应用统计技术

（修订版）

于善奇　编著

中国标准出版社

应用统计技术

（修订版）

于善奇　编著

中国标准出版社

应用统计技术

（修订版）

于善奇　编著

责任编辑　安　军　张尊生

*

中国标准出版社出版

北京复兴门外三里河北街 16 号

邮政编码：100045

电　　话：68522112

中国标准出版社秦皇岛印刷厂印刷

新华书店北京发行所发行　各地新华书店经售

*

开本 850×1168 1/32　印张 10.75　字数 260 千字

1999 年 3 月第一版　2001 年 5 月第二版

*

ISBN 7-5066-1857-5/Z・328

印数 3 001—6 000　定价 28.00 元

序　言

本书是数理统计学在工业生产与管理领域中的应用，要求读者具备高等数学和概率论的基础知识，这对于理解所提问题的实质和通晓所述方法的原理，无疑是必要的。对于不具备上述数学基础的管理人员，可以略去原理，掌握方法，弄清例题，就不难应用于实际。

数理统计及其应用作为一门课程，其内容十分丰富，本书主要介绍统计理论在产品质量分析中的应用；假设检验的原理与方法；质量控制的原理与方法；抽样检验的原理与方法；验收标准及其应用；正交试验设计与分析；相关与回归分析方法；统计预测方法。

本书目的之一，是为工科与管理类院校的师生提供选修教材。数理统计类书版本很多，但与应用密切结合者相对较少。本书目的之二，是为企业或管理部门的工程师们提供常用的管理技术，力求理论联系实际，提高管理水平。本书目的之三，是为贯彻 ISO 9000 系列标准及认证的企业，提供常用的统计技术。

本书在写作过程中，参考了若干中外著名学者的论著，在此表示衷心的感谢。

由于作者才疏学浅，纰漏之处在所难免，恳请读者批评指正。

编著者　于善奇

于北京工业大学管理楼教授室

2001.2.25 修稿

目　　录

第一章 统计技术基础

§1.1 随机变量及其数字特征

(一) 随机变量

1. 随机变量概念

众所周知,同种型号的显像管,其使用寿命是不同的。尽管生产显像管的工艺过程、原材料、设备、操作人员等条件基本相同,然而显像管的使用寿命也不可能完全相同。这说明,显像管的使用寿命是一个变量,这个变量的取值是不确定的,或者说是随机的,这种变量就叫做随机变量。

由于显像管寿命是以时间为单位进行计量的,它的取值可以是连续的。通常称这种可以连续取值的随机变量为连续型随机变量。如果随机变量的取值是离散的,通常称为离散型随机变量。例如,掷一枚骰子出现的点数;铸件表面的砂眼数;玻璃制品表面的气泡数等,这些变量都是离散型随机变量。

习惯上,把连续型随机变量的取值叫做计量值;把离散型随机变量的取值叫做计数值。计数值还可以进一步区分为计件值和计点值。例如,一批产品中的合格品数,是计件值;铸件表面的砂眼数,是计点值。

通常用英文大写字母 X,Y 等表示随机变量,用小写字母 x,y 等表示相应随机变量的取值。

2. 随机变量的概率分布

设 X 为离散型随机变量，可能的取值为 $x_1, x_2, x_3, \cdots$。如果

$$P(X = x_i) = p_i, i = 1,2,3\cdots \tag{1-1}$$

其中 p_i 已知，且 $\Sigma p_i = 1$，则称上式为随机变量 X 的概率分布。

例如，用 X 表示掷一枚骰子出现的点数，它可取 1,2,3,4,5,6 共 6 个自然数，并且取每个自然数的概率都等于 $\frac{1}{6}$。即 X 的概率分布为

$$P(X = k) = \frac{1}{6}, \quad k = 1,2,\cdots,6。$$

对于任一实数 x，我们考虑事件"$X \leqslant x$"的概率。如果 X 是离散型随机变量，则有

$$P(X \leqslant x) = \sum_{x_i \leqslant x} P(X = x_i) \tag{1-2}$$

不难看到，$P(X \leqslant x)$ 是 x 的函数，我们称它为随机变量 X 的分布函数，记作

$$F(x) = P(X \leqslant x)。 \tag{1-3}$$

如果 X 是连续型随机变量，则它的分布函数也可以用下 $P(X \leqslant x)$ 来刻画，即

$$F(x) = P(X \leqslant x) = \int_{-\infty}^{x} f(x)\mathrm{d}x, \tag{1-4}$$

其中右端的被积函数 $f(x)$ 称为连续型随机变量 X 的概率密度。

(1) 概率密度函数 $f(x)$ 的性质

① $f(x) = F'(x)$；

② $f(x) \geqslant 0$；

③ $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$。

(2) 分布函数 $F(x)$ 的性质

设 X 为随机变量，其分布函数为 $F(x)$，则

① $0 \leqslant F(x) \leqslant 1, F(-\infty) = 0, F(+\infty) = 1$；

② $F(x)$是x的不减函数；

③ $F(x)$至多有可列个间断点。

（二）期望

1．离散型随机变量的期望

设X为离散型随机变量，其概率分布为

$$P(X=x_i)=p_i,\quad i=1,2,3,\cdots。$$

如果级数$\sum x_i p_i$绝对收敛，则称$\sum x_i p_i$为随机变量X的期望，记作$E(X)$。即

$$E(X)=\sum_i x_i p_i。\tag{1-5}$$

2．连续型随机变量的期望

设X为连续型随机变量，其概率密度函数为$f(x)$。如果积分$\int_{-\infty}^{\infty} xf(x)\mathrm{d}x$绝对收敛，则称这个积分值为随机变量$X$的期望，记作$E(X)$。即

$$E(X)=\int_{-\infty}^{\infty} xf(x)\mathrm{d}x \tag{1-6}$$

3．期望的性质

设X,Y为随机变量，a为常数，则有

(1) $E(a)=a$；

(2) $E(a\cdot X)=a\cdot E(X)$；

(3) $E(X\pm Y)=E(X)\pm E(Y)$；

(4) 当X,Y相互独立时，$E(X\cdot Y)=E(X)\cdot E(Y)$。

（三）方差

随机变量离差平方的期望，称为随机变量的方差。随机变量X的方差记作$D(X)$，即

$$D(X)=E[X-E(X)]^2。\tag{1-7}$$

1．离散型随机变量的方差

设X为离散型随机变量，其概率分布为

$$P(X=x_i)=p_i,\quad i=1,2,3,\cdots,$$

则
$$D(X)=\sum_i[x_i-E(X)]^2\cdot p_i。\tag{1-8}$$

2. 连续型随机变量的方差

设 X 为连续型随机变量，其概率密度为 $f(x)$，则

$$D(X)=\int_{-\infty}^{\infty}[x-E(X)]^2\cdot f(x)\mathrm{d}x\tag{1-9}$$

3. 方差的性质

设 X,Y 为随机变量，a,b 为常数，则

(1) $D(a)=0$；

(2) $D(a\cdot X)=a^2\cdot D(X)$；

(3) $D(X\pm a)=D(X)$；

(4) 当 X,Y 相互独立时，

$$D(aX\pm bY)=a^2\cdot D(X)+b^2\cdot D(Y)；$$

(5) $D(X)=E(X^2)-E^2(X)$。

随机变量 X 的期望也可以记作 $E(X)=\mu_x$；方差也可以记作 $D(X)=\sigma_x{}^2$。在这里，下标 x 表示属于随机变量 X 的期望或方差。同理，μ_y 表示随机变量 Y 的期望；$\sigma_y{}^2$ 表示 Y 的方差。

此外，在应用中常遇到标准差这个概念。随机变量 X 方差的算术平方根称为随机变量的标准差，记作 $\sqrt{D(X)}$ 或 σ_x。

§1.2 几种常用的概率分布

(一) 超几何分布

1. 超几何分布的概念

设有一批产品，批量 N 为有限数，假定其中含有 D 个不合格品。从批中随机抽取容量为 n 的样本，则这 n 个样本中含有的不合格品个数是一个随机变量，不妨用 X 表示这个随机变量。n 个样本中恰含有 d 个不合格品的概率可用下式计算

$$P(X=d)=\frac{C_D^d \cdot C_{N-D}^{n-d}}{C_N^n}=\frac{\binom{D}{d}\binom{N-D}{n-d}}{\binom{N}{n}}, \qquad (1\text{-}10)$$

在(1－10)式中，如果给定 N,D 和 n，则 $P(X=d)$是 d 的函数。记作

$$h(d,n,D,N)=\frac{\binom{D}{d}\binom{N-D}{n-d}}{\binom{N}{n}}, \qquad (1\text{-}11)$$

其中，d 可取 $0,1,2,\cdots,\min(D,n)$。对于批量为 N 的一批产品，从批中随机抽取容量为 n 的样本，所有可能的抽法总数等于

恰有 0 个不合格品的抽法总数

＋恰有 1 个不合格品的抽法总数

＋恰有 2 个不合格品的抽法总数

＋…………………………………

＋恰有 $\min(D,n)$个不合格品的抽法总数，

即

$$\binom{N}{n}=\sum_{d=0}^{\min(D,n)}\binom{D}{d}\binom{N-D}{n-d},$$

将上式两端同除以$\binom{N}{n}$，得

$$\sum_{d=0}^{\min(D,n)} h(d;n,D,N)=1。 \qquad (1\text{-}12)$$

根据离散型随机变量概率分布的概念可知，$P(X=d)=h(d;n,D,N)$是随机变量 X 的概率分布。这个分布称为超几何分布，它揭示了在容量为 n 的样本中含有不合格品个数的分布规律。

如前所述，容量为 n 的样本中含有不合格品的个数 X 是一个离散型随机变量，它的期望和方差分别由下面的式子给出：

$$E(X)=\mu_x=\frac{nD}{N},\qquad(1\text{-}13)$$

$$D(X)=\sigma_x^2=\frac{nD}{N}\cdot\frac{N-D}{N}\cdot\frac{N-n}{N-1}。\qquad(1\text{-}14)$$

2. 超几何分布的重要特性

(1) n 与 D 互换后，其值不变，即

$$h(d;n,D,N)=h(d;D,n,N)。\qquad(1\text{-}15)$$

(2) 以$(N-D)$替换 D，以$(n-d)$替换 d，其值不变，即

$$h(d;n,D,N)=h(n-d;n,N-D,N)。\qquad(1\text{-}16)$$

3. 超几何分布的图形表示

由(1-10)式表示的超几何分布，给定 N,D 和 n 后，$P(X=d)$ 是 d 的函数。因为 X 是离散型随机变量，所以它的图形是由若干个离散点构成，为醒目起见，相邻的点用线段连接。图 1-1 绘出了 N,D,n 不全相同的超几何分布图形。

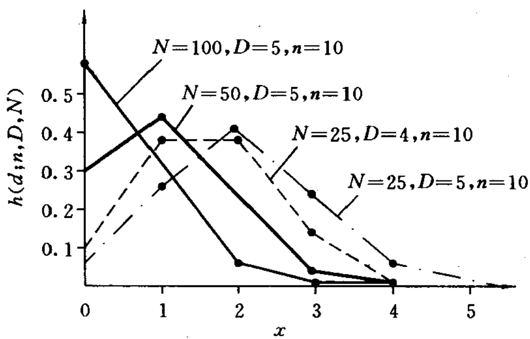

图 1-1　超几何分布的图形

例 1　一批产品，批量为 100 件。已知批不合格品率为 0.01，从批中抽取 5 件，求其中含有不合格个数的概率分布。

解　设样本中含有的不合格品个数为 X，$N=100$，$D=(100)(0.01)=1$，$n=5$。

$$P(X=0)=\frac{\binom{1}{0}\binom{99}{5}}{\binom{100}{5}}=0.95,$$

$$P(X=1)=\frac{\binom{1}{1}\binom{99}{4}}{\binom{100}{5}}=0.05。$$

（二）二项分布

1. 二项分布的概念

一批产品，批量为无限大，假定产品总体的不合格品率为 p。从总体中随机抽取容量为 n 的样本，由于总体是无限的，所以样本中含有的不合格品个数恰等于 d 这个事件的概率为

$$P(X=d)=\binom{n}{d}p^{d}(1-p)^{n-d} \tag{1-17}$$

二项分布的期望和方差由下式给出：

$$E(X)=\mu_x=np \tag{1-18}$$

$$D(X)=\sigma_x^2=np(1-p) \tag{1-19}$$

2. 二项分布的图形表示

在二项分布中，给定 n 和 p 后，$P(X=d)$是 d 的函数。d 的可能取值为 $0,1,2,\cdots,n$。所以，二项分布的图形由$(n+1)$个离散点构成。图 1-2 绘出了 n 的值不全相同的二项分布图形。

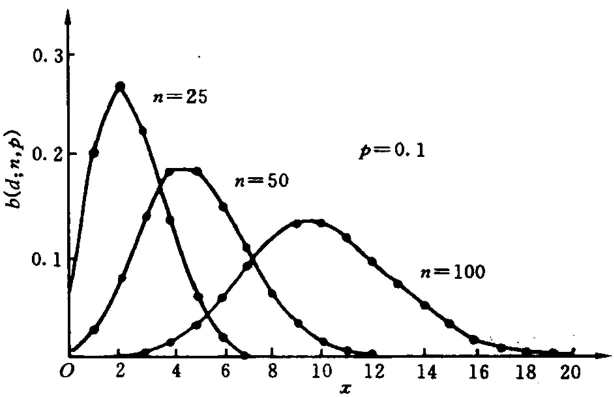

图 1-2　二项分布的图形

例 2　某种产品的日产量很大,不合格品率为 0.01。把日产量看作一批,从中抽取 3 个单位产品,求样本中含有的不合格品个数的概率分布。

解　$N=\infty, n=3, p=0.01$。由(1—17)式可得:

$$P(X=0)=\binom{3}{0}(0.01)^0(0.99)^3=0.970299$$

$$P(X=1)=\binom{3}{1}(0.01)^1(0.99)^2=0.029403$$

$$P(X=2)=\binom{3}{2}(0.01)^2(0.99)^1=0.000297$$

$$P(X=3)=\binom{3}{3}(0.01)^3(0.99)^0=0.000001$$

(三) 泊松分布(S. D. Poisson,1781—1840)

1. 泊松分布的概念

如果单位产品的缺陷数满足下列三个假定,则说单位产品的缺陷数服从泊松分布。

(1) 在单位产品很小的面积(长度或体积,下同)上,出现两个

或两个以上缺陷的概率很小，在极限的状态下可以略去不计。

(2) 在任一很小的面积上，出现一个缺陷的概率仅与面积成正比。

(3) 在任一很小的面积上是否出现缺陷，与另一很小的面积上是否出现缺陷相互独立。

设 X 表示单位产品上的缺陷数，并且满足上述三个假定，则 X 服从泊松分布。理论研究和长期实践都表明，事件"$X=d$"的概率 $P(X=d)$ 可用下式表出。

$$P(X=d)=\frac{\lambda^d}{d!}e^{-\lambda} \tag{1-20}$$

其中 λ 为单位产品缺陷数的期望。

泊松分布的期望和方差分别由下式给出。

$$E(X)=\mu_x=\lambda \tag{1-21}$$

$$D(X)=\sigma_x^2=\lambda \tag{1-22}$$

2. 泊松分布的图形表示

在泊松分布中，给定 λ 后，$P(X=d)$ 是 d 的函数。d 的可能取值为 0，1，2，…。所以，泊松分布由无穷个离散点构成。图 1-3 绘出了不同 λ 值的泊松分布图形。

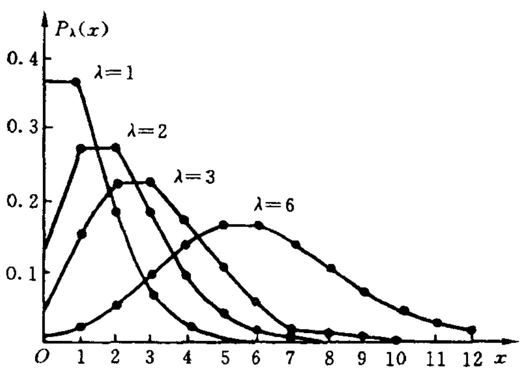

图 1-3 泊松分布图形

3．泊松分布的计算

泊松分布的计算，可以使用功能齐全的计算器，也可以利用泊松分布表。参见附表2。

（四）正态分布

1．正态分布的概念

任何产品，由于受制造过程中各种因素的干扰，即使生产过程稳定，产品的质量特征值仍然存在差异。这种差异是正常的，无差异是不可能的。问题在于如何研究这些差异。

设产品的某个质量特征值为 X，如果 X 的概率密度 $f(x)$ 可用下式表示，即

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<\infty \tag{1-23}$$

则称随机变量 X 服从正态分布，记作 $X\sim N(\mu,\sigma^2)$。容易证明，无论参数 μ 和 σ 取何值，都有 $f(x)>0(\sigma>0)$，且

$$\int_{-\infty}^{\infty}f(x)\mathrm{d}x=1 \tag{1-24}$$

事实上，令 $\frac{x-\mu}{\sigma}=t$，则

$$\int_{-\infty}^{\infty}f(x)\mathrm{d}x=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\cdot\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t=\frac{1}{\sqrt{2\pi}}\cdot\sqrt{2\pi}=1$$

所以，(1-23)式可以作为正态随机变量 X 的概率分布。

正态分布的两个参数 μ 和 σ 具有特别意义，即

$$E(X)=\mu \tag{1-25}$$

$$D(X)=\sigma^2 \tag{1-26}$$

2．正态分布的图形表示

由(1-23)式可知，给定 μ 和 σ 后，$f(x)$ 是 x 的连续函数。图1-4绘出了 $\mu=0$，σ 取不同值的三条正态分布密度曲线。