A FURTHER NOTE ON THE ESTIMATION OF EFFECTIVE NUMBER OF FACTORS BY MEANS OF PROBABILITY-GENERATING FUNCTIONS[1]

TENG-SHAN WENG[2]

(Received Feb. 21, 1967)

In a previous note (Weng, 1966) we arrived at two formulas for estimating effective number of factors or gene-pairs under two given genetic models designated as Model I and Model II, respectively. We now propose another genetic model, namely Model III, thereby to derive a new formula for the same purpose.

As before, we consider a quantitative character or trait of polygenic nature in a certain diploid organism and also consider a crossing between two pure lines L_1 and L_2 of this organism. Suppose that there are n independent contributing factors involved in this cross so that we can designate the genotypes of L_1 and L_2 respectively by $a_1 a_1 a_2 a_2 \ldots a_n a_n$ and $A_1 A_1 A_2 A_2 \ldots A_n A_n$, where the genotype of L_1 is assumed to have a smaller genotypic value as compared with that of L_2. Suppose further that there is no interaction between the contributing factors. We then proceed to estimate the effective number n of factors under the framework of the following model:

Model III: Let the genotypes of L_1 and L_2 be redesignated as $b_1 b_1 b_2 b_2 \ldots b_\kappa b_\kappa c_1 c_1 c_2 c_2 \ldots c_m c_m$ and $B_1 B_1 B_2 B_2 \ldots B_\kappa B_\kappa C_1 C_1 C_2 C_2 \ldots C_m C_m$ ($\kappa+m=n$, $0 \leqslant \kappa \leqslant n$, $0 \leqslant m \leqslant n$), respectively, where we assume that each of the B genes is lacking in dominance relative to its allele b so that the B genes have each an additive and equal effect e, and that each of the C genes is dominant over its allele c so that the gene-pairs of the form CC or Cc have each an additive and equal dominance effect $d=2e$. Hence if we denote the difference of the genotypic values of L_1 and L_2 as D, we have

$$D=2\kappa e+md=2(\kappa+m)e=2ne.$$

We thus see that

$$e=D/(2n) \qquad (1)$$

(1) Paper No. 61 of the Scientific Journal Series, Institute of Botany, Academia Sinica.
(2) Associate Research Fellow, Institute of Botany, Academia Sinica

and

$$d=\mathrm{D}/n. \tag{2}$$

Now our problem is to estimate n and κ, subject to the restriction $n-\kappa=m$. To this effect, let K and M be the random variables representing respectively the number of B genes and the number of gene-pairs of the form CC or Cc in each individual of the F_2 population of the cross $L_1\times L_2$. Then the probability-generating functions of these random variables are given respectively by

$$\mathrm{P_K}(t)=\sum_{i=0}^{2\kappa}\binom{2\kappa}{i}\left(\frac{1}{2}\right)^{2\kappa}t^i=\left[\frac{1}{2}(t+1)\right]^{2\kappa} \tag{3}$$

and

$$\mathrm{P_M}(t)=\sum_{j=0}^{m}\binom{m}{j}\left(\frac{3}{4}\right)^{j}\left(\frac{1}{4}\right)^{m-j}t^j=\left[\frac{1}{4}(3t+1)\right]^{m}, \tag{4}$$

where t is a dummy variable such that $-h<t<h$ for some real number $h>0$.

We now consider the random varialbe $\mathrm{N}=\mathrm{K}+\mathrm{M}$, which represents the total number of B genes and gene-pairs of the form CC or Cc in F_2. Since K and M are stochastically independent, it follows that N has the probability-generating function

$$\mathrm{P_N}(t)=\mathrm{P_K}(t)\cdot\mathrm{P_M}(t)=\left[\frac{1}{2}(t+1)\right]^{2\kappa}\left[\frac{1}{4}(3t+1)\right]^{m}. \tag{5}$$

By virtue of (3), (4) and (5) we find the expectations and the first factorial moments of K, M and N to be

$$E(\mathrm{K})=\frac{d}{dt}\mathrm{P_K}(t)\,|_{t=1}=\kappa,$$

$$E[\mathrm{K}(\mathrm{K}-1)]=\frac{\mathrm{d}^2}{dt^2}\mathrm{P_K}(t)\,|_{t=1}=\frac{1}{2}\kappa(2\kappa-1),$$

$$E(\mathrm{M})=\frac{d}{dt}\mathrm{P_M}(t)\,|_{t=1}=\frac{3}{4}m,$$

$$E[\mathrm{M}(\mathrm{M}-1)]=\frac{d^2}{dt^2}\mathrm{P_M}(t)\,|_{t=1}=\frac{9}{16}m(m-1),$$

$$E(\mathrm{N})=\frac{d}{dt}\mathrm{P_N}(t)\,|_{t=1}=k+\frac{3}{4}m,$$

$$E[\mathrm{N}(\mathrm{N}-1)]=\frac{d^2}{dt^2}\mathrm{P_N}(t)\,|_{t=1}=\frac{1}{2}\kappa(2\kappa-1)+\frac{3}{2}\kappa m+\frac{9}{16}m(m-1).$$

It follows immediately that

$$\mathrm{var}(\mathrm{K})=E[\mathrm{K}(\mathrm{K}-1)]+E(\mathrm{K})-E^2(\mathrm{K})=\frac{1}{2}\kappa. \tag{6}$$

Similary,

$$\mathrm{var}(\mathrm{M})=\frac{3}{16}m, \tag{7}$$

and

$$\begin{aligned}\mathrm{var}(\mathrm{N})&=\frac{1}{2}\kappa+\frac{3}{16}m\\&=\mathrm{var}(\mathrm{K})+\mathrm{var}(\mathrm{M}).\end{aligned} \tag{8}$$

The above result shows that var(N) can be partitioned into two components, namely var(K) and var(M). Accordingly, the genetic variance of F_2, to be denoted as $V_{g_{F_2}}(N)$, can also be partitioned into two components, to be denoted respectively as $V_{g_{F_2}}(K)$ and as $V_{g_{F_2}}(M)$, which are the counterparts of var(K) and var(M), respectively Let us express this relation as follows:

$$V_{g_{F_2}}(N)=V_{g_{F_2}}(K)+V_{g_{F_2}}(M). \tag{9}$$

Now, referring to (1) and (2), we can calculate $V_{g_{F_2}}(K)$ and $V_{g_{F_2}}(M)$ as follows:

$$V_{g_{F_2}}(K)=\mathrm{var}(eK)=\mathrm{var}\left(\frac{D}{2n}K\right)=\frac{D^2}{4n^2}\mathrm{var}(K)=\frac{\kappa D^2}{8n^2}, \tag{10}$$

$$V_{g_{F_2}}(M)=\mathrm{var}(dM)=\mathrm{var}\left(\frac{D}{n}M\right)=\frac{3mD^2}{16n^2}=\frac{3(n-\kappa)D^2}{16n^2}. \tag{11}$$

We note here that $m=n-\kappa$. Substituting (10) and (11) into (9). we obtain

$$V_{g_{F_2}}(N)=\frac{(3n-\kappa)D^2}{16n^2}. \tag{12}$$

Consider now a backcrossing of the type $F_1\times L_1$, *i.e.*, $B_1b_1B_2b_2\ldots B\kappa\, b\kappa\, C_1c_1C_2c_2\ldots C_mc_m\times b_1b_1b_2b_2\ldots b_\kappa b_\kappa c_1c_1c_2c_2\ldots c_mc_m$. Let K′ and M′ be random variables respresenting respectively the number of B genes and the number of gene-pairs of the form *Cc* in each individual of the backcross $F_1\times L_1$. Then the probability-generating functions of K′ and M′ can be obtained as follows:

$$P_{K'}(t)=\sum_{i=0}^{\kappa}\binom{\kappa}{i}\left(\frac{1}{2}\right)^{\kappa}t^i=\left[\frac{1}{2}(t+1)\right]^{\kappa},$$

$$P_{M'}(t)=\sum_{j=0}^{m}\binom{m}{j}\left(\frac{1}{2}\right)^{m}t^j=\left[\frac{1}{2}(t+1)\right]^{m}.$$

Here the random variables K′ and M′ are stochastically independent, hence the probability-generating function of the random variable $N'=K'+M'$ is given by

$$P_{N'}(t)=\left[\frac{1}{2}(t+1)\right]^{\kappa+m}.$$

Making use of these probability-generating functions and following a procedure similar to what we have undergone so far, we find that the genetic variance of $F_1\times L_1$, denoted by $V_{g_{B_1}}(N')$, as follows:

$$V_{g_{B_1}}(N')=\frac{(4n-3\kappa)D^2}{16n^2}. \tag{13}$$

Solving Equations (12) and (13) simultaneously for n and κ we obtain

$$n=\frac{5D^2}{16[3V_{g_{F_2}}(N)-V_{g_{B_1}}(N')]}, \tag{14}$$

$$\kappa=\frac{5[4V_{g_{F_2}}(N)-3V_{g_{B_1}}(N')]D^2}{16[3V_{g_{F_2}}(N)-V_{g_{B_1}}(N')]^2}. \tag{15}$$

In practice the parameter D may be estimated by the difference between the respective sample means of L_1 and L_2. Thus, denoting by $\hat{D}$, $\bar{L}_1$ and $\bar{L}_2$ the sample estimate of D and the sample means of L_1 and L_2, respectively, we have

$$\hat{D}=\bar{L}_2-\bar{L}_1. \tag{16}$$

Further, we can, under the assumption of no interaction between the genotypes and the environmental effects and of no epistasis between the gene-pairs of the genotypes, decompose the phenotypic variances of F_2 and $F_1\times L_1$, denoted respectively by V_{F_2} and V_{B_1}, as follows:

$$V_{F_2}=V_{g_{F_2}}+V_e,$$
$$V_{B_1}=V_{g_{B_1}}+V_e,$$

where V_e denotes the environmental variance. Let V_{F_1} denote the phenotypic variance of the F_1 population (*i.e.*, $L_1\times L_2$). Then we should have $V_{F_1}=V_e$, because all the F_1 individuals have the same genotype. Therefore, we can estimate $V_{g_{F_2}}$ and $V_{g_{B_1}}$ respectively by

$$\hat{V}_{g_{F_2}}=\hat{V}_{F_2}-\hat{V}_{F_1} \tag{17}$$

and

$$\hat{V}_{g_{B_1}}=\hat{V}_{B_1}-\hat{V}_{F_1}, \tag{18}$$

where $\hat{V}_{g_{F_2}}$, $\hat{V}_{g_{B_1}}$, $\hat{V}_{F_2}$, $\hat{V}_{B_1}$ and $\hat{V}_{F_1}$ are the sample estimates of $V_{g_{F_2}}(N)$, $V_{g_{B_1}}(N')$, V_{F_2}, V_{B_1} and V_{F_1}, respectively. Denoting by $\hat{n}$ and $\hat{\kappa}$ the sample estimates of n and κ, respectively, and referring to Equations (14), (15), (16), (17) and (18), we obtain, finally,

$$\hat{n}=\frac{5\hat{D}^2}{16(3\hat{V}_{g_{F_2}}-\hat{V}_{g_{B_1}})}$$
$$=\frac{5(\bar{L}_2-\bar{L}_1)^2}{16(3\hat{V}_{F_2}-\hat{V}_{B_1}-2\hat{V}_{F_1})}, \tag{19}$$
$$\hat{\kappa}=\frac{5(4\hat{V}_{g_{F_2}}-3\hat{V}_{g_{B_1}})\hat{D}^2}{16(3\hat{V}_{g_{F_2}}-\hat{V}_{g_{B_1}})^2}$$
$$=\frac{5(4\hat{V}_{F_2}-3\hat{V}_{B_1}-3\hat{V}_{F_1})(\bar{L}_2-\bar{L}_1)^2}{16(3\hat{V}_{F_2}-\hat{V}_{B_1}-2\hat{V}_{F_1})^2} \tag{20}$$

Incidentally, letting $\hat{m}$ denote the sample estimate of m, we get

$$\hat{m}=\hat{n}-\hat{\kappa}. \tag{21}$$

These are the results we wanted.

介紹利用機率母函數以估算有效因子總數的另一個方法

翁 登 山

設 L_1, L_2 爲兩個品系，其因子型值之差 D 係受 n 對微量因子所控制。再設 L_1, L_2 之因子型各爲 $b_1 b_1 b_2 b_2 \ldots b_k b_k c_1 c_1 c_2 c_2 \ldots c_m c_m$ 和 $B_1 B_1 B_2 B_2 \ldots B_k B_k C_1 C_1 C_2 C_2 \ldots C_m C_m$ $(k+m=n,\ 0 \leqslant k \leqslant n,\ 0 \leqslant m \leqslant n)$，其中 B 因子對於 b 因子無顯性，但每一 B 因子均具有相等的加法效應 e, C 因子對於 c 因子爲顯性，而且每一對 CC 或 Cc 均具有相等的顯性加法效應 $d=2e$，基於上述的遺傳模型，我們推知

$$e=\frac{D}{2n}, \qquad d=\frac{D}{n}.$$

如是可利用機率母函數及 L_1, L_2 雜交後代之統計資料求算 n, k, m 估計值如下：

$$\hat{n}=\frac{5(\bar{L}_2-\bar{L}_1)^2}{16(3\hat{V}_{F_2}-\hat{V}_{B_1}-2\hat{V}_{F_1})},$$

$$\hat{k}=\frac{5(4\hat{V}_{F_2}-3\hat{V}_{B1}-3\hat{V}_{F_1})(\bar{L}_2-\bar{L}_1)^2}{16(3\hat{V}_{F_2}-\hat{V}_{B_1}-2\hat{V}_{F_1})^2},$$

$$\hat{m}=\hat{n}-\hat{k}.$$

$\bar{L}_1=L_1$ 之樣本平均；$\bar{L}_2=L_2$ 之樣本平均；$\hat{V}_{F_1}=L_1\times L_2$ 的 F_1 之外表型樣本變方 $\hat{V}_{B_1}=F_1\times L_1$ [即 $(L_1\times L_2)\times L_1$] 之外表型樣本變方，$\hat{V}_{F_2}=L_1\times L_2$ 的 F_2 之外表型樣本變方。

Literature Cited

WENG, TENG SHAN. A note on the estimation of effective number of factors by means of probability-generating functions. Botanical Bulletin of Academia Sinica, 7: 101-104, 1966.