

# 基因\*(有效因子)模型及其辨识

陈 剑 郑维敏 王永县

(清华大学经济管理学院系统工程博士点, 北京)

数量性状受多基因控制且其表现型受环境影响较大。为了更加有效地分析数量性状的表现, 本文构建了数量性状的基因模型。由于该模型既包含了基因的加性作用, 也涉及到显性偏差和上位性偏差, 从而能够比较全面地刻划数量性状群体的分布形态。文章还给出了辨识模型参数的方法及一些性质的讨论。通过对一系列实例的计算并与Castle-Wright公式的结果进行了比较。本文所构建的模型产生的误差明显比Castle-Wright公式的小。可见该模型确实能够提供更多有用的信息。

**关键词** 基因模型; 群体分布; 参数辨识

作物的多数性状属数量性状, 受多基因控制且环境作用不可忽视。正是这些因素造成了育种过程的极端复杂性。为了更加有效地分析数量性状的表现, 提高育种工作的效率。建立数量性状群体分布的数学模型是有益的。利用模型有助于我们进行杂交组合配置、后代选择等问题的处理。从而进一步增强育种工作的预见性。从1921年<sup>[2]</sup>开始, 这方面已经开展了许多工作, 其中以Castle-Wright公式<sup>[3]</sup>最为出色。它是在基因仅有累加作用等一系列假定基础上得出的。用公式估计出控制性状的基因数目。然后, 运用估计出的基因数目分析该数量性状的表现, 以期能对育种工作有所指导。但是, 对数量性状的研究表明, 基因除了简单的累加作用外, 还存在显性和上位性作用。它们分别表达了等位基因间和非等位基因间的相互作用。在群体上就表现出各种不同的分布形态。这就是说, 以往的方法在刻划性状表现型分布上是片面的。从而极大地限制了其结果的应用。

基于上述分析, 本文构建了数量性状的基因模型, 它既考虑到基因型的加性作

用, 也涉及到显性偏差和上位性偏差, 用不同的参数表示, 将基因的不同作用与不同的分布形状联系起来。通过辨识的方法求出控制性状的基因数目。其结果明显优于Castle-Wright公式。同时, 我们还讨论了辨识方法的一些性质, 并探讨了亲本为一类非极端类型的情况。

## 数量性状的表达

性状表现出的数值称为该性状的表型值。它是基因型与环境共同作用的结果。所以, 表型值  $x$  按其发生的原因可以分解成两部分: 基因型值  $G$  和环境偏差  $E$ 。一般可以将  $x$  描述成  $G$  和  $E$  的线性函数<sup>[4]</sup>:

$$x = G + E \quad (1)$$

进一步涉及到基因型的效应问题, 关于  $G$  可再作分解。

### (一) 加性效应值(育种值)

本文于1989年9月18日收到。

\* 这里所指实为有效因子<sup>[1]</sup>, 后面若无说明意义均同。

本文得到南农马育华教授、北农大张树榛教授和北京师院郭平仲教授的指导和鼓励, 并得到中国自然科学基金资助, 特此致谢。

假定基用作用可加, 这样在一个基因座内每个基因型中的两个基因效应之和为该基因型的加性效应值。同型合子  $AA$ ,  $aa$  的加性效应值分别为  $2qa$  和  $-2pa$ , 杂型合子  $Aa$  的加性效应值为  $(q-p)a$ 。其中,  $a$  为基因  $A$  的代换平均效应。该基因座内  $A$  和  $a$  的基因频率分别为  $p$  和  $q$ 。

**(二) 显性偏差**

显性偏差因同一基因座内等位基因间相互作用而产生, 表示为:

$$G = H + D \quad (2)$$

$H$  是加性效应值, 而  $D$  是显性偏差, 它依赖于显性程度和群体内基因频率, 当显性不存在时, 加性效应值与基因型值一致。

**(三) 上位性偏差**

它由不同基因座间非等位基因间的互作产生的。设  $G_A$  是一个基因座的基因型值,  $G_B$  是第二个基因座的基因型值,  $G$  是这两个基因座集合的基因型值。那么:

$$G = G_A + G_B + I_{AB} \quad (3)$$

式中:  $I_{AB}$  称为上位性偏差

由上述分解, 基因型值  $G$  可表示成:

$$G = H + D + I \quad (4)$$

从而表型值的表达式为:

$$x = H + D + I + E \quad (5)$$

其中:  $H$ 、 $D$ 、 $I$  是离散随机变量,  $H$  为加性效应值,  $D$  为显性偏差,  $I$  为上位性偏差。  $E$  为环境偏差, 通常设它服从  $N(0, V^2)$  且  $E$  与  $H$ 、 $D$ 、 $I$  相互独立。

**基因模型 ( $F_2$  代群体分布模型)**

由数学表达式 (5), 我们将推导数量性状  $F_2$  代群体分布的数学模型。首先作如下假定:

1°  $EFF$  为每一个别基因效应, 且各个基因作用大小相同<sup>[5]</sup>;

2° 基因间没有连锁<sup>[5]</sup>;

3° 两个亲本是纯合的, 且为两极端类型<sup>[5]</sup>;

4° 为了充分反映式 (5), 又便于数学表达及应用于预测。我们规定某一数量性状的基因对数由三部分组成 (设它共有  $N$  对基因, 分解为  $N_1, N_2, N_3$ , 满足  $0 \leq N_1, N_2, N_3 \leq N$  且  $N_1 + N_2 + N_3 = N$ )。第一部分的基因对数为  $N_1$ , 这  $N_1$  对基因没有显性及上位性作用, 它反映了式 (5) 中  $H$  的一部分  $H_1$ 。第二部分的基因对数为  $N_2$ , 设这部分中不同基因座间无上位性偏差, 而每一基因座内的基因有显性关系。它们体现了式 (5) 中  $H$  的又一部分  $H_2$ , 还贡献了式 (5) 中的  $D$ 。最后一部分的基因对数为  $N_3$ , 它们每两对基因间存在互作, 从而提供了式 (5) 中  $H$  的剩余部分  $H_3$  ( $H_1 + H_2 + H_3 = H$ ) 及上位性偏差  $I$ 。这三部分基因集合构成  $N$  对基因在整体上就产生了式 (5) 的结果。  $N_2, N_3$  数值的大小反映了显性偏差和上位性偏差的程度。

**(一) 基因型值分布函数**

1. 第一部分: 由假定这部分  $N_1$  对基因没有显性和上位性效应, 先设  $N_1 = 1$ ,  $F_2$  代分布列为:

基因型	$aa$	$Aa$	$AA$
状态	$X_0$	$X_1$	$X_2$
状态值	0	$EFF$	$2EFF$
概率	1/4	1/2	1/4

写成概率的形式:

$$pr_j(X_j) = \binom{2}{j} \left(\frac{1}{2}\right)^2 \quad (6)$$

式中:  $j = 0, 1, 2$

推广到一般情况:

$$pr_j(X_j) = \binom{2N}{j} \left(\frac{1}{2}\right)^{2N} \quad (7)$$

式中:  $j = 0, 1, 2, \dots, 2N$

2. 第二部分：这部分  $N_2$  对基因进一步可分成两部分( $KX, KF, 0 \leq KX, KF \leq N_2$ 且  $KX + KF = N_2$ )，其中一部分为正显性 ( $KX$ )，另一部分为负显性 ( $KF$ )，且都为完全显性。

①正显性(基因A相对a显性)：设  $KX = 1$ ，则  $F_2$  代分布列为：

基因型	aa	(Aa, AA)
状态	$X_0$	$X_1$
状态值	0	EFF
概率	1/4	3/4

表示成概率的形式：

$$pr_2(X_j) = \binom{1}{j} \left(\frac{3}{4}\right)^j \left(\frac{1}{4}\right)^{1-j} \quad (8)$$

式中： $j = 0, 1$

扩展到一般情况：

$$pr_2(X_j) = \binom{KX}{j} \left(\frac{3}{4}\right)^j \left(\frac{1}{4}\right)^{KX-j} \quad (9)$$

式中： $j = 0, 1, \dots, KX$

②负显性：类似①的求法得到：

$$pr_3(X_j) = \binom{KF}{j} \left(\frac{1}{4}\right)^j \left(\frac{3}{4}\right)^{KF-j}$$

式中： $j = 0, 1, \dots, KF$

3. 第三部分：这部分  $N_3$  对基因中每两对基因间存在上位作用。上位作用有几种不同的类型。这里仅考虑两类(隐性上位和显性上位)。这样，我们又将  $N_3$  分解成为  $2KY$  和  $2KS$ ，其中  $0 \leq 2KY, 2KS \leq N_3$  且  $2KY + 2KS = N_3$ 。因考虑两对非等位基因间的上位作用，故在  $KX$  和  $KY$  前加上 2 倍。

①隐性上位：设  $KY = 1$ ， $F_2$  代分布列为：

基因型	(aabb, aaB_)	A_bb	A_B_
状态	$X_0$	$X_1$	$X_2$
状态值	0	EFF	2EFF

概率 1/4 3/16 9/16  
用概率的形式表示为：

$$pr_4(X_j) = \sum_{\substack{j_1=0 \\ 2j_1+j_2=j \\ j_1+j_2 \leq 1}}^1 \sum_{j_2=0}^1 \binom{1}{j_1, j_2} \left(\frac{9}{16}\right)^{j_1} \cdot \left[\left(\frac{3}{16}\right)^{j_2} \left(\frac{1}{4}\right)^{1-j_1-j_2}\right] \quad (11)$$

式中： $j = 0, 1, 2$

$$\binom{n}{j_1, j_2} = \frac{n!}{j_1! j_2! (n-j_1-j_2)!} \quad (12)$$

推广到一般情形：

$$pr_4(X_j) = \sum_{\substack{j_1=0 \\ 2j_1+j_2=j \\ j_1+j_2 \leq KY}}^{KY} \sum_{j_2=0}^{KY} \binom{KY}{j_1, j_2} \left(\frac{9}{16}\right)^{j_1} \cdot \left(\frac{3}{16}\right)^{j_2} \left(\frac{1}{4}\right)^{KY-j_1-j_2} \quad (13)$$

式中： $j = 0, 1, 2, \dots, 2KY$

类似①的做法可得到  $2KS$  对具有显性上位作用基因其  $F_2$  代各类型出现的概率：

$$pr_5(X_j) = \sum_{\substack{j_1=0 \\ 2j_1+j_2=j \\ j_1+j_2 \leq KS}}^{KS} \sum_{j_2=0}^{KS} \binom{KS}{j_1, j_2} \left(\frac{3}{4}\right)^{j_1} \cdot \left(\frac{3}{16}\right)^{j_2} \left(\frac{1}{16}\right)^{KS-j_1-j_2} \quad (14)$$

式中： $j = 0, 1, 2, \dots, 2KS$

4. 基因型值G的分布：综合上述所做假设有：

$$\begin{cases} 0 \leq N_1, KX, KF, 2KY, 2KS \leq N \\ N_1 + KX + KF + 2KY + 2KS = N \end{cases} \quad (15)$$

又基因间没有连锁，不同基因座效应通过相加组合。所以，具有相同状态  $X_K$  的各类基因型出现的概率  $pr(X_K)$  可以由离散卷积公式求出：

$$pr(X_K) = pr_1(X_{i_1}) \oplus pr_2(X_{i_2}) \oplus pr_3(X_{i_3}) \oplus pr_4(X_{i_4}) \oplus pr_5(X_{i_5})$$

$$= \sum_{\substack{i_1=0 \\ i_1+i_2+i_3+i_4+i_5=K}}^{2N_1} \sum_{i_2=0}^{KX} \sum_{i_3=0}^{KF} \sum_{i_4=0}^{2KY} \sum_{i_5=0}^{2KS}$$

$$\cdot \left\{ \left[ \binom{2N_1}{i_1} \left(\frac{1}{2}\right)^{2N_1} \right] \left[ \binom{KX}{i_2} \left(\frac{3}{4}\right)^{i_2} \left(\frac{1}{4}\right)^{KX-i_2} \right] \right.$$

$$\cdot \left[ \binom{KF}{i_3} \left(\frac{1}{4}\right)^{i_3} \left(\frac{3}{4}\right)^{KF-i_3} \right]$$

$$\cdot \left[ \sum_{\substack{j_1=0 \\ 2j_1+j_2=i_4 \\ j_1+j_2 \leq KY}}^{KY} \sum_{j_2=0}^{KY} \binom{KY}{j_1, j_2} \left(\frac{9}{16}\right)^{j_1} \right.$$

$$\cdot \left. \left[ \sum_{\substack{j_1=0 \\ 2j_1+j_2=i_5 \\ j_1+j_2 \leq KS}}^{KS} \sum_{j_2=0}^{KS} \binom{KS}{j_1, j_2} \left(\frac{3}{4}\right)^{j_1} \right] \right\}$$

$$\cdot \left. \left[ \sum_{\substack{j_1=0 \\ 2j_1+j_2=i_5 \\ j_1+j_2 \leq KS}}^{KS} \sum_{j_2=0}^{KS} \binom{KS}{j_1, j_2} \left(\frac{3}{16}\right)^{j_2} \left(\frac{1}{16}\right)^{KS-j_1-j_2} \right] \right\}$$

$$\cdot \left. \left[ \sum_{\substack{j_1=0 \\ 2j_1+j_2=i_5 \\ j_1+j_2 \leq KS}}^{KS} \sum_{j_2=0}^{KS} \binom{KS}{j_1, j_2} \left(\frac{3}{4}\right)^{j_1} \right] \right\}$$

(16)

其中:  $X_K = \min(p_1, p_2) + |p_1 - p_2| \times K/M$

$$M = 2N_1 + KX + KF + 2KY + 2KS$$

$$K = 0, 1, 2, \dots, M$$

$p_1, p_2$  分别为两极端亲本性状值

⊕为卷积符

容易得出, 数量性状  $F_2$  代基因型值  $G$  的概率密度函数为:

$$f_G(x) = \sum_{K=0}^M pr(X_K) \delta(x - X_K) \quad (17)$$

其中:  $\delta(x - X_K)$  为  $\delta$  函数, 满足:

$$\delta(x - X_K) = \begin{cases} 1 & x = X_K \\ 0 & \text{其他} \end{cases} \quad (18)$$

### (二) 环境变量 $E$ 的分布

由式(5)即可得出环境变量的概率密度函数:

$$f_E(x) = \frac{1}{\sqrt{2\pi}V_e} e^{-\frac{x^2}{2V_e^2}} \quad (19)$$

### (三) 表型值 $x$ 的分布

由式(5), 依据概率论原理, 数量性状  $F_2$  代表型值分布密度函数由下式确定:

$$f_{F_2}(\theta, x) = \int_{-\infty}^{\infty} [f_G(\theta, y) \cdot f_E(y - x)] dy \quad (20)$$

式中:  $\theta = \{N_1, KX, KF, KY, KS\}^T$  为参数向量。

利用  $\delta$  函数的性质容易得出:

$$f_{F_2}(\theta, x) = \sum_{i=0}^M pr(\theta, X_i) \cdot \frac{1}{\sqrt{2\pi}V_e} e^{-\frac{(x - X_i)^2}{2V_e^2}} \quad (21)$$

由式(21)可看出, 当环境方差已知时, 数量性状  $F_2$  代群体分布由基因数目确定, 故将式(21)称为基因模型。

## 参数辨识方法

在辨识模型参数之前, 首先利用亲本  $P_1, P_2$  和  $F_1$  代的样本估计出两亲本性状值及环境方差。估计式如下:

$$\hat{p}_1 = \frac{1}{N_{P_1}} \sum_{i=1}^{N_{P_1}} x_{P_1, i} \quad (22)$$

$$\hat{p}_2 = \frac{1}{N_{P_2}} \sum_{i=1}^{N_{P_2}} x_{P_2, i} \quad (23)$$

$$\hat{V}_2^2 = \frac{1}{3} \left\{ \frac{1}{N_{P_1} - 1} \sum_{i=1}^{N_{P_1}} (x_{P_1,i} - \hat{p}_1)^2 + \frac{1}{N_{P_2} - 1} \sum_{i=1}^{N_{P_2}} (x_{P_2,i} - \hat{p}_2)^2 + \frac{1}{N_{F_1} - 1} \sum_{i=1}^{N_{F_1}} \left( x_{F_1,i} - \frac{1}{N_{F_1}} \sum_{j=1}^{N_{F_1}} x_{F_1,j} \right)^2 \right\} \quad (24)$$

式中:  $N_{P_1}$ 、 $N_{P_2}$ 、 $N_{F_1}$  分别为  $P_1$ 、 $P_2$  及  $F_1$  代样本个数。

$x_{P_1,i}$ 、 $x_{P_2,j}$ 、 $x_{F_1,k}$  分别为  $P_1$  第  $i$  个,  $P_2$  第  $j$  个及  $F_1$  代第  $k$  个样本性状值。

关于模型参数辨识, 即式(21)中  $N_1$ 、 $KX$ 、 $KF$ 、 $KY$ 、 $KS$  估计问题。思路是用  $F_2$  代样本构成的经验分布与式(21)比较, 取使两者差距最小的一组参数作为我们要求的估计值。因此, 需要定义一个损失函数  $L(\theta, x)$ 。它表达在不同参数下, 模型与实际群体分布的差异, 表示为:

$$L(\theta, x) = \int_{-\infty}^{\infty} [f_{F_2}(\theta, x) - f'_{F_2}(x)]^2 dx \quad (25)$$

式中:  $f_{F_2}(\cdot)$ 、 $f'_{F_2}(\cdot)$  分别为理论和经验分布密度函数。

取满足下式的参数值作为估计值:

$$L(\hat{\theta}, x) = \min_{\theta} L(\theta, x) \quad (26)$$

为了方便计算, 可用下式作为近似:

$$L(\hat{\theta}, x) = \min_{\theta} \sum_{i=1}^{N_T} \{ [F_{F_2}(\theta, x^i) - F_{F_2}(\theta, x^{i-1})] / \Delta x^i - [F'_{F_2}(x^i) - F'_{F_2}(x^{i-1})] / \Delta x^i \}^2 \Delta x^i \quad (27)$$

其中:  $F_{F_2}(\theta, x) = \int_{-\infty}^x f_{F_2}(\theta, y) dy$

$$F'_{F_2}(x) = \int_{-\infty}^x f'_{F_2}(y) dy$$

$N_T$  为区间个数

$\Delta x^i = x^i - x^{i-1}$  为第  $j$  个区间长度

由概率论, 容易得出当  $N_T \rightarrow \infty$  时, 式(27)将收敛到式(25)。下面, 我们给出两个关于统计方面的性质。性质 1: 由式(22)~(24)得出的  $\hat{p}_1$ 、 $\hat{p}_2$  和  $\hat{V}_2^2$  是无偏和一致估计; 性质 2: 经验分布  $F'_{F_2}(x)$ , 当样本量  $N_{F_2} \rightarrow \infty$  时将收敛到总体分布。

由这两个性质自然得到如下定理。定理: 设数量性状表型值由式(5)表达, 且满足第三节假定 1°~4°, 同时控制性状的基因对数  $N$  是有限的。则有: (1) 该性状  $F_2$  代分布密度函数理论上由式(21)确定; (2) 总能找到一组  $\hat{\theta}$  满足式(26); (3) 当样本量趋于无穷大时, 即  $N_{P_1} \rightarrow \infty$ 、 $N_{P_2} \rightarrow \infty$ 、 $N_{F_1} \rightarrow \infty$ 、 $N_{F_2} \rightarrow \infty$ , 估计值  $\hat{\theta}$  将收敛到真实的参数值  $\theta$ 。

定理的结论是显然的, 故略去证明(详见[6])。最后, 我们讨论一种特殊情况。在这种情况下, 两个亲本  $P_1$  和  $P_2$  基因型中都既有类似  $A_i A_i$  的基因对, 也有类似  $a_i a_i$  的基因对。换句话说, 这两亲本的性状值满足:

$$p_2^* < p_1, p_2 < p_1^* \quad (28)$$

其中: 性状值为  $p_1^*$  的亲本基因型中基因对全为  $A_i A_i$ ,  $i = 1, 2, \dots, N$ ; 性状值取  $p_2^*$  的亲本基因型中基因对全为  $a_i a_i$ ,  $i = 1, 2, \dots, N$ 。

满足式(28)的亲本称为非极端类型亲本。也就是说上述情况不满足假定 3°。但是若满足下述约束, 我们仍可利用这两个亲本及其杂交后  $F_1$  和  $F_2$  代的资料辨识出模型的参数。假设两亲本有如下特点: 在同一个基因座上(设第  $i$  个基因座,  $i = 1, 2, \dots, N$ )若亲本  $P_1$  取基因对  $A_i A_i$ , 则亲本  $P_2$  取基因对  $a_i a_i$ , 反之亦然。显然, 符合这一条件的两亲本杂交后  $F_1$  代所有基

因座上的基因对都是杂合的(即 $A_i a_i, i = 1, 2, \dots, N$ )。这与由假定1°—4°得出的结果相同。因此,  $F_2$ 代分布密度函数仍将服从式(21)。剩下的问题是需要确定个别基因作用EFF。它由这两亲本性状值及这两个亲本各类型有利基因对相差对数决定。为此, 我们引入另外几个参数 $n_1$ 、 $KX_1$ 、 $KF_1$ 、 $KY_1$ 、 $KS_1$ 分别表示两亲本的各类型有利基因对相差数目(设 $p_1 > p_2$ ,  $n_1$ 表示亲本 $P_1$ 基因型中 $N_1$ 类基因对类似 $A_i A_i$ 的比亲本 $P_2$ 要多 $n_1$ 对。其他4个参数的意义类似)。利用这几个参数, 将式(16)中说明部分作如下修改(当 $|p_1 - p_2|$ 太小时估计结果不可靠):

$$\left\{ \begin{aligned} X_K &= \min(p_1, p_2) + [K - (N_1 - n_1 + KY - KY_1 + KS - KS_1 + \frac{1}{2}(KX - KX_1 + KF - KF_1))] \\ &\quad \times EFF \end{aligned} \right. \quad (29)$$

$$EFF = |p_1 - p_2| / M_1$$

$$M_1 = 2n_1 + KX_1 + KF_1 + 2KY_1 + 2KS_1$$

代入式(27)即可求出一组参数 $\hat{\theta}_1$  ( $\theta_1 = [N_1, n_1, KX, KX_1, KF, KF_1, KY, KY_1, KS, KS_1]^T$ , 代替式(27)中的 $\theta$ ), 且仍具备原有的估计性质。

### 实例及结论

表 1 基于分布函数估计与Castle-Wright估计实例比较

Table 1 Comparison of estimation based on distribution function with Castle-Wright estimation

名称 Name	来源 Source	估计结果 Result of estimation	损失函数 Loss function
Leaf-margin	Donald R.	$N = 2, KF = 6, KY = 2$ ( $N_1 = 4$ )	0.0187 (0.0653)
Leaf and stem surface	Donald R.	$N = 4, KX = 1, KF = 5$ $KS = 1(N_1 = 4)$	0.0183 (0.0706)
Leaf thickness	Donald R.	$KF = 6, KY = 2$ ( $N_1 = 3$ )	0.0146 (0.0887)
Leaf cuticle	Donald R.	$N = 3, KX = 3, KY = 1$ $KF = 4(N_1 = 3)$	0.0133 (0.0386)
Length of stomatal apparatus	Donald R.	$KF = 6, KS = 1$ ( $N_1 = 3$ )	0.0069 (0.0278)
Mass Fla	Tan W.Y.	$N = 6, KF = 6$ ( $N_1 = 7$ )	0.0121 (0.0221)
Md Fla	Tan W.Y.	$N = 6, KF = 6$ ( $N_1 = 7$ )	0.0099 (0.0257)
Mass Md	Tan W.Y.	$N = 1, KX = 1, KF = 1$ $KY = 4, KY = 2(N_1 = 3)$	0.0040 (0.0287)
玉米棒长	刘祖洞	$N = 1, KF = 2, KS = 2$ ( $N_1 = 5$ )	0.00136 (0.01790)

\* 带括号的是Castle-Wright公式的结果

Data in bracket are the results of Castle-Wright formula estimation

依据前述的估计方法, 我们编制了相应的计算机程序, 并将它应用于几个实例。从计算结果(表1)可以看出, 应用本文所建模型得出的损失函数均低于 Castle-

Wright 公式的。说明该模型能够更好地反映实际的分布。这将为有效地分析数量性状的遗传分布提供更为可靠的理论依据。

## 参 考 文 献

- [1] Mather, K. et al.: *Biometrical Genetics*, 3rd edn., London, 1980.  
 [2] Castle, W.E.: *Science*, 54:223, 1921.  
 [3] Wright, S.: *Genetics*, 19:506—536, 1934.  
 [4] 郭平仲: 数量遗传分析, 北京师范学院出版社, 1987。  
 [5] 马育华: 植物育种的量遗传学基础, 江苏科技出版社, 1982。  
 [6] 陈 剑: 作物良种培育遗传过程的系统分析及控制, 清华大学博士论文, 1989。  
 [7] Donald, R. et al.: *The American Naturalist*, 77:63—69, 1943.  
 [8] Tan, W.Y. et al.: *Biometrics*, 28:1073—1090, 1972.  
 [9] 刘祖洞等: 遗传学, 上册, 高等教育出版社, 1979。

## GENE MODEL AND ITS IDENTIFICATION

Chen Jian Zheng Weimin Wang Yongxian

(Systems Engineering Lab, School of Economics and Management, Tsinghua University, Beijing)

Quantitative characters are controlled by polygene and their phenotypic values are influenced by environmental factors. In order to analyze quantitative character effectively, a gene model is established, which can describe the population distribution of a quantitative character quite well because it not only involves the additive effect of genes, but also considers the dominance deviation and epistatic deviation. In this paper, we also present a method for identifying the parameters of the model. From the results of some examples, the model is superior to Castle-Wright formula, and it may provide more useful information to breeders.

## Key words

Gene model; population distribution; parameter identification