Author's personal copy

Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 52 (2008) 1281–1291
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Non-parametric Bayesian estimation for multitype branching processes through simulation-based methods

M. González, J. Martín, R. Martínez*, M. Mota

Department of Mathematics, University of Extremadura, 06071 Badajoz, Spain

Received 28 February 2006; received in revised form 23 May 2007; accepted 12 June 2007
Available online 16 June 2007

Abstract

The problem of statistical inference from a Bayesian outlook is studied for the multitype Galton–Watson branching process, considering a non-parametric framework. The only data assumed to be available are each generation's population size vectors. The Gibbs sampler is used in estimating the posterior distributions of the main parameters of the model, and the predictive distributions for as yet unobserved generations. The algorithm provided is independent of whether the process becomes extinct or not. The method is illustrated with simulated examples.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Multitype Galton–Watson branching process; Non-parametric Bayesian inference; Gibbs sampler

1. Introduction

Branching processes are stochastic models which arise in the context of population dynamics. They are the useful models to describe the extinction/growth of populations (see Haccou et al., 2005). Branching models have been applied to biological problems in such fields as epidemiology, genetics, and cell dynamics, for example, including the evolution of infectious diseases (e.g., Mode and Sleemam, 2000; Ball et al., 2004), sex-linked genes (e.g., González et al., 2006), and stem cells (e.g., Yakovlev and Yanev, 2006). Further examples of the application of branching processes in biology are reviewed in the recent monographs by Kimmel and Axelrod (2002) and Pakes (2003). There have also been significant applications of these models in other areas such as algorithm and data structures (e.g., Montpetit et al., 1992; Devroye, 1998) and economics (e.g., Dion and Epps, 1999).

We shall here focus on the multitype Galton–Watson branching process, adopting the nomenclature of population dynamics, although, as was noted above, the results are applicable to many other fields. This model is a generalization of the classical Galton–Watson process used to describe the evolution of populations where several types of individuals coexist, and each specimen, independently of the rest, is able to give rise to offspring of any type. Consequently, the descendence of the individuals of a type is represented by random vectors with a common probability distribution which is referred as the *offspring distribution* or *law* of this type. The main properties of the multitype Galton–Watson branching process have been widely studied (see, for example, Mode, 1971). Much of the interest of these studies is

* Corresponding author. Tel.: +34 924289300x9141; fax: +34 924272911.
E-mail address: rmartinez@unex.es (R. Martínez).

0167-9473/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2007.06.008

Author's personal copy

focused on the extinction probability, specifically on determining whether this probability is equal to unity, in which case extinction is certain.

The key parameter in determining whether or not extinction is certain is the real maximum-modulus eigenvalue, ρ, of the mean matrix, $M = (m_{ij})$, whose element (i, j) is the mean number of type j descendants given rise to by a type i specimen. Roughly speaking, one can say that extinction occurs with probability 1 if and only if $\rho \leqslant 1$. Otherwise the process grows exponentially. Nevertheless, while ρ is the parameter of prime interest in classifying the process, a full characterization is only possible through the offspring distributions.

Hence, the focus of inferential studies of multitype branching processes has been on M, ρ, and the offspring distributions, under the assumption that data from the first N generations are available. Work from a frequentist viewpoint includes that of Asmussen and Keiding (1978) and Nanthi (1982). Both provide estimators of such parameters together with their asymptotic distributions, conditional on non-extinction of the process. The assumption $\rho > 1$ is implicit in this approach, i.e., the process must not become extinct for the estimators to be useful. It is not possible, therefore, to decide on the basis of these estimators whether the process will become extinct almost surely ($\rho \leqslant 1$) or will have a possibility of survival ($\rho > 1$).

Bayesian inference, on the contrary, provides an appropriate framework in which to study this problem. The posterior distribution of the parameters of interest provides enough information not only to decide about the extinction or survival of the process, but also to infer the most probable values of those parameters. For the multitype Galton–Watson branching process, a Bayesian analysis could be made by generalizing the methods that Guttorp (1991) and more recently Mendoza and Gutiérrez-Peña (2000) have put forward for the simple Galton–Watson process. Following these authors and with no additional hypotheses, i.e., using a non-parametric framework, given a proper conjugate prior, one can obtain the posterior distributions of the offspring laws in a closed form, irrespective of whether the process becomes extinct or not. There is no closed form for the posterior distributions of the parameters M or ρ, which usually must be approximated via Monte Carlo methods.

From a practical point of view, however, there is one serious drawback in both the frequentist and the Bayesian approaches for working in a non-parametric framework, the whole family tree up to some given generation must be observed. That is, for every element k in the support of an offspring distribution, the number of individuals of the corresponding type having vector k of descendants must be available for each generation. These variables will be denoted here by $Z_i(n, k)$, with n standing for the generation and i the type of progenitor. In most populations, it is not possible to observe these variables, only the vector of the total number of individuals of the different types in each generation.

In the present work, we study the problem of Bayesian inference for the main parameters of a multitype branching processes, i.e., the offspring distributions, M, and ρ, considering a non-parametric framework. Unlike other inferential studies of multitype branching processes, the only data we assume to be available is the vector of the population size in each generation, whereas the variables $Z_i(n, k)$ are considered unobservable. We make up for this lack of information by using the Gibbs sampler to approximate for each generation n the distribution of the variables $Z_i(n, k)$, with k taking values in the support of the offspring distribution of type i. At the same time, we approximate the posterior distributions of the offspring laws. Once this first objective has been attained, it will not be difficult to at least approximately obtain the posterior distributions of the mean matrix M and its real maximum-modulus eigenvalue ρ, as well as the predictive distribution for as yet unobserved generations. This algorithm is independent of whether or not the process becomes extinct, and consequently provides a rule for deciding on this dichotomy even though one has only a very small data set.

The work is organized into four sections and an Appendix. The multitype Galton–Watson branching process is defined mathematically in Section 2. Section 3 presents the Gibbs-sampler-based algorithm and the inferential procedures for the parameters of interest and the unobserved generations. In Section 4 we illustrate the accuracy of the methods by means of simulated examples, and discuss some technical aspects of the procedure such as some measurements of the convergence and efficiency of the algorithm, a sensitivity analysis of the prior parameters, and information concerning the computational complexity of the statistical procedure. Finally, Appendix presents the proof of a theoretical result needed for the development of the algorithm.

2. The probability model

Mathematically, the *multitype Galton–Watson branching process* (MP) is a sequence $\{Z(n)\}_{n \geqslant 0}$ of m-dimensional vectors, with m the number of different types of individuals coexisting in the population. One defines $Z(0)$ to be an

Author's personal copy

arbitrary fixed vector, and for $n = 0, 1, \ldots$

$$Z(n+1) = (Z_1(n+1), \ldots, Z_m(n+1)) = \sum_{i=1}^{m} \sum_{j=1}^{Z_i(n)} Y_{ij}(n) \tag{1}$$

($\sum_1^0 = \mathbf{0}$, the null vector), where $\{Y_{ij}(n): i = 1, \ldots, m;\ j = 1, 2, \ldots;\ n = 0, 1, \ldots\}$ is a sequence of non-negative integer-valued independent m-dimensional random vectors. Moreover, these vectors are identically distributed for each type. The offspring distribution for a given type i, $i = 1, \ldots, m$, is denoted by $p_i = (p_{ik} : k \in \mathscr{S}_i)$ with $\mathscr{S}_i$ standing for the support, i.e., for a fixed vector k, $p_{ik} = \Pr(Y_{ij}(n) = k)$ for every $n = 0, 1, \ldots,\ j = 1, 2, \ldots$. Intuitively, p_{ik} represents the probability that a type i individual has offspring given by the vector k, and $Z_i(n)$ represents the number of individuals (specimens) of type i in the nth generation.

The mean matrix $M=(m_{ij})_{i,j=1,\ldots,m}$, described intuitively in the previous section, is defined by $m_{ij}=E[Z_j(1)|Z(0)=\varepsilon_i]$, where $\varepsilon_i = (\delta_{i1}, \ldots, \delta_{im})$, with δ_{ij} being Kronecker's delta. In terms of this matrix, MPs have been classified as positively regular, periodic, or reducible, and, in terms of the magnitude of their real maximum-modulus eigenvalue ρ, they are said to be subcritical if $\rho < 1$, critical if $\rho = 1$, and supercritical if $\rho > 1$ (see Mode, 1971).

It is immediate that $Z(n) = \mathbf{0}$ for some n implies $Z(n + r) = \mathbf{0}$ for every $r \geqslant 0$. One thus defines the extinction probability

$$q = \Pr(Z(n) = \mathbf{0} \text{ for some } n = 1, 2, \ldots).$$

As indicated in Section 1, under certain conditions that we shall omit for simplicity, the following result is significant in determining the extinction probability as well as clearly showing the importance of the parameter ρ:

$$q < 1 \text{ if and only if } \rho > 1.$$

Finally, the variables $Z_i(n, k)$, $i = 1, \ldots, m$, $n \geqslant 0$, and $k \in \mathscr{S}_i$, are defined by

$$Z_i(n, k) = \sum_{j=1}^{Z_i(n)} I_{\{Y_{ij}(n)=k\}},$$

with I_A standing for the indicator function of the set A. Recall that, intuitively, $Z_i(n, k)$ represents the number of type i individuals in generation n whose offspring vector is k.

3. Using the Gibbs sampler to approximate posterior distributions

As noted in Section 1, non-parametric inference in branching processes from both the Bayesian and the frequentist standpoints requires a great amount of information for the estimation of the main parameters, particularly the offspring law. Specifically, for MPs, there is the usually unrealistic requirement that the variables $Z_i(n, k)$ must be observed. Restricting ourselves to the Bayesian standpoint, we describe an algorithm based on the Gibbs sampler to approximate the posterior offspring distribution of a MP only requiring to observe $\mathscr{Z}(N) = \{Z(0), \ldots, Z(N)\}$. Using this algorithm and Monte Carlo method, we also approximate the distribution of $Z(N + r)$ for any $r > 0$, and the posterior distribution of the real maximum-modulus eigenvalue, ρ.

Let $p = (p_1, \ldots, p_m)$ denote the offspring distribution. Our aim is to approximate the posterior distribution of p after observing $\mathscr{Z}(N)$, denoted by $p|\mathscr{Z}(N)$. There is no closed form expression for this distribution. To implement the Gibbs sampler, we consider an augmented parameter vector $(p, \mathscr{Z}(N, \mathscr{S}))$, where $\mathscr{Z}(N, \mathscr{S})$ is the vector $(Z_i(n, k); n = 0, \ldots, N - 1, k \in \mathscr{S}_i, i = 1, \ldots, m)$. The unobserved vector $\mathscr{Z}(N, \mathscr{S})$ is considered a latent vector and is required to make inferences about the offspring laws. It is easy to verify that this vector is related to $\mathscr{Z}(N)$, for every $i = 1, \ldots, m$, by the expressions

$$Z_i(n) = \sum_{k \in \mathscr{S}_i} Z_i(n, k) \quad \text{and} \quad Z(n+1) = \sum_{i=1}^{m} \sum_{k \in \mathscr{S}_i} k Z_i(n, k), \quad n = 0, \ldots, N - 1. \tag{2}$$

Author's personal copy

Therefore, it is only necessary to know the full conditional posterior distributions

$$p|(\mathscr{Z}(N), \mathscr{Z}(N, \mathscr{S})) \quad \text{and} \quad \mathscr{Z}(N, \mathscr{S})|(\mathscr{Z}(N), p).$$

Notice that, by (2), the posterior distribution $p|(\mathscr{Z}(N), \mathscr{Z}(N, \mathscr{S}))$ is the same as $p|\mathscr{Z}(N, \mathscr{S})$. Since we are approaching the problem from a Bayesian perspective, the Dirichlet distributions (the natural conjugate family) are a convenient class of prior distributions of p_i, for each $i = 1, \ldots, m$. We therefore take p_i to follow a Dirichlet distribution of parameter $\alpha_i = (\alpha_{ik}, k \in \mathscr{S}_i)$, $\alpha_{ik} > 0$, i.e., its density function has the form

$$f(p_i) \propto \prod_{k \in \mathscr{S}_i} p_{ik}^{\alpha_{ik}-1}.$$

Since individuals of different types reproduce independently, we thus assume that

$$f(p) \propto \prod_{i=1}^{m} \prod_{k \in \mathscr{S}_i} p_{ik}^{\alpha_{ik}-1}.$$

Furthermore, since the likelihood function of p based on $\mathscr{Z}(N, \mathscr{S})$ satisfies

$$L(p|\mathscr{Z}(N, \mathscr{S})) \propto \prod_{n=0}^{N} \prod_{i=1}^{m} \prod_{k \in \mathscr{S}_i} p_{ki}^{Z_i(n,k)}$$

(for $m = 1$ see Mendoza and Gutiérrez-Peña, 2000), it follows that, for each $i = 1, \ldots, m$, the distribution of $p_i|\mathscr{Z}(N, \mathscr{S})$ is also Dirichlet of parameter $\alpha_i(\mathscr{Z}(N, \mathscr{S})) = (\alpha_{ik} + \sum_{n=0}^{N-1} Z_i(n, k), k \in \mathscr{S}_i)$, i.e.,

$$f(p_i|\mathscr{Z}(N, \mathscr{S})) \propto \prod_{k \in \mathscr{S}_i} p_{ik}^{\alpha_{ik}+\sum_{n=0}^{N-1} Z_i(n,k)-1},$$

and

$$f(p|\mathscr{Z}(N, \mathscr{S})) \propto \prod_{i=1}^{m} \prod_{k \in \mathscr{S}_i} p_{ik}^{\alpha_{ik}+\sum_{n=0}^{N-1} Z_i(n,k)-1}.$$

With respect to $\mathscr{Z}(N, \mathscr{S})|(\mathscr{Z}(N), p)$, since individuals reproduce independently and the multitype Galton–Watson branching process is a Markov chain, one has the following result (the proof is given in Appendix):

$$P(\mathscr{Z}(N, \mathscr{S})|(\mathscr{Z}(N), p)) = \prod_{n=0}^{N-1} P(Z(n, \mathscr{S})|(Z(n), Z(n+1), p)),$$

where $P(\cdot|\cdot)$ denotes the conditional distribution, and $Z(n, \mathscr{S})$ is the vector $(Z_i(n, k), k \in \mathscr{S}_i, i = 1, \ldots, m)$. From (2) and the definition of $Z_i(n, k)$, it is easy to see that, for each $i = 1, \ldots, m$, $(Z_i(n, k), k \in \mathscr{S}_i)|(Z(n), p)$ follows a multinomial distribution with size $Z_i(n)$ and probabilities p_i. One also has that

$$P(Z(n, \mathscr{S})|(Z(n), Z(n+1), p)) \propto P((Z(n+1), Z(n, \mathscr{S}))|(Z(n), p)),$$

where $P((Z(n+1), Z(n, \mathscr{S}))|(Z(n), p))$ is equal to

$$\begin{cases} 0 & \text{if } Z(n+1) \neq \sum_{i=1}^{m} \sum_{k \in \mathscr{S}_i} k Z_i(n, k), \\ \prod_{i=1}^{m} P(Z_i(n, k), k \in \mathscr{S}_i|(Z(n), p)) & \text{if } Z(n+1) = \sum_{i=1}^{m} \sum_{k \in \mathscr{S}_i} k Z_i(n, k). \end{cases}$$

Therefore, to sample from $\mathscr{Z}(N, \mathscr{S})|(\mathscr{Z}(N), p)$ it is enough to sample, for each $n = 0, \ldots, N-1$, from $(Z(n+1), Z(n, \mathscr{S}))|(Z(n), p)$, and, as we have shown, this can be done by conveniently normalizing the probabilities given by the multinomial distribution with size $Z_i(n)$ and probabilities p_i, for every $i = 1, \ldots, m$. Notice that the smaller the cardinals of the sets $\mathscr{S}_i$ and the population size, the computationally faster the sampling.

Author's personal copy

In summarized form, given $\mathscr{S}_i$ and α_i, for every $i = 1, \ldots, m$, by means of the Gibbs sampler, one calculates a sequence of vectors $\{p^{(l)}\}_{l \geqslant 0}$ as follows. The initial $p^{(0)}$ is obtained by sampling, for each $i = 1, \ldots, m$, from the Dirichlet distribution with parameters α_i (prior distribution), and then applying the following algorithm:

Fixed $p^{(0)}$
Do $l = 1$
Generate $\mathscr{Z}^{(l)}(N, \mathscr{S}) \sim \mathscr{Z}(N, \mathscr{S})|(\mathscr{Z}(N), p^{(l-1)})$
Generate $p^{(l)} \sim p|\mathscr{Z}^{(l)}(N, \mathscr{S})$
Do $l = l + 1$

For a run of the sequence $\{p^{(l)}\}_{l \geqslant 0}$, one chooses $Q+1$ vectors of the form $\{p^{(L)}, p^{(L+G)}, \ldots, p^{(L+QG)}\}$, where G is a batch size. Since $\alpha_{ik} > 0$ for all $k \in \mathscr{S}_i$ and $i = 1, \ldots, m$, from the properties of the Dirichlet and multinomial distributions, one deduces that the Markov chain $\{p^{(l)}\}_{l \geqslant 0}$ is ergodic. Therefore the vectors $\{p^{(L)}, p^{(L+G)}, \ldots, p^{(L+QG)}\}$ are considered independent samples from $p|\mathscr{Z}(n)$ if G and L are large enough (see Tierney, 1994). Since these vectors could be affected by the initial state $p^{(0)}$, the algorithm is applied T times, yielding a final sample of length $T(Q+1)$. With this sample, one approximates the distribution function of $p|\mathscr{Z}(N)$ by means of the empirical distribution function.

Remark 1. The foregoing algorithm allows to perform Bayesian inference on the offspring laws based on the observation of just the population size in each generation. Notice that the algorithm applies even though one only observes the total number of parents of each type ($\sum_{n=0}^{N-1} Z(n)$) and the total number of descendants of each type ($\sum_{n=1}^{N} Z(n)$). Indeed, the posterior distribution $p|(\sum_{n=0}^{N-1} Z_i(n,k), k \in \mathscr{S}_i, i = 1, \ldots, m)$ is proportional to the product of Dirichlet distributions with parameters $\alpha_i(\mathscr{Z}(N, \mathscr{S}))$. The transition distribution $(\sum_{n=0}^{N-1} Z_i(n,k), k \in \mathscr{S}_i, i = 1, \ldots, m)|(\sum_{n=0}^{N-1} Z(n), \sum_{n=1}^{N} Z(n), p)$ is also determined by taking into account the probabilities given by the multinomial distribution with size $\sum_{n=0}^{N-1} Z_i(n)$ and probabilities p_i, for every $i = 1, \ldots, m$, and the constraint $\sum_{i=1}^{m} \sum_{k \in S_i} \sum_{n=0}^{N-1} k Z_i(n,k) = \sum_{n=1}^{N} Z(n)$. Also notice, however, that although from a practical viewpoint this is a better situation because it requires less information, computationally the algorithm takes longer to run.

As was noted above, the magnitude of the real maximum-modulus eigenvalue ρ of M plays a major role in the behaviour of the extinction of the process. Therefore, when $\mathscr{Z}(N)$ is observed, the posterior distribution function of $M|\mathscr{Z}(N)$ is approximated from $p|\mathscr{Z}(N)$ and then $\rho|\mathscr{Z}(N)$ from $M|\mathscr{Z}(N)$, using Monte Carlo method and the empirical distribution function. This procedure leads us to adopt the following rule based on $\Pr(\rho \leqslant 1|\mathscr{Z}(N))$ for deciding whether or not the population becomes extinct almost surely: if this value is greater than or equal to 0.5 one decides for extinction and otherwise for exponential growth.

Regardless of whether or not the population eventually becomes extinct almost surely, it is also of interest to forecast the population sizes in future generations. Therefore, for a fixed $r > 0$, we approximate the predictive distribution $Z(N+r)|\mathscr{Z}(N)$ by applying the Monte Carlo method and the empirical distribution function, simulating r generations of a MP starting from $Z(N)$ and with reproduction law sampled from $p|\mathscr{Z}(N)$. In the sense that this algorithm only requires $\mathscr{Z}(N)$ to be observed instead of $\mathscr{Z}(N, \mathscr{S})$, it improves that proposed in Mendoza and Gutiérrez-Peña (2000) for the classical Galton Watson branching process.

4. Some simulated examples

In this section, we shall describe the application of the above algorithm to simulated data. To this end, we considered a MP with $m = 2$ and supports $\mathscr{S}_i = \{(0,0), (0,1), (1,0), (1,1)\}$, for each of $i = 1, 2$, for three different cases: one subcritical with $\rho = 0.8$, one critical, and one supercritical with $\rho = 1.13$. For these models, we simulated 10 generations starting with $Z(0) = (2, 0)$. Table 1 presents the population sizes obtained.

Notice that the data for the three processes are very similar, and it would be difficult to determine at a glance whether extinction will occur on the basis of these observations. Assuming that there was no prior information available, in all the cases we considered Dirichlet prior distributions with parameters $\alpha_i = (\alpha_{ik} = \frac{1}{2}, k \in \mathscr{S}_i)$, $i = 1, 2$, as is suggested in Berger and Bernardo (1992). We then applied the algorithm of the previous section, taking $L = 1000$, $G = 10$, and $Q = T = 100$. As a test of the convergence of the resulting probabilities to the stationary distribution, Figs. 1–3 show the Gelman–Rubin–Brooks diagnostic plots for $p = (p_1, p_2)$ in each of the three cases. For simplicity

Author's personal copy

Table 1
Simulated data

Case	Subcritical		Critical		Supercritical	
n	$Z_1(n)$	$Z_2(n)$	$Z_1(n)$	$Z_2(n)$	$Z_1(n)$	$Z_2(n)$
1	1	2	2	1	0	2
2	2	1	2	1	2	1
3	1	2	1	1	1	1
4	2	0	0	1	1	2
5	1	2	1	0	2	2
6	1	3	1	0	3	1
7	1	2	1	1	1	3
8	3	1	1	1	2	3
9	1	0	1	2	4	2
10	0	1	0	2	1	2

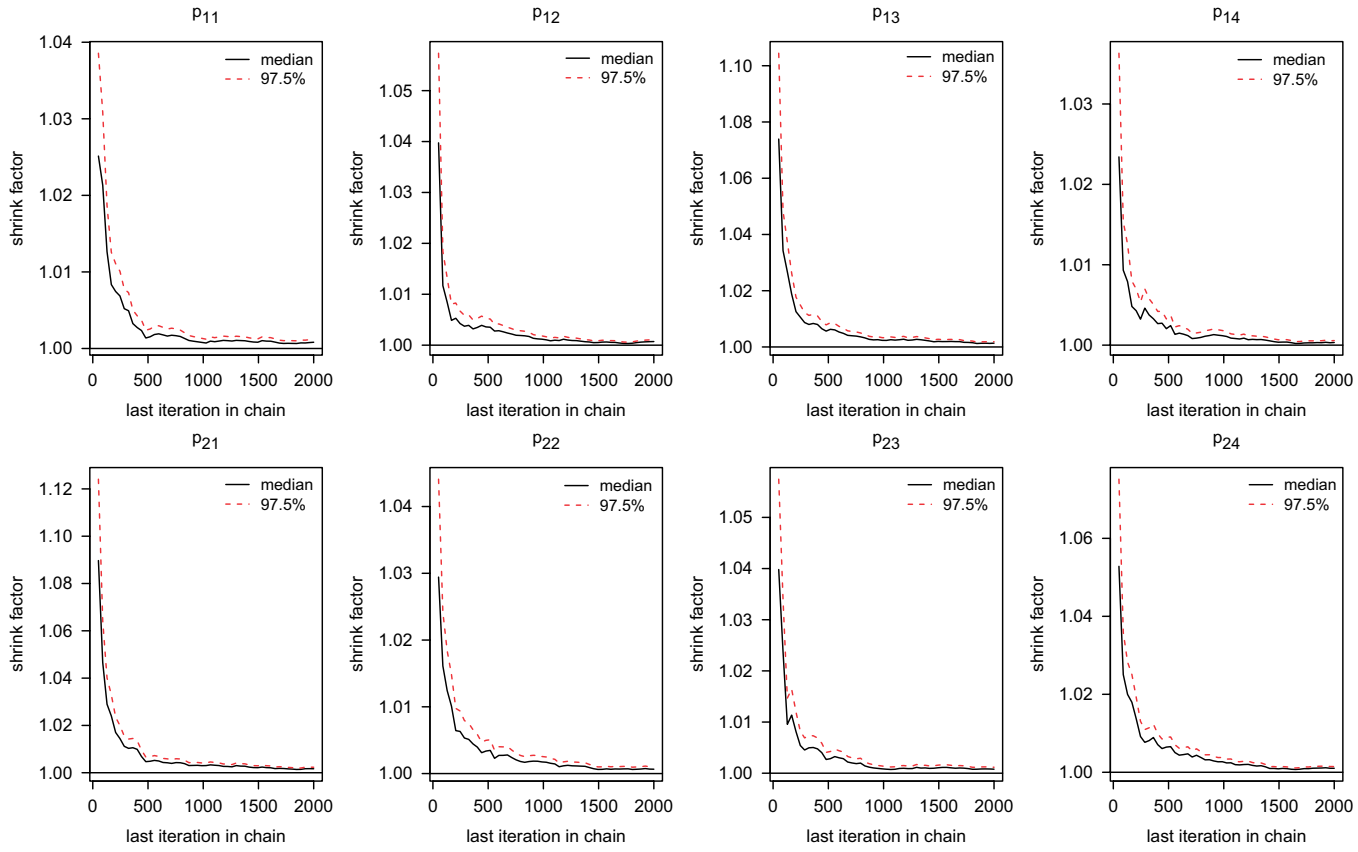

Fig. 1. Subcritical case: Gelman–Rubin–Brooks diagnostic plots.

of notation, we have used $1, \ldots, 4$ to label the states of $\mathcal{S}_i = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$, $i = 1, 2$, so that, for example, $p_{12} = \Pr(Y_{1j}(n) = (0, 1))$. Table 2 gives the estimated potential scale reduction factor together with an upper limit. Because the estimated scale reduction factors are close to unity, they suggest that further simulations will not improve the values of the listed scalar estimators (see Gelman and Rubin, 1992; Brooks and Gelman, 1998, and more recently El Adlouni et al., 2006). Table 2 also gives the autocorrelation values for iterations 1001–2000 at lags 1 and 10. The lag-1 autocorrelation values indicate that a batch procedure is necessary, and the lag-10 values that the batch size $G = 10$ is sufficient.

Author's personal copy

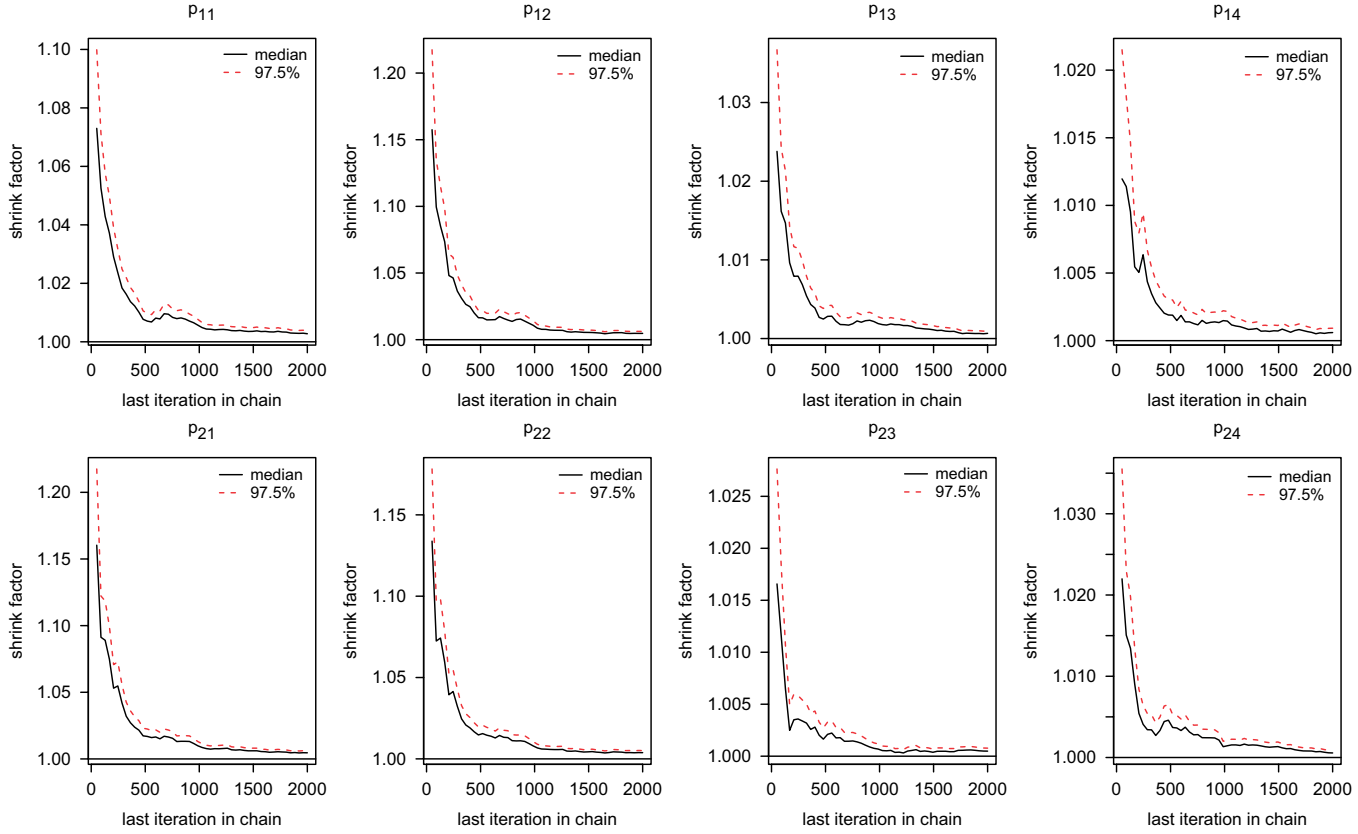

Fig. 2. Critical case: Gelman–Rubin–Brooks diagnostic plots.

To evaluate the algorithm's efficiency, we extracted some summary statistics for the posterior distribution of ρ in the subcritical, critical, and supercritical cases (Table 3). Note that, due to the batch procedure, the time-series standard errors (TSSE) are very close to the Monte-Carlo standard errors (MCSE). Also, in all three cases, the standard errors (MCSE and TSSE) are less than 5% of the posterior standard deviation (SD), indicative of the number of observations considered seeming to be a reasonable choice.

Fig. 4 shows the empirical posterior distribution functions of ρ. From these, one can approximate $\Pr(\rho \leqslant 1 | \mathscr{Z}(N))$ by 0.61, 0.54, and 0.33, for the subcritical, critical, and supercritical cases, respectively. Applying the rule proposed in the previous section, one concludes that the population becomes extinct almost surely for the subcritical and critical cases, but could grow exponentially for the supercritical case. As is usual in branching process theory, the critical case is special, as is manifest in our context in the value of $\Pr(\rho \leqslant 1 | \mathscr{Z}(N))$ close to 0.5.

Finally, Fig. 5 illustrates the predictive distribution of the total number of individuals in the 11th generation. As expected, the predicted total population size for the subcritical case is less than for the critical case, which in turn is less than for the supercritical case.

In these simulated examples, we assumed a flat prior distribution. In order to study the influence of the prior parameters, we performed a discrete sensitivity analysis. Specifically, we were interested in determining how $\Pr(\rho \leqslant 1 | \mathscr{Z}(N))$ changes. In the subcritical case, it is important to establish when $\Pr(\rho \leqslant 1 | \mathscr{Z}(N))$ becomes less than 0.5. Fundamentally, this probability will decline if, for each individual's type, one increases the weight assigned to the state (1, 1). We therefore re-ran the simulation for several values of $\alpha_{1,(1,1)}$, $\alpha_{2,(1,1)}$. For simplicity of exposition, we here consider $\alpha_{1,(1,1)} = \alpha_{2,(1,1)}$ and denote it by α. Fig. 6 shows the evolution of $\Pr(\rho \leqslant 1 | \mathscr{Z}(N))$ as a function of α. As can be seen, in the subcritical case, the change occurs at around $\alpha = 0.95$, so that one can conclude that the subcritical case is not very sensitive to the prior parameters. We also made similar studies for the critical (again taking $\alpha_{1,(1,1)} = \alpha_{2,(1,1)} = \alpha$) and supercritical (now increasing the weight assigned to the state (0,0), i.e., $\alpha_{1,(0,0)} = \alpha_{2,(0,0)} = \alpha$) cases. Fig. 6 shows that, while neither is the supercritical case very sensitive, the critical case does seem to be sensitive.

The critical case thus leads us to conclude that care must be taken to use informative prior distributions.

Author's personal copy

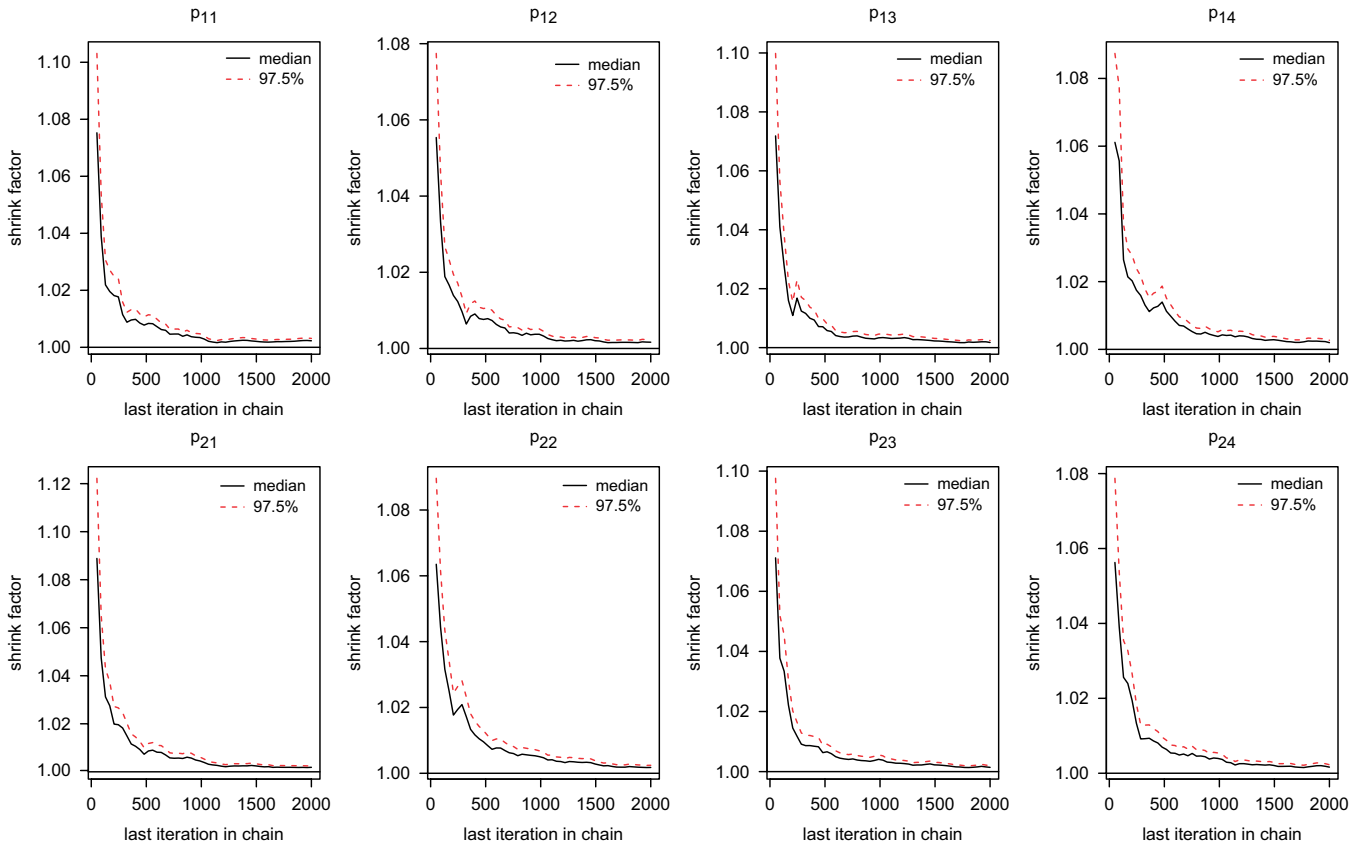

Fig. 3. Supercritical case: Gelman–Rubin–Brooks diagnostic plots.

Table 2
Potential scale reduction factor and autocorrelation for p

	Subcritical case				Critical case				Supercritical case			
	Potential scale reduction		Autocorrelation		Potential scale reduction		Autocorrelation		Potential scale reduction		Autocorrelation	
	Est.	97.5%	lag1	lag10	Est.	97.5%	lag1	lag10	Est.	97.5%	lag1	lag10
p_{11}	1.00	1.00	0.2705	0.0062	1.00	1.00	0.3963	0.0816	1.00	1.00	0.5271	0.0228
p_{12}	1.00	1.00	0.3430	−0.0038	1.00	1.01	0.6232	0.1538	1.00	1.00	0.4660	0.0146
p_{13}	1.00	1.00	0.4743	0.0110	1.00	1.00	0.3008	0.0127	1.00	1.00	0.4500	0.0217
p_{14}	1.00	1.00	0.2533	−0.0027	1.00	1.00	0.2046	0.0034	1.00	1.00	0.5368	0.0211
p_{21}	1.00	1.00	0.5632	0.0140	1.00	1.01	0.6711	0.1615	1.00	1.00	0.5364	0.0353
p_{22}	1.00	1.00	0.3122	0.0057	1.00	1.01	0.5692	0.1265	1.00	1.00	0.5474	0.0255
p_{23}	1.00	1.00	0.3094	0.0065	1.00	1.00	0.2156	0.0047	1.00	1.00	0.5221	0.0152
p_{24}	1.00	1.00	0.4830	−0.0038	1.00	1.00	0.3410	0.0078	1.00	1.00	0.4887	0.0165

Table 3
Summary statistics for the posterior distribution of ρ

	MEAN	SD	MCSE	TSSE
Subcritical case	0.97025	0.10681	0.00107	0.00105
Critical case	0.98708	0.11147	0.00111	0.00107
Supercritical case	1.04225	0.10045	0.00100	0.00099

Author's personal copy

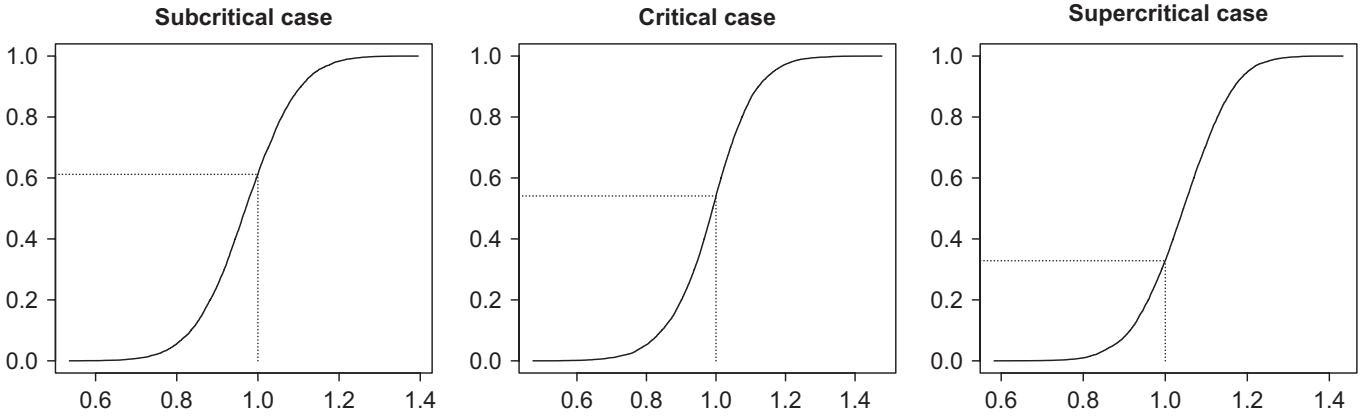

Fig. 4. Empirical posterior distribution function of ρ.

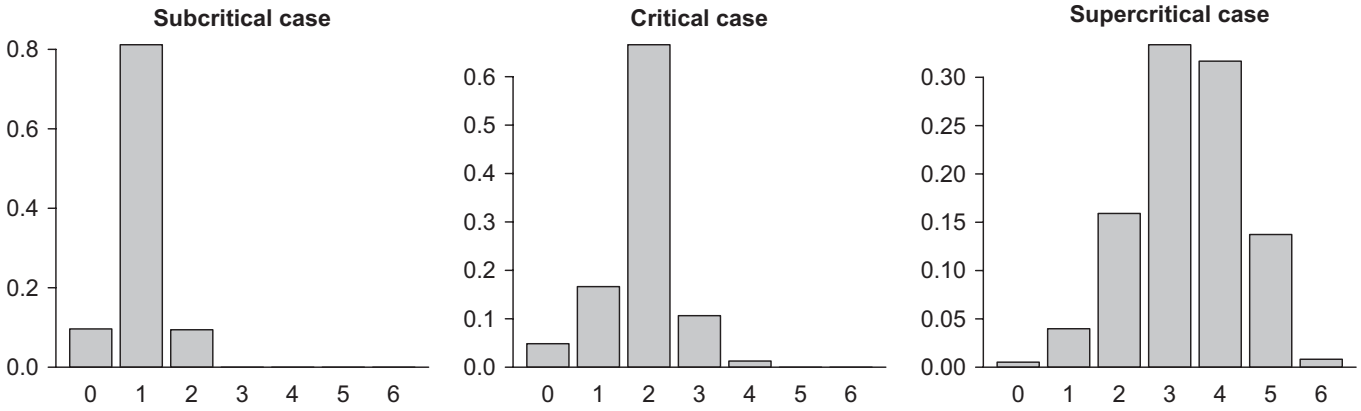

Fig. 5. Histogram of the estimated predictive distribution of $Z_1(11) + Z_2(11)|\mathscr{Z}(10)$.

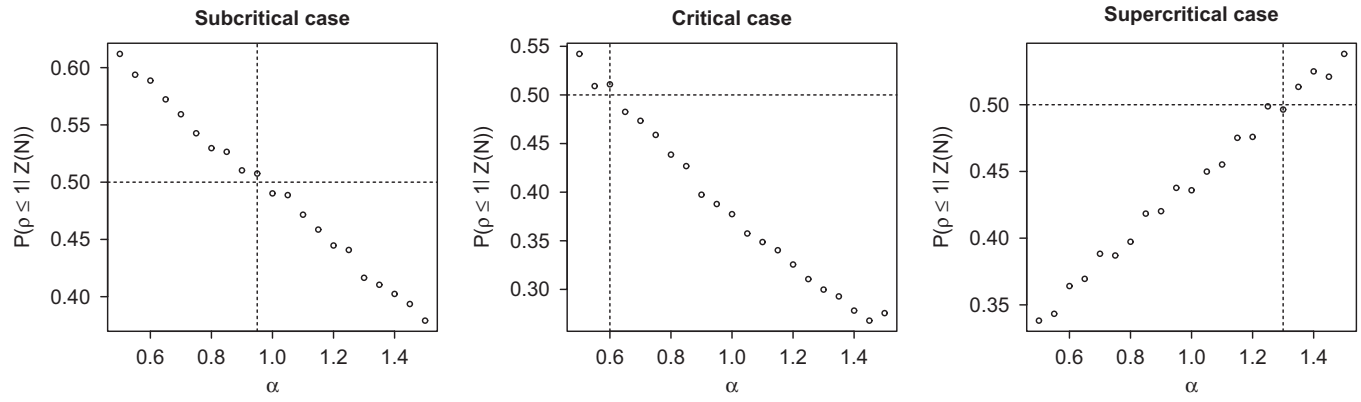

Fig. 6. Sensitivity analysis.

Remark 2. The computer simulations were programmed in the statistical computing and graphics language and environment **R** ("*GNU S*") (see R Development Core Team, 2007). Computation times for each chain were 11.33, 10.31, and 18.68 s for the subcritical, critical, and supercritical simulations, respectively, on an Intel Pentium IV CPU running at 3 GHz with 512 MB of RAM. The convergence diagnostic used the CODA package for **R** (see Plummer et al., 2006).

Author's personal copy

Acknowledgements

The authors are grateful to the anonymous referees for their comments and suggestions which have considerably improved this paper.

Research supported by the Ministerio de Educación y Ciencia and the FEDER through the Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica, grants MTM2006-08891 and TSI2004-06801-C04-03.

5. Appendix

We here prove the equality

$$P(\mathscr{Z}(N,\mathscr{S})|(\mathscr{Z}(N),p)) = \prod_{n=0}^{N-1} P(Z(n,\mathscr{S})|(Z(n),Z(n+1),p)),$$

recalling that $\mathscr{Z}(N,\mathscr{S}) = (Z_i(n,k); n=0,\ldots,N-1, k\in\mathscr{S}_i, i=1,\ldots,m)$, $\mathscr{Z}(N) = (Z(0),\ldots,Z(N))$, $Z(n,\mathscr{S}) = (Z_i(n,k); k\in\mathscr{S}_i, i=1,\ldots,m)$ and $p=(p_1,\ldots,p_m)$.

First, applying the multiplication rule, one has that the left-hand side of the above identity is equal to

$$\prod_{n=0}^{N-1} P(Z(n,\mathscr{S})|\mathscr{Z}(N), Z(s,\mathscr{S}), s=n+1,\ldots,N-1,p). \tag{3}$$

Given $n=0,\ldots,N-1$, and the vectors $(z(l), l=0,\ldots,N)$ and $(z_i(s,k), s=0,\ldots,N-1, k\in\mathscr{S}_i, i=1,\ldots,m)$, for each $s=n,\ldots,N-1$, write $z(s,\mathscr{S}) = (z_i(s,k), k\in\mathscr{S}_i, i=1,\ldots,m)$ and

$$A_{z(s)}^{z(s,\mathscr{S})} = \left\{ \sum_{j=1}^{z_i(s)} I_{\{Y_{ij}(s)=k\}} = z_i(s,k), k\in\mathscr{S}_i, i=1,\ldots,m \right\},$$

with $z_i(s)$ being the ith coordinate of vector $z(s)$, $i=1,\ldots,m$.

Therefore, from (3), it is enough to prove

$$\begin{aligned}&\Pr(A_{z(n)}^{z(n,\mathscr{S})}|Z(l)=z(l), l=0,\ldots,N, A_{z(s)}^{z(s,\mathscr{S})}, s=n+1,\ldots,N-1,p)\\ &\quad=\Pr(A_{z(n)}^{z(n,\mathscr{S})}|(Z(n)=z(n),Z(n+1)=z(n+1),p)).\end{aligned} \tag{4}$$

Since the Y vectors are independent,

$$\Pr(Z(l)=z(l), l=0,\ldots,N, A_{z(s)}^{z(s,\mathscr{S})}, s=n,\ldots,N-1|p)$$

is the product of the probabilities

$$\Pr(Z(l)=z(l), l=0,\ldots,n+1, A_{z(n)}^{z(n,\mathscr{S})}|p)$$

and

$$\Pr(Z(s)=z(s), A_{z(s)}^{z(s,\mathscr{S})}, s=n+1,\ldots,N-1, Z(N)=z(N)|B),$$

with $B=(Z(n+1)=z(n+1),p)$. Also, one has that

$$\begin{aligned}\Pr(Z(l)=z(l), l=0,\ldots,n+1, A_{z(n)}^{z(n,\mathscr{S})}|p) &= \Pr(Z(l)=z(l), l=0,\ldots,n|p)\\ &\quad\times\Pr(A_{z(n)}^{z(n,\mathscr{S})}, Z(n+1)=z(n+1)|(Z(n)=z(n),p)).\end{aligned}$$

From the Markov property, it follows that

$$\begin{aligned}\Pr(Z(l)=z(l), l=0,\ldots,N, A_{z(s)}^{z(s,\mathscr{S})}, s=n+1,\ldots,N-1|p) &= \Pr(Z(l)=z(l), l=0,\ldots,n|p)\\ &\quad\times\Pr(Z(n+1)=z(n+1)|(Z(n)=z(n),p))\\ &\quad\times\Pr(Z(s)=z(s), A_{z(s)}^{z(s,\mathscr{S})}, s=n+1,\ldots,N-1, Z(N)=z(N)|B).\end{aligned}$$

Author's personal copy

The left-hand side of (4) can therefore be written as

$$\frac{\Pr(A_{z(n)}^{z(n,\mathscr{S})},\, Z(n+1)=z(n+1)|(Z(n)=z(n),\, p))}{\Pr(Z(n+1)=z(n+1)|(Z(n)=z(n),\, p))}$$

and the proof is complete. □

References

Asmussen, S., Keiding, H., 1978. Martingale central limit theorems and asymptotic estimation theory for multitype branching processes. Adv. in Appl. Probab. 10, 109–129.

Ball, F., Britton, T., Lyne, O.D., 2004. Stochastic multitype epidemics in a community of households: estimation of threshold parameter R^* and secure vaccination coverage. Biometrika 91 (2), 345–362.

Berger, J., Bernardo, J.M., 1992. Ordered group reference prior with application to a multinomial problem. Biometrika 79, 25–37.

Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. J. Comput. Graph. Statist. 7, 434–455.

Devroye, L., 1998. Branching processes and their applications in the analysis of tree structures and tree algorithms. In: Habib, M., McDiarmid, C., Ramirez-Alfonsin, J., Reed, B., (Eds.), Probabilistic Methods for Algorithmic Discrete Mathematics, vol. 16. Springer, Berlin, pp. 249–314.

Dion, J.P., Epps, T.P., 1999. Stock prices as branching processes in random environments: estimation. Comm. Statist. Simulation Comput. 28 (4), 957–975.

El Adlouni, S., Favre, A.C., Bobe, B., 2006. Comparison of methodologies to assess the convergence of Markov Chain Monte Carlo methods. Comput. Statist. Data Anal. 50 (10), 2685–2701.

Gelman, A., Rubin, D., 1992. Inference from iterative simulation using multiple sequences. Statist. Sci. 7 (4), 457–511.

González, M., Hull, D.M., Martínez, R., Mota, M., 2006. Bisexual branching processes in a genetic context: the extinction problem for Y-linked genes. Math. Biosci. 202, 227–247.

Guttorp, P., 1991. Statistical Inference for Branching Processes. Wiley, New York.

Haccou, P., Jagers, P., Vatutin, V., 2005. Branching processes: variation, growth and extinction of populations. Cambridge University Press, Cambridge.

Kimmel, M., Axelrod, D.E., 2002. Branching Processes in Biology. Springer, Berlin.

Mendoza, M., Gutiérrez-Peña, E., 2000. Bayesian conjugate analysis of the Galton–Watson process. Test 9 (1), 149–171.

Mode, C., 1971. Multitype Branching Processes. Elsevier, Amsterdam.

Mode, C.J., Sleemam, C.K., 2000. Stochastic Processes in Epidemiology. World Scientific, Singapore.

Montpetit, M., Haccoun, D., Deslauriers, G., 1992. A branching process analysis of the stack algorithm for variable channel conditions. IEEE Trans. Inform. Theory 38, 1845–1848.

Nanthi, K., 1982. Estimation of the variance for the multitype Galton–Watson process. J. Appl. Probab. 19, 421–426.

Pakes, A., 2003. Biological applications of branching processes. In: Shanbhag, D.N., Rao, C.R., (Eds.), Handbook of Statistic, vol. 21 Stochastic Processes: Modelling and Simulation, Elsevier Science, Amsterdam, pp. 693–773.

Plummer, M., Best, N., Cowles, K., Vines, K., 2006. coda: output analysis and diagnostics for MCMC. R package version 0.10-7.

R Development Core Team, 2007. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. URL: http://www.R-project.org.

Tierney, L., 1994. Markov chain for exploring posterior distribution (with discussion). Ann. Statist. 22, 1701–1762.

Yakovlev, A., Yanev, N., 2006. Branching stochastic processes with immigration in analysis of renewing cell populations. Math. Biosci. 203, 37–63.