Adv. Appl. Prob. **6**, 463–488 (1974)
Printed in Israel

THE SAMPLING THEORY OF SELECTIVELY NEUTRAL ALLELES

G. A. WATTERSON, *Monash University*

Abstract

In a paper of the same title, Ewens has considered the problem of inferring whether the genotypic frequencies observed in a small sample are consistent with a particular population model, in which all types are selectively neutral. The present paper gives the general theory of such sampling schemes, when the sampling is from either a deterministic, or a stochastically varying, population.

The theory is illustrated on a population model which was introduced by Karlin and McGregor. Various results are proved for both the population, and its sample. It is shown that Ewens' sampling theory applies to this population model, and conjectures as to why different population models may have the same sampling theory are made.

SAMPLING OF TYPES; OCCUPANCY PROBLEM; GENETIC POPULATION MODELS; ESTIMATION; GOODNESS OF FIT; SELECTIVE NEUTRALITY HYPOTHESIS

1. Sampling from a given population

Consider a population of individuals, each of which can be one of the genetic types A_1, A_2, $A_3, \cdots$. For definiteness, the individuals may be thought of as monoecious haploids (or haploid gametes, from diploids). We may describe the composition of the population in various ways, the most obvious being in terms of the frequencies, or relative frequencies, of the various possible types. However, in situations in which recurrent mutation or migration allows new types to continually appear in the population, there is no upper limit to the number of possible types, and it may be preferable simply to distinguish individuals of different types, without classifying them more thoroughly. In that case, the population may be described by the (unordered) set of type-frequencies, without stating which frequency belongs to which type. Equivalently, the population may be thought of as being partitioned into

$\beta(1)$ types each having one representative only,

$\beta(2)$ types each having two representatives, and, in general,

$\beta(j)$ types each having j representatives.

The population size, M say, is then given by

Received 16 October 1973.

$$M = 1\beta(1) + 2\beta(2) + 3\beta(3) + \cdots. \tag{1.1}$$

On the other hand, the number, K, of distinct types, each represented by one or more individuals in the population, is

$$K = \beta(1) + \beta(2) + \beta(3) + \cdots, \tag{1.2}$$

with, of course, $K \leqq M$. The two extreme cases are:

(i) if the population consists of M individuals all of the same type, then $\beta(M) = 1$, $\beta(j) = 0$ for $j \neq M$, and $K = 1$; and

(ii) if all individuals are of different types, then

$$\beta(1) = M, \ \beta(j) = 0 \ \text{for} \ j \neq 1, \ \text{and} \ K = M.$$

If the above method of describing a population is useful, then equally, the same method is of use in describing a sample. It may well be that an experimenter does not even know how many, or what types there are in a population. But he may be able to recognise, in his sample of size r, say, that there are k distinct types represented, $\alpha(1)$ types being represented once only, $\alpha(2)$ types appearing twice, and so on. Thus, corresponding to the population description $(\beta(1), \beta(2), \beta(3), \cdots)$, there is the sample description $(\alpha(1), \alpha(2), \alpha(3), \cdots)$, where the sample size is given by

$$r = 1\alpha(1) + 2\alpha(2) + 3\alpha(3) + \cdots \tag{1.3}$$

and the number of different types in the sample is

$$k = \alpha(1) + \alpha(2) + \alpha(3) + \cdots. \tag{1.4}$$

If the sample had been chosen at random, without replacement, then the sample composition would approach the population composition as the sample size approaches the population size; in fact

$$\alpha(j) = \beta(j) \ \text{for all} \ j, \ \text{when} \ r = M. \tag{1.5}$$

If sampling is at random with replacement, (1.5) is not true. We shall discuss large-sample results shortly, but first we quote the distribution, and its probability generating function (p.g.f.), for the sample variates $\boldsymbol{\alpha} = (\alpha(1), \alpha(2), \cdots, \alpha(r))$ given the population composition $\boldsymbol{\beta} = (\beta(1), \beta(2), \cdots, \beta(M))$.

Theorem 1. For random sampling with replacement, the sample variates $\boldsymbol{\alpha}$, in a sample of size r, have joint probability distribution

$$P(\boldsymbol{\alpha} \mid \boldsymbol{\beta}) = r!\left[\prod_{j=1}^{r} (j!^{\alpha(j)}\alpha(j)!)\right]^{-1} [1^{\alpha(1)}, 2^{\alpha(2)}, \cdots, r^{\alpha(r)}] \tag{1.6}$$

(where $\boldsymbol{\alpha}$ satisfies (1.3) and $\boldsymbol{\beta}$ satisfies (1.1)), and the joint p.g.f. is

$$E\left(\prod_{j=1}^{r} s_j^{\alpha(j)} \,\middle|\, \boldsymbol{\beta}\right) \tag{1.7}$$
$$= r! \text{ coefficient of } \phi^r \text{ in } \prod_{i=1}^{M} \left(1 + \sum_{j=1}^{r} s_j(\phi i/M)^j/j!\right)^{\beta(i)}.$$

Remark. The notation in (1.6) involves the augmented monomial symmetric functions (see David and Barton (1962), pp. 319–321); here,

$$[1^{\alpha(1)}, 2^{\alpha(2)}, \cdots, r^{\alpha(r)}] = \prod_{j=1}^{r} (\alpha(j)!)\, \Sigma \left(\prod_{i,k} x_{i,k}{}^{n_{i,k}}\right) \tag{1.8}$$

where $x_{i,k} = i/M$ for $k = 1, 2, \cdots, \beta(i)$, $i = 1, 2, \cdots, M$, and the summation is over all partitions of r, into non-negative integers $n_{i,k}$, such that

$$r = \sum_{i=1}^{M} \sum_{k=1}^{\beta(i)} n_{i,k} = \sum_{j=1}^{r} j\alpha(j), \tag{1.9}$$

i.e., such that $\alpha(j)$ of the $n_{i,k}$ are equal to j, $j = 1, 2, \cdots, r$.

The proofs of our major results will be relegated to appendices: for Theorem 1, see Appendix 1.

Theorem 1 has many consequences. From (1.7) we can obtain the moments of the sample variates $\boldsymbol{\alpha}$, and the p.g.f. and moments of linear combinations of $\alpha(1), \alpha(2), \cdots, \alpha(r)$, in particular of k given by (1.4).

The results quoted in the following corollaries may be obtained from (1.7).

Corollary 1. For the sample distribution in Theorem 1, the first and second order moments are

$$E(\alpha(j) \mid \boldsymbol{\beta}) = \sum_{i=1}^{M} \beta(i) \binom{r}{j} \left(\frac{i}{M}\right)^j \left(1 - \frac{i}{M}\right)^{r-j}, \qquad j = 1, 2, \cdots, r \tag{1.10}$$

and

$$E(\alpha(j_1)\alpha(j_2) \mid \boldsymbol{\beta})$$
$$= \sum_{i=1}^{M} \sum_{l=1}^{M} \beta(i)[\beta(l) - \delta_{i,l}] \binom{r}{j_1, j_2, r-j_1-j_2} \left(\frac{i}{M}\right)^{j_1} \left(\frac{l}{M}\right)^{j_2} \left(1 - \frac{i+l}{M}\right)^{r-j_1-j_2} \tag{1.11}$$
$$+ \delta_{j_1,j_2} E(\alpha(j_1) \mid \boldsymbol{\beta}), \qquad j_1, j_2 = 1, 2, \cdots, r,$$

where $\delta_{i,l}$ is Kronecker's delta. More generally, the nth order moments of the $\alpha(j)$'s are nth order expressions in the $\beta(i)$'s.

Corollary 2. The number, k, of types represented in the sample of Theorem 1 has p.g.f.

$$E(s^k \mid \boldsymbol{\beta}) = \frac{r!}{M^r} \text{ coefficient of } \phi^r \text{ in } \prod_{i=1}^{M} (1 + s(e^{\phi i} - 1))^{\beta(i)} \tag{1.12}$$

with mean

$$E(k \mid \boldsymbol{\beta}) = K - \sum_{i=1}^{M} \beta(i)(1 - i/M)^r, \tag{1.13}$$

and variance

$$\begin{aligned} \operatorname{var}(k \mid \boldsymbol{\beta}) = &\sum_{i=1}^{M} \sum_{l=1}^{M} \beta(i)[\beta(l) - \delta_{i,l}] \left(1 - \frac{i+l}{M}\right)^r \\ &+ \sum_{i=1}^{M} \beta(i)(1 - i/M)^r \left(1 - \sum_{i=1}^{M} \beta(i)(1 - i/M)^r\right), \end{aligned} \tag{1.14}$$

where (1.1), (1.2), (1.3) and (1.4) hold.

Some asymptotic results follow. From Corollary 2 we get Corollary 3.

Corollary 3. The number of types, k, in the sample of Theorem 1, is a consistent (and mean-square convergent) estimator of K, the number of types in the population, as $r \to \infty$.

More general asymptotic results, which correspond to (1.5), but in the case of sampling with replacement, may be motivated intuitively. We have seen in Corollary 3 that for very large samples (perhaps considerably larger than the population itself), all types will eventually appear in the sample. If the sample size increases still further, even types rare in the population will have many representatives in the sample, so we may expect $\alpha(j) \to 0$ as $r \to \infty$ for fixed j, and type frequencies proportional to the sample size, say $j = pr$ will predominate. In fact from Corollary 1 we may deduce Corollary 4.

Corollary 4. For the sample of Theorem 1, consider two proportions $0 < p_1 < p_2 < 1$ such that Mp_1 and Mp_2 are not integers. Then the quantity $\sum_{rp_1 < j < rp_2} \alpha(j)$, the number of types whose sample relative frequencies lie between p_1 and p_2, tends, in probability and in mean square, to the corresponding population quantity $\sum_{Mp_1 < i < Mp_2} \beta(i)$ as $r \to \infty$.

If Mp_1, say, happened to be an integer, the limiting expression in Corollary 4 would require an extra term $\frac{1}{2}\beta(Mp_1)$ and similarly if Mp_2 were an integer.

The above results have implications in the "occupancy problem"; see David and Barton (1962). This problem is usually discussed only in the context of sampling of equally likely types. This is the special case of the above theory when $\beta(i) = 0$ for all $i = 1, 2, \cdots, M$ except for $\beta(M/K) = K$, each of the K types having M/K representatives in the population. The variate k would then be the number of "occupied boxes", and $\alpha(j)$ would be the number of boxes having j "balls" (out of the r in the sample). The moments in Corollary 1 have been found, even in the general case, by Tukey (1949), Section 16.

We do not intend to discuss these aspects further, nor to pursue a deeper study of statistical inference questions for the population parameters $\boldsymbol{\beta}$. Our main interest lies in sampling from randomly varying populations, and so in the next section, we study population models leading to random compositions $\boldsymbol{\beta}$.

2. Models for randomly varying populations

Model 1. Karlin and McGregor (1967) introduced various models to describe the continuing presence of genetic heterogeneity. This is due to mutation, which ntroduces a stream of new types into a population and counteracts the loss of heterogeneity due to random genetic drift.

The first model we shall consider here is one in which mutants arise according to a Poisson process, parameter v. The times between mutations are thus independent, exponentially distributed with mean $1/v$. Once a mutant appears, it is the initial ancestor for a sub-population of identical descendents, all having the same birth rate (λ) and death rate (also equal to λ).

Although such a "linear birth and death process" will eventually become extinct, yet the continuing introduction of further mutant types means that gradually, the total population increases in size unboundedly.

Karlin and McGregor showed that, at time $t \geqq 0$, *the number of mutant types represented by j individuals*, ($\beta_t(j)$ in our notation, $N_t^*(j)$ in their's) *has a Poisson distribution with mean*

$$E(\beta_t(j)) = v\lambda^{-1}\left(\frac{\lambda t}{1+\lambda t}\right)^j/j, \qquad j = 1,2,3,\cdots, \tag{2.1}$$

and that $\beta_t(1)$, $\beta_t(2)$, $\beta_t(3),\cdots$ *are independent.*

(Their Equations (2.9), (2.10) appear to contain a slip, corrected in our (2.1).) For simplicity, we choose the time scale so that $\lambda = 1$; thus (2.1) becomes

$$E(\beta_t(j)) = v\mu^j/j, \qquad j = 1,2,3,\cdots, \tag{2.2}$$

in which $\mu = t/(1+t)$ and the mutation rate v is the only free parameter. Notice that $E(\beta_t(j)) \to v/j$ as $t \to \infty$. We call the independent Poisson variates $\beta_t(1)$, $\beta_t(2),\cdots$ subject to (2.2), *Model* 1.

In this model, the population size, M_t, is the sum of infinitely many, independent, random variables:

$$M_t = \sum_{j=1}^{\infty} j\beta_t(j), \tag{2.3}$$

and so too is the number of different types in the population:

$$K_t = \sum_{j=1}^{\infty} \beta_t(j). \tag{2.4}$$

M_t varies in time as a linear birth and death process, with "immigration", and is a special case of "Process B" in Karlin and McGregor (1958). Kendall (1948) found the distribution of M_t (see (2.5) below).

We can prove the following theorem. (For this section, proofs of the major results are contained in Appendix 2.)

Theorem 2. For population Model 1, at finite time t, M_t has a negative binomial distribution with mean $E(M_t) = \nu\mu/(1-\mu) = \nu t$, variance $\nu t(1+t)$:

$$(2.5) \qquad P(M_t = n) = \binom{\nu+n-1}{n}\mu^n(1-\mu)^\nu, \qquad n = 0,1,2,\cdots.$$

K_t has a Poisson distribution with mean and variance equal to

$$(2.6) \qquad E(K_t) = \nu \log(1+t).$$

Further, for the partial sums,

$$(2.7) \qquad P\left(\sum_{j=1}^{m} j\beta_t(j) = n\right) = \exp\left\{-\nu\sum_{j=1}^{m}\mu^j/j\right\}\mu^n\binom{\nu+n-1}{n} \qquad \text{for } n = 0,1,2,\cdots,m,$$

and

$$(2.8) \qquad P\left(\sum_{j=1}^{m}\beta_t(j) = n\right) = \exp\left\{-\nu\sum_{j=1}^{m}\mu^j/j\right\}\left(\nu\sum_{j=1}^{m}\mu^j/j\right)^n/n!, \qquad \text{for } n = 0,1,2,\cdots.$$

We remark that (2.7) is not the full distribution. The probability that $\sum_1^m j\beta_t(j) = n$, for $n > m$, is a more complicated expression. On the other hand, (2.8) is the full (Poisson) distribution for $\sum_1^m \beta_t(j)$. The theorem is not completely new; certainly the study of the distributions of M_t and K_t was undertaken by Karlin and McGregor (1958), (1967) for even more general models. Related work has been done by Singer (1970), (1971), who has made a detailed study of a quantity $S(t)$, the number of positive $\beta_t(j)$'s at time t. We will not require the quantity $S(t)$ in this paper.

As much of our subsequent work will be conditional on a given population size M, and in particular, conditional on a *large* population size, we now consider such conditional situations. In view of (2.5) and (2.6), we can deduce that both M_t and K_t will diverge (almost surely) to $+\infty$ as $t \to \infty$ (i.e., as $\mu \to 1$). Hence it is reasonable to assume a large population for large t values.

Theorem 3. If $\beta(1), \beta(2), \cdots$ are independent Poisson variates, with $E(\beta(j)) = \nu\mu^j/j$, then:

(i) The joint distribution of $\beta(1), \beta(2), \cdots, \beta(M)$, conditional on a given value of $M = \sum_{j=1}^{\infty} j\beta(j)$, is

(2.9)
$$P(\beta(1) = \beta_1, \beta(2) = \beta_2, \cdots, \beta(M) = \beta_M \mid M) = \prod_{j=1}^{M} [(v/j)^{\beta_j}/\beta_j!] \Big/ \binom{v + M - 1}{M}$$

where $\Sigma_1^M j\beta_j = M$. Of course, $\beta(M + 1) = \beta(M + 2) = \cdots = 0$.

(ii) The joint p.g.f. conditional on M is

(2.10)
$$E\left(\prod_{j=1}^{M} s_j^{\beta(j)} \,\Big|\, M\right) = \text{coefficient of } \phi^M \text{ in } \exp\left\{ v \sum_{j=1}^{M} s_j \phi^j / j \right\} \Big/ \binom{v + M - 1}{M}.$$

Notice that (2.9) and (2.10) are now free of the time parameter $\mu = t/(1 + t)$.

The following corollary provides answers to the problem of finding the distribution of the number of types, K, represented in a population of given size M.

Corollary 1. Under the assumptions of Theorem 3,

(i) the distribution and p.g.f. of $K = \Sigma_{j=1}^M \beta(j)$, given that $M = \Sigma_1^\infty j\beta(j)$, are

(2.11)
$$P(K = n \mid M) = v^n \left| S_M^{(n)} \right| / (v)_M, \qquad n = 1, 2, \cdots, M,$$

and

(2.12)
$$E(s^K \mid M) = \binom{vs + M - 1}{M} \Big/ \binom{v + M - 1}{M} = (vs)_M / (v)_M,$$

where $S_M^{(n)}$ is a Stirling number of the first kind, and

(2.13)
$$(v)_M = v(v + 1)(v + 2) \cdots (v + M - 1) = \Gamma(v + M)/\Gamma(v).$$

(ii) If, as $M \to \infty$, then $v \to 0$ so that $v \log M$ remains fixed, then $K - 1$ has an asymptotic Poisson distribution mean $v \log M$.

If, as $M \to \infty$, v remains fixed, then K has an asymptotic normal distribution with mean and variance equal to $v \log M$.

Information on the moments of the population variates, conditional on a given size, are contained in the following. We denote factorial powers by $x^{[n]} = x(x - 1) \cdots (x - n + 1)$.

Corollary 2. Under the assumptions of Theorem 3,

(i) for non-negative integers $n_1, n_2, \cdots, n_M$ with $n_1 + 2n_2 + \cdots + Mn_M = N$, the joint factorial moments are non-zero for $M \geqq N$:

(2.14)
$$E\left(\prod_{j=1}^{M} (\beta(j))^{[n_j]} \,\Big|\, M\right) = \prod_{j=1}^{M} (v/j)^{n_j} \binom{v + M - N - 1}{M - N} \Big/ \binom{v + M - 1}{M},$$

(ii) the factorial moments of $K = \Sigma_{j=1}^{M} \beta(j)$, given $\Sigma_1^\infty j\beta(j) = M$, are

$$E(K^{[n]} \mid M) = \mu_{[n],M} \quad \text{say,} \tag{2.15}$$

where $\mu_{[n],M}$ is given explicitly by

$$\mu_{[n],M} = \sum_{j=1}^{M} v^j j^{[n]} \left| S_M^{(j)} \right| / (v)_M, \tag{2.16}$$

or, recursively, by $\mu_{[0],M} \equiv 1$ together with

$$\mu_{[n],M} = \sum_{j=0}^{n-1} \binom{n-1}{j} v^{n-j} (\psi^{(n-1-j)}(v+M) - \psi^{(n-1-j)}(v)) \mu_{[j],M}, \tag{2.17}$$

wherein the derivatives of the Digamma function satisfy

$$\psi^{(n-1-j)}(v+M) - \psi^{(n-1-j)}(v) = (-1)^{n-1-j}(n-1-j)! \sum_{i=0}^{M-1} (v+i)^{-n+j}. \tag{2.18}$$

The moment formulas (2.14), and (2.17), look rather unattractive in general. Of course, particular cases are quite simple, especially

$$\begin{aligned} E(K \mid M) = \mu_{[1],M} &= v \sum_{i=0}^{M-1} \frac{1}{v+i} \\ &\sim v \log M \quad \text{as} \quad M \to \infty, \end{aligned} \tag{2.19}$$

$$\begin{aligned} E(K^{[2]} \mid M) = \mu_{[2],M} &= -v^2 \sum_{i=0}^{M-1} \frac{1}{(v+i)^2} + \left(\sum_{i=0}^{M-1} \frac{v}{v+i} \right)^2 \\ &\sim (v \log M)^2 \quad \text{as} \quad M \to \infty, \end{aligned} \tag{2.20}$$

and the asymptotic result assuming $M - N$ is large,

$$E\left(\prod_{j=1}^{M} (\beta(j))^{[n_j]} \,\Big|\, M \right) \sim \prod_{j=1}^{M} (v/j)^{n_j} \left(1 - \frac{N}{M}\right)^{v-1}. \tag{2.21}$$

Not all of the distributions and moment formulas above are new. Indeed, in a different notation, the results (2.11), (2.19) and (2.20) are essentially identical with Ewens' (1972) Equations (23), (11), and (24) respectively. However, that they apply under the conditions of Theorem 3, i.e., for the population of Model 1, is believed to be new. Ewens (1972) would call a special case of (2.14), namely

$$E(\beta(j) \mid M) = (v/j) \binom{v+M-j-1}{M-j} \Big/ \binom{v+M-1}{M} \tag{2.22}$$

$$\sim (v/j)(1-(j/M))^{v-1} \quad \text{as} \quad M \to \infty, \text{for } 0 < j/M < 1, \tag{2.23}$$

the "frequency spectrum" of the population, this being the large-population approximation to the mean number of types whose relative frequency is j/M.

However, (2.23) is not a good approximation when j/M is very close to 1 (or 1 itself) and when $v < 1$.

Many of the above results are familiar in a setting different from that of Model 1 itself. Therefore, we briefly outline two other models of relevance.

Model 2. In the same paper which contained Model 1, Karlin and McGregor (1967) discussed a population of constant size (N in their notation, although we shall use M here). The model, there called Model II, postulated that at each time $t = 1,2,3,\cdots$, one individual died and was immediately replaced by a (possibly mutant) copy of one of the existing individuals. The mutation rate, v, was per individual, per unit time. To make it comparable with the rate v in Model 1, which was per population per unit time (but in continuous time), we shall replace the Model II quantity $Nv/(1-v)$, by v itself. Then, also, $N/(1-v)$ will be replaced by $v + M$. After making these substitutions, we shall call the resulting model, *when its stationary distribution has been reached, Model* 2.

We make the following conjecture.

Conjecture 1. At a particular time, Model 2 is identical in distribution with Model 1, given that both populations have the same finite size.

In support of this conjecture, we first remark that when Model 1 was conditioned on a fixed size, its distribution's time-dependence disappeared (see Theorem 3 above). This is the reason for the stationarity assumption being imposed on Model 2.

Next, Karlin and McGregor (1967) have proved, for Model 2, formulas for $E(\beta(j) \mid M)$ and for some moments of K, given M, (see e.g., their (3.7), (3.8), (4.34)) which are exactly equivalent to our results (2.23), (2.19), (2.20) for Model 1. Most impressively, their (4.35) is equivalent to our (2.17) for the general moments of K, given M (provided an undefined coefficient in their (4.35), c_l, is a Stirling number of the second kind, $\mathscr{S}_k^{(l)}$).

We remark that Karlin and McGregor (1967) suggested and subsequently proved (private communication) the validity of a central limit theorem for K in Model 2. Certainly, for Model 1, we have seen in Corollary 1 of Theorem 3 that K is asymptotically normal if v is fixed and $M \to \infty$. Finally, Trajstman (1974) and the present author have independently proved Conjecture 1, after this paper was first submitted. Thus Theorems 2 and 3 (and corollaries) are applicable for both Model 1 and Model 2.

Model 3. A model similar to Model 2, but one in which generations are kept completely distinct, has been studied by Kimura and Crow (1964), Ewens (1964), (1972), and Karlin and McGregor (1967), (1972). In terms of his notation, Ewens

(1972) considers a population of size $2N$, with mutation rate u per individual, per unit time (i.e., per generation). This was reparametrized by putting $\theta = 4N_e u$, where N_e was the effective diploid population size. To make allowances for the different effective sizes in Models 1 and 2 on the one hand, and this non-overlapping generation model on the other hand, we shall identify our notation with Ewens' as follows: $M = 2N$, $v = \theta$, for our sample size $r = 2n$, and K and k have the same interpretations in each paper. Surprisingly little is known about the population variates *even in the stationary case*, which we will take as our *Model* 3. By the use of diffusion theory, Ewens has established, for Model 3, the asymptotic formula (2.23), but he doubted that K has the distribution (2.11) (cf. his (1972) Equation (27)). More firmly, he states that K has the mean given in (2.19) above (cf. his (14)), but he is less sure that its variance is consistent with our (2.19), (2.20). This aspect has also been studied by simulation, in Guess and Ewens (1972), and in Ewens and Gillespie (1974).

Karlin and McGregor (1972) rigorously established the distribution of a random sample, size r, drawn from the population of Model 3. In the next section, we show that this sampling distribution is, for large populations, the same as for sampling from Model 1. This is the most compelling evidence, at present available, in support of the conjecture which follows.

Conjecture 2. At a particular time, Model 3 is asymptotically equivalent in distribution to Model 1 (and to Model 2), conditional on large populations of equal size.

A proof of Conjecture 2 may well arise from considering the (probably common) stationary distribution of three diffusion models, approximating the three models above. We leave to a further paper an analysis of the links between the three models and the logarithmic series distribution, see Watterson (1974).

3. Sampling from randomly varying populations

We remark, at the outset, that in order to test the validity of a stochastic population model, it is probably preferable to take samples from the population at several different times. This would be true, for instance, if one wanted to estimate mutation and selection rates, because these factors influence the "dynamics" of the population. However, in this paper, we follow Ewens (1972), and restrict ourselves to a single sample at a particular time. Thus we are operating in the realm of "static" statistics, not of time series analysis.

Consider a population of given size M, but whose descriptive variates $\beta(1)$, $\beta(2), \cdots, \beta(M)$ have a joint probability generating function

$$g_M(s_1, s_2, \cdots, s_M) = E\left(\prod_{j=1}^{M} s_j^{\beta(j)} \middle| M\right). \tag{3.1}$$

Suppose that a random sample, of size r, is chosen (with replacement) from the population. Then the validity of the following theorem is immediately clear from Theorem 1.

Theorem 4. The joint p.g.f. for the sample variates, from a population described by (3.1), is given by

$$E\left(\prod_{j=1}^{r} s_j^{\alpha(j)} \middle| M\right) = r! \quad \text{coefficient of } \phi^r \text{ in } g_M(t_1, t_2, \cdots, t_M), \tag{3.2}$$

where

$$t_i = 1 + \sum_{j=1}^{r} s_j(\phi i/M)^j/j!. \tag{3.3}$$

Straight away, we turn to the application of this theorem to the population Model 1 of the previous section. For that model, the population p.g.f. is given by (2.10), conditional on the population size M. We thus obtain Corollary 1.

Corollary 1. The joint p.g.f. for the sample variates, sampling from Model 1 subject to a given population size M, is

$$\begin{aligned} &E\left(\prod_{j=1}^{r} s_j^{\alpha(j)} \middle| M\right) \\ &= r! \text{ coefficient of } \phi^r \xi^M \text{ in } \exp\left\{v \sum_{i=1}^{M} t_i \xi^i/i\right\} \Big/ \binom{v+M-1}{M} \end{aligned} \tag{3.4}$$

where t_i is given by (3.3).

Perhaps the major theorem of the paper is the next one, which establishes that in sampling from Model 1, and assuming the population is large, the sample variates have asymptotically the same distributional form as the population variates, except for the conditioning on a given sample size. The result assumes that v, the mutation rate for the whole population, remains fixed as M increases. This is the usual assumption used in getting diffusion theory approximations.

Theorem 5. Consider sampling from Model 1 conditional on a given population size M. Then the sample variates $\alpha(1), \alpha(2), \cdots, \alpha(r)$ are asymptotically (as $M \to \infty$) distributed as independent Poisson variates, means $E(\alpha(j)) = v/j$, but conditioned by the event $r = \sum_{j=1}^{r} j\alpha(j)$.

The proof is given in Appendix 3. The conclusion of the theorem may be re-stated that, asymptotically as $M \to \infty$,

$$E\left(\prod_{j=1}^{r} s_j^{\alpha(j)} \middle| r, M\right) \to \text{coefficient of } \phi^r \text{ in } \exp\left\{v \sum_{j=1}^{r} s_j \phi^j/j\right\} \Big/ \binom{v+r-1}{r} \tag{3.5}$$

and

$$(3.6) \qquad P(\alpha(1) = \alpha_1, \cdots, \alpha(r) = \alpha_r \,|\, r, M) \to \prod_{j=1}^{r} [(\nu/j)^{\alpha_j}/\alpha_j!\,] \Big/ \binom{\nu + r - 1}{r},$$

because the α's asymptotically satisfy the assumptions of Theorem 3, with M replaced by r. Thus (3.5) is a reinterpretation of (2.10), and (3.6) of (2.9).

For the same reason, we can re-interpret all the results stemming from Theorem 3 (its Corollaries 1 and 2, etc.) as statements about the *sample* variates. Thus we have Corollary 1.

Corollary 1. Under the assumptions of Theorem 5, if the sample contains $k = \sum_{j=1}^{r} \alpha(j)$ different types, then k has: (i) as asymptotic distribution, (2.11), (ii) the p.g.f. of (2.12), (iii) the factorial moments of (2.19), (2.20), (2.15), provided M is replaced by r throughout. Moreover, keeping ν fixed as $r \to \infty$, k is asymptotically normal, with mean and variance $\nu \log r$.

The sample variates $\alpha(j)$ have the factorial moments (2.14), and asymptotically as $r \to \infty$, the factorial moments (2.21), (2.23), with M replaced by r throughout.

In connection with Conjecture 2, we asserted that for large populations, the sampling distributions from Model 1 and Model 3 are equal. This may be rigorously proved without relying on the truth of Conjecture 2, and the assertion is contained in the next theorem. Its proof is given in Appendix 3, and relies heavily on the work of Karlin and McGregor (1972).

Theorem 6. The conclusions of Theorem 5, its corollary, and the limits (3.5) and (3.6), are valid for sampling from a Model 3 population.

To round out this series of theorems, a re-statement of Conjecture 1 leads to Conjecture 3.

Conjecture 3. The conclusions of Theorem 5, its corollary, and the limits (3.5) and (3.6), are valid for sampling from a Model 2 population.

In view of subsequent proofs of Conjecture 1, Conjecture 3 is true also.

4. Statistical inference problems

The entire work of Ewens (1972), on estimation and hypothesis testing, may be applied to Model 1 as well as to Model 3. This, because of Theorem 6. As Conjecture 3 is also true, his methods also apply for Model 2. This indicates that there is some "robustness" in the large population theory.

Conversely, the new interpretation of the sampling distribution, as in Theorem 5, means that we can make some contributions to the study of inference problems.

First, we re-state some of Ewens' major points. Throughout, we shall be assuming the asymptotic ($M \to \infty$) form for the sampling distribution.

From (3.6) the sample likelihood is

$$P(\alpha(1) = \alpha_1, \alpha(2) = \alpha_2, \cdots, \alpha(r) = \alpha_r \,|\, r) = \prod_{j=1}^{r} [(\nu/j)^{\alpha_j}/(\alpha_j!)] \Big/ \binom{\nu + r - 1}{r}$$

which factorizes, according to the sample version of (2.11), into

$$\{\nu^k |S_r^{(k)}| / (\nu)_r\} \left\{ r! \Big/ \left\{ \prod_{j=1}^{r} (j^{\alpha_j}\alpha_j!) \right\} |S_r^{(k)}| \right\} \tag{4.1}$$
$$= P(k\,|\,r) \left\{ r! \Big/ \prod_{j=1}^{r} (j^{\alpha_j}\alpha_j!) \,|S_r^{(k)}| \right\}$$

where $k = \alpha_1 + \alpha_2 + \cdots + \alpha_r$. The first factor involves k and ν, while the second factor must be the conditional probability

$$P(\alpha(1) = \alpha_1, \alpha(2) = \alpha_2, \cdots, \alpha(r) = \alpha_r \,|\, k, r)$$
$$= r! \Big/ \left\{ \prod_{j=1}^{r} (j^{\alpha_j}\alpha_j!) \right\} |S_r^{(k)}|. \tag{4.2}$$

Notice that (4.2) does not involve ν. Thus, as Ewens had found, k is a sufficient statistic for ν. He also showed that the maximum likelihood estimator for ν is got by solving the equation

$$k = E(k\,|\,r) = \nu \sum_{i=0}^{r-1} \frac{1}{\nu + i},$$

(cf. (2.19) above). No doubt the maximum likelihood estimator is a consistent estimator; more simply, in view of Theorem 5, Corollary 1, k itself is approximately normally distributed in large samples, with mean and variance $\nu \log r$, so that $k/\log r$ is consistent for ν as $r \to \infty$.

An approximately 95% confidence interval for ν would be

$$[(k/\log r) - 1.96(\sqrt{k}/\log r),\ (k/\log r) + 1.96(\sqrt{k}/\log r)].$$

However, this approximation requires r to be large. Ewens' (1972) computer program to calculate the distribution of k may be used, especially if r is not sufficiently large.

Having used the information in k to estimate the mutation parameter ν, Ewens turned to the remaining sample data for a test of the adequacy of Model 3 to fit the data. Of course, in view of Theorem 6, and Conjecture 3, the models 1, 2 and 3 (at least) cannot be distinguished using a single sample at a particular time. It is hoped that a goodness of fit statistic might distinguish the stationary distributions of models having no selective forces (e.g., Models 1, 2 and 3 of this paper) from those in which selective forces tend to produce, eventually, few types each of high frequency, or perhaps a single high frequency type, with many low frequency types completing the population.

The test statistic Ewens (1972) used was

$$B = -\sum_{i=1}^{k} x_i \log x_i,$$

where x_i, $i = 1, 2, \cdots, k$, are the relative frequencies of types present in the sample. As $\alpha(j)$ is the number of types having relative frequency j/r in the sample, B may be written

$$B = -\sum_{j=1}^{r} \alpha(j) \frac{j}{r} \log \frac{j}{r}, \tag{4.3}$$

a linear function of the sample variates $\alpha(j)$. Its distribution, conditional on r and k, is independent of the nuisance parameter v, provided the models 1, 2 or 3 fit the data well. In theory, the distribution of B may be derived from the following theorem, whose proof is given in Appendix 4.

Theorem 7. For the sample from a large, Model 1 or 3 population, the following results hold conditional on the sample, of size r, containing k different types.

(i) The p.g.f. of $\alpha(1), \alpha(2), \cdots, \alpha(r)$ is

$$\begin{aligned} &E\left(\prod_{j=1}^{r} s_j^{\alpha(j)} \,\Big|\, r, k\right) \\ &\quad = r! \text{ coefficient of } \phi^r \text{ in } \left(\sum_{j=1}^{r} s_j \phi^j / j\right)^k / [k! |S_r^{(k)}|], \end{aligned} \tag{4.4}$$

corresponding to the distribution (4.2).

(ii) The joint factorial moments of the sample variates are

$$E\left(\prod_{j=1}^{r} (\alpha(j))^{[n_j]} \,\Big|\, r, k\right) = r^{[N]} \prod_{j=1}^{r} (1/j)^{n_j} \left| S_{r-N}^{(k-n)} \right| / \left| S_r^{(k)} \right|, \tag{4.5}$$

where $n = \sum_{j=1}^{r} n_j$, $N = \sum_{j=1}^{r} j n_j$.

(iii) The moment generating function of B is

$$\begin{aligned} &E(e^{\xi B} \mid r, k) = r! \text{ coefficient of } \phi^r \text{ in} \\ &\quad \left(\sum_{j=1}^{r} \exp\{-\xi(j/r)\log(j/r)\}\phi^j/j\right)^k / [k! \mid S_r^{(k)} \mid]. \end{aligned} \tag{4.6}$$

Unfortunately, even large-sample results ($r \to \infty$) for the distribution of B seem difficult to obtain analytically, and we cannot improve, here, on Ewens' (1972) computer program method for evaluating the mean and variance of B (in terms of Stirling numbers, logarithms, etc.)

There may be some merit in considering other goodness-of-fit test statistics, e.g., the linear combination $\sum_{j=1}^{r} \alpha(j) j^2$, or the non-linear $\sum_{j=1}^{r} (\alpha(j) - E(\alpha(j)))^2 / \operatorname{var}(\alpha(j))$, in which the moments are calculated from (4.5), conditional on r and k. However, it would seem wise to develop a test statistic powerful against alternative models involving selection. We hope to take up this matter elsewhere. Certainly, the model in Theorem 2.1 of Karlin and McGregor (1967), was general enough to allow the inclusion of, say, density-dependent selection in the birth and death processes for mutant sub-populations.

Proofs of the major results, together with a few additional remarks, may be found in the following Appendices.

Appendix 1

Proof of Theorem 1

In the population, there are $\beta(i)$ types each having relative frequency i/M. Denote those types by $A_{i,1}, A_{i,2}, \cdots, A_{i,\beta(i)}$, and their relative frequencies by $x_{i,1} = x_{i,2} = \cdots = x_{i,\beta(i)} = i/M$. Further, suppose that the numbers of these types which appear in the sample (size r) are $n_{i,1}, n_{i,2}, \cdots, n_{i\,\beta(i)}$ respectively. As sampling is with replacement, the joint distribution of the $n_{i,l}$, $l = 1, 2, \cdots, \beta(i)$, $i = 1, 2, \cdots, M$ is the multinomial

$$r!\left[\prod_i \prod_l n_{i,l}!\right]^{-1} \prod_{i=1}^{M} \prod_{l=1}^{\beta(i)} x_{i.l}^{n_{i,l}}.$$

The probability that $\alpha(j)$ of the $n_{i,l}$'s should equal j, $j = 1, 2, \cdots, r$, is then the sum of the probabilities

$$r!\left[\prod_{j=1}^{r} (j!)^{\alpha(j)}\right]^{-1} \prod_{i=1}^{M} \prod_{l=1}^{\beta(i)} x_{i.l}^{n_{i,l}}$$

summed over all partitions of $r = \sum_{i=1}^{M} \sum_{l=1}^{\beta(i)} n_{i,l}$ such that $\alpha(j)$ of the $n_{i,l}$ equal j, that is, over all partitions consistent with $r = \sum_{j=1}^{r} j\alpha(j)$. Using the notation in (1.8), the equation (1.6) is thus proved.

To show that (1.7) is the p.g.f., we have, by definition and (1.6), that

$$E\left(\prod_{j=1}^{r} s_j^{\alpha(j)} \,\middle|\, \boldsymbol{\beta}\right) = \sum_{\boldsymbol{\alpha}} \frac{r!\,(\prod_{j=1}^{r} s_j^{\alpha(j)})}{\prod_{j=1}^{r} (j!^{\alpha(j)} \alpha(j)!)} [1^{\alpha(1)}, 2^{\alpha(2)}, \cdots, r^{\alpha(r)}]$$

where the summation is over all partitions $\boldsymbol{\alpha}$, consistent with $r = \sum_{j=1}^{r} j\alpha(j)$. Now the above expression would be

$$r!\left[\prod_{i=1}^{M}\prod_{l=1}^{\beta(i)}\left(1+\sum_{j=1}^{r}s_jx_{i,l}^j/j!\right)-1\right]$$

if the summation was not so constrained; see the generating function formula on p. 321 of David and Barton (1962). To single out only those terms for which $r = \sum_{j=1}^r j\alpha(j)$ holds, we use the dummy variable ϕ so that, replacing s_j by $s_j\phi^j$ in the above, the required terms will come from terms of the form

$$\prod_{j=1}^{r}(s_j\phi^j)^{\alpha(j)} = \phi^r\prod_{j=1}^{r}s_j^{\alpha(j)}.$$

Thus

$$E\left(\prod_{j=1}^{r}s_j^{\alpha(j)}\,\Big|\,\beta\right)$$
$$= r!\text{ coefficient of }\phi^r\text{ in }\prod_{i=1}^{M}\prod_{l=1}^{\beta(i)}\left(1+\sum_{j=1}^{r}\phi^js_jx_{i,l}^j/j!\right).$$

Recalling that $x_{i,l} = i/M$ for each $l = 1,2,\cdots,\beta(i)$, we see that (1.7) is obtained.

Remark. Using the notation in Abramowitz and Stegun (1965), p. 823, (1.6) may be written

$$P(\boldsymbol{\alpha}\,|\,\boldsymbol{\beta}) = (r:\alpha(1),\ \alpha(2),\cdots,\alpha(r))'\,[1^{\alpha(1)},\ 2^{\alpha(2)},\cdots,r^{\alpha(r)}].$$

Proof of Corollary 1 (*Theorem* 1)

Differentiating (1.7) with respect to s_j, and then putting $s_1 = s_2 = \cdots = s_r = 1$, yields

$E(\alpha(j)\,|\,\boldsymbol{\beta}) = r!$ coefficient of ϕ^r in

$$\sum_{i=1}^{M}\beta(i)(\phi i/M)^j/j!\left\{\prod_{l=1}^{M}\left(1+\sum_j(\phi l/M)^j/j!\right)^{\beta(l)}\right\}\left\{1+\sum_j(\phi i/M)^j/j!\right\}^{-1}.$$

Since we only require terms in ϕ^r or lesser degree, the sum $1+\sum_{j=1}^r(\phi l/M)^j/j!$ and $e^{\phi l/M}$ are equivalent for our purpose. Thus

$$\begin{aligned} E(\alpha(j)\,|\,\beta) &= r!\text{ coefficient of }\phi^r\text{ in}\\ &\qquad\sum_{i=1}^{M}\beta(i)\,(\phi i/M)^j/j!\exp\left\{\sum_{l=1}^{M}\phi l\beta(l)/M\right\}e^{-\phi i/M}\\ &= r!\sum_{i=1}^{M}\beta(i)\,(i/M)^j/j!\text{ coefficient of }\phi^{r-j}\text{ in }e^{\phi(1-i/M)}, \end{aligned}$$

as $\sum_l l\beta(l) = M$.

From this, (1.10) is obtained by using the expansion of the remaining exponential term. Similarly (1.11) may be obtained.

Proof of Corollary 2 (*Theorem* 1)

Put each $s_j = s$ in (1.7). Because $k = \sum_j \alpha(j)$, one obtains

$$E\left(\prod_{j=1}^{r} s^{\alpha(j)} \,\middle|\, \boldsymbol{\beta}\right)$$

$$\equiv E(s^k \mid \boldsymbol{\beta}) = r! \text{ coefficient of } \phi^r \text{ in } \prod_{i=1}^{M} \left(1 + s \sum_{j=1}^{r} (\phi i/M)^j/j!\right)^{\beta(i)}.$$

Again, as only terms in ϕ^r (or lesser degree) are relevant, the expression $\sum_{j=1}^{r} (\phi i/M)^j/j!$ may be replaced by $e^{\phi i/M} - 1$. Hence (1.12) is obtained. Suitable differentiations of (1.12) with respect to s, at $s = 1$, yield (1.13) and (1.14) in a similar way to the proof of Corollary 1 above. The relations $K = \sum_{i=1}^{M} \beta(i)$ and $M = \sum_{i=1}^{M} i\beta(i)$ are required, to complete the proof.

Proof of Corollary 3 (*Theorem* 1)

Because $K - k \geqq 0$, we see that

$$\begin{aligned} E(K-k) &= 0\,P(K-k=0) + 1P(K-k=1) + 2P(K-k=2) + \cdots \\ &\geqq 1\,P(K-k>0). \end{aligned}$$

However, from (1.13),

$$E(K-k) = \sum_{i=1}^{M} \beta(i)(1 - i/M)^r \to 0 \qquad \text{as } r \to \infty,$$

because each term $(1 - i/M)^r \to 0$. Thus, as $r \to \infty$,

$$P(K - k > 0) \to 0, \text{ i.e., } k \to K \text{ in probability.}$$

It can easily be shown that $\operatorname{var}(k \mid \beta) \to 0$ as $r \to \infty$ so $k \to K$ in mean-square.

Proof of Corollary 4 (*Theorem* 1)

Using (1.10) we find that

$$\text{(A1.1)} \quad E\left(\sum_{rp_1<j<rp_2} \alpha(j) \,\middle|\, \boldsymbol{\beta}\right) = \sum_{i=1}^{M} \beta(i) \left\{\sum_{rp_1<j<rp_2} \binom{r}{j} \left(\frac{i}{M}\right)^j \left(1 - \frac{i}{M}\right)^{r-j}\right\}.$$

The expression in braces is the probability that a binomially distributed variate lies within (rp_1, rp_2); in the present case this may be approximated by the probability that a *normal* variate, mean ri/M and standard deviation $(r(i/M)(1-(i/M)))^{\frac{1}{2}}$ lies within (rp_1, rp_2). As r increases, the probability approaches 1 or 0, depending on whether the mean ri/M lies between rp_1 and rp_2, or not, i.e., whether i lies between Mp_1 and Mp_2, or not.

Hence (A1.1) approaches $\sum_{Mp_1<i<Mp_2} \beta(i)$.

Using (1.11), similar arguments show that the variance of $\Sigma_{rp_1<j<rp_2}\alpha(j)$ approaches 0 as $r \to \infty$, so that mean square convergence is true.

Appendix 2

Proof of Theorem 2

For Model 1, the series of non-negative terms $\Sigma_{j=1}^{m} j\beta_t(j)$ either converges, or diverges to $+\infty$, as $m \to \infty$. For finite times t, the variances of the terms $j\beta_t(j)$, namely $j^2(\nu/j)\mu^j$, form a convergent series $\Sigma_{j=1}^{\infty} j\nu\mu^j$, since $\mu < 1$. Applying Kolmogorov's inequality, we see that $M_t = \Sigma_{j=1}^{\infty} j\beta_t(j)$ is almost surely convergent. (Remark: putting $\mu = 1$, corresponding to $t = \infty$, yields that $M_t = \infty$ almost surely.)

Because almost sure convergence implies convergence in distribution, we have that

$$P(M_t = n) = \lim_{m\to\infty} P\left(\sum_{j=1}^{m} j\beta_t(j) = n\right). \tag{A2.1}$$

Now the joint p.g.f. of the independent Poisson variates $\beta_t(1), \beta_t(2), \cdots, \beta_t(m)$ is

$$\begin{aligned} E\left(\prod_{j=1}^{m} s_j^{\beta_t(j)}\right) &= \prod_{j=1}^{m} \exp\left\{\frac{\nu}{j}\mu^j(s_j - 1)\right\} \\ &= \exp\left\{-\nu\sum_{j=1}^{m}\mu^j/j\right\}\exp\left\{\nu\sum_{j=1}^{m} s_j\mu^j/j\right\}. \end{aligned} \tag{A2.2}$$

Putting $s_j = s^j$, we get

$$E(s^{\Sigma_1^m j\beta_t(j)}) = \exp\left\{-\nu\sum_{j=1}^{m}\mu^j/j\right\}\exp\left\{\nu\sum_{j=1}^{m}(s\mu)^j/j\right\}.$$

Provided $m \geqq n$, the probability that $\Sigma_1^m j\beta_t(j) = n$ holds may be obtained as the coefficient of s^n in

$$\exp\left\{-\nu\sum_{1}^{m}\mu^j/j\right\}\exp\left\{\nu\sum_{1}^{\infty}(s\mu)^j/j\right\} = \exp\left\{-\nu\sum_{1}^{m}\mu^j/j\right\}(1 - s\mu)^{-\nu},$$

because the infinite sum $\Sigma_1^{\infty}(s\mu)^j/j = -\log(1 - s\mu)$ is equivalent to the finite sum $\Sigma_1^m(s\mu)^j/j$ for our purpose.

Thus

$$P\left(\sum_{1}^{m} j\beta_t(j) = n\right) = \exp\left\{-\nu\sum_{1}^{m}\mu_j/j\right\}\binom{\nu+n-1}{n}\mu^n$$

because

$$(1 - s\mu)^{-\nu} = \sum_{n=0}^{\infty}\binom{\nu+n-1}{n}(s\mu)^n \text{ by Taylor's Theorem.}$$

This proves (2.7).

Substituting (2.7) into (A2.1) yields (2.5).

The proof of (2.8), and (2.6), is more elementary. Because the $\beta_t(j)$'s are independent Poisson variates, $\sum_{j=1}^{m} \beta_t(j)$, and K_t itself, will have Poisson distributions, with means being the sum of the means of the $\beta_t(j)$'s. In particular,

$$E(K_t) = \sum_{j=1}^{\infty} \nu\mu^j/j = -\nu \log(1-\mu) = \nu \log(1+t).$$

Remark. (2.7) and (2.8) are, of course, also valid if $\mu = 1$ (or even, if $\mu > 1$, formally). It is strange that, when $\mu = 1$ and $\nu = 1$, the variate $\sum_{j=1}^{m} j\beta_t(j)$ allots *equal* probabilities to each value $n = 0, 1, 2, \cdots, m$. But the values $n > m$ are not, of course, equally likely.

Proof of Theorem 3.

The conditional distribution (2.9) follows by dividing the unconditional multivariate Poisson probability

$$P(\beta(1) = \beta_1, \beta(2) = \beta_2, \cdots, \beta(M) = \beta_M)$$
$$= \exp\left\{-\nu \sum_{1}^{M} (\mu^j/j)\right\} \prod_{j=1}^{M} [(\nu\mu^j/j)^{\beta_j}/\beta_j!]$$

by the probability of $\sum_{j=1}^{M} j\beta(j) = M$ obtained from (2.7) with $n = m = M$, and cancelling the μ terms since $\sum_1^M j\beta_j = M$ is being assumed.

By singling out from A2.2 those terms consistent with the conditioning event (by means of the dummy variable ϕ) and dividing by (2.7) with $n = m = M$, we get (2.10).

Remark. In the notation of the cycle indicator of the symmetric group (see Riordan (1958), p. 68, (3a)), (2.10) may be written

$$\text{(A2.3)} \qquad E\left(\prod_{j=1}^{M} s_j^{\beta(j)} \,\middle|\, M\right) = C_M(\nu s_1, \nu s_2, \cdots, \nu s_M)/(\nu)_M.$$

Proof of Corollary 1 (*Theorem* 3).

(i) (2.12) follows from (2.10) by putting $s_j = s$ for each j, and noting that $\exp(\nu \sum_{j=1}^{M} s\phi^j/j)$ is equivalent, for our purpose, with

$$\exp\left(\nu s \sum_{j=1}^{\infty} \phi^j/j\right) = \exp(-\nu s \log(1-\phi)) = (1-\phi)^{-\nu s}.$$

(2.11) comes from (2.12) by observing the expansion

$$\begin{aligned}(\nu s)_M &= (-1)^M(-\nu s)(-\nu s-1)(-\nu s-2)\cdots(-\nu s-M+1)\\ &= (-1)^M(-\nu s)^{[M]}\\ &= (-1)^M \sum_{n=0}^{M} S_M^{(n)}(-\nu s)^n \qquad \text{(A2.4)}\\ &= \sum_{n=0}^{M} \left|S_M^{(n)}\right|(\nu s)^n.\end{aligned}$$

(See Abramowitz and Stegun (1965) p. 824.)

(ii) From (2.12),

$$E(s^K \mid M) = (\nu s)_M/(\nu)_M = \frac{\Gamma(\nu s+M)}{\Gamma(\nu s)}\,\frac{\Gamma(\nu)}{\Gamma(\nu+M)}.$$

But for $M \to \infty$,

$$\frac{\Gamma(\nu s+M)}{\Gamma(\nu+M)} \sim M^{\nu(s-1)} = \exp\{\nu \log M(s-1)\}$$

and if $\nu \to 0$ as $M \to \infty$,

$$\frac{\Gamma(\nu)}{\Gamma(\nu s)} = \frac{\Gamma(\nu+1)/\nu}{\Gamma(\nu s+1)/\nu s} \to s.$$

Thus

$$E(s^K \mid M) \sim \exp\{\nu \log M(s-1)\}s,$$

i.e.,

$$E(s^{K-1} \mid M) \sim \exp\{\nu \log M(s-1)\},$$

the p.g.f. of a Poisson distribution, mean $\nu \log M$. However, if $\nu = O(1)$, write $Z = (K-\sigma^2)/\sigma$ where $\sigma^2 = \nu \log M \to \infty$ as $M \to \infty$. Then

$$E(e^{\theta Z}) = e^{-\theta\sigma}E(e^{\theta K/\sigma})$$

which, by (2.12) and (2.13),

$$\begin{aligned}&= e^{-\theta\sigma}\,\frac{\Gamma(\nu e^{\theta/\sigma}+M)}{\Gamma(\nu+M)}\,\frac{\Gamma(\nu)}{\Gamma(\nu e^{\theta/\sigma})}\\ &\sim e^{-\theta\sigma}M^{\nu(e^{\theta/\sigma}-1)}\cdot 1 \text{ as } M \to \infty(\sigma \to \infty),\\ &= e^{-\theta\sigma}\exp\{\nu \log M(e^{\theta/\sigma}-1)\}\\ &= e^{-\theta\sigma}\exp\{\sigma^2(\theta/\sigma + \tfrac{1}{2}\theta^2/\sigma^2 + \tfrac{1}{6}\theta^3/\sigma^3 + \cdots)\}\\ &= \exp\{\tfrac{1}{2}\theta^2 + O(1/\sigma)\}.\end{aligned}$$

This is asymptotically the m.g.f. of a standard normal variate, so that K is asymptotically normal, with mean and variance equal to $\sigma^2 = v \log M$.

Proof of Corollary 2 (*Theorem* 3)

(i) To prove (2.14), we start from (2.10). Differentiating with respect to s_j, n_j times, $j = 1, 2, \cdots, M$ and then putting $s_j = 1$ for each j, yields

$$\begin{aligned} &E\left(\prod_{j=1}^{M} \beta(j)^{[n_j]} \,\middle|\, M\right) \\ &\quad = \text{coefficient of } \phi^M \text{ in } \prod_{j=1}^{M} (v\phi^j/j)^{n_j} \exp\left\{v \sum_{j=1}^{M} \phi^j/j\right\} \Big/ \binom{v+M-1}{M}. \end{aligned}$$

But $\prod_{j=1}^{M} (\phi^j)^{n_j} = \phi^N$ where $N = \sum_1^M j n_j$, and up to terms of degree M in ϕ, the exponential term agrees with

$$\exp\left\{v \sum_1^{\infty} \phi^j/j\right\} = (1-\phi)^{-v},$$

in which ϕ^{M-N} has a coefficient $\binom{v+M-N+1}{M-N}$. Hence (2.14) is obtained.

The result may also be achieved from (A2.3), using Equation (6) of Riordan (1958), p. 70, to achieve the differentiations.

(ii) The explicit form (2.16) for the moments of K comes by differentiating (2.12), using the expansion (A2.4). The iterative form (2.17) may be derived as follows. Differentiating (2.12) once with respect to s yields

$$\begin{aligned} E(Ks^{K-1} \mid M) &= \frac{d}{ds}(\Gamma(vs+M)/\Gamma(vs))/(v)_M \\ &= v(\psi(vs+M) - \psi(vs))(\Gamma(vs+M)/\Gamma(vs))/(v)_M \end{aligned}$$

where $\psi(x) = \Gamma'(x)/\Gamma(x)$. Using Leibnitz's rule for differentiating this product a further $n-1$ times, and then putting $s = 1$, yields (2.17). The formula (2.18) is a standard identity for derivatives of the Digamma function (see Abramowitz and Stegun (1965), (6.4.6)).

Appendix 3

The proof of Theorem 4 and its Corollary are immediate from Theorem 1 and Equations (3.1) and (2.10). Of greatest importance is the proof of Theorem 5. This plays the same role for Model 1 as does the addendum by Karlin and McGregor (1972) to Ewens' work on Model 3. Whereas their proof required a double induction, the present proof is a direct one.

Proof of Theorem 5

The sample p.g.f. is given in (3.4) as

$$E\left(\prod_{j=1}^{r} s_j^{\alpha(j)} \middle| M\right)$$

$$= r!\text{ coefficient of } \phi^r\, \xi^M \text{ in } \exp\left\{v \sum_{i=1}^{M} t_i \xi^i / i\right\} \bigg/ \binom{v+M-1}{M}$$

where $t_i = 1 + \sum_{j=1}^{r} s_j(\phi i/M)^j/j!$. Thus the exponent is

$$v\sum_{i=1}^{M} t_i \xi^i/i = v\sum_{i=1}^{M} \xi^i/i + v\sum_{j=1}^{r} s_j(\phi/M)^j \left[\sum_{i=1}^{M} i^{j-1}\xi^i\right]/j!$$

which, up to terms involving ξ^M, is equivalent to

$$v\sum_{i=1}^{\infty} \xi^i/i + v\sum_{j=1}^{r} s_j(\phi/M)^j \left[\sum_{i=1}^{\infty} i^{j-1}\xi^i\right]/j!.$$

But $v\sum_{i=1}^{\infty} \xi^i/i = -\log(1-\xi)^v$, and (see Abramowitz and Stegun (1965), p. 825)

$$\sum_{i=1}^{\infty} i^{j-1}\xi^i = \sum_{i=0}^{j-1} \mathscr{S}_{j-1}^{(i)} \xi^i \frac{d^i}{d\xi^i}\left(\frac{1}{1-\xi}\right) - \delta_{1,j}$$

$$= \sum_{i=0}^{j-1} \mathscr{S}_{j-1}^{(i)} \xi^i (1-\xi)^{-i-1} i! - \delta_{1,j}.$$

Hence the sample p.g.f. may now be written

$$E\left(\prod_{j=1}^{r} s_j^{\alpha(j)} \middle| M\right) = r!\text{ coefficient of } \phi^r\xi^M \text{ in}$$

$$(1-\xi)^{-v}\exp\left\{v\sum_{j=1}^{r} s_j(\phi/M(1-\xi))^j \sum_{i=0}^{j-1} \mathscr{S}_{j-1}^{(i)}\xi^i(1-\xi)^{j-i-1} i!/j!\right.$$

$$\left. - vs_1(\phi/M(1-\xi))(1-\xi)\right\} \bigg/ \binom{v+M-1}{M}.$$

Note that we have associated a factor $1/M(1-\xi)$ with each ϕ-term in the exponent; thus the coefficient of ϕ^r will include a factor $[1/M(1-\xi)]^r$, which we now separate out explicitly:

$$E\left(\prod_{j=1}^{r} s_j^{\alpha(j)} \middle| M\right) = r!\, M^{-r}\text{ coefficient of } \phi^r\, \xi^M \text{ in}$$

$$(1-\xi)^{-v-r}\exp\left\{v\sum_{j=1}^{r} s_j\phi^j \sum_{i=1}^{j-1} \mathscr{S}_{j-1}^{(i)}\xi^i(1-\xi)^{j-i-1} i!/j!\right.$$

$$\left. - vs_1\phi(1-\xi)\right\} \bigg/ \binom{v+M-1}{M}.$$

Choosing a term ξ^{M-n} from the factor $(1-\xi)^{-v-r}$ and a complementary term ξ^n from the exponential, yields

$$E\left(\prod_{j=1}^{r} s_j^{\alpha(j)}\,\middle|\, M\right) = r!\,M^{-r}\sum_{n=0}^{M} \text{coefficient of } \phi^r\,\xi^n \text{ in}$$

$$\binom{\nu+r+M-n-1}{M-n}\exp\left\{\nu\sum_{j=1}^{r} s_j\phi^j\sum_{i=0}^{j-1}\mathscr{S}_{j-1}^{(i)}\xi^i(1-\xi)^{j-i-1}i!/j!\right.$$

$$\left.-\,\nu s_1\phi(1-\xi)\right\}\bigg/\binom{\nu+M-1}{M}.$$

All terms in the exponent are of the form $\phi^j\xi^i$ with $i \leqq j$, and as we seek the coefficient of $\phi^r\xi^n$, only terms with $n \leqq r$ are obtainable and relevant. But for $n \leqq r$, we have

$$r!\,M^{-r}\binom{\nu+r+M-n-1}{M-n}\bigg/\binom{\nu+M-1}{M}$$

$$= M^{-r}\,\frac{\Gamma(\nu+r+M-n)}{\Gamma(1+M-n)}\,\frac{\Gamma(1+M)}{\Gamma(\nu+M)}\bigg/\binom{\nu+r-1}{r}$$

$$\sim M^{-r}M^{\nu+r-1}M^{1-\nu}\bigg/\binom{\nu+r-1}{r} \text{ as } M\to\infty,\ (n,r,\nu \text{ fixed})$$

$$= 1\bigg/\binom{\nu+r-1}{r}.$$

Thus, as $M \to \infty$,

$$E\left(\prod_{j=1}^{r} s_j^{\alpha(j)}\,\middle|\, M\right) \to \sum_{n=0}^{\infty} \text{coefficient of } \phi^r\xi^n \text{ in}$$

$$\exp\left\{\nu\sum_{j=1}^{r} s_j\phi^j\sum_{i=0}^{j-1}\mathscr{S}_{j-1}^{(i)}\xi^i(1-\xi)^{j-i-1}i!/j! - \nu s_1\phi(1-\xi)\right\}\bigg/\binom{\nu+r-1}{r},$$

and to add up all the coefficients of ξ^n, we simply put $\xi = 1$ in the function itself. Noting that $\mathscr{S}_{j-1}^{(j-1)} = 1$, we get the required limit as in (3.5). This has the interpretation claimed for it in Theorem 5 because of Theorem 3, with M replaced by r.

Remarks. The above theorem could be re-phrased as a theorem concerning the sampling distribution's behaviour as *time* increases, not conditionally on a given (large) sized population.

In Model 1, as $t \to \infty$, $\mu \to 1$ and M diverges to $+\infty$ in probability (a consequence of its negative binomial distribution (2.5)), so that the conclusion of Theorem 5 would apply, in a sense, to sampling from the "limiting population".

The asymptotic sample distribution agrees exactly with the population distribution, when the sample size r, equals M, the population size. Thus it would seem that the asymptotic sampling distribution should apply well if sampling is *without* replacement.

Proof of Theorem 6

Denote the numbers of individuals of the various types present in the sample by $n_1, n_2, \cdots, n_k$, where k is the (random) number of types. For Model 3, Karlin and McGregor (1972) showed that the unordered set $\{n_1, n_2, \cdots, n_k\}$ had likelihood, for a sample of size $r = n_1 + n_2 + \cdots + n_k$, of

$$\text{(A3.1)} \qquad P(v; n_1, n_2, \cdots, n_k) = \frac{r!}{n_1 n_2 \cdots n_k} \frac{1}{\alpha_1 ! \alpha_2 ! \cdots \alpha_p !} \frac{v^k}{L_r(v)}$$

in which p denotes the number of distinct numerical values amongst the frequencies $n_1, \cdots, n_k$, and $\alpha_1, \alpha_2, \cdots, \alpha_p$ (here) denote the numbers of the n_j's taking those distinct values, respectively. Also

$$L_r(v) \equiv v(v+1) \cdots (v+r-1) = (v)_r = \binom{v+r-1}{r} r!.$$

By describing the sample solely in terms of our variates $\alpha(1), \alpha(2), \cdots, \alpha(r)$, we find that $n_1 n_2 \cdots n_k = 1^{\alpha(1)} 2^{\alpha(2)} \cdots r^{\alpha(r)}$, that $\alpha_1 ! \ \alpha_2 ! \cdots \alpha_p ! = \alpha(1)! \ \alpha(2)! \cdots \alpha(r)!$ and that $v^k = \prod_{j=1}^r (v^{\alpha(j)})$. Then (A3.1) becomes the right-hand side of (3.6), as required.

Remark. The random quantity p above is the object of study by Singer (1970), (1971), for various populations (rather than samples).

Appendix 4

Proof of Theorem 7

(i) The p.g.f. (4.4) may be proved directly from (4.2), or by singling out from (3.5) those terms consistent with $\sum_{j=1}^r \alpha(j) = k$ (i.e., of kth degree in the s_j's), and dividing the result by

$$\text{(A4.1)} \qquad P(k \mid r) = v^k \left| S_r^{(k)} \right| / (v)_r$$

(as given in (4.1), analogous to (2.11) for the population K).

(ii) Differentiating the p.g.f. (4.4) with respect to s_j, n_j times, $j = 1, 2, \cdots, r$, and putting $s_1 = s_2 = \cdots = s_r = 1$, yields

$$\text{(A4.2)} \qquad \begin{aligned} &E\left(\prod_{j=1}^r (\alpha(j))^{[n_j]} \,\middle|\, r,k \right) \\ &\quad = r! \text{ coefficient of } \phi^r \text{ in } k^{[n]} \prod_{j=1}^r (\phi^j / j)^{n_j} \left(\sum_{j=1}^r \phi^j / j \right)^{k-n} / [k! \left| S_r^{(k)} \right|] \end{aligned}$$

where $n = \sum_{j=1}^r n_j$. But if $N = \sum_{j=1}^r j n_j$ then

$$\prod_{j=1}^{r} (\phi^j/j)^{n_j} = \phi^N \prod_{j=1}^{r} (1/j)^{n_j}$$

and also, for our purposes, $(\sum_{j=1}^{n} \phi^j/j)^{k-n}$ and

$$\left(\sum_{j=1}^{\infty} \phi^j/j\right)^{k-n} = [-\log(1-\phi)]^{k-n}$$

are equivalent. By a formula in Section 24.1.3 of Abramowitz and Stegun (1965),

$$[-\log(1-\phi)]^{k-n} = (-1)^{k-n}(k-n)! \sum_{m=k-n}^{\infty} S_m^{(k-n)}(-\phi)^m/m!.$$

Marshalling these facts together, and after some simplifications (A4.2) yields (4.5).

(iii) Since

$$B = -\sum_{j=1}^{r} \alpha(j) \frac{j}{r} \log \frac{j}{r},$$

then

$$e^{\xi B} = \exp\left\{-\xi \sum_{j=1}^{r} \alpha(j) \frac{j}{r} \log \frac{j}{r}\right\} = \prod_{j=1} \left(\exp\left\{-\xi \frac{j}{r} \log \frac{j}{r}\right\}\right)^{\alpha(j)}.$$

Thus $E(e^{\xi B})$ may be calculated from (4.4) by the substitution

$$s_j = \exp\left\{-\xi \frac{j}{r} \log \frac{j}{r}\right\}.$$

The resulting expression is (4.6).

Remarks. In the notation of Abramowitz and Stegun (1965) p. 823, the distribution (A4.2) for the numbers of types in the sample may be written

$$P(\alpha(1) = \alpha_1,\ \alpha(2) = \alpha_2, \cdots, \alpha(r) = \alpha_r \,|\, r, k) = (r; \alpha_1, \alpha_2, \cdots, \alpha_r)^* / \left|S_r^{(k)}\right|. \tag{A4.3}$$

The same reference indicates that (4.4) must be the p.g.f., and that (A4.3) is a proper probability distribution because the $(r; \alpha_1, \cdots, \alpha_r)^*$ sum to $\left|S_r^{(k)}\right|$ over all permissible choices of the α's, consistent with r and k.

Acknowledgment

I am pleased to acknowledge my debt to Warren Ewens, who pioneered the subject and stimulated my interest at a seminar.

References

ABRAMOWITZ, M. AND STEGUN, I. A. (1965) *Handbook of Mathematical Functions*. Dover, New York.

DAVID, F. N. AND BARTON, D. E. (1962) *Combinatorial Chance*. Charles Griffin, London.

EWENS, W. J. (1964) The maintenance of alleles by mutation. *Genetics* **50**, 891–898.

EWENS, W. J. (1972) The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* **3**, 87–112.

EWENS, W. J. AND GILLESPIE, J. H. (1974) Simulation results for the neutral allele model. To appear.

GUESS, H. A. AND EWENS, W. J. (1972) Theoretical and simulation results relating to the neutral allele theory. *Theor. Pop. Biol.* **3**, 434–447.

KARLIN, S. AND MCGREGOR, J. (1958) Linear growth, birth and death processes. *J. Math. Mech.* **7**, 643–662.

KARLIN, S. AND MCGREGOR, J. (1967) The number of mutant forms maintained in a population. *Proc. 5th Berkeley Symp. Math. Statist. Prob.* **IV**, 415–438.

KARLIN, S. AND MCGREGOR, J. (1972) Addendum to a paper of W. Ewens. *Theor. Pop. Biol.* **3**, 113–116.

KENDALL, D. G. (1948) On some modes of population growth leading to R. A. Fisher's logarithmic series distribution. *Biometrika* **35**, 6–15.

KIMURA, M. AND CROW, J. (1964) The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.

RIORDAN, J. (1958) *An Introduction to Combinatorial Analysis*. John Wiley, New York.

SINGER, B. (1970) Some asymptotic results in a model of population growth. I. *Ann. Math. Statist.* **41**, 115–132.

SINGER, B. (1971) Some asymptotic results in a model of population growth. II. *Ann. Math. Statist.* **42**, 1296–1315.

TRAJSTMAN, A. C. (1974) On a conjecture of G. A. Wattersen. *Adv. Appl. Prob.* **6**, 489–493.

TUKEY, J. W. (1949) Moments of random group size distributions. *Ann. Math. Statist.* **20**, 523–539.

WATTERSON, G. A. (1974) Models for the logarithmic species abundance distributions. *Theor. Pop. Biol.* To appear.