Model Fitting and Inference for Infectious Disease Dynamics

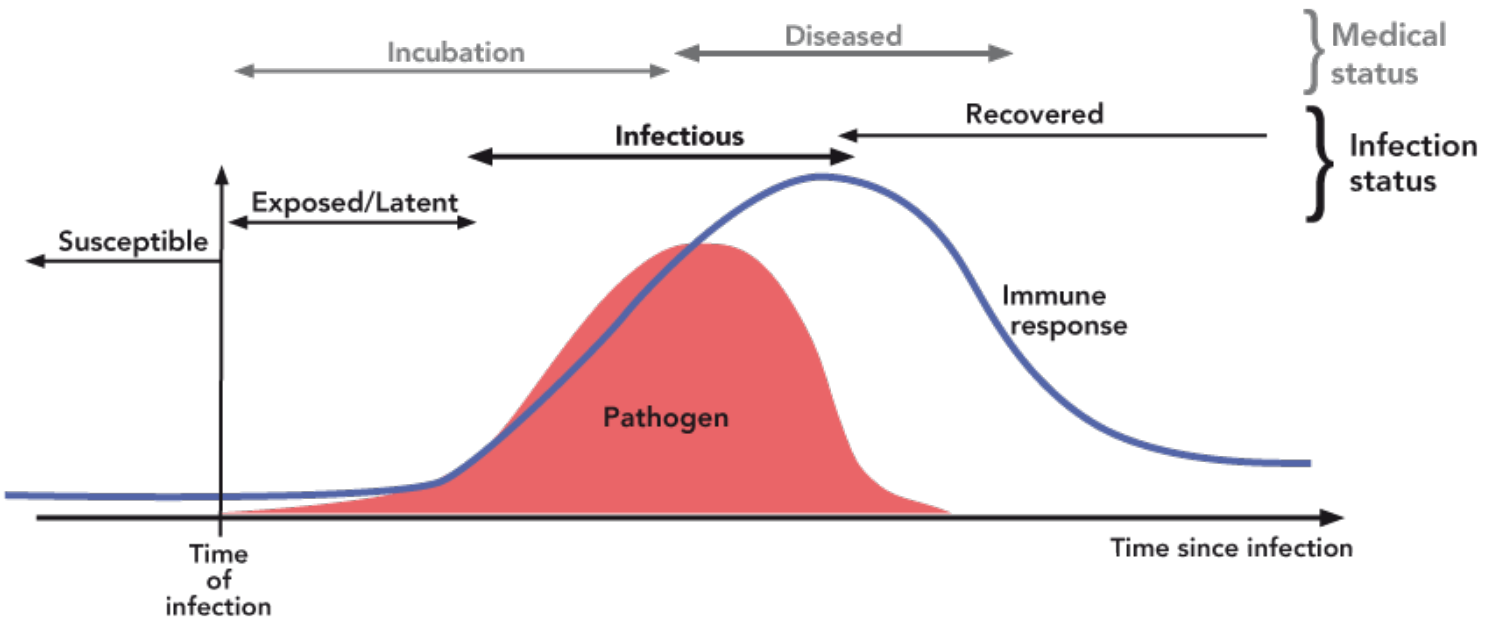
Introduction to Stochastic Modelling

centre *for the* mathematical modelling *of* infectious diseases

Objectives

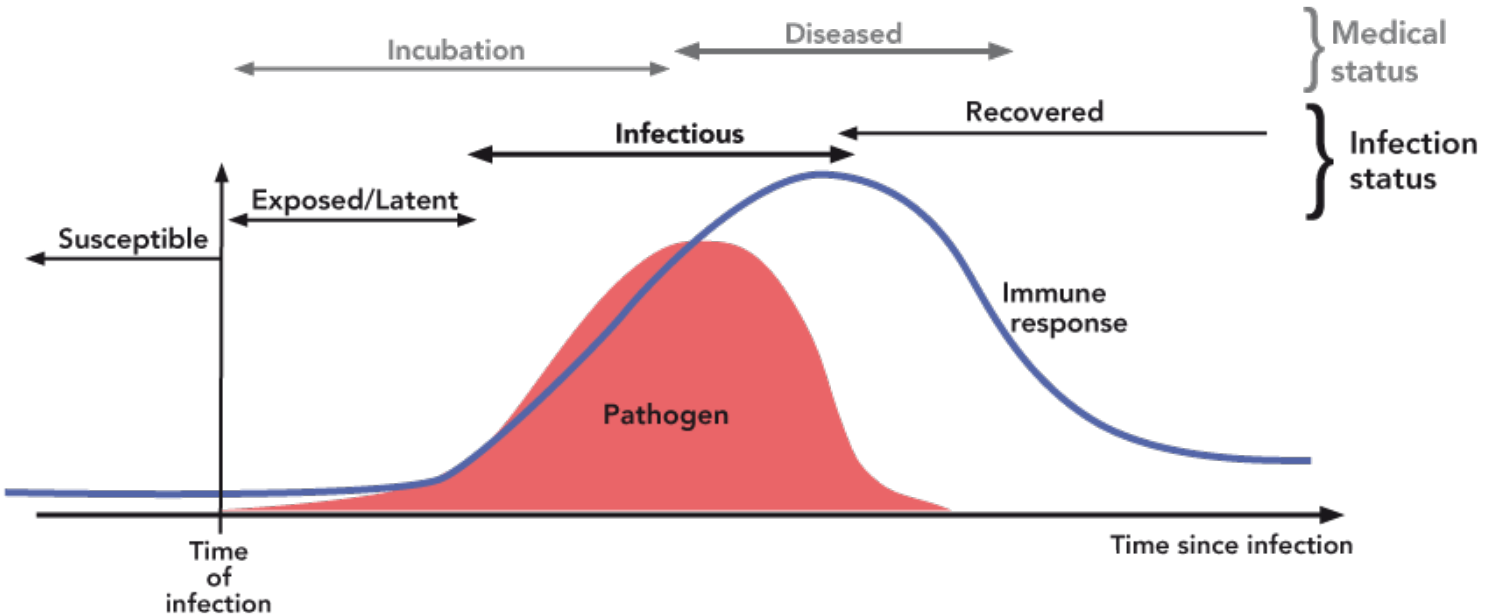
- Introduce basic mathematical and statistical concepts of infectious disease modelling.
- Define the structure and formalism of models.
- Present the relationship between deterministic and stochastic models. Why using stochastic modelling?
- Show how these models can be fitted rigorously to data.

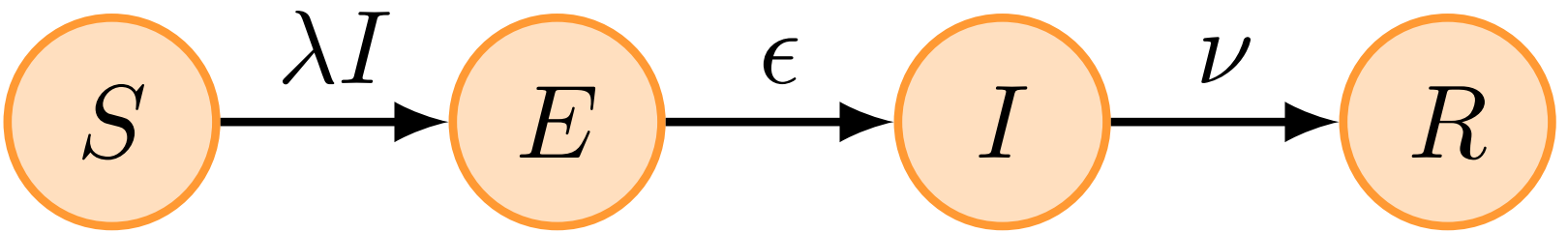
Infection dynamics



Keeling & Rohani (2008)

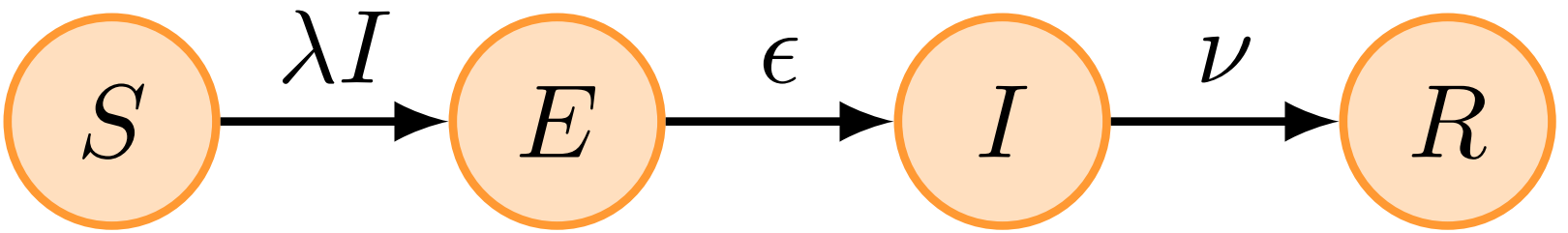
Infection dynamics





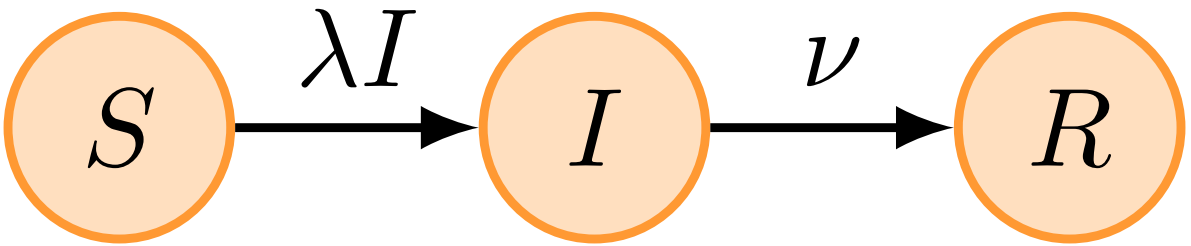
Keeling & Rohani (2008)

The SEIR model



- λ: rate at which two individuals come into contact.
- λI: per capita force of infection.
- ϵ: inverse of the latent period.
- ν: inverse of the infectious period.

The deterministic SIR model

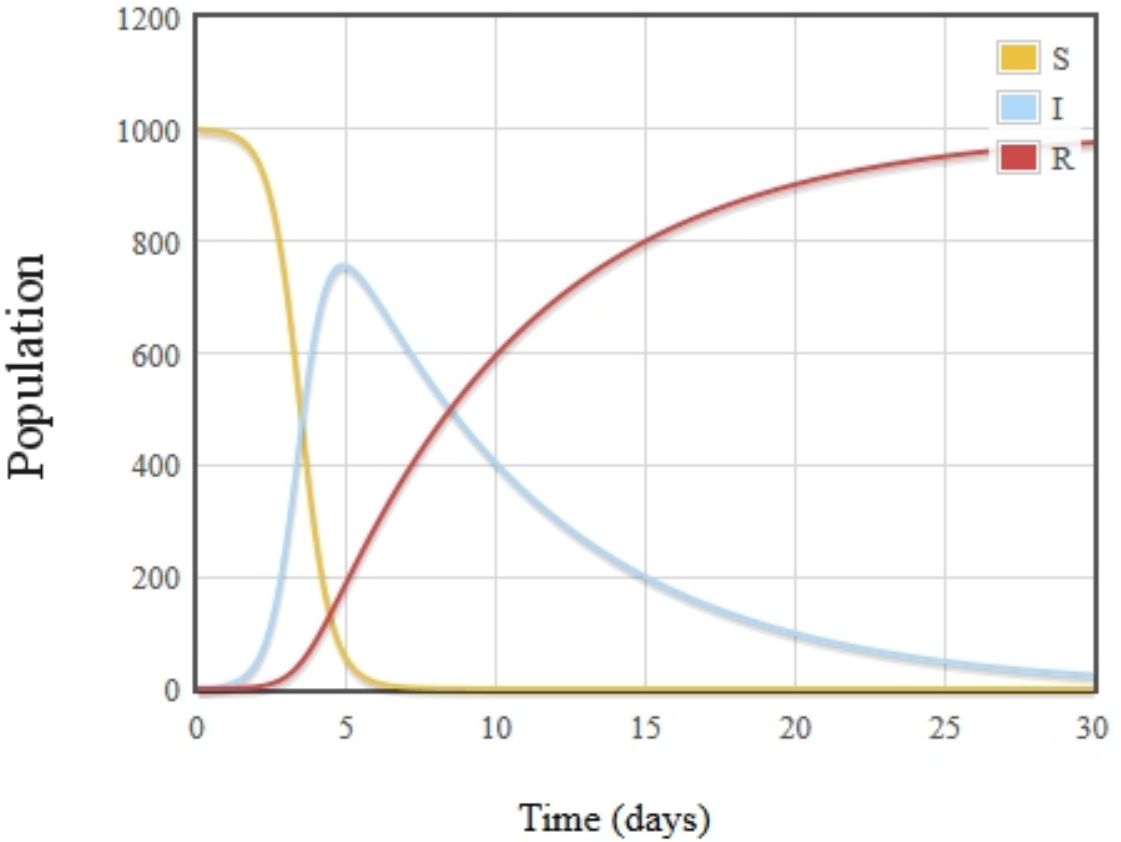


$$\begin{cases} \dfrac{\mathrm{d}S(t)}{\mathrm{d}t} = -\lambda S(t) I(t) \\ \dfrac{\mathrm{d}I(t)}{\mathrm{d}t} = \lambda S(t) I(t) - \nu I(t) \\ \dfrac{\mathrm{d}R(t)}{\mathrm{d}t} = \nu I(t) \end{cases} \tag{1}$$

With $\big(S(0), I(0), R(0)\big) = (S_0, I_0, 0)$
and constant population size $S(t) + I(t) + R(t) = \Omega$.

Kermack & McKendrick (1927)

The deterministic SIR model



Kermack & McKendrick (1927)

Two important results

- An epidemic can occur only if $S_0 > \nu/\lambda$, when the population initially susceptible is above a critical size.
- At the end of the epidemic, it remains S_∞ susceptible individuals, with S_∞ solution of the equation:

$$S_\infty = \Omega + \frac{\nu}{\lambda} \ln(\frac{S_\infty}{S_0}). \quad (2)$$

- Not all susceptibles will get infected during the epidemic.

Kermack & McKendrick (1927)

Why using stochastic models?

1. It is the natural way to describe how an epidemic disease spreads.
2. Some phenomena are stochastic by nature and cannot be described in a deterministic setting.
3. Take into account the variability of the epidemic process when estimating and forecasting.

Why using stochastic models?

1. It is the natural way to describe how an epidemic disease spreads.
2. Some phenomena are stochastic by nature and cannot be described in a deterministic setting.
3. Take into account the variability of the epidemic process when estimating and forecasting.

Why using stochastic models?

1. It is the natural way to describe how an epidemic disease spreads.
2. Some phenomena are stochastic by nature and cannot be described in a deterministic setting.
3. Take into account the variability of the epidemic process when estimating and forecasting.

Three sources of stochasticity

1. Demographic (internal).
   Depends on the epidemiogical process.
2. Environmental (external).
   Acts on the epidemiogical process.
3. Observation.
   Does not change the epidemiological dynamics.

Observation stochasticity

- Diagnostic errors: false positive and false negative.
- Incomplete reporting of cases: 60% for measles.
- Fluctuations of the reporting rate: change in the number of GPs in the surveillance system.
- The reporting rate at time t is $\rho_t \sim \text{Gamma}(1/\phi, \rho\phi)$.
- Conditioning on ρ_t and the incidence in the model (X_t), the observed incidence $Y_t|\rho_t, X_t \sim \text{Poisson}(\rho_t X_t)$.
- Conditioning on X_t, Y_t follows a negative-binomial distribution with:

$$\text{E}[Y_t|X_t] = \rho X_t \text{ and } \text{Var}[Y_t|X_t] = \rho X_t + \phi\rho^2 X_t^2.$$

- This corresponds to an overdispersed observation process.

Fine & Clarkson (1982), Bretó et al. (2009)

Observation stochasticity

- Diagnostic errors: false positive and false negative.
- Incomplete reporting of cases: 60% for measles.
- Fluctuations of the reporting rate: change in the number of GPs in the surveillance system.
- The reporting rate at time t is $\rho_t \sim \text{Gamma}(1/\phi, \rho\phi)$.
- Conditioning on ρ_t and the incidence in the model (X_t), the observed incidence $Y_t|\rho_t, X_t \sim \text{Poisson}(\rho_t X_t)$.
- Conditioning on X_t, Y_t follows a negative-binomial distribution with:

$$\text{E}[Y_t|X_t] = \rho X_t \text{ and } \text{Var}[Y_t|X_t] = \rho X_t + \phi\rho^2 X_t^2.$$

- This corresponds to an overdispersed observation process.

Fine & Clarkson (1982), Bretó et al. (2009)

Observation stochasticity

- Diagnostic errors: false positive and false negative.
- Incomplete reporting of cases: 60% for measles.
- Fluctuations of the reporting rate: change in the number of GPs in the surveillance system.
- The reporting rate at time t is $\rho_t \sim \text{Gamma}(1/\phi, \rho\phi)$.
- Conditioning on ρ_t and the incidence in the model (X_t), the observed incidence $Y_t|\rho_t, X_t \sim \text{Poisson}(\rho_t X_t)$.
- Conditioning on X_t, Y_t follows a negative-binomial distribution with:

$$\text{E}[Y_t|X_t] = \rho X_t \text{ and } \text{Var}[Y_t|X_t] = \rho X_t + \phi\rho^2 X_t^2.$$

- This corresponds to an overdispersed observation process.

Fine & Clarkson (1982), Bretó et al. (2009)

Environmental stochasticity

- Stochastic fluctuations of environmental factors (e.g. temperature, humidity) lead to stochastic fluctuations of transmission parameters.

$$\begin{cases} \dfrac{\mathrm{d}S(t)}{\mathrm{d}t} = -\lambda(1+F\xi)S(t)I(t) \\ \dfrac{\mathrm{d}I(t)}{\mathrm{d}t} = \lambda(1+F\xi)S(t)I(t) - \nu I(t) \\ \dfrac{\mathrm{d}R(t)}{\mathrm{d}t} = \nu I(t) \end{cases} \tag{3}$$

where ξ is a random variable and F is a forcing constant.

Bretó et al. (2009)

Demographic stochasticity

- Results from the discrete nature of individuals in the population: $\big(S(t), I(t), R(t)\big) \in \mathbb{N}^3$.
- Mechanistic modelling of random events at the individual level:
  - Infectious period with mean ν^{-1} and variance σ^2.
  - Number of contacts $\sim$ Poisson process with intensity β.
- One can compute the distribution of the final size of the epidemic $(S_0 - S_\infty)$.
- To go further, let's assume an exponentially distributed infectious period: $\nu^{-2} = \sigma^2$.
- Memory-less property of the exponential distribution: $\{\big(S(t), I(t), R(t)\big) : t \geq 0\}$ becomes a discrete state, continuous time Markov process.

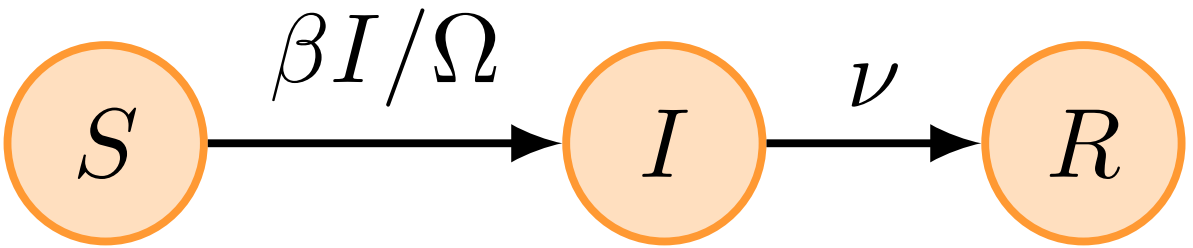
Andersson & Britton (2000)

Demographic stochasticity

- Results from the discrete nature of individuals in the population: $\big(S(t), I(t), R(t)\big) \in \mathbb{N}^3$.
- Mechanistic modelling of random events at the individual level:
  - Infectious period with mean ν^{-1} and variance σ^2.
  - Number of contacts $\sim$ Poisson process with intensity β.
- One can compute the distribution of the final size of the epidemic $(S_0 - S_\infty)$.
- To go further, let's assume an exponentially distributed infectious period: $\nu^{-2} = \sigma^2$.
- Memory-less property of the exponential distribution: $\{\big(S(t), I(t), R(t)\big) : t \geq 0\}$ becomes a discrete state, continuous time Markov process.

Andersson & Britton (2000)

SIR with demographic stochasticity



Event	Transition	Transition probability within $[t, t + \mathrm{d}t]$
Infection	$(s, i, r) \to (s - 1, i + 1, r)$	$\frac{\beta}{\Omega} si \, \mathrm{d}t + o(\mathrm{d}t)$
Recovery	$(s, i, r) \to (s, i - 1, r + 1)$	$\nu i \, \mathrm{d}t + o(\mathrm{d}t)$

Bartlett (1949)

Doob-Gillespie algorithm

- Proposed by Joseph L. Doob in 1940s and popularized by Daniel T. Gillespie in 1970s.
- Given a state $\mathbf{x}$ and a time t, simulate the following event: *No transition occurs during the time interval* $[t, t+\tau[$ *and the transition* T_μ *occurs at the time* $t+\tau$.
- One can show that τ and μ are two random variables with probability density:

$$p(\tau) = a_0(\mathbf{x}) \exp(-a_0(\mathbf{x})\tau),\ \tau > 0, \quad (5)$$

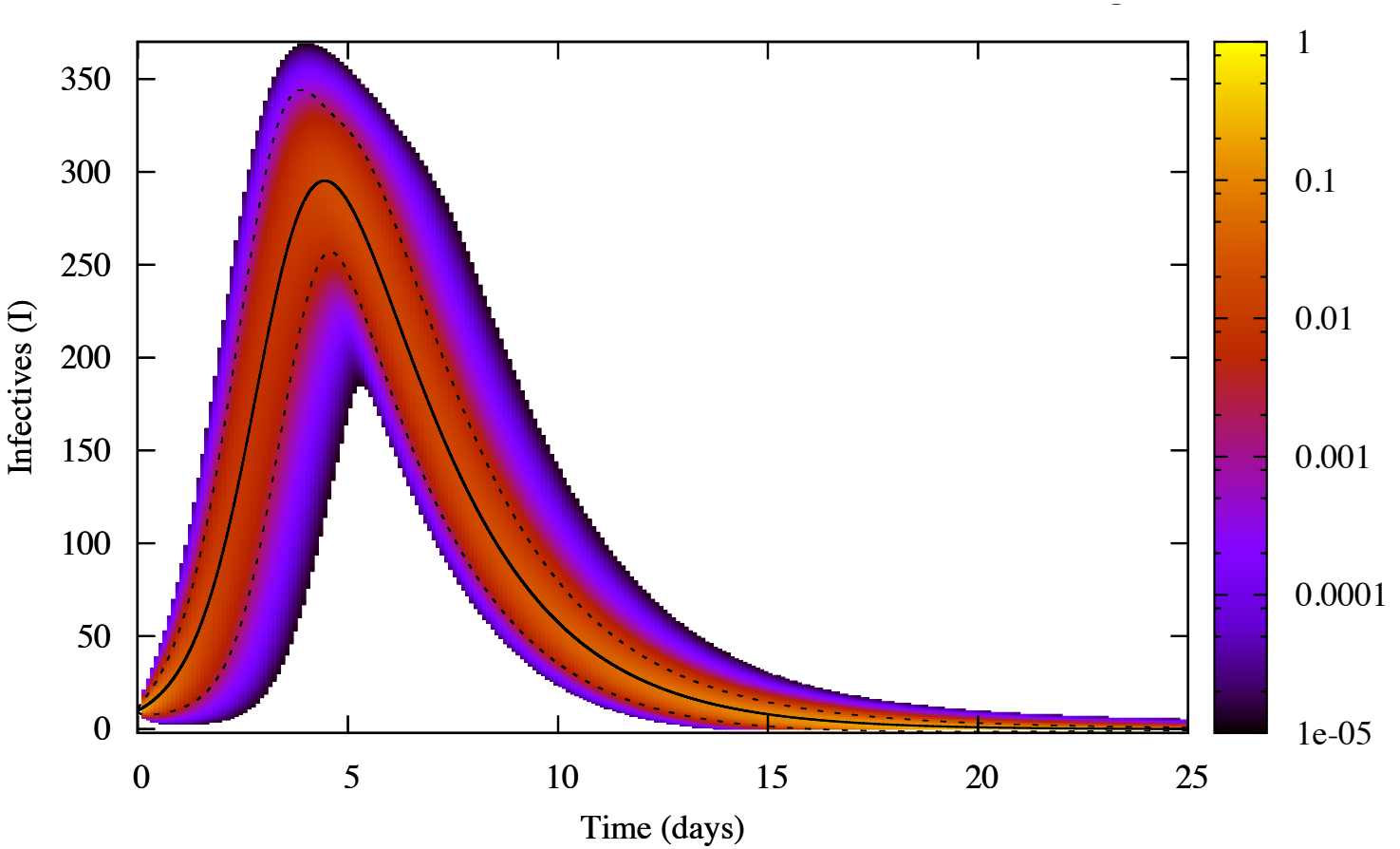
$$p(\mu) = \frac{a_\mu(\mathbf{x})}{a_0(\mathbf{x})},\ \mu = 1, \ldots, M, \quad (6)$$

where $a_0(\mathbf{x}) = \sum_{m=1}^{M} a_m(\mathbf{x})$.

- Slow for large populations because $\mathrm{E}[\tau] = 1/a_0(\mathbf{x}) \propto 1/\Omega$.

Doob (1942), Gillespie (1977)

Doob-Gillespie vs deterministic approximation



Large initial condition I_0

Herd immunity threshold

- In the deterministic SIR model, an epidemic can occur if $S_0 > \Omega\nu/\beta$.
- If $I_0 = O(1)$, we have $S_0 \simeq \Omega$ and the condition above becomes $\beta/\nu > 1$.
- We note $R_0 = \beta/\nu$ the basic reproduction number: *Average number of secondary cases generated by a primary case in a fully susceptible population.*
- If $R_0 > 1$, a proportion $V_c = 1 - 1/R_0$ of the population needs to be vaccinated to prevent an epidemic.
- The R_0 of influenza is between 1.5 and 2, so between 30% and 50% of the population needs to be vaccinated.
- For childhood diseases (e.g. measles), R_0 is above 10, so more than 90% of the population needs to be vaccinated.

Herd immunity threshold

- In the deterministic SIR model, an epidemic can occur if $S_0 > \Omega\nu/\beta$.
- If $I_0 = O(1)$, we have $S_0 \simeq \Omega$ and the condition above becomes $\beta/\nu > 1$.
- We note $R_0 = \beta/\nu$ the basic reproduction number: *Average number of secondary cases generated by a primary case in a fully susceptible population.*
- If $R_0 > 1$, a proportion $V_c = 1 - 1/R_0$ of the population needs to be vaccinated to prevent an epidemic.
- The R_0 of influenza is between 1.5 and 2, so between 30% and 50% of the population needs to be vaccinated.
- For childhood diseases (e.g. measles), R_0 is above 10, so more than 90% of the population needs to be vaccinated.

Herd immunity threshold

- In the deterministic SIR model, an epidemic can occur if $S_0 > \Omega\nu/\beta$.
- If $I_0 = O(1)$, we have $S_0 \simeq \Omega$ and the condition above becomes $\beta/\nu > 1$.
- We note $R_0 = \beta/\nu$ the basic reproduction number: *Average number of secondary cases generated by a primary case in a fully susceptible population.*
- If $R_0 > 1$, a proportion $V_c = 1 - 1/R_0$ of the population needs to be vaccinated to prevent an epidemic.
- The R_0 of influenza is between 1.5 and 2, so between 30% and 50% of the population needs to be vaccinated.
- For childhood diseases (e.g. measles), R_0 is above 10, so more than 90% of the population needs to be vaccinated.

Initial extinction risk

- In the stochastic SIR model, a major epidemic can still be avoided even if $R_0 > 1$.
- The initial phase of the epidemic can be described by a birth-death process: individuals live for a random duration (infectious period) during which they give birth (infect) according to a Poisson process with intensity β.
- The probability P_{ext} that the birth-death process stops after a *finite* number of generations is q^{I_0}, where q is the unique root in $[0, 1[$ of $s = f(s)$, with

$$f(s) = \int_0^\infty e^{-\beta t(1-s)} g_I(t) dt, \quad |s| \leq 1, \tag{10}$$

  where g_I is the pdf of the infectious period.
- For DDJMPs, we can set $g_I(t; k, \nu) = \Gamma(k, k\nu), k \in \mathbb{N}^*$:

Becker (1977), Andersson & Britton (2000)

Initial extinction risk

- In the stochastic SIR model, a major epidemic can still be avoided even if $R_0 > 1$.
- The initial phase of the epidemic can be described by a birth-death process: individuals live for a random duration (infectious period) during which they give birth (infect) according to a Poisson process with intensity β.
- The probability P_{ext} that the birth-death process stops after a *finite* number of generations is q^{I_0}, where q is the unique root in $[0, 1[$ of $s = f(s)$, with

$$f(s) = \int_0^\infty e^{-\beta t(1-s)} g_I(t) dt, \quad |s| \leq 1, \tag{10}$$

  where g_I is the pdf of the infectious period.
- For DDJMPs, we can set $g_I(t; k, \nu) = \Gamma(k, k\nu), k \in \mathbb{N}^*$:

Becker (1977), Andersson & Britton (2000)

Initial extinction risk

- In the stochastic SIR model, a major epidemic can still be avoided even if $R_0 > 1$.
- The initial phase of the epidemic can be described by a birth-death process: individuals live for a random duration (infectious period) during which they give birth (infect) according to a Poisson process with intensity β.
- The probability P_{ext} that the birth-death process stops after a *finite* number of generations is q^{I_0}, where q is the unique root in $[0, 1[$ of $s = f(s)$, with

$$f(s) = \int_0^\infty e^{-\beta t(1-s)} g_I(t) dt, \quad |s| \leq 1, \tag{10}$$

where g_I is the pdf of the infectious period.

- For DDJMPs, we can set $g_I(t; k, \nu) = \Gamma(k, k\nu), k \in \mathbb{N}^*$:

Becker (1977), Andersson & Britton (2000)

Initial extinction risk

- In the stochastic SIR model, a major epidemic can still be avoided even if $R_0 > 1$.
- The initial phase of the epidemic can be described by a birth-death process: individuals live for a random duration (infectious period) during which they give birth (infect) according to a Poisson process with intensity β.
- The probability P_{ext} that the birth-death process stops after a *finite* number of generations is q^{I_0}, where q is the unique root in $[0, 1[$ of $s = f(s)$, with

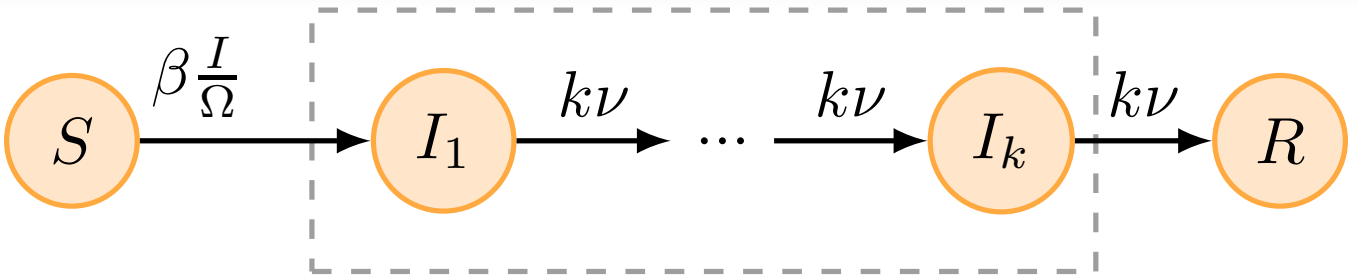
$$f(s) = \int_0^\infty e^{-\beta t(1-s)} g_I(t) dt, \quad |s| \leq 1, \tag{10}$$

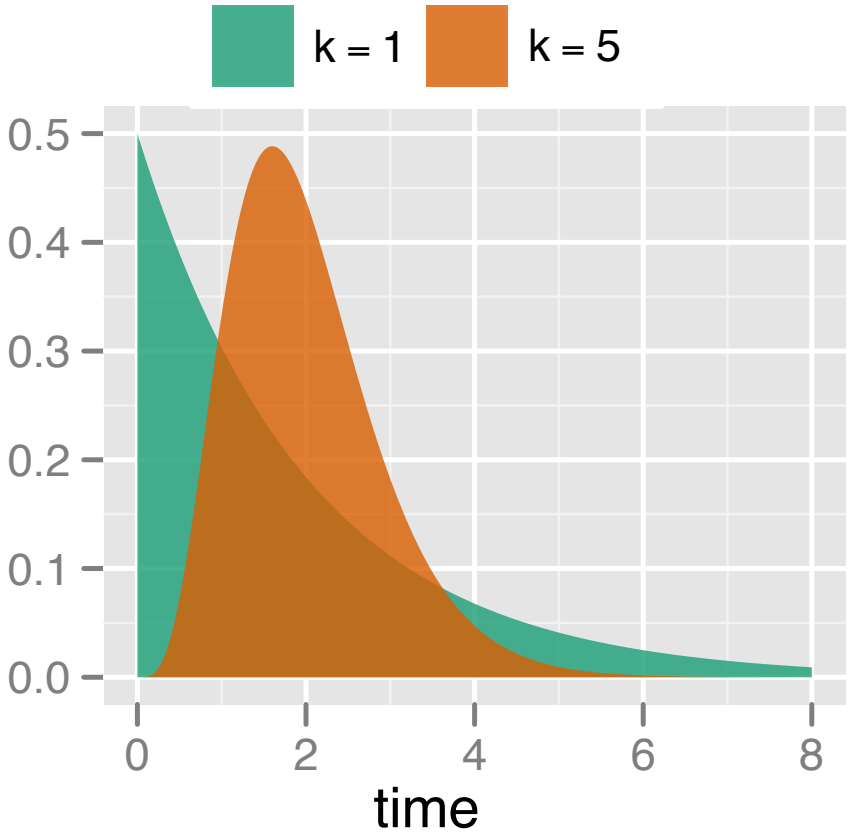
where g_I is the pdf of the infectious period.

- For DDJMPs, we can set $g_I(t; k, \nu) = \Gamma(k, k\nu), k \in \mathbb{N}^*$:

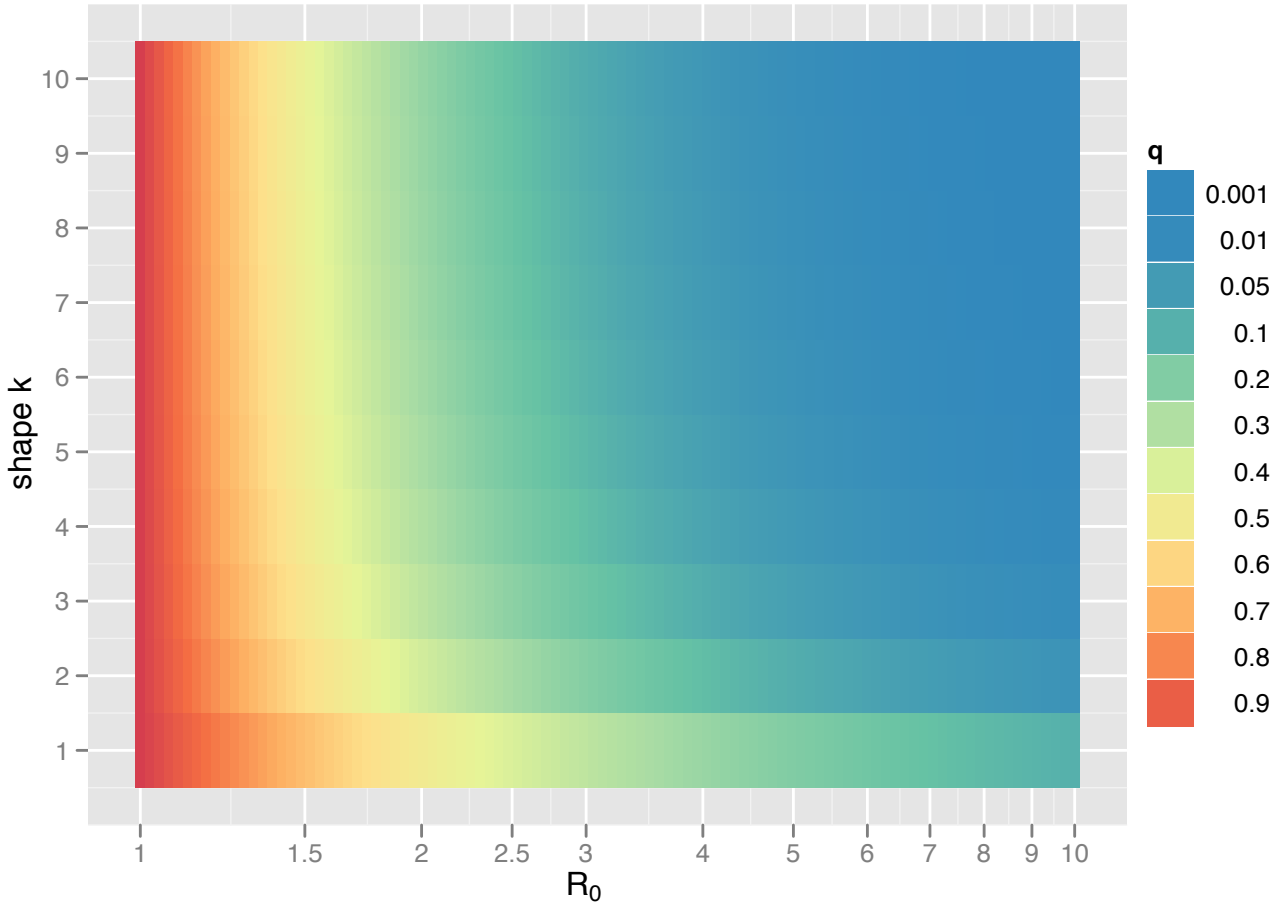
Becker (1977), Andersson & Britton (2000)

Erlang distribution





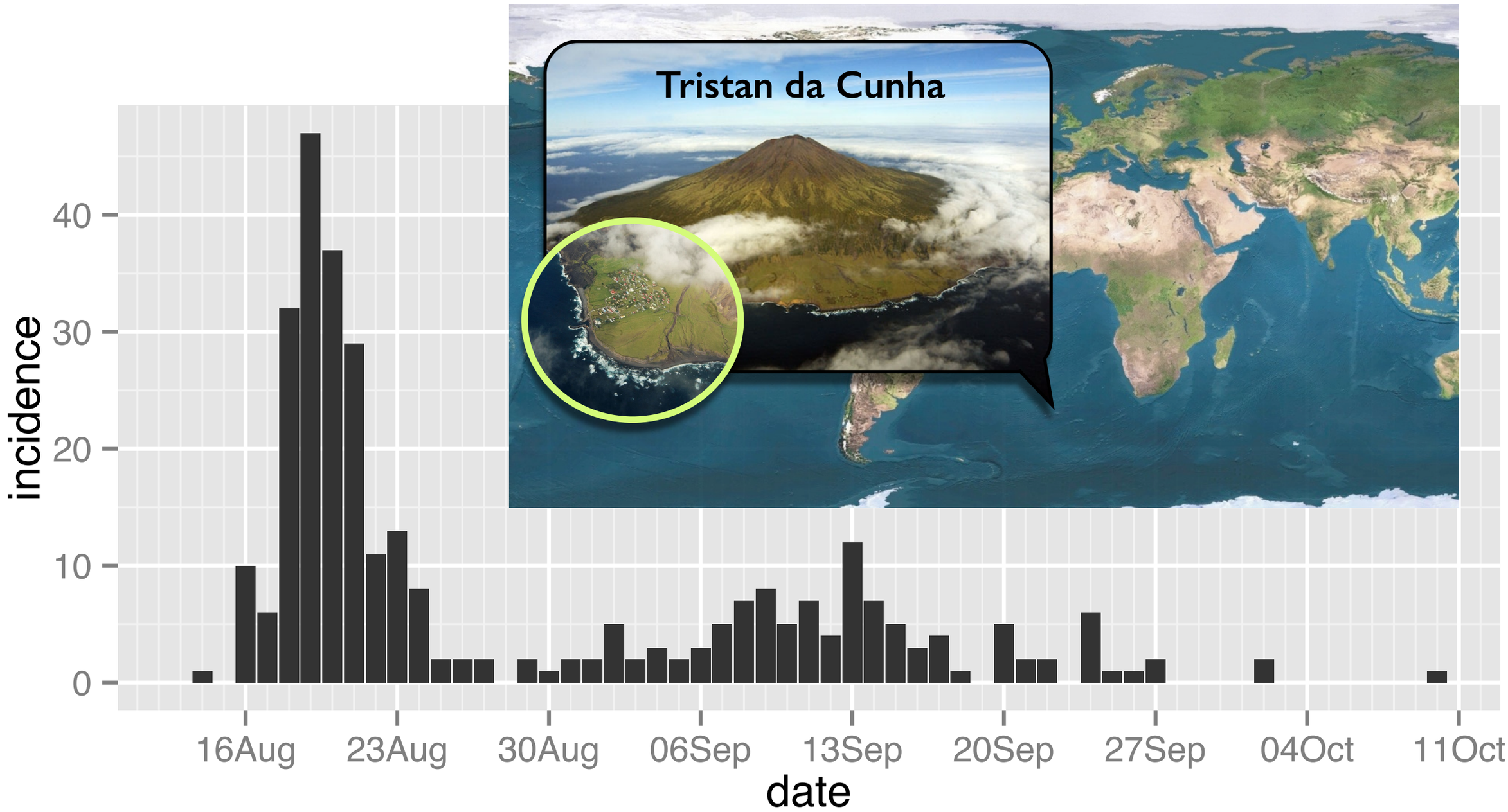
Initial extinction risk



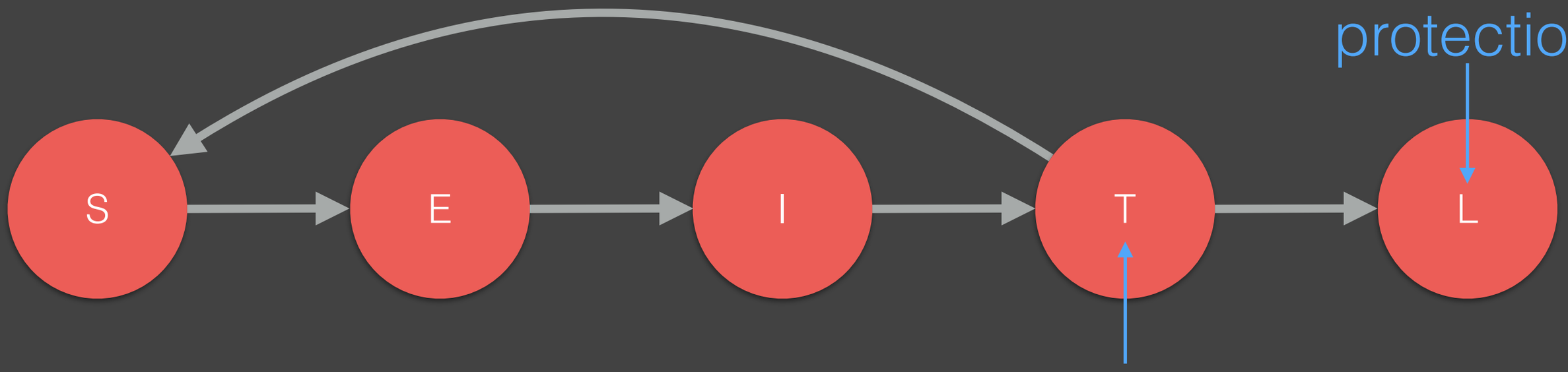
Pathogens with more variable infectious periods have a higher risk of initial extinction.

Inference for small population outbreaks

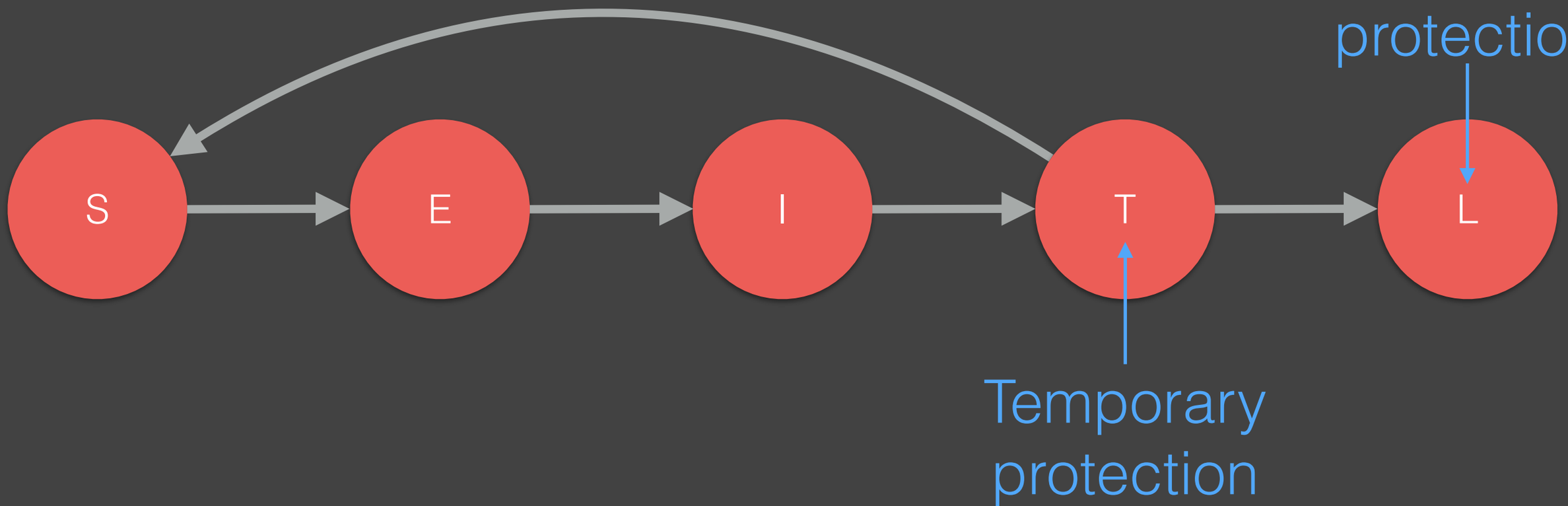
284 ind - 32% reinfected



One possible model…



One possible model…



Already implemented as a fitmodel!