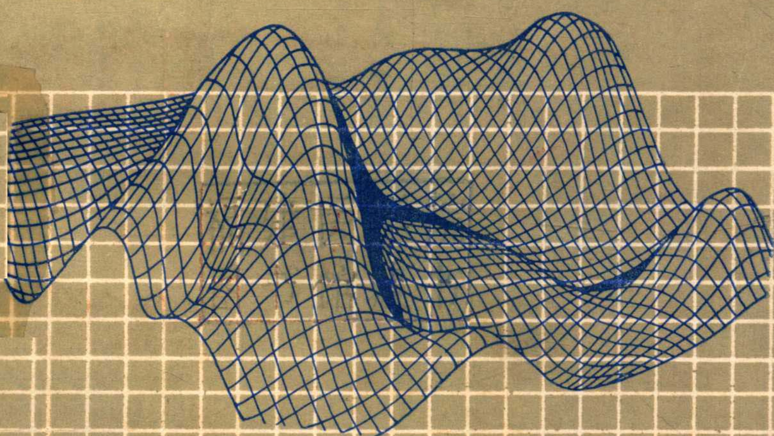
统计方法
在刑事技术中的应用

李从珠



群众出版社

统计方法
在刑事技术中的应用

李从珠

（公安机关 内部发行）

群众出版社
一九八六年·北京

统计方法
在刑事技术中的应用
李从珠

群众出版社出版发行
通县振兴印刷厂印刷

787×1092毫米 32开本 8.625印张 插4页 180千字
1986年6月第1版 1986年6月北京第1次印刷

统一书号：13067·83 定价：1.95元
（公安机关 内部发行）

前言

数理统计在国民经济的许多方面得到越来越广泛地应用，这是众所周知的。但是在刑事技术方面的应用，在我国还是一个新领域。作者自1977年以来同周口地区公安处协作，在公安部126所的领导与支持下，试用数理统计的方法不仅在案例分析中取得了较好的效果，而且在现场破案中有了新的突破，为刑事技术提供了一个新工具。为使广大公安干警掌握这个新工具，在刑事技术中更广泛地应用统计方法，本人编写了这本小册子。

本书共分两部分，第一部分主要介绍概率论中的一些基本知识，为掌握统计方法提供必要的理论基础。第二部分主要介绍在刑事技术中已经得到应用的一些统计方法。所涉及的实例数据主要是周口地区公安处及126所提供的。

在本书的编写过程中，得到126所刘云起副所长，国际政治学院陈树林副院长，中国科学院系统科学研究所张里千研究员，河南省公安厅技术科王乃斌科长和周口地区公安处王清举副处长等的鼓励与支持，特别是中国科技大学陈希孺教授仔细校阅了这本小册子。对此，作者表示衷心的感谢。

由于水平所限，书中一定存在着不少缺点和错误，敬请读者批评指正。

作　者

1982年9月1日

序

虽然概率论和数理统计的方法在人类活动的各个领域中都得了程度不等的应用，它在刑事侦查技术上发挥作用，相对说来还是比较新鲜的事情，这种应用所以可能，是因为衡量人体各部分的特征指标，以及人在特定环境下活动所留下的痕迹，都是有联系的。随机变量或带有很大的随机性成分，用数理统计的方法对现场数据进行科学分析，可以在一定程度上排除随机因素的干扰，提取出有用的信息，这对于缩小侦查目标，检证由别的途径获得的证据以至提出一定的设想，都是有意义的。当然，统计方法只是一种归纳式的分析方法，它只能减轻以及正确估量偶然因素的影响而不能完全消除这种影响。不言而喻，用统计方法分析得出的结果只能作为参考而不能据以定案。

李从珠同志这本书以简练的笔法和不大的篇幅，介绍了概率论和数理统计的基本概念和常用方法，值得提出的是，本书在介绍这些方法时，结合了一些现实的而非虚构的例子，这不仅有助于使读者更清楚地理解方法，而且能使他对这些方法产生一种现实感，这对于真正掌握统计方法的精神是很重要的。

本书不仅适合于有关的专业干部阅读，就是对于数理统计方法感兴趣的一般读者，也能从其中找到一些有益的东西。

陈希孺

1982年9月8日

于烟台

目　录

第一部分　统计方法的基础知识

第二部分　统计方法在刑事技术中的应用

第一部分　统计方法的基础知识

第一章　概率论中的一些基本概念

当我们观察自然界和社会上发生的多种现象时，会发现，有一类现象，在一定的条件下，必然发生。如："太阳从东边出来。""在阶级社会中，刑事案件一定会有。""同性电荷互相排斥。"等等。这类现象称为必然现象。另一类现象是，在一定条件下一定不发生，如："太阳从西边出来。""异性电荷互相排斥。"等等，这类现象称为不可能现象。以上两类现象形式上虽然不同，但却有其共同点，那就是事先对其发生与否可以做出肯定的回答，所以这两类现象统称为确定性现象。然而在自然界和社会上还存在另一类与确定性现象有着本质区别的现象，它在一定的条件下可能发生也可能不发生，例如，从一群人中随便找出一人，事先无法知道他手指中斗型纹的个数，事先也无法知道他的血型；对一孕妇，事先也无法知道她生男还是生女等等。这类现象每次试验(或观测)的结果，都随"机会"而变，我们称这样的现象为随机现象。

人们长期的实践和研究发现，随机现象虽然每次试验的结果具有不确定性，但大量的重复试验和观测就会发现，它呈现某种规律性。如，一孕妇生男生女虽事先难于预料，但大量观察就会发现男女大体各半。又如，任找一个中国人，事先虽无法确知他斗型纹的手指个数，但大量观察表明，中

国人斗型纹约占50%，等等。这种在大量重复试验和观测中事物所呈现出的固有规律性，我们称之为统计规律性。概率论和数理统计正是研究随机现象统计规律的一门数学分支。

§1.1 概率的统计定义

在一定的条件下，可能发生也可能不发生的事件叫做随机事件；在一定条件下必然发生的事件叫做必然事件；在一定条件下必然不发生的事件叫做不可能事件。以后我们分别以字母Ω和Φ表示必然事件和不可能事件；以字母A、B、C、…，表示随机事件。

我们以抛掷一分币为例，说明概率的统计定义。首先对"在一定条件下"这句话做点解释，在抛分币的例中是指：分币是匀称的，将分币放在手心上用一定的力量上抛，让分币自由落在具有弹性的地面上等等。以A记"正面朝上"这一事件，显然，在上述条件实现一次时，事件A是否发生是不确定的，然而，大量重复试验抛掷分币时，事件A发生的次数(称为频数)，就体现出一定的规律性，即"正面朝上"出现的次数约占试验次数的一半。写成式子为：

$$\text{事件A发生的频率}=\frac{\text{频　数}}{\text{试验总次数}}\approx\frac{1}{2}$$

历史上许多人作过这个试验，下面列出他们中一些试验的记录。

从表1.1.1可以看出，抛掷次数越多，频率越接近0.5。

定义 1.1 在一定的条件下，重复作几次试验，μ为n次试验中事件A发生的次数，如果随着n逐渐增大，频率μ/n稳定地在某一数值P附近摆动，且摆动的幅度越来越小，则

表 1.1.1

试验者	试验次数 n	A 出现的次数 μ	频率 $=\frac{\mu}{n}$
德摩尔根	2048	1061	0.5181
蒲丰	4040	2048	0.5069
皮尔逊	12000	6019	0.5016
皮尔逊	24000	12012	0.5005
维尼	30000	14994	0.4998

称数值 P 为随机事件 A 在该条件下发生的概率，记作 $P(A)=P$。

这个定义称为概率的统计定义。

从定义中可以看到，数值 P 就是在该条件下刻画事件 A 发生可能性大小的一个数量指标。

由于频率 μ/n 总是介于 0 和 1 之间，从定义可知，对任意事件 A，皆有

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

$$P(\Omega)=1$$

$$P(\Phi)=0$$

在刑事技术中，许多情况下，都是采用概率的统计定义。如公安部 1953—1955 年，组织调查中国人手指纹型分布状况。得出斗型纹约占 50%，这个 50% 可以认为是斗型纹出现的概率，在刑事技术中这种用频率近似代替概率的情况是经常出现的。

§ 1.2 概率的古典定义

概率统计定义的优点是直观明了，但用来计算事件的概

率却非常不方便，即便做了大量试验，也只能得出概率的近似值。而概率的古典定义在应用中计算就比较方便。

在介绍概率的古典定义之前，先介绍点排列组合知识。

1.排列

例1.2.1　问1、2、3、4能组成多少个无重复的两位数、三位数、四位数？并排出来。

解：1、2、3、4共能组成以下 $4\times3=12$ 个无重复数字出现的两位数，它们是：

12	21	31	41
13	23	32	42
14	24	34	43

这是因为，十位数有四种可能性，而当十位数定了之后，个位数只有三种可能性。

同理，1、2、3、4共能组成以下

$$4\times3\times2=24$$

种无重复的三位数，它们是：

$$\begin{cases}123\\124\end{cases}\quad\begin{cases}213\\214\end{cases}\quad\begin{cases}312\\314\end{cases}\quad\begin{cases}412\\413\end{cases}$$

$$\begin{cases}132\\134\end{cases}\quad\begin{cases}231\\234\end{cases}\quad\begin{cases}321\\324\end{cases}\quad\begin{cases}421\\423\end{cases}$$

$$\begin{cases}142\\143\end{cases}\quad\begin{cases}241\\243\end{cases}\quad\begin{cases}341\\342\end{cases}\quad\begin{cases}431\\432\end{cases}$$

不难看出，组成的无重复的四位数是

$4\times3\times2\times1=4!=24$ 种。

由上例的讨论可以看出，1、2、3、4、5共能组成 5×4 个无重复的二位数；$5\times4\times3$ 个无重复的三位数；$5\times4\times3\times2$

个无重复的四位数；5!个无重复的五位数。

定义 1.2.1　将 n 个元素按一定的次序排列叫做 n 个元素的一个全排列。

从 n 个元素中取出 m 个（$m<n$，不允许重复）按一定的次序排列，叫做从 n 个元素取 m 个元素的选排列（简称排列）。

n 个元素的全排列总数为 n!

n 中选 m 的选排列总数记为 A_n^m，则

$$A_n^m=n(n-1)(n-2)\cdots(n-m+1)$$

例1.2.2　假定北京和天津之间有 8 个火车客运站，问要印多少种不同的火车票？

解：包括北京、天津在内共有 10 个站，任意两个站如甲和乙，甲到乙是一种车票，乙到甲是另一种不同的火车票，所以共要印

$$A_{10}^2=10\times9=90$$

种不同的火车票。

例1.2.3　问 1、2、3、4 能组成多少个二位数、三位数？（每个数码允许重复取）

解：此题与例1.2.1不同，例1.2.1中，1、2、3、4 四个数不允许重复取，而这里允许重复，所以对两位数来说，十位数有四种可能的取法，个位数也同样有四种取法，所以共能组成

$$4\times4=4^2$$

种两位数，

同理，可以组成

$$4\times4\times4=4^3$$

种不同的三位数，

可以组成

$$4\times4\times4\times4=4^4$$

种不同的四位数。

一般地有

定义 1.2.2 从 n 个元素取出 m 个(允许重复)元素的排列总数为 n^m 种。

例1.2.4 某城用六位电话号码，而首位不允许用0，问该城最多能安多少台“直拨”电话机。

解：首位不允许用0，它只能是1至9这九个数字中的一个，第二位可以是0、1、…、9十个数字中的任意一个，同样，第三位、四位、五位、六位都有十种可能，故该城最多能安装的“直拨”电话台数是

$$9\times10\times10\times10\times10\times10=900000$$

2组合

从例1.2.2知，北京至天津间的火车票共有90种，但却只有 90/2＝45 种票价，这是因为“北京至天津”和“天津至北京”车票不同，但票价却是一样的。这实际上是可以看成从10个元素(车站)中选取二个元素，有几种选法的问题。这称为从10个元素中选取二个元素的组合问题。

定义 1.2.3 从 n 个元素中取出 m 个元素，不管顺序并成一组，叫做从 n 个元素取 m 个元素的组合。其组合总数记为 C_n^m。

排列与次序有关，如 ab 和 ba 是不同的排列，但却是同样的组合。

再如从30个人中选取3个人照像，问共有多少种照

法？因照像与次序有关，同样的三个人，站的(或坐的)次序不同，其对应的照法也不同，故这是排列问题，所以共有

$$A_{30}^{3}=30\times29\times28=24360\text{ 种照法。}$$

若从 30 人中选 3 人组成一个劳动队，问有多少种组织法？这与照像不同，它与次序无关，只要三个人是一样的，就是一个劳动队，故共有

$$C_{30}^{3}=A_{30}^{3}/3!=8120\text{种组织方法。}$$

这就是说，排列数是组合数所取出元素的阶乘倍，所以一般的有

$$C_n^m=A_n^m/m!=\frac{n!}{m!(n-m)!}$$

可以证明：

$$C_n^m=C_n^{n-m},$$

$$C_n^m+C_n^{m-1}=C_{n+1}^m,$$

(这两个公式直接由组合的表达式立即可得。)

$$2^n=\sum_{m=0}^{n}C_n^m$$

($\sum$表示相加的意思，$\sum\limits_{m=0}^{n}C_n^m=C_n^0+C_n^1+\cdots C_n^n$)

例1.2.5 八男五女组成的劳教队，选三个组长

(1) 必须是一女二男，共有多少种选法？

(2) 必须是二女一男，共有多少种选法？

(3) 至少有一女，共有多少种选法？

解：(1) 必须是一女二男，这一女是从五个女的中间选出的，故有C_5^1种选法，二男是从八个男的中间选出的，故有

C_8^2种选法，而女的任意一种选法都对应着男的C_8^2种，所以共有

$$C_5^1 \cdot C_8^2 = 5 \times 28 = 140$$

种选法，同理有

(2) $C_5^2 \cdot C_8^1 = 80$

(3) "至少一女"，这句话是指正好一个女的，或正好两个女的，或三个全是女的。故共有

$$C_5^1 \cdot C_8^2 + C_5^2 \cdot C_8^1 + C_5^3 \cdot C_8^0 = 140 + 80 + 10 = 230$$

种选法。

或这样解：三个组长全是男的选法有

$$C_5^0 \cdot C_8^3 = 56$$

而三个组长从 13 个人中任选的选法有

$$C_{13}^3 = 286$$

所以至少一女的选法有

$$C_{13}^3 - C_5^0 \cdot C_8^3 = 286 - 56 = 230$$

两种解法结果是一样的。但后者较简便。

例 1.2.6 有 10 人的会议，开会前两两互相握手并互相赠送照片一张，问整个会议上的人共握了几次手？共送了几张照片？读者自己算算看？

(答案为：$C_{10}^2 = 45$，$A_{10}^2 = 90$)

有了排列和组合的知识，回过来我们介绍概率的古典定义。

概率的古典定义是根据问题本身所具有的某种"对等性"，分析事物的本质，来直接计算概率的。

如果一个试验满足两条

1° 所有可能的试验结果只有有限个。

2° 每一个试验结果出现的可能性是一样的。

这样的试验，称为古典试验。

对于古典试验中的随机事件 A，它的概率定义为：

$$P(A)=\frac{m}{N}$$

N 表示，该试验中所有可能出现结果的总数目。m 表示事件 A 包含的试验结果数。

这种定义概率的方法称为概率的古典定义。

依此定义知，掷一均匀钱币，

$$P(\text{正面朝上})=\frac{1}{2}$$

掷一均匀骰子，

$$P(\text{出现偶数点})=\frac{3}{6}=\frac{1}{2}$$

$$P(\text{出现的点数}>4)=\frac{2}{6}=\frac{1}{3}$$

为了熟悉概率的古典定义，我们再举几个例子。

例 1.2.7 号码锁有八个拨盘，每个拨盘上有 0 至 9 共 10个数字。当这八个拨盘上的数字组成某一个 8 位数（开锁号码）时，锁才能打开，如果不知道这个开锁号码，问一次把锁打开的概率是多少？

解：在这个号码锁上，总共可组成10^8种 8 位数字(这是允许重复的情况)。而一次打开锁，即是一次就拨到开锁号码，这只有一种可能。所以

$$P(\text{一次把锁打开})=\frac{1}{10^8}=0.00000001=10^{-8}$$

可见，不知道开锁号码，一次把锁打开几乎是不可能的。