

FIXATION IN CONDITIONAL BRANCHING PROCESS MODELS IN POPULATION GENETICS

THOMAS PRINCE* AND

NEVILLE WEBER,** *University of Sydney*

Abstract

An alternative version of the necessary and sufficient condition for almost sure fixation in the conditional branching process model is derived. This formulation provides an insight into why the examples considered in Buckley and Seneta (1983) all have the same condition for fixation.

Keywords: Fixation; branching process; Wright–Fisher model; bisexual model

2000 Mathematics Subject Classification: Primary 92D25
Secondary 60J80

1. Introduction

Donnelly (1986) discussed a general exchangeable model used to study gene survival in populations whose size changes without density dependence. He provided necessary and sufficient conditions on the rate of growth of the population size for the maintenance of genetic diversity.

In this paper we focus on conditional branching process models. These models were first considered by Moran and Watterson (1959), and fixation was investigated for generations of a fixed size by Karlin and McGregor (1964). Buckley and Seneta (1983) studied the variable population size situation and provided a sufficient condition for almost sure fixation. This condition was noted in Donnelly (1986) as equivalent to the necessary and sufficient conditions given in Donnelly (1986).

By applying the techniques developed in Karlin and McGregor (1964), we obtain an alternative formulation of the necessary and sufficient condition for fixation. One consequence of this formulation is a unified way of dealing with the three special conditional branching process models considered by Buckley and Seneta (1983). The formulation also has a natural extension to conditional bisexual branching process models. These results complement the fixation results for the class of exchangeable bisexual models considered by Möhle (1997).

2. The general exchangeable model

Assume that the population evolves in discrete generations. Let M_n denote the size of generation n . Assume that $1 < M_n < \infty$, $n = 0, 1, 2, \dots$. Consider a haploid asexual population of genes. At a single locus of interest there exist two neutral alleles, a and A . Let X_n denote the number of type a alleles in generation n , $0 < X_0 < M_0$.

Received 2 February 2007; revision received 2 November 2007.

* Postal address: School of Mathematics and Statistics F07, University of Sydney, NSW 2006, Australia.

** Email address: neville@maths.usyd.edu.au

Conditional on M_n and M_{n+1} , let $v_i^{(n)}$ denote the number of offspring of individual i in generation n . Assume that the offspring are of the same allelic type as the parent and that $\{v_i^{(n)}\}$ are exchangeable with $v_1^{(n)} + v_2^{(n)} + \dots + v_{M_n}^{(n)} = M_{n+1}$. Furthermore, assume that the distribution of $\{v_i^{(n)}\}$, given M_0, M_1, \dots , depends only on M_n and M_{n+1} .

Let $Y_n := X_n/M_n$ and $\mathcal{F}_n := \sigma\{X_0, M_0, \dots, X_n, M_n\}$. Then, as noted by Donnelly (1986), (Y_n, \mathcal{F}_n) is a martingale and since $|Y_n| \leq 1$, we have $Y_n \rightarrow Y$ almost surely (a.s.). The probability of fixation is $P(Y = 0) + P(Y = 1)$.

Let

$$V_k := E(v_1^{(k)}(v_1^{(k)} - 1) \mid M_k, M_{k+1}) \quad \text{and} \quad W_k := \frac{M_k V_k}{M_{k+1}(M_{k+1} - 1)}.$$

Let ‘ $\stackrel{\text{A.S.}}{=}$ ’ denote almost sure equality.

Theorem 2.1. (Donnelly (1986).) *A necessary and sufficient condition for certain fixation is*

$$\prod_{k=0}^{\infty} (1 - W_k) \stackrel{\text{A.S.}}{=} 0.$$

If $W_k < 1$ a.s. for all k then this condition is equivalent to $\sum_{k=0}^{\infty} W_k \stackrel{\text{A.S.}}{=} \infty$.

Note that the condition $W_k < 1$ a.s. for all k is needed. If $M_n = 2^{n+1}$ and each allele produces zero or four offspring, each with probability $\frac{1}{2}$, then the population is fixed at generation 1, but $W_k = 3/(2^{k+2} - 1)$, so $\sum_{k=0}^{\infty} W_k < \infty$ a.s.

3. Conditional branching process models

Let $R_i^{(n)}$ denote the number of offspring produced by the i th individual in generation n , where $R_i^{(n)}$, $n = 0, 1, \dots$, $i = 1, 2, \dots$, are independent and identically distributed random variables. Then, given M_n and M_{n+1} ,

$$v_i^{(n)} \stackrel{\text{D}}{=} \left(R_i^{(n)} \mid \sum_{i=1}^{M_n} R_i^{(n)} = M_{n+1} \right), \tag{3.1}$$

where ‘ $\stackrel{\text{D}}{=}$ ’ denotes equality in distribution. Each branching process model $(R_i^{(n)}, M_n)_{i,n}$ induces an exchangeable model via (3.1). Thus, the class of conditional branching process models is a subclass of the class of exchangeable models. In particular, if the offspring distribution is Poisson then the conditional branching process model reduces to the classical Wright–Fisher model with

$$P(X_{n+1} = j \mid X_n, M_n, M_{n+1}) = \binom{M_{n+1}}{j} \left(\frac{X_n}{M_n}\right)^j \left(1 - \frac{X_n}{M_n}\right)^{M_{n+1}-j}.$$

Therefore, in order to analyze a conditional branching process model, we can use the theory of exchangeable models or the special structure of the branching process model.

Theorem 3.1. (Buckley and Seneta (1983).) *Let*

$$V'_n := \text{var} \left(R_1^{(n)} \mid \sum_{i=1}^{M_n} R_i^{(n)} = M_{n+1}, M_n, M_{n+1} \right) \quad \text{and} \quad W'_{n+1} := \frac{V'_n}{M_n - 1} \left(\frac{M_n}{M_{n+1}} \right)^2.$$

Then fixation is certain if $\sum_{k=0}^{\infty} W'_{k+1} \stackrel{\text{A.S.}}{=} \infty$.

Donnelly (1986) commented that the Buckley and Seneta (1983) condition is in fact necessary and sufficient for certain fixation. To see this note that

$$\begin{aligned} V_n &= E(v_1^{(n)}(v_1^{(n)} - 1) \mid M_n, M_{n+1}) \\ &= V'_n - E(v_1^{(n)} \mid M_n, M_{n+1}) + (E(v_1^{(n)} \mid M_n, M_{n+1}))^2 \\ &= V'_n + \frac{M_{n+1}}{M_n^2}(M_{n+1} - M_n). \end{aligned}$$

Thus,

$$1 - W_n = \frac{1 - M_n^{-1}}{1 - M_{n+1}^{-1}}(1 - W'_{n+1})$$

and, hence,

$$\prod_{n=0}^k (1 - W_n) = \frac{(1 - 1/M_0)}{(1 - 1/M_{k+1})} \prod_{n=0}^k (1 - W'_{n+1}).$$

Adapting the technique of Karlin and McGregor (1964), we develop an alternative form for the above condition. We will exclude trivial cases where individuals can have only zero or one offspring. In these cases $V_n = W_n = 0$ and fixation is not certain. Let $f(w) := \sum_{j=0}^\infty P(R_1^{(1)} = j)w^j$ denote the common offspring probability generating function (PGF).

Theorem 3.2. Fixation is certain for the conditional branching process model if and only if

$$\prod_{n=0}^\infty \frac{(M_n - 1)K_n}{1 + (M_n - 1)K_n} \stackrel{\text{A.S.}}{=} 0,$$

where

$$K_n := \frac{\text{Coefficient of } w^{M_{n+1}-2} \text{ in } f(w)^{M_n-2}(f'(w))^2}{\text{Coefficient of } w^{M_{n+1}-2} \text{ in } f(w)^{M_n-1}f''(w)}.$$

If $K_n > 0$ a.s. for all n then this condition is equivalent to $\sum_{n=0}^\infty (1 + (M_n - 1)K_n)^{-1} \stackrel{\text{A.S.}}{=} \infty$.

Proof. From

$$\begin{aligned} P(v_1^{(n)} = j \mid M_n, M_{n+1}) &= P\left(R_1^{(n)} = j \mid \sum_{i=1}^{M_n} R_i^{(n)} = M_{n+1}, M_n, M_{n+1}\right) \\ &= \frac{P(R_1^{(n)} = j, \sum_{i=2}^{M_n} R_i^{(n)} = M_{n+1} - j \mid M_n, M_{n+1})}{P(\sum_{i=1}^{M_n} R_i^{(n)} = M_{n+1} \mid M_n, M_{n+1})} \\ &= \frac{(\text{Coefficient of } s^j \text{ in } f(s))(\text{Coefficient of } t^{M_{n+1}-j} \text{ in } (f(t))^{M_n-1})}{\text{Coefficient of } w^{M_{n+1}} \text{ in } (f(w))^{M_n}} \\ &= \frac{\text{Coefficient of } s^j t^{M_{n+1}-j} \text{ in } f(s)(f(t))^{M_n-1}}{\text{Coefficient of } w^{M_{n+1}} \text{ in } (f(w))^{M_n}} \\ &= \frac{\text{Coefficient of } r^j w^{M_{n+1}} \text{ in } f(rw)(f(w))^{M_n-1}}{\text{Coefficient of } w^{M_{n+1}} \text{ in } (f(w))^{M_n}}, \end{aligned}$$

it follows that

$$G(r) := \sum_j P(v_1^{(n)} = j \mid M_n, M_{n+1}) r^j = \frac{\text{Coefficient of } w^{M_{n+1}} \text{ in } f(rw)(f(w))^{M_n-1}}{\text{Coefficient of } w^{M_{n+1}} \text{ in } (f(w))^{M_n}}.$$

Differentiating $G(r)$ twice with respect to r gives

$$\begin{aligned} G''(r) &= \frac{M_{n+1}(M_{n+1} - 1)}{M_n} \\ &\times \frac{\text{Coefficient of } w^{M_{n+1}-2} \text{ in } f''(rw)(f(w))^{M_n-1}}{\text{Coefficient of } w^{M_{n+1}-2} \text{ in } (f''(w)(f(w))^{M_n-1} + (M_n - 1)f(w)^{M_n-2}(f'(w))^2)}. \end{aligned}$$

Thus,

$$\begin{aligned} V_n &= E(v_1^{(n)}(v_1^{(n)} - 1) \mid M_n, M_{n+1}) \\ &= G''(1) \\ &= \frac{M_{n+1}(M_{n+1} - 1)}{M_n} \frac{1}{1 + (M_n - 1)K_n}, \end{aligned}$$

and the result follows by applying Theorem 2.1.

If $K_n = 0$ for some n then $W_n = 1$ and fixation is certain. If $K_n = K > 0$ then $K M_n/2 \leq (1 + K(M_n - 1)) \leq (K + 1)M_n$, as $M_n \geq 2$, so in this case $\sum_{n=0}^\infty M_n^{-1} \stackrel{\text{A.S.}}{=} \infty$ is necessary and sufficient for certain fixation.

Corollary 3.1. *If the offspring PGF satisfies*

$$(f'(w))^2 = Kf(w)f''(w)$$

for some constant $K > 0$ then $\sum_{n=0}^\infty M_n^{-1} \stackrel{\text{A.S.}}{=} \infty$ is necessary and sufficient for certain fixation in the conditional branching process model.

Solving the differential equation gives

$$f(w) = \begin{cases} e^{\mu(w-1)} & \text{for } K = 1, \\ \left(1 + \frac{\mu(K-1)}{K}(w-1)\right)^{K/(K-1)} & \text{for } K \neq 1. \end{cases}$$

For $K = 1$, we obtain the Poisson offspring distribution and the classical Wright–Fisher model. If $0 < K < 1$, we have the negative binomial offspring distribution, and if $K = 1 + (d - 1)^{-1}$ for some integer d then the offspring distribution is binomial with parameters $(d, \mu/d)$.

Corollary 3.2. *If the offspring PGF is a polynomial, $f(w) = p_0 + p_1w + \dots + p_gw^g$ of degree g , and if there exists a constant $K_1 < \infty$ such that $p_1^2 \leq K_1p_0p_2$ then $\sum_{n=0}^\infty M_n^{-1} \stackrel{\text{A.S.}}{=} \infty$ is sufficient for certain fixation in the conditional branching process model.*

Proof. First note that $(f'(w))^2$ and $f(w)f''(w)$ are both polynomials of the same degree. Compare the coefficients term by term for w^j . For $j = 0$, we need to compare p_1^2 with $2p_0p_2$,

and by assumption there exists a finite constant such that $K_1 p_0 p_2 \geq p_1^2$. By considering each term in turn for $j \geq 1$, it is easy to see that

$$(f'(w))^2 \leq K_2 f(w) f''(w),$$

where $K_2 \geq \max(2g, K_1/2)$. Thus,

$$\sum_{n=0}^{\infty} (1 + (M_n - 1)K_n)^{-1} \geq (K_2 + 1)^{-1} \sum_{n=0}^{\infty} M_n^{-1},$$

and the result follows.

Donnelly (1986) commented that the ‘offspring distribution plays far too prominent a role’ in determining necessary and sufficient conditions for certain fixation. ‘Conditions (. . .) which depend only on the population size process would be of more interest.’ Corollaries 3.1 and 3.2 go some way towards addressing this comment.

4. Extension to bisexual models

Daley (1968) introduced the bisexual Galton–Watson branching process whereby at each generation, F_n females and M_n males form $Z_n = \mathcal{L}(F_n, M_n)$ mating pairs, each of which independently produce a random number of offspring. The function $\mathcal{L}(x, y) = \min(x, y)$ denotes the perfect fidelity mating function. One variation examined by Daley has mating pairs producing offspring according to the PGF $f(s)$ and then each offspring is independently classified as male, with probability α , or female, with probability $1 - \alpha$. This structure naturally extends to produce a bisexual conditional branching process model.

Möhle (1997) extended the ideas of Donnelly (1986) to a bisexual model based on the structure of Moran (1958). Möhle obtained necessary and sufficient conditions for certain fixation and applied these to the bisexual analogues of the Wright–Fisher model and Moran models. He did not consider the conditional branching process model, but it can be considered as a special case of the general bisexual exchangeable model.

Using Möhle’s notation, we will briefly describe his general bisexual model. Consider a diploid, dioecious population of genes which, at a single locus, has two neutral alleles, a and A . Males and females in generation n are randomly paired to form N_n pairs. As time goes from n to $(n + 1)$, V_n pairs are removed from the population, and U_n newborn pairs are added:

$$N_{n+1} = N_n + U_n - V_n, \quad n \geq 0, N_n \geq 2, U_n \geq 1.$$

Each child receives one gene from its mother and one gene from its father randomly, independently of the process. A pair alive at time n is called an n -pair. An individual alive at time n is called an n -individual. The sons and daughters produced by the n -individuals are randomly paired to form U_n new pairs. Let $v_j^{(n)}$ and $\mu_j^{(n)}$ respectively denote the number of daughters and sons of the j th n -pair, and suppose that $\sum_{j=1}^{N_n} v_j^{(n)} = \sum_{j=1}^{N_n} \mu_j^{(n)} = U_n$. Define the offspring vectors $\mathbf{K}^{(n)} := (v_1^{(n)}, \mu_1^{(n)}, v_2^{(n)}, \mu_2^{(n)}, \dots)$, with the convention $v_j^{(n)} := \mu_j^{(n)} := 0$ for $j > N_n$. We assume that, conditional on $U := (N_0, U_0, U_1, \dots)$ and $V := (V_0, V_1, \dots)$, the vectors $\mathbf{K}^{(n)}$ are independent. Furthermore, we assume exchangeability. Conditioned on U and V and for fixed n , let π be a permutation on \mathbb{N} with $\pi(j) = j$ for all $j > N_n$. Then $\pi \mathbf{K}^{(n)} := (v_{\pi(1)}^{(n)}, \mu_{\pi(1)}^{(n)}, v_{\pi(2)}^{(n)}, \mu_{\pi(2)}^{(n)}, \dots)$ has the same distribution as $\mathbf{K}^{(n)}$. It is initially assumed that both alleles are present in the population.

We are concerned with certain fixation on a genetic level. Define H_n to be the probability that two genes chosen from generation n at random with replacement are of different allelic types.

Lemma 4.1. (Möhle (1997, Lemma 4.2).) *For the model described above, certain fixation occurs if and only if $\lim_{n \rightarrow \infty} H_n = 0$.*

Next apply the arguments in Lemma 2.3 and Theorem 2.1 of Donnelly (1986).

Theorem 4.1. *Fixation is certain if and only if $\sum_{n=0}^{\infty} W_n^* \stackrel{\text{A.S.}}{=} \infty$, where*

$$W_n^* := \text{P}(\text{two distinct genes in generation } n + 1 \text{ have the same ancestor gene in generation } n \mid U, V).$$

Note that since each individual receives a gene from its mother and father, the genes in generation n cannot all be descended from a single $(n + 1)$ -ancestor gene, so $W_n^* < 1$ a.s. Our aim is to evaluate W_n^* . Fix $n \geq 1$. Choose two n -genes (with replacement) and consider their ancestral genes in generation $k \in \{0, \dots, n - 1\}$.

Define

$$\zeta_k^{(n)} := \zeta_k := \begin{cases} 1 & \text{if the two ancestor genes are the same,} \\ 2 & \text{if the ancestor genes are distinct, but belong to the same individual,} \\ 3 & \text{if the ancestor genes belong to different individuals of the same } k\text{-pair,} \\ 4 & \text{if the ancestor genes belong to different } k\text{-pairs.} \end{cases}$$

Let $G_n := \text{P}(E_n \mid U, V)$, where E_n denotes the event that the two individuals of a randomly chosen newborn $(n + 1)$ -pair are brother and sister. Furthermore, let $W_n := \text{P}(F_n \mid U, V)$, where F_n denotes the event that two randomly chosen individuals belonging to different $(n + 1)$ -pairs are either both newborn children of the same n -pair, or one individual is newborn and a child of the other individual.

Lemma 4.2. (Möhle (1997, Lemma 4.8).) *Fix n . Conditional on U and V , the backward Markov chain $\{\zeta_k\}$, $k \in \{n, n - 1, \dots, 0\}$ has initial distribution*

$$\text{P}(\zeta_n = i \mid U, V) \stackrel{\text{A.S.}}{=} \begin{cases} \frac{1}{4N_n} & \text{for } i \in \{1, 2\}, \\ \frac{1}{2N_n} & \text{for } i = 3, \\ 1 - \frac{1}{N_n} & \text{for } i = 4, \end{cases}$$

and transition matrix $\mathbf{P} = (P_{ij})_{1 \leq i, j \leq 4}$ with $P_{ij} := \text{P}(\zeta_{k-1} = j \mid \zeta_k = i, U, V)$, given by

$$\mathbf{P} \stackrel{\text{A.S.}}{=} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 - \frac{U_{k-1}}{N_k} & \frac{U_{k-1}}{N_k} & 0 \\ \frac{1}{4} \frac{G_{k-1}U_{k-1}}{N_k} & \frac{1}{4} \frac{G_{k-1}U_{k-1}}{N_k} & 1 - \frac{U_{k-1}}{N_k} + \frac{G_{k-1}U_{k-1}}{2N_k} & \frac{(1 - G_{k-1})U_{k-1}}{N_k} \\ \frac{W_{k-1}}{4} & \frac{W_{k-1}}{4} & \frac{W_{k-1}}{2} & 1 - W_{k-1} \end{pmatrix}.$$

Using the transition matrix in Lemma 4.2, we are able to evaluate W_n^* .

Lemma 4.3. *For the above bisexual model,*

$$W_n^* = \frac{N_n E[(v_1^{(n)} + \mu_1^{(n)})(v_1^{(n)} + \mu_1^{(n)} - 1) \mid U, V]}{4N_{n+1}(4N_{n+1} - 1)} + \frac{2U_n(N_{n+1} - U_n)}{(4N_{n+1} - 1)N_n N_{n+1}}.$$

Proof. Two distinct genes in generation $n + 1$ can either:

- (i) belong to the same individual with probability $1/(4N_{n+1} - 1)$,
- (ii) belong to different individuals of the same pair with probability $2/(4N_{n+1} - 1)$, or
- (iii) belong to different $(n + 1)$ -pairs with probability $4(N_{n+1} - 1)/(4N_{n+1} - 1)$.

Then,

$$\begin{aligned} W_n^* &= P(\text{two distinct genes in generation } n + 1 \text{ have the same ancestor} \\ &\quad \text{gene in generation } n \mid U, V) \\ &= \sum_{i=1}^3 P(\text{two distinct genes in generation } n + 1 \text{ have the same ancestor} \\ &\quad \text{gene in generation } n \mid U, V, A_i) P(A_i \mid U, V), \end{aligned}$$

where A_1 is the event that the $(n + 1)$ -genes belong to the same individual, A_2 is the event that the $(n + 1)$ -genes belong to different individuals of the same pair, and A_3 is the event that the genes come from different $(n + 1)$ -pairs. Using Lemma 4.2, we obtain

$$\begin{aligned} W_n^* &= 0 \frac{1}{4N_{n+1} - 1} + \frac{G_n U_n}{4N_{n+1}} \frac{2}{4N_{n+1} - 1} + \frac{W_n}{4} \frac{4(N_{n+1} - 1)}{4N_{n+1} - 1} \\ &= \frac{1}{4N_{n+1} - 1} \left(\frac{G_n U_n}{2N_{n+1}} + (N_{n+1} - 1)W_n \right) \\ &= \frac{N_n E[(v_1^{(n)} + \mu_1^{(n)})(v_1^{(n)} + \mu_1^{(n)} - 1) \mid U, V]}{4N_{n+1}(4N_{n+1} - 1)} + \frac{2U_n(N_{n+1} - U_n)}{(4N_{n+1} - 1)N_n N_{n+1}}, \end{aligned}$$

using Lemmas 4.3 and 4.5 of Möhle (1997) to evaluate G_n and W_n .

Consider a bisexual Galton–Watson process, where each of the N_n pairs at generation n produce a random number of offspring $R_i^{(n)}$, $i = 1, \dots, N_n$. We condition on the requirement that $\sum_{i=1}^{N_n} R_i^{(n)} = 2N_{n+1}$. Theorem 4.1 can now be applied. Note that we are dealing with nonoverlapping generations, so $U_n = N_{n+1}$. The techniques in the proof of Theorem 3.2 are used to evaluate the relevant conditional expectation.

Theorem 4.2. *Provided that there exists a nonzero probability of an individual pair having two or more offspring, certain fixation for the bisexual conditional branching process model occurs if and only if*

$$\sum_{n=0}^{\infty} (1 + (N_n - 1)K_n')^{-1} \stackrel{\text{a.s.}}{=} \infty,$$

where

$$K_n' = \frac{\text{Coefficient of } w^{2N_{n+1}-2} \text{ in } f(w)^{N_n-2} (f'(w))^2}{\text{Coefficient of } w^{2N_{n+1}-2} \text{ in } f(w)^{N_n-1} f''(w)},$$

provided that $K_n' > 0$ a.s. for all n .

Proof. Since $U_n = N_{n+1}$, the result follows by evaluating the conditional expectation in Lemma 4.3. To do this, replace M_{n+1} by $2N_{n+1}$ and M_n by N_n in the proof of Theorem 3.2 and apply Theorem 4.1.

Thus, for the Poisson, binomial, and negative binomial offspring distributions, we have certain fixation if and only if $\sum_{n=0}^{\infty} N_n^{-1} \stackrel{\text{A.S.}}{=} \infty$.

Acknowledgement

The authors would like to thank an anonymous referee for constructive comments that have improved the presentation of this paper.

References

- BUCKLEY, M. J. AND SENETA, E. (1983). The genetic balance between varying population size and selective neutrality. *J. Math. Biol.* **17**, 217–222.
- DALEY, D. J. (1968). Extinction conditions for certain Galton–Watson branching processes. *Z. Wahrscheinlichkeitsthe.* **9**, 315–322.
- DONNELLY, P. (1986). A genealogical approach to variable-population-size models in population genetics. *J. Appl. Prob.* **23**, 283–296.
- KARLIN, S. AND MCGREGOR, J. (1964). Direct product branching processes and related Markov chains. *Proc. Nat. Acad. Sci. USA* **51**, 598–602.
- MÖHLE, M. (1997). Fixation in bisexual models with variable population sizes. *J. Appl. Prob.* **34**, 436–448.
- MORAN, P. A. P. (1958). Random processes in genetics. *Proc. Camb. Philos. Soc.* **54**, 60–71.
- MORAN, P. A. P. AND WATTERSON, G. A. (1959). The genetic effects of family structure in natural populations. *Austral. J. Biol. Sci.* **12**, 1–15.