

PARTITIONING GENETIC VARIANCES AND COVARIANCES UNDER
RECURRENT BACKCROSSING IN HAPLOIDS

June, 1956

Douglas S. Robson
Cornell University

BU-69-M

Introduction

The simultaneous use of several different breeding systems greatly increases the precision and amount of information obtained from quantitative inheritance studies. We have already demonstrated (BU-65-M) that, barring linkage and other complicating factors, the random mating of full sibs derived from an initial cross of two haploids permits the estimation of all types of genetic variance components. As a supplement to this breeding system we shall now examine the experimental procedure of recurrently backcrossing to each of the two parents. Environmental effects are ignored, and we restrict our attention to the computation of means, variances, and covariances among relatives.

The single locus problem

If the two parents, say P_0 and P_1 , differ genetically at a single locus then the F_1 generation consists of the two parental genotypes, say A_0 and A_1 , mixed with equal frequency. Backcrossing the F_1 population to the parent A_0 we obtain the generation B_1^0 consisting of two families, one family resulting from the cross $A_0 \times A_0$ and one from the cross $A_1 \times A_0$. These two families appear with equal frequency in generation B_1^0 , hence the total gene frequency now becomes $\frac{3}{4}A_0 : \frac{1}{4}A_1$. Backcrossing again to A_0 ; i.e., $B_1^0 \times P_0$, gives the generation B_2^0 consisting of two equally frequent family-groups descending from the two, equally frequent, B_1^0 -families. There are still but two genetically distinct families, however, since there were but two different genotypes in generation B_1^0 . The family consisting entirely of A_0 individuals, resulting from crossing a B_1^0 individual of genotype A_0 onto the parent P_0 , appears with frequency $3/4$ in generation B_2^0 ; the family consisting of $\frac{1}{2}A_0$ and $\frac{1}{2}A_1$ individuals, resulting from crossing a B_1^0 individual of genotype A_1 onto the parent P_0 , appears with frequency $1/4$ in generation B_2^0 . The total gene frequency in B_2^0 is therefore $\frac{7}{8}A_0 : \frac{1}{8}A_1$.

In general, the total gene frequency in generation B_{k-1}^0 is $(1 - \frac{1}{2^k})A_0 + \frac{1}{2^k}A_1$ and the frequency of the two genetically distinct families is $(1 - \frac{1}{2^{k-1}}) \{ A_0 \} : \frac{1}{2^{k-1}} \{ \frac{1}{2}A_0 + \frac{1}{2}A_1 \}$. These two families in generation B_{k-1}^0 produce two family groups in any subsequent generation; thus, in generation B_k^0 the family group consisting entirely of A_0 individuals appears with frequency $1 - \frac{1}{2^{k-1}}$ while the family group consisting of the two equally frequent families $[A_0]$ and $[\frac{1}{2}A_0 + \frac{1}{2}A_1]$ appears with probability $\frac{1}{2^{k-1}}$. Likewise, the family structure in generation B_m^0 , $m \geq k$, when related to generations B_{k-1}^0 and B_k^0 becomes

$$B_m^0: (1 - \frac{1}{2^{k-1}}) \{ A_0 \} + \frac{1}{2^{k-1}} \left\{ \frac{1}{2} [A_0] + \frac{1}{2} \left[(1 - \frac{1}{2^{m-k+1}}) A_0 + \frac{1}{2^{m-k+1}} A_1 \right] \right\}$$

Thus, B_m^0 consists of two family groups, designated by $\{ \}$ -brackets, which trace back to the two genetically distinct families of generation B_{k-1}^0 ; within each $\{ \}$ -group the ancestry may be traced back to a $[]$ -family of generation B_k^0 to form a $[]$ -group within a $\{ \}$ -group in generation B_m^0 . An entirely analogous argument gives the structure of generation B_m^1 obtained by backcrossing m times to the parent P_1 as

$$B_m^1: (1 - \tau_{k-1}) \{ A_1 \} + \tau_{k-1} \left\{ \frac{1}{2} [A_1] + \frac{1}{2} \left[(1 - \tau_{m-k})(A_1) + \tau_{m-k} \left(\frac{1}{2}A_1 + \frac{1}{2}A_0 \right) \right] \right\}$$

where $\tau_x = \frac{1}{2^x}$.

Following the approach in BU-65-M we may now introduce the model for genotypic values,

$$A_j = a_0 + (1-2_j)a,$$

and compute the $Cov(k, k-1; n, m)$ of (k, m) - and (k, n) -family means within family groups tracing back to the families of generation B_{k-1}^0 . For example, in the B_i^0 series

$$\begin{aligned} \text{Cov}_0(2,1;3,2) &= \frac{1}{2} \{0\} + \frac{1}{2} \left\{ \frac{1}{2} [a_0+a] [a_0+a] + \frac{1}{2} \left[\frac{3}{4} (a_0+a) + \frac{1}{4} (a_0-a) \right] \left[\frac{1}{2} (a_0+a) + \frac{1}{2} (a_0-a) \right] \right. \\ &\quad \left. - \left[\frac{7}{8} (a_0+a) + \frac{1}{8} (a_0-a) \right] \left[\frac{3}{4} (a_0+a) + \frac{1}{4} (a_0-a) \right] \right\} \\ &= \frac{1}{16} a^2 \end{aligned}$$

More generally, we shall see that for the case of $N = 1$ locus

$$\text{Cov}_0(k,k-1;n,m) = \text{Cov}_1(k,k-1;n,m) = \frac{a^2}{2^{m+n-k+1}} .$$

This result is obtained as a special case of the following results derived for an arbitrary number N of loci.

N unlinked loci

As in BU-65-M the array of genotypic values in generation B_m^0 for the case where the original parents differ genetically at N unlinked loci is obtained as the product

$$\begin{aligned} &\prod_{i=1}^N \left[(1 - \tau_{k-1}) \left\{ [a_{i0}+a_i] \right\} + \tau_{k-1} \left\{ \frac{1}{2} [a_{i0}+a_i] + \frac{1}{2} [a_{0i}+(1-\tau_{m-k})a_i] \right\} \right] \\ &= \sum_{v=0}^N \tau_{k-1}^v (1-\tau_{k-1})^{N-v} \sum_{I_v \text{ in } I_N} \left\{ \tau_v \sum_{\alpha=0}^v \sum_{I_\alpha \text{ in } I_v} \left[\sum_{\beta=0}^\alpha (1-\tau_{m-k})^\beta \sum_{I_\beta \text{ in } I_\alpha} \sum_{\gamma=0}^{v-\alpha} \right. \right. \\ &\quad \left. \left. \sum_{I_\gamma \text{ in } I_v - I_\alpha} \sum_{\delta=0}^{N-v} \sum_{I_\delta \text{ in } I_N - I_v} \varepsilon_{I_\beta + I_\gamma + I_\delta} \right] \right\} \end{aligned}$$

where

$$\varepsilon_{I_\beta + I_\gamma + I_\delta} = \prod_{i \in I_\beta + I_\gamma + I_\delta} a_i \prod_{j \in I_N - I_\beta - I_\gamma - I_\delta} a_{0j} .$$

The corresponding expression for B_m^1 is obtained simply by appending the coefficient $(-1)^{\beta+\gamma+\delta}$ to $g_{I_\beta+I_\gamma+I_\delta}$.

The covariance of []-family means within { }-families in generations B_m^0 and B_n^0 is then computed as

$$\begin{aligned} \text{Cov}_0(k, k-1; m, n) = & \sum_{v=0}^N \tau_{k-1}^v (1-\tau_{k-1})^{N-v} \sum_{I_v \text{ in } I_N} \left\{ \tau_v \sum_{\alpha=0}^v \sum_{I_\alpha \text{ in } I_v} \left[\sum_{\beta=0}^\alpha (1-\tau_{m-k})^\beta \right. \right. \\ & \sum_{I_\beta \text{ in } I_\alpha} \sum_{\gamma=0}^{v-\alpha} \sum_{I_\gamma \text{ in } I_v-I_\alpha} \sum_{\delta=0}^{N-v} \sum_{I_\delta \text{ in } I_N-I_v} \left. \left. g_{I_\beta+I_\gamma+I_\delta} \right] \cdot \left[\sum_{\beta=0}^\alpha (1-\tau_{n-k})^\beta \right. \right. \\ & \left. \left. \sum_{I_\beta \text{ in } I_\alpha} \sum_{\gamma=0}^{v-\alpha} \sum_{I_\gamma \text{ in } I_v-I_\alpha} \sum_{\delta=0}^{N-v} \sum_{I_\delta \text{ in } I_N-I_v} g_{I_\beta+I_\gamma+I_\delta} \right] \right\} \\ & - \sum_{v=0}^N \tau_{k-1}^v (1-\tau_{k-1})^{N-v} \sum_{I_v \text{ in } I_N} \left\{ \sum_{\alpha=0}^v (1-\tau_{m-k+1})^\alpha \sum_{I_\alpha \text{ in } I_v} \sum_{\delta=0}^{N-v} \right. \\ & \left. \sum_{I_\delta \text{ in } I_N-I_v} g_{I_\alpha+I_\delta} \right\} \cdot \left\{ \sum_{\alpha=0}^v (1-\tau_{n-k+1})^\alpha \sum_{I_\alpha \text{ in } I_v} \sum_{\delta=0}^{N-v} \sum_{I_\delta \text{ in } I_N-I_v} g_{I_\alpha+I_\delta} \right\}. \end{aligned}$$

By an argument very similar to that given in BU-65-M, it can be shown that in $\text{Cov}_0(k, k-1; n, m)$ the coefficient of any squared term $g_{I_v}^2$ reduces to

$$(1 - \tau_m - \tau_n + \tau_{m+n-k})^v - (1 - \tau_m - \tau_n + \tau_{m+n-k+1})^v$$

and the coefficient of the cross product $g_{I_\alpha} \cdot g_{I_\beta}$, $I_\alpha \neq I_\beta$, may be written as

$$[(1-\tau_m)^{\alpha-\delta} (1-\tau_n)^{\beta-\delta} + (1-\tau_m)^{\beta-\delta} (1-\tau_n)^{\alpha-\delta}] \cdot [(1-\tau_m - \tau_n + \tau_{m+n-k})^\delta - (1-\tau_m - \tau_n + \tau_{m+n-k+1})^\delta]$$

where δ is the number of elements common to both I_α and I_β ; i.e., δ is the number of elements in the intersection $I_\alpha \cap I_\beta$. In $\text{Cov}_1(k, k-1; n, m)$ the coefficients are the same except for a factor of $(-1)^{\alpha+\beta}$. We see, then, that

only cross products of the form $g_{I_\alpha} \cdot g_{I_\beta}$ with $I_\alpha \cap I_\beta = \emptyset$ vanish from either of these covariances. In general, however, the experimenter will be concerned only with the sum and difference of these two covariances; we note that $\text{Cov}_0(k, k-1; n, m) + \text{Cov}_1(k, k-1; n, m)$ contains no terms for which $\alpha + \beta$ is odd, while $\text{Cov}_0(k, k-1; n, m) - \text{Cov}_1(k, k-1; n, m)$ contains no terms for which $\alpha + \beta$ is even.

The covariance $\text{Cov}_j(k; n, m)$ among (k, n) - and (k, m) -family means is computed as

$$\text{Cov}_j(k; n, m) = \sum_{i=1}^k \text{Cov}_j(i, i-1; n, m)$$

Thus, the coefficient of $g_{I_\alpha}^2$ in $\text{Cov}_j(k; n, m)$ is

$$\begin{aligned} \sum_{i=1}^k \left[(1 - \tau_m - \tau_n + \tau_{m+n-i})^v - (1 - \tau_m - \tau_n + \tau_{m+n-i+1})^v \right] \\ = (1 - \tau_m - \tau_n + \tau_{m+n-k})^v - (1 - \tau_m)^v (1 - \tau_n)^v \end{aligned}$$

and the coefficient of $g_{I_\alpha} \cdot g_{I_\beta}$, $I_\alpha \cap I_\beta = I_\delta \neq \emptyset$, is then

$$\left[(1 - \tau_m)^{\alpha-\delta} (1 - \tau_m)^{\beta-\delta} + (1 - \tau_m)^{\beta-\delta} (1 - \tau_n)^{\alpha-\delta} \right] \cdot \left[(1 - \tau_m - \tau_n + \tau_{m+n-k})^\delta - (1 - \tau_m)^\delta (1 - \tau_n)^\delta \right].$$

with the factor $(-1)^{\alpha+\beta}$ in the case $j = 1$.

The genetic variances $V_j(k, k-1; m)$, $V_j(k; m)$ are obtained by putting $n = m$ in $\text{Cov}_j(k, k-1; m, n)$ and $\text{Cov}_j(k; m, n)$, respectively, $j = 0, 1$. For example, the total genetic variance in generation B_k^j is $V_j(k; k) = \text{Cov}_j(k; k, k)$ and the average variance within the families in B_k^j determined by parents in B_{k-1}^j is simply $V_j(k; k) - V_j(k-1; k)$.

The population mean for generation B_k^j , denoted by \bar{B}_k^j , is given by

$$\begin{aligned} \bar{B}_k^0 &= \sum_{v=0}^N (1 - \tau_k)^v \sum_{I_\alpha \text{ in } I_N} g_{I_\alpha} \\ \bar{B}_k^1 &= \sum_{v=0}^N (-1)^v (1 - \tau_k)^v \sum_{I_\alpha \text{ in } I_N} g_{I_\alpha} \end{aligned}$$

Hence,

$$\frac{1}{2}(\bar{B}_k^0 + \bar{B}_k^1) = \sum_{v=0}^{\lfloor \frac{N}{2} \rfloor} (1-\tau_k)^{2v} \sum_{I_{2v} \text{ in } I_N} \varepsilon_{I_{2v}}$$

and

$$\frac{1}{2}(\bar{B}_k^0 - \bar{B}_k^1) = \sum_{v=1}^{\lfloor \frac{N}{2} \rfloor} (1-\tau_k)^{2v-1} \sum_{I_{2v-1} \text{ in } I_N} \varepsilon_{I_{2v-1}}$$

References

Robson, D. S. Random mating among full sibs, an experimental technique for estimating genetic variance components in a haploid population. BU-65-M, April, 1956.