

# A NOTE ON THE ESTIMATION OF EFFECTIVE NUMBER OF FACTORS BY MEANS OF PROBABILITY- GENERATING FUNCTIONS

TENG-SHAN WENG<sup>(1)</sup>

(Received December 2, 1965)

By a factor we shall mean any gene-pair or locus which is responsible for a particular character of a certain organism. Now consider a metric character of polygenic nature in a certain organism and consider a crossing between two pure line  $L_1$  and  $L_2$  of the organism. Suppose that there are  $n$  independent contributing factors in the cross, so that we can designate the genotypes of  $L_1$  and  $L_2$  respectively by  $a_1a_1a_2a_2\dots a_na_n$  and  $A_1A_1A_2A_2\dots A_nA_n$ ,  $L_1$  being considered as having a smaller genotypic value in comparison with  $L_2$ . Let  $D$  denote the difference between the genotypic values of  $L_1$  and  $L_2$ . Assuming that there is no interaction between the contributing factors, we then proceed to the estimation of  $n$  in accordance with the following two models:

**Model I:** Assume that there is no dominance and that each of the  $A$  genes has an additive and equal effect  $e$ ; *i. e.*,

$$(1) \quad e = \frac{D}{2n}.$$

Then, for some  $r$  such that  $0 \leq r \leq 2n$ , the relative frequency of an  $F_2$  individual having  $r$   $A$  genes and  $2n - r$   $a$  genes in its genotype will be

$$(2) \quad \binom{2n}{r} \left(\frac{1}{2}\right)^{2n}.$$

The probability-generating function for the distribution of  $A$  genes is then given as follows:

$$(3) \quad P_N(t) = \left(\frac{1}{2}\right)^{2n} \sum_{r=0}^{2n} \binom{2n}{r} t^r = \left[\frac{1}{2}(t+1)\right]^{2n}$$

where  $N$  is the random variable representing the number of  $A$  genes, and  $t$  is a dummy variable. The average number of  $A$  genes in the  $F_2$  population is

$$(4) \quad E(N) = \frac{d}{dt} P_N(t) \Big|_{t=1} = n.$$

---

<sup>(1)</sup> Paper No. 46 of the scientific Journal Series. Associate Research Fellow, Institute of Botany, Academia Sinica.

The first factorial moment of  $N$  is

$$\begin{aligned}
 (5) \quad E(N(N-1)) &= \left. \frac{d^2}{dt^2} P_N(t) \right|_{t=1} \\
 &= \left(\frac{1}{2}\right)^{2n} \sum_{r=2}^{2n} r(r-1) \binom{2n}{r} \\
 &= \frac{n(2n-1)}{2},
 \end{aligned}$$

whence it follows that

$$\begin{aligned}
 (6) \quad \text{var}(N) &= E(N(N-1)) + E(N) - E^2(N) \\
 &= \frac{n}{2}.
 \end{aligned}$$

Hence the genetic variance, denoted by  $V_g$ , of  $F_2$  is given by

$$\begin{aligned}
 (7) \quad V_g &= \text{var}(eN) = \text{var}\left(\frac{D}{2n} \cdot N\right) \\
 &= \frac{D^2}{8n}.
 \end{aligned}$$

Thus  $n$  can be expressed in terms of  $D$  and  $V_g$ , as follows:

$$(8) \quad n = \frac{D^2}{8V_g}.$$

**Model II:** Assume that there is dominance and that each gene-pair  $AA$  or  $Aa$  has an additive and equal dominance effect

$$(9) \quad d = 2e = \frac{D}{n}.$$

Then, for some  $k$  satisfying  $0 \leq k \leq n$ , the relative frequency of an  $F_2$  individual having  $k$  gene-pairs of the form  $AA$  or  $Aa$  and  $n-k$  gene-pairs of the form  $aa$  in its genotype is given by

$$(10) \quad \binom{n}{k} \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{n-k},$$

and the probability-generating function for the distribution of the gene-pair  $AA$  or  $Aa$  will be

$$P_M(t) = \sum_{k=0}^n \binom{n}{k} \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{n-k} t^k = \left[\frac{1}{4}(3t+1)\right]^n$$

where  $M$  is the random variable representing the number of gene-pairs of the form  $AA$  or  $Aa$ . The average number of such gene-pairs in the  $F_2$  population is

$$(11) \quad E(M) = \left. \frac{d}{dt} P_M(t) \right|_{t=1} \\ = \frac{3n}{4}.$$

The first factorial moment of  $M$  is

$$(12) \quad E(M(M-1)) = \left. \frac{d^2}{dt^2} P_M(t) \right|_{t=1} \\ = \frac{9n(n-1)}{16}.$$

It then follows that

$$(13) \quad \text{var}(M) = E(M(M-1)) + E(M) - E^2(M) \\ = \frac{3n}{16}.$$

Hence we obtain

$$(14) \quad V_g = \text{var}(dM) = \text{var}\left(\frac{D}{n} \cdot M\right) \\ = \frac{3D^2}{16n},$$

and so we can solve Equation (14) for  $n$  as follows:

$$(15) \quad n = \frac{3D^2}{16V_g}.$$

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

$$(16) \quad \hat{D} = \bar{L}_2 - \bar{L}_1.$$

Furthermore, 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 variance of  $F_2$  population, denoted by  $V_{F_2}$  as follows:

$$(17) \quad V_{F_2} = V_g + V_e,$$

where

$V_g$  = Genetic variance,

and

$V_e$  = Environmental variance.

Since all of the  $F_1$  individuals have the same genotype, it is reasonable to express  $V_g$  by the variance of the  $F_1$  population, denoted by  $V_{F_1}$ . Thus we can estimate  $V_g$  by

$$(18) \quad \hat{V}_g = \hat{V}_{F_2} - \hat{V}_{F_1},$$

where  $\hat{V}_g$ ,  $\hat{V}_{F_2}$ , and  $\hat{V}_{F_1}$  are the estimates of  $V_g$ ,  $V_{F_2}$ , and  $V_{F_1}$  respectively. Finally, the substitution of Equations (16) and (18) into Equations (8) and (15) leads respectively to

$$(19) \quad \hat{n} = \frac{(\bar{L}_2 - \bar{L}_1)^2}{8(\hat{V}_{F_2} - \hat{V}_{F_1})},$$

and

$$(20) \quad \hat{n} = \frac{3(\bar{L}_2 - \bar{L}_1)^2}{16(\hat{V}_{F_2} - \hat{V}_{F_1})},$$

where  $\hat{n}$  denotes the estimate of  $n$ .

## 介紹利用機率母函數以估算有效因子 總數的一個方法

翁 登 山

如果一個育種計劃的目標是在於改進某種作物、家畜、或家禽的定量形質 (metric character)，則立計劃之前似乎應該考慮到究竟有多少有效的因子控制着該形質，如是才能預計育種的年限和安排育種的步驟。通常有效因子的總數越大，遺傳的效率越低，育種的年限也須要增長，同時在步驟方面也得根據集團育種的原理加以調整。本文於此介紹一個估算有效因子總數的方法，主要的構想是把  $F_2$  集團中每一因子型所含有的有效因子之數看作一個機率變數，藉以界定該變數的機率母函數並藉以探討各種可能的因子型在  $F_2$  集團中的分佈狀態，而利用  $F_1$  及  $F_2$  集團的觀察資料以估算控制該定量形質的有效因子之總數。