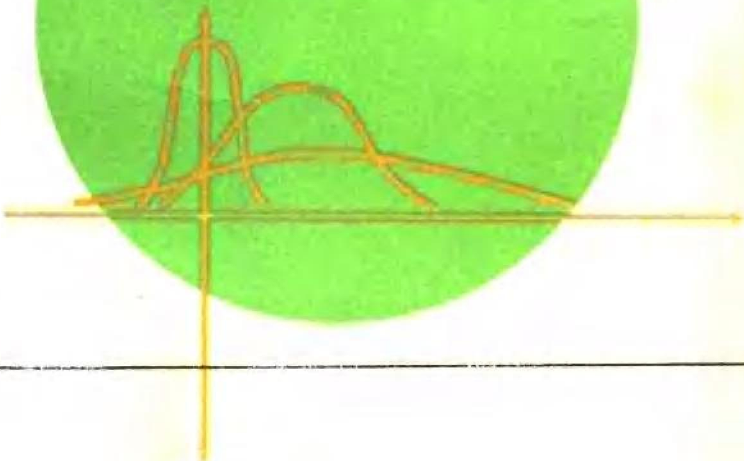
生物统计



钟平安编著

湖南科学技术出版社

农业数理统计丛书

生 物 统 计

钟 平 安 编著

湖南科学技术出版社

生 物 统 计
钟 平 安 编 著
责任编辑：贺晓兴
*
湖南科学技术出版社出版
（长沙市展览馆路14号）
湖南省新华书店发行 湖南省新华印刷二厂印刷
*
1983年11月第1版第1次印刷
开本：87×1092毫米 1/32 印张：19.75 字数：456,000
印数：1—4,200
统一书号：16204·126 定价：2.35元

目　　录

第一章　基础知识

第一节　概率运算

一、事件及其概率

生物统计是概率论与数理统计应用于生物科学的产物。

生物现象很多都是“随机”的。例如玉米种子播种以后，每粒种子可能发芽，也可能不发芽。杂种的紫花豌豆，它的后代可能是紫花，也可能不是紫花。我们把那些具有多种可能发生的结果，而对于究竟发生哪一种结果，事先不能肯定的现象，称为随机现象。

对于这些随机现象进行个别观察时，虽然不能看出它是否一定发生。但是，当我们进行了多次观察后，会发现它们是存在着一定的规律性的。例如玉米种子，对其中一粒来讲，播种以后可能发芽，也可能不发芽。但对一批种子来讲，其发芽率是有一定规律的。概率论与数理统计就是从数量上研究大量随机现象规律性的科学。

随机现象的每一种表现或结果，叫作一个随机事件。某田块每亩中粘虫幼虫的条数是一个随机现象。“幼虫在600条以上”是一个事件。“幼虫在3,000条到4,000条之间”也是一个事件。通常就用大写拉丁字母A，B，C，……表示事件。

有些现象，只有一种可能的结果（一个事件）。例如，没有水分，种子肯定不会发芽。纯种的紫花豌豆，它的后代一定是紫花。这些都是确定性现象。

在一定条件下，必然会发生的事件，叫必然事件。一般记作Ω。在一定条件下，必然不会发生的事件，叫不可能事件。一般记作ϕ。必然事件和不可能事件实质上都是确定性现象的表现。但是，把它们看作是随机事件的两种特例，对于分析问题是有利的。

考察某品种小麦的发芽情况时，分别抽取5粒、10粒、50粒、100粒、300粒、600粒种子，在相同的条件下，进行观察。得到表1.1的统计结果：

表1.1　　小麦种子发芽情况表

种子总粒数 n	5	10	50	100	300	600
种子发芽数 m	5	8	44	91	272	542
种子发芽频率 $\frac{m}{n}$	1	0.8	0.88	0.91	0.907	0.903

在表1.1中，种子总粒数是n，发芽种子数为m，我们把$\frac{m}{n}$叫作n粒种子发芽的频率。显然，随着n的增大，$\frac{m}{n}$愈来愈稳定地在0.9左右摆动。我们自然会认为：用0.9这个数值去表示抽取一粒种子时抽到能发芽种子的可能性大小是适当的。这种描述随机事件在试验结果中出现的可能性大小的数值，叫作概率。

在多次重复进行同一试验时，随机事件A发生的频率$\frac{m}{n}$所稳定接近的值P，叫作随机事件A的概率。记作

$$P(A)=P\approx\frac{m}{n}$$

这里我们要注意，某事件发生的概率是$\frac{1}{10}$时，并不意味着

作10次试验，这一事件准有一次发生。而是意味着：如果耐心地就试验作几十次、几百次……，则这事件发生的次数愈接近试验次数的$\frac{1}{10}$。如重复试验了9次，这事件都没有出现，就认为第10次试验时，这事件一定会出现，那就错了。因为第10次试验这事件出现的概率和前9次试验的成败毫无关系，其概率仍为$\frac{1}{10}$。

对于这种定义，可能有人会感到遗憾。认为这个P值，通常只是一个未知常数，而我们只可能求得其近似值，这似乎总是一个缺点。但事实上，即使我们直接用度量衡的工具或仪器对于一定对象的某些数量进行度量时，所得的数值，也都只是该数量的客观大小的近似值而已。这是由于我们所使用的度量衡工具或仪器的精确度，总是有一定限度的缘故。这完全不影响我们所得到的数值的实际意义。我们可以用增加度量衡工具或仪器的精确度的方法，使所得度量结果的数值，满足我们所需要的准确程度。我们也同样可以用增加试验次数n的方法，使所得到的频率数值满足我们所需要的准确程度。

从概率的定义出发，概率有以下三个基本性质：

1.任何事件的概率都在0与1之间，即

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

2.必然事件的概率等于1，即

$$P(\Omega)=1$$

3.不可能事件的概率等于0，即

$$P(\phi)=0$$

二、互不相容事件的概率加法定理

如果事件A与事件B不可能在同一试验结果中都发生，则称

事件A与事件B是两个互不相容事件。例如 A：种子发芽数等于6，B：种子发芽数小于6，是两个互不相容事件。因为种子发芽数等于6的A事件出现了，则种子发芽数小于6的B事件不可能也在结果中出现。

如果事件A与事件B互不相容，则事件A或事件B出现其一（记作$A+B$，叫A与B的和事件）的概率，等于该两事件的概率之和，即

$$P(A+B)=P(A)+P(B)$$

事实上，设在n次试验中，事件A出现m_1次，事件B出现m_2次，由于事件A与事件B互不相容，所以在事件A出现的m_1次中，事件B不会出现。同时在事件B出现的m_2次中，事件A也不出现。因此在n次试验中，事件A或事件B出现的次数共为m_1+m_2次，故知事件A或事件B出现其一的频率为：

$$\frac{m_1+m_2}{n}=\frac{m_1}{n}+\frac{m_2}{n}$$

根据事件的频率稳定接近概率这一特点，显然$A+B$、A、B的频率的稳定值——概率$P(A+B)$、$P(A)$、$P(B)$就存在下列关系：

$$P(A+B)=P(A)+P(B)$$

推论：如果事件A_1，A_2，…，A_n等n个事件，两两互不相容，则

$$P(A_1+A_2+\cdots+A_n)=P(A_1)+P(A_2)+\cdots+P(A_n)$$

运用这条定理，必须充分注意所给条件。

例1.1　一批安江香柚，其中有7%是二等品，3%是三等品，其余都是一等品，求出现非一等品的概率。

设　“抽取一只香柚是二等品”为事件A，“抽取一只香柚

是三等品”为事件B，则“抽取一只香柚是非一等品”，就是事件$A+B$。由于事件A和事件B是互不相容事件，所以

$$P(A+B)=P(A)+P(B)=7\%+3\%=10\%$$

例1.2 张三打下麻雀的概率是80%，李四打下麻雀的概率是70%，两人同时开枪，打下麻雀的概率是多少？

设 张三打下麻雀为事件A，李四打下麻雀为事件B，则“打下麻雀”是事件$A+B$，由于事件A和事件B没有互不相容的条件（可以由两人同时打下），因此不能运用上述定理，否则会出现：$P(A+B)=P(A)+P(B)=80\%+70\%=1.5$ 得到打下麻雀的概率是大于1的错误结论。

我们把A不出现的事件，称为A的逆事件，记作$\bar{A}$，于是可以得到

$$P(A)+P(\bar{A})=1$$

这是由于事件A或者出现，或者不出现，二者必居其一。也就是说：事件A与事件$\bar{A}$不可能同时出现，也不可能都不出现。根据A与$\bar{A}$不可能同时出现，所以A与$\bar{A}$是互不相容事件，可以得到

$$P(A+\bar{A})=P(A)+P(\bar{A})$$

又根据A与$\bar{A}$不可能都不出现，亦即$(A+\bar{A})$是必然事件，可以得到

$$P(A+\bar{A})=1$$

因此 $$P(A)+P(\bar{A})=1$$

这一性质，对于我们以后的概率计算，提供了很大的方便。例如事件A的概率，很难直接计算，而其逆事件$\bar{A}$的概率计算却很容易，这时，可先求$P(A)$，然后再由

$$P(A)+P(\bar{A})=1$$

求出 $$P(A)=1-P(\bar{A})$$

推广到一般情况，若A_1、A_2、$\cdots A_n$，是一个互不相容事件组，则有

$$P(A_1+A_2+\cdots+A_n)=P(A_1)+P(A_2)+\cdots+P(A_n)$$

当 $A_1+A_2+\cdots+A_n$构成必然事件时，则

$$P(A_1)+P(A_2)+\cdots+P(A_n)=1$$

满足这一条件的诸事件A_1、A_2、$\cdots A_n$，叫完全事件系。

三、独立事件的概率乘法定理

如果事件B的发生或不发生不影响事件A发生的可能性，则称事件A对于事件B是独立的。反之，如果事件B的发生，影响到事件A的发生，则称事件A对于事件B是相依的。

假使有10粒种子，其中有两粒不能发芽，其余8粒能发芽。如果我们由其中抽出一粒种子后仍将该种子放回，然后进行第二次抽取，则（B）第一次抽中不发芽种子（概率是2/10）的事件，与(A)第二次抽出不发芽种子(概率是2/10)的事件，是互相独立的两个事件。如果第一次抽出一粒不发芽种子后，不将该粒种子放回，则（B）第一次抽出不发芽种子（概率是2/10），与（A）第二次抽出不发芽种子（概率是1/9），不是相互独立的两个事件，而是相依的两个事件。

如果事件A与事件B相互独立，则事件A与事件B同时出现（记作$A\cdot B$，叫做A与B的积事件）的概率等于该两事件的概率之积，即

$$P(A\cdot B)=P(A)\cdot P(B)$$

事实上，设事件A在n_1次试验中出现了m_1次，事件B在n_2次试验中出现了m_2次，当n_2与n_1都充分大时，则

$$P(A)\approx\frac{m_1}{n_1}\qquad P(B)\approx\frac{m_2}{n_2}$$

由于事件A与事件B相互独立，在两项试验同时进行n次，事件A出现的次数为$n\frac{m_1}{n_1}$次时，则事件$A\cdot B$只能在A出现的$n\frac{m_1}{n_1}$次中出现其$\frac{m_2}{n_2}$次，即$n\frac{m_1}{n_1}\cdot\frac{m_2}{n_2}$次。因此，在$n$次试验中，事件$A$与事件$B$同时出现的次数有$n\frac{m_1}{n_1}\cdot\frac{m_2}{n_2}$次。根据概率的意义，事件$A$与事件$B$同时出现的概率是：

$$P(A\cdot B)\approx\frac{n\cdot\frac{m_1}{n_1}\cdot\frac{m_2}{n_2}}{n}=\frac{m_1}{n_1}\cdot\frac{m_2}{n_2}\approx P(A)\cdot P(B)$$

推论：如果事件A_1，A_2，…，A_n相互独立，则

$$P(A_1\cdot A_2\cdots A_n)=P(A_1)\cdot P(A_2)\cdots P(A_n)$$

如果$P(A\cdot B)=P(A)\cdot P(B)$，则$A$与$B$相互独立。

独立事件的概率乘法定理与互不相容事件的概率加法定理，是概率运算中的两条最基本的定理。下面举例说明这些定理的用法。

例1.3　如果一批种子播下后成为壮苗的概率是0.6，依次播下三粒这样的种子，求只有第一粒成为壮苗的概率？

设：(A)第一粒成为壮苗、(B)第二粒不成为壮苗、(C)第三粒不成为壮苗，只有第一粒成为壮苗的事件可为积事件A、B、C，由于A、B、C相互独立，有

$$P(A\cdot B\cdot C)=P(A)\cdot P(B)\cdot P(C)$$

而　$P(A)=0.6$　$P(B)=1-P(\bar{B})=1-0.6=0.4$

$P(C)=1-P(\bar{C})=0.4$

所以　$P(A\cdot B\cdot C)=0.6\times0.4\times0.4=0.096$

例1.4　对于上面的问题，如果是问，有一粒成为壮苗的概

率是多少？

则设(A_1)只有第一粒成为壮苗、(A_2)只有第二粒成为壮苗、(A_3)只有第三粒成为壮苗。有一粒成为壮苗的事件为和事件$A_1+A_2+A_3$。由于A_1、A_2、A_3互不相容，有

$$P(A_1+A_2+A_3)=P(A_1)+P(A_2)+P(A_3)$$

由例1.3知：$P(A_1)=0.096$　同理$P(A_2)=0.096$

$$P(A_3)=0.096$$

所以　$P(A_1+A_2+A_3)=0.096\times 3=0.288$

例1.5　如果用这批种子每穴播三粒，待出苗后再间苗，每穴留一株较强壮的苗，求每穴有壮苗的概率？

设　每穴有一粒成为壮苗、每穴有两粒成为壮苗，每穴有三粒成为壮苗的事件分别记为A_1，A_2，A_3。每穴有壮苗的事件为和事件$A_1+A_2+A_3$。由于A_1、A_2、A_3互不相容，有

$$P(A_1+A_2+A_3)=P(A_1)+P(A_2)+P(A_3)$$

由例1.4知　$P(A_1)=0.288$

显见　A_3每穴有三粒成为壮苗的情况是：第一粒成壮苗、第二粒成壮苗，第三粒成壮苗，这是同时出现的事件。由于这三个事件相互独立，且每粒成为壮苗的概率是0.6，所以

$$P(A_3)=0.6\times 0.6\times 0.6=0.216$$

A_2：每穴有两粒成为壮苗的情况可分为：第一、二粒成为壮苗第三粒不成壮苗，第一、三粒成为壮苗第二粒不成壮苗，第二、三粒成为壮苗第一粒不成壮苗，且这些事件互不相容。每一事件出现的概率是：$0.6\times 0.6\times 0.4=0.144$

从而　$P(A_2)=0.144\times 3=0.432$

故　$P(A_1+A_2+A_3)=P(A_1)+P(A_2)+P(A_3)$

$$=0.288+0.432+0.216=0.936$$

对这个问题，比较简单的解法是：先计算每穴没有壮苗的

概率。每穴没有壮苗的事件是:“第一粒不是壮苗”、“第二粒不是壮苗”，“第三粒不是壮苗”,这是三个同时出现的事件。这三个事件相互独立，每穴没有壮苗的事件为这三个事件之积。并且，不是壮苗的概率为 $1-0.6=0.4$， 所以每穴没有壮苗的概率是：

$$0.4\times0.4\times0.4=0.064$$

每穴有壮苗是每穴没有壮苗的逆事件。所以，每穴有壮苗的概率是：

$$1-0.064=0.936$$

与前面计算的结果完全相同。

第二节 随机变量

一、随机变量的概率分布

随机事件可以用量来描述。例如：播5粒种子，其中有ξ粒发芽这一事件，就可用量ξ来表示：

“$\xi=0$”，表示事件“没有一粒发芽”；

“$\xi=1$”，表示事件“有一粒发芽”；

“$\xi=2$”，表示事件“有二粒发芽”；

“$\xi\leqslant3$”，表示事件“有不多于三粒发芽”；

……

又如某品种玉米穗位的高低，用量η表示。$120\leqslant\eta<130$表示“玉米穗位在120到130厘米之间”这一事件。

注意：$120\leqslant\eta<130$是一般的写法，因为习惯上的分组数列，总是包括左端点。

因为事件是随机的，用来表示随机事件的量只能是变量，但事件的概率又是确定的，所以这个量取各种值的概率也是确定的。

以确定的概率取不同值的变量，叫随机变量。

引入了随机变量，就能够充分利用数学工具来研究随机现象。

如果随机变量可能的取值，能按次序一一列举（有限个或可数个），这样的随机变量称为离散型随机变量。如发芽种子量就是离散型随机变量。

有些随机变量的取值，充满了一个区间，无法按次序一一列举，如穗位的高低，亩产量的大小等等，这些叫连续型随机变量。

为了全面掌握一个离散型随机变量ξ的统计规律，就必须知道：

（1）ξ的所有可能取值：x_1，x_2，…，x_n…是什么？

（2）ξ取每一可能值的概率是多少？

即要知道

$P(\xi=x_1)$，$P(\xi=x_2)$，…$P(\xi=x_n)$，…

我们常把上面两点写成：

ξ	x_1	x_2	…	x_n
$p(\xi)$	$p(\xi=x_1)$	$p(\xi=x_2)$	…	$p(\xi=x_n)$

这称为离散型随机变量ξ的分布列。掌握了分布列，我们就掌握了离散型随机变量ξ的概率分布规律。

如果记

$P(\xi=x_1)=P_1$，$P(\xi=x_2)=P_2$，…，$P(\xi=x_n)=P_n$，…

分布列就可以简写成

$$\begin{pmatrix} x_1, & x_2, & \cdots, & x_n\cdots \\ p_1, & p_2, & \cdots, & p_n\cdots \end{pmatrix}$$

或者合并写成：

$$P(\xi=x_i)=p_i \quad (i=0,1,2,3,\cdots,n,\cdots)$$

分布列的这几种表达方式，形式虽然不同，实质是一样的。

根据概率的基本含义，不难得出，任一概率分布，都应有以下两条性质：

(1) 随机变量取任何值时其概率都不会是负数，即

$$p_1\geqslant 0,\ p_2\geqslant 0,\ \cdots p_n\geqslant 0,\ \cdots$$

(2) 随机变量取遍所有可取的值时，相应的概率之和等于1,即

$$p_1+p_2+\cdots+p_n+\cdots=1$$

连续型随机变量的可能取值是充满区间的。因此不能象离散型随机变量那样用分布列来描述其概率分布。

对于连续型随机变量,我们所关心的是ξ落在某一范围内的概率，即：$p(x_1\leqslant\xi<x_2)$

通常我们用概率分布密度函数$f(x)$（简称分布密度）来描述连续型随机变量的概率分布。

分布密度$f(x)$是一条平面曲线。事件ξ取值落在区间$[x_1、x_2)$内的概率$p(x_1\leqslant\xi<x_2)$就是由分布密度曲线$f(x)$下、区间$[x_1、x_2)$上所形成的曲边梯形的面积，如图1.1。

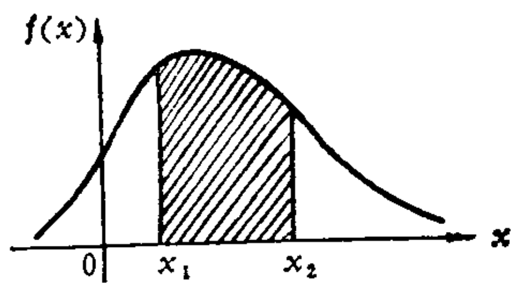


图1.1　分布密度函数$f(x)$

用数学式表示，就是

$$p(x_1\leqslant\xi<x_2)=\int_{x_1}^{x_2}f(x)dx=\text{阴影部分的面积}$$

由于连续型随机变量ξ所取的值为一个范围或整个实数，且所有可能取值相应的概率之和等于1，所以曲线与x轴所夹的面积等于1。