ELSEVIER

Mathematical Biosciences 156 (1999) 21–40

Mathematical Biosciences
an international journal

On the quasi-stationary distribution of the stochastic logistic epidemic

Ingemar Nåsell *

The Royal Institute of Technology, Department of Mathematics, 10044 Stockholm, Sweden

Received 14 January 1998; accepted 26 May 1998

Abstract

An approximation is derived for the quasi-stationary distribution of the stochastic logistic epidemic in the intricate case where the transmission factor R_0 lies in the transition region near the deterministic threshold value 1. An approximation for the expected time to extinction from quasi-stationarity in the same parameter region is also given. © 1999 Elsevier Science Inc. All rights reserved.

Mathematics subject classification: 60J27; 92D30

Keywords: Quasi-stationary distribution; Time to extinction; Asymptotic approximation; Transition region; SIS model

1. Introduction

We analyze the stochastic logistic epidemic, also known as the stochastic SIS model. Here, the letters S and I stand for susceptible and infected, respectively, and the model designation SIS indicates the successive states of an individual. A recovered individual is again susceptible. The SIS model is used for endemic infections that do not confer immunity.

The epidemiologically most important features of this model are the quasi-stationary distribution and the time to extinction. Neither of them can be

* E-mail: ingemar@math.kth.se

PII: S 0 0 2 5 - 5 5 6 4 (9 8) 1 0 0 5 9 - 7

determined explicitly, so the analysis is directed at finding useful approximations. Deterministic versions of the model cannot be used for this analysis.

Qualitative results are important for all models in mathematical epidemiology, with thresholds of deterministic models as the outstanding example. For many models one can introduce a transmission factor that traditionally is denoted by R_0 and referred to as the basic reproduction ratio. The deterministic model has a threshold at $R_0 = 1$. For an endemic infection the deterministic model predicts that the infection will die out of its own if $R_0 < 1$, while a stationary endemic infection level is possible if $R_0 > 1$.

It is natural to search for qualitative features also in stochastic models. Such features would then, in some sense, generalize the threshold concept for deterministic models, as discussed by Nåsell [1]. We note that the population size N is a parameter in the stochastic model but not in its deterministic counterpart. The reason for this is that the deterministic version of the stochastic model can be derived as an approximation of the latter under the assumption that N approaches infinity.

It is shown by Nåsell [2] for the stochastic SIS model that both the quasi-stationary distribution and the expected time to extinction from quasi-stationarity have three qualitatively different behaviours as a function of the two parameters N and R_0. The corresponding regions of parameter space are defined mathematically in terms of conditions for validity of asymptotic approximations as $N \to \infty$. Two of the regions are natural counterparts to the two cases in the deterministic case. In one of them $R_0 > 1$ is fixed and in the other one $R_0 < 1$ is fixed as $N \to \infty$. We describe these regions by saying that R_0 is distinctly larger than 1 or distinctly smaller than 1. The third region is a transition region between these two near the deterministic threshold value $R_0 = 1$. It is found by the rescaling $R_0 = 1 + \rho/\sqrt{N}$ that makes R_0 a function of N in such a way that R_0 approaches 1 as N approaches infinity. The third region is defined by the requirement that ρ is fixed as $N \to \infty$. This region becomes empty in the limit as $N \to \infty$. This explains why it is absent from the deterministic model.

The stochastic SIS model has been treated previously by a number of authors; see [3–10,2]. Among these authors, Nåsell is the only one to deal with the transition region.

The behaviour of the SIS model in the transition region is more intricate than the behaviour when R_0 is distinctly different from 1. The approximations of the quasi-stationary distribution and of the time to extinction in the transition region given by Nåsell [2] are rather crude. One purpose of the present paper is to present improvements of these results. A second purpose is to summarize previous and new approximations in each of the three parameter regions.

It is shown by Nåsell [11] that the transition region is relatively narrow for the stochastic SIS model. This means that the results in the transition region are of minor interest epidemiologically. However, there are other endemic infections where the transition region is very wide and therefore of considerable epidemiological interest. One example is the endemic SIR model for infections with immunity where one also accounts for demographic influences as in Ref. [11]. One can hope that the insight into the behaviour of the quasi-stationary distribution and of the time to extinction in the transition region that the present paper gives for the SIS model will help in analyzing more difficult problems like those posed by the endemic SIR model.

The SIS model is introduced in Section 2 as a finite-state birth-and-death process with an absorbing state at the origin. Conditions satisfied by its quasi-stationary distribution are given in two different forms, and an important relation is given between the quasi-stationary distribution and the time to extinction from quasi-stationarity. Section 3 deals with two birth-and-death processes that are approximations of the SIS model. These two processes lack absorbing states and have stationary distributions (called $p^{(0)}$ and $p^{(1)}$) that are given in explicit form. Earlier work involving these two distributions was summarized by Kryscio and Lefèvre [8]. The two stationary distributions are useful approximations of the quasi-stationary distribution (called q), one in each of the two parameter regions where R_0 is distinctly different from 1. Approximations of the two stationary distributions are needed for the derivation of an approximation of the quasi-stationary distribution. A summary of previously available and newly derived approximations of the two stationary distributions $p^{(0)}$ and $p^{(1)}$, as well as of the quasi-stationary distribution q, are given in Section 4. The three distributions $p^{(0)}$, $p^{(1)}$ and q depend in important ways on three sums denoted respectively by $S^{(0)}$, $S^{(1)}$ and S. Approximations of these three sums are summarized in Section 5. Approximations of the expected time to extinction from quasi-stationarity are summarized in Section 6. Derivations of the new approximations of the distribution $p^{(0)}$ and of its associated sum $S^{(0)}$ in the transition region are given in Section 7. Similarly, derivations of the new approximations of the quasi-stationary distribution q and of its associated sum S in the transition region are contained in Section 8.

2. Model formulation

The model that we study was first introduced in a deterministic version by Ross [12]. The stochastic version was formulated by Weiss and Dishon [3], who apparently were unaware of the work by Ross. The stochastic model is a finite-state univariate birth-and-death process with an absorbing state at the origin, We use one state variable, namely the number of infected individuals $I(t)$ at time t. It takes values in the state space $\{0, 1, \dots, N\}$, where N is the total constant number of individuals in the population. The state probabilities are denoted by $p_n(t) = P(I(t) = n)$. They depend on the initial distribution $\{p_n(0)\}$.

The model accounts for two basic events, i.e. infection of a susceptible individual and recovery of an infected individual. Three parameters are used, namely the population size N, the contact rate β, and the recovery rate γ. We introduce $R_0 = \beta/\gamma$. It is straightforward to show that the deterministic version of our model has a threshold at $R_0 = 1$.

The hypotheses of the model are set down by specifying the transition rates as follows:

Event	Transition	Transition rate
Infection of susceptible	$n \to n+1$	$\lambda_n = n\beta(1 - n/N)$
Recovery of infected	$n \to n-1$	$\mu_n = n\gamma$

The state 0 is absorbing and the states $n, 1 \leqslant n \leqslant N$, are transient. The stationary distribution is the trivial one that puts probability 1 at the origin. It is thus non-informative. Instead of the stationary distribution, we study the quasi-stationary one. It is supported on the set of transient states.

The Kolmogorov forward equations for the state probabilities $p_n(t)$ can be written in the form

$$p' = pA, \tag{2.1}$$

where $p(t)$ is a row vector with components $p_n(t), n = 0, 1, \ldots, N$, and A is a square matrix of order $N+1$, defined by

$$A = \begin{pmatrix} -\kappa_0 & \lambda_0 & 0 & \ldots & 0 \\ \mu_1 & -\kappa_1 & \lambda_1 & \ldots & 0 \\ 0 & \mu_2 & -\kappa_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & -\kappa_N \end{pmatrix},$$

with $\kappa_n = \lambda_n + \mu_n$.

The quasi-stationary distribution is a stationary conditional distribution. To define it we study the process $\{I(t)\}$ conditioned on non-extinction at time t. We use $q_n(t)$ to denote the corresponding state probabilities and $q(t)$ to denote the row vector of values of $q_n(t)$, $n = 1, 2, \ldots, N$. Note that $q(t)$ depends on the initial distribution $q(0)$. We derive a system of differential equations for $q(t)$ that is closely related to the system in Eq. (2.1). Let $p_Q(t)$ denote the vector containing all components of $p(t)$ except the first one, and assume that the initial state m is positive. We then have $q(t) = p_Q(t)/(1 - p_0(t))$. By differentiating this relation and using the equation $p_0' = \mu_1 p_1$ we get

$$q' = \frac{p_Q'}{1 - p_0} + \mu_1 q_1 \frac{p_Q}{1 - p_0}. \tag{2.2}$$

This leads to the system of equations

$$q' = qA_Q + \mu_1 q_1 q,$$

where A_Q is the N-by-N matrix formed from A by deleting the first row and the first column. The quasi-stationary distribution is the stationary solution of this system of differential equations. Thus, it satisfies the relation

$$qA_Q = -\mu_1 q_1 q. \tag{2.3}$$

This shows that the quasi-stationary distribution q is a left eigenvector of the matrix A_Q corresponding to the eigenvalue $-\mu_1 q_1$. This fact can be used for numerical evaluation of the quasi-stationary distribution. It can be shown that the eigenvalue $-\mu_1 q_1 = -\gamma q_1$ is the largest eigenvalue of the matrix A_Q.

It is shown in Ref. [2] that the quasi-stationary distribution q_j satisfies the relation

$$q_j = \gamma(j)\alpha(j)R_0^{j-1}q_1, \quad j = 1, 2, \ldots, N, \tag{2.4}$$

where

$$\gamma(j) = \frac{1}{j}\sum_{k=1}^{j}\delta(k), \quad \delta(k) = \frac{1 - \sum_{l=1}^{k-1} q_l}{\alpha(k)\, R_0^{k-1}}, \tag{2.5}$$

$$\alpha(j) = \frac{N!}{(N-j)!N^j} \tag{2.6}$$

and

$$q_1 = \frac{1}{S} \quad \text{with} \quad S = \sum_{j=1}^{N}\gamma(j)\alpha(j)R_0^{j-1}. \tag{2.7}$$

Expressions (2.4)–(2.7) do not give the quasi-stationary distribution in explicit form since $\gamma(j)$ in Eq. (2.5) depends on the q_l-values. The expressions can, however, be used to determine the quasi-stationary distribution numerically by iteration. The q_j-values at any stage of the iteration are found by solving Eq. (2.4) and Eq. (2.7) with the γ-values found from Eq. (2.5) by using the q_l-values of the previous stage of iteration. The same approach forms the basis for our analytic approximation of the quasi-stationary distribution in the transition region in Section 8.

The time to extinction τ is a random variable that depends on the initial distribution of infected individuals. If it is known at some time that infection exists in a community and that it has existed for a long time, then we can conclude that the distribution of the number of infected individuals $I(t)$ is well described by the quasi-stationary distribution. Therefore it is especially interesting to study the time to extinction starting from the quasi-stationary distribution. We denote the time to extinction from this initial distribution by τ_Q. It turns out that the distribution of this random variable is pleasingly simple from a mathematical standpoint. Indeed, the distribution of τ_Q is then exponential, and its expected value is equal to

$$\mathrm{E}\tau_Q = \frac{1}{\gamma q_1} = \frac{S}{\gamma}. \tag{2.8}$$

Thus, the expected time to extinction from quasi-stationarity is inversely proportional to the probability q_1. It can therefore be determined if the quasi-stationary distribution is known. The fact that it also is inversely proportional to the recovery rate γ is no surprise, since $1/\gamma$ is a natural time unit for the process.

3. Two approximations of the SIS model: exact results

We describe the two approximations of the SIS process mentioned in the introduction. Both of the approximations are birth-and-death processes whose transition rates are close to those given in the previous section. Furthermore, both of them have non-degenerate stationary distributions that can be found explicitly. The state space of each of the approximations differs from the state space of the SIS model by not including the state 0. The first approximation can be interpreted as the SIS model with one permanently infected individual. In this approximation, every recovery rate $\mu_n = n\gamma$ is replaced by $(n-1)\gamma$, while all infection rates λ_n are unchanged. The second approximation is interpreted as the SIS model with the origin removed. In this approximation the recovery rate μ_1 is replaced by 0, while all other transition rates are unchanged.

The stationary distribution of the first approximation of the SIS model is denoted $p^{(1)} = (p_1^{(1)}, \ldots, p_N^{(1)})$. It is straightforward to verify that this stationary distribution is given by the explicit expression

$$p_j^{(1)} = \alpha(j) R_0^{j-1} p_1^{(1)}, \quad j = 1, 2, \ldots, N, \tag{3.1}$$

where $\alpha(j)$ is defined in Eq. (2.6) and

$$p_1^{(1)} = \frac{1}{S^{(1)}}, \quad \text{with} \quad S^{(1)} = \sum_{j=1}^{N} \alpha(j) R_0^{j-1}. \tag{3.2}$$

Furthermore, the stationary distribution of the second approximation of the SIS model (with the origin removed) is denoted $p^{(0)} = (p_1^{(0)}, \ldots, p_N^{(0)})$. This stationary distribution is given by

$$p_j^{(0)} = \frac{1}{j} \alpha(j) R_0^{j-1} p_1^{(0)}, \quad j = 1, 2, \ldots, N, \tag{3.3}$$

where

$$p_1^{(0)} = \frac{1}{S^{(0)}}, \quad \text{with} \quad S^{(0)} = \sum_{j=1}^{N} \frac{1}{j} \alpha(j) R_0^{j-1}. \tag{3.4}$$

Numerical illustrations of the two stationary distributions $p^{(0)}$ and $p^{(1)}$ and of the quasi-stationary distribution q are given in Figs. 1–3. All numerical evaluations have been done using Matlab on a PC. The evaluations of $p^{(0)}$ and $p^{(1)}$ are based on the explicit expressions in Eqs. (3.1) and (3.3). The evaluation

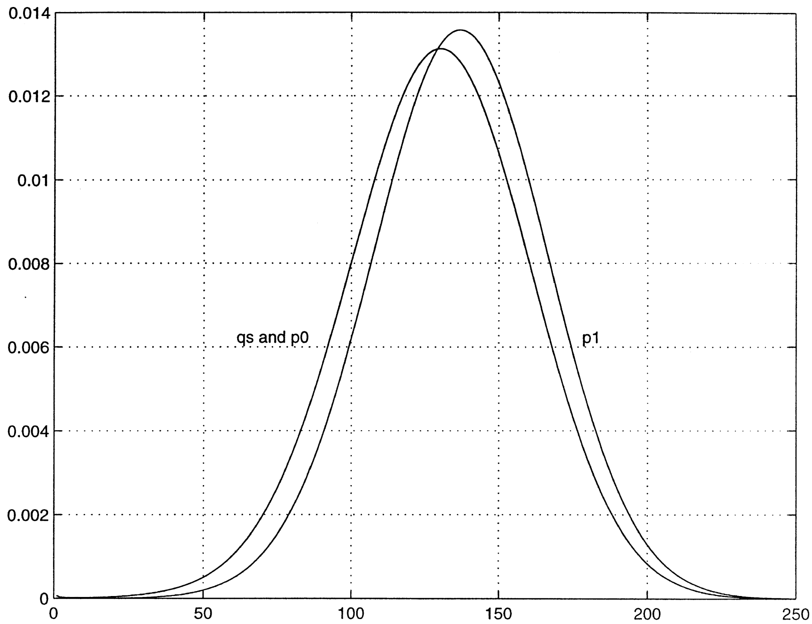

Fig. 1. The quasi-stationary distribution q is very close to the distribution $p^{(0)}$ and rather close to $p^{(1)}$ when R_0 is distinctly above the deterministic threshold. The parameters are $N = 1000$ and $\rho = 5$, and hence $R_0 = 1.16$.

of the quasi-stationary distribution q uses iteration of the Eqs. (2.4)–(2.7). One figure is devoted to each of the three parameter regions described in the introduction.

From a practical viewpoint it is useful to note the rule of thumb that for any finite value of N we are in the transition region if $|\rho| < 3$ and in either of the regions where R_0 is distinctly different from 1 if $|\rho| > 3$. The choice here of the bounding value 3 is somewhat arbitrary. It is related to the fact that a normally distributed random variable takes values within 3 standard deviations of its mean with high probability.

The plots strongly support the claim by Kryscio and Lefèvre [8] that the quasi-stationary distribution q is well approximated by the distribution $p^{(0)}$ when R_0 is distinctly larger than 1 and by the distribution $p^{(1)}$ when R_0 is distinctly smaller than 1. In the transition region, the quasi-stationary distribution q makes a transition from close to $p^{(1)}$ to close to $p^{(0)}$ as the value of R_0 grows past the value 1. The main result in this paper is an approximation of this behaviour. It is illustrated numerically by the dotted curve in Fig. 2. The approximation is given in Section 4 and derived in Section 8.

The figures give some understanding for the shapes of the distributions, but the explicit expressions for the two stationary distributions $p^{(0)}$ and $p^{(1)}$ do not.

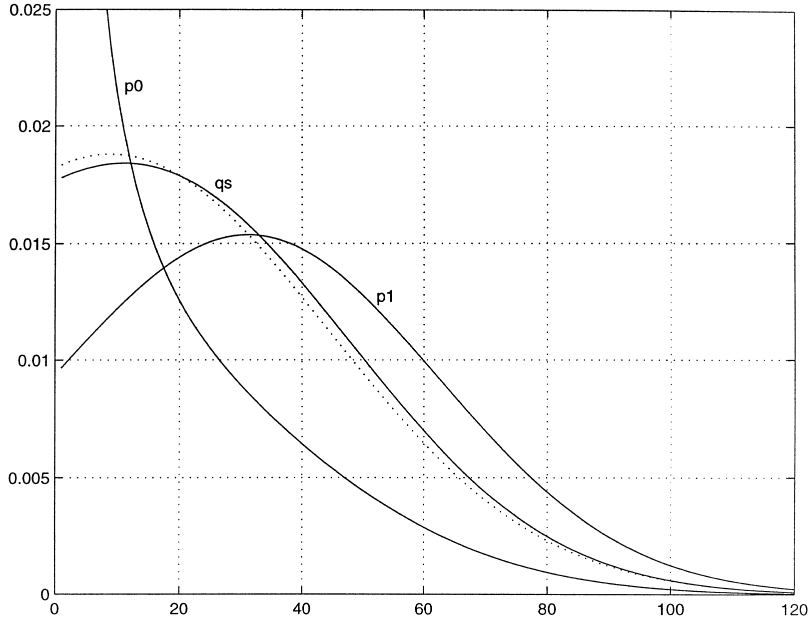

Fig. 2. The quasi-stationary distribution q differs from both of the approximations $p^{(0)}$ and $p^{(1)}$ in the transition region near the deterministic threshold. The approximation (4.7) of the quasi-stationary distribution is shown dotted. The parameters are $N = 1000$ and $\rho = 1$, and hence $R_0 = 1.03$.

In order to improve our understanding, we search for approximations of the exact solutions. These will also help the derivation of approximations of the quasi-stationary distribution (for which no explicit solution exists).

4. Approximations of the distributions $p^{(1)}$, $p^{(0)}$ and q: summary of results

Several approximations of the distributions $p^{(1)}$, $p^{(0)}$ and q have been derived in Ref. [2]. Some of them, but not all, have the mathematically desirable feature of being asymptotic as $N \to \infty$. We summarize some of these results here, together with the new results derived below. The new results deal with the distributions of $p^{(0)}$ and q in the transition region. The approximations that we give all apply to the body of the corresponding distribution, i.e. where $(j - \mu)/\sigma = \mathrm{O}(1)$ as $N \to \infty$, where μ and σ denote the mean and the standard deviation of the distribution.

The distribution $p^{(1)}$ is approximated by a normal distribution when R_0 is distinctly larger than 1, by a truncated normal distribution in the transition region, and by a geometric distribution when R_0 is distinctly smaller than 1. The three approximations can be expressed in the following way. In each case, the condition $N \to \infty$ is assumed.

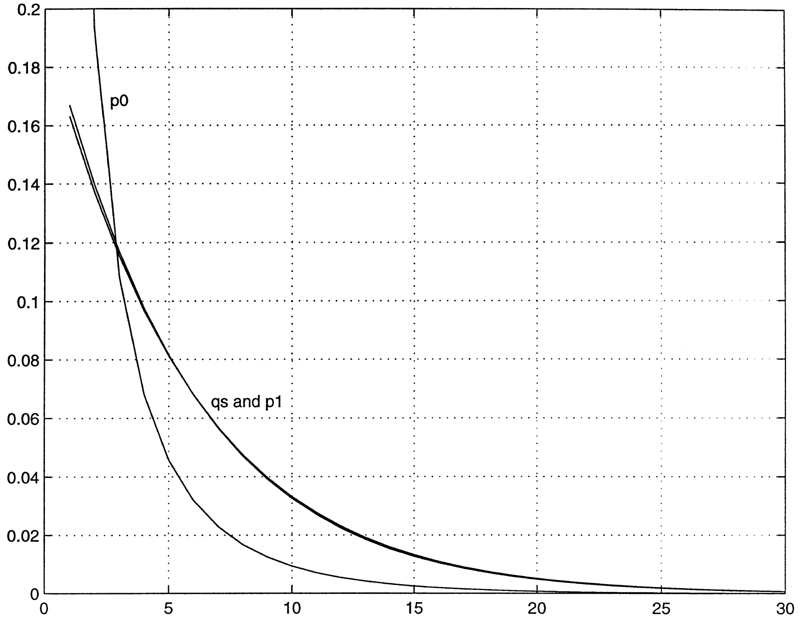

Fig. 3. The quasi-stationary distribution q is well approximated by the distribution $p^{(1)}$ but less well by the distribution $p^{(0)}$ when R_0 is distinctly below the deterministic threshold. The parameters are $N = 1000$ and $\rho = -5$, and hence $R_0 = 0.84$.

$$p_j^{(1)} \sim \begin{cases} \frac{1}{\sigma_1}\varphi(y_1(j)), & R_0 > 1, \ R_0 \text{ fixed}, \\ \frac{1}{\sigma}\frac{\varphi(y(j))}{\Phi(\rho)}, & \rho \text{ fixed}, \\ (1-R_0)R_0^{j-1}, & R_0 < 1, \ R_0 \text{ fixed}. \end{cases} \tag{4.1}$$

Here, $\varphi(x) = \exp(-x^2/2)/\sqrt{2\pi}$ denotes the normal density function and $\Phi(x) = \int_{-\infty}^{x} \varphi(t)\,\mathrm{d}t$ denotes the normal cumulative distribution function. Furthermore, the functions y_1 and y are defined in terms of the basic parameters N and R_0 via the parameters μ_1, σ_1, μ and σ as follows:

$$\mu_1 = N\frac{R_0 - 1}{R_0}, \ \ \sigma_1 = \sqrt{\frac{N}{R_0}}, \ \ y_1(j) = \frac{j - \mu_1}{\sigma_1}, \tag{4.2}$$

$$\mu = \sqrt{N}\rho, \ \ \sigma = \sqrt{N}, \ \ y(j) = \frac{j-\mu}{\sigma}. \tag{4.3}$$

The important probability $p_1^{(1)}$ of taking the value 1 belongs to the body of the distribution when R_0 is distinctly smaller than 1 and also in the transition region, but not when R_0 is distinctly larger than 1.

The three approximations are not uniform in R_0. This is illustrated by the fact that the geometric distribution that approximates $p^{(1)}$ when R_0 is distinctly smaller than 1 is useless for $R_0 = 1$. Ref. [2] gives an asymptotic approximation of the body of the distribution $p^{(1)}$ that is valid for all positive values of R_0. This approximation is uniform in R_0. Uniformity is desirable and mathematically powerful, but the resulting formula is more complicated and does not give the same insight as the nonuniform expressions that we emphasize in this paper.

The distribution $p^{(0)}$ is approximated by a normal distribution when R_0 is distinctly larger than 1, and by a log series distribution when R_0 is distinctly smaller than 1. The three approximations of $p^{(0)}$ can be expressed as follows:

$$p_j^{(0)} \sim \frac{1}{\sigma_1}\varphi(y_1(j)), \quad R_0 > 1, \; R_0 \text{ fixed}, \tag{4.4}$$

$$p_j^{(0)} \approx \frac{1}{(1/2 \cdot \log N + H_0(\rho))\varphi(\rho)} \frac{\varphi(y(j))}{j}, \quad \rho \text{ fixed}, \tag{4.5}$$

$$p_j^{(0)} \sim \frac{R_0^j}{j \log(1/(1-R_0))}, \quad R_0 < 1, \; R_0 \text{ fixed}. \tag{4.6}$$

The function H_0 is defined in Section 7 where it is also used to derive approximations of $S^{(0)}$ and $p^{(0)}$ in the transition region. Note that the two approximations of $p^{(1)}$ and $p^{(0)}$ for R_0 distinctly larger than 1 coincide.

Finally, the approximations of the quasi-stationary distribution q coincide with those of the distribution $p^{(1)}$ when R_0 is distinctly larger or smaller than 1. The three approximations of q take the following form:

$$q_j \approx \begin{cases} \frac{1}{\sigma_1}\varphi(y_1(j)), & R_0 > 1, \quad R_0 \text{ fixed}, \\ \frac{1}{\varphi(\rho)(1+\rho H(\rho))} \frac{1}{j}(1 - \frac{1}{R_1^j})\varphi(y(j)), & \rho \text{ fixed}, \\ (1-R_0)R_0^{j-1}, & R_0 < 1, \quad R_0 \text{ fixed}, \end{cases} \tag{4.7}$$

where

$$R_1 = 1 + \frac{\rho + 1/H(\rho)}{\sqrt{N}}. \tag{4.8}$$

The function H is defined in Section 8 where it is also used to derive approximations of the sum S and of the quasi-stationary distribution q in the transition region. The approximation of q in the transition region given by Nåsell [2] is crude and replaced by the above expression.

5. Approximations of the sums $S^{(1)}$, $S^{(0)}$ and S: summary of results

The sums $S^{(1)}$, $S^{(0)}$ and S, defined in Eqs. (3.2), (3.4) and (2.7), are important for our work. We summarize approximations of them in each of the three parameter regions. With R_0 distinctly larger than 1 we get

$$S^{(1)} \sim \frac{\sqrt{N}}{R_0} \frac{1}{\varphi(\beta_1)}, \quad R_0 > 1, \ R_0 \text{ fixed}, \tag{5.1}$$

$$S^{(0)} \sim \frac{1}{\sqrt{N}(R_0 - 1)} \frac{1}{\varphi(\beta_1)}, \quad R_0 > 1, \ R_0 \text{ fixed}, \tag{5.2}$$

$$S \approx \frac{R_0}{\sqrt{N}(R_0 - 1)^2} \frac{1}{\varphi(\beta_1)}, \quad R_0 > 1, \ R_0 \text{ fixed}, \tag{5.3}$$

where

$$\beta_1 = \operatorname{sign}(R_0 - 1)\sqrt{2N\left(\log R_0 - \frac{R_0}{R_0 - 1}\right)}. \tag{5.4}$$

Thus, the approximations of all three sums grow exponentially with N in this region.

With R_0 distinctly smaller than 1 we get

$$S^{(1)} \sim \frac{1}{1 - R_0}, \quad R_0 < 1, \ R_0 \text{ fixed}, \tag{5.5}$$

$$S^{(0)} \sim \frac{1}{R_0} \log \frac{1}{1 - R_0}, \quad R_0 < 1, \ R_0 \text{ fixed}, \tag{5.6}$$

$$S \approx \frac{1}{1 - R_0}, \quad R_0 < 1, \ R_0 \text{ fixed}. \tag{5.7}$$

Thus the approximations of all three sums are independent of N in this region.

Finally, we get the following results in the transition region:

$$S^{(1)} \sim \sqrt{N} H_1(\rho), \quad \rho \text{ fixed}, \tag{5.8}$$

$$S^{(0)} \approx \frac{1}{2} \log N + H_0(\rho), \quad \rho \text{ fixed}, \tag{5.9}$$

$$S \approx \sqrt{N} H(\rho) \quad \rho \text{ fixed}. \tag{5.10}$$

Here, the function H_1 is defined by the relation

$$H_1(\rho) = \frac{\Phi(\rho)}{\varphi(\rho)}, \tag{5.11}$$

while the functions H_0 and H are defined in Sections 7 and 8. The approximation of $S^{(1)}$ in the transition region is derived by Nåsell [2], while the approximations of $S^{(0)}$ and S in the same region are derived in Sections 7 and 8.

6. Approximations of the time to extinction from quasi-stationarity: summary of results

It was noted in Section 2 that the time to extinction from quasi-stationarity τ_Q has an exponential distribution whose expectation is equal to

$\mathrm{E}(\tau_Q) = 1/(\gamma q_1) = S/\gamma$. Approximations of the expected time to extinction from quasi-stationarity follow therefore directly from our approximations of the sum S. The approximations take the following form in the three parameter regions:

$$\mathrm{E}\tau_Q \approx \begin{cases} \frac{R_0}{\gamma\sqrt{N}(R_0-1)^2}\,\frac{1}{\varphi(\beta_1)}, & R_0 > 1, \ R_0 \text{ fixed}, \\ \frac{\sqrt{N}H(\rho)}{\gamma}, & \rho \text{ fixed}, \\ \frac{1}{\gamma(1-R_0)}, & R_0 < 1, \ R_0 \text{ fixed}. \end{cases} \tag{6.1}$$

Fig. 4 shows the expected time to extinction from quasi-stationarity in the transition region, and its approximation, as functions of N for the ρ-value 1 with $\gamma = 1$. The agreement is reasonably good over a large range of N-values, but we observe that the ratio between the numerically computed value and its approximation does not seem to approach the value 1 as N becomes large. This indicates that our approximation may not be asymptotic.

Figs. 5 and 6 show the expected time to extinction from quasi-stationarity and its approximation as functions of ρ in the transition region for the fixed value $N = 1000$, also with $\gamma = 1$. A linear vertical scale is used for ρ negative, and a logarithmic vertical scale for ρ positive. The approximation gets poorer

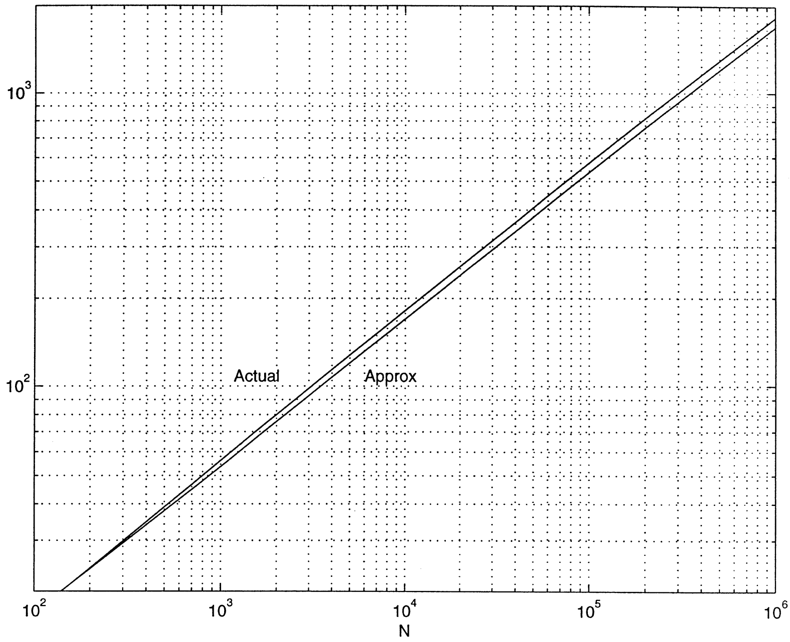

Fig. 4. The expected time to extinction from quasi-stationarity and its approximation (6.1) in the transition region are shown as functions of N for $\rho = 1$ and $\gamma = 1$.

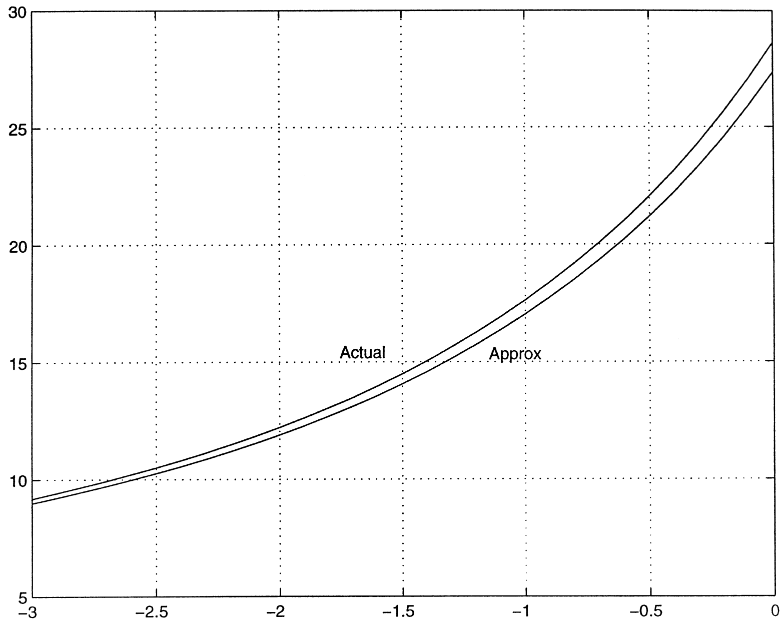

Fig. 5. The expected time to extinction from quasi-stationarity and its approximation (6.1) are shown as functions of ρ for $N = 1000$ and $\gamma = 1$.

as ρ increases. This is related to the fact that the approximation that we have found for the transition region is not uniform in R_0 or ρ.

7. Approximations of the sum $S^{(0)}$ and of the distribution $p^{(0)}$ in the transition region: derivations

The first step in this section is to derive the approximation (5.9) of the sum $S^{(0)}$ in the transition region. This represents an improvement over Nåsell [2], where no useful approximation was found for $S^{(0)}$ in the transition region. It does not appear possible to derive an approximation of $S^{(0)}$ directly from the definition (3.4) when ρ is fixed.

Our approach is based on three steps. The first step is to express $S^{(0)}$ as an integral of $S^{(1)}$ with respect to the parameter R_0, the second one is to insert a uniform approximation of $S^{(1)}$ into the integrand of this integral, and the third one is to approximate the resulting integral.

In order to express the relation between $S^{(0)}$ and $S^{(1)}$ we emphasize the functional dependence of each of the two sums on R_0 by writing $S^{(0)}(R_0)$ instead of $S^{(0)}$ and $S^{(1)}(R_0)$ instead of $S^{(1)}$. From their definitions in Eqs. (3.2) and (3.4) we find that $S^{(1)}$ equals the derivative with respect to R_0 of the product of R_0 and $S^{(0)}$

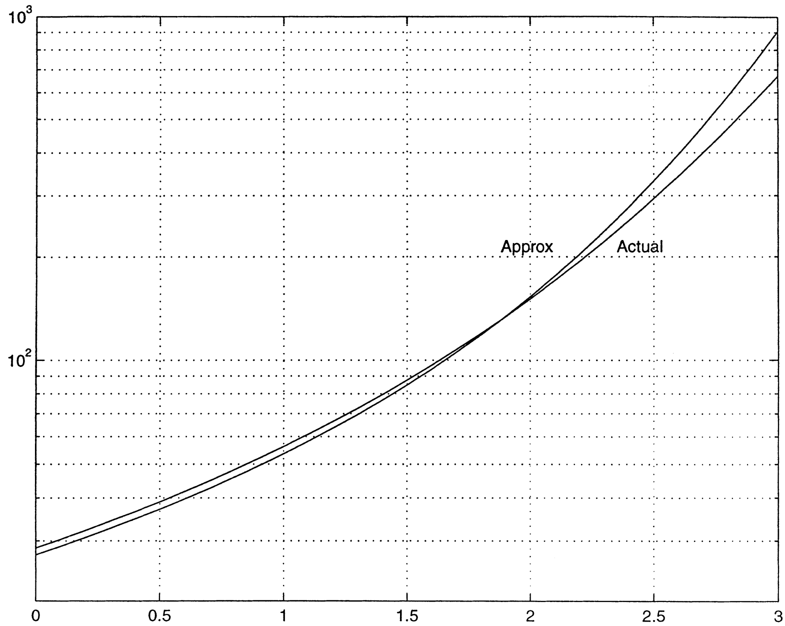

Fig. 6. The expected time to extinction from quasi-stationarity and its approximation (6.1) are shown as functions of ρ for $N = 1000$ and $\gamma = 1$.

$$S^{(1)}(R_0) = \frac{d}{\mathrm{d}R_0}\left(R_0 S^{(0)}(R_0)\right). \tag{7.1}$$

By noting that $R_0 S^{(0)}(R_0) = 0$ for $R_0 = 0$ we find from this relation that $S^{(0)}$ can be determined from $S^{(1)}$ by integration as follows:

$$S^{(0)}(R_0) = \frac{1}{R_0}\int_0^{R_0} S^{(1)}(x)\,\mathrm{d}x. \tag{7.2}$$

This integration interval covers two parameter regions, namely both the one where $x < 1$ is fixed as $N \to \infty$ and the transition region. A uniform approximation of $S^{(1)}$ that covers both of these regions was derived in Ref. [2]. It can be written in the form

$$S^{(1)}(R_0) \sim \frac{\sqrt{N}}{R_0}\frac{\Phi(\beta_2)}{\varphi(\beta_2)}, \quad 0 < R_0 < 1 \quad \text{or} \quad \rho = \mathrm{O}(1), \; N \to \infty, \tag{7.3}$$

where

$$\beta_2 = \frac{R_0 - 1}{R_0}\sqrt{N}. \tag{7.4}$$

By insertion of this asymptotic approximation of $S^{(1)}$ into the integral above we find after change of integration variable that

$$S^{(0)} \sim \frac{1}{R_0} \int_{-\infty}^{\beta_2} \frac{1}{1 - y/\sqrt{N}} \frac{\Phi(y)}{\varphi(y)} \, \mathrm{d}y$$

$$\sim \int_{-\infty}^{\rho} \frac{1}{1 - y/\sqrt{N}} \frac{\Phi(y)}{\varphi(y)} \mathrm{d}y, \quad \rho = \mathrm{O}(1), \ N \to \infty. \tag{7.5}$$

This completes the second of the three steps. Essentially the same result is derived by Nåsell [2].

The third step consists of finding an approximation of this integral. It uses both approximations and bounds of $\Phi(y)/\varphi(y)$. To describe them we define

$$a_k = (-1)^k 2^{k-1} \frac{\Gamma(k - 1/2)}{\Gamma(1/2)}, \quad k = 1, 2, \ldots, \tag{7.6}$$

and put

$$S_m(y) = \sum_{k=1}^{m} \frac{a_k}{y^{2k-1}} = -\frac{1}{y} + \frac{1}{y^3} - \frac{1 \cdot 3}{y^5} + \frac{1 \cdot 3 \cdot 5}{y^7} - \cdots + \frac{a_m}{y^{2m-1}}. \tag{7.7}$$

-0.5It is well known that the sum $S_m(y)$ gives an asymptotic approximation of $\Phi(y)/\varphi(y)$ as $y \to -\infty$, and also that $S_m(y)$ gives an upper bound of $\Phi(y)/\varphi(y)$ for $y < 0$ if m is odd and a lower bound if m is even. Derivations of these results are based on integration by parts of $\int_{-\infty}^{y} \exp(-t^2/2)\mathrm{d}t$. The results are given as formula (26.2.12) in Ref. [13]. By using the bounds one readily finds that

$$\left| \frac{\Phi(y)}{\varphi(y)} - S_m(y) \right| < -\frac{|a_{m+1}|}{y^{2m+1}}, \quad y < 0. \tag{7.8}$$

With these preparations we turn to the derivation of an approximation of the integral in Eq. (7.5). The interval of integration in the integral of Eq. (7.5) is written as a union of three subintervals by introducing ρ_a and ρ_b to satisfy the inequalities $-\infty < \rho_a < \rho_b < \rho$. We assume that ρ_a is a function of N such that $\rho_a < 0, \rho_a \to -\infty$ as $N \to \infty$ and $\rho_a = o(\sqrt{N})$. Furthermore, $\rho_b < 0$ is a constant, to be determined later. We can then write

$$S^{(0)} \sim I_1 + I_2 + I_3 + I_4, \tag{7.9}$$

where

$$I_1 = \int_{-\infty}^{\rho_a} \frac{1}{1 - y/\sqrt{N}} \frac{\Phi(y)}{\varphi(y)} \, \mathrm{d}y, \tag{7.10}$$

$$I_2 = \int_{\rho_a}^{\rho_b} \frac{1}{1 - y/\sqrt{N}} S_m(y) \, \mathrm{d}y, \tag{7.11}$$

$$I_3 = \int_{\rho_a}^{\rho_b} \frac{1}{1 - y/\sqrt{N}} \left(\frac{\Phi(y)}{\varphi(y)} - S_m(y) \right) \mathrm{d}y, \tag{7.12}$$

$$I_4 = \int_{\rho_b}^{\rho} \frac{1}{1 - y/\sqrt{N}} \frac{\Phi(y)}{\varphi(y)} \, \mathrm{d}y. \tag{7.13}$$

Here, m is a positive integer to be determined later.

Among these integrals, an asymptotic approximation of I_1 is found by replacing $\Phi(y)/\varphi(y)$ in the integrand by its asymptotic approximation $S_m(y)$. The integrand of the approximating integral is the same as the integrand of I_2. Since this integrand is rational, the integrals can be evaluated. The absolute value of the integral I_3 is bounded with the help of the inequality in Eq. (7.8). We shall determine m to minimize an asymptotic approximation of this bound. Finally, we replace the first factor in the integrand of I_4 by its asymptotic approximation 1. The resulting approximating integral is a function of ρ_b and ρ that is independent of N.

The absolute value of I_3 is bounded as follows:

$$|I_3| < -|a_{m+1}| \int_{\rho_a}^{\rho_b} \frac{1}{1 - y/\sqrt{N}} \frac{1}{y^{2m+1}} \, \mathrm{d}y. \tag{7.14}$$

By partial fraction expansion, integration, and substitution of the integration bounds, we find after leaving out terms that are o(1) as $N \to \infty$ that the upper bound of $|I_3|$ is asymptotically approximated by

$$b_m = \frac{|a_{m+1}|}{2m\rho_b^{2m}}. \tag{7.15}$$

By using the definition of a_k we find that

$$\frac{b_{m+1}}{b_m} = \frac{(2m+1)m}{(m+1)\rho_b^2}. \tag{7.16}$$

The value of m that minimizes b_m is the largest integer $m_b = m_b(\rho_b)$ that is smaller than or equal to the positive root of the equation in m that results by putting $b_{m+1} = b_m$. Thus we get

$$m_b(\rho_b) = \left[\frac{\rho_b^2 - 1 + \sqrt{(\rho_b^2 + 1)^2 + 4\rho_b^2}}{4} \right], \tag{7.17}$$

where the brackets indicate the largest integer less than or equal to the argument inside the brackets. The resulting asymptotic approximation of the error bound is b_{mb}. Our approximation of $I_1 + I_2$ can now be written

$$I_1 + I_2 \approx \int_{-\infty}^{\rho_b} \frac{1}{1 - y/\sqrt{N}} S_{mb}(y) \, \mathrm{d}y. \tag{7.18}$$

By partial fraction expansion of the integrand we find after integration and substitution of the integration bounds and leaving out all terms that are o(1) as $N \to \infty$ that this approximation of $I_1 + I_2$ is asymptotic to

$$\frac{1}{2}\log N - \log|\rho_b| - \sum_{k=2}^{m_b}\frac{a_k}{2k-2}\frac{1}{\rho_b^{2k-2}}. \tag{7.19}$$

Our results so far are based on the assumption that $\rho_b < \rho$. If instead $\rho \leqslant \rho_b$, then the integration interval from $-\infty$ to ρ is subdivided into two subintervals by the common boundary point ρ_a. It is straightforward to verify that the function H_0 that appears in the approximation (5.9) of $S^{(0)}$ can be defined as follows below in these two cases. We begin by defining an auxiliary function H_a:

$$H_a(\rho) = -\log|\rho| - \sum_{k=2}^{m_b}\frac{a_k}{2k-2}\frac{1}{\rho^{2k-2}}. \tag{7.20}$$

We can then use H_a to express H_0 as follows

$$H_0(\rho) = \begin{cases} H_a(\rho) & \rho \leqslant \rho_b, \\ H_a(\rho_b) + \int_{\rho_b}^{\rho}\frac{\Phi(y)}{\varphi(y)}\,\mathrm{d}y, & \rho > \rho_b. \end{cases} \tag{7.21}$$

This definition of H_0 is unique only after we specify the value of ρ_b. A guidance to the choice of ρ_b is given by the following table. It lists the number of terms m_b and the asymptotic approximation b_{mb} of the error bound for some possible values of ρ_b.

ρ_b	m_b	b_{m_b}
−3	4	2×10^{-3}
−4	8	3×10^{-5}
−5	12	2×10^{-7}
−6	18	6×10^{-10}

It appears from the table that the asymptotic approximation of the error bound can be made acceptably small by choosing the absolute value of ρ_b sufficiently large. We shall use $\rho_b = -4$ in our numerical illustrations.

Numerical evaluations indicate that the sum $S^{(0)}$ has an asymptotic approximation of the form

$$S^{(0)} \sim \frac{1}{2}\log N + \tilde{H}_0(\rho). \tag{7.22}$$

If this is true, then the approximation of $S^{(0)}$ given in Eq. (5.9) lacks the property of being asymptotic only because the function H_0 is not equal to $\tilde{H}_0$. Our arguments above indicate that the error that we commit by using H_0 instead of $\tilde{H}_0$ can be made acceptably small.

The distribution $p^{(0)}$ is given explicitly by Eq. (3.3). We derive here the approximation given in Eq. (4.5) in the transition region. The approximation is claimed to be valid in the body of the distribution where $j = \mathrm{O}(\sqrt{N})$.

To derive this result, we note from Ref. [2] that the factor $\alpha(j)R_0^{j-1}$ is approximated as follows:

$$\alpha(j)R_0^{j-1} \sim \frac{1}{R_0}\frac{\varphi(y_2(j))}{\varphi(\beta_3)}, \quad j = \mathrm{O}(\sqrt{N}), \quad \rho = \mathrm{O}(1), \quad N \to \infty, \tag{7.23}$$

where

$$y_2(j) = \frac{j - \mu_2}{\sigma_2}, \quad \mu_2 = N \log R_0, \quad \sigma_2 = \sqrt{N}, \tag{7.24}$$

and

$$\beta_3 = \sqrt{N} \log R_0. \tag{7.25}$$

By utilizing the fact that $\rho = \mathrm{O}(1)$ we find that

$$\alpha(j)R_0^{j-1} \sim \frac{1}{R_0}\frac{\varphi(y(j))}{\varphi(\rho)}, \quad j = \mathrm{O}(\sqrt{N}), \quad \rho = \mathrm{O}(1). \tag{7.26}$$

The approximation of $p_j^{(0)}$ in Eq. (4.5) now follows by insertion into Eq. (3.3) of this approximation of $\alpha(j)R_0^{j-1}$, and using the relation $p_1^{(0)} = 1/S^{(0)}$ and the approximation of $S^{(0)}$ in Eq. (5.9).

8. Approximations of the sum S and of the quasi-stationary distribution q in the transition region: derivations

In this section we define the function H and show that it leads to the approximation (5.10) of the sum S and the approximation of the quasi-stationary distribution q in the transition region given in Eq. (4.7). Our derivations do not allow us to conclude that the approximations are asymptotic.

The relations (2.4)–(2.7) show that the sum $S = 1/q_1$ is important for determining the quasi-stationary distribution q. In contrast to the case with the sums $S^{(0)}$ and $S^{(1)}$, no explicit expression is available for S.

The function H that appears in the approximation is defined implicitly for all real values of ρ by the following relation:

$$H(\rho) = \frac{1}{\rho + 1/H(\rho)} \int_{-1/H(\rho)}^{\rho} \frac{\Phi(y)}{\varphi(y)}\,\mathrm{d}y. \tag{8.1}$$

We begin by deriving an approximation of $\delta(k)$ in Eq. (2.5) for small values of k. To be specific, we assume that $k = \mathrm{O}(1)$ as $N \to \infty$. It follows from Eq. (4.1) that $p_l^{(1)} \sim \varphi(\rho)/(\sqrt{N}\Phi(\rho))$. This means that the l-dependence of the approximation is of smaller order of magnitude for small values of l. We start out by making the assumption that q_l behaves in a similar way, namely that $q_l \sim 1/(\sqrt{N}H(\rho))$, where it remains to show that the function H is given by Eq. (8.1). It follows that the numerator of $\delta(k)$ is asymptotically approximated by

$$1 - \sum_{l=1}^{k-1} q_l \sim 1 - (k-1)/(\sqrt{N}H(\rho)), \quad k = \mathrm{O}(1). \tag{8.2}$$

We now turn to the denominator of $\delta(k)$. With the restriction $k = \mathrm{O}(1)$ we find from Eq. (7.26) that

$$\alpha(k)R_0^{k-1} \sim \frac{1}{R_0}\frac{\varphi(y(k))}{\varphi(\rho)} \sim 1 + \frac{(k-1)\rho}{\sqrt{N}}, \quad k = \mathrm{O}(1). \tag{8.3}$$

We can therefore conclude that

$$\delta(k) \sim 1 - \frac{(k-1)(\rho + 1/H(\rho))}{\sqrt{N}}, \quad k = \mathrm{O}(1). \tag{8.4}$$

This result can be expressed in the form

$$\delta(k) \sim \frac{1}{R_1^{k-1}}, \quad k = \mathrm{O}(1), \;\; N \to \infty, \tag{8.5}$$

where we use the notation

$$R_1 = 1 + \frac{\rho + 1/H(\rho)}{\sqrt{N}}. \tag{8.6}$$

By inserting this approximation into the expression in Eq. (2.5) that defines $\gamma(j)$ in terms of $\delta(k)$ we find that

$$\gamma(j) \sim \frac{1}{j}\frac{R_1}{R_1 - 1}\left(1 - \frac{1}{R_1^j}\right), \quad j = \mathrm{O}(1), \;\; N \to \infty. \tag{8.7}$$

The next step in the approximation procedure consists in inserting this approximation of $\gamma(j)$ into the sum in Eq. (2.7) that defines S. This step has not been justified since the approximation of $\gamma(j)$ has been shown to be asymptotic only for small values of j. We get

$$S \approx \sum_1^N \frac{1}{j}\frac{R_1}{R_1 - 1}\left(1 - \frac{1}{R_1^j}\right)\alpha(j)R_0^{j-1}. \tag{8.8}$$

This approximation of the sum S can be expressed with the help of the sum $S^{(0)} = S^{(0)}(R_0)$. Thus

$$S \approx \frac{R_1}{R_1 - 1}\left[S^{(0)}(R_0) - \frac{1}{R_1}S^{(0)}\left(\frac{R_0}{R_1}\right)\right]. \tag{8.9}$$

Note now that $R_0/R_1 \sim 1 - 1/(\sqrt{N}H(\rho))$ as $N \to \infty$. By using the approximation of $S^{(0)}$ in (5.9) we find that

$$\begin{aligned} S &\approx \frac{\sqrt{N}}{\rho + 1/H(\rho)}\left(\frac{1}{2}\log N + H_0(\rho) - \frac{1}{2}\log N - H_0\left(-\frac{1}{H(\rho)}\right)\right) \\ &= \frac{\sqrt{N}}{\rho + 1/H(\rho)}\int_{-1/H(\rho)}^{\rho}\frac{\Phi(y)}{\varphi(y)}\,\mathrm{d}y. \end{aligned} \tag{8.10}$$

The probability q_j for small j-values can be approximated by using Eq. (2.4) with the above approximation of $\gamma(j)$. We get

$$q_j \sim \frac{1}{S}, \quad j = \mathrm{O}(1), \quad \rho = \mathrm{O}(1), \quad N \to \infty. \tag{8.11}$$

By using the approximation of S we find therefore that the assumption made at the beginning of this section, namely that $q_j \sim 1/(\sqrt{N}H(\rho))$ for $j = \mathrm{O}(1)$, leads to the relation (8.1) that defines the function H. This also establishes the approximation of the sum S given in Eq. (5.10).

The approximation in Eq. (4.7) of the quasi-stationary distribution in the transition region follows by inserting the approximations of $\gamma(j)$ in Eq. (8.7), of $\alpha(j)R_0^{j-1}$ in Eq. (7.26), and of $S = 1/q_1$ into the expression (2.4) for q_j.

A graphical comparison between a numerical evaluation of the quasi-stationary distribution and the approximation (4.7) in the transition region is contained in Fig. 2. The approximation is seen to agree surprisingly well with the distribution it approximates. In this figure, ρ equals 1. Comparisons for other ρ-values in the interval $-3 \leqslant \rho \leqslant 2$ show agreements that are reasonably close to what is exhibited by Fig. 2, while the agreements are somewhat poorer in the interval $2 < \rho \leqslant 3$.

References

[1] I. Nåsell, The threshold concept in stochastic epidemic and endemic models. in: D. Mollison (Ed.), Epidemic Models: Their Structure and Relation to Data, Cambridge University, Cambridge, 1995.

[2] I. Nåsell, The quasi-stationary distribution of the closed endemic SIS model, Adv. Appl. Prob. 28 (1996) 895.

[3] G.H. Weiss, M. Dishon, On the asymptotic behavior of the stochastic and deterministic models of an epidemic, Math. Biosci. 11 (1971) 261.

[4] D.J. Bartholomew, Continuous time diffusion models with random duration of interest, J. Math. Sociol. 4 (1976) 187.

[5] I. Oppenheim, K.E. Shuler, G.H. Weiss, Stochastic theory of nonlinear rate processes with multiple stationary states, Physica A 88 (1977) 191.

[6] J.A. Cavender, Quasi-stationary distributions of birth-and-death processes, Adv. Appl. Prob. 10 (1978) 570.

[7] R.H. Norden, On the distribution of the time to extinction in the stochastic logistic population model, Adv. Appl. Prob. 14 (1982) 687.

[8] R.J. Kryscio, C. Lefèvre, On the extinction of the S–I–S stochastic logistic epidemic, J. Appl. Prob. 26 (1989) 685.

[9] J.A. Jacquez, C.P. Simon, The stochastic SI model with recruitment and deaths. I. Comparisons with the closed SIS model, Math. Biosci. 117 (1993) 77.

[10] H. Andersson, B. Djehiche, A threshold limit theorem for the stochastic logistic epidemic, J. Appl. Prob. 35 (1998) 662.

[11] I. Nåsell, On the time to extinction in recurrent epidemics, to appear in J. Roy. Stat. Soc., Series B.

[12] R. Ross, The Prevention of Malaria, 2nd ed., John Murray, London, 1911.

[13] M. Abramowitz, I. Stegun, Handbook of Mathematical Functions, Dover, New York, 1965.