Studying behavior before absorption in stochastic epidemic models: Quasi-stationary and Ratio of Expectations distributions

MJ Lopez-Herrero
Complutense University of Madrid

Lherrero@ucm.es

MPDE, Aix-Marseille, 5th - 9th September 2016

Talk Schedule

1. Introduction
2. Quasi-stationarity
3. RE-distribution
4. Application to epidemic stochastic models
5. Conclusions and References

1. Introduction

We deal with continuous time Markov chains $\{X(t);\, t \geq 0\}$ (CTMC) with state space

$$S = S_A \cup S_T,$$

where S_T is a set of transient states and S_A is a set of absorbing states (e.g. $S_A = \{0\}$).

The stationary distribution is degenerate so the fundamental problem is:

There exist a need for probabilistic measures of the system behavior before absorption. Analogues of the stationary distribution of the irreducible case should be considered and compared.

In the spirit of the seminal work by Darroch and Seneta (1967), we will consider two possibilities:

- Quasi-stationary distribution.
- Ratio of expectations (RE) distribution.

A comparison between both distributions is justified only if the convergence to quasi-stationary regime is relatively fast.

2. Quasi-stationarity

The starting point is the conditional probability

$$u_i(t) = P\{X(t) = i \mid T > t\},\ i \in S_T,$$

where $T = \sup\{t \geq 0 \mid X(t) \in S_T\}$ denotes the absorption time.

Definition 1. Suppose that the chain starts with the initial distribution $a_i = P\{X(0) = i\}$, $i \in S_T$. If there exists a starting distribution $a_i = u_i$ such that

$$P\{X(t) = i \mid T > t\} = u_i,\ i \in S_T,$$

for all $t \geq 0$, then $\mathbf{u} = \{u_i;\ i \in S_T\}$ is called a quasi-stationary distribution.

There also exists a limiting interpretation which states that

$$\lim_{t \to \infty} P\{X(t) = i \mid T > t\} = u_i, \ i \in S_T,$$

independently of the initial distribution $\{a_i;\, i \in S_T\}$.

The above limiting result shows that the quasi-stationary distribution is a good measure of the system dynamics before absorption, but restricting only to those realizations in which the time to absorption is sufficiently large.

The existence and computation of the quasi-stationary distribution becomes difficult when S_T is infinite.

A. *S_T is finite and irreducible*

There exists a unique quasi-stationary distribution, but an analytical (explicit form) solution only exists in a few special cases.

Example 1. If $\{X(t);\, t \geq 0\}$ is a birth and death process with $S_A = \{0\}$, then

$$(1 - \delta_{iN})\mu_{i+1}u_{i+1} - (\lambda_i + \mu_i)u_i + \lambda_{i-1}u_{i-1} = -\mu_1 u_1 u_i,\ 1 \leq i \leq N,$$

where $S = \{0, ..., N\}$ and $\lambda_0 = 0$.

The above non-linear equation also holds when $S = \mathbb{N}$.

Computation, approximations and recursive methods

• The power method provides an iterative procedure for computing $\mathbf{u}$ and $-\alpha$ (associate eigenvalue).

• Several approximations and recursive methods are available for the birth and death process on $S = \{0, ..., N\}$, with $S_A = \{0\}$. A first approximation uses a birth and death process, $X^{(0)}(t)$, with the same rates except $\mu_1^{(0)} = 0$.

• A second approximation is based on the birth and death process, $X^{(1)}(t)$, with shifted death rates $\mu_i^{(1)} = \mu_{i-1}$, $1 \leq i \leq N$.

• A recursive scheme is based on the following formula:

$$u_i = u_1 \pi_i \sum_{k=1}^{i} \frac{1-(1-\delta_{k1}) \sum_{j=1}^{k-1} u_j}{\tau_k}, \; 2 \leq i \leq N.$$

B. S_T *is finite but reducible* (van Doorn and Pollett, 2008)

Suppose that S_T consists of L communicating classes S_k, for $1 \leq k \leq L$. A partial order on $\{S_k; 1 \leq k \leq L\}$ is defined by writing $S_i \prec S_j$ when class S_i is accessible from S_j. Let $-\alpha_k$ be the (negative) eigenvalue with maximal real part of the sub-generator Q_k corresponding to the states in S_k. Then, the eigenvalue of Q_{S_T} with maximal real part is obtained as $-\alpha$, where $\alpha = \min_{1 \leq k \leq L} \alpha_k$. We also define $I(\alpha) = \{k : \alpha_k = \alpha\}$ and $a(\alpha) = \min I(\alpha)$.

Theorem 1. If $-\alpha$ has a geometric multiplicity one, then the Markov chain has a unique quasi-stationary distribution $\{u_i; i \in S_T\}$ from which $S_{a(\alpha)}$ is accessible. The jth component of $\{u_i; i \in S_T\}$ is positive if and only if state j is accessible from $S_{a(\alpha)}$. A simple necessary and sufficient condition for establishing that $-\alpha$ has geometric multiplicity one is that $\{S_k; k \in I(\alpha)\}$ is linearly ordered, that is, $S_i \prec S_j \iff i \leq j$, for all $i, j \in I(\alpha)$.

3. RE-distribution

Let T_j be the time that the CTMC spends in state $j \in S_T$ before absorption.

Definition 2. Given that $X(0) = i \in S_T$, we define the RE-distribution, $\mathbf{P}_i = (P_i(j))$, as follows

$$P_i(j) = \frac{E_i[T_j]}{E_i[T]},\ i, j \in S_T.$$

The ratio of means distribution (Darroch and Seneta, 1967) corresponds with the unconditional version:

$$P(j) = \frac{\sum_{i \in S_T} a_i E_i[T_j]}{\sum_{i \in S_T} a_i E_i[T]},\ j \in S_T.$$

Remarks and computation

• The RE-distribution always exists provided that the expected time to absorption $E_i[T] < \infty$.

• The RE-distribution assigns positive probability to all state j accessible from the initial state i.

• Let us construct the ideal replicated model obtained by assuming that at each extinction the biological model restarts in the same initial state $i \in S_T$. The stationary distribution of this replicated (regenerative) model amounts to the RE-distribution $\mathbf{P}_i$.

For a fixed $j \in S_T$, a first-step argument yields

$$E_i\left[T_j\right] = \frac{\delta_{ij}}{q_i} + \sum_{\substack{k \in S_T \\ k \neq i}} \frac{q_{ik}}{q_i} E_k\left[T_j\right], \; i \in S_T.$$

4. Application to the SIS stochastic model

- Closed population model of N individuals.
  - Classified either as susceptible or infective individual.

- Susceptible can be infected, then they recover and return to the susceptible pool.

- Evolution of the epidemic:
  - Birth and death process $\{I(t);\, t \geq 0\}$.
  - $I(t)$: number of infective individuals at time t.
  - $S = \{0, 1, ..., N\}$ (0 is an absorbing state).

- Classical SIS rates
  - Infection rate $\lambda_i = \frac{\beta}{N} i(N-i)$.
  - Recovery rate $\mu_i = \gamma i$.

- $R_0 = \frac{\beta}{\gamma}$ denotes the transmission factor.

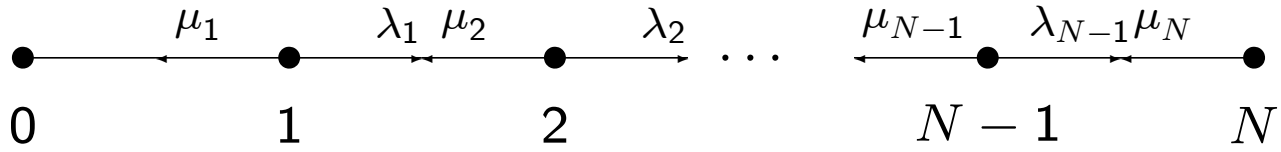


Figure 1. States and transitions of the birth and death model

In the case of a birth and death process, we have

$$E_0\left[T_j\right] = 0,$$
$$(\lambda_i + \mu_i)E_i\left[T_j\right] = \mu_i E_{i-1}\left[T_j\right] + \lambda_i E_{i+1}\left[T_j\right],\ i \neq j,\ 1 \leq i \leq N,$$
$$(\lambda_j + \mu_j)E_j\left[T_j\right] = \mu_j E_{j-1}\left[T_j\right] + \lambda_j E_{j+1}\left[T_j\right] + 1.$$

Theorem 2. For the SIS epidemic model, the RE-distribution reduces to

$$P_i(j) = \frac{\frac{1}{j}\sum_{k=1}^{\min(i,j)}\left(\frac{R_0}{N}\right)^{j-k}\frac{(N-k)!}{(N-j)!}}{\sum_{j=1}^{N}\frac{1}{j}\sum_{k=1}^{\min(i,j)}\left(\frac{R_0}{N}\right)^{j-k}\frac{(N-k)!}{(N-j)!}},\ 1 \leq j \leq N.$$

.

Stochastic ordering relationships

- For a birth and death process with $S = \{0, ..., N\}$ and $S_A = \{0\}$:

$$\mathbf{P}_1 = \mathbf{p}^{(0)},$$

$$\mathbf{p}^{(0)} \leq_{st} \mathbf{u},$$

$$\mathbf{P}_1 \leq_{st} \mathbf{u} \leq_{st} \mathbf{P}_N,$$

$$\mathbf{P}_i \leq_{st} \mathbf{P}_{i'},\ 1 \leq i \leq i' \leq N,$$

where $X \leq_{st} Y \Leftrightarrow F_X(x) \geq F_Y(x)$, for all $x \in \mathbb{R}$.

- For the SIS model:

$$\mathbf{u} \leq_{st} \mathbf{p}^{(1)},$$

$\mathbf{P}_N \leq_{st} \mathbf{p}^{(1)}$, for N fixed and R_0 sufficiently large,

$\mathbf{p}^{(1)} \leq_{st} \mathbf{P}_N$, for N fixed and R_0 sufficiently small.

Numerical example

$N = 100$, $\gamma = 1$ and several choices of $R_0 = \beta$.

$R_u < 1 \Rightarrow$ to use $\mathbf{u}$ is meaningful.

$\widehat{\mathbf{P}}$ and $\widetilde{\mathbf{P}}$ are mixtures of the RE-distributions.

$\left|\mathbf{p}^{(1)} - \mathbf{u}\right| = \max\limits_{1 \le j \le N} \left|p_j^{(1)} - u_j\right|$ (maximum pointwise distance).

R_0	0.5	0.9	1.0	1.3	1.5	2.0
$\left\|\mathbf{p}^{(1)} - \mathbf{u}\right\|$	0.00440	0.05091	0.08746	0.16147	0.11150	0.05890
$\|\mathbf{P}_1 - \mathbf{u}\|$	0.21231	0.30368	0.30295	0.07878	0.00357	4.267×10^{-8}
$\|\mathbf{P}_N - \mathbf{u}\|$	0.58829	0.37515	0.28886	0.03260	0.00107	1.038×10^{-8}
$\left\|\widehat{\mathbf{P}} - \mathbf{u}\right\|$	0.05134	0.09190	0.06356	0.00546	1.485×10^{-4}	9.214×10^{-10}
$\left\|\widetilde{\mathbf{P}} - \mathbf{u}\right\|$	0.05508	0.08627	0.06453	0.00232	5.591×10^{-6}	9.020×10^{-16}

Table 1. Distributions distances with respect to $\mathbf{u}$

Application to the SIR stochastic model

- Classical SIR rates: $\lambda_{ij} = \frac{\beta}{N} ij$ (infection rate) and $\mu_i = \gamma i$ (recovery rate).

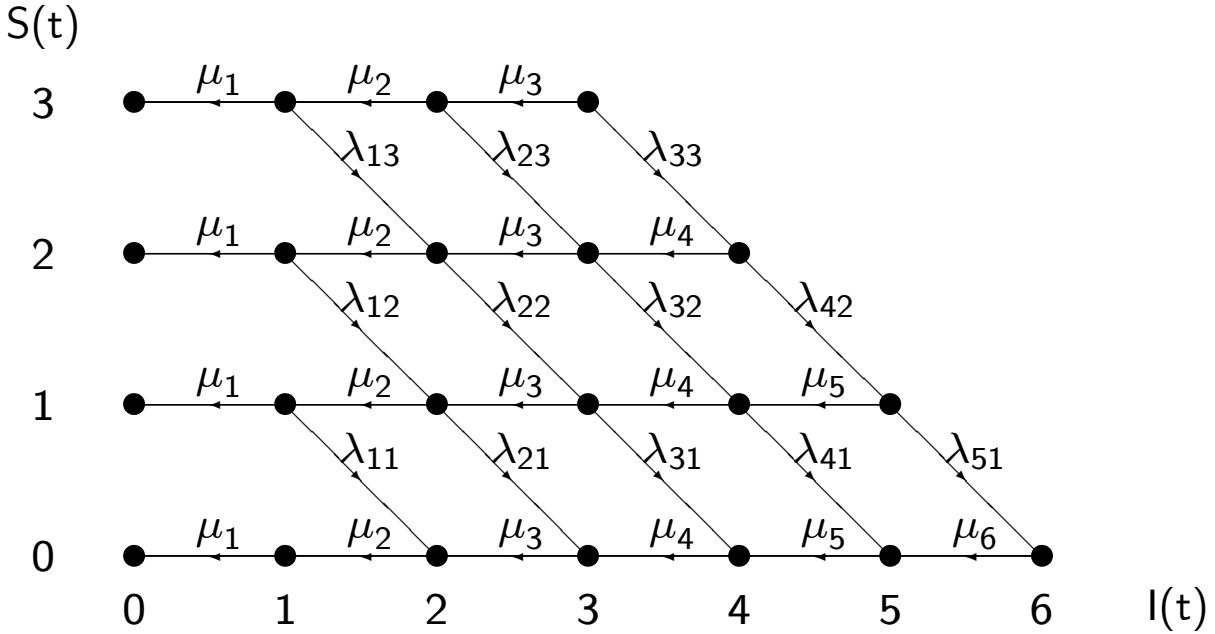


Figure 2. States and transitions of the SIR epidemic model

The quasi-stationary probabilities are

$$u_{(i,j)} = \delta_{(1,0)(i,j)},\ 0 \leq j \leq n,\ 1 \leq i \leq m+n-j,$$

because $\min_{(i,j)\in S_T}(\lambda_{ij}+\mu_i) = \lambda_{10}+\mu_1 = \gamma$, for $\mu_i = \gamma i$, and $1 \leq i \leq m+n$.

The RE-distribution of the SIR epidemic model is given by

$$P_{(m,n)}(i,j) = \frac{\frac{A_{ij}}{\lambda_{ij}+\mu_i}}{\sum_{j=0}^{n}\sum_{i=1}^{m+n-j}\frac{A_{ij}}{\lambda_{ij}+\mu_i}},\ (i,j) \in S_T,$$

where A_{ij} is the probability of reaching the state $(i,j) \in S$ starting from (m,n) before the extinction occurs.

Application to the SEIR stochastic model

- Classical SEIR rates: $\lambda_{ij} = \frac{\beta}{N} ij$ (infection rate), $\sigma_e = \sigma e$ (σ rate at wich exposed an individual becomes infective) and $\mu_i = \gamma i$ (recovery rate).

- Quasi-stationary distribution is almost degenerate: The quasi-stationary distribution assigns all its probability mass to one or two states.
  To state (0,0,1), i.e., 1 infective individuals and N-1 recovered, for $\gamma \leq \sigma$.
  To states (0,1,0) and (0,0,1); i.e, 1 exposed or infective individual and N-1 recovered, for $\gamma > \sigma$.

The RE-distribution of the SEIR $\mathbf{p}^{RE} = (p_{sei}^{RE} : (s,e,i) \in S_T)$ is given by

$$p_{sei}^{RE} = \frac{\frac{\theta_{sei}}{\beta_{si}+\sigma_e+\gamma_i}}{\sum_{(s',e',i')\in S_T} \frac{\theta_{s'e'i'}}{\beta_{s'i'}+\sigma_{e'}+\gamma_{i'}}}, \ (s,e,i) \in S_T,$$

where θ_{sei} is the probability of having a finite first passage time to state (s,e,i), for $(s,e,i) \in S_T$,.

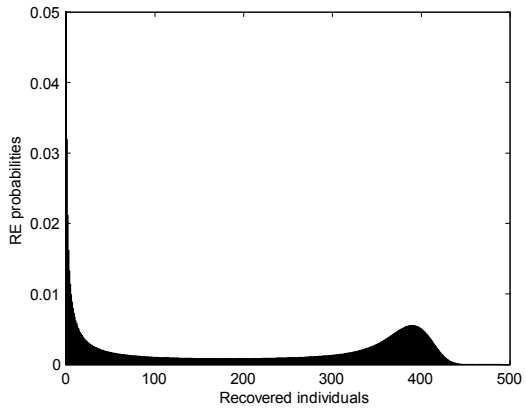
The expected values corresponding to the RE-distributions of the number of individuals E_{RE} and I_{RE} in the E and I classes quantify intuitively the mean number of exposed and infected individuals during an outbreak of the epidemic. They are related to the expected values of the final size $R(\infty)$ and the extinction time $L(\infty)$ through the relationships

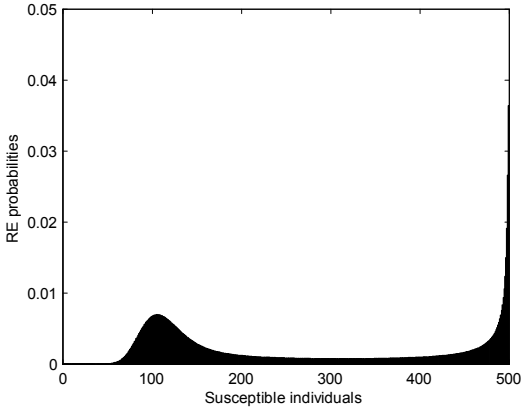
$$\begin{aligned} E_{(s_0,e_0,i_0)}[E_{RE}] &= \frac{E_{(s_0,e_0,i_0)}[R(\infty)] - i_0}{E_{(s_0,e_0,i_0)}[L(\infty)]} \times \frac{1}{\sigma}, \\ E_{(s_0,e_0,i_0)}[I_{RE}] &= \frac{E_{(s_0,e_0,i_0)}[R(\infty)]}{E_{(s_0,e_0,i_0)}[L(\infty)]} \times \frac{1}{\gamma}. \end{aligned}$$

Numerical example

$N = 500$, $\gamma = 1, \beta = 2.0$. Starting from a single infective individual: initial state (499,0,1).

- Mass functions for the R (recovered) and S (susceptible) classes, $\sigma = 5.0$

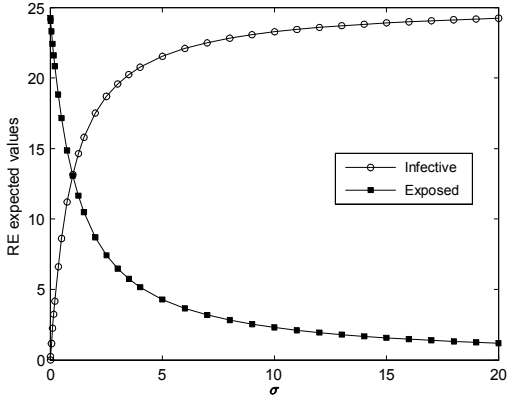




Numerical example

$N = 500$, $\gamma = 1, \beta = 2.0$. Starting from a single infective individual: initial state (499,0,1).

- Mean values of the RE distributions for the *E* and *I* classes versus σ



5. Conclusions and References

The talk has a two-fold objective: i) to consider the quasi-stationarity and the ratio of expectations as two conceptually different approaches for measuring the behavior of a biological system before reaching the absorbing states, and ii) to evaluate the possibility of using the RE-distribution as an approximation to the quasi-stationary distribution provided that the quasi-stationary regime has already been reached.

• The quasi-stationary distribution gives an excellent measure of the long-term behavior of the system. Due to the non-linear structure of the quasi-stationary equations, it is usually impossible to obtain explicit expressions. However, there exists a number of helpful results including recursive methods, approximations and asymptotic analysis.

• The RE-distribution gives an alternative to measure the system dynamics before absorption, despite of how long the absorption time is. Since the RE-distribution is governed by linear equations, it can typically be evaluated more simply. The RE-distribution assigns positive mass to all transient states. The main problem concerns the difficulties for managing info about the initial state.

Main references

Artalejo, J.R., Lopez-Herrero, M.J., 2010. Quasi-stationary and ratio of expectations distributions: A comparative study. J. Theor. Biol. 266, 264-274.

Artalejo, J.R., Economou, A., Lopez-Herrero, M.J., 2015. The stochastic SEIR model before extinction: Computational approaches. Appl. Math. Comput. 265, 1026-1043.

van Doorn, E.A., Pollett, P.K., 2013. Quasi-stationary distributions for discrete-state models. Eur. J. Oper. Res. 230, 1-14.