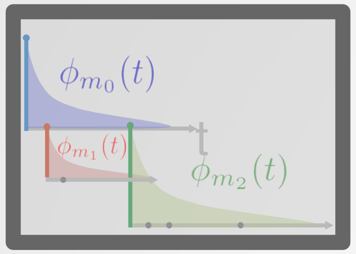
$$\lambda^I(t) = \beta \frac{S_t}{N} I_t$$

New recovery rate

$$\lambda^R(t) = \gamma I_t$$

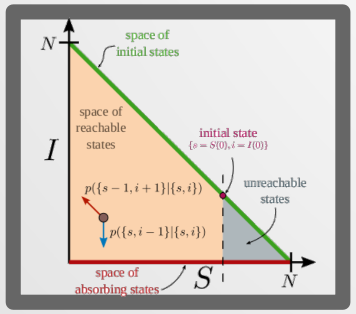


Presentation outline



Prerequisites: Hawkes point processes and SIR infectious models

Linking SIR and the Hawkes processes



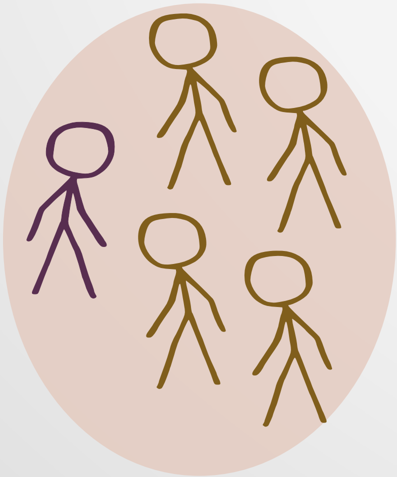
Computing the distribution of diffusion size

A finite population Hawkes model

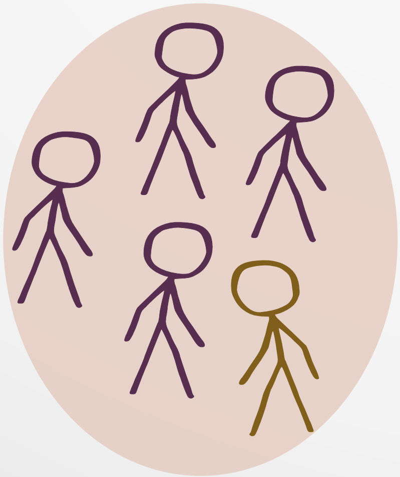
Goal: Introduce population size in Hawkes

HawkesN: modulate the event intensity by the size of the available population:

$$\lambda^H(t) = \left(1 - \frac{N_t}{N}\right) \underbrace{\left[\mu + \sum_{t_j < t} \phi(t - t_j)\right]}_{\textbf{Hawkes intensity}}$$

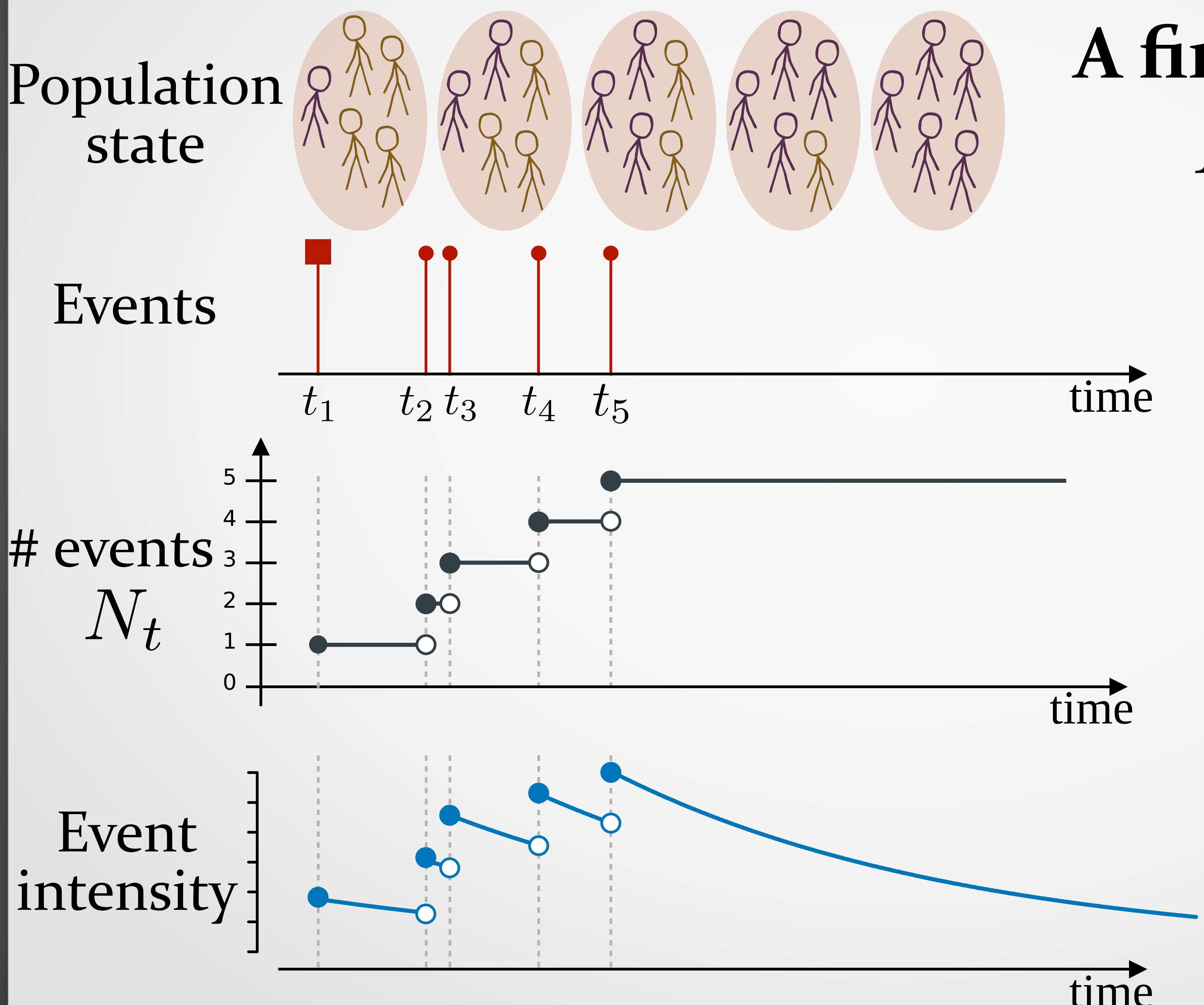


80%
susceptible



20%
susceptible

Example: a HawkesN diffusion



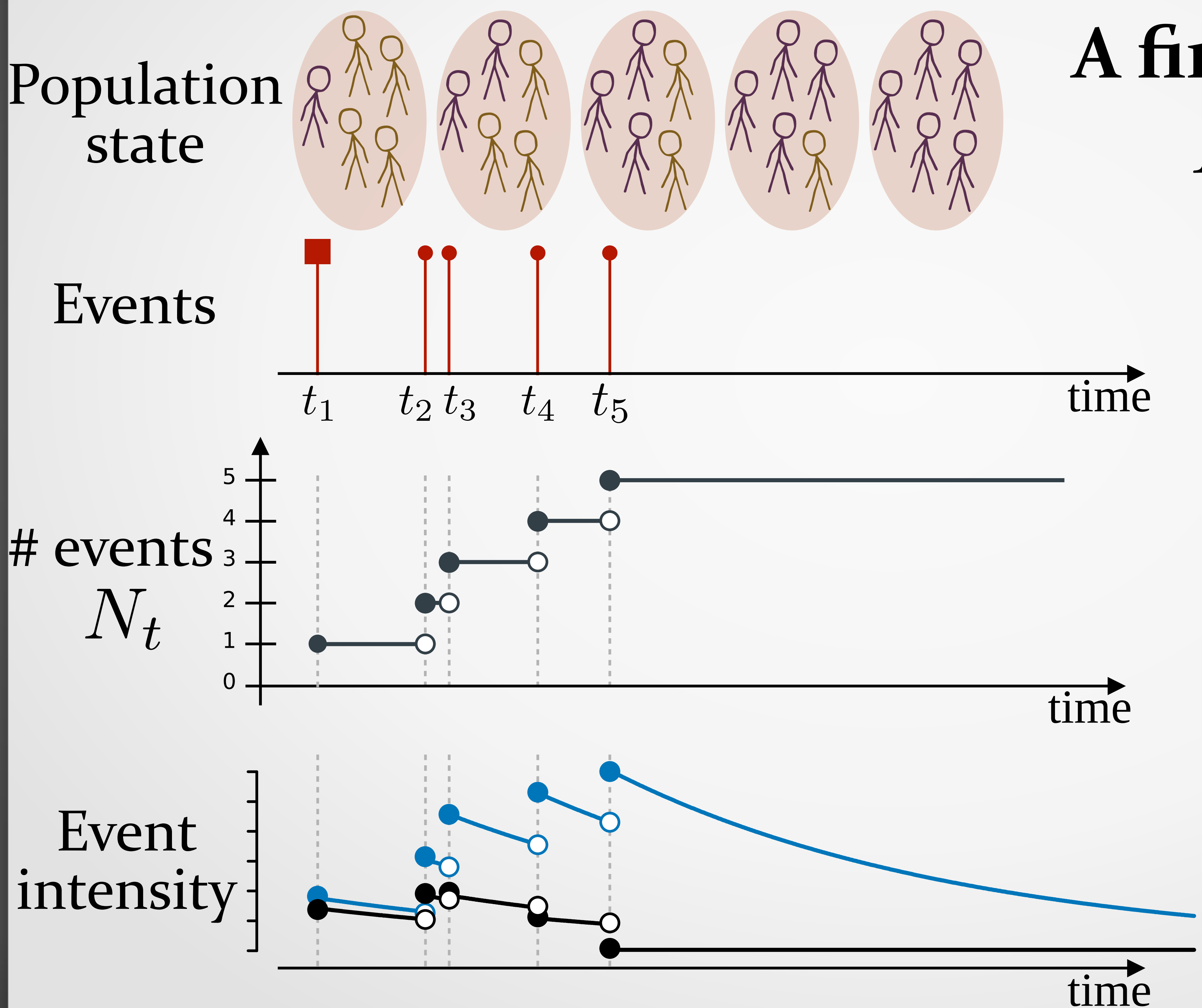
A finite population

$$N = 5$$

Hawkes

$$\lambda(t) = \mu + \sum_{t_j < t} \phi(t - t_j)$$

Example: a HawkesN diffusion



A finite population

$$N = 5$$

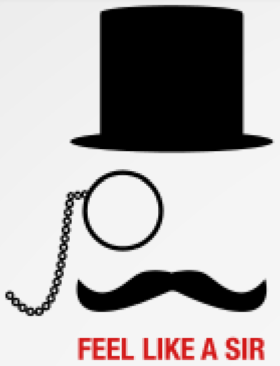
Hawkes

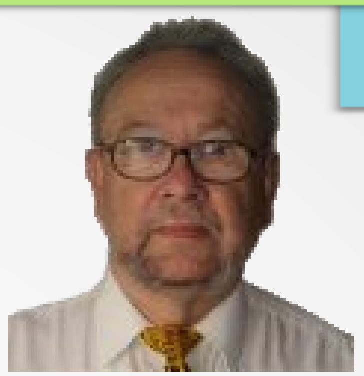
$$\lambda(t) = \mu + \sum_{t_j < t} \phi(t - t_j)$$

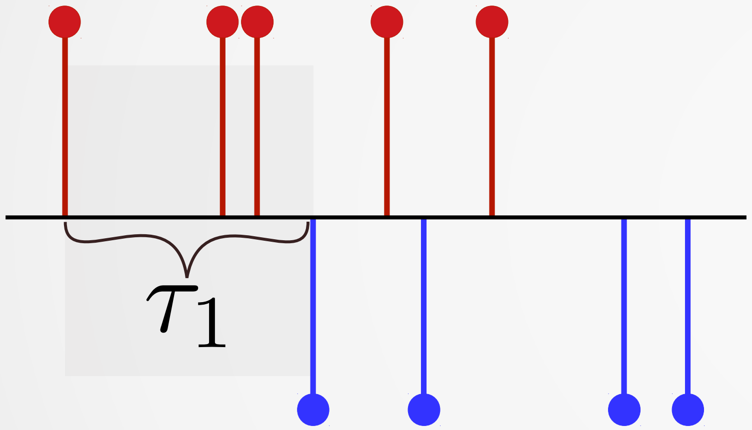
HawkesN

$$\lambda^H(t) = \left(1 - \frac{N_t}{N}\right) \lambda(t)$$

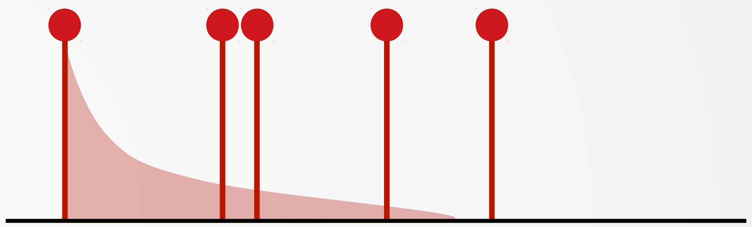
Linking SIR and Hawkes





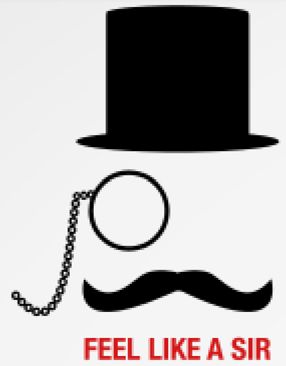


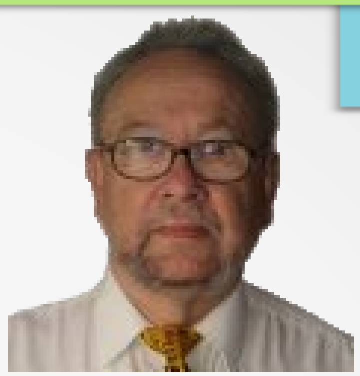
$SIR(\beta, \gamma)$

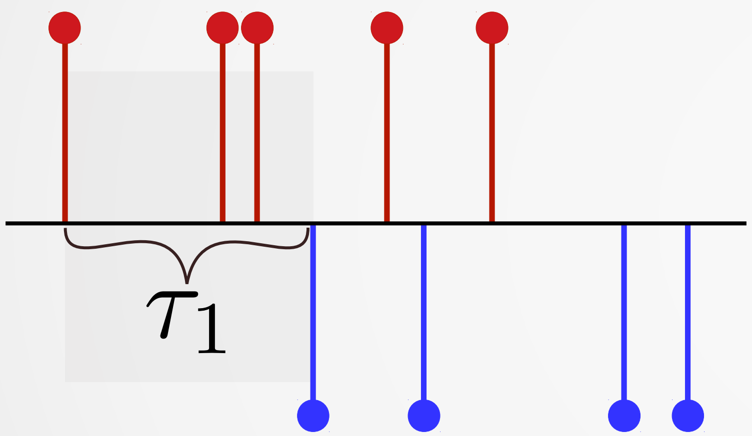


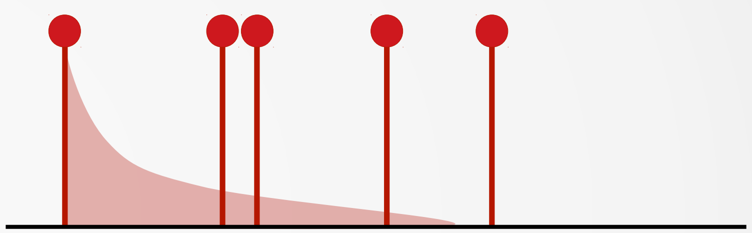
$HawkesN(\mu, \kappa, \theta)$

Linking SIR and Hawkes









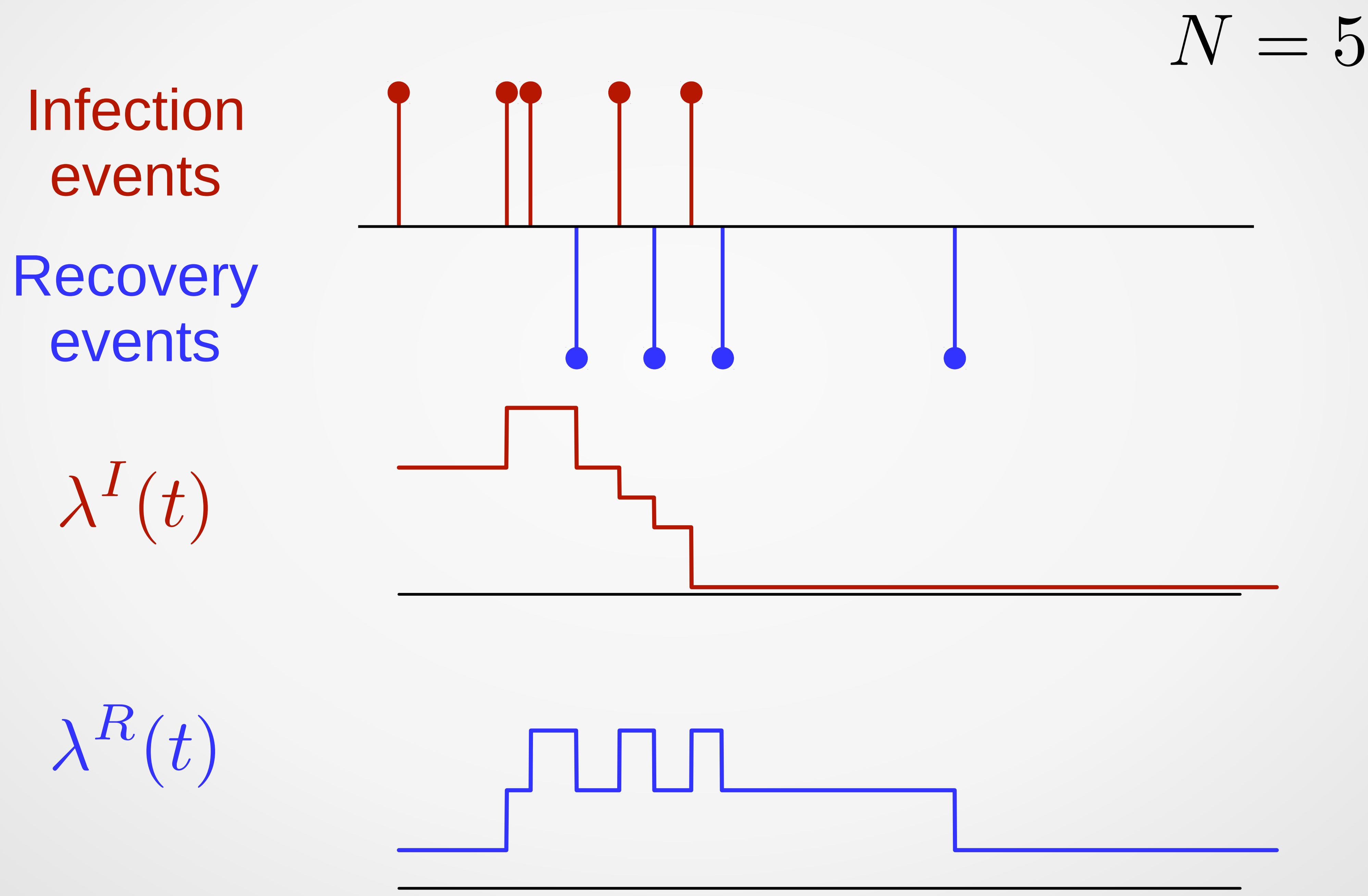
$$SIR(\beta, \gamma)$$

$$HawkesN(\mu, \kappa, \theta)$$

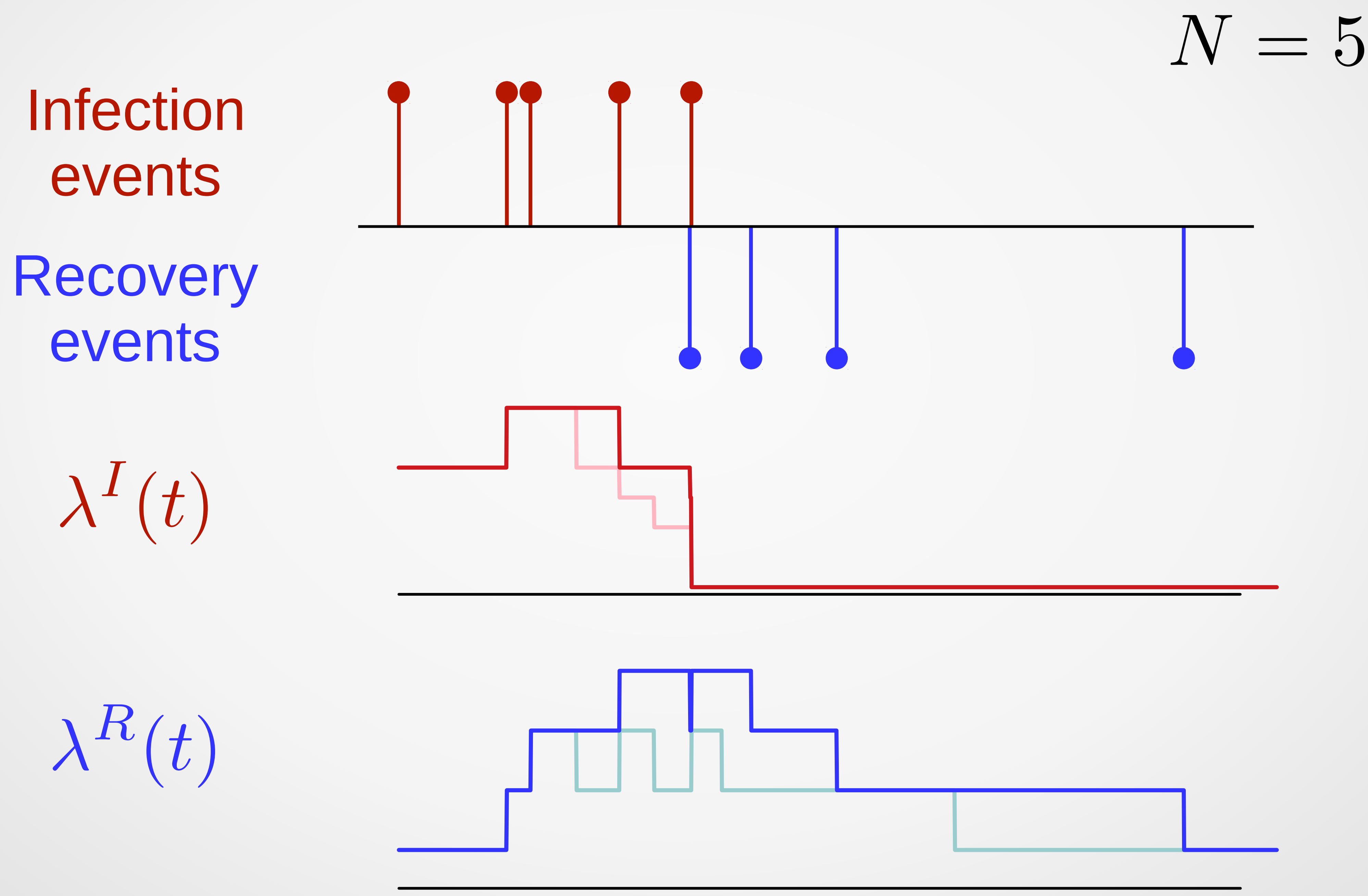
$$\mathbb{E}_{t^R}[\lambda^I(t)] = \lambda^H(t)$$

where $\mu = 0,\ \beta = \kappa\theta,\ \gamma = \theta$

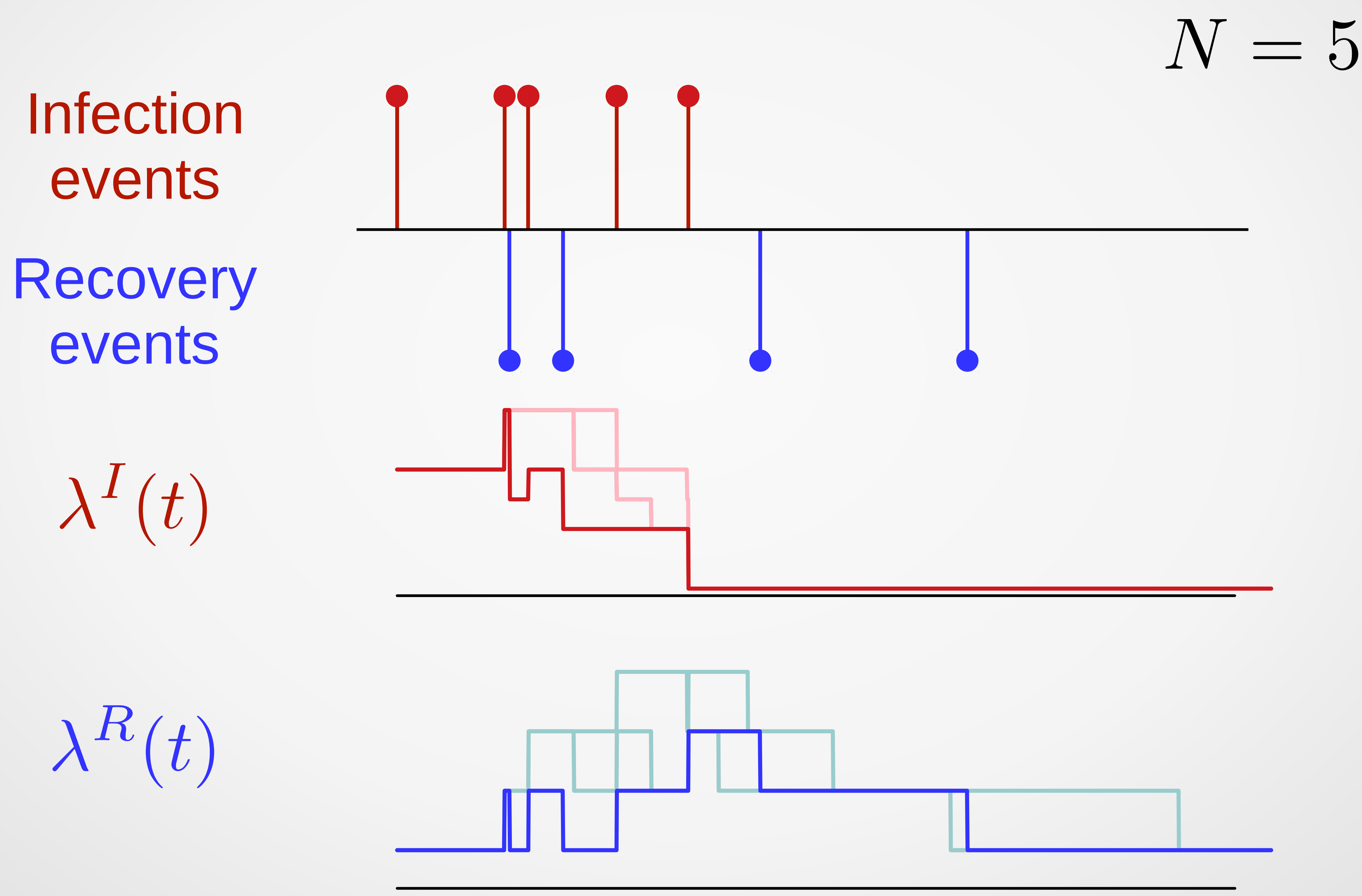
From SIR to HawkesN



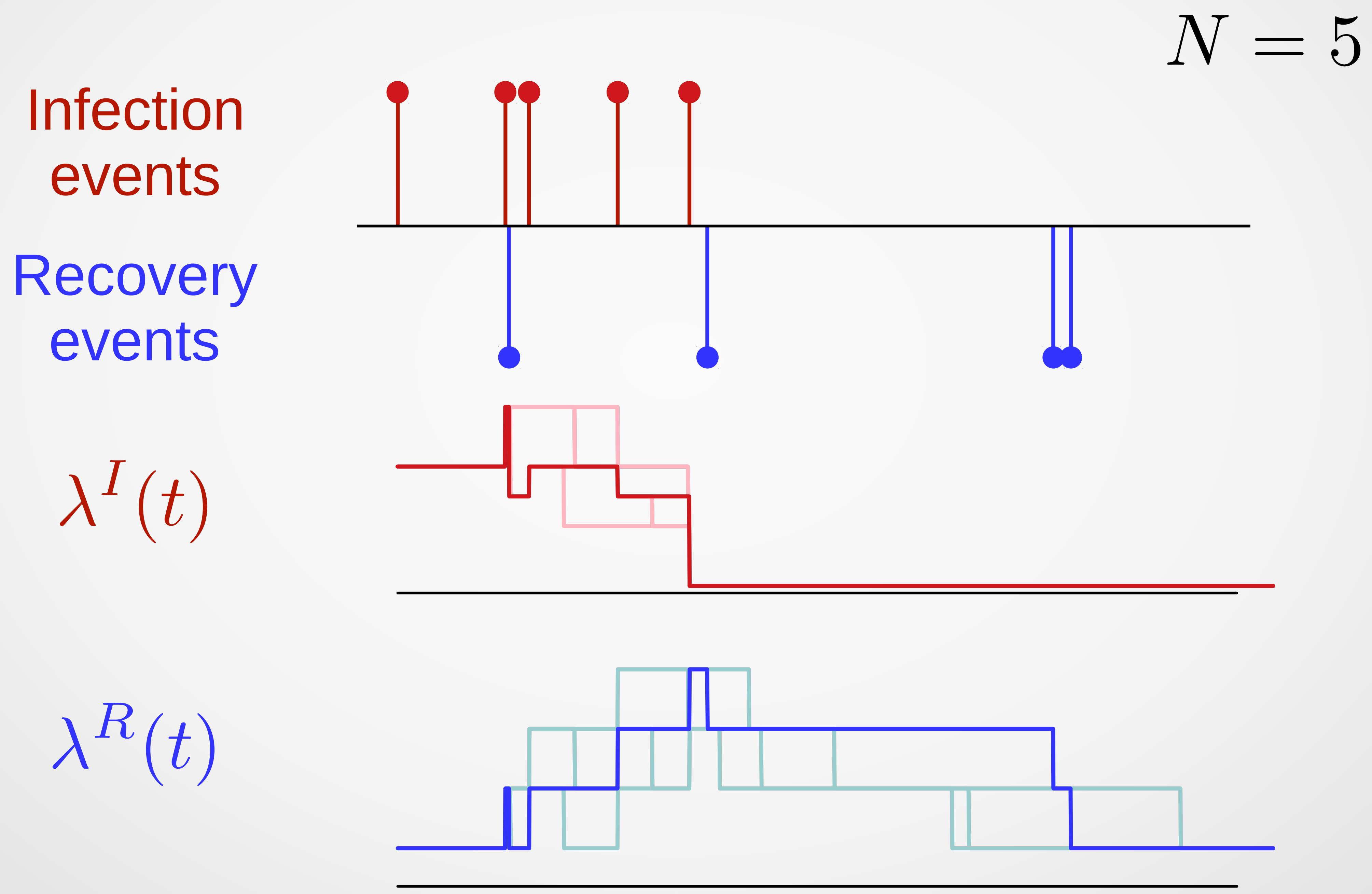
From SIR to HawkesN



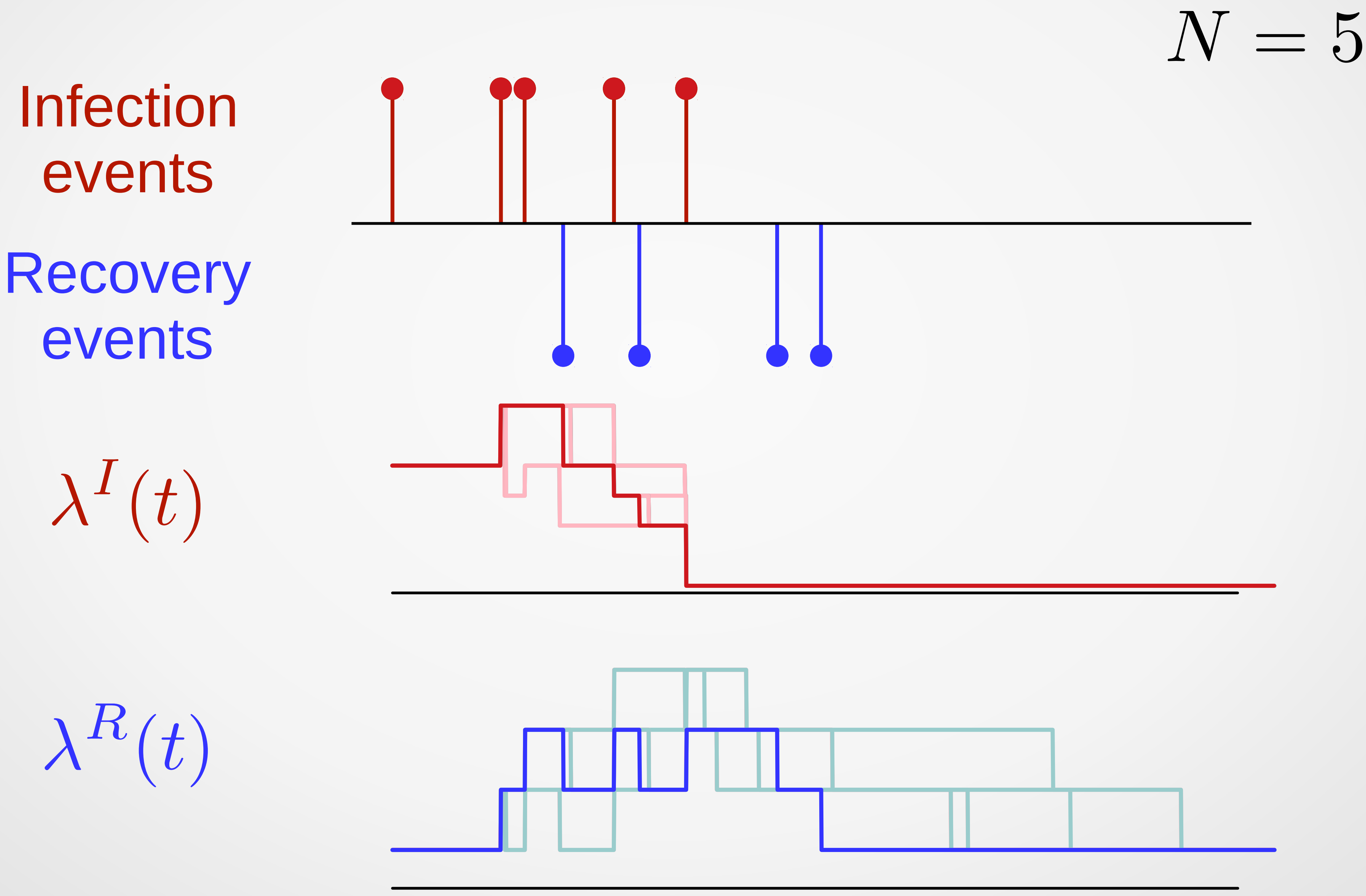
From SIR to HawkesN



From SIR to HawkesN

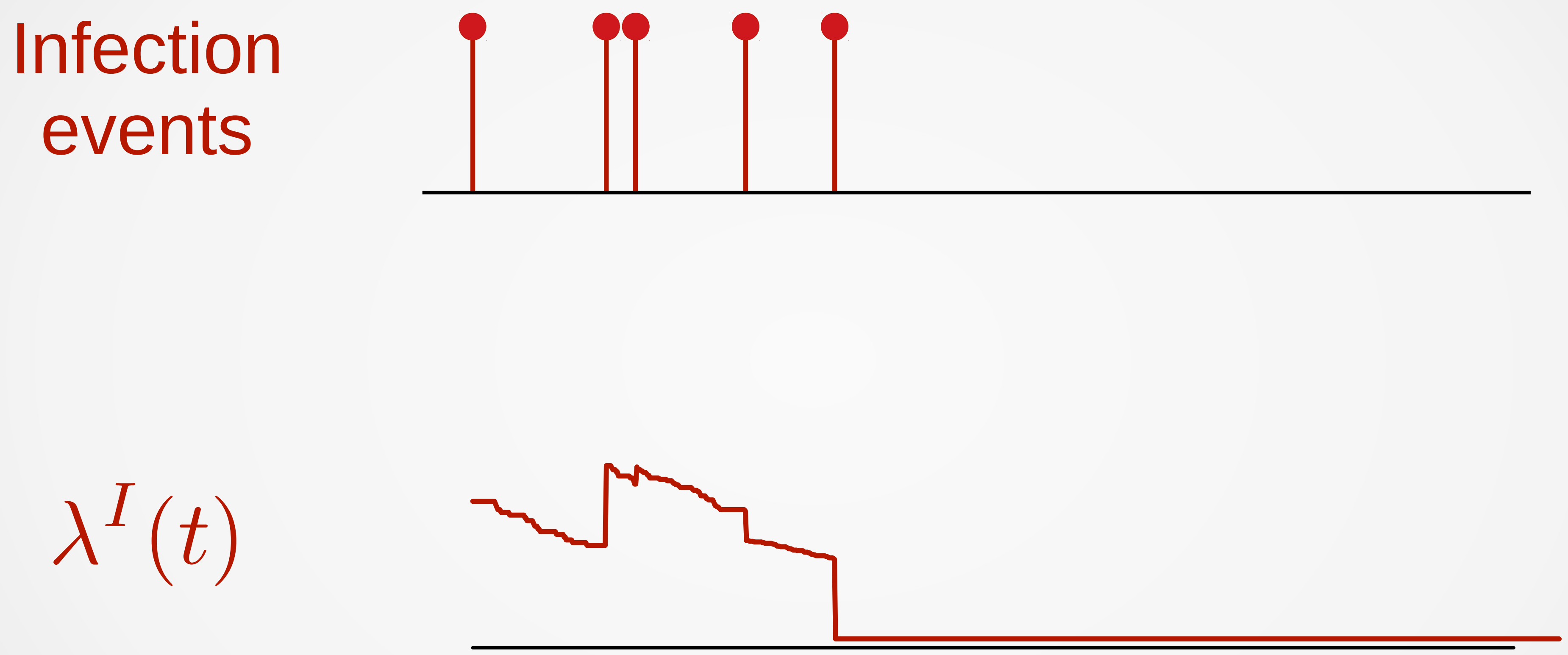


From SIR to HawkesN



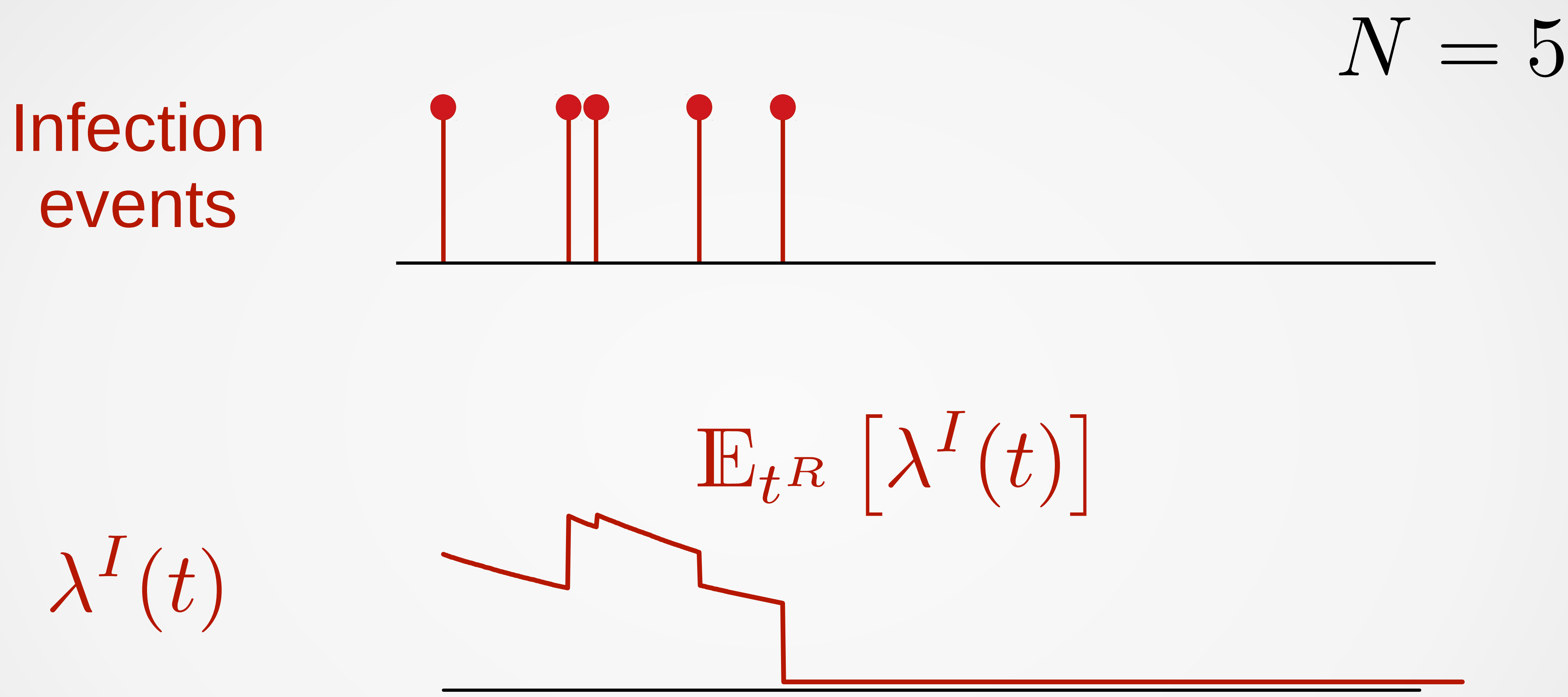
From SIR to HawkesN

$N = 5$



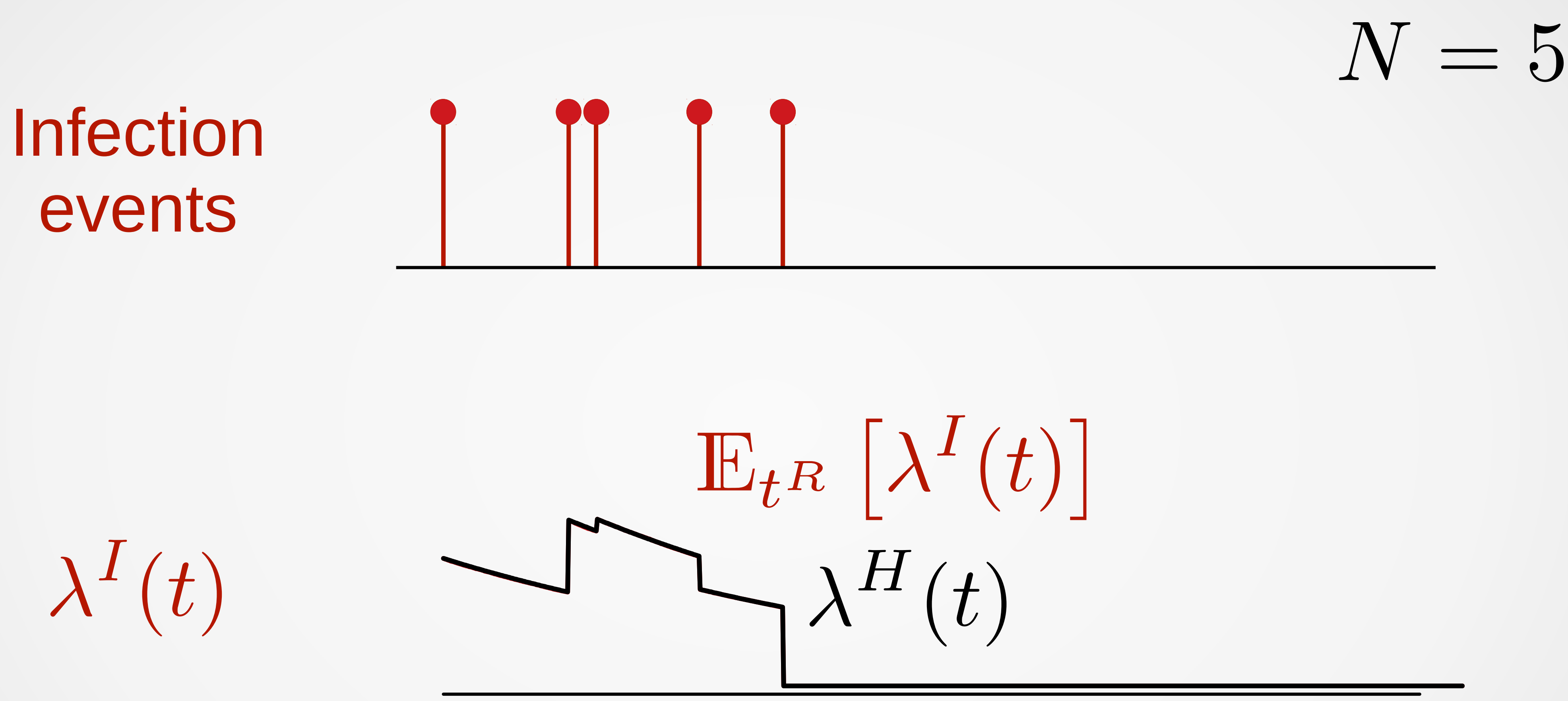
Aggregated over 50 recovery realizations

From SIR to HawkesN



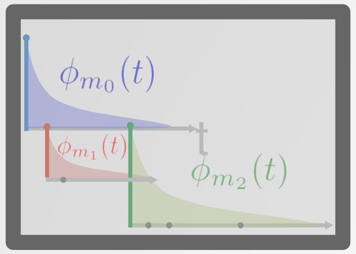
Aggregated over 10,000 recovery realizations

From SIR to HawkesN



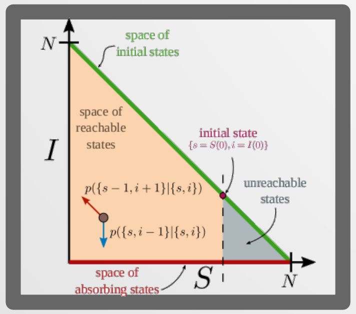
The event intensity of the equivalent HawkesN is the expected new infections intensity

Presentation outline



Prerequisites: Hawkes point processes and SIR infectious models

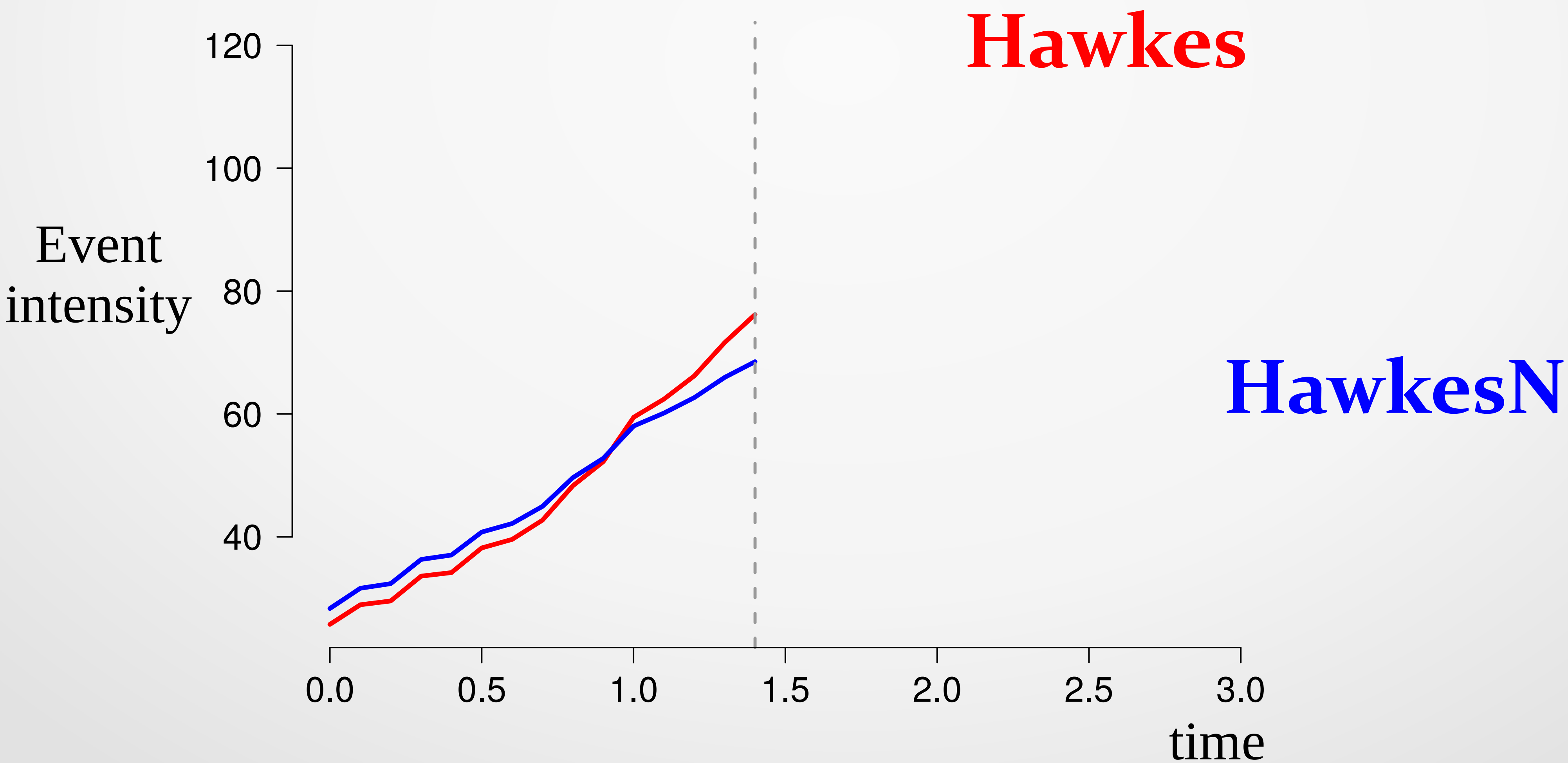
Linking SIR and the Hawkes processes



Computing the distribution of diffusion size

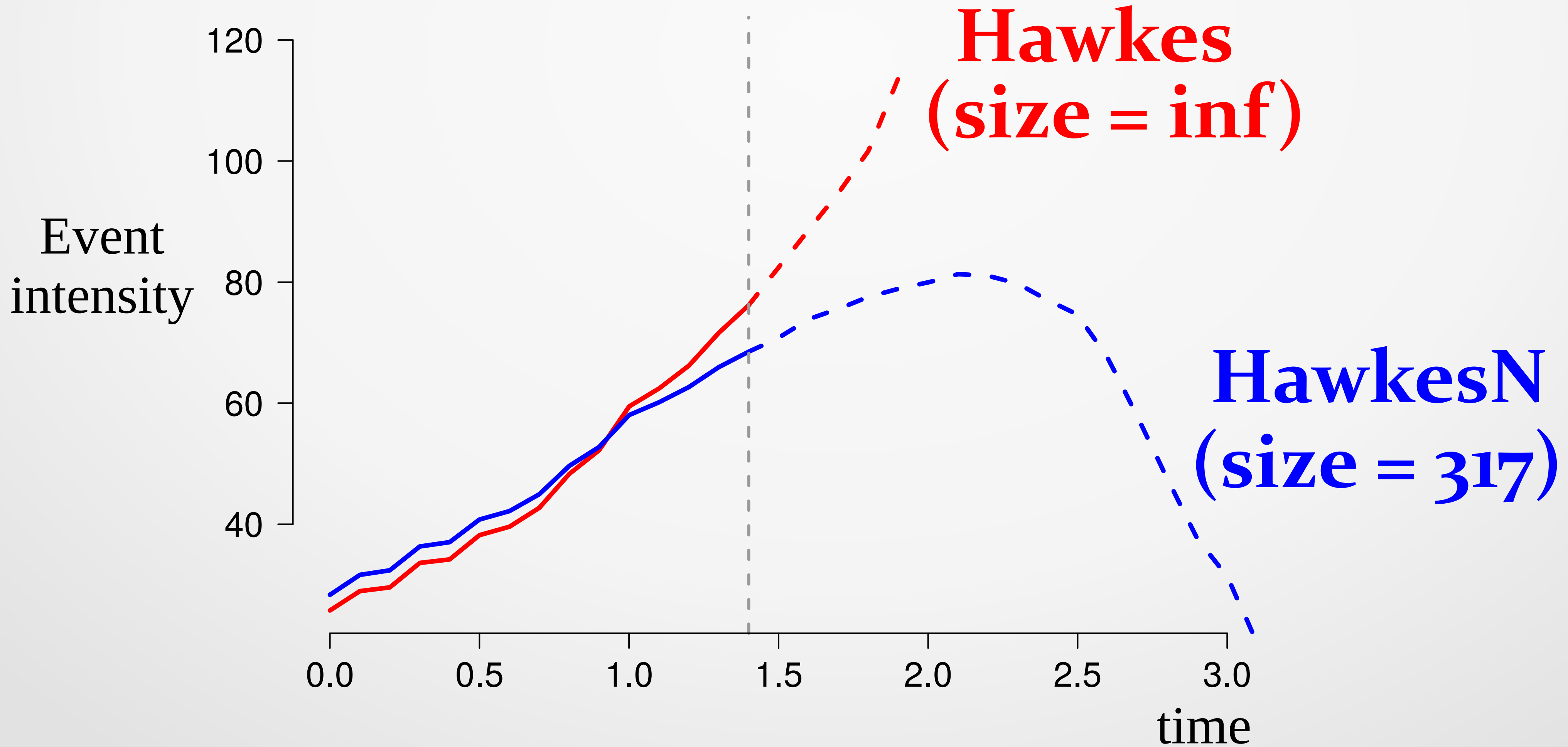
Hawkes and HawkesN in prediction

- 100 observed events;
- predict the final size of the cascade.



Hawkes and HawkesN in prediction

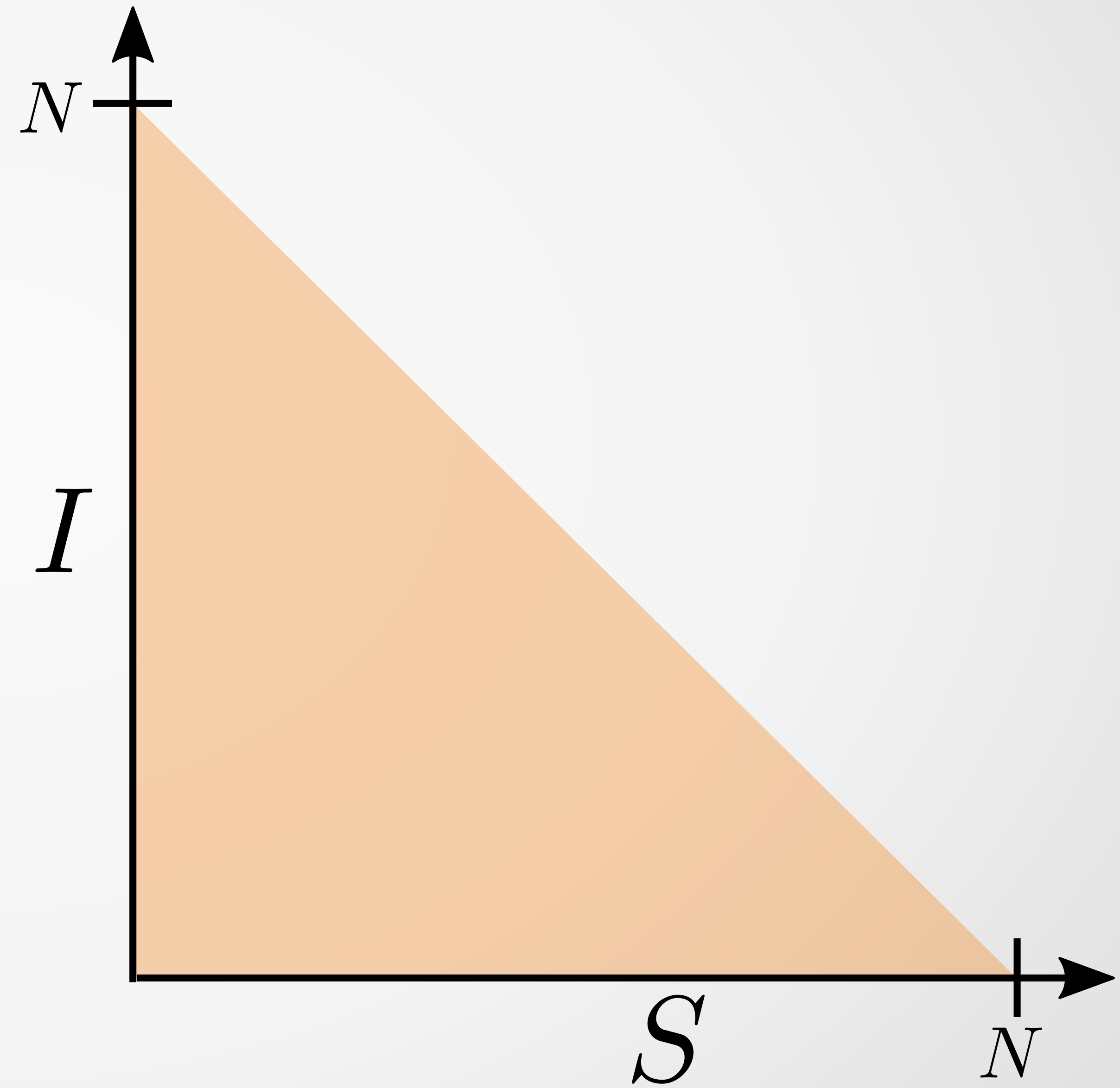
- 100 observed events;
- predict the final size of the cascade.



Distribution of total size

using an SIR Markov chain technique

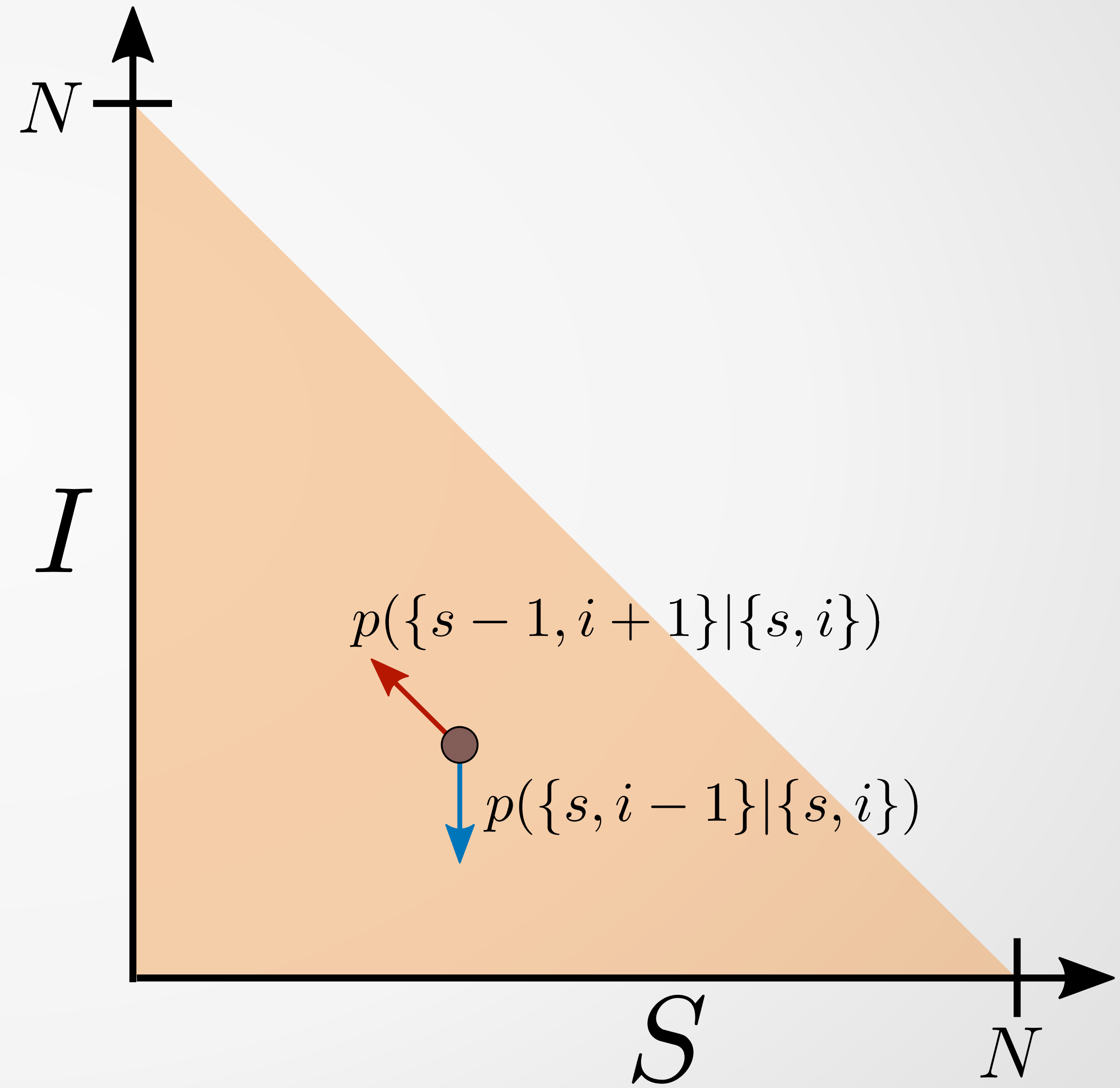
- 2-D space of (S, I)



Distribution of total size

using an SIR Markov chain technique

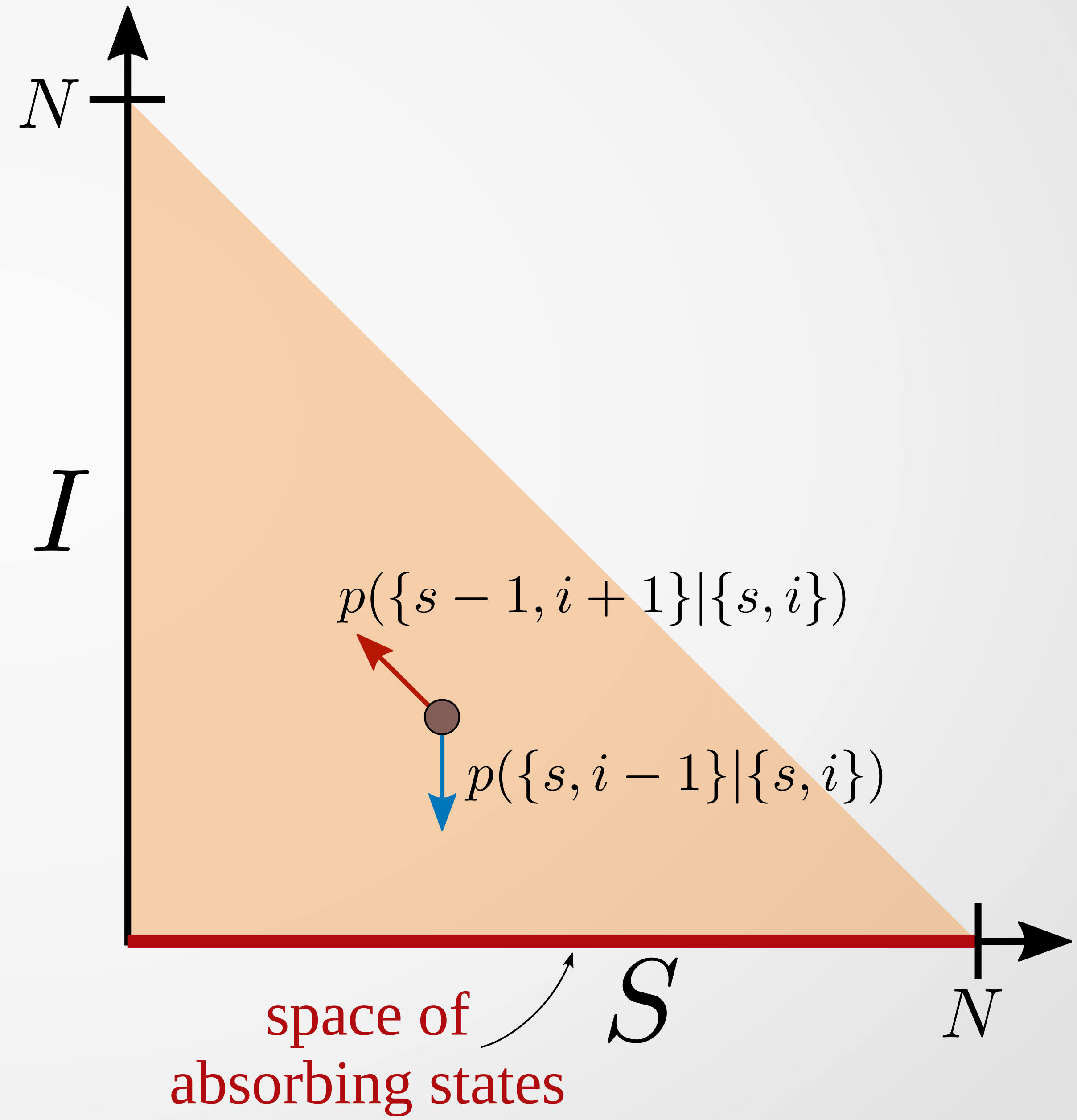
- 2-D space of (S, I)
- From $(S(t) = s, I(t) = i)$:
  - New infection → $(s-1, i+1)$
  - New recovery → $(s, i-1)$



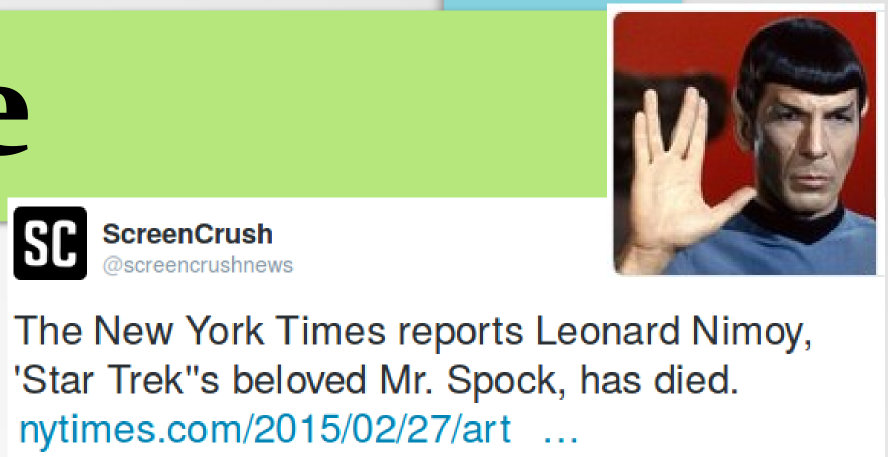
Distribution of total size

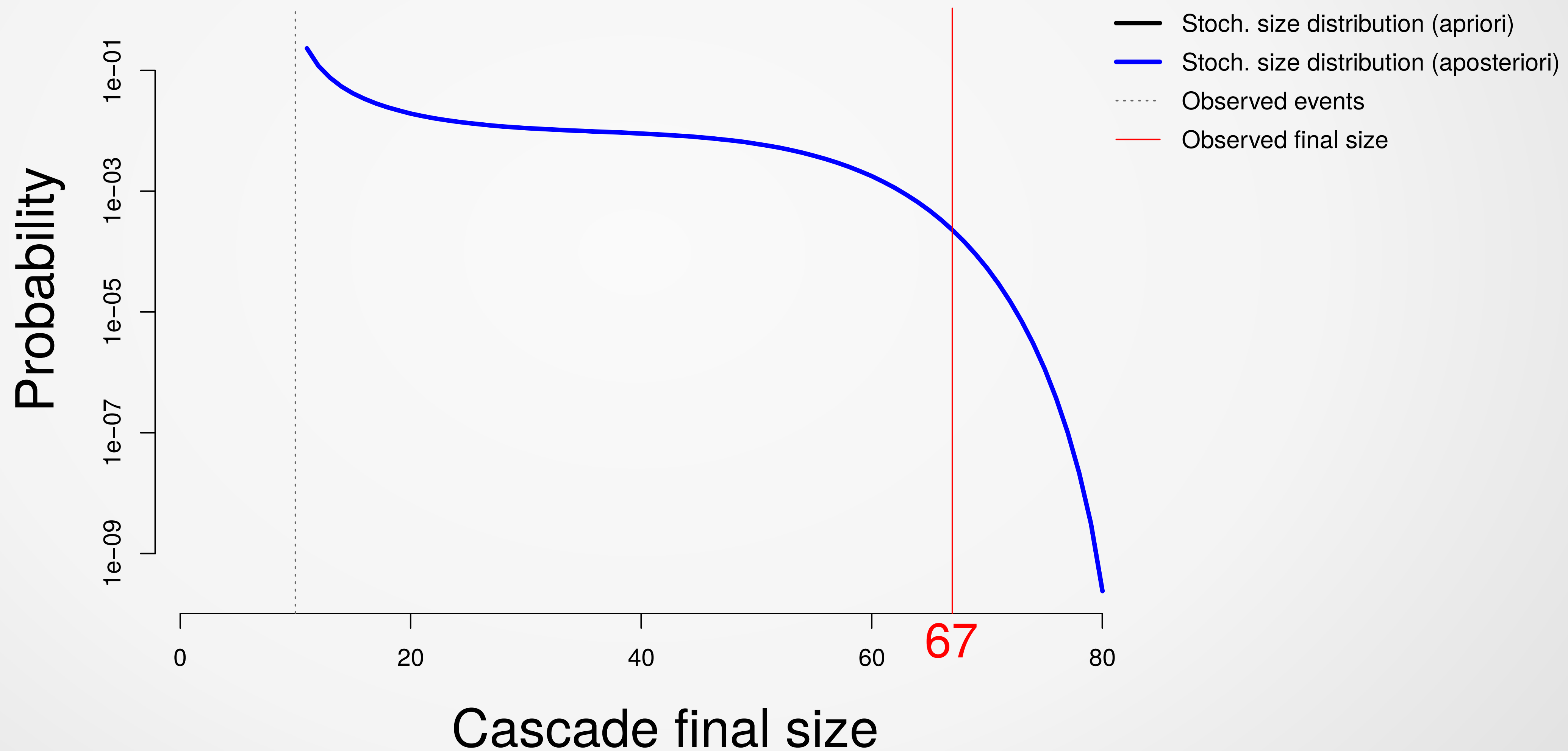
using an SIR Markov chain technique

- 2-D space of (S, I)
- From $(S(t) = s, I(t) = i)$:
  - New infection → $(s\text{-}1, i\text{+}1)$
  - New recovery → $(s, i\text{-}1)$
- States (s, 0) are absorbing
- Probability of total size is the probability of $N\text{-}s$



Example: a tweet cascade



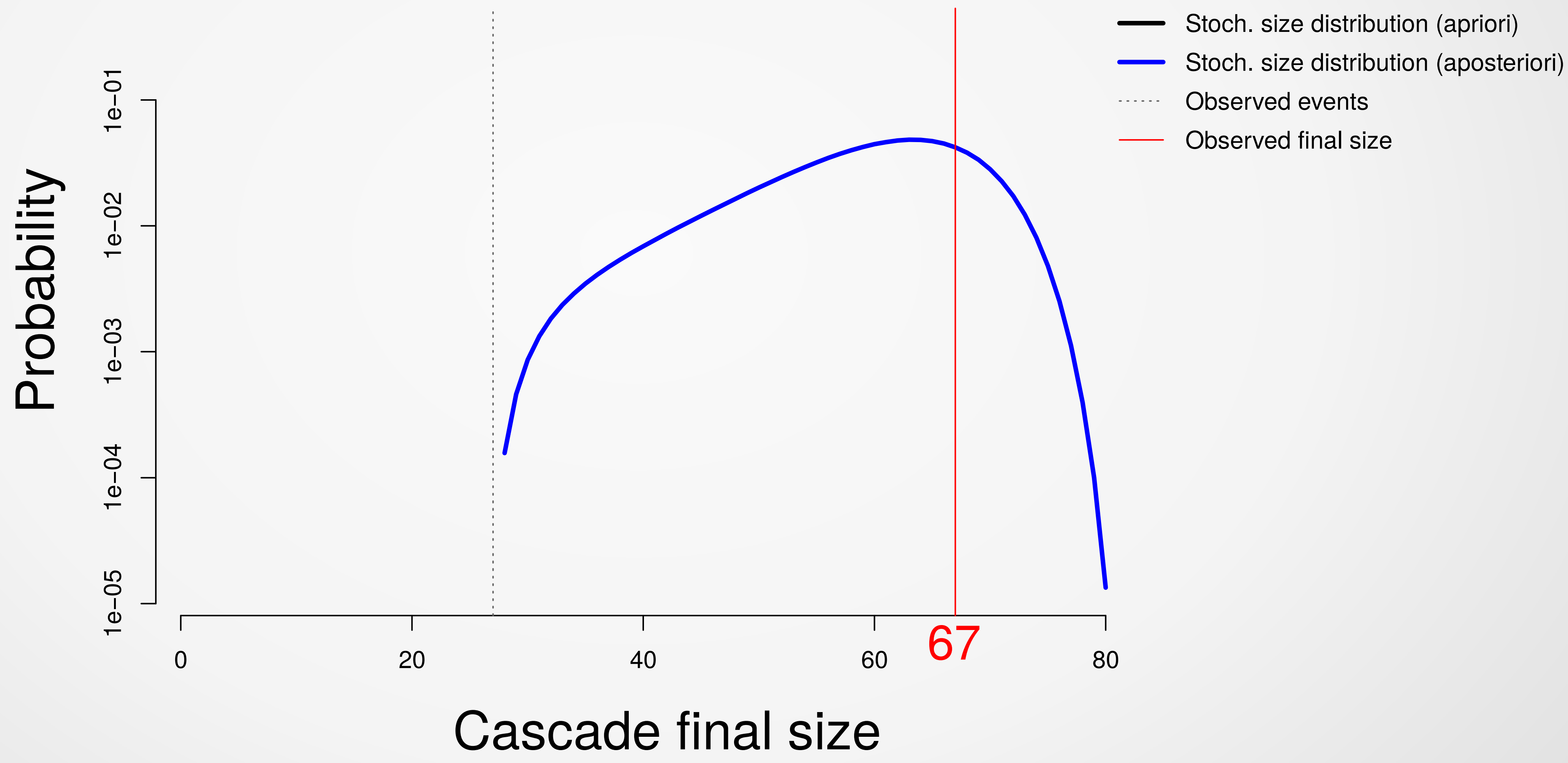


10 seen events

Example: a tweet cascade

ScreenCrush
@screencrushnews

The New York Times reports Leonard Nimoy, 'Star Trek''s beloved Mr. Spock, has died.
nytimes.com/2015/02/27/art …

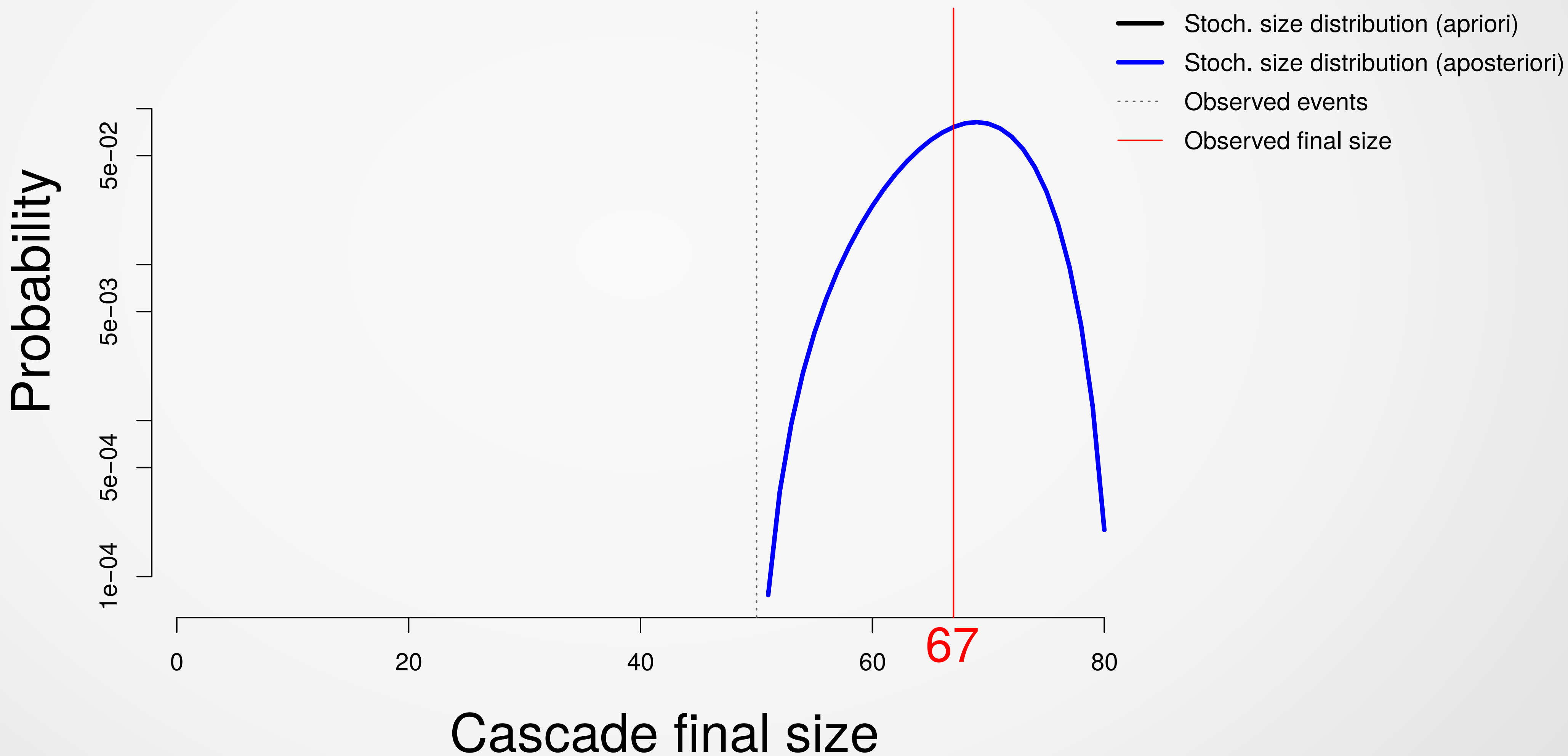


27 seen events

Example: a tweet cascade

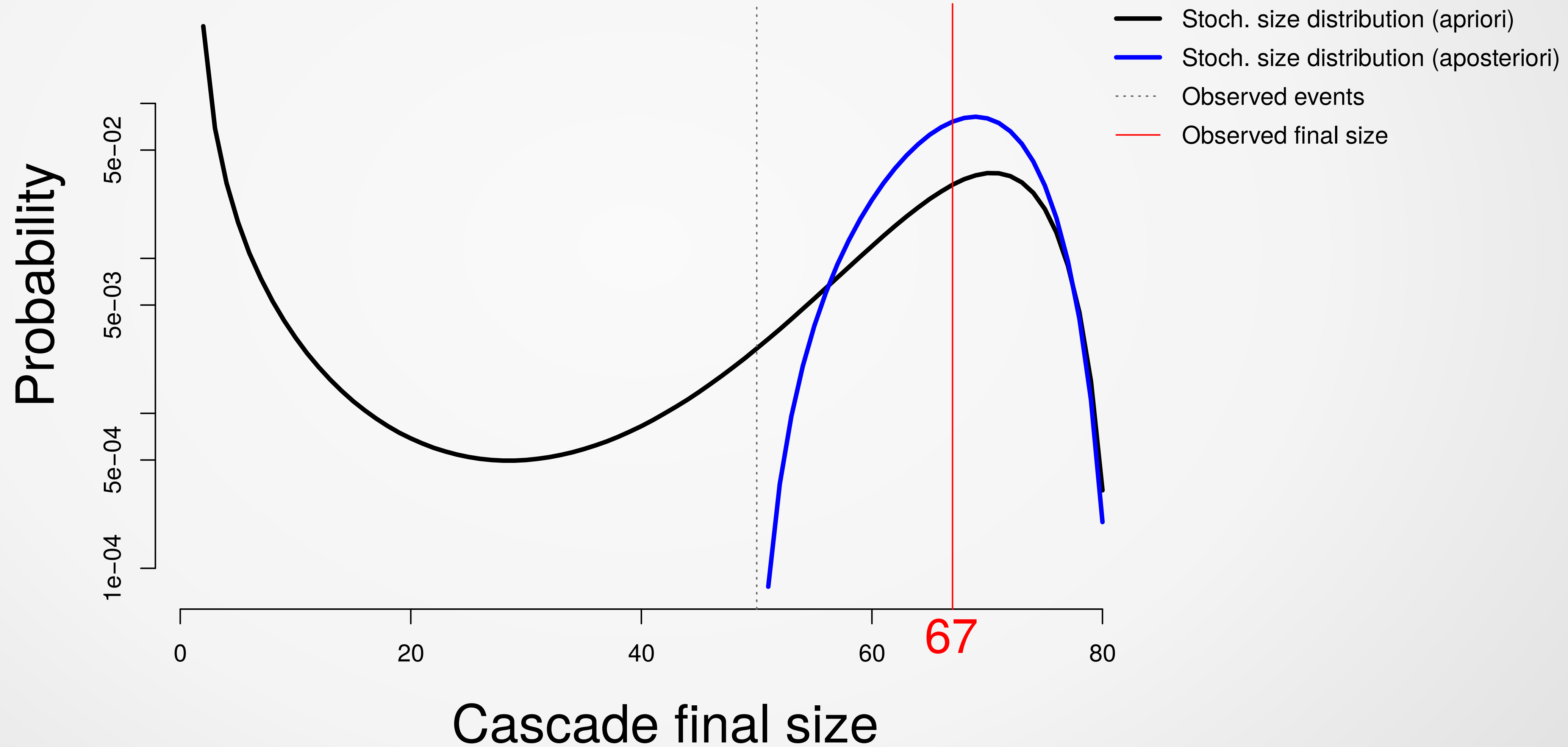
ScreenCrush
@screencrushnews

The New York Times reports Leonard Nimoy, 'Star Trek"s beloved Mr. Spock, has died.
nytimes.com/2015/02/27/art …



50 seen events

Example: a tweet cascade

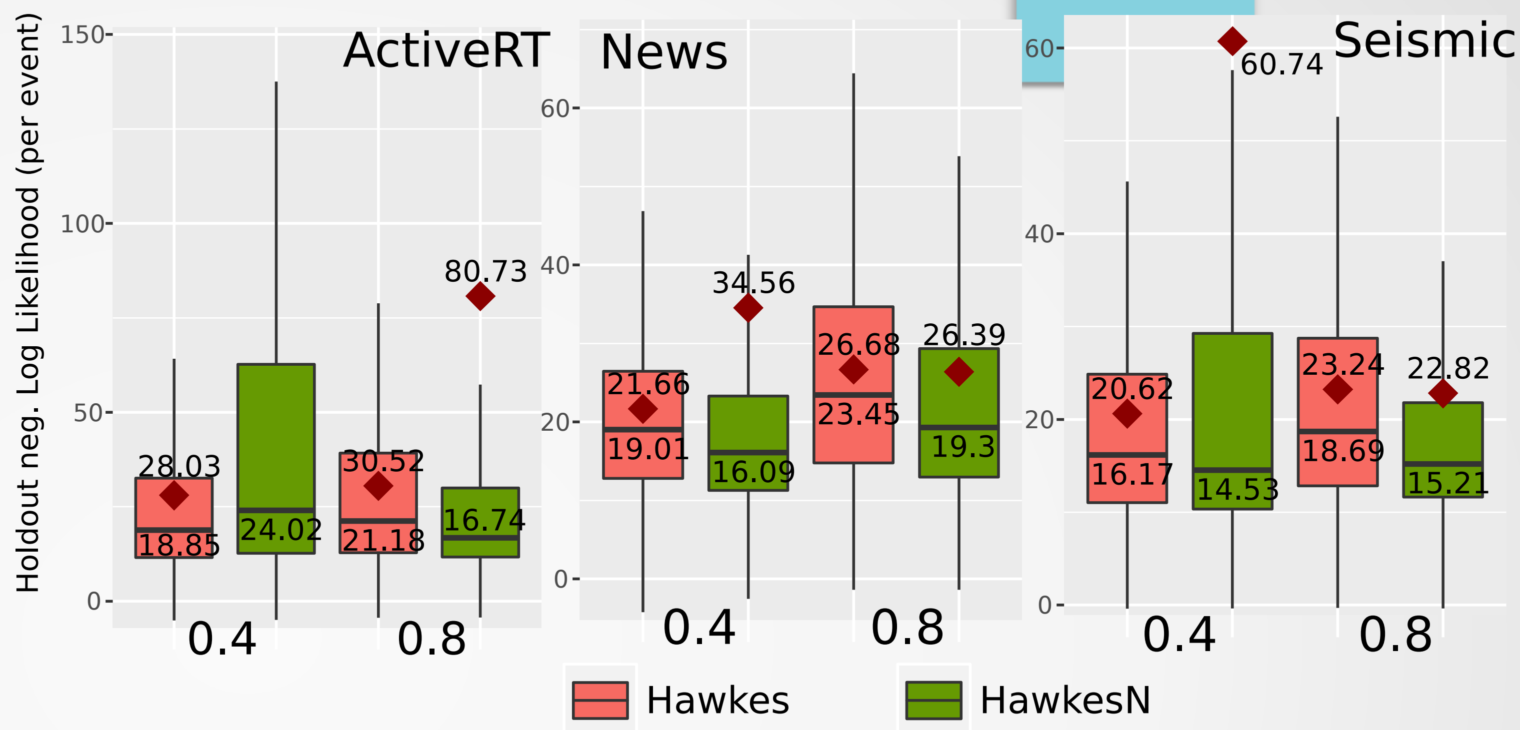


50 seen events

Explanation for the unpredictability of online popularity

HawkesN generalization

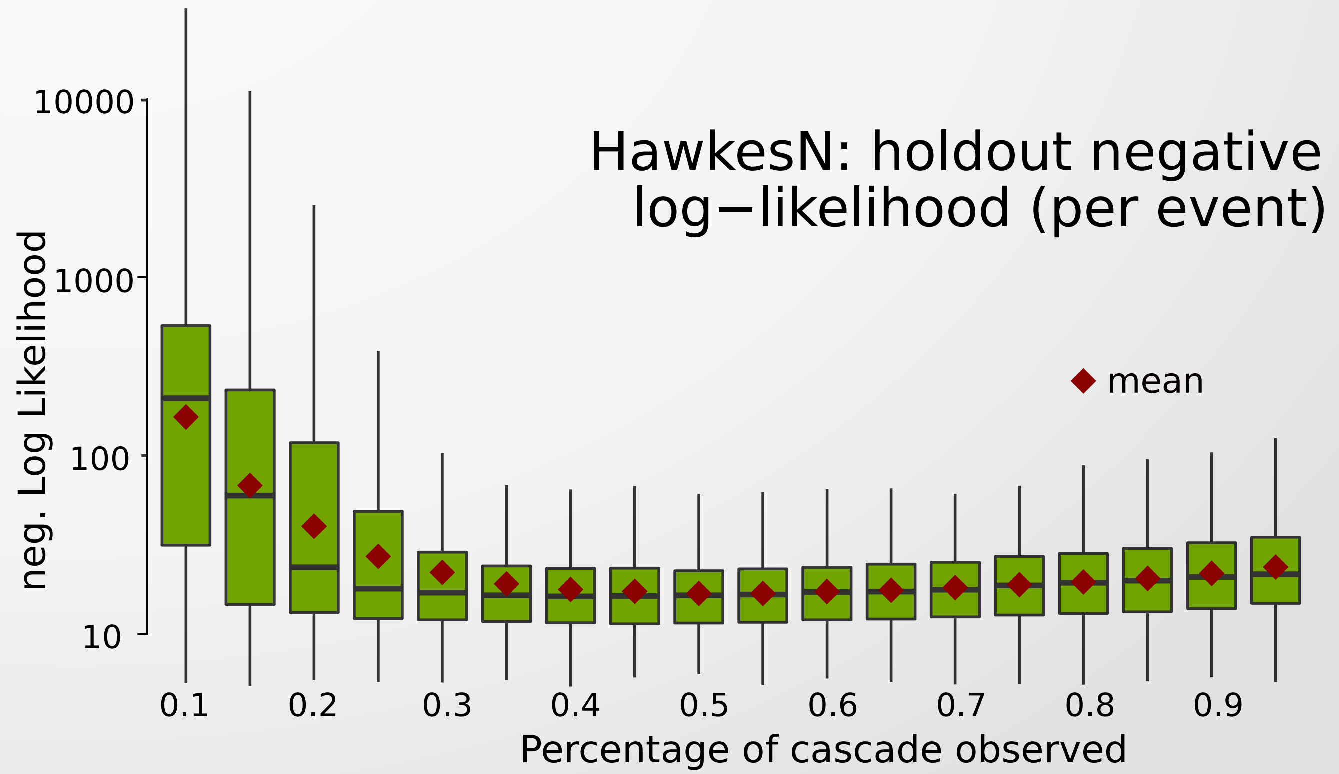
HawkesN generalizes better than Hawkes on real-life cascades



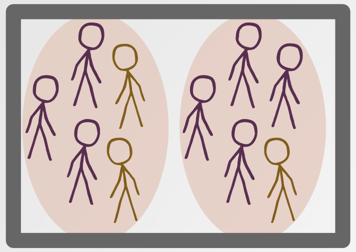
Caveat:

Estimating N from data is unreliable.

New statistic for diagnostic.

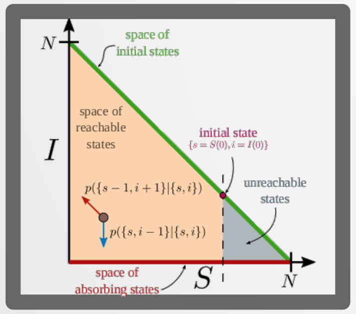


Summary



HawkesN: an extension of Hawkes accounting for a finite population

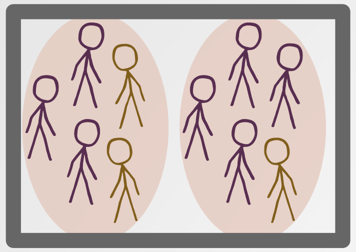
Connecting SIR epidemic models and HawkesN through the expected new infection intensity



A Markov Chain tool for computing the distribution of final size adapted to HawkesN

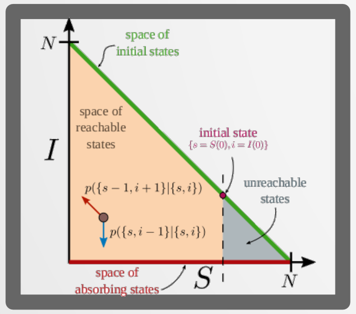
Limitations & future work: Fixed population, N estimated from each cascade, other kernels in HawkesN.

Thank you!



HawkesN: an extension of Hawkes accounting for a finite population

Connecting SIR epidemic models and HawkesN through the expected new infection intensity



A Markov Chain tool for computing the distribution of final size adapted to HawkesN

Limitations & future work: Fixed population, *N* estimated from each cascade, other kernels in HawkesN.

Data & code: https://github.com/computationalmedia/sir-hawkes

Supp: Estimating *I(0)* in HawkesN

Issue:

Recovery events are unobserved in HawkesN → the number of infected is unknown.

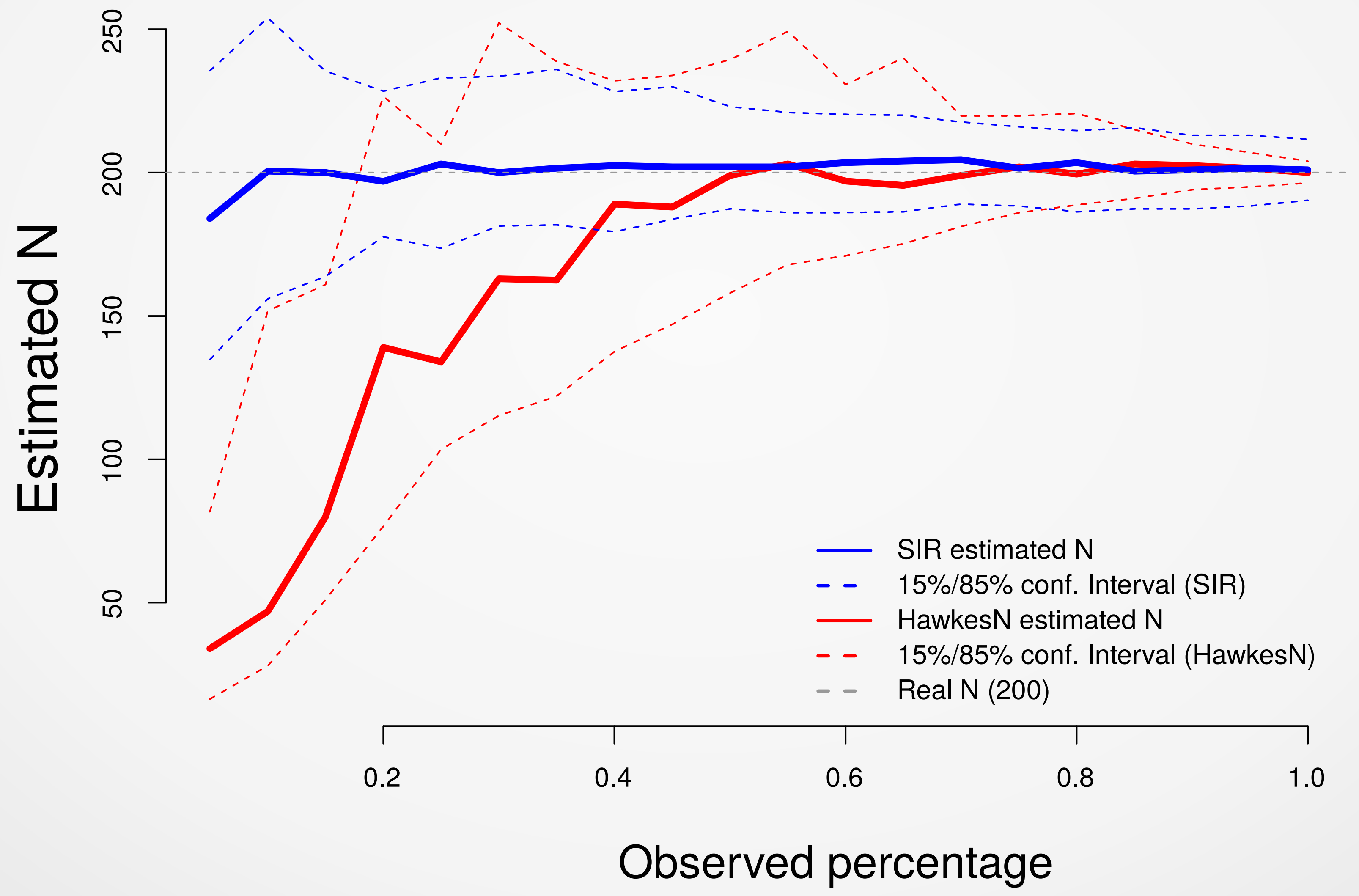
Solution:

Estimate its expected value

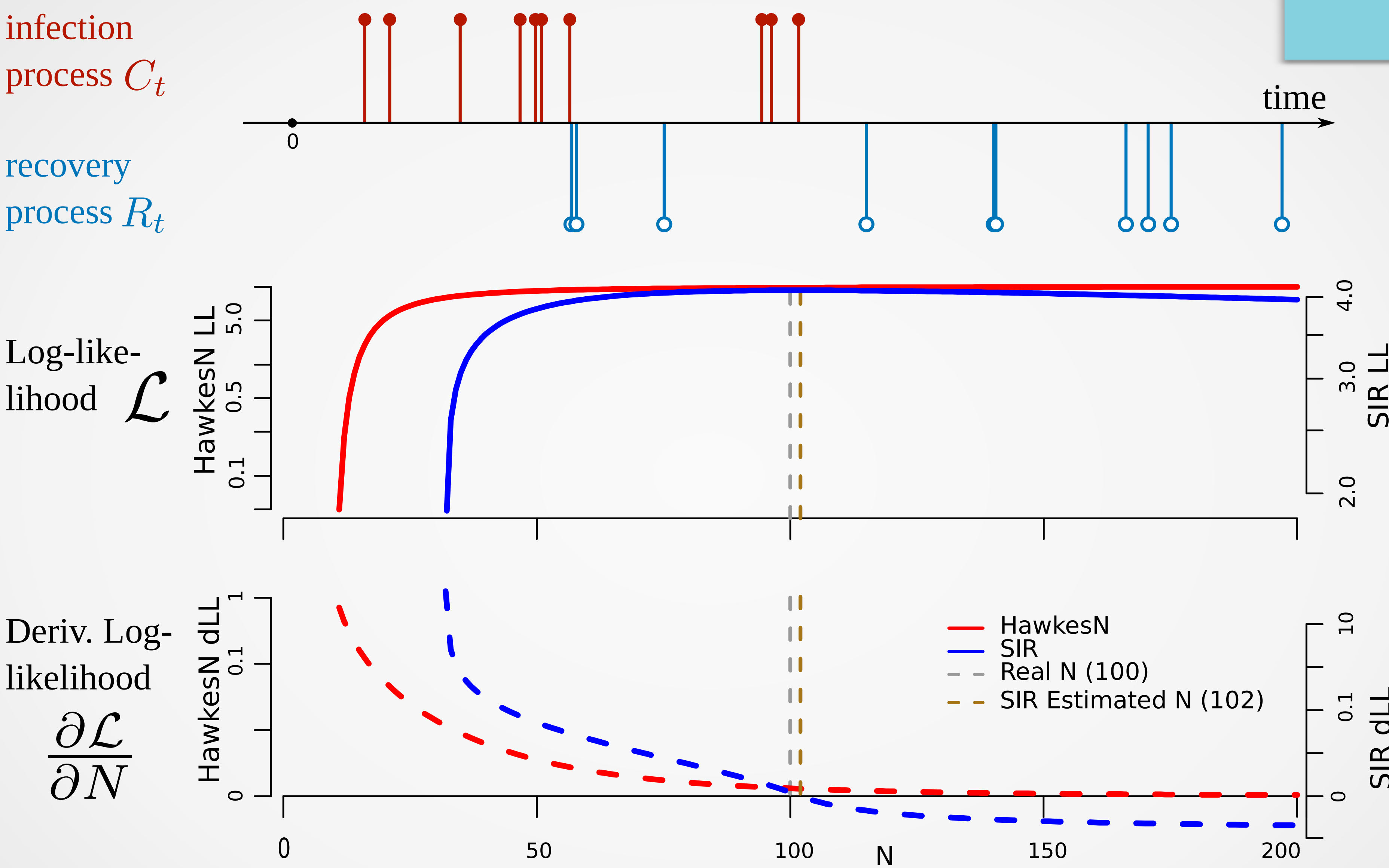
$$\mathbb{E}_{t^R}[I(0)] = \mathbb{E}_{t^R}\left[\sum_{j=1}^{l} \mathbb{1}(t_j^R > t_l)\right] = \sum_{j=1}^{l} e^{-\gamma(t_l - t_j^I)}$$

when $t_1, t_2, \ldots, t_l$ are the *l* observed events.

Supp: (under) Estimating N from data



Supp: Estimating N from data



$$\frac{\partial \mathcal{L}}{\partial N} \geq \frac{1}{N^2} \underbrace{\left(\frac{(n-1)n}{2} - \sum_{j=0}^{n-1} \sum_{l=j}^{n-1} l\kappa \left[e^{-\theta(t_l - t_j)} - e^{-\theta(t_{l+1} - t_j)} \right] \right)}_{\mathcal{S}(\kappa, \theta, \{t_1, t_2, \ldots, t_n\})}$$