Coupling a branching process to an infinite dimensional epidemic process

A. D. Barbour*
Universität Zürich

To Cindy Greenwood, for her 70th.

Abstract

Branching process approximation to the initial stages of an epidemic process has been used since the 1950's as a technique for providing stochastic counterparts to deterministic epidemic threshold theorems. One way of describing the approximation is to construct both branching and epidemic processes on the same probability space, in such a way that their paths coincide for as long as possible. In this paper, it is shown, in the context of a Markovian model of parasitic infection, that coincidence can be achieved with asymptotically high probability until M_N infections have occurred, as long as $M_N = o(N^{2/3})$, where N denotes the total number of hosts.

1 Introduction

The classical law of large numbers and central limit theorem have process analogues for many Markovian models arising in population ecology. The law of large numbers is replaced by a deterministic process, obtained by solving an appropriate system of ordinary or partial differential equations, and the central limit theorem is replaced by a diffusion approximation around

*Angewandte Mathematik, Winterthurerstrasse 190, CH–8057 ZÜRICH, Switzerland: a.d.barbour@math.uzh.ch

the deterministic limit. For many techniques and examples concerning such density dependent Markov population processes, see Kurtz (1976, 1981).

In the context of invasion biology, when the central question is whether the introduction of a small number of individuals of a species can lead to its becoming established in a new habitat, these large population approximations are no longer appropriate. The more natural process approximations, at least if spatial restrictions on mixing are not critical in such small populations, are now branching processes. These were introduced, in the context of epidemic theory, by Whittle (1955), Kendall (1956) and Bartlett (1956, p. 129); here, infected individuals play the part of the invading species, and those that are infected by an individual correspond to an individual's 'offspring'.

When considering the development of a single species as a branching process, the biological quantity R_0, the lifetime mean number of offspring of a single individual when unhampered by competition from others of the same species, is just the mean offspring number of the corresponding Galton–Watson process. The branching process criticality theorem then corresponds to the biological meta-theorem, that an invading population can only become established if its R_0 (in the context that it experiences upon invasion) exceeds 1. For models involving more species, the analogy is to multitype branching processes, and the dominant eigenvalue of the mean matrix of the branching process has a corresponding interpretation in the biological context. For more detailed discussions of such issues, see Heesterbeek (1992) and Diekmann & Heesterbeek (2000, Section 5.7).

Whittle (1955) was able to justify his birth and death approximation to the early stages of the Markovian SIR-epidemic, and hence his formula for the probability of a large epidemic occurring, by sandwiching the epidemic process, during its initial stages, between two birth and death processes with slightly differing transition rates. This can be interpreted in terms of a pathwise comparison of processes. Ball (1983) and Ball and Donnelly (1995) went rather further, using a coupling argument to link the epidemic process with an approximating branching process on one and the same probability space, in such a way that the paths of the two processes are identical for a certain (random) length of time. In particular, they showed that the total variation distance between the distributions of the paths of the branching and epidemic processes is small, up to the time at which $M = M_N$ infections

have taken place, for any choice $M_N = o(\sqrt{N})$. They also suggest that this range of M_N cannot be extended.

The coupling used by Ball and Donnelly is simple and natural, and it is somewhat surprising that accurate coupling is in fact possible, for some epidemic processes, over rather longer time intervals than they had supposed possible. This was first established by Barbour and Utev (2004), in the context of the discrete time Reed–Frost epidemic process. They showed that the branching process approximation to the path distribution actually has asymptotically small error in total variation for all choices of $M_N = o(N^{2/3})$. The essence of their argument lay in examining the likelihood ratio of the two processes along paths of given length, and showing that it was typically close to 1. In this paper, we show that similar arguments can also be applied to some continuous time models. We take as example the infinite dimensional BK-model, introduced in Barbour and Kafetzaki (1993) and subsequently generalized by Luchsinger (2002a,b), for describing the transmission of the parasitic disease schistosomiasis.

2 The BK-model

In the BK-model, N hosts are infected by parasites, with $X_j^N(t)$ hosts having j parasites at time t, for $j \in \mathbb{Z}_+$ and $t \geq 0$. The process evolves as a Markov jump process X^N in continuous time on the set $\mathcal{X} := \{(\xi_j \in \mathbb{Z}_+,\, j \geq 0)\colon\ \sum_{j\geq 0} \xi_j = N\}$, with transition rates given by

$$\begin{aligned} \xi &\to \xi + e(j-1) - e(j) \quad \text{at rate} \quad j\mu\xi_j, \quad j \geq 1; \\ \xi &\to \xi + e(j) - e(0) \quad \text{at rate} \quad \lambda\xi_0 \sum_{l\geq 1} (\xi_l/N) p_{lj}, \quad j \geq 1, \end{aligned}$$

for any $\xi \in \mathcal{X}$, where $e(j)$ denotes the unit vector in the j-th coordinate. The first of the transitions models the death of a parasite in one of the ξ_j hosts currently carrying j parasites, the parasites being assumed to have independent exponentially distributed lifetimes with mean $1/\mu$. The second transition models infection. Only currently uninfected hosts can be newly infected, and each makes contacts that could potentially lead to infection at rate λ, the chance of such a contact being made with a host carrying l parasites being (ξ_l/N) (homogeneous mixing of hosts). If there is such a contact between an uninfected host and an l-host, then j parasites are established

in the previously uninfected host with probability p_{lj}; in the BK-model, it is supposed that $p_{lj} = \mathbb{P}[U_l = j]$, for $U_l := \sum_{i=1}^{l} Y_i$, where the Y_i are independent and identically distributed random variables with mean θ and finite variance, implying that each of the l parasites transmits on average θ infective stages to the newly infected host at an infectious contact, independently of the others.

For a disease such as schistosomiasis, infection is actually indirect, and involves a host infecting suitable aquatic snails and these snails subsequently passing infection to other hosts. Thus the BK-model does not seem at first sight to be at all realistic. However, it can be thought of as an extreme case of a model incorporating features of the transmission process that were not present in many of the previous models: infection by parasites in groups, rather than singly, immunity in the definitive host (here, in the form of perfect concomitant immunity), explicit incorporation of the parasite burdens of individual hosts.

The model that results is interesting for a number of reasons. The first is that, although it is rather complicated, it is still simple enough for some analytic conclusions to be reached. For instance, it can be shown that the model has a 'law of large numbers' approximation for large N, in the form of the solution to an infinite system of differential equations, whose components approximate the proportions of hosts with different numbers of parasites. If $\theta > e$, this differential equation system has no (endemic) equilibrium solution that yields a finite mean number of parasites per host. In practice, the distribution of parasites among hosts is observed to be extremely irregular, so that such behaviour is very encouraging: most earlier models have tacitly predicted Poisson–like distributions, which are far from realistic, and those that have tried to account for the over-dispersed distributions observed have imposed a specific form for the distribution without proposing any mechanism that might generate it. Another feature is that, if $\theta < e$, there is exactly one equilibrium distribution of the differential equation system that has finite mean number of parasites per host, and that, in this equilibrium, the distribution of the number of parasites per host, conditional on the host being infected, depends only on the value of θ, and not on λ or μ.

For the purposes of this paper, it is the behaviour when few hosts are infected that is of primary relevance, with interest centering on questions such as the probability that the introduction of a single infected host can cause

endemic infection to become established. These are the kinds of problem that can be addressed by way of a branching process approximation. Here, we begin by proving an error bound for the approximation (Theorem 3.1) that is asymptotically valid in total variation for paths of length $o(N^{2/3})$ transitions as $N \to \infty$. The branching process in turn yields a criticality theorem, which, to a close approximation, describes whether or not endemic equilibrium is possible in the BK-model.

However, the approximating branching process — a Markov branching process with countably infinitely many types — itself displays unexpected critical behavour. If $\theta \leq e$, the branching process is super-critical, in the sense of having positive probability of growing indefinitely, if and only if $\lambda\theta/\mu > 1$. The quantity $\lambda\theta/\mu$ has an immediate interpretation, being the lifetime average number of offspring of a single parasite, where offspring is interpreted in terms of parasites successfully passed on to other hosts, and is therefore precisely the biological quantity R_0, as seen from the parasites' viewpoint. Its appearance as the criticality parameter is therefore exactly what one would expect. However, if $\theta > e$, the criticality parameter is $\lambda e \log\theta/\mu$, a fact that is much more difficult to interpret.

Another feature of the model is that the mean number of parasites develops in time with exponential rate $\lambda\theta - \mu$, whereas, if $\theta > e$, a super-critical process has a smaller exponential growth rate for the number of infected hosts. Thus, in such circumstances, the mean number of parasites per host increases ever faster. As a result, because deaths of parasites are counted as transitions, paths containing m_N transitions may contain many fewer infections — roughly speaking, one may well have only $M_N \approx m_N^\alpha$ infections, for some $\alpha < 1$. For such choices of the parameters, this makes the above theorem unsuitable for direct comparison with the results of Ball and Donnelly (1995). We therefore prove a second error bound in Theorem 3.2, which is expressed in terms of the asymptotics of M_N. Its proof turns out to be a relatively simple adaptation of that of Theorem 3.1. We conclude with Theorem 4.2, which establishes a rather stronger local statement, showing that, except on a set of asymptotically negligible probability, the ratio of the likelihoods under the two models of paths containing at most M infections typically differs from 1 by an amount of order $O\left((M^{2/3}/N)\sqrt{\log(N/M^{2/3})}\right)$.

3 Total variation approximation

The Markov branching process $X := (X_j(\cdot),\, j \geq 1)$ that approximates the BK-model is obtained from the process X^N by ignoring the 0-component, taking the countable set $\mathcal{X}^* := \{(\xi_j \in \mathbb{Z}_+,\, j \geq 1)\colon\ \sum_{j\geq 1} \xi_j < \infty\}$ as state space, and modifying the transition rates to

$$\begin{aligned}
\xi &\to \xi + e(j-1) - e(j) \quad \text{at rate} \quad j\mu\xi_j, \quad j \geq 2;\\
\xi &\to \xi - e(1) \quad \text{at rate} \quad \mu\xi_1,\\
\xi &\to \xi + e(j) \quad \text{at rate} \quad \lambda \sum_{l\geq 1} \xi_l p_{lj}, \quad j \geq 1,
\end{aligned}$$

for $\xi \in \mathcal{X}^*$. These rates are identical with those for X^N, except that, in the infection transition, the factor $\xi_0/N = 1 - S(\xi)/N$ is replaced by 1, where $S(\xi) := \sum_{j\geq 1} \xi_j$. This represents the fact that, in the branching approximation, the total proportion of infected hosts is considered to be vanishingly small. Clearly, this should make little difference to individual transitions if $S(\xi) \ll N$. The main result of this paper is to show that it makes little difference even for the distribution of whole path segments, considered as paths in $\mathcal{X}^*$, provided that the number of transitions m in the segment and the initial state $\xi^{(0)} \in \mathcal{X}^*$ are such that $m + S(\xi^{(0)}) \ll N^{2/3}$. We denote such a path by $\{(\xi^{(l)}, t^{(l)}),\, 0 \leq l \leq m\}$, where $t^{(0)} := 0$, and we let $\mathcal{F}_m$ denote the Borel σ-algebra of events generated by these paths. To avoid trivial exceptions caused by paths that are absorbed in $0 \in \mathcal{X}^*$ never making further jumps, we suppose that both processes, when in state 0, make 'jumps' to state 0 at unit rate.

Theorem 3.1 *Suppose that $\xi^{(0)} \in \mathcal{X}^*$, $N \geq 2$ and m are such that $S_m\sqrt{m} < N$, where $S_m := m + S(\xi^{(0)})$. Then, for any $A \in \mathcal{F}_m$, we have*

$$|\mathbb{P}[X^{*N} \in A] - \mathbb{P}[X \in A]| \ \leq\ 8\frac{S_m\sqrt{m}}{N},$$

where X^{*N} denotes the process X^N without the zero coordinate.

Proof. For $\xi \in \mathcal{X}^*$ with $1 \leq S(\xi) \leq N$, write

$$\begin{aligned}
\Lambda^N(\xi) &:= \lambda(1 - S(\xi)/N) \sum_{l\geq 1} \xi_l(1 - p_{l0}), \qquad \rho^N(\xi) \ =\ \mu \sum_{l\geq 1} l\xi_l + \Lambda^N(\xi);\\
\Lambda(\xi) &:= \lambda \sum_{l\geq 1} \xi_l(1 - p_{l0}), \qquad \rho(\xi) \ =\ \mu \sum_{l\geq 1} l\xi_l + \Lambda(\xi).
\end{aligned}$$

The quantities $\rho^N(\xi)$ and $\rho(\xi)$ respectively denote the overall jump rates of the processes X^N and X in state ξ, $\Lambda^N(\xi)$ and $\Lambda(\xi)$ the overall infection rates; for $\xi = 0 \in \mathcal{X}^*$, we set $\rho^N(0) = \rho(0) = \Lambda^N(0) = \Lambda(0) = 1$. Suppose that $m + S(\xi^{(0)}) \leq N$. Then, for a path with m transitions starting in $\xi^{(0)}$ at time 0 and then passing through the sequence of states $(\xi^{(l)}, 1 \leq l \leq m)$ at times $t^{(1)} < t^{(2)} < \cdots < t^{(m)}$, the likelihood ratio $d\mathbb{P}_{X^N}/d\mathbb{P}_X$ evaluated at such a path is just

$$L_m^N \;:=\; L_m^N(\xi^{(\cdot)}, t^{(\cdot)}) \;=\; \prod_{l=0}^{m-1} V_l^N, \tag{3.1}$$

where

$$\begin{aligned} V_l^N \;&:=\; V_l^N(\xi^{(l)}, \xi^{(l+1)}, t^{(l)}, t^{(l+1)}) \\ &:=\; \exp\{-(\Lambda^N(\xi^{(l)}) - \Lambda(\xi^{(l)}))(t^{(l+1)} - t^{(l)})\} \left(1 - \frac{S(\xi^{(l)})}{N}\right)^{u_l}, \end{aligned} \tag{3.2}$$

where

$$u_l \;=\; \begin{cases} 1 & \text{if} \quad S(\xi^{(l+1)}) = S(\xi^{(l)}) + 1; \\ 0 & \text{otherwise.} \end{cases}$$

Hence we have

$$L_{l+1}^N \;=\; L_l^N(1 + \eta_{l1}^N)(1 - \eta_{l2}^N),$$

with

$$\eta_{l1}^N \;:=\; \exp\{-(\Lambda^N(\xi^{(l)}) - \Lambda(\xi^{(l)}))(t^{(l+1)} - t^{(l)})\} - 1,$$

and

$$\eta_{l2}^N \;:=\; \frac{S(\xi^{(l)})}{N} \mathbf{1}\{u_l = 1\}.$$

Note that each of these quantities is zero if $\xi^{(l)} = 0 \in \mathcal{X}^*$.

The inequality

$$0 \;\leq\; \eta_{l2}^N \;\leq\; S(\xi^{(l)})/N \tag{3.3}$$

is immediate. Then, from the definitions of Λ^N and Λ, it follows directly that

$$\Lambda(\xi) - \Lambda^N(\xi) \;=\; \lambda \frac{S(\xi)}{N} \sum_{l \geq 1} \xi_l (1 - p_{l0}),$$

implying that

$$|\Lambda^N(\xi)/\Lambda(\xi) - 1| \;\le\; S(\xi)/N. \tag{3.4}$$

Furthermore, using the inequality $0 \le e^x - 1 \le 2x$ in $0 \le x \le 1$, if

$$\tilde{\eta}_{l1}^N \;:=\; |\Lambda^N(\xi^{(l)})/\Lambda(\xi^{(l)}) - 1|\Lambda(\xi^{(l)})(t^{(l+1)} - t^{(l)}) \;\le\; 1, \tag{3.5}$$

we also have

$$|\eta_{l1}^N| \;\le\; 2\tilde{\eta}_{l1}^N \;\le\; 2\{S(\xi^{(l)})/N\}e_l, \tag{3.6}$$

where

$$e_l \;:=\; \rho(\xi^{(l)})(t^{(l+1)} - t^{(l)}).$$

Hence, if (3.5) is satisfied, it follows that

$$|L_{l+1}^N - L_l^N| \;\le\; N^{-1}S(\xi^{(l)})\{1 + 2e_l\}L_l^N. \tag{3.7}$$

Now suppose that $(X^{(l)},\, l \ge 0)$ is a path resulting from a realization of the process X starting with $X(0) = \xi^{(0)}$, and that $(T^{(l)},\, l \ge 1)$ are the corresponding jump times: set $T^{(0)} = 0$. Then, defining

$$E_{l+1} \;:=\; \rho(X^{(l)})(T^{(l+1)} - T^{(l)}), \qquad l \ge 0, \tag{3.8}$$

we note that $\mathcal{L}(E_{l+1} \,|\, \mathcal{F}_l)$ is the standard exponential distribution for each l. Furthermore, recalling that we take $m + S(\xi^{(0)}) \le N$, the process $\{L_l^N(X^{(\cdot)}, T^{(\cdot)}),$ $0 \le l \le m\}$ is a non-negative martingale with $L_0^N = 1$ a.s. We now modify L^N slightly to a process $\widetilde{L}^N$, in such a way that $\widetilde{L}^N$ remains a martingale with mean 1 and is with high probability identical with L^N for a long time, but whose increments *always* satisfy the inequality

$$|\widetilde{L}_{l+1}^N - \widetilde{L}_l^N| \;\le\; 2N^{-1}S(X^{(l)})\{1 + 2E_l\}. \tag{3.9}$$

For technical reasons, this represents a useful improvement on having (3.7) satisfied with high probability.

To make this modification, it is clear that one should stop L^N as soon as it reaches a value greater than 2. Then we need also to protect against large values of the increment $T^{(l+1)} - T^{(l)}$, which may lead to (3.5) (and hence (3.7)) being violated. To do so, we note that, for any $t_0 > 0$,

$$\begin{aligned}
&\mathbb{E}\{V_l^N(X^{(l)}, X^{(l+1)}, T^{(l)}, T^{(l+1)})\mathbf{1}_{\{T^{(l+1)} - T^{(l)} > t_0\}} \,|\, \mathcal{F}_l\} \\
&\quad= \;\mathbb{P}[T^{(l+1)} - T^{(l)} > t_0]\, \exp\{-(\Lambda^N(X^{(l)}) - \Lambda(X^{(l)}))t_0\} \\
&\qquad\qquad \mathbb{E}\{V_l^N(X^{(l)}, X^{(l+1)}, T^{(l)} + t_0, T^{(l+1)}) \,|\, \mathcal{F}_l \cap \{T^{(l+1)} - T^{(l)} > t_0\}\} \\
&\quad= \;\mathbb{P}[T^{(l+1)} - T^{(l)} > t_0]\, \exp\{-(\Lambda^N(X^{(l)}) - \Lambda(X^{(l)}))t_0\},
\end{aligned}$$

where the last line uses the strong Markov property for X. Hence we may replace $V_l^N(X^{(l)}, X^{(l+1)}, T^{(l)}, T^{(l+1)})$ by $\exp\{-(\Lambda^N(X^{(l)})-\Lambda(X^{(l)}))t_0^{(l)}\}$ on the event $T^{(l+1)} - T^{(l)} > t_0^{(l)} := N/\{S(X^{(l)})\rho(X^{(l)})\}$, without altering its conditional expectation given $\mathcal{F}_l$. Now (3.5), and hence (3.9), are satisfied if $T^{(l+1)} - T^{(l)} \le t_0^{(l)}$, because $\Lambda(X^{(l)}) \le \rho(X^{(l)})$ and from (3.4). However, if $T^{(l+1)} - T^{(l)} > t_0^{(l)}$, it follows from the same observations that

$$0 \;\le\; \exp\{(\Lambda(X^{(l)}) - \Lambda^N(X^{(l)}))t_0^{(l)}\} - 1 \;\le\; e - 1 \;\le\; 2\frac{S(X^{(l)})}{N}E_{l+1},$$

because $T^{(l+1)} - T^{(l)} > t_0^{(l)}$ is equivalent to $E_{l+1}S(X^{(l)})/N > 1$. So if, at the first occasion that $E_{l+1}S(X^{(l)})/N > 1$, we replace V_l^N by

$$\exp\{(\Lambda(X^{(l)}) - \Lambda^N(X^{(l)}))N/(S(X^{(l)})\rho(X^{(l)}))\},$$

and then stop, we preserve (3.9) in this case also.

Hence we define

$$\widetilde{L}_l^N \;:=\; L^N_{l\wedge\tau_1^N\wedge\tau_2^N}(X^{(\cdot)}, T^{(\cdot)})\, C_l^N, \tag{3.10}$$

where

$$C_l^N \;:=\; \begin{cases} 1 & \text{if} \quad \tau_1^N > \min\{l, \tau_2^N\}; \\ \frac{1}{V^N_{\tau_1^N}} \exp\left\{\frac{N(\Lambda(X^{(\tau_1^N-1)})-\Lambda^N(X^{(\tau_1^N-1)}))}{S(X^{(\tau_1^N-1)})\rho(X^{(\tau_1^N-1)})}\right\} & \text{if} \quad \tau_1^N \le \min\{l, \tau_2^N\}, \end{cases}$$

and

$$\begin{aligned} \tau_1^N \;&:=\; \inf\{l \ge 1\colon\; E_l > N/S(X^{(l-1)})\}, \\ \tau_2^N \;&:=\; \inf\{l \ge 1\colon\; L_l^N > 2\}. \end{aligned}$$

We then observe that $\widetilde{L}^N$ is a martingale whose differences satisfy (3.9), and that $L_l^N = \widetilde{L}_l^N$ for all $0 \le l \le \min\{\tau_2^N, (\tau_1^N - 1)\}$. Note also that $S(X^{(l)}) \le S(\xi^{(0)}) + l \le S_m$ for all $0 \le l \le m$, so that, from (3.9),

$$|\widetilde{L}_{l+1}^N - \widetilde{L}_l^N| \;\le\; 2S_m N^{-1}(1 + 2E_l) \quad \text{for all } 0 \le l < m. \tag{3.11}$$

Now, because $\widetilde{L}_l^N$ is a martingale, it follows from (3.11) that

$$\mathbb{E}(\widetilde{L}_m^N - 1)^2 \;\le\; \frac{4mS_m^2}{N^2}\,\mathbb{E}(1 + 2E_1)^2 \;=\; \frac{52mS_m^2}{N^2}. \tag{3.12}$$

Then, for any $A \in \mathcal{F}_m$, we have

$$\begin{aligned}\mathbb{P}[X \in A] - \mathbb{P}[X^{*N} \in A] &= \mathbb{E}\{(1 - L_m^N)\mathbf{1}\{X \in A\}\} \le \mathbb{E}\{(1 - L_m^N)^+\} \\ &\le \mathbb{P}[\tau_1^N \le m] + \mathbb{P}[\{\tau_2^N \le m\} \cap \{\tau_1^N > m\}] + \mathbb{E}(1 - \widetilde{L}_m^N)^+. \quad (3.13)\end{aligned}$$

From the definition of τ_1^N, it is immediate that

$$\mathbb{P}[\tau_1^N \le m] \le me^{-N/S_m} \le 4e^{-2}\,\frac{S_m\sqrt{m}}{N} \quad (3.14)$$

if $S_m\sqrt{m} \le N$. Then we have

$$\mathbb{P}[\{\tau_2^N \le m\} \cap \{\tau_1^N > m\}] \le \mathbb{P}[\widetilde{L}_m^N - 1 > 1] \le \mathbb{E}(\widetilde{L}_m^N - 1)^+. \quad (3.15)$$

Finally, we have

$$\begin{aligned}\mathbb{E}(1 - \widetilde{L}_m^N)^+ + \mathbb{E}(\widetilde{L}_m^N - 1)^+ &= \mathbb{E}|1 - \widetilde{L}_m^N| \\ &\le \sqrt{\mathbb{E}(1 - \widetilde{L}_m^N)^2} \le 2\sqrt{13}S_m\sqrt{m}/N.\end{aligned}$$

It remains to note that $4e^{-2} + 2\sqrt{13} < 8$. □

In general, the bound given in the theorem provides useful information as long as $S_m\sqrt{m} \ll N$. In asymptotic terms, for fixed $\xi^{(0)}$, this allows paths of lengths $m_N = o(N^{2/3})$ as $N \to \infty$, with an error bound of order $O(N^{3\gamma/2-1})$ if $m_N \sim N^\gamma$ for some $\gamma < 2/3$.

If the Ball and Donnelly (1995) coupling is used to obtain error bounds, the resulting order $O(N^{2\gamma-1})$, if $\xi^{(0)}$ is fixed and $M_N \sim N^\gamma$, is at first sight not as sharp. However, there is an important difference between the two results: the theorem above has m_N, the total number of transitions, in the error bound, whereas the Ball and Donnelly coupling leads to an error expressed in terms of M_N, the number of births or infections. Now the total number of transitions includes all the parasite deaths, and if the mean number of parasites per host grows fast, as may be the case when $\theta > e$, m_N may be substantially bigger than M_N. Thus Theorem 3.1 is not strong enough to yield an obvious improvement. For this reason, we now bound the discrepancies in the likelihood ratio more carefully, basing the argument explicitly on the sequence of infection events. To this end, we let $\mathcal{H}_l$ denote the Borel σ-algebra of events generated by paths containing exactly l infection events; as before, to avoid trivial exceptions caused by paths that are absorbed in $0 \in \mathcal{X}^*$ having no further infections, we suppose that both processes, when in state 0, create 'pseudoinfections' at unit rate.

Theorem 3.2 *Suppose that* $\xi^{(0)} \in \mathcal{X}^*$, $N \geq 2$ *and* M *are such that* $S_M\sqrt{M} \leq N$, *where* $S_M := M + S(\xi^{(0)})$. *Then, for any* $A \in \mathcal{H}_M$, *we have*

$$|\mathbb{P}[X^{*N} \in A] - \mathbb{P}[X \in A]| \;\leq\; 8\frac{S_M\sqrt{M}}{N}.$$

Proof. The likelihood ratio at a path $\xi(\cdot)$ in $\mathcal{H}_M$ can be written, using (3.2), in the form

$$\begin{aligned} L'^N_M \;&:=\; L'^N_M(\xi(u),\, 0 \leq u \leq \sigma_M) \\ &:=\; \prod_{l=1}^{M} \left\{ \exp\left(-\int_{\sigma_{l-1}}^{\sigma_l} (\Lambda^N(\xi(u)) - \Lambda(\xi(u)))\, du \right) \left(1 - \frac{S(\xi(\sigma_l))}{N} \right) \right\} \end{aligned} \tag{3.16}$$

where $0 = \sigma_0 < \sigma_1 < \cdots$ denote the times of infection transitions. Hence, very much as before, we have

$$L'^N_l \;=\; L'^N_{l-1}(1 + \eta'^N_{l1})(1 - \eta'^N_{l2}),$$

with

$$\eta'^N_{l1} \;:=\; \exp\left(-\int_{\sigma_{l-1}}^{\sigma_l} (\Lambda^N(\xi(u)) - \Lambda(\xi(u)))\, du \right) - 1,$$

and

$$\eta'^N_{l2} \;:=\; \frac{S(\xi(\sigma_l))}{N},$$

each of these quantities being zero if $\xi^{(l)} = 0 \in \mathcal{X}^*$.

From (3.4), setting

$$\tilde{\eta}'^N_{l1} \;:=\; \int_{\sigma_{l-1}}^{\sigma_l} \{1 - \Lambda^N(\xi(u))/\Lambda(\xi(u))\}\Lambda(\xi(u))\, du \tag{3.17}$$

we have

$$0 \;\leq\; \eta'^N_{l1} \;\leq\; 2\tilde{\eta}'^N_{l1} \;\leq\; 2\{S(\xi(\sigma_{l-1}))/N\}e_l, \tag{3.18}$$

whenever $\tilde{\eta}'^N_{l1} \leq 1$, where

$$e_l \;:=\; \int_{\sigma_{l-1}}^{\sigma_l} \Lambda(\xi(u))\, du.$$

Hence, noting that $S(\xi(u)) \le S_M$ for $0 \le u \le \sigma_M$, if

$$e_l \;=\; \int_{\sigma_{l-1}}^{\sigma_l} \Lambda(\xi(u))\,du \;\le\; N/S_M \tag{3.19}$$

is satisfied, it follows that

$$|L'^N_l - L'^N_{l-1}| \;\le\; N^{-1}S_M\{1+2e_l\}L'^N_{l-1}, \quad 1 \le l \le M. \tag{3.20}$$

Now suppose that $(X(u),\, u \ge 0)$ is a path resulting from a realization of the process X starting with $X(0) = \xi^{(0)}$, and that $(\sigma_l,\, l \ge 1)$ are the corresponding times of births (infection transitions): set $\sigma_0 = 0$. Then, defining

$$E'_l \;:=\; \int_{\sigma_{l-1}}^{\sigma_l} \Lambda(X(u))\,du, \qquad l \ge 1, \tag{3.21}$$

we note that $\mathcal{L}(E'_{l+1}\,|\,\mathcal{H}_l)$ is the standard exponential distribution for each l. We now argue as before using the likelihood ratio martingales $\{L'^N_l(X(\cdot)),\, l \ge 0\}$ and $\tilde{L}'^N_l := L'^N_{l\wedge\tau'^N_1\wedge\tau'^N_2}(X(\cdot))\,C'^N_l$, where

$$\begin{aligned}\tau'^N_1 \;&:=\; \inf\{l \ge 1\colon\, E'_l > N/S_M\},\\ \tau'^N_2 \;&:=\; \inf\{l \ge 1\colon\, L'^N_l > 2\},\end{aligned}$$

and C'^N_l is the analogue of C^N_l in the previous section. Here, arguing much as before, on the event $\sigma_{l+1} > \tau_{l+1}$, where

$$\tau_{l+1} \;:=\; \inf\left\{t \ge \sigma_l\colon\, \int_{\sigma_l}^{t} \rho(\xi(u))\,du \ge N/S_M\right\},$$

we replace

$$\exp\left\{-\int_{\sigma_l}^{\sigma_{l+1}} (\Lambda^N(\xi(u)) - \Lambda(\xi(u)))\,du\right\}\left(1 - \frac{S(\xi(\sigma_l))}{N}\right)$$

by

$$\exp\left\{-\int_{\sigma_l}^{\tau_{l+1}} (\Lambda^N(\xi(u)) - \Lambda(\xi(u)))\,du\right\}$$

without altering the martingale property, and preserving (3.20) on this event. Hence, and since (3.19) is satisfied for all $0 \le l < \tau'^N_1$, it follows from (3.20) and the definition of τ'^N_2 that

$$|\tilde{L}'^N_{l+1} - \tilde{L}'^N_l| \;\le\; 2S_M N^{-1}(1+2E'_{l+1}) \quad \text{for all } 0 \le l < M. \tag{3.22}$$

The remaining argument is now exactly as before, using the martingale $\tilde{L}'_l$ to compare the probabilities $\mathbb{P}[X^{*N} \in A]$ and $\mathbb{P}[X \in A]$ for $A \in \mathcal{H}_M$. $\square$

Thus Theorem 3.2 yields bounds of order $O(N^{3\gamma/2-1})$, improving on the rate obtained using the Ball and Donnelly (1995) coupling, if $\xi^{(0)}$ is fixed and $M_N \sim N^\gamma$ for $\gamma < 3/2$, where M_N denotes the number of infection transitions.

The new argument exploits the fact that the life histories of individuals infected with a given number of parasites have identical distributions in both models, except for the infection events, so that the likelihood ratio is correspondingly simpler. The key element is then that the *difference* in infection rates between the two models is sufficiently small compared to the infection rate itself. The argument in Theorem 3.1 is less precise largely because, if the number of parasites is large, the bound (3.6) is rather pessimistic, since a potentially small factor $\Lambda(\xi)/\rho(\xi)$ is not being exploited.

4 Local approximation

It was argued in Barbour and Utev (2004) that total variation approximation is not necessarily the best measure of closeness, if statistical applications involving likelihoods are to be justified. It is much more natural to want to have local approximations, which ensure that the ratio of actual and approximate likelihood is very close to 1, except possibly on a set of very small probability. As a result, they defined a measure of relative closeness: probability measures P and Q on $\mathcal{F}$ are said to be ε-relatively close with tolerance η, $\mathbf{RC}\,(\varepsilon, \eta)$ for short, if there exists a set $B \in \mathcal{F}$ such that

$$P(B^c) \;\le\; \eta, \quad Q(B^c) \;\le\; \eta, \quad \sup_{x \in B} |\log((dP/dQ)(x))| \;\le\; \varepsilon.$$

In this section, we show that the branching process approximation of the previous sections is indeed relatively close, as long as $M_N \ll N^{2/3}$.

We begin with a minor modification of the bounded differences lemma for martingales.

Lemma 4.1 *If $(L_n, \mathcal{G}_n,\, n \ge 0)$ is a martingale, and if*

$$|L_{n+1} - L_n| \;\le\; a + bE_{n+1} \quad \textit{for each } n \ge 0,$$

where $\mathcal{L}(E_{n+1} \,|\, \mathcal{G}_n)$ *is the standard exponential distribution* $\exp(1)$ *for each* $n \geq 1$*, then*

$$(1): \quad \max\{\mathbb{P}[L_n - L_0 \leq -y], \mathbb{P}[L_n - L_0 \geq y]\} \;\leq\; \exp\left\{\frac{-3y^2}{8n\{(a+b)^2+b^2\}}\right\}$$

for all

$$0 \;\leq\; y \;\leq\; \frac{4n}{3}\varepsilon_0\{(a+b)^2+b^2\}/\max(a,b),$$

where $\varepsilon_0 > 1/15$ *is the constant defined by* $e^{\varepsilon_0}(1-\varepsilon_0)^{-3} = 4/3$*. Furthermore, for all* $y \geq 0$*,*

$$(2): \quad \max\{\mathbb{P}[L_n - L_0 \leq -y], \mathbb{P}[L_n - L_0 \geq y]\} \;\leq\; \exp\left\{\frac{-y}{15\max(a,b)\sqrt{n}} + \frac{2}{135}\right\}.$$

Proof. If X is any random variable with $\mathbb{E}X = 0$ and $|X| \leq a + bE$, where $E \sim \exp(1)$, then it follows that, for any $\theta > 0$,

$$\begin{aligned}
\mathbb{E}\{e^{\theta X}\} &\leq \mathbb{E}\left\{1 + \theta X + \tfrac{1}{2}\theta^2 X^2 e^{\theta|X|}\right\} \\
&\leq 1 + \tfrac{1}{2}\theta^2 \mathbb{E}\left\{(a+bE)^2 e^{\theta(a+bE)}\right\} \\
&= 1 + \tfrac{1}{2}\theta^2 e^{a\theta}\left\{\frac{a^2}{1-b\theta} + \frac{2ab}{(1-b\theta)^2} + \frac{2b^2}{(1-b\theta)^3}\right\} \\
&\leq \exp\left\{\frac{2}{3}\theta^2\{(a+b)^2+b^2\}\right\},
\end{aligned}$$

as long as $\theta\max(a,b) \leq \varepsilon_0$, with ε_0 defined as above. Hence, for any $n \geq 1$,

$$\begin{aligned}
\mathbb{E}\left\{e^{\theta(L_n - L_0)} \,|\, \mathcal{G}_{n-1}\right\} &= e^{\theta(L_{n-1}-L_0)}\mathbb{E}\left\{e^{\theta(L_n - L_{n-1})} \,|\, \mathcal{G}_{n-1}\right\} \\
&\leq \exp\left\{\frac{2}{3}\theta^2\{(a+b)^2+b^2\}\right\} e^{\theta(L_{n-1}-L_0)},
\end{aligned}$$

implying that

$$\mathbb{E}\left\{e^{\theta(L_n - L_0)}\right\} \;\leq\; \exp\left\{\frac{2}{3}\theta^2\{(a+b)^2+b^2\}\right\}\mathbb{E}\left\{e^{\theta(L_{n-1}-L_0)}\right\}$$

for all $n \geq 1$, and hence that

$$\mathbb{E}\left\{e^{\theta(L_n - L_0)}\right\} \;\leq\; \exp\left\{\frac{2}{3}n\theta^2\{(a+b)^2+b^2\}\right\}.$$

Hence, for any $y \geq 0$ and any θ such that $\theta \max(a,b) \leq \varepsilon_0$, we have

$$\mathbb{P}[L_n - L_0 \geq y] \;\leq\; \exp\left\{-y\theta + \frac{2}{3}n\theta^2\{(a+b)^2 + b^2\}\right\}.$$

Now, if $y \max(a,b) \leq (4\varepsilon_0/3)n\{(a+b)^2 + b^2\}$, we can take

$$\theta \;=\; \frac{y}{(4n/3)\{(a+b)^2 + b^2\}},$$

to give

$$\mathbb{P}[L_n - L_0 \geq y] \;\leq\; \exp\left\{\frac{-3y^2}{8n\{(a+b)^2 + b^2\}}\right\}.$$

On the other hand, for all $y \geq 0$ and $n \geq 1$, we can choose $\theta = 1/\{15\max(a,b)\sqrt{n}\}$, giving

$$\mathbb{P}[L_n - L_0 \geq y] \;\leq\; \exp\left\{\frac{-y}{15\max(a,b)\sqrt{n}} + \frac{2}{135}\right\}.$$

The same arguments also cover $\mathbb{P}[L_n - L_0 \leq -y]$ for the corresponding choices of y, since the conditions of the theorem apply equally well to the martingale $-L_n$. □

This lemma enables us to prove the following estimate of relative closeness.

Theorem 4.2 *Suppose that $\xi^{(0)} \in \mathcal{X}^*$, $N \geq 2$ and M are such that $\psi(M,N) := S_M\sqrt{M}/N \leq 1$, where $S_M := M + S(\xi^{(0)})$. Then, with respect to paths in $\mathcal{H}_M$, the processes X^{*N} and X are $\mathbf{RC}\,(\varepsilon^r_{M,N}, \eta^r_{M,N})$ relatively close for any choice of $r \geq 1$, where*

$$\begin{aligned}
\varepsilon^r_{M,N} &:= C_r\psi(M,N)\sqrt{\log(1/\psi(M,N))};\\
\eta^r_{M,N} &:= 2\psi(M,N)^r + e^{2/135}\exp\{-1/60\psi(M,N)\} + Me^{-N/S_M},
\end{aligned}$$

and $C_r := \sqrt{416r/3}$, provided that $M \geq (1/5)C_r^2 \log N$ and that $\varepsilon^r_{M,N} \leq 1$.

Proof. It was shown in the proof of Theorem 3.2 that the likelihoods of the processes X^{*N} and X are close; here, we tighten the argument. We start from (3.22), which states that

$$|\tilde{L}'^N_{l+1} - \tilde{L}'^N_l| \;\leq\; 2S_M N^{-1}(1 + 2E'_{l+1}) \quad \text{for all } 0 \leq l < M,$$

where $\mathcal{L}(E'_{l+1} \,|\, \mathcal{H}_l)$ is the standard exponential distribution for each l, and from the observation that, by the definition of $\tilde{L'}_M^N$, we have $\tilde{L'}_M^N = L'_M^N$ as long as $\min\{\tau'_1^N, \tau'_2^N\} > M$. Now it is immediate, as for (3.14), that

$$\mathbb{P}[\tau'_1^N \le M] \;\le\; Me^{-N/S_M}.$$

Then, from the definition of τ'_2^N, it follows that

$$\mathbb{P}[\{\tau'_1^N > M\} \cap \{\tau'_2^N \le M\}] \;\le\; \mathbb{P}[\tilde{L'}_M^N - 1 > 1].$$

Hence, to establish the desired relative closeness, we take

$$B^c \;:=\; \{\min(\tau'_1^N, \tau'_2^N) \le M\} \cup \{|\tilde{L'}_M - 1| > \varepsilon^r_{M,N}/2\},$$

(here using the assumption that $\varepsilon^r_{M,N} \le 1$) and bound the probabilities $\mathbb{P}[\tilde{L'}_M^N - 1 > 1]$ and $\mathbb{P}[|\tilde{L'}_M - 1| > \varepsilon^r_{M,N}/2]$ using Lemma 4.1 with $n = M$ and $2a = b = 4S_M/N$.

First, we use Lemma 4.1 (2) to give

$$\mathbb{P}[\tilde{L'}_M^N - 1 > 1] \;\le\; \exp\{-N/(60S_M\sqrt{M})\}e^{2/135}.$$

Then we use Lemma 4.1 (1) to show that

$$\mathbb{P}[|1 - \tilde{L'}_M^N| > y] \;\le\; 2\exp\{-3N^2y^2/(416MS_M^2)\},$$

provided that

$$0 \;\le\; y \;\le\; \frac{4M}{3}\,\varepsilon_0\,\frac{13S_M}{N}.$$

Hence we can take $y = \varepsilon^r_{M,N}/2$ if

$$C_r S_M\sqrt{M}\sqrt{\log N}/N \le 104MS_M/45N,$$

and thus if $M \ge (1/5)C_r^2 \log N$, giving

$$\mathbb{P}[|1 - \tilde{L'}_M^N| > \varepsilon^r_{M,N}/2] \;\le\; 2\left\{\frac{S_M\sqrt{M}}{N}\right\}^{(3/416)C_r^2} \;=\; 2\left\{\frac{S_M\sqrt{M}}{N}\right\}^{r}. \quad \square$$

Thus asymptotic relative closeness of order $O\{\psi(M,N)\sqrt{\log(1/\psi(M,N))}\}$ can be established with tolerance of arbitrarily small polynomial order in $\psi(M,N) = S_M\sqrt{M}/N$.

References

[1] F. G. Ball (1983) The threshold behaviour of epidemic models. *J. Appl. Probab.* **20**, 227–241.

[2] F. G. Ball & P. Donnelly (1995) Strong approximations for epidemic models. *Stoch. Procs Applics.* **55**, 1–21.

[3] A. D. Barbour & M. Kafetzaki (1993) A host–parasite model yielding heterogeneous parasite loads. *J. Math. Biol.* **31**, 157–176.

[4] A. D. Barbour & S. Utev (2004) Approximating the Reed-Frost epidemic process. *Stoch. Procs Applics* **113**, 173–197.

[5] M. S. Bartlett (1956) *An introduction to stochastic processes.* Cambridge University Press.

[6] O. Diekmann & J. A. P. Heesterbeek (2000) *Mathematical epidemiology of infectious diseases.* Wiley, New York.

[7] J. A. P. Heesterbeek (1992) R_0. CWI Amsterdam.

[8] D. G. Kendall (1956) Deterministic and stochastic epidemics in closed populations. *Proc. Third Berk. Symp. Math. Stat. Probab.* **4**, 149–165.

[9] T. G. Kurtz (1976) Limit theorems and diffusion approximations for density dependent Markov chains. *Mathematical Programming Study* **5**, 67–78.

[10] T. G. Kurtz (1981) *Approximation of population processes.* CBMS-NSF Regional Conference Series in Applied Mathematics 36, SIAM, Philadelphia.

[11] C. J. Luchsinger (2002a) Stochastic models of a parasitic infection, exhibiting three basic reproduction ratios. *J. Math. Biol.* **42**, 532–554.

[12] C. J. Luchsinger (2002b) Approximating the long-term behaviour of a model for parasitic infection. *J. Math. Biol.* **42**, 555–581.

[13] P. Whittle (1955) The outcome of a stochastic epidemic—a note on Bailey's paper. *Biometrika* **42**, 116–122.