SHULITONGJIYUCHANPINZHILIANG

许金钊 编译



# 数理统计与产品质量

中国标准出版社

# 数理统计与产品质量

许金钊　编译

中国标准出版社

**数 理 统 计 与 产 品 质 量**

许金钊　编译

·

中国标准出版社出版

（北京复外三里河）

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 6　字数 162,000

1986年 2 月第一版　1986年 2 月第一次印刷

印数 1—9,000

·

书号：15169·3-277　定价 1.85 元

·

标目 31—2

## 内 容 提 要

本书主要介绍如何正确运用数理统计方法，来解决机械制造工艺方面和机械产品质量方面的问题，从而达到提高产品质量、提高生产效率和降低生产成本的目的。内容有：一、概率论的基础知识；二、数理统计理论概述；三、机械制造工艺过程的精度分析；四、工艺研究中统计法的应用；五、工艺计算中统计法的应用；六、统计检查和质量控制等。书末还有附表，可供查用。

本书可供机械制造部门的质量管理人员、标准化工作者、数理统计人员和有关院校师生参考使用。

# 前　言

数理统计是一门根据实验数据或实际观测资料来研究随机变量的分布函数和参数特征的科学。数理统计所研究的对象与概率论一样是随机变量（或随机事件），所以前者是以后者作基础的。另一方面，要以有限的实际观测资料为依据来研究随机变量的客观规律，除了概率论外，还得建立一些有关局部与整体之间内在联系方面的理论，这就是数理统计所讨论的内容。

由于这种从部分观测资料去推断随机变量的客观规律的方法具有普遍而重要的意义，数理统计的应用日趋广泛，而且数理统计学本身也在不断的发展中。农业生产中农作物产量的提高，新品种的试验和推广，气象的天气预报，新药物的试制和应用，以及其他科学技术部门都得应用数理统计。

显然，机械制造工业也不能例外。应用数理统计来解决工艺问题日趋频繁。为此，编写本书供机械加工工厂的质量管理人员、产品检验人员以及大专院校机械制造专业的师生参考。

必须说明：编写本书的目的不在于探讨数理统计理论，而是介绍如何正确的使用这种统计方法来解决我们在机械制造工艺方面或机械产品质量方面可能遇到的问题，从而达到提高产品质量、提高生产效率和降低生产成本的目的。

本书承李继祯同志全面审校。但由于编译者的专业知识肤浅，生产实践更加贫乏，书中不足或错误之处一定不少，请读者批评指正。

**编译者**

1982年10月于长春

# 目　　录

# 第一章

# 概率论的基础知识

## §1 基本概念和定义

数理统计是以概率为基础的，在概率论中采用了一系列的特种概念，如试验（实验）、事件、随机变量、概率、频数和频率。

任何综合条件在实际现场的实现叫做试验或实验。

在一定综合条件实现的结果中，即在试验或实验的结果中，所产生的现象叫做事件。

进行多次重复试验所产生的现象或事件叫做大量事件。

如果每次试验时，事件 $A$ 一定发生，则这一事件叫做必然事件。

如果在已知试验条件下，事件 $B$ 显然不会发生，则称它为不可能事件。如果在一次试验中可能发生事件 $A$，或 $B$、或 $C$ 等，则这些事件称谓是或然的、或随机的。所以在试验中可能发生或者可能不发生的那种事件就叫做随机事件。例如在箱内装有 100 个零件，其中有一个是废品，从箱内取得这个报废零件将是随机事件，因为它可以出现也可以不出现。

"随机性"这个名词在任何情况下都不应该理解成为被研究事件是毫无规律的。辩证唯物主义告诉我们：任何现象，包括我们所说的随机现象在内，都包含着它自己所特有的或多或少的复杂因素。

随机变量是一种变化的量，它在试验结果中，可能采取这个或那个值，例如：在机床上加工出来的零件的实际尺寸是一个变量，它可

以具有在公差范围内的任何值，所以它是随机变量。

在研究大量现象时，任何随机事件或随机变量在试验过程中能够出现若干次，例如，在 $N$ 次试验中，事件 $A$ 出现了 $k$ 次。$k$ 次就叫做事件 $A$ 出现的频数，事件 $A$ 出现的频数与试验总次数 $N$ 之比叫做事件 $A$ 的出现频率（或称相对频数），用 $f_A$ 表示。

$$f_A=\frac{k}{N} \tag{1}$$

例如：在机床上加工 100 个零件。通过对这批零件的测量，指出 75 个零件的尺寸位于公差范围之内，其余 25 个零件的尺寸超出了公差范围。所以在 100 次试验中，出现合格零件这一事件 $A$ 的频率为：

$$f_A=\frac{75}{100}=0.75$$

出现废品这一事件 $B$ 的频率为：

$$f_B=\frac{25}{100}=0.25$$

如果随机事件有稳定的频率，则在大量试验时，随机事件出现的频率将波动在某一数值的附近。这一数值就叫做随机事件的概率。随机事件 $A$ 的概率用符号 $P$（$A$）表示，它是一个小于 1 的正数，又是表达随机事件 $A$ 的可能性的数量估计，概率是在大量试验中随机事件的频率波动所围绕的那个数值。

根据给定的概率定义，为了解决任何事件 $A$ 的概率，应该进行大量试验。按照试验结果计算频率 $f_A$，以便近似地决定概率 $P(A)$。

但是，也可以不进行试验来决定概率 $P$（$A$）。为此，采用了所谓概率的古典定义。

事件 $A$ 的概率等于对该事件的有利机会数 $m$ 与给定类型试验的所有可能机会数 $n$ 之比。即：

$$P(A)=\frac{m}{n} \tag{2}$$

这时假设所有机会数 $n$ 是有限的、等可能性的、互不相容的，而且也

是唯一可能的。

在给定的试验中，所有可能机会 $n$ 于每次试验时具有同等可能出现的机会称这些事件为等可能事件。当这些事件之一的出现排斥了其他事件出现的可能时，这些事件称谓互不相容事件。在给定类型试验中，当其他任何事件都不可能出现而仅有一个事件能出现，这个事件称谓唯一可能事件。

综合类型试验是指产生这种或那种事件的一定条件的综合，例如：在箱里有 100 个零件，在零件上编上号码 1 到 100，一下取得第 40 号零件的概率为若干？显然，有利于这事件出现的机会数等于 1，而该试验的所有可能机会数等于 100。这时，所有机会数是等可能的、互不相容的，并且也是唯一可能的。即：

$$P(A)=\frac{1}{100}$$

## §2 概率的性质及其加法定理

必然事件 $U$ 的概率等于 1，即：

$$P(U)=1 \tag{3}$$

因为必然事件的有利机会数 $m$ 等于所有可能机会数 $n$，即 $m=n$，由此得出：

$$\frac{m}{n}=1$$

不可能事件 $V$ 的概率等于零，即：

$$P(V)=0 \tag{4}$$

因为 $n$ 个可能机会中没有一个对不可能事件是有利的，即 $m=0$ 或 $\frac{m}{n}=0$。

任何随机事件 $A$ 的概率位于不可能事件 $V$ 的概率和必然事件 $U$ 的概率之间，或者说位于零和 1 之间，即：

$$0<P(A)<1 \tag{5}$$

因为 $m<n$。

当事件 $A$、事件 $B$ 和事件 $C$ 互不相容，即它们之中任何一个的出现排斥了其他事件出现的可能性，而且每一事件出现的概率分别等于 $P(A)$、$P(B)$、$P(C)$ 时，则这些互不相容事件的任一事件（$A$ 或 $B$ 或 $C$）出现的概率等于它们各自概率的总和。即：

$$P(A\text{或}B\text{或}C)=P(A)+P(B)+P(C) \qquad (6)$$

例如：100 个零件中有 6 个零件的尺寸小于最小极限尺寸，有 4 个零件的尺寸超过最大极限尺寸，则取得小于最小极限尺寸的零件这一事件 $A$ 的概率为：

$$P(A)=\frac{6}{100}$$

取得超过最大极限尺寸的零件这一事件 $B$ 的概率等于：

$$P(B)=\frac{4}{100}$$

取得不合格零件的概率，即超出极限尺寸范围这一事件（$A$ 或 $B$）的概率为：

$$P(A\text{或}B)=P(A)+P(B)=\frac{6}{100}+\frac{4}{100}=\frac{10}{100}$$

从概率加法定理导出两个推论：

1）对立事件概率之和等于 1。如果用 $P$ 表示任一事件出现的概率，而它不出现的概率用 $q$ 表示，则 $p+q=1$，因为事件要么出现，要么不出现。

2）对立事件中一事件的概率等于 1 减去反面事件的概率。即：

$$q=1-p \qquad (7)$$

例如：在箱里有 100 个零件，其中 90 个合格和 10 个不合格，取得合格零件的概率等于 0.9，反面事件，即取得不合格零件的概率为

$$q=1-p=1-0.9=0.1$$

## §3 大数定律

任何事件的概率和频率几乎相等，其接近程度随试验次数的增加而增长，这一推论从贝奴里定理导出。贝奴里定理指出：在大量试验 $N$ 次中，具有一定概率 $P$ 的事件 $A$ 的出现频率 $\frac{k}{N}$ 将与该事件概率的差别达到任意小的程度。

$$\lim P\left(\left|\frac{k}{N}-P\right|<\varepsilon\right)=1 \qquad 当 N\to\infty$$

式中 $\varepsilon$——任意小的正数。

贝奴里定理和它的概括叫做大数定律。根据这一定律，当试验次数增加，任何事件的频率接近于它的概率，并且试验次数越多，两者越靠近。即：

$$P\approx\frac{k}{N}=f_A$$

当事件的概率（$P$）为已知时，可从上式求得期望频率 $f_A$ 或一系列实验（$N$）中的频数 $k$。反之，当若干事件的概率为未知，则基于大数定律就能求得它们的近似值，这时必须进行足够多的随机而互相独立的试验，把获得的频率作为未知概率的估量。

大数定律是随机性和必然性间的辩证联系的表达法之一，它有重大的实际意义，并为采用统计概率法研究大量现象的规律性提供了理论依据。

## §4 随机变量和它的分布

随机变量分为离散的和连续的。

离散随机变量是这样一种变量，它在试验结果中仅能取得隔离的、正整的数值，而不能取得这些数值之间的中间值。比如：一批零件中不合格品的数目仅能是正数值1、2、3等，所以不合格品的数

目是离散随机变量。

连续随机变量是这些变量，它在试验结果中能取得一定界限范围内的任何数值，比如在机床上加工出来的零件的实际尺寸是连续随机变量，因为它能取得公差范围内的任何数值。

在试验中，随机变量取得这个或那个数值的可能性，是靠概率来估计的。随机变量取得的数值及其相应概率的综合，叫做随机变量的分布。随机变量的分布分为理论的和经验的两种。在理论分布中，随机变量的数值数目一般是无限的。在经验分布中，随机变量数值的数目是有限的，并由试验或实验的结果中获得。所以随机变量的经验分布是随机变量观察值及其相应频数或频率的综合。

随机变量的分布可以用表和图的形式来表达。离散随机变量的分布可以列成表 1 和表 2，或者在表 2 的基础上制成图（图 1）。

表 1　离散随机变量的理论分布

| $x_i$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ | …… | $x_{n-2}$ | $x_{n-1}$ | $x_n$ | |
|---|---|---|---|---|---|---|---|---|---|
| $p(x_i)$ | $p(x_1)$ | $p(x_2)$ | $p(x_3)$ | $p(x_4)$ | | $p(x_{n-2})$ | $p(x_{n-1})$ | $p(x_n)$ | $\sum_1^n p(x_i)=1$ |

表 2　离散随机变量的经验分布

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|---|
| $p(x)$ | 1/32 | 5/32 | 10/32 | 10/32 | 5/32 | 1/32 | $\sum_1^n p(x)=1$ |

连续随机变量的分布用表或图的形式来表达是有困难的，即使随机变量的分布位于一个很狭窄的范围内。对连续随机变量进行实际观察时，我们也会遇到一系列困难，如测量工具和设备的精度不足等。所以实际上，在研究连续随机变量时，把连续随机变量分成间段或区段，使区段的间隔（或间距）比量具的刻度值稍大，这样便补偿了测量精度的不足。之后，不按随机变量的实际值而根据区段来计算频数。换言之，不计算连续随机变量实际值的概率或频率，而计算其在各区段或间隔内那些值的概率（或频率）。这样，连续随机变量的分

布将有表 3 和表 4 的形式。

**表 3　连续随机变量的理论分布**

| $x_i$ | $p_i$ | $x_i$ | $p_i$ |
|---|---|---|---|
| 从$x_1$到$x_2$ | $p(x_1 \leqslant X < x_2)$ | …… | …… |
| 从$x_2$到$x_3$ | $p(x_2 \leqslant X < x_3)$ | 从$x_{m-1}$到$x_m$ | $p(x_{m-1} \leqslant X < x_m)$ |
| 从$x_3$到$x_4$ | $p(x_3 \leqslant X < x_4)$ | | |

**表 4　连续随机变量的经验分布**

| 间　　隔 | 频　数　$n$ | 频率 $n/N$ |
|---|---|---|
| 20～20.05 | 2 | 0.02 |
| 20.05～20.10 | 10 | 0.10 |
| 20.10～20.15 | 24 | 0.24 |
| 20.15～20.20 | 30 | 0.30 |
| 20.20～20.25 | 22 | 0.22 |
| 20.25～20.30 | 10 | 0.10 |
| 20.30～20.35 | 2 | 0.02 |
| 共　　计 | 100 | 1 |

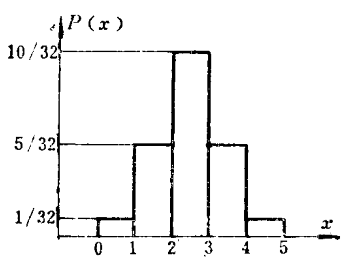


图 1　离散随机变量分布图

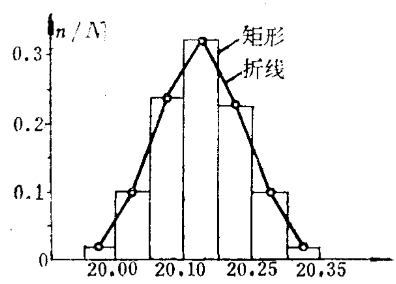


图 2　连续随机变量分布图

连续随机变量的经验分布（表 4）可以用阶梯图或折线图来表示（图 2）。阶梯图叫做矩形式分布，而折线图叫做多边形分布或经验曲线分布。

在连续随机变量的理论描述和研究中，很难把它们分成区段。因此，为了解决这种困难，引入了分布函数这个概念。

令 $X$ 代表随机变量，而 $x$ 表示随机变量的任何实际数，当 $X<x$ 时，用 $P(X<x)$ 表示这一事件 $(X<x)$ 的概率。显然，它是 $x$ 的函数，即：

$$P(X<x)=F(x) \tag{8}$$

$F(x)$ 叫做随机变量的概率分布函数或分布积分函数。这样，分布积分函数便能决定在试验中随机变量 $X$ 取得的任意小于实际测量值 $x$ $(-\infty<x<+\infty)$ 的概率了。如果随机变量的分布函数为已知，则可以认为随机变量本身是给定的了。

离散随机变量的分布积分函数 $F(x)$ 可以按分布表或分布图来决定。例如：根据图 1，任意值 $x$ 的 $F(x)$ 等于那些在 $x$ 点之左的 $X$ 值的各实际值 $x_i$ 概率之总和。其中如 $X<3$

$$P(X<3)=P(x=0)+P(x=1)+P(x=2)$$

$$=\frac{1}{32}+\frac{5}{32}+\frac{10}{32}=\frac{16}{32}$$

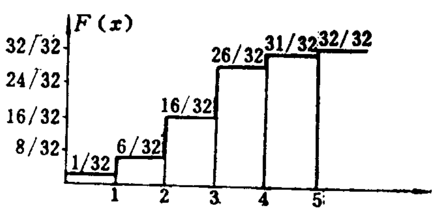


图 3 离散随机变量分布积分函数图

分布积分函数可以用图形来表示，如果以横坐标表示 $x$ 值，而以纵坐标表示 $F(x)$，则：

$$F(x)=P(X<x)$$

离散随机变量分布积分函数图将呈阶梯形的折线。像表 2 的分布，可用图 3 表示出来。这折线的纵坐标对任何 $x$ 值将代表以前各 $x$ 值的概率之总和，即 $F(x)=P(X<x)$。

如果已知 $F(x_1)$ 和 $F(x_2)$，即在积分折线上任意相应于横坐

标上两点的纵坐标，就能知道包括随机变量 $X$ 之值小于 $x_1$ 或 $x_2$ 这种事件的概率。因为：

$$F(x_1)=P(X<x_1)$$

$$F(x_2)=P(X<x_2)$$

已知这些概率，就能计算在试验中随机变量从 $x_1$ 到 $x_2$（包括 $x_1$，不包括 $x_2$）范围内的概率，即：$P(x_1\leqslant X<x_2)$。

显然，事件 $X<x_2$ 包含着两部分事件，即 $X<x_1$ 和 $x_1\leqslant X<x_2$。在概率加法定理的基础上，能写成：

$$P(X<x_2)=P(X<x_1)+P(x_1\leqslant X<x_2)$$

由此得：

$$\begin{aligned}P(x_1\leqslant X<x_2)&=P(X<x_2)-P(X<x_1)\\&=F(x_2)-F(x_1)\end{aligned}\qquad(9)$$

这样，随机变量 $X$ 在试验中出现于 $x_1$ 到 $x_2$ 之间的概率等于分布在该间隔上的积分函数的增量。

连续随机变量的分布积分函数图将不是阶梯折线，而呈逐渐增长的曲线。通过曲线上的每点都可作切线（图4）。

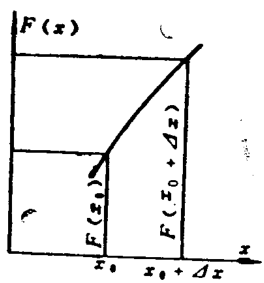


图 4　连续随机变量分布积分函数的曲线形状

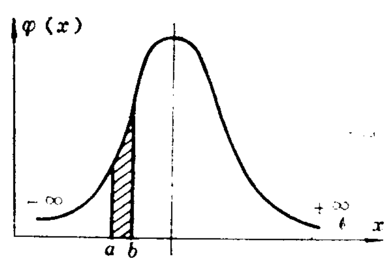


图 5　连续随机变量分布微分函数的曲线形状

在横坐标上任取两点 $x_0$ 和 $x_0+\Delta x$，在这两点上的函数纵坐标将为 $F(x_0)$ 和 $F(x_0+\Delta x)$。与上面所说的一样，可以写成：

$$P(x_0\leqslant X<x_0+\Delta x)=F(x_0+\Delta x)-F(x_0)\qquad(10)$$

连续随机变量分布积分函数是可微函数。积分函数的一次导数叫