参数统计教程

A Course in Parametric Statistics

韦博成　编著

高等教育出版社

参数统计教程

A Course in Parametric Statistics

韦博成　编著

高等教育出版社

内容提要

本书为概率统计专业的研究生教材，全书共分八章，比较全面系统地介绍了：常见的统计分布，充分统计量和信息函数，点估计的基本理论和方法，假设检验的理论、方法及其应用，区间估计及其应用，Bayes统计推断的基本概念和方法等。本书也可作为经济金融、生物医学、管理科学、工程技术等专业研究生的教学参考书，还可供相关专业的大学生、研究生、教师、科技人员和统计工作者参考。

图书在版编目(CIP)数据

参数统计教程／韦博成编著.—北京：高等教育出版社，2006.11

ISBN 7-04-020054-6

Ⅰ.参…　Ⅱ.韦…　Ⅲ.数理统计－高等学校－教材　Ⅳ.O212

中国版本图书馆 CIP 数据核字(2006)第 120772 号

策划编辑　杨　波　　责任编辑　高尚华　　封面设计　李卫青
责任绘图　尹文军　　版式设计　史新薇　　责任校对　王效珍
责任印制　陈伟光

出版发行	高等教育出版社	购书热线	010－58581118
社　　址	北京市西城区德外大街 4 号	免费咨询	800－810－0598
邮政编码	100011	网　　址	http://www.hep.edu.cn
总　　机	010－58581000		http://www.hep.com.cn
		网上订购	http://www.landraco.com
经　　销	蓝色畅想图书发行有限公司		http://www.landraco.com.cn
印　　刷	北京宝旺印务有限公司	畅想教育	http://www.widedu.com
开　　本	787×960　1/16	版　　次	2006 年 11 月第 1 版
印　　张	26.25	印　　次	2006 年 11 月第 1 次印刷
字　　数	440 000	定　　价	35.90 元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换。

物料号　20054-00

前　言

本书为概率统计专业的研究生教材，比较全面系统地介绍了数理统计的基本原理、基本方法及其应用。本书也可作为经济金融、生物医学、管理科学、工程技术等专业研究生的教学参考书，还可供相关专业的大学生、研究生、教师、科技人员和统计工作者参考。阅读本书需要有数学分析、线性代数，特别是概率论方面的基础知识，但不需要测度论方面的知识。掌握本书的内容，即可进一步学习统计学其他各分支的理论与方法，也可比较顺利地理解其他学科中用到的统计学基本概念。

本书定位为“中等水平、便于阅读、内容充实、有一定特色”的教材。希望在不很高的起点上，对数理统计的基本理论和方法有比较清楚、深入的阐述，对数理统计的实际背景和应用有适当的介绍。全书共分八章，包罗了通常高等数理统计的主要内容。第一章介绍常见的统计分布及有关问题，也介绍了非中心 Γ 分布、带有多余参数的指数族分布等内容；由于任何统计问题都涉及统计分布，因此本书单列一章，用较大篇幅介绍这方面的内容，这使仅有本科概率论基本知识的读者，对数理统计的研究对象有更多的了解，以便于后面集中精力去领会数理统计的基本概念和方法。第二章介绍充分统计量和分布族的信息函数，也介绍了 Basu 定理、Kullback 信息等内容；本章内容也相对较多，因为几乎任何统计问题都与充分统计量、样本信息等基本概念有关，本书也单列一章，用较大篇幅比较系统地介绍这些基本概念，这对非数学专业的读者进一步学习数理统计可能更加有用。本书第三至第五章介绍点估计的基本理论和方法，第三章介绍常用的点估计方法，也介绍了不变原理、子集参数的似然等内容，同时也配备了比较丰富的例题与习题；第四章单列一章，以比较初等的方法系统地介绍了“同变估计”及其求解方法；第五章介绍点估计的基本性质，也介绍了广义 C-R 型不等式等内容。第六章篇幅最大(与点估计的 3 章相当)，比较全面系统地介绍了假设检验的基本概念、基本理论和基本方法，并附有较多的应用方面的例题与习题；本章还比较详细地介绍了有广泛应用价值的 score 检验统计量。第七章介绍常用的区间估计方法及其应用，也介绍了单调似然比分布族参数的区间估计方法。第八章以较大的篇幅介绍 Bayes 统计推断的基本概念和方法，除了常见的参数估计与假设检验的 Bayes 方法

以外，本章还从同变原理出发，比较详细地介绍了位置、尺度参数分布族无信息先验分布的选取准则；同时也比较详细地阐述了 HPD 可信域的基本性质和求解方法。

本书初稿为我校研究生课程讲义，经作者多次讲授、不断修改、逐步形成；在成书过程中又作了全面的充实、加工与修订。厦门大学王海斌博士校阅了全书，并且提出了许多修改意见；香港中文大学博士生周影辉仔细校阅了全书，并且帮助演算了本书习题，提出了许多修改意见；博士生陆建也演算了部分习题；南京农业大学解锋昌副教授为本书绘制了全部图形；研究生章珏帮助打印了全书初稿，特此表示衷心的感谢！另外，在本书写作过程中，参考了国内外许多图书资料，受益匪浅，一并对这些作者表示衷心的感谢！

本书在写作过程中，自始至终得到高等教育出版社的关心与帮助，特别要感谢高等理工出版中心、数学分社的杨波同志，他们对本书的写作、审定与出版都给予大力的支持与帮助，特此表示衷心的感谢！同时也对审稿先生的厚爱与支持与表示衷心的感谢！

由于作者水平有限，难免有不妥与谬误之处，恳请同行专家和广大读者提出批评和建议。

韦博成

2006 年 6 月于东南大学

目 录

第一章 统计分布基础

对于一个随机变量，其分布函数完全描述了它的概率结构. 但在实际问题中，分布函数常常是未知的. 数理统计的中心任务，就是通过样本观察值，对总体的分布函数以及由此产生的问题进行合理的推断. 统计问题的基本提法是：给定样本 $X_1,\cdots,X_n$；大多情形为独立同分布(即 *i. i. d.*)样本，这时 X_1 服从未知分布 $F(x_1)$或 $F_\theta(x_1)$，θ 未知；如 X_1 服从正态分布 $N(\mu,\sigma^2)$，$\theta=(\mu,\sigma^2)$，或 Poisson 分布 $P(\lambda)$等等. 从样本 $X=(X_1,\cdots,X_n)$出发，对未知分布 $F(x_1)$或 $F_\theta(x_1)$进行统计推断(诸如参数估计、假设检验、回归分析等). 因此，随机变量及其分布函数，既是统计推断的目的，也是统计推断的基础. 一切统计问题都离不开随机变量及其分布，熟练掌握这方面的知识，对学好数理统计并应用于实践是非常必要和有益的，读者必须十分熟悉. 关于概率论的基础知识，读者可参见王梓坤(1979)，李贤平(1997)以及严士健，刘秀芳(1994).

本章第 1.1 节回顾了分布函数的基本性质；第 1.2、第 1.3 节分别介绍常见的离散型分布和连续型分布；第 1.4 节介绍一元非中心 Γ 分布及其有关分布；第 1.5 节介绍指数族分布以及带有多余参数的指数族分布；第 1.6 节介绍次序统计量的分布，并特别介绍了均匀分布和指数分布次序统计量的性质. 关于常见的统计分布，可参见方开泰，许建伦(1987)，Zacks(1981)，茆诗松等(1998).

1.1 随机变量及其分布函数

1.1.1 分布函数与分布密度

给定 n 维随机变量 X，其相应的概率空间记为 $X\sim(\mathscr{X},\mathcal{B}_X,P)$，其中 $\mathscr{X}$ 表示 $\mathbf{R}^n$ 上 X 的样本空间，$X\in\mathscr{X}$；$\mathcal{B}_X$ 表示 $\mathbf{R}^n$ 上的 Borel 域；P 表示$\mathcal{B}_X$ 上对应于 X 的概率测度，即对 $A\in\mathcal{B}_X$，$P(A)=P(X\in A)$表示 X 属于 A 的概率. 概率测度 P 也常表示为 P_θ、P_X 或 P_θ^X.

今设 $X\in\mathbf{R}^1$，则一元分布函数定义为

$$F(x)=P(X\leqslant x)=P\{X\in(-\infty,x]\}.$$

熟知，分布函数 $F(x)$ 为定义在 $(-\infty,+\infty)$ 上的非降，右连续函数，且有 $F(-\infty)=0$，$F(+\infty)=1$，并且满足 $F(x+0)=F(x)$，$F(x-0)=P(X<x)$，$P(X=x)=F(x)-F(x-0)$；若 x 为 $F(x)$ 的连续点，则 $P(X=x)=0$. 另外，单调函数 $F(x)$ 可扩张为 $\mathbf{R}^1$ 上的测度 $F(\cdot)$，且有 $F(A)=P(A)$（见王梓坤，1979）.

分布函数通常可分为绝对连续型、离散型和奇异型. 若存在密度函数 $f(x)$，使 $F(x)=\int_{-\infty}^{x}f(y)\mathrm{d}y$，$F'(x)=f(x)$，则称 $F(x)$ 为绝对连续型，且有 $P(X\in A)=\int_A f(y)\mathrm{d}y$. 若 $F(x)$ 连续但不存在密度函数，则称 $F(x)$ 为奇异型；奇异型分布实际中很少见到. 若随机变量 X 取有限或可数个值，则其分布函数为离散型；这时 $P(X=x_k)=p_k \stackrel{\mathrm{d}}{=} f(x_k)$，$k=0,\pm1,\pm2,\cdots$，$\sum_k p_k=1$. 例如 Poisson 分布：$P(X=k)\stackrel{\mathrm{d}}{=}f(k)=\mathrm{e}^{-\lambda}\cdot\lambda^k/k!=p_k(k=0,1,2,\cdots)$. 退化分布是离散型分布一个重要特例，这时 $P(X=0)=1$ 或 $P(X=a)=1$；其分布函数为

$$F(x)\stackrel{\mathrm{d}}{=}\Delta(x)\stackrel{\mathrm{d}}{=}\begin{cases}1, & x\geqslant 0,\\ 0, & x<0\end{cases}$$

或 $F(x)=\Delta(x-a)=I\{x\geqslant a\}$，其中 $I\{x\geqslant a\}$ 为示性函数. 对于一个集合 A，$I_A(x)\stackrel{\mathrm{d}}{=}I\{A\}$ 定义为 $I_A(x)=1$，$x\in A$；$I_A(x)=0$，$x\overline{\in}A$. 因此，一般情形下离散型分布的分布函数可表示为

$$F(x)=\sum_k p_k\Delta(x-x_k)=\sum_{x_k\leqslant x}p_k=\sum_{x_k\leqslant x}f(x_k). \qquad (1.1.1)$$

离散型分布的分布函数亦可表示为积分形式，今简要介绍 Lebesgue－Stieltjes 积分如下（可参见 Cramer，1946 第七章和第九章；Rao，1973 第二章）.

以下介绍 $\mathbf{R}^1$ 上的积分，$\mathbf{R}^n$ 上的积分完全类似. 给定 $\mathbf{R}^1$ 上的有界函数 $g(x)$ 和有限测度 $\mu(\cdot)$，将一个有穷或无穷区间 $[a,b]$ 分划为有限个互不相交的可测集 $\Delta x_i(i=1,\cdots,n)$ 之和. 对任一“分划”，记 $S_m=\sum_{i=1}^{n}m_i\mu(\Delta x_i)$ 和 $S_M=\sum_{i=1}^{n}M_i\mu(\Delta x_i)$，其中 $\mu(\Delta x_i)$ 为 Δx_i 的测度，m_i 和 M_i 分别为 $g(x)$ 在 Δx_i 上的下确界与上确界. 若 $\max_{1\leqslant i\leqslant n}\{\mu(\Delta x_i)\}\to 0$ 时，S_m 和 S_M 的极限存在且相等，则其极限称为函数 $g(x)$ 在区间 $[a,b]$ 上关于测度 $\mu(\cdot)$ 的 Lebesgue－Stieltjes 积分，并记为 $\int_a^b g(x)\mathrm{d}\mu(x)$ 或 $\int_a^b g(x)\mu(\mathrm{d}x)$. 以下介绍与分布函数有关的两种情况.

(1) 若 $\mu(\cdot)$ 为 Lebesgue 测度，Δx_i 为区间，则 $\mu(\Delta x_i)=\Delta x_i$ 为区间长度；$\int_a^b g(x)\mathrm{d}\mu(x) = \int_a^b g(x)\mathrm{d}x$．因此若 $F(x)$ 为绝对连续型分布函数，存在密度函数 $f(x)$，则有

$$F(x) = \int_{-\infty}^{x} f(y)\mathrm{d}y = \int_{-\infty}^{x} f(y)\mathrm{d}\mu(y). \tag{1.1.2}$$

(2) 若 $\mu(\cdot)$ 为计数测度．设点列 $\{x_k\}$ 有可数个值 $x_1,x_2,\cdots,x_k,\cdots$，计数测度 $\mu(A)$ 表示 A 所包含点列 $\{x_k\}$ 中点的个数．因此对区间 $\Delta x_i\to 0$ 有 $\mu(\Delta x_i)=1$，若 Δx_i 中包含一个 $\{x_k\}$ 中的点；$\mu(\Delta x_i)=0$，若 Δx_i 不包含 $\{x_k\}$ 中的点．对于定义在点列 $\{x_k\}$ 上的函数 $g(x)$，则由 Lebesgue - Stieltjes 积分的定义可知 $\int_a^b g(x)\mathrm{d}\mu(x) = \sum\limits_{x_k\in[a,b]} g(x_k)$．特别，对于离散型分布的分布函数 $F(x)$，由(1.1.1)式可知，若在点列 $\{x_k\}$ 上定义计数测度以及 $f(x_k)=p_k=P(X=x_k)$，则有

$$F(x) = \sum_{x_k\leqslant x} f(x_k) = \int_{-\infty}^{x} f(y)\mathrm{d}\mu(y). \tag{1.1.3}$$

比较(1.1.2)式和(1.1.3)式，通常把 $P(X=x_k)=p_k=f(x_k)$ 称为离散型分布的密度函数．例如 Poisson 分布的密度函数为 $f(x_k)=\mathrm{e}^{-\lambda}\cdot\lambda^{x_k}/x_k!$.

因此对于绝对连续型分布和离散型分布，一个函数的积分，即数学期望可表示为

$$E[\phi(X)] = \int \phi(x)f(x)\mathrm{d}\mu(x) = \int \phi(x)\mathrm{d}F(x),$$

其中 $\mathrm{d}F(x)=f(x)\mathrm{d}\mu(x)$.

特别，概率 $P(A)$ 可表示为 $P(A) = \int_X I_A(x)f(x)\mathrm{d}\mu(x) = \int_A f(x)\mathrm{d}\mu(x) = \int_A \mathrm{d}F(x)$.

另外，Lebesgue - Stieltjes 积分可推广到 n 维空间，可参见 Cramer (1946,第九章)．特别，若 $\mu(\cdot)$ 就取为概率测度 $P(\cdot)$，则有 $P(A) = \int_X I_A(x)\mathrm{d}P(x) = \int_A \mathrm{d}P(x)$；也有 $E[\phi(X)] = \int_X \phi(x)\mathrm{d}P(x)$.

1.1.2 反函数及分位数

在理论与实际问题中，经常要知道分布函数 $F(x)$ 的反函数的值，即 $F^{-1}(p)$ 的值，其中 $0<p<1$，(如 $p=0.05, p=0.95$ 等)；通常称为分位数或分位点．由于 $F(x)$ 非降，不一定连续，因此 $F^{-1}(p)$ 的值可能不存在或不唯一．以下定义保证了分位数的存在性和唯一性.

定义 1.1.1 分布函数 $F(x)$ 的 p 分位数或 p 分位点，定义为 $x_p \triangleq \inf\{x; F(x) \geqslant p\}$. ▌

由下确界的定义知 x_p 存在且唯一. 也容易证明，分位数 x_p 满足

i) 若 $x' < x_p$，则 $F(x') < p$；若 $x' \geqslant x_p$，则 $F(x') \geqslant p$；

ii) $F(x_p - 0) \leqslant p \leqslant F(x_p)$；若 x_p 为 $F(x)$ 的连续点，则 $F(x_p) = p$.

常用的分位点有 $x_{0.05}$，$x_{0.1}$，$x_{0.9}$，$x_{0.95}$，$x_{0.5}$ 等. $x_{0.5}$ 通常称为中位数，它有以下重要性质.

定理 1.1.1 函数 $g(c) = E|X - c|$ 在中位数 $c = x_{0.5}$ 时达到最小值.

证明 设 X 的分布函数为 $F(x)$，若 $c < x_{0.5}$，则

$$
\begin{aligned}
g(c) &= \int_{-\infty}^{+\infty} |x - c| \mathrm{d}F(x) \\
&= \int_{(-\infty, c)} (c - x) \mathrm{d}F(x) + \left(\int_{[c, x_{0.5})} + \int_{[x_{0.5}, +\infty)} \right) (x - c) \mathrm{d}F(x) \\
&= \int_{(-\infty, c)} [(x_{0.5} - x) - (x_{0.5} - c)] \mathrm{d}F(x) + \int_{[x_{0.5}, +\infty)} (x - x_{0.5} + x_{0.5} \\
&\quad - c) \mathrm{d}F(x) + \int_{[c, x_{0.5})} [(x_{0.5} - x) + 2(x - c) - (x_{0.5} - c)] \mathrm{d}F(x) \\
&= \int_{-\infty}^{+\infty} |x - x_{0.5}| \mathrm{d}F(x) + 2\int_{[c, x_{0.5})} (x - c) \mathrm{d}F(x) \\
&\quad + (x_{0.5} - c)[P\{X \geqslant x_{0.5}\} - P\{X < x_{0.5}\}].
\end{aligned}
$$

由 $c < x_{0.5}$ 及 $x_{0.5}$ 的定义知 $P\{X \geqslant x_{0.5}\} \geqslant \frac{1}{2}$ 和 $P\{X < x_{0.5}\} = F(x_{0.5} - 0) \leqslant \frac{1}{2}$，从而 $g(c) \geqslant E|X - x_{0.5}|$；反之，若 $c > x_{0.5}$，则

$$
\begin{aligned}
g(c) &= \int_{-\infty}^{+\infty} |x - c| \mathrm{d}F(x) = \left(\int_{(-\infty, x_{0.5}]} + \int_{(x_{0.5}, c]} \right) (c - x) \mathrm{d}F(x) \\
&\quad + \int_{(c, +\infty)} (x - c) \mathrm{d}F(x) \\
&= \int_{(-\infty, x_{0.5}]} [(x_{0.5} - x) + (c - x_{0.5})] \mathrm{d}F(x) + \int_{(c, +\infty)} [(x - x_{0.5}) \\
&\quad - (c - x_{0.5})] \mathrm{d}F(x) + \int_{(x_{0.5}, c]} [(x - x_{0.5}) + 2(c - x) \\
&\quad - (c - x_{0.5})] \mathrm{d}F(x) \\
&= \int_{-\infty}^{+\infty} |x - x_{0.5}| \mathrm{d}F(x) + 2\int_{(x_{0.5}, c]} (c - x) \mathrm{d}F(x) \\
&\quad + (c - x_{0.5})[P\{X \leqslant x_{0.5}\} - P\{X > x_{0.5}\}].
\end{aligned}
$$

由 $c > x_{0.5}$ 及 $x_{0.5}$ 的定义知 $P\{X \leqslant x_{0.5}\} \geqslant \frac{1}{2}$ 和 $P\{X > x_{0.5}\} \leqslant \frac{1}{2}$，从而

$g(c) \geqslant E|X - x_{0.5}|$. 总之，对于任何 c，$g(c) \geqslant g(x_{0.5})$，且当 $c = x_{0.5}$ 时等号成立. ∎

1.1.3 特征函数和数字特征

分布函数虽然完全描述了一个随机变量的概率结构，但也存在一些不能令人满意的地方. 比如，只能保证分布函数是单调非降、有界、右连续的，而不能保证它是一致连续和绝对连续的. 因此分布函数的分析性质并不完美，通常引进其他工具作为对分布函数的有益补充.

随机变量 X 的数学期望或均值为 $\mathrm{E}(X) = \mu$；方差为 $\mathrm{Var}(X) = \sigma^2$，而 σ 通常称为均方差或标准差. 矩母函数和特征函数分别定义为 $M(t) \stackrel{\mathrm{d}}{=} \mathrm{E}(\mathrm{e}^{tX})$ 和 $\varphi(t) \stackrel{\mathrm{d}}{=} \mathrm{E}(\mathrm{e}^{\mathrm{i}tX}) = M(\mathrm{i}t)$. 常用的性质有

i) $M^{(k)}(0) = \mathrm{E}(X^k) \stackrel{\mathrm{d}}{=} a_k$，$\varphi^{(k)}(0) = \mathrm{i}^k a_k$，$\mathrm{E}(X^k) = \mathrm{i}^{-k}\varphi^{(k)}(0)$；

ii) $\varphi(t) = 1 + \sum\limits_{k=1}^{\infty} a_k \dfrac{(\mathrm{i}t)^k}{k!}$，$M(t) = 1 + \sum\limits_{k=1}^{\infty} a_k \dfrac{t^k}{k!}$；

iii) 给定随机向量 $\xi = (\xi_1, \cdots, \xi_n)$，其分量 $\xi_1, \cdots, \xi_n$ 相互独立的充要条件为

$$\varphi_{\xi}(t_1, \cdots, t_n) = \varphi_{\xi_1}(t_1)\varphi_{\xi_2}(t_2)\cdots\varphi_{\xi_n}(t_n).$$

由性质 ii) 可知，特征函数和矩母函数的 Taylor 展开式的系数为随机变量 X 的各阶矩. 近年来，特征函数对数的 Taylor 展开式的系数也经常用到，今介绍如下.

定义 1.1.2 累积量(cumulant). $\log \varphi(t)$ 的 Taylor 展开式的系数称为累积量，若

$$\log \varphi(t) = \sum_{r=1}^{\infty} \kappa_r \frac{(\mathrm{i}t)^r}{r!},$$

则系数 κ_r 称为 r 阶累积量或半不变量. ∎

随机变量 X 的累积量 κ_r 与其各阶矩 a_k 有密切的关系. 其常用性质有

i) 累积量与各阶矩可互相表示. 其前 3 阶矩的关系为：

$\kappa_1 = a_1$，$\kappa_2 = m_2 = \mathrm{E}(X - \mathrm{E}X)^2$，$\kappa_3 = a_3 - 3a_1a_2 + 2(a_1)^3$；

$a_1 = \kappa_1$，$a_2 = \kappa_2 + (\kappa_1)^2$，$a_3 = \kappa_3 + 3\kappa_1\kappa_2 + (\kappa_1)^3$；

ii) 若 X，Y 独立，则有 $\kappa_r(X+Y) = \kappa_r(X) + \kappa_r(Y)$；

iii) $\kappa_r(X + C) = \kappa_r(X)$，$(r > 1)$，其中 C 为任意常数.

以下数字特征也是统计学中常见的.

随机变量 X 的三阶中心矩与四阶中心矩分别反映了密度函数的偏度与峰度. 其定义如下：

偏度系数 $\gamma_1 \stackrel{\mathrm{d}}{=} \alpha_3/\sigma^3$；峰度系数 $\gamma_2 \stackrel{\mathrm{d}}{=} \alpha_4/\sigma^4 - 3$；其中 $\alpha_k = \mathrm{E}(X -$

$EX)^k$；$\sigma^2=\mathrm{Var}(X)$. 容易验证，若 X 为 正态分布，则有 $\gamma_1=0$，$\gamma_2=0$. 因此正态分布可作为偏度与峰度的比较标准. 另外，密度函数的峰值称为众数(mode).

以下与条件期望有关的公式非常重要，今后将多次用到.

$$\mathrm{E}(X)=\mathrm{E}_T\{\mathrm{E}(X|T)\};\tag{1.1.4}$$

$$\mathrm{Var}(X)=\mathrm{E}_T\{\mathrm{Var}(X|T)\}+\mathrm{Var}_T\{\mathrm{E}(X|T)\}.\tag{1.1.5}$$

另外，若 $X=(X_1,\cdots,X_n)^{\mathrm{T}}$，$Y=(Y_1,\cdots,Y_m)^{\mathrm{T}}$ 为随机向量，则其期望与方差定义为 $\mathrm{E}(X)\stackrel{\mathrm{d}}{=}(\mathrm{E}X_1,\cdots,\mathrm{E}X_n)^{\mathrm{T}}$, $\mathrm{Cov}(X,Y)\stackrel{\mathrm{d}}{=}(\sigma_{ij})_{n\times m}$；$\sigma_{ij}=\mathrm{Cov}(X_i,Y_j)$；$\mathrm{Var}(X)\stackrel{\mathrm{d}}{=}\mathrm{Cov}(X,X)_{n\times n}$. 随机向量的期望与方差的常用公式有

i) $\mathrm{Cov}(X,Y)=\mathrm{E}[(X-\mathrm{E}X)(Y-\mathrm{E}Y)^{\mathrm{T}}]=\mathrm{E}(XY^{\mathrm{T}})-(\mathrm{E}X)(\mathrm{E}Y)^{\mathrm{T}}$；
$\mathrm{Var}(X)=\mathrm{E}(XX^{\mathrm{T}})-(\mathrm{E}X)(\mathrm{E}X)^{\mathrm{T}}$；

ii) $\mathrm{Cov}(AX,BY)=A\mathrm{Cov}(X,Y)B^{\mathrm{T}}$；

iii) $\mathrm{E}\|X\|^2=\|\mathrm{E}X\|^2+\mathrm{tr}[\mathrm{Var}(X)]$； (1.1.6)

iv) 设 $X=\begin{pmatrix}X_1\\X_2\end{pmatrix}$，则 $\mathrm{Var}(X)=\begin{pmatrix}\mathrm{Var}(X_1) & \mathrm{Cov}(X_1,X_2)\\ \mathrm{Cov}(X_2,X_1) & \mathrm{Var}(X_2)\end{pmatrix}$.

1.1.4 经验分布函数

经验分布函数是分布函数很好的近似. 考虑独立同分布的样本 $X_1,\cdots,X_n$，这时 X_1 服从未知分布 $F(x)=P(X_1\leqslant x)$. 给定 x，记 ν_n 表示 $X_1,\cdots,X_n$ 中 $\leqslant x$ 的个数. 易见，概率 $F(x)=P(X_1\leqslant x)$ 可用频率 ν_n/n 来逼近.

定义 1.1.3 给定独立同分布的样本 $X_1,\cdots,X_n$ 和 x，经验分布函数定义为 $F_n(x)\stackrel{\mathrm{d}}{=}n^{-1}\nu_n$. ▮

易见，经验分布函数可表示为

$$F_n(x)=n^{-1}\sum_{i=1}^{n}I\{X_i\leqslant x\}\stackrel{\mathrm{d}}{=}n^{-1}\sum_{i=1}^{n}Y_i,\tag{1.1.7}$$

其中 $Y_1,\cdots,Y_n$ 为独立同分布，取 0，1 两个值的随机变量；即 $P(Y_i=1)=P\{X_i\leqslant x\}=F(x)$，$P(Y_i=0)=P\{X_i>x\}=1-F(x)$. 并且有 $\mathrm{E}(Y_i)=F(x)$，$\mathrm{Var}(Y_i)=F(x)[1-F(x)]$.

定理 1.1.2 若 $X_1,\cdots,X_n$ 为 $i.i.d.$ 样本，$X_1\sim F(x)$，则对 $\forall x$ 有

(1) $F_n(x)\to F(x)(a.e.)$；

(2) $\sqrt{n}\{F_n(x)-F(x)\}\xrightarrow{L}N(0,F(x)[1-F(x)])$，

其中$(a.e.)$表示几乎处处收敛，$\xrightarrow{L}$表示依分布收敛.