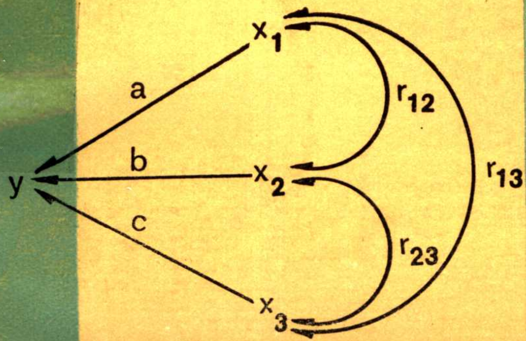

作物遺传参数統計法

兰巨生

河
人民出版社

作物遗传参数统计法

兰 巨 生

河 北 人 民 出 版 社

一九八二年·石家庄

作物遗传参数统计法

兰 巨 生

河北人民出版社出版（石家庄市北马路19号）

河北新华印刷一厂印刷　河北省新华书店发行

787×1092毫米 1/32　8 印张　163,000字　印数：1—3,240　1982年11月第1版

1982年11月第1次印刷　统一书号：16086·368　定价：0.68 元

前　　言

近年来，许多育种工作者根据统计遗传学的原理，提出了各种遗传设计方法和统计模型，用来研究农作物数量性状在不同条件下遗传和变异的规律，以便在实践中正确设计育种规划，选择育种方法，预测选择进展，提高工作效率。

本书广泛收集有关遗传设计方法和统计模型，并结合实际，举出例题，进行演算，简明扼要地向读者介绍了这方面的知识。只要具有高中数学水平，通过实际计算，就能够初步了解模型原理，掌握统计程序，在育种工作中逐步运用。

由于理论水平所限，书中缺点、错误在所难免，衷心希望提出宝贵意见。

兰巨生

1982年3月

目　　录

第一章　集合和概率

一、集合的概念

全国统编教材高中数学中关于集合的概念是这样写的：

把具有某种属性的一些对象看做一个整体便形成一个集合。

这就是说，集合是指一类事物，从性质上可资识别。它是一类有定义的事物的概括。例如，你在学校读书时，全班的男同学是一个集合，全班的女同学也是一个集合。集合中每一个个体，或者说每一个对象是这个集合中的元素。

你如果把全班同学看做一个集合，即它由另外若干个集合所构成，那么，这样的集合称为全集，男同学的集合和女同学的集合则是这个全集的子集。

你现在已参加了作物育种工作，例如在亲本圃里种着各种小麦品种，有的是抗条锈的，有的是抗叶锈的，有的什么也不抗，还有的则是兼抗条锈和叶锈的。用图表示，则如图1。

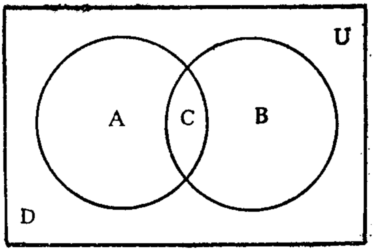

图 1

图中：A 表示抗条锈的品种；B 表示抗叶锈的品种；C表示兼抗两种锈病的品种；D 表示不抗病的品种。

A、B、C、D都是子集，U 则是一个全集。

为了运算方便，需要引进一些符号。

我们常用大写拉丁字母表示集合，用小写的拉丁字母表示元素，有时用具体的数字表示元素。如果 a 和 b 是集合 A 的元素，就说 a 和 b 属于集合 A，记为

$$a\in A \quad \text{和} \quad b\in A$$

设有 A 和 B 两个集合，如果集合 B 中的任一元素都是集合 A 的元素，那么集合 B 就称为集合 A 的子集，记为

$$B\subset A \quad \text{或} \quad A\supset B$$

（B 包含于 A）　　（A 包含 B）

例如

$$A=\{1,3,5,7,9\}$$

$$B=\{3,5\}$$

即

B 是 A 的子集。

由集合 A 和集合 B 的共同元素构成的集合，称为A 和B

的交集，即交集中的所有元素同时属于 A 和 B。A 和 B 的交集记为 $A\cap B$，用图表示则为图 2：

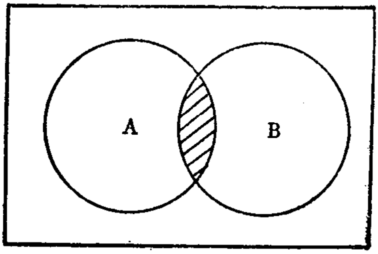

图 2

设

$$A=\{8,7,6,2,5\}$$

$$B=\{5,2,6,9,10\}$$

则

$$A\cap B=\{5,2,6\}$$

由属于 A 和属于 B 的一切元素所组成的集合，称为 A 与 B 的并集。记为

$$A\cup B$$

例如

$$A=\{1,2,3,4\}$$

$$B=\{5,6,7,\}$$

$$A\cup B=\{1,2,3,4,5,6,7\}$$

另外，不含任何元素的集合，称为空集合，记为 ϕ。设 A 是全集 U 中的某一个子集，现在另有一个子集，其中的元素，属于 U，但不属于 A，那么这个集合称为集合 A 的补集，记为 A'。例如，在一个原始材料圃中，全集 $U=\{$抗叶锈者，抗条锈者，抗秆锈者，抗白粉病者$\}$，假如 $A=\{$抗叶锈者，抗条锈者，抗秆锈者$\}$，那么 $A'=\{$抗白粉病者$\}$。

二、样本空间和事件

当我们分析研究结果时，会发现各种原始数据有一个重

要的特性，即不确定性。就是说，在实验过程中常碰到机遇因素在起作用。例如在实验室或大田中为某一个小麦品种接种条中 19 号菌种，叶片上可能出现有 5 种 结 果：

$S=\{0;1,2,3,4\}$ （0;1,2,3,4分别表示发病等级）

一个随机试验产生的各种可能的结果，构成一个集合，这个集合称为这个随机试验的样本空间。文献中常把样本空间记为 S。样本空间中的某一个元素，称为样点。例如 $S=\{0;1,2,3,4\}$ 中，0;1,2,3,4 都是样本空间 S 的样点。

实验中，我们经常关心的不是某一个具体的样点，而是要查明这个小麦品种对条中 19 号菌种是抗的？耐的？或是感病的。假如我们把 0；定为专化抗性的级别，把 1 、 2 两级定为具有一般抗性的级别， 3 、 4 则是感病的级别。这样，S 中的 5 个样点就构成 3 个子集：

$S=$｛抗病，具一般抗性，感病｝

以上 S 为全集，而把样本空间 S 的子集定义为事件。

三、概率简述

在一种比较稳定的条件下，重复做 n 次试验，假如事件 A 出现了 m 次，m 称为事件 A 在这种条件下出现的频数，m/n 则称为事件 A 发生的频率。如果经过一段长期重复试验，发现事件 A 发生的频率是相当稳定的，例如掷一个硬币出现正面或反面的频率为 1/2是相当稳定的，那么，我们就称这种稳定性的事件为随机事件。所谓随机事件就是说它的发生与客观上存在的机会密切相关。

随机事件 A 在大量试验中发生的频率稳定地在一个常

数 P 上下摆动，我们就把 P 称为随机事件 A 在这种试验条件下的概率，记为 $P(A)$

$$p(A)=\lim_{n\to\infty}\frac{A}{n}$$

它的含义是，当实验的次数 n 越来越大，趋近于无穷大时（Lim 为 Limit 的缩写，意为极限），事件 A 发生的频率 A/n，有一个稳定的摆动中心，它便是随机事件 A 的概率。

假设有一种矮秆小麦，在它的某一个染色体上携带着一个单主基因 r_3，控制着小麦的矮秆性状，那么，它的相对的等位基因 R_3，即是一个与高秆性状相联系的基因。现在以 r_3 矮秆小麦与 R_3 高秆小麦杂交，F_1 代基因型为 R_3r_3。这样，F_1 代产生的配子，只可能有两种，即

$$S=\{R_3, r_3\}$$

设 R_3 为事件 A，r_3 必为 A 的补集，记为

$$S=\{A, A'\}$$

A 事件出现 1 个配子，按机会来讲，A' 事件也会有一个配子，即 $A=1, A'=1$。那么，在 F_1 的千千万万个配子中，事件 A 出现 n 次，A' 也自然出现 n 次，事件 A 的概率

$$p(A)=\lim_{n+n\to\infty}\frac{n}{n+n}=\frac{1}{2}$$

一般情况下记为

$$p(R_3)=\frac{1}{1+1}=\frac{1}{2}$$

同样

$$p(r_3)=\frac{1}{1+1}=\frac{1}{2}$$

当 F_1 代杂种小麦进行自交时，就相当于父本产生的 R_3 与 r_3 配子以及母本产生的 R_3，r_3 配子的随机交配：

	R_3	r_3
R_3	R_3R_3	R_3r_3
r_3	r_3R_3	r_3r_3

由于等位基因 R_3 对 r_3 为显性，所以 R_3R_3，R_3r_3，r_3R_3 三种基因型的表现型是一样的，都是高秆的，只有 r_3r_3 基因型的小麦是矮秆的。那么，在许许多多 F_1 代植株进行自交时，矮秆后代出现的概率为

$$p(r_3r_3)=\frac{1}{1+3}=\frac{1}{4}$$

以上是我们根据已知条件做出的一种合乎逻辑的推理，这就是通常所说的事前概率。但是，在实验中我们对事前的条件往往一无所知，只能通过大量的实验来估计某一事件客观上存在的概率有多大。事实上，每次实验求得的只是某一事件发生的频率，我们只是通过频率近似地去估计某一事件的概率。

例如，现已查明在小麦群体中至少有三个矮秆基因控制着小麦的矮秆性状。今有高秆株与矮秆株所得的杂种一代小麦进行自交，在 F_2 代分离时，出现的高、矮秆小麦的比例因实

验次数不同，会有不同结果。

F_1代自交在F_2代分离时高秆和矮秆小麦的比例

实验次数	高秆株数	矮秆株数	比例
三次实验总计	96	41	2.34:1
十次实验总计	285	90	3.16:1
二十次实验总计	578	192	3.01:1

可以看到，随着实验次数增加，高、矮秆小麦的比例接近一个常数，亦即矮秆小麦的频率稳定地在某个常数左右摆动。这种概率称为实验概率或者称为经验概率。

$$\text{三次实验}\quad p(\text{矮秆})=\frac{41}{96+41}=29\%$$

$$\text{十次实验}\quad p(\text{矮秆})=\frac{90}{285+90}=24\%$$

$$\text{二十次实验}\quad p(\text{矮秆})=\frac{192}{578+192}=24.96\%\approx25\%$$

这样，我们确认纯合的矮秆基因型在F_2代群体中占25%，当然，表现型为高秆的小麦应占75%。由此推知，这一对杂交亲本的高、矮秆性状由一对主效等位基因所控制，而不是由两对或三对等位基因所控制的。

四、概率的加法和乘法

设S为样本空间，A和B是S中的两个事件，或称为两个子集，如果A和B不相交，则A和B两个事件为互斥事件，如果相交，则不是互斥事件，如图3所示。

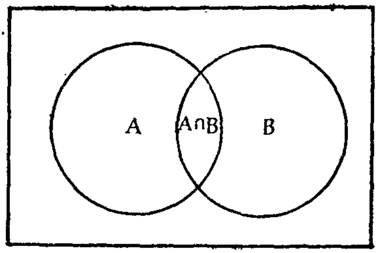

图 3

s样本空间中，事件 A 与 B 相交，相交部分为$A\cap B$·

现在求事件 A 和 B，或子集 A 和 B 的并集的概率有多少。

所谓 A 和 B 的的并集，即 $A\cup B$，是指图 4 中有阴影的部分。

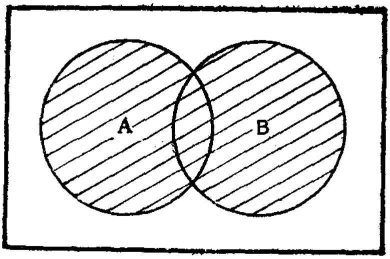

图 4

s 样本空间中A和B的并集

我们用 $P(A)$ 和 $P(B)$ 各自表示 A 和 B 的概率，欲求 A 和 B 并集的概率，应该等于 A 的概率加上 B 的概率。但是由于 A 和 B 相交，在 $P(A)+P(B)$ 中，相交的部分 $P(A\cap B)$，一共加上了两次，而不是一次，所以在总概率中减去一次相交部分的概率才行。这样，A 和 B 并集的概率应为

$$p(A\cup B)=p(A)+p(B)-p(A\cap B)$$

以上称为概率的加法法则。

由上图可以想到。若 A 与 B 互斥，即两个子集不相交，没有共同元素，那么

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

这是因为在 S 样本空间中，$A\cap B$ 是一个空集合，即

$$p(A\cap B)=p(\phi)=0$$

一般地，若 $A_1, A_2, A_3, \cdots\cdots, A_n$ 为互斥事件，就有

$$p(A_1\cup A_2\cup A_3\cdots\cdots\cup A_n)=p(A_1)+p(A_2)+p(A_3)+\cdots\cdots+p(A_n)$$

例如，小麦有一对等位基因控制着抗或不抗干热风的性状，在杂种二代群体中 RR，Rr, rR, rr 四种基因型出现的概率各为 25%，试问在一次选择中，正好选到 Rr 或 rR 杂合基因型的概率有多少？

设 A 为一次选到 Rr 的事件，B 为一次选到 rR 的事件，在一次选择中，选 A 时，即不会选到 B，反之亦然，A 和 B 为互斥事件。因此，所求的概率为

$$\begin{aligned}p(A\cup B)&=p(A)+p(B)\\&=25\%+25\%=50\%\end{aligned}$$

若 A 和 B 为互补事件，记为 A 和 A'，则有

$$p(A') = 1 - p(A)$$

因为 A 和 A' 为互补事件，所以 $A \cup A' = S$，另外 A 和 A' 不能相交，所以

$$p(A) + p(A') = 1$$

$$p(A') = 1 - p(A)$$

例如，某一性状为一对等位基因所控制时，在完全显性的情况下，隐性纯合基因型出现的概率为 25%，那么其它三种基因型出现的概率为

$$p(A') = 1 - p(A) = 1 - 25\% = 75\%$$

现在举一个适当的例子来说明公式

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

一副扑克牌共 52 张(除大、小王)，黑桃、梅花等各 13 张。所以摸一张牌正好是黑桃的概率应为 13/52，而正好摸到老 K 的概率（不论黑牌或是红牌）应该是 4/52。试问如果摸一次牌，正好是黑桃或者是老 K 的概率有多少？

设 A 为摸到黑桃的事件，B 为摸到老 K 的事件。有这样一种可能，即摸到一张牌时它既是黑桃，又是老 K，即事件 A 与 B 是相交的。可以想到，这种可能性为 1/52，即

$$p(A \cap B) = 1/52$$

所以
$$p(A \cup B) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52}$$

在样本空间 S 中，如果事件 A(A子集）的发生不受事件 B 的发生的影响，则称 A、B两个事件是独立的。

请注意独立事件与互斥事件是不同的。

若 A 与 B 为二事件，$A \cap B = \phi$，则称 A 与 B 为互斥事件，即一个事件发生时，另一个事件必不能发生。

若

$$A \cap B \neq \phi$$

而且

$$p(A \cap B) = p(A) \cdot p(B)$$

则称 A、B 两事件独立。这个公式表明的是概率的乘法法则。

例如，投掷一枚硬币，连续三次出现正面的概率是多少？

设三次投掷硬币，每次的事件依次为 A, B, C，各次事件的概率 $p(A), p(B), p(C)$ 皆为 $1/2$，每个事件发生时，都与另两个事件无关，它们相互是独立的，因此

$$p(A \cap B \cap C) = p(A) \cdot p(B) \cdot p(C)$$

$$= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

又例如，在一种二倍体农作物中，有一个品种是早熟、抗病和高秆的，现在与一个矮秆、晚熟和感病的品种进行杂交，想选出一个早熟、抗病、矮秆的类型。假设控制这三个性状的主效基因各有一个，而且位于三条不时的染色体上。这样，在减数分裂时，三个基因的分配是独立的，基因之间的相互结合是随机的。如以 E/e 表示早熟性的等位基因，R/r 表示抗病性的等位基因，H/h 表示高、矮秆等位基因，它们出现在父母本配子中的机会是相等的，一个配子中出现 E 或是 e，出现 R 或是 r，出现 H 或是 h 的概率都是 $1/2$。那

么，E（早熟）R（抗病）h（矮秆）三个基因出现在同一个配子中的概率即为

$$p(E\cap R\cap h)=\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{8}$$

其它配子的概率也都是1/8。这些配子的类型包括 *ERH*, *ERh*, *ErH*, *Erh*, *erh*, *erH*, *eRh*, *eRH*, 一共8种。父本有8种，母本也有8种。父母本的配子进行随机交配时，会出现13页的情况。

我们所要选择的基因型位于上表的二行二列的小方格内，它在杂交第二代的群体中只占到$\frac{1}{8}\times\frac{1}{8}=\frac{1}{64}$，即

$$p(ERh\cap ERh)=\frac{1}{8}\times\frac{1}{8}=\frac{1}{64}$$

但是，在 F_2 群体中，当我们用肉眼选择时，有一些杂合基因型与纯合基因型从表型上是无法分辨的。所以入选的材料，可能包含有 *EErRhh*、*EeRRhh*、*EeRrhh* 等杂合基因型。只有继续自交，从分离群体中才能选出比较稳定的纯合基因型。