测量误差的统计分布和检验

张方仁　张金通　编著

中 国 计 量 出 版 社

测量误差的
统计分布和检验

张方仁　　张金通　编著

中国计量出版社

新登(京)字 024 号

内容提要

本书主要介绍测量误差的各种统计分布、异常值的探测、数据诊断和近代拟合检验方法等.各章都附有实际应用的算例，兼顾数学理论和实际应用，书末还编制了 31 个附表，表中大部分是作者的研究成果，其中大多数在其他文献中尚未见到。

本书可作为高等院校有关专业的教学参考书，也可供从事数据 处理的工程技术人员参考。

测量误差的统计分布和检验

张方仁　张金通　编著

责任编辑:方荣颐　栾桂芬

*

中国计量出版社出版

北京和平里西街甲 2 号

河北省三河县潮河印刷厂印刷

新华书店北京发行所发行

*

开本 787×1092　1/16　印张 14 字数 350 千字

1992 年 3 月第 1 版　1992 年 3 月第 1 次印刷

印数　1—3000

ISBN　7—5026—0493—6/TB・383

定价　9.50 元

前　言

本书主要讨论测量误差的各种统计分布、它们的定义、性质、相互间的关系及其应用、数据诊断和探测、近代的拟合检验方法等。

全书共分四章，第一章概率统计的基本知识。第二章测量误差的统计分布（指常用的正态分布、χ^2、t、F 分布等以外），包括 τ 分布，瑞利分布、伽玛分布、贝塔分布、相关系数分布、顺序量分布和极差分布等，并列举了实际例子，编制了各种数表（这些数表中的大多数在其它文献中尚未见到过，对实际工作很有好处）；第三章异常值探测和数据诊断，论述了残差的 统计性质和应用 ，各种探测异常值的方法，影响函数和它在数据诊断中的应用等；第四章近代拟合方法，讨论了夏皮罗-威尔克检验的原理和方法并对其系数作了扩充（偏、峰态检验的新进展，A^2 和 W^2 检验等等）。最后是 31 个附表为各种应用数表。

数学理论和应用技术兼容是本书努力的目标。

一般数学理论书籍其公式推导过于简略，阐述也太概括，于应用方法则更为简略，工程技术人员难以理解和接受。本书从应用的角度出发，讲清应用方法的同时，对一些重要的公式和理论概念作适当的推演和阐述，力求推证详细，叙述通俗，并结合实际应用例子，使从事实际工作者易于接受，不仅使他们能掌握这些数学方法，且在理论上得到一定的提高，这是其一。

其二是各种常用的统计分布（不含 χ^2、t、F 分布），目前的书刊上往往只介绍它们的定义、性质，而没有刊出有关的应用数表，对于应用很不方便。因为一个科学计算问题中，统计检验、拟合检验等仅仅是其中一个部分，如果每次应用时，均列公式于计算程序中计算，往往使程序复杂，增加计算时间，故一般不应采用这样的方法。鉴于此，作者编算了这些数表，相信在“统计检验”中一定会起到好作用。

第三，书中有些数表，目前书刊中只编算了一部分，作者近年来进行这方面的工作，完成了它们的编算工作。例如，相关系数检验临界值表，国内、外书刊上都是 $\rho=0$ 的表（David 于 1938 年编制出），对于 ρ 为其它数值的表一直未见刊出，而 $\rho\neq0$ 的相关系数检验在很多领域中会遇到，例如在测绘专业中，相邻三角形闭合差的相关系数 $\rho=-\frac{1}{3}$。作者编出了 $-1<\rho<+1$ 的相关系数检验临界值表，，对于这方面的检验无疑是有益的，它还可以用来作相关系数的区间估计。再如夏皮罗-威尔克正态性检验法自 1965 年问世以来，只有容量 $n\leqslant50$ 的系数表，我们常要作 $n>50$ 的似合检验，而且这种方法是正态性检验方法中功效最好的方法之一，本书编制了 $50<n\leqslant300$ 的系数表，还有一些其他数表，不一一列举。

本书系中国科学院测量与地球物理研究所动力大地测量学开放研究实验室和武汉测绘科技大学大地系合作课题的研究成果。

本课题的研究和本书的撰写得到许厚泽研究员、朱灼文研究员、刘大杰教授和樊功瑜教授的关心和支持，也得到方荣颐高级工程师的热心帮助，在此一并致谢。

由于作者水平有限，难免有错误和欠妥之处，恳请读者指正。

作者　1991 年 7 月于武昌

目　录

第一章　概率统计的基本知识

§1.1　随机事件和概率

一、随机事件和样本空间

在自然界和人类社会中，人们观察到的现象可分为两类，一类是在一定条件下它们是否出现完全可以预言的，例如“在一个大气压下，水在100℃时会沸腾”。“在没有外力作用下，作等速直线运动的物体必然继续作等速直线运动”等等，这种在一定条件下必然出现的现象称为必然事件。显然，它的反面是不可能事件。另一类事件是事前不可预言的，在一定条件下，它们可能出现也可能不出现，例如“掷一枚质地均匀的硬币，结果可能是正面朝上，也可能正面朝下”等等，这类现象称为随机事件，简称事件。

我们常常通过随机试验来观察随机事件的统计规律性。设 E 为一试验，它可在相同条件下重复进行，如不能事先准确地预言某一次试验的结果，则称为随机试验。以 ω 表示它的一个可能结果，称 ω 为 E 的一基本事件，全体基本事件的集合 $\Omega=\{\omega\}$ 称为样本空间或基本事件空间。

例 1.1　E——掷一枚质地均匀的硬币观察出现的面；ω_1——正面朝上；ω_2——正面朝下；于是 $\Omega=\{\omega_1,\omega_2\}$。

例 1.2　E——在一个箱子中装有10球，球上标以数字1，2，…，10，若从箱子中随机地取一个球，令 ω_i——球上的数字为 i，则 $\Omega=\{\omega_1,\cdots,\omega_{10}\}$。

基本事件是不可能再分的事件，一般的事件总是由若干基本事件组合而成的，称为复合事件，它是 Ω 的一个子集。显然是必然事件，用 Φ 表示不可能事件。

二、事件的关系和运算

一个样本空间 Ω 中有很多随机事件，我们往往要通过对较简单事件规律的研究去掌握更复杂事件的规律。为此，需要研究事件之间的关系和运算。

(1)如果事件 A 发生必然导致 B 发生，就说 B 包含 A 或 A 含于 B，记作为 $A\subset B$，见图1.1(a)。如果 $A\subset B$ 且 $B\subset A$，就说 A 与 B 相等，记作 $A=B$。

(2)“两事件 A、B 中至少有一个发生”也是一个事件，称它为 A 与 B 的和，记作 $A\cup B$，或 $A+B$，见图1.1(b)中阴影部分。

(3)“两事件 A、B 同时发生”也是一个事件，称它为 A 与 B 的交，记为 $A\cap B$(或 AB)，见图1.1(c)中阴影部分。

(4)“事件 A 发生而事件 B 不发生”也是一个事件，称它为 A 与 B 的差，记为 $A-B$，见图1，1(d)中阴影部分。

(5)称“事件 Ω 与 A 的差”为 A 的逆事件，记作 $\overline{A}$ 见各1.1(e)中阴影部分。

(6)如果事件 A 和 B 不能同时发生，也就是说 $AB=\Phi$，则说 A 与 B“互不相容”，见图 1.1(f)。

如果 n 个事件 $A_1,\cdots,A_n$ 中任意两个事件是互不相容的，就说 $A_1,\cdots,A_n$ 互不相容。

根据集合论知识，不难发现事件间的关系和运算与布尔代数中集合间的关系和运算之间是完全可以互相类比的，为了便于对照，把它们的术语列于表 1.1。

表 1.1

符　号	集合论	概率论
Ω	空间	样本空间；必然事件
Φ	空集	不可能事件
$\omega\in\Omega$	Ω 中的点(或元素)	样本点
$\{\omega\}$	单点集	基本事件
$A\subset\Omega$	Ω 的子集 A	事件 A
$A\subset B$	集合 A 包含在集合 B 中	事件 A 含于事件 B
$A=B$	集合 A 与 B 相等	事件 A 与 B 相等
$A\cup B$	集合 A 与 B 之和	事件 A 与 B 至少有一发生(事件 A 与 B 之和)
$A\cap B$	集合 A 与 B 之交	事件 A 与 B 同时发生(事件 A 与 B 之交)
$\bar{A}$	集合 A 的余集	事件 A 的逆事件
$A-B$	集合 A 与 B 之差	事件 A 发生而事件 B 不发生(事件 A 与 B 之差)
$A\cap B=\Phi$	集合 A 与 B 没有公共元素	事件 A 与 B 互不相容

三、 概率的公理化定义和概率的性质

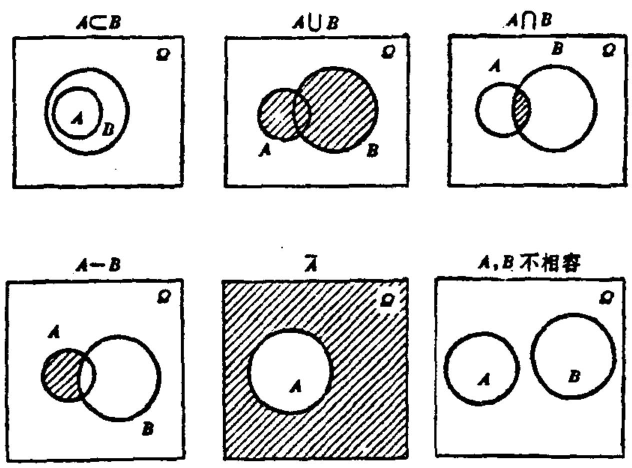

图 1.1 ($A\cup B$, $A\cap B$, $A-B$, $\bar{A}$ 分别为图中阴影部分)

近代概率论的公理结构，给出了事件与概率的严格定义。

定义 1.1　设 Ω 是抽象的点 ω 的集，Ω 中的一些子集所成的集 $\mathscr{F}$ 称为 Ω 中的一个 σ—代数，如果 $\mathscr{F}$ 满足

(i) $\Omega\in\mathscr{F}$；

(ii) 如 $A\in\mathscr{F}$，则 $\bar{A}\in\mathscr{F}$；

(iii) 如 $A_n\in\mathscr{F}$ $(n=1,2,\cdots)$，则 $\bigcup_{n=1}^{\infty}A_n\in\mathscr{F}$。

定义 1.2、设 $P(A)(A\in\mathscr{F})$ 是定义在 σ—代数上的实数集函数，如果它满足下列条件

(i) 对每个 $A\in\mathscr{F}$，有 $0\leqslant P(A)\leqslant 1$；

(ii) $P(\Omega)=1$；

(iii)如 $A_n \in \mathscr{F}$ (n $=1,2,\cdots$) ,且 $A_iA_j=\Phi(i\neq j)$,则有 $(P(\bigcup_{n=1}^{\infty}A_n)=\sum_{n=1}^{\infty}P(A_n)$ 就称它为 $\mathscr{F}$ 上的概率,而称 $\mathscr{F}$ 中的集为事件,三元体 $(\Omega,\mathscr{F},P)$ 为概率空间。

根据定义 1.2,可得概率的性质如下

(1) $P(\Phi)=0$

(2)如果 A,B 为两事件,且 $A\supset B$,则 $0\leqslant P(A-B)=P(A)-P(B)$,从而 $P(A)\geqslant P(B)$

(3)对任一事件 A,有 $P(\overline{A}) = 1 - P(A)$

(4)对任意 n 个事件 A_1、……、A_n,有

$$P(A_1 + A_2 +,\cdots\cdots. + A_n) \leqslant P(A_1) + P(A_2) +,\cdots\cdots\cdots, + P(A_n)$$

(5)对 n 个事件 $A_1,\cdots,A_n$,若 $A_iA_j=\Phi(i\neq j)$,则有

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

(6)对任意两个事件 A 和 B,有

$$P(A + B) = P(A) + P(B) - P(AB) \qquad (1.1)$$

(7)对任意三个事件 A,B,C,有

$$P(A + B + C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC) \qquad (1.2)$$

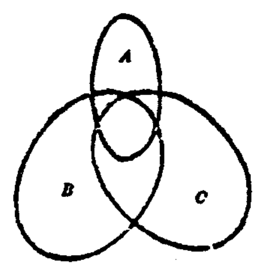

图 1.2

这个公式由图 1.2 很容易导出。

(8)对任意 n 个事件 $A_1\cdots A_n$,有

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n}P(A_i) - \sum_{1\leqslant i<j\leqslant n} P(A_iA_j) + \sum_{1\leqslant i<j<k\leqslant n} P(A_iA_jA_k) -,\cdots\cdots, + (-1)^{n-1}P(\bigcap_{i=1}^{n} A_i) \qquad (1.3)$$

(1.3)式可根据(1.1)和(1.2)用归纳法推导出来,这个公式也称为概率的一般加法公式。

四、 条件概率

在实际问题中,一般除了要知道事件 A 的概率 $P(A)$外,有时还需要知道在"事件 B 已发生"这一条件下,事件 A 发生的概率,一般地说,这两者未必相同。我们把后者称作条件概率,记为 $P(A|B)$。

我们可以通过图 1.3 来解释 $P(A|B)$的含意。当已知 B 已发生,表示样本点的可能区域是 B,此时的样本空间已不是原来的样本空间 Ω,已缩小为 B,这时 A 发生等价于样本点落在 AB 的区域内,由概率的定义,若 $P(B)>0$,则

$$P(A|B) = \frac{P(AB)}{P(B)} \qquad (1.4)$$

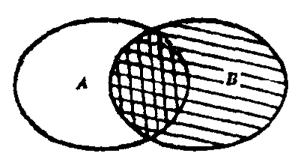

图 1.3

通常用公式(1.4)来定义条件概率。由(1.4)式可以验证条件概率有如下的性质

(i) $0\leqslant P(A|B)\leqslant 1$;

(ii) $P(\Omega|B)=1$

(iii)若 $A_1,\cdots,A_n,\cdots$是互不相容的事件,则

$$P(\bigcup_{n=1}^{\infty} A_n | B) = \sum_{n=1}^{n} P(A_n | B) \tag{1.5}$$

除此以外，条件概率还有几个很重要的公式。

定理 1.1(乘法公式) 设 $A_1,\cdots,A_n$ 为 n 个事件，$n\geq 2$，且 $P(A_1A_2,\cdots\cdots,A_{n-1})>0$，则

$$P(A_1,\cdots\cdots,A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots\cdots P(A_n|A_1A_2\cdots\cdots A_{n-1}) \tag{1.6}$$

定理 1.2(全概率公式) 设 $A_1,A_2\cdots$ 为有限个或无穷个互不相容的事件，且 $\bigcup_{n=1}^{\infty} A_n=\Omega$，$P(A_i)>0$，$i=1,2,\cdots\cdots$。则对任一事件 A，有

$$P(A) = \sum_{n} P(A_n)P(A|A_n) \tag{1.7}$$

定理 1.3(贝叶斯公式) 设 $A_1,A_2,\cdots\cdots$ 为有限或无穷多个互不相容的事件，$\bigcup_{n=1}^{\infty} A_n=\Omega$，$P(A_n)>0$，$n=1,2\cdots\cdots$，则对任一事件 A，$P(A)>0$ 有

$$P(A_m|A) = \frac{P(A|A_m)P(A_m)}{\sum_{n=1}^{\infty} P(A_n)P(A|A_n)}, \qquad m = 1,2,\cdots\cdots \tag{1.8}$$

五、 独立性

若事件 A 发生与否和事件 B 发生与否无关，则称事件 A 与 B 是相互独立的。这时 $P(A|B)=P(A)$，由公式(1.4)得 $P(AB) = P(B)P(A|B) = P(B)P(A)$

于是在一般的教科书上就用这个公式来定义独立性。

定义 1.3 设 A、B 为两事件，如满足

$$P(AB) = P(A)P(B) \tag{1.9}$$

则称事件 A 与 B 是相互独立的。

若 A 与 B 是相互独立的，则不难推出下列三对事件 $(A,\bar{B})$，$(\bar{A},B)$，$(\bar{A},\bar{B})$ 也分别是相互独立的。

定义 1.4 设 $A_1\cdots A_n$ 是 n 个事件，我们说这些事件是相互独立的，如果对任意的 $S(2\leqslant s\leqslant n)$ 任意 $1\leqslant i_1<i_2\cdots<i_s\leqslant n$，有

$$P(A_{i_1}\cdots A_{i_s}) = \prod_{j=1}^{s} P(A_{i_j}) \tag{1.10}$$

"独立性"这个概念在概率统计中是一个极其重要的概念，必须引起读者的注意，同时，还要提及一点，由事件 $A_1\cdots A_n$ 的两两独立不能推出它们的相互独立性。

§1.2 随机变量及其概率分布

一、 随机变量的定义

上节我们讨论了随机事件及其概率，读者可能会注意到，在某些例子中，如例 1.2，随机事件和实数之间存在着客观的联系。因此，我们可以采取另一种方法来描述这样的随机现象。令

X 为随机抽取的球上的号码，则 X 的取值是 $1,2,\cdots\cdots,10$ 中的任一个，并且基本事件 $\{\omega_i\}$ 可以通过 X 来表示，即 $\omega_i=\{X=i\}$，$i=1,2,\cdots\cdots,10$ 并可知 $P(\omega_i)=P(X=i)=1/10$，$i=1,2,\cdots\cdots,10$

我们看到用 X 这个变量来描述例 1.2 中的随机现象是很方便的，由于这个变量的取值事先是无法确定的，且以一定的概率来取值的，故称之为"随机变量"。对于如例 1.1 这样的随机现象，因为其中设有出现任何数字，我们可以这样来处理，若令出现正面时 X 取数为"1"，出现反面时 X 取数为"0"，于是 $\omega_i=\{X=i\}$，$i=0,1$，$P(\omega_i)=P(X=i)=\frac{1}{2}$，$i=0,1$。上述的例子中随机变量的取数都是整数值，还有一些随机变量，它们的取值可能是某个区间中的任何一个数。

随机变量与通常的变量有本质上的区别，通常变量的取值是完全确定的，而随机变量的取值，在每次试验之前是不能确定的，它们的取值要受到随机因素的影响，依赖于试验的结果，并且以一定的概率来取各种数值的。

随机变量在试验之前是一个不确定的量，它可能取各种不同的值，但在试验之后，它只能取这些可能值中的一个，这个值叫做随机变量的观察值。

随机变量按其取值的情况不同，可分为离散型随机变量和连续型随机变量。

二、 离散型随机变量

定义 1.5 如果随机变量 X 只能取有限个或可数个值，并且以各种确定的概率取这些不同的值，则称 X 为离散型随机变量。

设 X 的取值为 $x_1,x_2,\cdots\cdots$，相应的概率为 $P_i=P(X=x_i)$，$i=1,2,\cdots\cdots$，显然 $\{p_i\}$ 满足

(i) $p_i\geqslant 0$，$i=1,2,\cdots\cdots$；

(ii) $\sum_{i=1}^{\infty}p_i=1$

通常用一个二行的数组或一个表格的形式来表示

$$\begin{pmatrix} x_1 & x_2 & x_3 & \cdots\cdots \\ p_1 & p_2 & p_3 & \cdots\cdots \end{pmatrix} \tag{1.11}$$

或(1.11)′

X	x_1	x_2	x_3	$\cdots\cdots$
$P(X=x_i)$	p_1	p_2	p_3	$\cdots\cdots$

称(1.11)或(1.11)′为随机变量 X 的分布列，也称为分布律，简称为分布。

三、 连续型随机变量

如果随机变量 X 的取值充满某一区间，并且 X 的值落在任何一个子区间内的概率都是确定的，这样的随机变量称为连续型随机变量。可用数学的语言来描述其定义如下：

定义 1.6　一个随机变量称为连续型的，如果存在一个非负可积函数 $f(x)$，使得

$$P(a < X < b) = \int_a^b f(x)\mathrm{d}x \tag{1.12}$$

对一切 $-\infty < a < b < +\infty$ 成立，称 $f(x)$ 为 X 的分布密率函数或简称为密度函数、分布密度。

由于事件 $\{-\infty < X < +\infty\}$ 的概率为 1，故

$$\int_{-\infty}^{\infty} f(x) = 1 \tag{1.13}$$

若设连续型随机变量 X 包含某一值 a 的子区间的概率为 p，则随着这个子区间的无限缩小，可知 $p \to 0$，从而求得 $P(X=a)=0$。这就是说，对于连续型的随机变量，它取任何值的概率恒为零，所以描述连续型的随机变量，不能用(1.11)的形式，只能用(1.12)的办法，但有一种统一的方法，它既适用于离散型随机变量，又适用于连续型随机变量，这种方法就是分布函数。

四、 分布函数

定义 1.7　设 X 为随机变量，令

$$F(x) = p(X \leqslant x), \qquad -\infty < x < +\infty \tag{1.14}$$

则称 $F(x)$ 是 X 的分布函数

若 X 是离散型随机变量，(1.11)是它的取值和相应的概率，它的分布函数为

$$F(x) = \sum_{i:x_i \leqslant x} p_i \tag{1.15}$$

由(1.15)可知，$F(x)$ 与(1.11)之间可以相互唯一确定。

若 X 是连续型随机变量，由(1.12)和(1.14)有

$$F(x) \int_{-\infty}^{x} f(t)\mathrm{d}t \tag{1.16}$$

在这种情况下，可以证明 $F(x)$ 和 $f(x)$ 能够相互唯一地确定。这说明分布函数既可以描述离散型随机变是，也可以描述连续型随机变量。

分布函数 $F(x)$ 有如下的性质：

(1)分布函数是非降函数，即当 $x<y$ 时，$F(x) \leqslant F(y)$。这是因为事件 $\{X \leqslant x\} \subset \{Y \leqslant y\}$。

(2)$F(-\infty)=0$，这是因为 $\{X \leqslant -\infty\}$ 是一个不可能事件。

(3)$F(+\infty)=1$，这是因为 $\{X \leqslant +\infty\}$ 是一个必然事件。

(4)X 落在某一区间的概率为

$$p(a < X \leqslant b) = F(b) - F(a) \tag{1.17}$$

除离散型变量和连续型变量外，还有混合型变量。实际上，还存在一类理论上很有价值，然而实际问题中很少出现的"奇异型"分布函数，这些我们均不作介绍了。

定义 1.8　两个随机变量 X 和 Y 若有相同的分布函数，则记作为 $X \stackrel{d}{=} Y$。

有相同分布函数的两个随机变量，可以代表完全不同的实际问题。

随机变量的函数(若有意义)仍是随机变量。如 X 是随机变量，则 X^2、e^x，$1/(1+x^2)$，$\sin X$ 等仍为随机变量。若 X 和 Y 是随机变量，则 $X+Y$、XY、X/Y(若 $Y \neq 0$)等也是随机变量。

§1.3 随机变量的数字特征

为了能够全面地描述随机变量 X 的概率性质，自然应该知道 X 的分布函数。但是对一般随机变量要完全确定它们的分布函数并不容易，而且在许多实际问题中，也不需要知道它们的一切概率性质，只需要知道随机变量的某些特征也就够了，所谓数字特征，就是反映随机变量的某方面的特征的一些数，如平均值，最大可能值等。

一、 数学期望

定义 1.9　若 X 为离散型随机变量，它的概率分布为(1.11)，当 $\sum_{i=1}^{\infty}|x_i|p_i < \infty$ 时，它的数学期望是

$$E(X) = \sum_{i=1}^{\infty} x_i p_i \tag{1.18}$$

它反映了 X 取值的平均，为此，称 $E(X)$为 X 的均值，也称为分布的均值。

定义 1.10　设 X 为连续型随机变量，它的分布密度为 $f(x)$，当 $\int_{-\infty}^{\infty}|x|f(x)\mathrm{d}x < \infty$ 时，它的数学期望是

$$E(X) = \int_{-\infty}^{\infty} xf(x)\mathrm{d}x \tag{1.19}$$

同离散型随机变量一样，连续型随机变量 X 的数学期望 $E(x)$是 X 的可能取值(关于概率)的平均。

由数学期望的定义，可知它有如下的性质：

(i)常数的数学期望等于该常数。若 C 为常数，则 $E(C)=C$

(ii) 若 C 为常数则有

$$E(X + C) = E(X) + C \tag{1.20}$$

(iii)若 C 为常数，则有

$$E(CX) = CE(X) \tag{1.21}$$

(iv)若 X 和 y 是两个随机变量，则

$$E(X \pm Y) = E(X) \pm (E(Y) \tag{1.22}$$

(v)若随机变量 X 和 y 互相独立，则

$$E(XY) = E(X)E(Y) \tag{1.23}$$

二、方　差

除数学期望外，还需要一个反映随机变量取值波动大小的特征数，在实际中应用最广的是方差和标准差。

定义 1.11　若随机变是 X 的数学期望($E(X)$存在，且 $E[X-E(X)]^2<\infty$，则称 $E[X-E$

$(X)]^2$为 X 的方差，记作 Var(X)或 $D(X)$，方差的平方根$\sqrt{D(X)}$ 称为标准差，记为 σ_X，简记为 σ。

当 X 是离散型随机变量时，

$$\mathrm{Var}(X) = \sum_i [x_i - E(X)]^2 \cdot p_i \tag{1.24}$$

当 X 是连续型随机变量时，

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} [x - E(x)]^2 f(x)\mathrm{d}x \tag{1.25}$$

当 Var$(X)=\infty$时，称 X 的方差不存在。

由方差的定义，可以证明它有如下的性质：

(1)如果 X 的方差 Var(x)存在，则

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X) \tag{1.26}$$

其中 a,b 为常数，特别当 $a=0$ 时，有 $Var(X)=0$，

(2)方差计算中，常用下面的公式

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 \tag{1.27}$$

(3)若 X 方差 Var(X)存在，则对任意 $\varepsilon>0$，有

$$p\{|X - E(X)| \geqslant \varepsilon\} \leqslant \mathrm{Var}(X)/\varepsilon^2 \tag{1.28}$$

(4)若 Var$(X)=0$，则 $p(X=E(X))=1$

三、 高阶矩

对数学期望和方差作进一步推广，可得随机变量的高阶矩。

定义 1.12　记 $\mu'_r(b)\equiv E[(X-b)^r]$，称为随机变量 X 关于 b 的 r 阶矩(若存在)。特别，当 $b=0$ 时，称为 X 的 r 阶原点矩，并简记成 μ'_r；当 $b=E(X)$时，称为 X 的 r 阶中心矩，并简记成 μ_r。

显然，$E(X)$一阶原点矩，Var(X)是二阶中心矩，而一阶中心矩恒为零。μ_r 存在当且仅当 μ'_r 存在。若 μ_r 存在则一切 $\mu_s(s<r)$都存在。在三阶和四阶矩中，比较重要的是偏态系数和峰态系数 。

定义 1.13　$\gamma_1=\mu_3/\mu_2^{3/2}$ 称为随机变量 X 的偏态系数；$\gamma_2=\mu_4/\mu_2^2-3$ 称为随机变量的峰态系数；$C_x=\sigma_x/E(x)$称为随机变量 X 的变异系数。

利用 γ_1 和 γ_2，可以检验一个分布是否正态分布或对分布函数进行分类。若随机变量 X 的取值较小，则相应的 $E(X)$ 和 σ_x 均较小；若 X 的取值较大，则相应的 $E(X)$和 σ_x 也较大，直接比较两者的标准差是不合理的，这时采用变异系数就可以互相比较了。

四、 其它数字特征

定义 1.14　众数是指使得频率函数或分布密度达到极大值的点。详细地说，当 X 为离散型随机变量时，若 $p_j\geqslant p_i$ 对一切 $i\neq j$ 成立，则称 x_j 为众数；当 X 为连续型随机变量时，若 $f(x_0)=\max\limits_x f(x)$，则称 x_0 为 X 的众数。

定义 1.15 给定常数 $0<p<1$,若存在 a_p,使得

$$P(X < a_p) \leqslant P \leqslant P(x \leqslant a_p)$$

则称 a_p 为随机变量 X 的 P 分位点。当 $P=\frac{1}{2}$ 时,相应 $a_{1/2}$ 叫做随机变量 X 的中位数。

中位数是分布的"中点",也是刻划随机变量"均值"的一种方法,对于那些数学期望不存在的随机变量,中位数常起了数学期望的作用。

§1.4 矩母函数和特征函数

虽然随机变量可以用分布函数来描述,但有时直接处理分布函数会遇到一些困难,这时可设法把分布函数转化为另外一种形式,使后者比较容易处理。矩母函数和特征函数就是由于这个需要而产生的。

定义 1.16 随机变量 X 的矩母函数 $\mathrm{M}(t)$ 定义为

$$\mathrm{M}(t) = E(\mathrm{e}^{tx}), \qquad -\infty < t < \infty \tag{1.29}$$

这里 e 是自然对数的底。

若 X 为连续型随机变量,其分布密度为 $f(x)$,则

$$\mathrm{M}(t) = \int_{-\infty}^{\infty} \mathrm{e}^{tx} f(x)\mathrm{d}x \tag{1.30}$$

如果上式右边积分存在,并且允许在积分号内进行微分,我们得到 $\mathrm{M}'(t) \int_{-\infty}^{\infty} x\mathrm{e}^{tx} f(x)\mathrm{d}x$,

$\mathrm{M}'(0) = E(x)$; $\mathrm{M}''(t) = \int_{-\infty}^{\infty} x^2\mathrm{e}^{tx} f(x)\mathrm{d}x$, $M''(0) = E(X^2)$ 一般地有 $\mu'_r = \mathrm{M}^{(r)}(0)$, $\mathrm{M}^{(r)}(0)$ 表示 $M(t)$ 的 r 阶导数在 $t=0$ 时的取值。可见,只要有了矩母函数 $M(t)$,通过微分运算可方便地求得各阶原点矩,用这种方法常常比直接计算 $E(X^r)$ 要方便。但可惜不是每一个分布函数都存在矩母函数的,于是人们就寻找对一切分布都存在的工具,由此产生了特征函数。记 $i=\sqrt{-1}$,故 $\mathrm{e}^{itx}=\cos tx + i\sin tx$,而

$$E(\mathrm{e}^{itx}) = E(\cos tx) + iE(\sin tx) \tag{1.31}$$

定义 1.17 随机变量 X 的特征函数定义为 $\Phi(t)=E(\mathrm{e}^{itx})$。当 X 为离散随机变量时

$$\Phi(t) = \sum_k (\cos tx_k)p_k + i\sum_k (\sin tx_k)p_k = \sum_k \mathrm{e}^{itx}p_k \tag{1.32}$$

当 X 为连续型随机变量时

$$\Phi(t) = \int_{-\infty}^{\infty} (\cos tx) f(x)\mathrm{d}x + i\int_{-\infty}^{\infty} (\sin tx) f(x)\mathrm{d}x = \int_{-\infty}^{\infty} \mathrm{e}^{itx} f(x)\mathrm{d}x \tag{1.33}$$

由于 $|\mathrm{e}^{itx}|=[\cos^2 tx+\sin^2 tx]^{1/2}=1$,所以任何随机变量的特征函数都是存在的,特征函数具有很多好的性质,在此仅仅列出而不给出证明。

(1) $\Phi(t)$ 是一个有界的连续函数,$\Phi(t)\leqslant 1$ 对一切 t 成立。

(2) $\Phi(0)=1$

(3)若随机变量 X 有 r 阶矩存在,则 $\mu'_r=\frac{1}{i^r}\Phi^{(r)}(0)$;反之,由 $\Phi^{(r)}(0)$ 存在不一定能保证 X 有 r 阶矩,但可以证明 X 有 $r-1$ 阶矩存在。

(4)若随机变量 X 的各阶矩都存在，则它的特征函数为

$$\Phi(t) = 1 + \sum_{r=1}^{\infty} \mu'_r \frac{(it)^r}{r!} \tag{1.34}$$

(5)若随机变量 X 的特征函数为 $\Phi(t)$，则 $Y=a+bX(a,b$ 为实数)的特征函数为

$$\psi(t) = e^{iat}\Phi(bt)。$$

(6)若 X_1 和 X_2 的两个互相独立的随机变量，它们的特征函数分别 $\Phi_{x_1}(t)$ 和 $\Phi_{x_2}(t)$，则 $Y=X_1+X_2$ 的特征函数为 $\quad \Phi_y(t) = \Phi_{x_1}(t)\cdot\Phi_{x_2}(t)$

(7)特征函数与分布函数是一一对应的。即：如果 X 的分布函数和特征函数分别是 $F(x)$ 和 $\Phi(t)$，Y 的分布函数和特征函数分别是 $G(y)$ 和 $\psi(t)$，则 $F\equiv G$ 且仅当 $\Phi\equiv\psi$，这是一条极为重要的性质，它表明处理特征函数等价于处理分布函数。

§1.5　随机向量及其概率分布

本节是将前几节的内容推广到多个随机变量(称为随机向量)的情形。

一、 随机向量

在有些随机现象中，每次试验的结果需要同时用几个随机变量来描述。例如炮弹在地面的命中点的位置是由两个随机变量(两个坐标)来确定的：飞机(的重心)在空中的位置是由三个随机变量(三个坐标)来确定的、等等。因此需要拓广随机变量的概念至随机向量。

定义 1.18　若 $X_1,X_2,\cdots\cdots,X_n$ 是 n 个随机变量，由它们组成的一个数组 $X=(X_1,X_2,\cdots\cdots,X_n)$，叫做($n$ 维)随机向量。

当 $n=1$ 时，随机向量就化为随机变量，

在实际应用中，最重要的随机向量是离散型的或是连续型的。

定义 1.19　设 $X=(X_1,\cdots\cdots,X_n)$ 为一随机向量，若存在有限个或可列个 n 维数组(向量) $a_1=(a_{11},\cdots\cdots,a_{1n})$，$a_2=(a_{21},\cdots\cdots,a_{2n})$，$\cdots\cdots$，记

$$P(X=a_i) = P(X_1=a_{i1},\cdots\cdots,X_n=a_{in}) = P_i \tag{1.35}$$

且满足 $p_1+p_2+\cdots\cdots=1$，则称 X 为离散型随机向量。

$$\begin{pmatrix} a_1 & a_2 & \cdots\cdots \\ p_1 & p_2 & \cdots\cdots \end{pmatrix} \tag{1.36}$$

称为它的密度矩阵。若存在一个非负函数 $f(x)=f(x_1,\cdots\cdots x_n)$ 使得对一切 $-\infty<a_i\leqslant b_i<\infty$，$i=1,\cdots\cdots n$，均有

$$p(a_1<X_1<b_1,\cdots\cdots,a_n<X_n<b_n) = \int_{a_1}^{b_1},\cdots\cdots,\int_{a_n}^{b_n} f(x_1,\cdots\cdots,x_n)dx_1\cdots\cdots dx_n \tag{1.37}$$

则称 X 为连续型随机向量，$f(x_1,\cdots\cdots,x_n)$ 称为它的分布密度，也称为联合分布密度。

二、 分布函数

离散型随机向量的统计性质由它的密度矩阵完全确定。同样，连续型随机向量由它的分布

密度完全确定。但两者都可以用多元分布函数来描述它们的统计性质。

定义 1.20 设 $X=(X_1,\cdots,X_n)$为一随机向量,对任一 n 维向量$(x_1,\cdots\cdots,x_n)$,令函数

$$F(x_1,\cdots\cdots,x_n)=p(X_1\leqslant x_1,\cdots\cdots,X_n\leqslant x_n) \tag{1.38}$$

称为随机向量 X 分布函数,也称为联合分布函数。

分布函数 $F(x_1\cdots\cdots,x_n)$有如下的性质。

(1)对每个 i,F 是 x_i 的单调不降右连续函数。

(2)$F(-\infty,x_2,\cdots\cdots,x_n)=F(x_1,-\infty,x_3,\cdots\cdots,x_n)=\cdots\cdots=F(x_1,x_2,\cdots\cdots,x_{n-1},-\infty)=0$

(3)$F(\infty,\cdots\cdots,\infty)=1$

(4)设 X 为具有密度矩阵(1.36)的离散型随机向量,则它的分布函数为

$$F(x_1,\cdots\cdots,x_n)=\sum_{a_i\leqslant x}p_i \tag{1.39}$$

(5)设 X 为连续型随机向量,其分布密度为 $f(x_1,\cdots\cdots x_n)$则它的分布函数为

$$f(x_1,\cdots\cdots x_n)=\int_{-\infty}^{x_1}\cdots\cdots\int_{-\infty}^{x_n}f(t_1,\cdots\cdots,t_n)\mathrm{d}t_1\cdots\cdots\mathrm{d}t_n \tag{1.40}$$

三、 边缘分布

设 $X=(X_1,\cdots X_n)$是一个 n 维随机向量,它的部分随机变量组成的子随机向量的分布叫做边缘分布。与边缘分布相对应,称 X 的分布为 $X_1,\cdots,X_n$ 的联合分布。不失一般性,由前 m 个随机变量组成的子随机向量 $X^{(1)}=(X_1,\cdots,X_m)$,于是可得

(1)若 X 的分布函数是 $F(x_1,\cdots,x_n)$,则 $X^{(1)}$ 的分布函数是

$$F_1(x_1,\cdots,x_m)=F(x_1,\cdots,x_m,+\infty,\cdots,+\infty) \tag{1.41}$$

(2)若 X 的分布密度是 $f(x_1,\cdots,x_n)$,则 $X^{(1)}$ 也有分布密度,且等于

$$f_1(x_1,\cdots\cdots,x_m)=\int_{-\infty}^{\infty}\cdots\cdots\int_{-\infty}^{\infty}f(x_1,\cdots\cdots,x_m,x_{m+1}\cdots x_n)dx_{m+1}\cdots dx_n \tag{1.42}$$

从上述两个性质可看到,$X^{(1)}$ 的分布可完全由 X 的分布函数决定。但应注意:不同的分布函数可以有相同的边缘分布函数。也就是说,由相同的边缘分布可以构造出不同的联合分布。近年来,这个问题已成为一个专门的分支,概率统计学家们创造的许多方法,从边缘分布来构造联合分布使其满足希望的性质。

四、 独立性

定义 1.21 设 $X=(X_1,\cdots\cdots,X_n)$分布函数为 $F(x_1\cdots\cdots,x_n)$,$X_1,\cdots\cdots,X_n$ 的边缘分布函数分别为 $F_1(x_1),\cdots\cdots,F_n(x_n)$,若对任意实数 $x_1,\cdots\cdots,x_n$ 有

$$F(x_1,\cdots\cdots,x_n)=F_1(x_1),\cdots\cdots,F_n(x_n) \tag{1.43}$$

则称 $X_1,\cdots\cdots,X_n$ 相互独立。

独立性定义对于离散型和连续型随机向量分别有如下等价形式:

定理 1.4 设 $X=(X_1,\cdots\cdots,X_n)$为离散型随机向量,则 $X_1,\cdots\cdots,X_n$ 相互独立的充要条件是

$$P(X_1=a_1,\cdots\cdots,X_n=a_n)=P(X_1=a_1),\cdots\cdots,P(X_n=a_n) \tag{1.44}$$