7

A genetics model

7.1 Introduction

Consider a population of N independent individuals. At each time $k \in \{0, 1, 2, \dots\}$ each individual can be in one of n states. The total number, N, of individuals in the population remains constant in time. However, the distribution of the N individuals among the n states changes.

We suppose that initially all random variables are defined on a probability space $(\Omega, \mathcal{F}, P)$. For $1 \leq i, j \leq n$, p_{ji} is the probability that an individual in the population will jump from state i at time $k-1$ to state j at time k. That is, we suppose each individual in the population behaves like an independent time-homogeneous Markov chain with transition matrix $P = (p_{ji})$.

Note $\sum_{j=1}^{n} p_{ji} = 1$.

Write $p_j = (p_{1j}, p_{2j}, \dots, p_{nj})'$ for the j-th column of P.

Write $\Pi(N)$ for the set of all partitions of N into n summands; that is, $z \in \Pi(N)$ if $z = (z_1, z_2, \dots, z_n)$, where each z_i is a nonnegative integer and $z_1 + z_2 + \cdots + z_n = N$.

Write $X(k) = (X_1(k), X_2(k), \dots, X_n(k)) \in \Pi(N)$ for the distribution of the population at time k.

It is easily checked that

$$E[X(k) \mid X(k-1)] = PX(k-1). \tag{7.1.1}$$

However, the population is sampled by withdrawing (with replacement), at each time k, M individuals from the population and observing to which state they belong. That is, at each time k a sample

$$Y(k) = (Y_1(k), Y_2(k), \dots, Y_n(k)) \in \Pi(M)$$

is obtained, where $\Pi(M)$ is the set of partitions of M.

Clearly this sequence of samples, $Y(0), Y(1), Y(2), \dots$ enables us to revise our estimates of the state $X(k)$.

7.2 Recursive estimates

For

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n) \in R^n \text{ and}$$

$$s = (s_1, s_2, \dots, s_n) \in \Pi(N),$$

https://doi.org/10.1017/CBO9780511755330.008 Published online by Cambridge University Press

write

$$F(\alpha, s) = \prod_{j=1}^{n} \langle p_j, \alpha \rangle^{s_j},$$

where $\langle , \rangle$ denotes the scalar product in R^n.

For $r = (r_1, r_2, \ldots, r_n) \in \Pi(N)$ write

$$p_{rs} = P(X(k) = r \mid X(k-1) = s).$$

Then p_{rs} is the coefficient of $\alpha_1^{r_1} \alpha_2^{r_2} \ldots \alpha_n^{r_n}$ in $F(\alpha, s)$. That is,

$$p_{rs} = (r_1! r_2! \ldots r_n!) \frac{\partial^N}{\partial \alpha_1^{r_1} \partial \alpha_2^{r_2} \ldots \partial \alpha_n^{r_n}} F(\alpha, s). \tag{7.2.1}$$

For $y = (y_1, y_2, \ldots, y_n) \in \Pi(M)$ write $\begin{pmatrix} & M & \\ y_1 & y_2 & \ldots & y_n \end{pmatrix}$ for the multinomial coefficient $\dfrac{M!}{y_1! y_2! \ldots y_n!}$. This is just the number of ways of selecting y_1 objects from M into state 1, y_2 into state 2 and so on.

Then, under the original probability measure P,

$$P(Y(k) = y \mid X(k) = r) = \begin{pmatrix} & M & \\ y_1 & y_2 & \ldots & y_n \end{pmatrix} \left(\frac{r_1}{N}\right)^{y_1} \left(\frac{r_2}{N}\right)^{y_2} \cdots \left(\frac{r_n}{N}\right)^{y_n}.$$

Write $\mathcal{G}_k$ for the complete σ-field generated by $X(0), X(1), \ldots, X(k)$ and $Y(0), Y(1), Y(2), \ldots, Y(k-1)$.

$\mathcal{Y}_k$ will denote the complete σ-field generated by $Y(0), Y(1), Y(2), \ldots, Y(k)$. We wish to introduce a new probability measure $\overline{P}$ under which the probability of withdrawing an element in any one of the n states is just $1/n$. For this define factors

$$\gamma_k(Y(k)) = \left(\frac{1}{n}\right)^M \left(\frac{X_1(k)}{N}\right)^{-Y_1(k)} \left(\frac{X_2(k)}{N}\right)^{-Y_2(k)} \cdots \left(\frac{X_n(k)}{N}\right)^{-Y_n(k)},$$

and write

$$\Lambda_k = \prod_{\ell=0}^{k} \gamma_k.$$

A new probability measure can be defined by putting $\left.\dfrac{d\overline{P}}{dP}\right|_{\mathcal{G}_k} = \Lambda_k$.

Lemma 7.2.1 *For $y \in \Pi(M)$, $r \in \Pi(N)$,*

$$\overline{P}(Y(k) = y \mid \mathcal{G}_k) = \begin{pmatrix} & M & \\ y_1 & y_2 & \ldots & y_n \end{pmatrix} \left(\frac{1}{n}\right)^M.$$

Proof $\overline{P}(Y(k) = y \mid \mathcal{G}_k) = \overline{E}[I(Y(k) = y) \mid \mathcal{G}_k]$ and by a version of Bayes' Theorem (4.1.1), this is

$$= \frac{E[\Lambda_k I(Y(k) = y) \mid \mathcal{G}_k]}{E[\Lambda_k \mid \mathcal{G}_k]}.$$

https://doi.org/10.1017/CBO9780511755330.008 Published online by Cambridge University Press

Now γ_k is the only factor of Λ_k not $\mathcal{G}_k$-measurable, so this is

$$= \frac{E[\gamma_k I(Y(k) = y) \mid \mathcal{G}_k]}{E[\gamma_k \mid \mathcal{G}_k]}.$$

The denominator $E[\gamma_k \mid \mathcal{G}_k]$ equals

$$\left(\frac{1}{n}\right)^M E\left[\left(\frac{X_1(k)}{N}\right)^{-Y_1(k)} \left(\frac{X_2(k)}{N}\right)^{-Y_2(k)} \cdots \left(\frac{X_n(k)}{N}\right)^{-Y_n(k)} \mid \mathcal{G}_k\right],$$

and the only variables not $\mathcal{G}_k$-measurable are $Y_1(k), \ldots, Y_n(k)$. Consequently, this conditional expectation is

$$= \left(\frac{1}{n}\right)^M \sum_{y \in \Pi(M)} \binom{M}{y_1 \ldots y_n} = 1.$$

The numerator is

$$E[\gamma_k I(Y(k) = y) \mid \mathcal{G}_k] = \binom{M}{y_1\ y_2\ \ldots\ y_n} \left(\frac{1}{n}\right)^M.$$

Consequently,

$$\overline{P}(Y(k) = y \mid \mathcal{G}_k) = \binom{M}{y_1\ y_2\ \ldots\ y_n} \left(\frac{1}{n}\right)^M$$

$$[3pt] = \overline{P}(Y(k) = y).$$

That is, under $\overline{P}$ the n states are i.i.d. with probability $1/n$. ■

Remark 7.2.2 Under $\overline{P}$, $\overline{P}(X(k) = r \mid X(k-1) = s)$ is still p_{rs} given by (7.2.1). However, as we saw in Lemma 7.2.1,

$$\begin{aligned} \overline{P}(Y(k) = y \mid \mathcal{G}_k) &= \overline{P}(Y(k) = y \mid X(k) = r) \\ &= \overline{P}(Y(k) = y) = \binom{M}{y_1\ y_2\ \ldots\ y_n} \left(\frac{1}{n}\right)^M. \end{aligned}$$

□

To return from $\overline{P}$ to P the inverse density must be introduced. That is, with

$$\overline{\gamma}_k = \gamma_k^{-1} = \left(\frac{1}{n}\right)^{-M} \left(\frac{X_1(k)}{N}\right)^{Y_1(k)} \left(\frac{X_2(k)}{N}\right)^{Y_2(k)} \cdots \left(\frac{X_n(k)}{N}\right)^{Y_n(k)},$$

$$\overline{\Lambda}_k = \Lambda_k^{-1} = \prod_{\ell=0}^{k} \overline{\gamma}_\ell,$$

the probability P can be defined by putting $\left.\frac{dP}{d\overline{P}}\right|_{\mathcal{G}_k} = \overline{\Lambda}_k.$

https://doi.org/10.1017/CBO9780511755330.008 Published online by Cambridge University Press

If $\{\phi_k\}$ is a $\{\mathcal{G}_k\}$ adapted process then Bayes' Theorem (4.1.1) implies

$$E[\phi_k \mid \mathcal{Y}_k] = \frac{\overline{E}[\overline{\Lambda}_k \phi_k \mid \mathcal{Y}_k]}{\overline{E}[\overline{\Lambda}_k \mid \mathcal{Y}_k]}.$$

$\overline{E}[\overline{\Lambda}_k \phi_k \mid \mathcal{Y}_k]$ is, therefore, an unnormalized conditional expectation of ϕ_k given $\mathcal{Y}_k$. The denominator $\overline{E}[\overline{\Lambda}_k \mid \mathcal{Y}_k]$ is a normalizing factor.

For $r \in \Pi(N)$ write $q_r(k) = \overline{E}[\overline{\Lambda}_k I(X(k) = r) \mid \mathcal{Y}_k]$. Note that $\sum_{r \in \Pi(N)} I(X(k) = r) = 1$ so that $\sum_{r \in \Pi(N)} q_r(k) = \overline{E}[\overline{\Lambda}_k \mid \mathcal{Y}_k]$.

We then have the following recursion.

Theorem 7.2.3 *If*
$Y(k) = (Y_1(k), Y_2(k), \ldots, Y_n(k)) = (y_1, y_2, \ldots, y_n) \in \Pi(N)$,

$$q_r(k) = n^{-M} \left(\frac{r_1}{N}\right)^{y_1} \left(\frac{r_1}{N}\right)^{y_2} \cdots \left(\frac{r_n}{N}\right)^{y_n} \sum_{s \in \Pi(N)} p_{rs} q_s(k-1).$$

(Note we take $0^0 = 1$.*)*

Proof

$$\begin{aligned}
q_r(k) &= \overline{E}[\overline{\Lambda}_k I(X(k) = r) \mid \mathcal{Y}_k] \\
&= \overline{E}[\overline{\Lambda}_k I(X(k) = r) \mid \mathcal{Y}_{k-1}, Y(k) = (y_1, y_2, \ldots, y_n)] \\
&= \overline{E}[\overline{\Lambda}_{k-1} \overline{\gamma}_k I(X(k) = r) \mid \mathcal{Y}_{k-1}, Y(k) = (y_1, y_2, \ldots, y_n)] \\
&= n^{-M} \left(\frac{r_1}{N}\right)^{y_1} \left(\frac{r_2}{N}\right)^{y_2} \cdots \left(\frac{r_n}{N}\right)^{y_n} \overline{E}[\overline{\Lambda}_{k-1} I(X(k) = r) \\
&\quad \times \left(\sum_{s \in \Pi(N)} I(X(k-1) = s)\right) \mid \mathcal{Y}_{k-1}] \\
&= n^{-M} \left(\frac{r_1}{N}\right)^{y_1} \left(\frac{r_2}{N}\right)^{y_2} \cdots \left(\frac{r_n}{N}\right)^{y_n} \overline{E}[\overline{\Lambda}_{k-1} \\
&\quad \times \left(\sum_{s \in \Pi(N)} (X(k-1) = s)\right) p_{rs} \mid \mathcal{Y}_{k-1}] \\
&= n^{-M} \left(\frac{r_1}{N}\right)^{y_1} \left(\frac{r_2}{N}\right)^{y_2} \cdots \left(\frac{r_n}{N}\right)^{y_n} \sum_{s \in \Pi(N)} p_{rs} q_s(k-1).
\end{aligned}$$

■

Remarks 7.2.4

$$\begin{aligned}
P(X(k) = r \mid \mathcal{Y}_k) &= E[I(X(k) = r) \mid \mathcal{Y}_k] \\
&= \frac{q_r(k)}{\sum_{s \in \Pi(N)} q_s(k)}.
\end{aligned}$$

To obtain the expected value of $X(k)$ given the observations $\mathcal{Y}_k$ we consider the vector of values $r = (r_1, r_2, \ldots, r_n)$ for any $r \in \Pi(N)$. Then

$$E[X(k) \mid \mathcal{Y}_k] = \frac{\sum_{r\in\Pi(N)} q_r(k)\cdot r}{\sum_{s\in\Pi(N)} q_r(k)}.$$

Unfortunately this does not have the simple form of (7.1.1).

Also note that the transition probabilities p_{rs} can be re-estimated using the techniques described in Chapter 2 of [10]. □

7.3 Approximate formulae

Unfortunately the recursion for $q_r(k)$ given by Theorem 7.2.3 is not easily evaluated. One approximation would be to use a smaller value N' for N in the summation. To obtain nontrivial partitions of N' into n summands, N' should be greater than n. Substitution of the observed $Y(0), Y(1), Y(2), \ldots$ then would give a sequence of approximate distributions.

Alternatively, one could replace the martingale "noise" in the dynamics of $X(k)$ by Gaussian noise ([23]). To describe this, first suppose the n states of the individuals in the population are identified with the unit (column) vectors $e_1, \ldots, e_n$, $e_i = (0, \ldots, 1, 0, \ldots, 0)'$ of R^n. Let $X^i(k) \in \{e_1, \ldots, e_n\}$ denote the state of the i-th individual at time k. Then for each i, $1 \le i \le N$, $X^i(k)$ behaves like a Markov chain on $(\Omega, \mathcal{F}, P)$ with transition matrix P. Consequently,

$$X^i(k) = PX^i(k-1) + M^i(k), \tag{7.3.1}$$

where $E[M^i(k) \mid \mathcal{G}_{k-1}] = E[M^i(k) \mid X^i(k-1)] = 0$.

Write $p(0) = (p_1(0), \ldots, p_n(0))' = E[X^i(0)]$. Then from (7.3.1)

$$E[X^i(k)] = p(k) = P^k p(0).$$

For (column) vectors $x, y \in R^n$ write $x \otimes y = xy'$ for their Kronecker, or tensor, product, and $\operatorname{diag} x$ for the matrix with x on the diagonal.

Then, because $X^i(k)$ is one of the unit vectors $e_1, \ldots, e_n$,

$$\begin{aligned} X^i(k) \otimes X^i(k) &= \operatorname{diag} X^i(k) \\ &= P \operatorname{diag} X^i(k-1)P' + M^i(k) \otimes (PX^i(k-1)) \\ &\quad + (PX^i(k-1)) \otimes M^i(k) + M^i(k) \otimes M^i(k) \\ &= \operatorname{diag} PX^i(k-1) + \operatorname{diag} M^i(k). \end{aligned}$$

Taking the expectation, we have

$$\begin{aligned} E[M^i(k) \otimes M^i(k)] &= \operatorname{diag} Pp(k-1) - P \operatorname{diag} p(k-1)P' \\ &= Q(k), \text{ say.} \end{aligned}$$

For $i \neq j$ the processes X^i and X^j are independent.

Define

$$X(k) = \frac{\sum_{i=1}^{N} X^i(k)}{N},$$

$$M(k) = \frac{\sum_{i=1}^{N} M^i(k)}{N}.$$

The (vector) process $X(k)$ describes the actual distribution of the population at time k. Its components sum to unity and

$$X(k) = PX(k-1) + M(k). \tag{7.3.2}$$

Also, by independence, $E[M(k) \otimes M(k)]$ is also equal to the matrix $Q(k)$.

The suggestion made in [23] is to replace the martingale increments $M(k)$ in (7.3.2) by independent (vector) Gaussian random variables $W(k)$ of mean 0 and covariance $Q(k)$. Write $\phi_k(w)$ for the normal density on R^n of mean 0 and covariance $Q(k)$.

That is, suppose the signal process $\overline{X}(k)$, taking values in R^n, has dynamics

$$\overline{X}(k) = P\overline{X}(k-1) + W(k).$$

For $y = (y_1, y_2, \ldots, y_n) \in \Pi(M)$ and $x = (x_1, x_2, \ldots, x_n) \in R^n$, $x \neq 0$, define

$$\rho(x, y) = |x|^{-M}|x_1|^{y_1}|x_2|^{y_2}\ldots|x_n|^{y_n};$$

set $\rho(0, y) = 0$ for $y \in \Pi(M)$.

The observation process still gives rise to $Y(0), Y(1), \ldots, Y(k) \in \Pi(M)$ and for $y \in \Pi(M)$, $x \in R^n$ we suppose

$$P(Y(k) = y \mid X(k) = x) = \binom{M}{y_1\ y_2\ \ldots\ y_n}\rho(x, y).$$

Starting with the probability $\overline{P}$, now define $\overline{\gamma}_k = n^{-M}\rho(\overline{X}(k), Y(k))$, and $\overline{\Lambda}_k = \prod_{\ell=0}^{k}\overline{\gamma}_\ell$.

Again P can be defined in terms of $\overline{P}$ by setting $\left.\frac{dP}{d\overline{P}}\right|_{\mathcal{G}_k} = \overline{\Lambda}_k$.

Suppose $f : R^n \to R$ is any measurable "test" function. Consider

$$E[f(\overline{X}(k)) \mid \mathcal{Y}_k] = \frac{\overline{E}[\overline{\Lambda}_k f(\overline{X}(k)) \mid \mathcal{Y}_k]}{\overline{E}[\overline{\Lambda}_k \mid \mathcal{Y}_k]}.$$

Suppose there is an unnormalized conditional density $q_k(x)$ such that

$$\overline{E}[\overline{\Lambda}_k f(\overline{X}(k)) \mid \mathcal{Y}_k] = \int_{R^n} f(x)q_k(x)dx.$$

The next result gives a recursion for q_k which is the analog of Theorem 7.2.3.

Theorem 7.3.1

$$q_k(z) = n^{-M}\rho(z, y)\int_{R^n}\phi_k(z - Px)q_{k-1}(x)ds.$$

Proof

$$
\begin{aligned}
\overline{E}[\overline{\Lambda}_k f(\overline{X}(k)) \mid \mathcal{Y}_k] &= \int_{R^n} f(z) q_k(z) \mathrm{d}z \\
&= n^{-M} \overline{E}[\overline{\Lambda}_{k-1} \rho(\overline{X}(k), Y(k)) f(\overline{X}(k)) \mid \mathcal{Y}_k] \\
&= n^{-M} \overline{E}[\overline{\Lambda}_{k-1} \rho(P\overline{X}(k-1) + W(k), y) \\
&\qquad \times f(P\overline{X}(k-1) + W(k)) \mid \mathcal{Y}_{k-1}, Y(k) = y] \\
&= n^{-M} \overline{E}[\overline{\Lambda}_{k-1} \rho(P\overline{X}(k-1) + W(k), y) \\
&\qquad \times f(P\overline{X}(k-1) + W(k)) \mid \mathcal{Y}_{k-1}] \\
&= n^{-M} \overline{E}[\,\overline{\Lambda}_{k-1} \rho(P\overline{X}(k-1) + W(k), y) \\
&\qquad \times f(Px + w) \phi_k(w) q_{k-1}(x) \mid \mathcal{Y}_{k-1}] \\
&= n^{-M} \iint \rho(z, y) f(z) \phi_k(z - Px) q_{k-1}(x) \mathrm{d}z \mathrm{d}x.
\end{aligned}
$$

As this identity holds for all such f the result follows. ■

https://doi.org/10.1017/CBO9780511755330.008 Published online by Cambridge University Press