On the joint asymptotic distribution of additive genotype for polygenic characters

G. M. Tallis

Department of Statistics, University of Adelaide, G.P.O. Box 498, Adelaide, S.A. 5001, Australia

1. Introduction

In the Appendix of an important paper “Selection and Genetic Variability”, Bulmer (1971), and again in chap. 8 of his book The Mathematical Theory of Quantitative Genetics, Bulmer (1980), Bulmer discusses the asymptotic distribution of a single quantitative character for two related individuals chosen at random from the population. The character is assumed to be under the control of many loci and the basic conclusion is that the bivariate central limit theorem applies at equilibrium when, effectively, loci behave independently. However, when linkage effects destroy this independence relationship between loci, a major condition of the theorem is violated and the resulting joint distribution is not normal.

This is a very disturbing assertion since the normal distribution, in all its forms, is fundamental to most work in quantitative genetics. Further, it is counter-intuitive.

The difficulty seems to be that the assumption of independence between bivariate pairs, which is a sufficient condition of the central limit theorem, is used as a necessary condition. In the following sections I will use results on m-dependence to show that, for a polygenic character under panmixia, the joint distribution of this character among related individuals can be anticipated to be normal under very general conditions.

2. m-Dependent sequences

Consider the sequence of random vectors

\[ \mathcal{S}(X, Y) = \{(X_1, Y_1), (X_2, Y_2), \ldots\}, \]

then \( \mathcal{S} \) will be called m-dependent if \( (X_i, Y_i) \) is independent of \( (X_{i+m+j}, Y_{i+m+j}) \), \( j = 1, 2, \ldots, \) and when \( i \geq m \) of \( (X_{i-m-j}, Y_{i-m-j}), j = 1, 2, \ldots, i-m-1 \). Notice that this definition covers independent sequences when \( m = 0 \).

Now if \( S_N(X) = \sum_1^N X_i, S_N(Y) = \sum_1^N Y_i \), it is required to investigate the asymptotic distribution of \( \left( N^{-1/2} S_N(X), N^{-1/2} S_N(Y) \right) \) under m-dependence.

Clearly, when \( m = 0 \), the standard Central Limit Theorem applies. Otherwise, let \( N = kn \) and break the sets of random variables \{\( X_1, X_2, \ldots, X_N \)\}, \{\( Y_1, Y_2, \ldots, Y_N \)\} into k groups of size n, the \( S_N(X) \) and \( S_N(Y) \) being the sums
of the $k$ group sums. Of course, the groups are not independent, so $k$ auxiliary groups are formed from the original ones by omitting the first $m$ and the last $m$ variables of each group.

Group sums are now independent and the normed sums, $N^{-1/2}S^*_N(X)$ and $N^{-1/2}S^*_N(Y)$ over the auxiliary groups can be shown to be asymptotically normally distributed by standard argument, and under suitable assumptions concerning moments up to the third order.

By invoking the further assumption that the $X_i$ and $Y_i$ are uniformly bounded it is easily shown that the asymptotic distribution of $(N^{-1/2}S_N(X), N^{-1/2}S_N(Y))$ is the same as that for $(N^{-1/2}S^*_N(X), N^{-1/2}S^*_N(Y))$. The limiting operation is $n \to \infty$, $k \to \infty$, $k/n \to 0$; however the details of this demonstration are omitted. Instead, a particularly elegant and concise result on $m$-dependence will be quoted from Hoeffding and Robbins (1948).

**Theorem.** Let $\mathcal{S}(X, Y)$ be an $m$-dependent sequence as defined above, and put

$$A_i = V[X_{i+m}] + 2 \sum_{j=1}^{m} C[X_{i+m-j}, X_{i+m}]$$

$$B_i = C[X_{i+m}, Y_{i+m}] + \sum_{j=1}^{m} \{C[X_{i+m-j}, Y_{i+m}] + C[X_{i+m}, Y_{i+m-j}]\}$$

$$C_i = V[Y_{i+m}] + 2 \sum_{j=1}^{m} C[Y_{i+m-j}, Y_{i+m}]$$

Then, if

(a) $E[X_i] = E[Y_i] = 0$, $\max\{E|X_i|^3, E|Y_i|^3\} \leq R^3 < \infty$, $\forall i$

(b) $\lim_{p \to \infty} p^{-1} \sum_{i=1}^{p} A_{i+p} = A$, $\lim_{p \to \infty} p^{-1} \sum_{i=1}^{p} B_{i+p} = B$, $\lim_{p \to \infty} p^{-1} \sum_{i=1}^{p} C_{i+p} = C$, $\forall i$

as $N \to \infty$ $(N^{-1/2}S_N(X), N^{-1/2}S_N(Y))$ has a limiting normal distribution with mean vector $(0, 0)$ and covariance matrix

$$\begin{pmatrix} A & B \\ B & C \end{pmatrix}$$

The above results make it clear that a limiting normal distribution for normed sums of $m$-dependent random variables can be anticipated under rather general conditions.

3. Genetic applications

In order to apply the results of the previous section, let $X$ and $Y$ be two possibly related individuals and consider a single character, $\mathcal{K}$, under the control of many genes. Conceptually, chromosome pairs are put end to end so that locus pairs controlling $\mathcal{K}$ can be linearly ordered. The sequence $\mathcal{S}(X, Y)$ of (2) can be interpreted as a listing of gene effects at the various loci for $X$ and $Y$. Thus, specifically, $(X_i, Y_i)$ represents the gene effects at locus $i$ for $X$ and $Y$ respectively.

Since $X$ and $Y$ may be related, $X_i$ and $Y_i$ may not be independent because of a non-zero probability that genes at locus $i$ in $X$ and $Y$ will be common by descent. Further, loci are located on chromosomes and, but for the phenomenon