群体遗传及其程序设计

毛盛贤　黄远樟　编

北京师范大学出版社

高等学校教学用书

群体遗传及其程序设计

毛盛贤　　黄远樟　编

北京师范大学出版社

高等学校教学用书
群体遗传及其程序设计
毛盛贤　黄远樟　编

*

北京师范大学出版社出版
新华书店总店科技发行所发行
宝坻第十印刷厂印刷

开本：787×1092　1/16　印张：19.125　字数：441千
1991年2月第1版　　1991年2月第1次印刷
印数：1—1 200

ISBN7-303-01031-9/Q·20
定价：5.00元

前　言

（一）

我们学过的孟德尔遗传学，是以单个家系为繁殖单位，研究基因由亲代至子代的传递和表达的规律。现在要学的群体遗传学，是以一个群体的所有家系为繁殖单位，研究基因在世代间的传递和表达的规律。在理论上，群体遗传学的最终目的是阐明生物进化的机制；在实践上，为动、植物和微生物育种以及人类某些遗传病产生的原因和遗传咨询提供理论依据。

（二）

纵观群体遗传学的发展史，可追溯到达尔文在《物种起源》（1859）中提出的自然选择学说。他主张生物进化主要是自然选择的结果，特别强调微小变异的累积在进化上的重要性。

1900年，孟德尔的颗粒遗传理论的重新发现，标志现代遗传学的开始。但由于孟德尔学派是以质量性状（受单个基因效应大的一个或少数几个基因座控制）为研究对象，与主张微小变异在进化上具有重要性的自然选择学说结合不起来。这时以 J.B.S.We'don为代表的生统学派坚信并已实际证实，在微小的连续变异中，存在遗传变异分量。于是，在连续变异是否能遗传和其在进化中的重要性，孟德尔学派和生统学派展开了激烈的争论。

在争论过程中，G.V.Yele(1906) 提出了多因子假说，认为连续变异可由效应微小且可加的许多基因产生。这样，用生统方法从遗传上解释连续变异的原因才成为可能。

在上述条件下，由R.A.Fisher (1930)，J.B.S.Haldane (1924—1932) 以及S.Wright (1931) 等奠定了群体遗传学的基础，即认为突变提供群体的遗传变异，经自然选择使群体遗传组成发生变化，于是出现了生物进化。至于遗传漂变对进化的重要性，虽有过强调它的S.Wright和忽视它的R.A.Fisher间的争论，但现在看来，它应占有相当份量。

随着分子遗传学的进展，现在可从生物大分子水平对群体进行遗传变异的研究。关于遗传多态现象的维持和分子进化的机制，木村资生（1968）采取与自然选择为基础的达尔文进化理论完全相反的立场，提出了以遗传漂变为基础的中性基因的分子进化学说。

（三）

本书在为本科生讲授选修课和为遗传专业研究生所编讲义的基础上编写而成。目的是介绍群体遗传的基本原理和方法，以及群体遗传的计算机计算和模拟技术。全书分两篇：上篇，群体遗传基本原理（第一至第九章）；下篇，群体遗传程序设计（第十至第十四章）。

第一章介绍群体遗传学必备的概率统计基础。第二章讨论群体的遗传组成，重点阐明

在随机交配的不同情况下（二倍体群体、复等位基因、性连锁基因、自交不亲和、多倍体和两基因座）的遗传平衡条件，以及在近交下遗传变异与近交系数的关系。第三章涉及突变、选择、迁移和遗传漂变诸因素对生物进化的效应或影响。第四章是以群体遗传学的理论，说明物种是如何形成或生物是如何进化的。第五章简述群体遗传学和人类生活的关系。

以上各章考虑的都是质量性状。从第六章至第九章，考虑的则是受许多基因控制和易受环境影响的数量性状。在讨论数量性状遗传模型的基础上，主要结合动物育种，讨论了遗传参数（遗传率、重复率、育种值、表型相关、遗传相关、环境相关）及其估算，选择原理和选择指数。

下篇主要是根据上篇介绍的原理和方法，如何在计算机上实现群体遗传的计算和模拟。首先介绍与群体遗传有关的矩阵运算和概率统计的基本程序块，然后介绍有代表性的群体遗传的计算和模拟程序。

多数章节后附有练习题，以帮助读者了解对内容掌握的程度。少数章节未附练习题，或是为了避免与计算机部分重复，或是只具有复习性质，对群体遗传本身并非重点（如第一章）。书末附有问题解答方法和答案，以供读者参考。

上、下篇分别由毛盛贤、黄远樟执笔。限于编者水平，书中内容可能有不少缺点和错误，请读者批评指正。

作者于北京师范大学生物系

1988.8.28

目　　录

下篇 群体遗传程序设计

上篇　群体遗传基本原理

第一章　群体遗传学的概率统计基础

要学好群体遗传学，除具备普通遗传学基础外，还应了解一些概率统计基础。现就以后在学习群体遗传学时要涉及到的概率统计作一简单介绍。

§1　随机事件及其概率

一、随机事件

某事物发生的一种情况或实验中得到的一种结果称为一个事件。在一定条件下必然发生或不可能发生的事件分别称为必然事件或不可能事件。例如，同型合子自交的后代为同型合子是必然事件，同型合子自交的后代为杂型合子是不可能事件。这两类事件统称为确定事件。

在一定条件下可能发生也可能不发生的事件称为随机事件。例如，杂型合子（Aa）自交时，自交一代中一个个体基因型组合的可能结果是AA、Aa、aA或aa，这些可能结果都叫随机事件。

自交一代中这些随机事件有如下性质：根据孟德尔原理，每一事件的发生都是等可能的（各占1/4）；任两事件都不可能同时发生，它们是互斥的；每次试验至少有一事件发生，即它们构成一个完备群。也就是说，事件AA、Aa、aA和aa构成互斥的且等可能的完备群。构成这种完备群的事件称为基本事件。

由基本事件构成的事件叫复合事件。例如，令A对a呈完全显性，在上述自交一代中抽到“显性个体”的事件A_就称为由事件AA、Aa和aA构成的复合事件。

对随机事件进行个别观测时，得到什么样的结果似乎是莫测的。但作大量观测时，一般来说，就可探索出其发生的规律性。例如，令上述杂型合子Aa是两同型合子AA（开红花）和aa（开白花）豌豆的杂交一代。在Aa的自交一代中，就一株豌豆来说，随机抽到的可能是开红花的，也可能是开白花的，看不出什么规律性；但是，对大量个体作调查时，就会发现：

$$红花植株:白花植株\approx 3:1$$

的孟德尔式分离比。一般来说，调查的个体数越多，这两类植株的比就越接近这两个数。也就是说，随机事件的发生是有规律的。许多生命现象具有随机性，配子和合子的形成都是随机的。遗传学的三大基本规律，实际上是有关随机事件的遗传规律。

二、概率定义

试验中事件A发生的概率（或简称A的概率）是事件A发生的次数与试验中所有可能

事件总次数之比。假定在N次试验中出现事件A为N_A次（这些事件的出现是等可能的），那么事件A的概率

$$p(A)=N_A/N$$

也就是说，在随机试验中事件A的概率是其发生可能性大小的量度。

例1　抛一次质地均匀的钱币，“正面向上”的概率为多少？

这里的随机试验是抛钱币。可能结果是“正面向上”和“反面向上”。事件A是“正面向上”。假定N＝2，则$N_A=1$。于是“正面向上”的概率

$$p(A)=N_A/N=1/2$$

例2　根据孟德尔原理，杂型合子Aa与隐性同型合子aa杂交（测交）产生数目相等的两种后代Aa和aa。同样，“后代为Aa”的概率

$$p(A)=N_A/N=1/2$$

例3　群体中有120只羊，其中39只为雌羊，81只为雄羊。令事件F为“抽到一雌羊”，M为“抽到一雄羊”，则相应的概率

$$p(F)=M_F/N=39/120=13/40$$

$$p(M)=M_M/N=81/120=27/40$$

由上可知，某事件的概率就是该事件发生的理想化比率（或相对频率）。因此，任一事件A的概率的取值只能在零和1之间，即

$$0\leqslant p(A)\leqslant 1$$

$$p(\text{必然事件})=1$$

$$p(\text{不可能事件})=0$$

我们根据基本事件讨论了概率的基本概念。现转向复合事件概率的讨论。

三、复合事件——事件积的概率

考虑有800个小孩的一个群体，依性别和是否患某种遗传病这两个性状进行分类。令：A，患某种遗传病；$\bar{A}$，正常；B，男孩；$\bar{B}$，女孩。结果见表1-1。

表 1-1　800个小孩的二向分类（假设数据）

疾病 \ 性别	（男孩B）	女孩（$\bar{B}$）	行总和
患病（A）	$N_{AB}(70)$	$N_{A\bar{B}}(50)$	$N_A(120)$
正常（$\bar{A}$）	$N_{\bar{A}B}(330)$	$N_{\bar{A}\bar{B}}(350)$	$N_{\bar{A}}(680)$
列总和	$N_B(400)$	$N_{\bar{B}}(400)$	$N(800)$

由表1-1，事件AB（患病男孩）表示事件A和事件B同时（或相继）发生的一个复合事件，称为事件A和B的积。事件的积是一个事件。$A\bar{B}$（患病女孩）、$\bar{A}B$（正常男孩）和$\bar{A}\bar{B}$（正常女孩）都是两相应事件的积。显然，根据概率定义，事件积的概率

$$p(AB)=N_{AB}/N=70/800$$

类似可求出$p(A\bar{B})$、$p(\bar{A}B)$和$p(\bar{A}\bar{B})$。

1．条件概率　在事件A发生的条件下事件B发生的概率称为“给定A时B的条件概

率"，记作$p(B|A)$。

计算$p(B|A)$的值，直接由表1－1得：

$$p(B|A)=N_{AB}/N_A=70/120$$

或者，因为

$$p(AB)=N_{AB}/N \qquad p(A)=N_A/N$$

$$\therefore\ p(B|A)=\frac{N_{AB}}{N_B}=\frac{N\cdot p(AB)}{N\cdot p(A)}=\frac{p(AB)}{p(A)}=\frac{70/800}{120/800}=\frac{70}{120}$$

即 $p(B|A)=p(AB)/p(A)$

同理，

$$p(A|B)=p(AB)/p(B)$$

2．事件独立　如果给定B时A的条件概率等于A的概率，称事件A独立于事件B，记作

$$p(A|B)=p(A)$$

这说明，A发生的概率不受B事件发生与否的影响。显然，如果A独立于B，则A也独立于$\bar{B}$，即

$$p(A|B)=p(A)=p(A|\bar{B})$$

其中B和$\bar{B}$互为对立（或互补）事件。

在实际问题中，为确定事件A是否独立于事件B，要计算$p(A|B)$和$p(A)$。如果这两个值相等，就说明A独立于B。例如，根据表1－1：

$$p(A|B)=70/400 \qquad p(A)=120/800=60/400$$

由于$p(A|B)$不等于$p(A)$，该遗传病依赖于孩子的性别（这两个值差异的显著性检验见该章§3）。

3．乘法定理　事件积AB的概率等于A的概率与给定A时B的条件概率之乘积，即

$$p(AB)=p(A)p(B|A)$$

证

$$p(AB)=N_{AB}/N=\frac{N_A}{N}\cdot\frac{N_{AB}}{N_A}=p(A)p(B|A)$$

证毕。

因为事件AB与事件BA相同，所以乘法定理的另一公式

$$p(AB)=p(B)p(A|B)$$

对于3个和4个事件的乘法定理公式是：

$$p(ABC)=p(A)p(B|A)p(C|AB)$$

$$p(ABCD)=p(A)P(B|A)p(C|AB)p(D|ABC)$$

通过式$p(AB)=p(A)p(B|A)$的重复替换可证明以上二式成立。例如

$$\begin{aligned}p(ABC)&=p(A)p(BC|A)\\&=p(A)p(B|A)p(C|AB)\end{aligned}$$

4。独立事件的乘法定理　如果各事件彼此独立，以上乘法定理诸公式就成为

$$p(AB)=p(A)p(B)$$

$$p(ABC)=p(A)p(B)p(C)$$

$$p(ABCD)=p(A)p(B)p(C)p(D)$$

四、复合事件——事件和的概率

参考表1－1，事件（A＋B）表示事件A、B和AB中每次发生其一的复合事件，称为事件A和B的和。事件的和是一个事件。

1．加法定理　事件和（A＋B）的概率为

$$p(A+B)=p(A)+p(B)-p(AB)$$

证：

如表1－1，事件和（A＋B）会在什么情况下发生呢？为此，把试验中的所有可能结果分成四类：①发生AB、A和B；②发生$A\bar{B}$和A但不发生B；③发生$\bar{A}B$和B但不发生A；④发生$\bar{A}\bar{B}$但不发生A和B。因此，事件和（A＋B）只发生在前三类。

（A＋B）的概率是，依定义：由表1－1，

$$N_{A\bar{B}}=N_A-N_{AB}\qquad N_{\bar{A}B}=N_B-N_{\bar{A}\bar{B}}$$

$$\therefore\quad p(A+B)=(N_A+N_B-N_{AB})/N=N_A/N+N_B/N-N_{AB}/N$$

即$p(A+B)=p(A)+p(B)-p(AB)$

例　在表1－1中，A＝患病小孩，B＝男孩，所以p(A＋B)是患病女孩、正常男孩或患病男孩的概率。这里，我们有：

$$p(A+B)=p(A)+p(B)-p(AB)$$

$$=\frac{120}{800}+\frac{400}{800}-\frac{70}{800}=\frac{450}{800}$$

对于三个和四个事件的加法定理是：

$$\begin{aligned}p(A+B+C)=&p(A)+p(B)+p(C)\\&-p(AB)-p(BC)-p(AC)+p(ABC)\\p(A+B+C+D)=&p(A)+p(B)+p(C)+p(D)\\&-p(AB)-p(AC)-p(AD)-p(BC)-p(BD)-p(CD)\\&+p(ABC)+p(ABD)+p(ACD)+p(BCD)\\&-p(ABCD)\end{aligned}$$

2．互斥事件的加法定理　如果有两个事件，一个发生而另一个必然不发生，或者说在一次试验里只能发生其中之一，则称它们为互斥事件。例如表1-1中的B和$\bar{B}$是两互斥事件，所以$N_{B\bar{B}}=0$和$p(B\bar{B})=0$。

显然，如果以上加法定理中的各事件A、B（不是表1-1中的A、B）、C、D为互斥事件，则有关诸项的事件积的概率都等于零。于是有：

$$p(A+B)=p(A)+p(B)$$

$$p(A+B+C)=p(A)+p(B)+p(C)$$

$$p(A+B+C+D)=p(A)+p(B)+p(C)+p(D)$$

五、概率运算的分配律

在代数运算中有分配律

$$A(B+C)=AB+AC$$

$$(A+B)(C+D)=AC+AD+BC+BD$$

同样在概率运算中也有分配律

$$p[A(B+C)]=p(AB+AC)$$

$$p[(A+B)(C+D)]=p(AC+AD+BC+BD)$$

根据表1－1，我们有

$$p[A(B+\bar{B})]=p(AB+A\bar{B})$$

$$=p(AB)+p(A\bar{B})=\frac{70}{800}+\frac{50}{800}=\frac{120}{800}$$

在该情况下，$(B+\bar{B})=I$，是一必然事件，所以

$$p[A(B+\bar{B})]=p(AI)=p(A)=\frac{120}{800}$$

从而验证了运用分配律计算概率的有效性。

六、概率定理在遗传上的应用举例

例1　设一随机交配群体的基因型为AA、Aa和aa，相应的基因型频率为P、H和Q。试求交配为$Aa\times Aa$且子代为AA的概率多少？

解：设$Aa\times Aa$交配为事件A，该交配产生的子代AA为事件B，则依题意

$$p(AB)=p(A)p(B|A)$$

而　$p(A)=H^2$；$Aa\times Aa\rightarrow\frac{1}{4}AA+\frac{1}{2}Aa+\frac{1}{4}aa$，即$p(B|A)=1/4$。所以

$$p(AB)=p(A)p(B|A)=\frac{1}{4}H^2$$

例2　独立遗传的、具有4对基因差异的两同型合子杂交，试求F_2中隐性纯合体的概率。

解：两同型合子杂交后的F_1的基因型，比如说AaBbCcDd。因为F_1形成配子时，任一配子含A的概率$p(A)=1/2$，含a的概率$p(a)=1/2$；实际上，任一配子含每对等位基因中的任一基因的概率都等于1/2，即$p(A)=p(a)=p(B)=\cdots=p(d)=1/2$。又因有关4个基因座的基因位于不同的染色体上，它们进入一个配子的事件互为独立，故基因a、b、c和d同时进入一个配子的概率$p(abcd)=p(a)p(b)p(c)p(d)=(1/2)^4=1/16$。显然，含有a、b、c和d基因的雌配子和雄配子的概率都为1/16。

F_1的雌、雄配子结合产生F_2。由于这二者的结合是随机的，即为两独立事件，所以F_2是隐性同型合子的概率为

$$p[(abcd)♀(abcd)♂]=p(abcd)♀p(abcd)♂$$

$$=\left(\frac{1}{16}\right)^2=\frac{1}{256}$$

例3　在开红花的豌豆杂型合子Cc的自交一代中，有三类基因型：1/4CC＋1/2Cc＋1/4cc。那么自交一代表型为开红花的概率为

$$p(CC+Cc)=p(CC)+p(Cc)=1/4+1/2=3/4$$

例4　孟德尔的豌豆试验，籽粒颜色黄（Y）对绿（y）呈显性，籽粒形状圆（R）对皱（r）呈显性。对杂型合子F_1RrYy自交产生F_2的期望比如表1－2。

表 1-2

	$\frac{3}{4}$R_(圆)	$\frac{1}{4}$rr(皱)
$\frac{3}{4}$Y_(黄)	$\frac{9}{16}$R_Y_(圆、黄)	$\frac{3}{16}$rrY_(皱、黄)
$\frac{1}{4}$yy(绿)	$\frac{3}{16}$R_yy(圆、绿)	$\frac{1}{16}$rryy(皱、绿)

所以在F_2中，出现圆形籽粒的概率

$$\begin{aligned}p(R\_) &= p(R\_Y\_ + R\_yy)\\ &= p(R\_Y\_) + p(R\_yy) = 9/16 + 3/16 = 3/4\end{aligned}$$

出现皱形籽粒的概率

$$\begin{aligned}p(rr) &= p(rrY\_ + rryy)\\ &= p(rrY\_) + p(rryy) = 3/16 + 1/16 = 1/4\end{aligned}$$

出现皱形或黄色籽粒的概率（这两性状不独立）

$$\begin{aligned}p(rr + Y\_) &= p(rr) + p(Y\_) - p(rrY\_)\\ &= 1/4 + 3/4 - 3/16 = 3/16\end{aligned}$$

出现圆形或黄色籽粒的概率

$$\begin{aligned}p(R\_ + Y\_) &= p(R\_) + p(Y\_) - p(R\_Y\_)\\ &= 3/4 + 3/4 - 9/16 = 15/16\end{aligned}$$

例 5　设一群体的基因型为AA、Aa和aa，频率相应为P、H和Q。试求在随机交配条件下，子代为AA、Aa和aa的概率各为多少？

这需要综合应用概率的加法和乘法定理。首先考虑随机交配时有哪几种交配类型：

$$\begin{array}{lll} M_1 = AA \times AA & M_2 = AA \times Aa & M_3 = AA \times aa\\ M_4 = Aa \times AA & M_5 = Aa \times Aa & M_6 = Aa \times aa\\ M_7 = aa \times AA & M_8 = aa \times Aa & M_9 = aa \times aa \end{array}$$

其次要考虑在以上九种交配类型中，哪几种交配类型产生子代AA。显然是M_1、M_2、M_4和M_5。由于产生AA的这 4 种交配类型是两两互斥的，所以它们产生子代 AA 的概率，应用加法定理：

$$\begin{aligned}p(\text{子代}AA) &= p(\text{交配}AA \times AA \cdot \text{子代}AA)\\ &+ p(\text{交配}AA \times Aa \cdot \text{子代}AA)\\ &+ p(\text{交配}Aa \times AA \cdot \text{子代}AA)\\ &+ p(\text{交配}Aa \times Aa \cdot \text{子代}AA)\end{aligned}$$

再次要考虑各交配类型产生子代为AA的概率。因子代AA的概率受交配类型的影响，即一定的交配类型和其产生的子代AA这两事件不独立，故求各交配类型产生子代为AA的概率时，要用乘法定理，例如

$$p(\text{交配}AA \times AA \cdot \text{子代}AA) = p(\text{交配}AA \times AA)\,p(\text{子代}AA \mid \text{交配}AA \times AA)$$

在交配AA×AA中，因AA(♀)和AA(♂)互为独立事件，故其交配概率

$$p(\text{交配}AA \times AA) = p(AA)\,p(AA) = P^2$$

又AA×AA→AA，即在该交配中只产生子代AA，故

$$p(\text{子代}AA \mid \text{交配}AA \times AA) = 1$$

同理，可求出其它类型的交配概率及在相应交配类型下所产生各类子代的条件概率。现汇总于表 1－3。由该表可得

$$p(\text{交配}AA \times AA \cdot \text{子代}AA) = p(\text{交配}AA \times AA)\,p(\text{子代}AA \mid \text{交配}AA \times AA)$$
$$= P^2 \cdot 1 = P^2$$

$$p(\text{交配}AA \times Aa \cdot \text{子代}AA) = p(\text{交配}AA \times Aa)\,p(\text{子代}AA \mid \text{交配}AA \times Aa)$$
$$= HP \cdot \frac{1}{2} = \frac{1}{2}PH$$

$$p(\text{交配}Aa \times AA \cdot \text{子代}AA) = p(\text{交配}Aa \times AA)\,p(\text{子代}AA \mid \text{交配}Aa \times AA)$$
$$= HP \cdot \frac{1}{2} = \frac{1}{2}PH$$

表 1-3　各交配类型概率及其各类子代的条件概率

交配类型	交配频率	子代条件概率		
		AA	Aa	aa
$M_1 = AA \times AA$	P^2	1	0	0
$M_2 = AA \times Aa$	HP	1/2	1/2	0
$M_3 = AA \times aa$	PQ	0	1	0
$M_4 = Aa \times AA$	HP	1/2	1/2	0
$M_5 = Aa \times Aa$	H^2	1/4	1/2	1/4
$M_6 = Aa \times aa$	HQ	0	1/2	1/2
$M_7 = aa \times AA$	QP	0	1	0
$M_8 = aa \times Aa$	QH	0	1/2	1/2
$M_9 = aa \times aa$	Q^2	0	0	1

$$p(\text{交配}Aa \times Aa \cdot \text{子代}AA) = p(\text{交配}Aa \times Aa)\,p(\text{子代}AA \mid \text{交配}Aa \times Aa)$$
$$= H^2 \cdot \frac{1}{4} = \frac{1}{4}H^2$$

于是以上群体在随机交配条件下产生子代为AA的概率

$$p(\text{子代}AA) = P^2 + \frac{1}{2}PH + \frac{1}{2}PH + \frac{1}{4}H^2$$
$$= P^2 + PH + \frac{1}{4}H^2$$

同理，

$$p(\text{子代}Aa) = HP + PQ + \frac{1}{2}H^2 + HQ + PQ$$

$$p(\text{子代}aa) = \frac{1}{4}H^2 + HQ + Q^2$$

§2　二项分布和多项式概率

在遗传学中，有时是研究一次试验（或一次交配）中可能出现的结果和概率。例如，