Quasi-Stationary Distributions and Time-Reversion in Genetics
Author(s): E. Seneta
Published by: Blackwell Publishing for the Royal Statistical Society
Accessed: 03/06/2010 21:44

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at [http://www.jstor.org/page/info/about/policies/terms.jsp](http://www.jstor.org/page/info/about/policies/terms.jsp). JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at [http://www.jstor.org/action/showPublisher?publisherCode=black](http://www.jstor.org/action/showPublisher?publisherCode=black).

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.
Quasi-stationary Distributions and Time-reversion in Genetics

By E. Seneta

Australian National University, Canberra

[Read at a Research Methods Meeting of the Society, December 1st, 1965, Professor P. Armitage in the Chair]

Summary

Finite absorbing Markov chains appear frequently as models for genetic processes. Although in such chains absorption in a finite time is certain, this time may be long, and it is then of interest to consider quasi-stationary distributions as describing the process in such situations; this is also true of certain infinite chains. Two such distributions are discussed, and derived for three particular genetic models as diffusion approximations. The distribution regarded as more satisfactory has interesting properties as regards "time-reversion" of the genetic process. A general discussion of this concept leads to the quasi-stationary generalization of Wright's formula for stationary distributions.

1. Introduction

If we are interested only in the behaviour within the set of transient states \( T \) of a finite absorbing Markov chain, we may without loss of generality write its transition matrix as one describing a process with only one absorbing state, viz.

\[
P = \{p_{ij}\} = \begin{bmatrix} 1 & 0^T \\ p & Q \end{bmatrix}, \quad p \neq 0,
\]

where the state 0 is absorbing, and the transient set \( T \) is \( \{1, 2, \ldots, n\} \). We shall consider only chains where \( Q \) is indecomposable and non-cyclic (as is generally the case in genetic models).

From the theory of non-negative matrices, \( Q \) has a largest (in modulus) unique positive eigenvalue \( p_1 \), with corresponding positive left eigenvector \( v \) and positive right eigenvector \( w \). We may consider both of these vectors to be normalized in such a way that

\[
\sum_{i=1}^{n} v_i = 1 \quad \text{and} \quad \sum_{i=1}^{n} w_i v_i = 1,
\]

where \( w_i, v_i \) denote the \( i \)th components of \( w \) and \( v \) respectively.

Darroch and Seneta (1965) have exploited these properties in order to investigate "quasi-stationary" distributions over \( T \), which the Markov chain may approach when the time to absorption is long. The most fruitful approach was to consider only those realizations of the chain for which the system stayed within the transient states for at least a specified time \( n \), and then let \( n \to \infty \). This led in particular to two distributions
over $T$, viz. $\{v_i\}$ and $\{v_iw_i\}$, the first of which was obtained as

$$\lim_{n \to \infty} \Pr(\text{process in state } j \text{ at time } n | \text{absorption has not occurred by time } n) = \lim_{n \to \infty} \frac{\sum_{i \in T} \pi_i P_{ij}^{(n)}}{\sum_{i \in T} \pi_i (1 - p_{i0}^{(n)})} = v_j,$$  

where $\pi'$ is the initial distribution over $T$. This has the property that it is independent of the initial distribution and, moreover, has a stationary interpretation, since if the conditional absolute probability distribution is considered, then this may be stationary if and only if it is $\nu'$; this is a direct consequence of the invariant equation

$$\nu'Q = \rho_1 \nu'.$$  

Thus $\nu'$ is a direct analogue, at least apparently, of the limiting-stationary distribution of a regular chain.

The distribution $\{v_i\}$ has implicitly received some mention in the literature of mathematical genetics (Wright, 1931; Fisher, 1958; Moran, 1962, Chapter IV); in particular, diffusion approximations to it for particular genetic models have been found. Ewens (1965) has discussed these cases, and has given this distribution the name “asymptotic conditional”. For simplicity, we shall henceforth refer to the distribution $\{v_i\}$ as the limiting conditional distribution.

On the other hand, Darroch and Seneta found that the distribution $\{w_i v_i\}$ could be obtained as a limit of several quite different expressions. The one most suited for the present study, and similar to (1.1) in its geometric rate of convergence, was

$$\lim_{m \to \infty} \lim_{n \to \infty} \Pr(\text{process in state } j \text{ at time } m < n | \text{absorption has not occurred at time } n) = \lim_{m \to \infty} \lim_{n \to \infty} \frac{\sum_{i \in T} \pi_i P_{ij}^{(m)}(1 - p_{j0}^{(n-m)})}{\sum_{i \in T} \pi_i (1 - p_{i0}^{(n)})} = w_j v_j.$$  

Thus $\{w_i v_i\}$ may be crudely described as the distribution of the random variable at time $m$ ($m$ large), given that absorption has not yet taken place and will not take place for a long time. Apart from the independence of $\pi'$, in fact $\{w_i v_i\}$ is rather more relevant than $\{v_i\}$, in the sense that (1.1) is a “degenerate” case of (1.3). Moreover, the distribution $\{w_i v_i\}$, like $\{v_i\}$, has a stationarity property: it is the unique stationary vector of the regular “reverse” Markov chain defined on the transient states $T$ only, by the “reverse” transition matrix

$$P^* = \{p_{ij}^*\}, \text{ where } p_{ij}^* = v_j p_{ji} / \rho_1 v_i.$$  

This property is the crux of much of the present paper, and is closely related to the generalization of the theory of Markov chains to arbitrary non-negative matrices. A physical interpretation of the stationarity property is given in Section 6. For convenience, we shall subsequently refer to the distribution $\{w_i v_i\}$ as the product distribution.

The product distribution possesses a further property not mentioned by Darroch and Seneta, which is of some genetic importance, and which we now briefly discuss.
Consider an absorbing Markov chain with several absorbing states, whose transition matrix may be written

\[ P = \begin{bmatrix} I & 0 \\ R & Q \end{bmatrix}. \]

If there are \( p \) absorbing states, \( I \) is the \( p \times p \) identity matrix, \( Q \) corresponds to the set \( T \) as before, and \( R \) is the matrix of absorption transitions, the generalization of the \( p \) given previously. We shall once more take \( Q \) indecomposable and non-cyclic. Breny (1962) has shown that if we consider the process conditional on the absorbing state \( k \) (1 ≤ \( k \) ≤ \( p \)) being given, then the conditional process is once more a Markov chain, with, of course, only one absorbing state, and transition matrix

\[ P(k) = \begin{bmatrix} 1 & 0' \\ R(k) & Q(k) \end{bmatrix}. \]

Here

\[ Q(k) = D_k^{-1} QD_k \quad \text{and} \quad R(k) = D_k^{-1} R e \]

\[ D_k = \text{diag}[f'_1[I - Q]^{-1} R f_k, ..., f'_n[I - Q]^{-1} R f_k], \]

where \( e \) is the column vector containing units in every position, and \( f'_i \) is the row vector with a unit in the \( i \)th position and zeros elsewhere.

We are concerned only with the transformation of \( Q \); in particular, we notice that under the similarity transformation by \( D_k \) the eigenvectors of \( Q \) are transformed as follows:

\[ v' \rightarrow v'D_k, \quad w \rightarrow D_k^{-1} w. \]

Thus, although the quasi-stationary distribution \( \{v_i\} \) is changed, the product distribution is clearly invariant. The reason for this may be seen intuitively from (1.3): since we expect absorption a long time after \( \{w_i v_i\} \) has become established, the eventual absorbing state does not affect the distribution. Also, the eigenvalues are invariant under the similarity transformation, so that the rates of approach to the two quasi-stationary distributions remain unchanged for the conditional process. It is clear that the product distribution possesses a quite desirable invariance property which is not shared by the limiting conditional distribution.

2. THE METHOD OF DIFFUSION APPROXIMATIONS

In general, the eigenvectors, and so the stationary distributions, are difficult to find for genetic Markov chain models, although it is sometimes possible to find the corresponding eigenvalues. For a fuller discussion of this topic, the reader is referred to Moran (1962). In order to obtain some information about the model, it is usual to approximate it by a diffusion process in \((0, 1)\), described by the Fokker–Planck equation

\[ \frac{\partial}{\partial t} \phi(x, t) = -\frac{\partial}{\partial x} \{\alpha(x) \phi(x, t)\} + \frac{1}{2} \frac{\partial^2}{\partial x^2} \{\beta(x) \phi(x, t)\}, \quad (2.1) \]

where the density function \( \phi(x, t) \) describes at time \( t \) the distribution of a variate \( x \), related in a certain way to the original Markov variate. The coefficients \( \alpha(x) \) and \( \beta(x) \) are measures of rate of drift and spread by diffusion respectively. (A discussion of the applicability and adequacy of such an approximation may be found in Moran (1962), and the recent papers of Ewens.)
Of interest to us, although rarely appearing in the genetical literature, is the adjoint of (2.1), viz. the backward Kolmogorov equation

$$-\frac{\partial}{\partial \tau} \psi(y, \tau) = \alpha(y) \frac{\partial}{\partial y} \{\psi(y, \tau)\} + \frac{1}{2} \beta(y) \frac{\partial^2}{\partial y^2} \{\psi(y, \tau)\}, \tag{2.2}$$

where $\alpha(y)$ and $\beta(y)$ are as before. The function $\psi(y, \tau)$ should be written more fully as $\psi(x, t; y, \tau)$ and is the probability density that at time $t > \tau$ the random variable assumes the value $x$, given that at time $\tau$ it assumes the value $y$.

If we try to find solutions of (2.1) by separation of variables and write

$$\phi(x, t) = X(x)T(t),$$

we obtain the equation

$$\frac{1}{T} \frac{dT}{dt} = \frac{1}{X(x)} \left[ -\frac{d}{dx} \{\alpha(x) X(x)\} + \frac{1}{2} \frac{d^2}{dx^2} \{\beta(x) X(x)\} \right].$$

Since both sides are functions of independent variables only, we put both sides equal to a constant $-\lambda$, and find

$$-\frac{d}{dx} \{\alpha(x) X(x)\} + \frac{1}{2} \frac{d^2}{dx^2} \{\beta(x) X(x)\} = -\lambda X(x). \tag{2.3}$$

Similarly, from (2.2) we obtain by writing

$$\psi(y, \tau) = Y(y) \mathcal{S}(\tau)$$

that

$$\alpha(y) \frac{d}{dy} Y(y) + \frac{1}{2} \beta(y) \frac{d^2}{dy^2} Y(y) = -\lambda' Y(y). \tag{2.4}$$

Although we have used different constants $\lambda$ and $\lambda'$ in (2.3) and (2.4) in fact we must take $\lambda = \lambda'$ (in which case (2.3) and (2.4) are adjoint) since these equations correspond respectively to the left and right invariant equations of $P$ for the sub-unit eigenvalues, i.e. they correspond to the invariant equations

$$v'Q = \rho v'$$

and

$$Qw = \rho w,$$

where $Q, v', w, \rho$ are as in Section 1. Since the $\lambda'$s are the analogues of the $\rho$'s, to obtain corresponding eigenfunctions $X(y), Y(y)$ to $v', w$ we must consider the same characteristic value $\lambda$ in both (2.3) and (2.4). (Note that when dealing with a chain with one or two absorbing states, the diffusion process must be arranged so that one or both, respectively, of the boundaries 0 and 1 are absorbing (exit).)

Since from the above $T(t) = e^{-\lambda t}$, it is clear that the quantities $e^{-\lambda}$ correspond to the eigenvalues $\rho$ of the matrix, and to obtain our quasi-stationary distributions, since we consider maximal $\rho_t, \rho_t$, we must consider the smallest $\lambda_t$. In practice, as shown in the next Section, we obtain this from the knowledge of $\rho_t$ to the same degree of accuracy as the diffusion approximation. Moreover, we require that $X_t(x) > 0$ and $Y_t(y) > 0$ on $(0, 1)$, by analogy with the discrete time case, and also

$$\int_{0^+}^{1-} X_t(x) dx = \int_{0^+}^{1-} X_t(x) Y_t(x) dx = 1,$$
corresponding to the normalization of the eigenvectors $v'$ and $w$ in this case. In the examples we consider, these are also integrability conditions and are sufficient to pick out the required particular solutions of (2.3) and (2.4); so that we obtain the diffusion density approximation to the quasi-stationary distributions \{v_i\} and \{w_i v_i\}.

3. THE REVERSE EQUATION

Because of the discussion of Section 1 regarding the time-reversed process, we expect that the product density will satisfy an ordinary stationary equation of form

\[-\frac{d}{dx} \{x_R(x)f(x)\} + \frac{1}{2} \frac{d^2}{dx^2} \{\beta_R(x)f(x)\} = 0, \tag{3.1}\]

where $\alpha_R(x)$ and $\beta_R(x)$ are respectively the drift and diffusion coefficients for the diffusion approximation to the reverse process whose transition matrix is (1.4). In attempting to find such an equation we may either substitute $f(x)/X_i(x)$ into equation (2.3) or substitute $f(y)/X_i(y)$ into (2.4). The general theory of second-order linear homogeneous equations guarantees that the equation for $f(x)$ so obtained in each case can be written in perfect differential form in two ways (Ince, 1965, Section 5.3). The initial questions of interest are therefore as follows.

(a) Can the equation so obtained for $f(x)$ be written in the special perfect differential form (3.1) in each case?

(b) If so, will the two equations obtained be the same?

(c) Is it possible to arrive, by some other method, at an equation of form (3.1) satisfied by the product density, which differs from both obtained by the above method?

We are able to prove the following theorem, which we merely state, as the proof is straightforward.

Theorem A. Let $v(x)$ be any solution (not necessarily $X_i(x)$) of (2.3) and $w(y)$ any solution (not necessarily $Y_i(y)$) of (2.4), with $\lambda = \lambda'$. Then the substitution $\phi(x) = f(x)/w(x)$ into the differential form

\[
w(x) \left[-\frac{d}{dx} \{x_R(x)\phi(x)\} + \frac{1}{2} \frac{d^2}{dx^2} \{\beta_R(x)\phi(x)\} + \lambda \phi(x)\right]
\]

and the substitution $\psi(x) = f(x)/v(x)$ into the differential form

\[
v(x) \left[\alpha(x) \frac{d}{dx} \{\psi(x)\} + \frac{1}{2} \beta(x) \frac{d^2}{dx^2} \{\psi(x)\} + \lambda \psi(x)\right]
\]

both lead to differential forms of type

\[-\frac{d}{dx} \{x_R(x)f(x)\} + \frac{1}{2} \frac{d^2}{dx^2} \{\beta_R(x)f(x)\}. \tag{3.2}\]

With the first substitution,

\[
\begin{align*}
\alpha_{R_1}(x) &= \beta(x) - w(x)^{-1} w'(x) + \alpha(x), \\
\beta_{R_1}(x) &= \beta(x).
\end{align*}
\]

(3.3)

With the second substitution,

\[
\begin{align*}
\alpha_{R_1}(x) &= \beta'(x) + \beta(x) v(x)^{-1} v'(x) - \alpha(x), \\
\beta_{R_1}(x) &= \beta(x).
\end{align*}
\]

(3.4)
Thus
\[ \beta_{R_1}(x) = \beta_{R_2}(x) = \beta(x) = \beta_R(x). \]

**Corollary.** Putting \( v(x) = X_1(x), w(x) = Y_1(x) \) (and \( \lambda = \lambda_1 \)) we obtain that the product density \( f(x) = X_1(x) Y_1(x) \) is a solution of two equations of form (3.1), with the same "diffusion" coefficients but (apparently) different "drift" coefficients.

In fact, we shall show in Section 6 that \( \alpha_{R_1}(x) = \alpha_{R_2}(x) \), when we take up the problem of the interpretation of this "reverse" equation. For the moment, we shall consider three examples from genetical theory which support the validity of this statement.

4. THE WRIGHT MODEL

We shall take two cases of the genetic model due to Wright (1931). This model considers a haploid population of size \( N \), and a locus with two possible alleles \( A \) and \( a \), and supposes that the number \( i \) of \( A \) alleles in each non-overlapping generation is a Markovian variate with transition matrix

\[ P = \{ p_{ij} \} = \begin{pmatrix} N \\ j \end{pmatrix} \beta_j (1 - \beta)^{N-j}, \quad (4.1) \]

where
\[ \beta_i = iN^{-1}(1 - \alpha_1 - \alpha_2) + \alpha_a. \]

Here \( \alpha_2 \) is the mutation rate from \( a \) to \( A \) and \( \alpha_1 \) is the reverse mutation rate. The latent roots of the transition matrix (4.1) are well known to be

\[ \tau_k = (1 - \alpha_1 - \alpha_2)^k \begin{pmatrix} N \\ k \end{pmatrix} k! N^{-k} \quad (k = 0, 1, 2, \ldots, N). \]

To apply the diffusion approximation, we must assume that \( \alpha_1 \) and \( \alpha_2 \) are of order of magnitude no larger than \( N^{-1} \) (Ewens, 1965), and for convenience we write \( \beta_1 = \alpha_1 N, \beta_2 = \alpha_2 N \). It is usual to measure time in units of \( N \) generations, and write \( x = iN^{-1} \) which is taken to be a continuous variate on \([0, 1]\); the relevant coefficients for the diffusion process on \((0, 1)\) approximating (4.1) are then

\[ \alpha(x) = \beta_2 - (\beta_1 + \beta_2) x, \quad \beta(x) = x(1-x). \]

4.1. No Mutation

When \( \alpha_1 = \alpha_2 = 0 \), both 0 and 1 are clearly absorbing barriers, and we have

\[ \begin{cases} \alpha(x) = 0, \\ \beta(x) = x(1-x), \end{cases} \quad (4.2) \]

and

\[ \rho_1 = (1 - N^{-1}) \]

for the matrix \( P \). Then since \( e^{-\lambda_1/N} \approx 1 - N^{-1} \) i.e. \( \lambda_1 \approx -N \log(1 - N^{-1}) \), we have

\[ \lambda_1 = 1 \quad (4.2a) \]

to the same degree of accuracy as the diffusion approximation (Ewens, 1965).
We now solve (2.3) and (2.4) with the values of $\alpha(x)$, $\beta(x)$ and $\lambda$ given by (4.2) and (4.2a), and obtain the relevant solutions (under the integrability conditions)

$$X_1(x) = 1 \quad (x \in (0, 1))$$

and

$$Y_1(y) = 6y(1-y) \quad (y \in (0, 1)).$$

The first of these is well known (Fisher, 1930). The diffusion approximation to the product distribution \{w_1, v_1\} is therefore the beta density

$$f(x) = 6x(1-x) \quad (x \in (0, 1)).$$

This example, in particular, supports our view that the product density provides a somewhat more satisfactory description of quasi-stationary behaviour than the limiting conditional distribution.

On application of the formulae (3.3) and (3.4) with $v(x) = X_1(x)$, $w(x) = Y_1(x)$, we obtain the same sets of coefficients in both cases:

$$\begin{align*}
\alpha_R(x) &= 1 - 2x, \\
\beta_R(x) &= x(1-x).
\end{align*} \quad (4.3)$$

### 4.2. One-way Mutation

When $\alpha_1 > 0, \alpha_2 = 0$, absorption eventually occurs at $x = 0$. We have

$$\begin{align*}
\alpha(x) &= -\beta_1 x, \\
\beta(x) &= x(1-x),
\end{align*} \quad (4.4)$$

and

$$\rho_1 = 1 - \alpha_1 = 1 - \beta_1 N^{-1}.$$ 

Hence

$$\lambda_1 \approx -N \log (1 - \beta_1/N) = \beta_1 \quad (4.4a)$$

to the same accuracy as the diffusion approximation. Substituting the values (4.4) and (4.4a) into (2.3) and (2.4), we obtain the relevant normalized solutions

$$\begin{align*}
X_1(x) &= 2\beta_1(1-x)^{\beta_1-1} \quad (x \in (0, 1)), \\
Y_1(y) &= (1 - 2\beta_1) y \quad (y \in (0, 1)).
\end{align*}$$

The first of these is originally due to Wright (1931). The diffusion approximation to the product distribution is the beta density

$$f(x) = \frac{\Gamma(2 + 2\beta_1)}{\Gamma(2\beta_1)} x(1-x)^{\beta_1-1} \quad (x \in (0, 1)).$$

On application of formulae (3.3) and (3.4) with $v(x) = X_1(x)$, $w(x) = Y_1(x)$, we again obtain the same coefficients in both cases:

$$\begin{align*}
\alpha_R(x) &= 1 - x(1+\beta_1), \\
\beta_R(x) &= x(1-x).
\end{align*} \quad (4.5)$$
5. Behaviour of a Mutant Gene with Selective Disadvantage

A model for describing the survival of a mutant gene which has been used by Fisher (1958, Chapter IV) is the simple branching process, some details of which are given below; for further information we refer to the early part of Harris (1963).

Supposing that we begin with a single ancestor (in this case the mutant gene), which produces \( j \) offspring with probability given by the coefficient of \( s^j \) in the probability generating function

\[
F(s) = \sum_{j=0}^{\infty} p_j s^j \quad (0 < p_0 < 1),
\]

and all succeeding offspring reproduce independently in the same way, then if we denote by \( Z_n \) the number of individuals in the \( n \)th generation, we have

\[
\sum_{j=0}^{\infty} \Pr(Z_n = j) s^j = F_n(s),
\]

where \( F_n(s) \) is the \( n \)th functional iterate of \( F \) on \( s \). It is well known that, if

\[
F'(1) = m < 1,
\]

extinction (absorption) is bound to occur in finite mean time. If \( m = 1 \), then extinction is bound to occur, although the mean time to extinction is infinite and, if \( m > 1 \), there is a positive probability of survival. Since \( F'(1) = m \) is the mean number of offspring per individual, and in the genetic context \( F(s) \) is the probability-generating function of the number of representations of the gene appearing in the first generation offspring of an individual mutant, we may interpret \( m < 1 \) to mean that the mutant has a selective disadvantage and \( m > 1 \) that it has a selective advantage.

Fisher (1958) has considered the particular case \( F(s) = e^{m(s-1)} \) for \( m = 1 \) and \( m > 1 \), but has not treated the case \( m < 1 \). Yaglom (1947) has extended this work to more general offspring-distributions, and has also shown that if \( m < 1 \) and \( F''(1-\to) < \infty \), then

\[
\lim_{n \to \infty} \Pr(Z_n = j \mid Z_n > 0) = v_j > 0,
\]

where \( \sum v_j = 1 \). As the simple branching process is an absorbing infinite Markov chain, with absorbing state 0 and transient states \( T = \{1, 2, \ldots\} \), this is just the limiting expression (1.1) (for Yaglom's theorem may be shown to hold for any initial probability distribution, at least provided it has finite first moment). However, the distribution \( \{v_j\} \), as in the case of a finite chain, has been obtained only for a few special cases of the (infinite) \( Q \)-matrix, i.e. for certain special forms of \( F(s) \).

It can, in fact, be shown that under Yaglom's conditions, apart from the analogy in the absorption behaviour with the finite chain case,

(a) that \( v' \) is in fact the left eigenvector of \( Q \) corresponding to the eigenvalue \( m \), which thus plays the role of the largest eigenvalue, and

(b) \( v_j > 0 \) for \( j = 1, 2, 3, \ldots \) if \( Q \) is indecomposable—the necessary and sufficient condition for which is \( p_j > 0, p_j > 0 \) for some \( j > 1 \).

The right eigenvector \( w \) corresponding to \( m \) in this case is a constant multiple of \( w = [1, 2, 3, \ldots] \). It is also a consequence of the above-mentioned theorem of Yaglom that

\[
\infty > \sum_{1}^{\infty} iv_1 = K^{-1} > 0.
\]
We therefore choose \( w \) as \( K[1, 2, 3, \ldots]' \), so that
\[
\sum_{1}^{\infty} w_i v_i = 1,
\]
and then the product sequence \( \{w_i v_i\} \equiv \{K v_i\} \) forms a distribution over \( T \). This is, as expected, the product distribution defined by (1.3). The details of what has been stated here without proof will appear elsewhere.

It is therefore clear that, although the state space in this process is not finite, the behaviour of the process under Yaglom's conditions is very similar to the finite case. It is not possible to manipulate the process so that it is approximated by a diffusion process on \((0,1)\), nor can we use any of the justifications for the diffusion process in the same way as for the finite case (see, e.g. Ewens, 1965). However, Feller (1951) has shown that, if new units are introduced for measuring time and population such that during small intervals of time the changes of population size will be small, then a diffusion approximation, with a known estimate of the error, is possible.

In particular, if \( m = 1 - \delta \) where \( \delta \) is small and positive, and an individual in the old counting and the time of one generation are both \( \delta \), then the density function \( \phi(x, t) \) of the modified variate \( x \) at modified time \( t \) satisfies
\[
\frac{\partial \phi(x, t)}{\partial t} = \frac{\partial}{\partial x}\{x \phi(x, t)\} + \frac{1}{2} \sigma^2 \frac{\partial^2}{\partial x^2}\{x \phi(x, t)\} \quad (0 < x < \infty),
\]
where \( \sigma^2 \) is the variance of the offspring distribution with generating-function \( F(s) \). The estimate of the error in making the approximation is \( O(\delta) \). We thus have
\[
\begin{align*}
\alpha(x) &= -x, \\
\beta(x) &= \sigma^2 x,
\end{align*}
\]
so that
\[
e^{-\lambda_1 t} = m, \quad \lambda_1 = -\delta^{-1} \log(1 - \delta) = 1 + O(\delta),
\]
i.e. \( \lambda_1 = 1 \) to the same accuracy as the diffusion approximation. Proceeding as outlined in Section 2 (the only difference being that the diffusion is now on \((0, \infty)\)), we obtain the relevant normalized solutions
\[
X_1(x) = (2/\sigma^2) e^{-2x/\sigma^2} \quad (x \in (0, \infty)),
\]
\[
Y_1(y) = 2y/\sigma^2 \quad (y \in (0, \infty)).
\]
The diffusion approximation to the product density is then
\[
f(x) = (2/\sigma^2)^3 x e^{-2x/\sigma^2} \quad (x \in (0, \infty)).
\]
Notice that the value \( Y_1(y) \propto y \) agrees with the true value of \( w \propto [1, 2, 3, \ldots]' \). Moreover, we once more obtain the same coefficients from (3.3) and (3.4):
\[
\begin{align*}
\alpha_R(x) &= \sigma^2 - x, \\
\beta_R(x) &= \sigma^2 x.
\end{align*}
\]
A slightly more general form of (5.1) has actually been solved completely by Feller (1939) under the initial condition
\[
\phi(x, 0) = \{1 \text{ for } x = N, 0 \text{ otherwise}\}.
\]
This solution, for (5.1), is
\[
\phi(x, t) = \frac{4Ne^{-t}}{\sigma^2(e^{-t} - 1)^2} \exp \left[ \frac{2(Ne^{-t} + x)}{\sigma^2(e^{-t} - 1)} \right] \sum_{s=0}^{\infty} \frac{1}{s!(s+1)!} \left[ \frac{2}{\sigma^2(e^{-t} - 1)} \sqrt{(xNe^{-t})} \right]^{2s}.
\] (5.4)

From this the limiting interpretations of our two quasi-stationary distributions in the sense of (1.1) and (1.3) may be obtained. For example,
\[
\lim_{t \to \infty} \left\{ \phi(x, t) \int_0^\infty \phi(x, t) \, dx \right\} = (2/\sigma^2) e^{-2\alpha x/\sigma^2} X_1(x) \quad (x \in (0, \infty)),
\]
and in a similar way the product density \( Y_1(x) \) can be found directly from (5.4).

6. THE EXTENSION OF WRIGHT'S FORMULA AND INTERPRETATION OF THE “REVERSE” PROCESS

From the theory of second-order homogeneous linear differential equations, it is well known that if \( v(x) \) is the general solution of equation (2.3), then the general solution \( w(x) \) of (2.4) is related by the formula \( \lambda = \lambda' \) (in (2.3) and (2.4))
\[
v(x) = \frac{w(x)}{\beta(x)} \exp \left\{ 2 \int_0^x \frac{\alpha(x)}{\beta(x)} \, dx \right\}
\] (6.1)
(see Ince, 1956, Section 5.3). The formula also relates particular solutions of the two equations, and if \( v(x) \) is taken as \( X_1(x) \), the eigenfunction corresponding to \( \lambda \), it follows from the spectral theory of such equations that we can write
\[
X_1(x) = \text{const} \frac{Y_1(x)}{\beta(x)} \exp \left\{ 2 \int_0^x \frac{\alpha(x)}{\beta(x)} \, dx \right\},
\] (6.2)
where \( Y_1(x) \) is the eigenfunction of the adjoint equation corresponding to \( \lambda \). We then have

**Lemma 1**

\[
\alpha_{R_1}(x) = \alpha_{R_1}(x).
\]

**Proof.** From (3.3) and (3.4),
\[
\alpha_{R_1}(x) = \beta(x) \{ Y_1(x) \}^{-1} Y'_1(x) + \alpha(x),
\]
\[
\alpha_{R_1}(x) = \beta'(x) + \beta(x) \{ X_1(x) \}^{-1} X'_1(x) - \alpha(x).
\]

Now from (6.2),
\[
\log X_1(x) = \text{const} + \log Y_1(x) + 2 \int_0^x \frac{\alpha(x)}{\beta(x)} \, dx - \log \beta(x).
\]
Differentiating and rearranging,
\[
X_1(x) \{ X_1(x) \}^{-1} = Y'_1(x) \{ Y_1(x) \}^{-1} + 2\alpha(x)/\beta(x) - \beta'(x)/\beta(x),
\]
whence
\[
\beta'(x) + \beta(x) \{ X_1(x) \}^{-1} X'_1(x) - \alpha(x) = \beta(x) + \{ Y_1(x) \}^{-1} Y'_1(x) + \alpha(x),
\]
so that
\[
\alpha_{R_1}(x) = \alpha_{R_1}(x).
Lemma 2.

\[ f(x) = \text{const} \left( \frac{Y_1(x)}{\beta(x)} \right)^2 \exp \left\{ 2 \int_{x}^{\cdot} \frac{\alpha(x)}{\beta(x)} \, dx \right\}, \quad (6.3a) \]

\[ f(x) = \text{const} \left( \frac{X_1(x)}{\beta(x)} \right)^2 \beta(x) \exp \left\{ -2 \int_{x}^{\cdot} \frac{\alpha(x)}{\beta(x)} \, dx \right\}, \quad (6.3b) \]

where \( f(x) \) is the product distribution \( X_1(x) Y_1(x) \). This follows immediately from (6.2).

The formulae (6.2), (6.3a) and (6.3b) are clearly the quasi-stationary generalization of Wright’s (1937) formula for diffusion processes possessing an ordinary stationary distribution \( \phi(x) \), given by

\[ \phi(x) = \frac{\text{const}}{\beta(x)} \exp \left\{ 2 \int_{x}^{\cdot} \frac{\alpha(x)}{\beta(x)} \, dx \right\}, \quad (6.4) \]

which can be obtained from (6.2) merely by making the correct substitution in (6.2) for the “right” eigenfunction \( Y_1(x) \), viz. \( Y_1(x) = \text{const} \).

It remains only to discuss the interpretation of the “reverse” equation

\[ \frac{\partial}{\partial t} f(x, t) = -\frac{\partial}{\partial x} \left( \alpha_H(x) f(x, t) \right) + \frac{1}{2} \frac{\partial^2}{\partial x^2} \left( \beta_H(x) f(x, t) \right). \]

The “stationary” form is

\[ 0 = -\frac{d}{dx} \left( \alpha_H(x) f(x) \right) + \frac{1}{2} \frac{d^2}{dx^2} \left( \beta_H(x) f(x) \right), \]

of which we know the product distribution \( X_1(x) Y_1(x) \) is a solution.

Let us suppose that the coefficients are such as to describe a diffusion process which admits a stationary distribution (on the same interval as the “absorbing” equation from which the \( \alpha_H(x) \) and \( \beta_H(x) \) were obtained). This is, in fact, true for the three examples we have considered. Then this stationary distribution is given by Wright’s formula (6.4), and is

\[ \phi(x) = \frac{\text{const}}{\beta_H(x)} \exp \left\{ 2 \int_{x}^{\cdot} \frac{\alpha_H(x)}{\beta_H(x)} \, dx \right\}, \]

and since

\[ \int_{x}^{\cdot} \frac{\alpha_H(x)}{\beta_H(x)} \, dx = \int_{x}^{\cdot} \left[ Y_1'(x) (Y_1(x))^{-1} + \frac{\alpha(x)}{\beta(x)} \right] \, dx \]

\[ = \log Y_1(x) + \int_{x}^{\cdot} \frac{\alpha(x)}{\beta(x)} \, dx, \quad (6.5a) \]

it follows from (3.3) that

\[ \phi(x) = \frac{\text{const}}{\beta(x)} \left( Y_1(x) \right)^2 \exp \left\{ 2 \int_{x}^{\cdot} \frac{\alpha(x)}{\beta(x)} \, dx \right\}. \]

Similarly, by using (3.4),

\[ \int_{x}^{\cdot} \frac{\alpha_H(x)}{\beta_H(x)} \, dx = \log \beta(x) + \log X_1(x) - \int_{x}^{\cdot} \frac{\alpha(x)}{\beta(x)} \, dx \quad (6.5b) \]
gives
\[ \phi(x) = \text{const} \{ X_1(x) \}^2 \beta(x) \exp \left\{ -2 \int_{x}^{\alpha(x)} \frac{\alpha(x)}{\beta(x)} \, dx \right\}. \]

Thus, from (6.3a) and (6.3b), this stationary distribution is the product distribution.

The problem of whether the "reverse" equation always has a stationary distribution can be resolved by application of the Feller boundary criteria, which essentially involve (6.5a) and (6.5b). However, this will be taken up elsewhere. Assuming the existence of a stationary distribution (as for the three cases we have considered) it is logical to expect that the "reverse" process is the diffusion approximation to (1.4). However, this is, generally speaking, a conjecture, since if we apply our methods to the case when a true stationary distribution exists, the "reverse" equation is merely the same as our initial equation.

An alternative approach to both problems is to show that the "reverse" process is the diffusion approximation to the irreducible chain (1.4). The general proof of this fact appears difficult. However, for absorbing chains of the particular type which often occurs in genetics, when the diffusion approximation is obtained by normalizing the state variables by the largest (as in Section 4), the following sketch-proof has been suggested by W. J. Ewens (for any particular normalization, a similar argument appears possible).

For the chain whose matrix is (1.4),
\[ P^* = \{ p_{ij}^* \}, \quad \text{where} \quad p_{ij}^* = v_j p_{ji} / \rho_i \quad (i, j = 1, \ldots, n), \]
the drift and diffusion coefficients \( \alpha^*(x), \beta^*(x) \) for the relevant diffusion approximation may be obtained from the mean and mean-square increases (from arbitrary state \( i \in T \)), namely \( \Sigma_i (j-i) p_{ij}^* \) and \( \Sigma_j (j-i)^2 p_{ij}^* \). By writing \( i/n = x, j/n = x + \delta x \), we get
\[
\alpha^*(x) \approx \int_{\delta x} (\delta x) \frac{v(x+\delta x)}{v(x)} \, p_{x+\delta x}(-\delta x) \, d(\delta x)
\]
\[
= \int_{\delta x} (\delta x) \left\{ 1 + (\delta x) \frac{v'(x)}{v(x)} + O(\delta x)^2 \right\} \times \{ p_x(-\delta x) + (\delta x) p_x^2(-\delta x) + O(\delta x)^3 \} \, d(\delta x)
\]
\[
= -\alpha(x) + \beta(x) \frac{v'(x)}{v(x)} + \beta'(x) + O(\delta x),
\]
and
\[
\beta^*(x) \approx \int_{\delta x} (\delta x)^2 \frac{v(x+\delta x)}{v(x)} \, p_{x+\delta x}(-\delta x) \, d(\delta x)
\]
\[
= \int_{\delta x} (\delta x)^2 \left\{ 1 + (\delta x) \frac{v'(x)}{v(x)} + O(\delta x)^2 \right\} \times \{ p_x(-\delta x) + (\delta x) p_x^2(-\delta x) + O(\delta x)^3 \} \, d(\delta x)
\]
\[
= \beta(x) + O(\delta x).
\]

Here \( \alpha(x) \) and \( \beta(x) \) are the relevant coefficients for the forward process, and \( v(x) \) (the diffusion approximation to \( v' \)) can be replaced by \( X_1(x) \) as before. Moreover, the approximation \( O(\delta x) \) is no worse than that used to arrive at the forward equations.
Thus
\[ \alpha^*(x) = -\alpha(x) + \beta(x) X'(x)(X_1(x))^{-1} + \beta'(x) \]
and
\[ \beta^*(x) = \beta(x), \]
so that
\[ \alpha^*(x) = \alpha_R(x), \quad \beta^*(x) = \beta(x) \]
as required.

Finally, for completeness, we shall attempt to explain, in intuitive terms, why the product distribution should be the stationary distribution of the reverse process. This is most simply done by referring to the discrete chain and considering the limiting expression (1.3). A similar discussion is contained in the paper of Darroch and Seneta (1965). We have that
\[ \lim_{m \to \infty} \lim_{n \to \infty} \Pr(\text{process in state } j \text{ at time } m < n | \text{absorption has not occurred at time } n) = w_j v_j. \]

If the \( Q \) matrix were stochastic, then the conditioning would be irrelevant, \( w_j = 1 \), and the above expression would be a statement of the ergodic property of a regular Markov chain, of which the reverse matrix \( P^* \) is the matrix of the (normally defined) reverse chain. In the present case, however, the matrix \( Q \) is more general than this, and the above equation with the conditioning limit taken first, rather than (1.1), is the logical generalization of the ergodic property to a sub-stochastic matrix.

It is therefore natural to expect that, just as the stationary distribution of an ordinary regular chain is also the stationary distribution of the reverse chain, then, in the more general case, the product distribution will be the stationary distribution of the “reverse” chain \( P^* \). This last is defined analogously to the normal reverse chain, but only on the transient states, with the dividing element \( r_1 \) “compensating” for the probability efflux out of \( T \), and thus ensuring stochasticity of the new matrix. Essentially, this is an aspect of the generalization of Markov chain theory to matrices with uniformly bounded row sums.

The role of the weighting factors \( w_t \) on the (possibly) intuitively simpler distribution \( \{v_t\} \) to form the distribution \( \{w_t v_t\} \) is more clearly seen when one considers analogous theory for matrices with a denumerably infinite state space, and in particular the simple branching process and the random walk on the half-line. This will be developed in a forthcoming paper with D. Vere-Jones.

ACKNOWLEDGEMENTS

The author is indebted to Dr W. J. Ewens and the referee for several helpful suggestions.

REFERENCES


**Discussion on Mr Seneta’s Paper**

Professor D. G. Kendall: The idea of a quasi-stationary distribution has been on the agenda since M. S. Bartlett’s famous paper on competitive and predatory biological systems (Bartlett, 1957). We are very much in debt to our Australian colleagues for hammering out a careful formulation of this very important idea and for developing so many interesting applications of it. Darroch, Ewens and Seneta have concentrated most of their theoretical work on cases in which there is either a finite number of states or a continuum of states, so that they are working either with finite matrices or with differential equations. It therefore seems worth while setting out the corresponding theory for the case of a countable infinity of states. Here the basic ideas are derived from the work of Vere-Jones (1961, 1962). I shall therefore expect to be told that everything I have to say is well known to the rest of the Australian school. It cannot be said, however, to be well known in this country and, as I have always found the work of Vere-Jones particularly startling and attractive, I welcome the opportunity of sketching some of it here.

Let A denote an infinite matrix of non-negative elements \((a_{ij}: i, j = 1, 2, \ldots)\); we suppose that all its iterates \(A^n\) formed by multiplication have finite elements \((a_{ij}^{(n)}: i, j = 1, 2, \ldots)\), and that \(A\) is irreducible and aperiodic (i.e. \(a_{ij}^{(n)}\) is positive for at least one positive \(n = n(i, j)\), and for each \(j\) the set of positive \(n\)-values for which \(a_{ij}^{(n)} > 0\) has unity as greatest common divisor). Note that \(A\) is not necessarily a matrix of transition probabilities.

Now the numbers \(f(n) = a_{j1}^{(n)}\) form (for each \(j\)) a super-multiplicative sequence; \(f(m+n) \geq f(m)f(n)\). It follows from an extension of a classical theorem of analysis (Kingman, 1963a, p. 355) that the limit of the \(n\)th root of \(f(n)\) exists, and that

\[
0 < \lim_{n \to \infty} (a_{j1}^{(n)})^{1/n} = \sup_{n \geq 1} (a_{j1}^{(n)})^{1/n} = 1/R_j < \infty,
\]

(Because \(a_{j1}^{(n)}\) is positive for at least one positive \(n\) it is clear that \(R_j\) cannot be infinite; it might, however, be zero.) The basic theorem of Vere-Jones is that \(R_j\ does not depend on \(j\), and he further shows that if \(R = R_j\ then the series

\[
\sum_n a_{ij}^{(n)} r^n
\]

are all convergent for \(r < R\), and all divergent for \(r > R\).

In the critical case when \(r = R\ the series again all behave in the same way, whatever \(i\) and \(j\) may be, and there are now just three possibilities:

(i) the series converge, so that \(a_{ij}^{(n)} R^n \to 0\ as n \to \infty\);

(ii) the series diverge, but still \(a_{ij}^{(n)} R^n \to 0\ as n \to \infty\);

(iii) the series diverge, and \(a_{ij}^{(n)} R^n \to y_i x_j\ as n \to \infty\, where the numbers

\[
x_1, x_2, \ldots, y'_1, y'_2, \ldots
\]

are all positive and \(\sum x_i y_i = 1\).
Clearly it is only in case (i) that we can have \( R = 0 \). The matrix \( A \) is called \( R \)-transient in case (i), \( R \)-null-recurrent in case (ii), and \( R \)-positive-recurrent in case (iii). It will be noticed that the important phenomenon of \( R \)-positive-recurrence can be detected by examining the behaviour of the matrix elements for just one pair \( (i, j) \), with \( i = j \) if desired; thus a necessary and sufficient condition is that \( q^{(n)}_{ii} / R^n \) should not tend to zero as \( n \) tends to infinity.

However, in the application of the theory it is important that one should not be required to know all the matrix iterates; fortunately this is not necessary, because Vere-Jones has also provided a sort of spectral characterization; in fact the vectors

\[
\mathbf{x} = (x_1, x_2, ...) \quad \text{and} \quad \mathbf{y} = (y_1, y_2, ...)
\]

in the \( R \)-positive-recurrent case satisfy the linear equations

\[
\sum_i x_i a_{ij} = x_j / R, \tag{2}
\]

\[
\sum_i a_{ij} y_j = y_i / R, \tag{3}
\]

and apart from constant factors they are the unique non-negative (necessarily therefore strictly positive) solutions to (2) and (3). The constant factor is effectively determined by the condition \( \sum x_i y_i = 1 \). In this way, therefore, the limits at (iii) can in principle be found if we know that we are in the \( R \)-positive-recurrent case.

It is a very important complementary fact that the \( R \)-positive-recurrent matrices \( A \) can be identified by making use of the equations (2) and (3):

\[
\text{For some finite positive } R \text{ let the equations (2) and (3) possess non-negative non-null solutions } \mathbf{x} \text{ and } \mathbf{y} \text{ such that } \sum x_i y_i \text{ is convergent. Then } R = R_i \text{ in (1), the matrix } A \text{ is } R \text{-positive-recurrent, and if } \mathbf{x} \text{ and } \mathbf{y} \text{ are normalized so that } \sum x_i y_i = 1 \text{ then (iii) holds with this } \mathbf{x} \text{ and } \mathbf{y}. 
\]

**Proof.** Let \( 0 < R < \infty \) and let the irreducible aperiodic matrix \( A \) with non-negative elements and finite iterates be such that (2) has a solution \( \mathbf{x} \) where each \( x_i \geq 0 \) and some \( x_i > 0 \). Then

\[
\sum_i x_i a_{ij}^{(n)} = x_j / R^n, \tag{4}
\]

and so (by irreducibility) every \( x_i > 0 \). Now put

\[
(Q^n)_{ij} = q_{ij}^{(n)} = R^n x_i a_{ij}^{(n)} / x_j \quad (n = 0, 1, 2, ...); \tag{5}
\]

then these form a system of Markov transition probabilities satisfying the Chapman–Kolmogorov equation, and \( Q \) will be irreducible and aperiodic in virtue of the assumptions about \( A \). Such a chain may or may not possess a positive invariant measure; that is, there may or may not exist positive finite numbers \( m_i \) (with \( \sum m_i < \infty \)) such that

\[
\sum_i m_i q_{ij} = m_j. \tag{6}
\]

The necessary and sufficient condition for this to be so has been discovered by Harris (1957) and Veech (1963). (See also Pruitt (1964).) In particular, it is **sufficient** that there should be only finitely many non-zero entries in each column of \( Q \) (i.e. that there should be only finitely many non-zero entries in each row of \( A \)). The question of the existence of such an \( m \) is exactly the question of the existence of a non-negative non-null solution \( y \) to (3), for such a \( y \) must then necessarily have strictly positive elements, and \( m_i = x_i y_i \) will be a solution to (6), and conversely. Now, suppose that \( y \) exists (i.e. that the Harris–Veech condition is satisfied) and further that \( \sum x_i y_i < \infty \). Then \( \sum m_i < \infty \); this, however, ensures that the \( Q \)-chain is positive recurrent in the usual sense, and so by the Kolmogorov limit theorem we shall have \( q_{ij}^{(n)} \to m_j / \sum m_i \). On normalizing \( x \) and \( y \) so that \( \sum x_i y_i = 1 \) (as can certainly be done), and expressing \( q_{ij}^{(n)} \) in terms of \( a_{ij}^{(n)} \), we obtain (iii), from which it is now obvious that \( R \) must be the \( R_j \) of (1). The proof is complete.
Now let us identify $A$ with either (1) an irreducible transient aperiodic stochastic matrix, or (2) the perhaps sub-stochastic matrix associated with a maximal inter-communicating set $C$ of states of this character, and suppose that it has been shown to be $R$-positive-recurrent either directly or by means of the criterion just given. In view of the probabilistic interpretation it will be convenient to write $P$ for $A$, etc. We shall consider the behaviour of the Markov process $(Z_t; t \geq 0)$ associated with $P$ under the following two side-conditions: that it was in state $i$ in $C$ at a remote time $-T$ in the past, and that it will be in state $j$ in $C$ at a remote time $+S$ in the future. If $t_1 < t_2 < \cdots < t_r$ and if $k_1, k_2, \ldots, k_r$ are arbitrary states in $C$, then we shall have

$$
Pr(Z_{t_n} = k_n \text{ for } n = 1, 2, \ldots, r \mid Z_{-T} = i, Z_{+S} = j) = p_{k_1}^{(T+t_1)} p_{k_2}^{(t_2-t_1)} \cdots p_{k_r}^{(t_r-t_{r-1})} p_{k_n}^{(S-t_{r-1})}
$$

In this last expression we multiply numerator and denominator by

$$
R^{T+S} = R^{T+t_1} R^{t_2-t_1} \cdots R^{t_r-t_{r-1}} R^{S-t_r},
$$

and let $S$ and $T$ tend to $+\infty$. The limit we get for the conditional probability will be

$$
\pi_{k_1} \Pi_{k_2}^{(t_2-t_1)} \cdots \Pi_{k_r}^{(t_r-t_{r-1})}
$$

where $\pi$ is the probability distribution $\pi_k = x_k y_{k^*}$, and $\Pi$ is the stochastic matrix with elements

$$
(\Pi)_{k_1 k_2} = \frac{R}{y_{k_1}} p_{k_1 k_2} y_{k_2}.
$$

Thus, under this limiting two-point boundary condition, the chain behaves as if $\pi$ were its stationary distribution and $\Pi$ were its transition matrix, the conditioned chain being Markovian like the original one.

This appears to be the general formulation of the theory of quasi-stationary distributions in the denumerable case. Analogous results for a continuous time-parameter can be constructed by making use of Kingman's variants of Vere-Jones's theorems (Kingman, 1963a, b).

An example of this procedure can be found by reinterpreting the analysis in a recent paper of mine about super-critical branching processes with a positive chance of extinction (Kendall, 1965a). Here $p_j$ is the chance that an individual will have $j$ children, and the only assumptions are that

$$
p_0 > 0 \quad \text{and} \quad 1 < \sum j p_j < \infty,
$$

although for conformity with the present discussion we must add further conditions on the $(p_j)$ to make the set $j \neq 0$ intercommunicating and aperiodic. The set $C$ of states with labels $j = 1, 2, \ldots$ (thus excluding the state of extinction $j = 0$) is then to be identified with the set $C$ in the preceding paragraphs. It is found in Kendall (1965a) that this system is $R$-positive-recurrent with $R = 1/f(q)$, where $f(s) = \sum p_j s^j$ and $q$ is the chance of extinction.

The vector $y$ is easily found but the vector $x$ is difficult to compute from (2); in principle, it can be determined by a functional equation. The quasi-stationary distribution is of the form

$$
\pi_i = i q^{i-1} x_i \quad (i = 1, 2, \ldots)
$$

One special case is worked out by Kendall (1965a) in full to illustrate the theory. It is a consequence of these results that

$$
p_{i,j}^{(n)} = i q^{i-1} x_j \beta^n(1 + o(1)) \quad (i, j \neq 0; n \to \infty).
$$

In two very remarkable papers Karlin and McGregor (1966a, b) have recently given a complete series expansion for $p_{i,j}^{(n)}$ in powers of $\beta^n$, but it is not seen how the second and later terms in their expansion can be interpreted from the present point of view.
In conclusion, I should like to mention another conditional distribution which I have recently found, again in a biological context. For the simple (separable) birth process in continuous time we know from martingale theory that

$$\log Z_t = \lambda t + \log W + o(1) \quad (t \to \infty)$$

(11)

with probability one, where $W$ is a random variable whose distribution is known (Blackwell and Kendall, 1964). Now $W$ just locates the random linear asymptote to the stochastic logarithmic growth curve; suppose we fix this asymptote, thus prescribing the way the process is to behave at $t = \infty$; it can then be shown (Kendall, 1965b) that there exists a well-behaved conditional distribution for the sample paths, and up to a random time change the conditioned process is a Poisson process. This kind of result is particularly useful because we can now write down any property of the Poisson process we like, and by integration obtain new information about the simple birth process. Sometimes what we get is fairly trite, sometimes less so. Thus, if we transform the law of the iterated logarithm for the Poisson process, we find that we can replace the $o(1)$ in (11) by

$$\zeta(t) e^{-\lambda t} \sqrt{\frac{2 \log \lambda t}{W}},$$

(12)

where almost surely the function $\zeta(t)$ has the whole segment $[-1, 1]$ as its exact cluster set when $t \to \infty$.

Further examples of this sort of behaviour would be worth looking for. I have examined the simple birth and death process in the same way but unfortunately the results are much more complicated.

It gives me great pleasure to propose this vote of thanks. It will be obvious that I have found the first few pages of this paper very stimulating, and I look forward to hearing the comments of others on a by no means exhausted topic.

Professor J. Gani: It is a pleasure for me to express appreciation of this paper by Mr Seneta on the application of quasi-stationary distribution methods to diffusion approximations in genetic models. I should at the same time like to place on record my thanks to Professor Whittle for his spirited reading of this work to us in London. I hope I am voicing more than my personal opinion here tonight when I say how greatly I welcome the presentation, if necessary by proxy, of research by our Australian colleagues at the Society’s meetings. I shall look forward to a continuation of this policy, and even perhaps to its extension to all Fellows of the Society outside Britain.

Mr Seneta has made a valuable addition to the understanding of certain problems in mathematical genetics to whose study Professor P. A. P. Moran and his school have contributed so much in this last decade at Canberra. An account of these may be found in Moran’s book referred to above, and in his own, Wautherson’s and Ewens’s subsequent papers.

While at Canberra, I too became interested in a genetic model of Moran’s with overlapping generations, which contrasts with Wright’s non-overlapping generations model considered by Seneta in Section 4 of his paper. The effects of this difference on the stationary gene frequency distribution have already been pointed out by Moran when mutation in two directions is allowed (cf. pp. 131 et seq. of his book).

Moran’s model consists of a Markov chain embedded in a birth–death process representing the random fluctuation of gene frequencies in a population of haploid monoeocious individuals subject to mutation and selection. If $N$ is the total number of gametes $a, A$, the system is in state $j$ when $j$ of these are $a$, and $N-j$ are $A$. If mutations from $a$ to $A$, and $A$ to $a$ are allowed, and the probabilities of these are $\alpha_1, \alpha_2$ $(0 \leq \alpha_1, \alpha_2 \leq 1)$, the transition probabilities $p_{jk}(\alpha_1, \alpha_2) (j, k = 0, 1, \ldots, N)$ for the Markov matrix are

$$p_{j,j-1} = \frac{j}{N} q_j, \quad p_{j,j} = \frac{j}{N} p_j + \left(1 - \frac{j}{N}\right) q_j, \quad p_{j,j+1} = \left(1 - \frac{j}{N}\right) p_j, \quad p_{ik} = 0 \quad \text{for} \quad |k-j| > 1,$$
where
\[ p_j = \frac{1}{N} (1 - \alpha_2) + \left( 1 - \frac{j}{N} \right) \alpha_2, \quad q_j = \frac{1}{N} \alpha_1 + \left( 1 - \frac{j}{N} \right) (1 - \alpha_2). \]

Moran (1958) had correctly surmised the form of the eigenvalues for \( j = 1, 2 \), and in that paper’s appendix E. J. Hannan obtained their exact values for all \( j \) when there was no mutation (\( \alpha_1 = \alpha_2 = 0 \)). My own modest contribution (Gani, 1961) was to find the general form of the eigenvalues for all \( j \) as
\[ \rho_j(\alpha_1, \alpha_2) = 1 - \frac{j}{N} (\alpha_1 + \alpha_2) - \frac{j(j - 1)}{N^2} (1 - \alpha_1 - \alpha_2). \]

I propose to outline briefly my working through of Seneta’s method in this particular overlapping generations model to obtain the quasi-stationary distributions of the relevant diffusion approximations. For this purpose, we allow only one-way mutation from \( A \) to \( a \), and let \( \alpha_1 = 0 \), \( \alpha_2 = \alpha \) for simplicity, and \( \alpha N = \beta \), \( x = j/N \).

(a) No mutation. Here \( \alpha = 0 \) and \( x = 0 \), both absorbing barriers. We find
\[ \alpha(x) = 0, \quad \beta(x) = 2x(1-x), \quad \rho_1 = 1 - 2/N^2, \]
and since here\(\exp(-\lambda_1/N^2) \simeq 1 - 2/N^2 \), then \( \lambda_1 \simeq 2 \). Hence, as in Seneta’s example,
\[ X_t(x) = 1 \quad (x \in (0, 1)), \]
\[ Y_t(y) = 6y(1-y) \quad (y \in (0, 1)), \]
and
\[ f(x) = 6x(1-x). \]

(b) One-way mutation. For \( \alpha > 0 \), absorption will occur at \( x = 0 \).

We find
\[ \alpha(x) = -\beta x, \quad \beta(x) = 2x(1-x), \quad \rho_1 = 1 - \alpha/N, \]
and since \( \exp(-\lambda_1/N^2) \simeq 1 - \beta/N^2 \), then \( \lambda_1 \simeq \beta \).

Thus
\[ X_t(x) = \beta(1-x)^{\beta-1} \quad (x \in (0, 1)), \]
\[ Y_t(y) = (1-\beta) y \quad (y \in (0, 1)), \]
and
\[ f(x) = (1+\beta) \beta x(1-x)^{\beta-1} \quad (x \in (0, 1)). \]

We see that whereas the case of no mutation leads to identical quasi-stationary distributions in both the overlapping and non-overlapping generations models, these distributions are different when one-way mutation is allowed.

May I conclude by congratulating Mr Seneta once again on his most interesting paper; I have much pleasure in seconding the vote of thanks.

The vote of thanks was put to the meeting and carried unanimously.

Dr D. M. G. Wishart: The idea of a quasi-stationary distribution may also be useful in the study of birth and death processes or simple queues. Suppose we have a birth and death process on the integers \( \{0, 1, \ldots, N\} \) with a slight drift away from the origin, and we are interested in the first passage time to \( N \) from \( 0 \). The state probabilities are given by
\[ p(t) = p(0) \exp(tA), \]
or, using Jensen’s trick (Jensen, 1953; Keilson and Wishart, 1964),
\[ p(t) = p(0) \exp (-\nu t(I-B)), \]
where \( B \) is a stochastic matrix. By appropriate re-ordering of the states, \( B \) may be put in the form considered by Seneta, and a quasi-stationary distribution on \( \{0, 1, \ldots, N-1\} \) determined from the sub-matrix \( Q \) of \( B \) associated with the motion among these states.

\(\dagger\) Time is measured in units of \( N^2 \) birth–death events.
Mr R. Morley Jones: I have always felt a little uneasy about the diffusion approximation to genetical processes because, as has been known for a long time, they do not represent the situation accurately near the ends, however large the population. The approach to a diffusion process is, so to speak, non-uniform in these regions, and strictly speaking a different form of approximation, due to Fisher, is required there. (Perhaps I should say also due to Fisher.)

But the exact stationary or quasi-stationary distributions cannot usually be found explicitly. For the process with two-way mutation, whose transition matrix is given in equation (4.1), however, Feller’s derivation of the latent roots also yields the latent column vectors in terms of the elements of an upper triangular matrix \( \{ a_{nj} \} \) with the diagonal elements all unity. Thus the \( ih \)th element of the \( k \)th latent column vector \( (k = 0, 1, \ldots, N) \)

\[
x_{ik} = \sum_{\nu=0}^{k} a_{\nu i} L_{(\nu)}
\]

The latent row vectors can, of course, be obtained in principle by inverting the matrix \( \{ a_{nj} \} \). But in fact one need only invert \( \{ a_{mj} \} \), yielding another upper triangular matrix \( \{ b_{mj} \} \), and then it is not difficult to show that the \( jh \)th element of the \( k \)th latent row vector is

\[
y_{kj} = \sum_{\nu=j}^{N} (-1)^{j+\nu} \frac{\nu!}{\nu!} \binom{\nu}{j} b_{\nu k}
\]

Further, \( b_{00} \) is the \( v \)th factorial moment of the stationary distribution. It is possible to calculate the lower moments explicitly in this way, although there are easier ways of doing so.

Dr H. D. Miller: What I had intended to say has already essentially been said by Professor Kendall. However, it is perhaps worth saying anyhow since my own remarks are connected with the finite chain rather than with the denumerable case discussed by Professor Kendall.

The author has observed that \( (w_j, v_j) \) is the unique stationary vector of a “reverse” Markov chain defined on the transient states. However, there is also a “forward” chain intimately associated with the transient states and \( (w_j, v_j) \) turns out to be the unique stationary vector of this chain too.

Let \( \mathbf{Q} \) be a non-negative square matrix, not necessarily sub-stochastic, and suppose that \( \mathbf{Q} \) is indecomposable and non-cyclic. If \( (w_j) \) is the positive right eigenvector of \( \mathbf{Q} \) associated with the maximal positive eigenvalue \( \rho_1 \), then define \( \mathbf{W} = \text{diag}(w_j) \). It is easy to show that the matrix \( \mathbf{S} = \rho_1^{-1} \mathbf{W}^{-1} \mathbf{QW} \) is a stochastic matrix whose unique stationary vector is \( (w_j, v_j) \). Correspondingly, let \( \mathbf{V} = \text{diag}(v_j) \), where \( (v_j) \) is the positive left eigenvector associated with \( \rho_1 \), and let \( \mathbf{Q}' \) be the transpose of \( \mathbf{Q} \). Then, as the author observes in (1.4), \( \mathbf{S}^* = \rho_1^{-1} \mathbf{V}^{-1} \mathbf{Q} \mathbf{V} \) is also a stochastic matrix with unique stationary vector \( (w_j, v_j) \). Hence, associated with the matrix \( \mathbf{Q} \) there are two Markov chains each having the stationary vector \( (w_j, v_j) \); one is a “forward” chain defined by \( \mathbf{S} \), and the other is a “reverse” chain defined by \( \mathbf{S}^* \).

I have been at a loss for an interpretation of the forward chain defined by \( \mathbf{S} \) and am now grateful to Professor Kendall for supplying one in terms of observing the process at present conditional on starting a long time ago and finishing a long time in the future.

Mr D. J. Daley: What I have to say is mostly heuristic, though some of it can be made rigorous in special cases. It relates to the idea of diffusion rather than quasi-stationarity, which means that my comments have links with the examples rather than the general theory of Mr Seneta’s paper. Perhaps a closer link is to be found via Professor Bartlett’s paper (Bartlett, 1957) already referred to by Professor Kendall, for some motivation for these remarks may be found there.
When we consider the behaviour of random walks in the plane which occur as the embedded jump skeletons of some stochastic models for population processes (cf. Bartlett (1960)), we commonly make an approximation by forming differential equations with solution curves \( \lambda(x, y) = \text{const} \), where the constant is a function of the initial conditions and equal to \( \lambda_0 = \lambda(x_0, y_0) \), say. The diffusion idea which I wish to mention is due originally to Professor Kendall (Daley and Kendall, 1965), and can be briefly explained as follows. As the process evolves from its initial point \((x_0, y_0)\) it will more or less follow its deterministic path \( \lambda(x, y) = \lambda_0 \), but there will be diffusion around the path. Therefore, if the total effect of individual fluctuations can be determined, then (provided that the deterministic path is in fact the mean sample path) both the mean and variance of the state of the process can be estimated, approximately, by the deterministic path and the total mean square of the fluctuations, respectively. In particular, if the ultimate behaviour of the process is determined by absorption in (for example) one of the axes, then if the deterministic path intersects the axis its intersection is roughly the mean point of absorption, and the variation in \( \lambda(x, y) \) accumulated as the process follows the path \( \lambda(x, y) = \lambda_0 \) can be made to provide an estimate of the variance of the point of absorption. These assertions are stronger than what appears simpler to prove, namely, that under appropriate conditions the deterministic path (suitably normalized) is the limit to which the sample paths of a sequence of similar processes converge in probability.

A feature of Mr Seneta’s genetic examples is that if \( N \) denotes the time to extinction and it is supposed that \( N \) has finite variance, then \( E(N) \) and \( \text{var} (N) \) are of the same order. This is certainly not true of some of the processes I have in mind, for example, the general epidemic process. The link with diffusion in the genetic context is that \( E(\lambda(x, y)) \) deviates slowly from its initial value \( \lambda_0 \); it would be nice to obtain a diffusion-type equation reflecting this behaviour, though the boundary conditions would appear somewhat artificial in such a representation. Where there does appear to be more hope of getting such equations with natural boundary conditions (and where also the notion of quasi-stationarity is obviously relevant) is in simple stochastic models for prey–predator processes: here the deterministic paths \( \lambda(x, y) = \text{const} \) are tangential to the regions of the axes where ultimate absorption may be anticipated.

There followed a brief exchange:

Dr I. J. Good: In applications of this work, the branching process must be nearly critical in order to justify the assumption that it will survive for ever. This being so, I should like to ask whether the introduction of the \( w \)'s into the model can make much practical difference. This is intended as a genuine question and not a rhetorical one.

Mr E. M. L. Beale: That can be answered only by considering some examples.

Professor P. Whittle: In the case of the branching process, the \( w \) will increase linearly whenever \( m \) is less than 1. Now \( m \) can be near 1, in which case one expects that the process will continue for a long period; or \( m \) can be substantially less than 1, in which case one expects only a short run. In either case, the \( w \) show the same behaviour.

Dr I. J. Good: If a tree is known not to be extinct and to be likely to survive for a long time, then it will be a large tree. For small trees the introduction of the factors \( w \) would make a big difference, but perhaps not for large trees.

Professor D. G. Kendall: Could we look at a specific example? Consider the super-critical branching process for which \( f(0) > 0 \) (where \( f(s) \) is the basic probability-generating function whose iterates describe the behaviour of the process). Then if \( q \) is the chance of extinction, we know that (for a single initial parent)

\[
\Pr(Z_n \to 0) = q, \quad \Pr(Z_n \to \infty) = 1 - q,
\]
and $0 < q < 1$. Here the criticality-parameter $m$ can have any value in the range $1 < m \leq \infty$. The transient set of states is now $E = \{1, 2, 3, \ldots\}$; for simplicity let us take this to be inter-communicating. Then the theory I have sketched tells us that
\[
\lim_{T, S \to \infty} \Pr(Z_t = j|Z_{-T} = j_1, Z_{+S} = j_2) =fq^{j_1-1}x_j \quad (j \in E)
\]
whatever $j_1$ and $j_2$ (in $E$) may be. Here the $x$'s are uniquely determined by a set of linear equations. In the simplest case, when $f(\alpha)$ has the linear-fractional form $(\alpha + \beta s)/(\gamma + \delta s)$, and only then, the $x$'s do not depend on $f$ at all: $x_j = (1-q)^j$. It is tempting to compare this quasi-stationary distribution with the classical situation of a conditional distribution where the condition is (as it were) that a certain continuous random variable is to have a stated value. There are indications (unpublished work by Lamperti and by Snell on the related boundary theory) that the comparison may be a very fair one, but we do not understand the connection fully yet.

Now all this holds whatever the value of the criticality-parameter $m$ (provided that it is greater than unity). I am not quite clear what are the $\nu$'s here, and what are the $w$'s, but if Dr Good can say then he should be able to check his intuitions about what happens when $m$ is nearly 1, and what happens when $m$ is much greater than 1. I hope he will do this. The question he has raised seems to me extremely interesting.

Dr I. J. Good: When $m > 1$, which is the case discussed by Professor Kendall, let the probability that there are precisely $j$ individuals in the $r$th generation, conditional on the past but not on the future, be $p_j$. Now the likelihood of $j$, given that the tree never becomes extinct, is $1-q^j$, since this is the probability that the tree never becomes extinct, given $j$. So the conditional probability of $j$ is proportional to $p_j(1-q^j)$. Since $q < 1$, then, if $j$ is large enough, the conditional probabilities are not much affected by the presence of the factor $1-q^j$. Moreover, the probability is small that $j$ is positive but less than any assigned value. So I think a part of the answer to my original question is that the $w$'s have little effect when $m>1$.

When $m<1$, then ultimate extinction is almost certain, and so at first sight it seems impractical to make use of probabilities conditional on immortality. But in statistical mechanics there is a kind of precedent for this, since the Second Law of Thermodynamics is often regarded as rigorously true, even though entropy must decrease as well as increase in the very long run. This might be not too bad an analogy when $m<1$, but only just less. I think this is the case where the speaker's model makes the most sense from a practical point of view. The theory is interesting in any case.

Mr E. Seneta: In reply to Professor Kendall, I think he has very succinctly summarized the general formulation of the theory of quasi-stationary distributions in the denumerable case.

The theory of Vere-Jones is clearly the extension of the classical theorems of the Perron–Frobenius theory of non-negative matrices which is the fundamental tool of my joint paper with J. N. Darroch. In a sense Vere-Jones's theory is "most powerful" for the case when the irreducible aperiodic matrix $A$ is $R$-positive, which of course always holds true when $A$ is finite, and hence is particularly appropriate to the present context under the condition of $R$-positivity.

However, one may obtain further results, analogous to the finite case, if one imposes further restrictions on the matrix $A$, dictated by further comparison with what is true in this case, and also practical considerations with regard to the underlying physical situation. Thus, one may be essentially interested in the problem of absorption from the set $C$ into a set of finite states. In this case the matrix $A$ will be strictly sub-stochastic, representing one step transitions between the non-absorbing states. Secondly, in the finite case, it is
feasible to consider conditional probabilities of the kind (1.1) and (1.3), i.e.

\[
\frac{p_{ij}^{(p)}}{\sum_{k \in U} p_{ik}^{(p)}}
\]

(1)

\[
p_{ij}^{(m)} \left( \sum_{k \in U} p_{ik}^{(n-m)} \right) \sum_{k \in U} p_{ik}^{(n)} (n > m),
\]

(2)

because absorption from C is certain, and hence both numerators and denominators tend to zero. In order to obtain analogous results in the denominator case, therefore, we must either assume that the structure of C is such that absorption is certain, or if it is not, then it is necessary to work on the condition that absorption will eventually occur. This also appears the situation of most interest in biology.

Using this approach, and the results of a new paper of Vere-Jones (1965) on the ergodic properties of non-negative matrices (which incidentally contains the result proved above by Professor Kendall), we have been able to obtain many further results as regards limits of types (1) and (2), and also for an arbitrary initial distribution (Seneta and Vere-Jones, 1965).

Our approach has been complicated by other situations which have no counterpart for the finite case. The one that is immediately obvious is the fact that although the matrix A is sub-stochastic and R-positive, the unique positive left eigenvector \( \{x_i\} \) is not necessarily summable, even when absorption is certain. It is found that the existence of the limit of (1) is very closely related to this requirement, and (1) may tend to the elements of this eigenvector even if the R-positivity restriction on A is removed. On the other hand, the existence of the limit of type (2) is heavily dependent on the requirement of R-positivity, which ensures that \( \sum_{i \in U} x_i y_i < \infty \).

Another notable source of divergence from the finite chain theory is the effect of arbitrary initial distribution. Under the assumptions that A is R-positive,

\[
\sum_{i \in U} x_i < \infty,
\]

and absorption is certain, it is easy to prove that the quantities (1) and (2) tend to limits exactly analogous to the finite case, viz. \( x_i \) and \( x_i y_i \) respectively. However, it is in general necessary to impose further restrictions on an initial distribution in this case, to ensure that the limits are the same. For instance, it is shown for the branching process that a particular choice of initial distribution over C may change the limit of (1).

Using these techniques we have been able to study in some detail the simple branching process for all three cases \( m \geq 1 \) (\( m = f'(1) \)), and for the case \( m > 1 \) have obtained the result pointed out above by Professor Kendall. When \( m < 1 \), the R-positivity of an irreducible aperiodic A is ensured by a condition much weaker than assuming the analyticity at \( s = 1 \) of the offspring distribution \( f(s) \), which is the assumption made for this case by Karlin and McGregor, in order to obtain the spectral expansion for \( p_{ij}^{(p)} \). As a matter of fact, due to the special structure of A for a general branching process, the irreducibility (and aperiodicity) restriction on A is easily removable.

In conclusion to my reply to Professor Kendall, I would like to say that I was most interested in his remarks leading to his equation (7), and subsequent introduction of the chain,

\[
\Pi_{y_2 y_2} = \frac{R}{y_1} p_{y_2 y_2 y_2}.
\]

Since A is R-positive, this chain is positive-recurrent (with stationary distribution \( \{x_i, y_i\} \)) and hence its reverse chain is uniquely defined by

\[
\Pi x_{y_2 y_2} = \frac{R x_{y_2} p_{y_2 y_2}}{x_{y_2}},
\]

(3)
which is precisely the chain mentioned in my paper in equation (1.4). It is moreover not difficult to see by substitution in equation (7) that
\[ \Pr(Z_{tn} = k_n \text{ for } n = 1, 2, \ldots, t | Z_{-t} = i, Z_{+t} = j) \]
under the same limiting two-point boundary condition is
\[ \pi_{i \rightarrow j} \Pi_{i \rightarrow i+1} \cdots \Pi_{i \rightarrow i-t} \]
so that under this condition the chain in a sense behaves as the stationary case of the reverse chain with transition matrix \( \Pi^* \).

I have very little to add to Professor Gani's observations on the overlapping generation model of Moran. In the diffusion approximation this model differs essentially from Wright's only in that the \( \beta_1 \) and \( \beta_2 \) of the latter are replaced by \( \frac{1}{2} \beta_1 \) and \( \frac{1}{2} \beta_2 \) whenever these occur; a reason for this has been given by Moran on p. 133 of his book, as well as that for measuring time in units of \( N^2 \) birth–death events. I feel that, mathematically, the Wright and Moran transition matrices are much closer than would appear at first sight, and that the diffusion approximation to the Wright model, by emphasizing the "local" transitions and destroying the other possibilities to some extent, in a sense reduces its matrix to a random-walk one, very similar to Moran's.

I feel that Mr Morley Jones's uneasiness about the validity of the diffusion approximation near the ends—shared by a number of mathematical geneticists—may be to some extent relieved by the recent papers of Ewens devoted in part to this subject. In the paper of this author quoted in the above references, a fairly complete analysis of corrections to the diffusion approximations which I have discussed is given. In general the conclusions of this paper are that "the genetical relevance of the correction factor is negligible", and "the diffusion approximations, when applicable, supply excellent approximations". In another paper (Ewens, 1964) numerical computations appear to indicate (rather surprisingly) that the errors of certain quantities tend to be smaller near the ends than in the middle!

I find Dr Wishart's comment very useful in that I was not aware of Jensen's paper. Results for the general continuous-time Markov chain with a finite number of states—under the condition that every non-absorbing state can be reached from any other—have been obtained (Seneta, 1964) but have not been published, and are completely analogous to the discrete time results. In this connection, I would also like to mention a paper of Mandl (1960), who has effectively studied the limiting conditional behaviour as \( t \to \infty \) of
\[ \frac{\sum_{i \in \Omega} \pi_i p_{i\ell}(t)}{\sum_{i \in \Omega} \pi_i (1 - p_{i\ell}(t))} \]
in the general case, viz. without the restriction that all transient states communicate. Moreover, Jensen's trick seems to be related to the proof of theorem 1 of Mandl's paper.

Of interest also in this connection may be a simple unpublished result of mine on the simple continuous-time random walk on the half-line with absorbing barrier at the origin. For this case the infinitesimal matrix corresponding to the transient states \( T \) is
\[
C = \begin{bmatrix}
-(a+b) & b & 0 & \cdots \\
a & -(a+b) & b & \cdots \\
0 & a & -(a+b) & b & \cdots \\
& \ddots & \ddots & \ddots & \ddots \\
& & \ddots & \ddots & \ddots & \ddots \\
& & & \ddots & \ddots & \ddots \\
& \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\end{bmatrix}
\]
If $a > b > 0$, then
\[
\lim_{t \to \infty} \frac{p_{ij}(t)}{1-p_{ij}(t)} = (1 - \beta)^a j \beta^{j-1}
\]
for $i, j = 1, 2, \ldots$, where $\beta = (b/a)^a$. However,
\[
\lim_{s \to \infty} \lim_{t \to \infty} \frac{p_{ij}(s)}{1-p_{ij}(t)} \cdot \frac{1-p_{ij}^{(s-k)}}{1-p_{ij}^{(t-k)}} = 0
\]
(in contrast to finite state-space results) for $i, j = 1, 2, \ldots$.

In reply to Dr Miller, I would like to point out, however, that of the two chains in question, with $\{w_t, v_t\}$ as unique stationary distribution, one is merely the reverse chain of the other, and it is easy to check that a reverse chain of a regular chain has precisely the same unique stationary distribution. I find the backward chain $S^*$ more intuitively pleasing since it results from an inversion of $Q$ by a distribution with a definite intuitive meaning, whereas it is rather more difficult to find the intuitive meaning for the forward chain $S$. However, I agree with Dr Miller that Professor Kendall's remarks resolve this problem admirably.

I agree with Mr Daley that Bartlett's book on stochastic processes for population models provides an extremely fertile source of ideas concerning quasi-stationarity, and was in fact the initial point of my research into this subject with Dr Darroch. Whenever I return to Sections 3.4 and 4.3 of this work their content seems to present a new view of the idea. However, I am not sufficiently familiar with the later parts of this book nor the joint paper with Professor Kendall cited to reply at any length to Mr Daley's comments. The idea of using diffusion approximations to consider epidemic processes has also been put forward by Professor Moran, but I am not aware of developments in this direction.

I feel that I may be able to shed a little more light on the discussion between Dr Good, Mr Beale, Professor Whittle and Professor Kendall on the effect of the $w$'s into the branching process model, based on my work with Dr Vere-Jones (Seneta and Vere-Jones, 1965).

Whether absorption is certain or not the analogue of equation (1.2) of the present paper is clearly
\[
\lim_{m \to \infty} \lim_{n \to \infty} P[\text{process in state } j \text{ at time } m < n/\text{absorption has not occurred by time } n \text{ and is certain to occur eventually}]
\]
\[
= \frac{\sum_{k \in T} p_{ij}^{(n)} q_{k}}{\sum_{k \in T} p_{ij}^{(n)} q_{k}},
\]
where $q_j$ is the probability of ultimate absorption starting from state $j$. For an $R$-positive matrix corresponding to $T$ (the transient set) when $n \to \infty$, $m \to \infty$ this can be shown to approach the limit
\[
w_j v_j
\]
independently of $i$, where $v_i$ is the unique left eigenvector and $w_j$ the unique right eigenvector corresponding to the eigenvalue $1/R$. When $m > 1$ for the branching process, the matrix is $R$-positive, $q_j = q^j$ and
\[
w_j = \text{const } jq^{j-1},
\]
so that the limit is the same as that stated above by Professor Kendall. Hence the linear increase of the $w_j$'s themselves when $m < 1$ (when $q = 1$) is destroyed and, rather, there is a rapid decrease.

However, for the branching process when $m > 1$, the analogue of (1.1) is
\[
\lim_{n \to \infty} \frac{p_{ij}^{(n)} q^j}{\sum_{k \in T} p_{ik}^{(n)} q^k} = v_j q^j,
\]
where \( v_i \) is normalized so that
\[
\sum_{i=1}^{\infty} v_i q^i = 1.
\]
Denoting by \( \{e_j\} \) the limiting conditional distribution \( \{v_j q^j\} \) we have that \( w_i v_j = \text{const} \cdot j e_j \)
so that the modifying effect of the \( w \)'s is still a linear one on the \( e_j \)'s. In this sense, the effect of the \( w \)'s is therefore precisely the same as when \( m < 1 \).

I agree completely with Dr Good, however, on the assertion that in the branching process when \( m < 1 \) the quasi-stationary distributions “make sense” only when \( m \) is very close to unity. In some work on bounds extending that of Heathcote and Seneta (1966), I have been able to show that for cases of interest as \( m \to 1 \), the mean time to extinction increases only very slowly, e.g. for the Poisson offspring distribution with \( m = 0.99 \), it is the vicinity of six generations, only, although the variance increases more rapidly.

In conclusion, I would like to express my appreciation to Professor Whittle for being kind enough to read my paper in my absence.

REFERENCES IN THE DISCUSSION


