GONGKELEI YANJIUSHENG
SHUXUE KECHENG FUDAOSHU

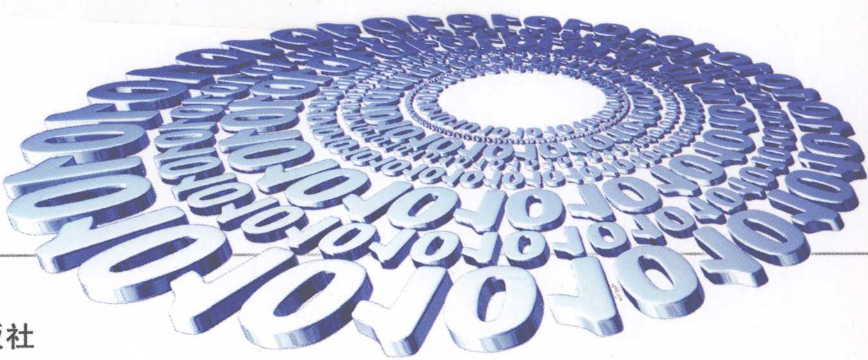
工科类研究生数学课程辅导书

高等工程数学学习指导

[下]

胡庆军　冯良贵/ 编著

GAODENG GONGCHENG SHUXUE
XUEXI ZHIDAO



国防科技大学出版社

工科类研究生数学课程辅导书

高等工程数学

学习指导

（下）

国防科技大学出版社
·长沙·

内容简介

本书是工科类硕士研究生数学公共课程《高等工程数学》的学习参考书,分上、下两册。上册(矩阵理论)包括五章:线性空间与线性变换、方阵的相似化简、矩阵分析及其应用、矩阵分解及其应用、矩阵的广义逆及其应用;下册(数理统计)包括六章:统计量及其分布、参数估计、假设检验、线性回归分析、方差分析、多元统计分析简介。

本书是广大工科类(包括工程类)硕士研究生在学习研究生课程《高等工程数学》或《矩阵理论》、《数理统计》的学习辅导书,也适合理工科院校高年级本科生阅读,并可供青年科技工作者和工科院校有关专业教师查阅和参考。

图书在版编目(CIP)数据

高等工程数学学习指导(下)/胡庆军,冯良贵编著.—长沙:国防科技大学出版社,2010.2

ISBN 978-7-81099-603-7

Ⅰ.高… Ⅱ.①胡… ②冯… Ⅲ.工程数学—研究生—教学参考资料 Ⅳ.TB11

中国版本图书馆 CIP 数据核字(2009)第 022493 号

国防科技大学出版社出版发行

电话:(0731)84572640　邮政编码:410073

http://www.gfkdcbs.com

责任编辑:耿　筠　　责任校对:唐卫葳

新华书店总店北京发行所经销

国防科技大学印刷厂印装

*

开本:787×960　1/16　印张:30.25　字数:576 千

2010 年 2 月第 1 版第 1 次印刷　印数:1-1000 册

ISBN 978-7-81099-603-7

全套定价:76.00 元

前　言

本书是工科类硕士研究生数学公共课程《高等工程数学》的学习参考书，分上、下两册。上册（矩阵理论）包括五章：线性空间与线性变换、方阵的相似化简、矩阵分析及其应用、矩阵分解及其应用、矩阵的广义逆与 Kronecker 积；下册（数理统计）包括六章：统计量及其分布、参数估计、假设检验、线性回归分析、方差分析、多元统计分析简介。每一章按节细分，每一节的具体安排：内容归纳（包括基本概念、主要结论、补充结论和注释）、例题解析、习题、习题解答或参考答案。

本书是作者担任《高等工程数学》课程讲授十多年的经验积累和总结，并汇集了教科书[1-2]的绝大多数习题，且做了较详细的解答；同时还收集了该课程多年的博士生入学考试试题，并按章、节归纳在各个部分作为习题。上册包括 190 多个例题和 250 余道习题；下册包括 180 多个例题和 310 余道习题，其中带“*”号的例题或习题偏难。另外，在下册的附录中还给出了工科硕士生《高等工程数学》课程学习三套模拟考试题及参考解答和工科博士生入学考试《高等工程数学》二套仿真试题及参考解答。

本书条理清楚，层次分明，表达清晰，并具有如下特点：对于每一节，注释中点出了该节内容的重点或难点；主要结论对应到教科书中的有关引理和定理，补充结论对应到教科书中的有关命题、推论和作者总结出来的常用结论，且主要结论和补充结论中的大多数命题，均可从例题和习题中找到证明方法；部分例题是从主要结

论或补充结论中的命题挑选出来改造而成的，并给出了相应的证明方法，以此来化解工科学生面对数学定理证明难读懂的障碍，启发他们的数学思维，提高他们的数学逻辑推理能力；对部分例题和习题给出了一题多解法，以强调对已学知识的灵活运用和各方面知识的综合运用，扩大读者的知识面；对于证明或推导题，强调解题的切入点和解题思路，且推导过程细致，易读易懂；习题中有许多计算题，以强调读者的计算化简能力。

本书是广大工科类（包括工程类）硕士研究生在学习研究生课程《高等工程数学》或《矩阵理论》、《数理统计》时的学习辅导书，也适合理工科院校高年级本科生阅读，并可供青年科技工作者和工科院校有关专业教师查阅和参考。

由于作者水平有限，纰漏之处在所难免，恳望读者批评指正。

作　者

2009.12

目　录

下册：数理统计

第1章 数理统计的基本概念和抽样分布

1.1 总体、样本和统计量

1.1.1 内容归纳

1. 内容提纲

总体、个体、未知参数、参数空间、样本、样本值、样本容量、统计量、样本均值、样本方差、样本标准差、样本变异系数、样本 k 阶矩、样本 k 阶中心矩、顺序统计量、样本中位数、样本极差、最小顺序统计量、最大顺序统计量、经验分布函数、直方图.

2. 基本概念

1) 所考察的随机现象的某项数量指标取值的全体(或:所研究对象的全体)称为**总体**;总体中的每个元素称为**个体**,每一个个体是一个实数.一个总体对应一个概率分布,总体对应的数量指标通常用大写字母表示,记为 X、Y 或 Z 等.通常会说总体 X 服从某一个分布,即指总体对应的数量指标 X 服从某一个分布,且视 X 是一个随机变量.

2) 已知总体 X 的分布形式, 即 $X \sim F(x;\theta)$,但 θ 是未知的,称 θ 为**未知参数**;未知参数 θ 在对应的总体(或:对应的实际问题)中,应是一个确定的值,只是我们不知道.但是,对应的总体其参数 θ 有一个允许范围(记为 Θ),称 Θ **参数空间**.即在对应的实际问题中,θ 是属于Θ 中的某一点.

3) 重复独立地对总体 X 进行 n 次抽样(对总体 X 进行观测或试验以获取数据),所得结果依次记为 $X_1,X_2,\cdots,X_n$,并称为来自总体 X 的一组**简单随机样本**,简称为**样本**,称 n 为**样本容量**.

4) 设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本，$g(x_1,x_2,\cdots,x_n)$ 是 n 元实值函数，若 $g(X_1,X_2,\cdots,X_n)$ 是随机变量，且不含任何未知参数，则称 $g(X_1,X_2,\cdots,X_n)$ 为**统计量**.

5) 设 $X_1,X_2,\cdots,X_n$ 是总体 X 的样本，常用统计量有：

样本均值：$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$；

样本方差：$S^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$；

样本标准差：$S=\sqrt{S^2}$；

修正的样本方差：$\tilde{S}^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$；

样本 k 阶矩：$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k,\ k=1,2,\cdots$；

样本 k 阶中心矩：$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k,\ k=2,3,\cdots$；

最小顺序统计量：$X_{(1)}=\min\{X_1,X_2,\cdots,X_n\}$；

最大顺序统计量：$X_{(n)}=\max\{X_1,X_2,\cdots,X_n\}$.

6) 称 $\frac{S}{\bar{X}}$ 为**样本变异系数**，它反映了总体 X 的变异系数 $\frac{\sqrt{D(X)}}{E(X)}$（设 $E(X)\neq 0$）的信息，是衡量总体分布散布程度的一个统计量，是以样本均值为单位来度量的，是一个相对量.

7) 设 $x_1,x_2,\cdots,x_n$ 为样本 $X_1,X_2,\cdots,X_n$ 的任一组样本值，将 $x_1,x_2,\cdots,x_n$ 按大小重新排列，记为 $x_{(1)}\leqslant x_{(2)}\leqslant\cdots\leqslant x_{(n)}$，当 $(X_1,X_2,\cdots,X_n)$ 取值 $(x_1,x_2,\cdots,x_n)$ 时，令

$$X_{(k)}=x_{(k)},\quad k=1,2,\cdots,n.$$

则称 $X_{(1)},X_{(2)},\cdots,X_{(n)}$ 为样本 $X_1,X_2,\cdots,X_n$ 的一组**顺序统计量**；称 $X_{(k)}$ 为**第 k 个顺序统计量**（对于 $k=1,2,\cdots,n$）.请注意，$X_{(1)}$ 和 $X_{(n)}$ 就是前面第5)条中出现的最小顺序统计量和最大顺序统计量.

8) **样本中位数**：$m_{0.5}\triangleq\begin{cases}X_{(\frac{n+1}{2})}, & \text{当 } n \text{ 为奇数}\\ \frac{1}{2}[X_{(\frac{n}{2})}+X_{(\frac{n}{2}+1)}], & \text{当 } n \text{ 为偶数}\end{cases}$；

样本极差：$\hat{D}\triangleq X_{(n)}-X_{(1)}$.

9) 记总体 X 的分布函数为 $F(x)$（$=P\{X\leqslant x\}$）（常称总体 $F(x)$），$X_1,X_2,$

$\cdots, X_n$ 是 X 的样本，$x_1, x_2, \cdots, x_n$ 为样本值，对任给定实数 x，记

$$V_n(x) = \text{在 } n \text{ 次重复独立观测中事件}\{X \leqslant x\}\text{出现的次数}$$
$$= \text{样本 } X_1, X_2, \cdots, X_n \text{ 中小于等于 } x \text{ 的个数}.$$

则称函数

$$F_n(x) \triangleq \frac{1}{n} V_n(x), \quad -\infty < x < +\infty$$

为(对应总体 X 的样本容量为 n 的)**经验分布函数**.

10) 设总体 $X \sim f(x)$(密度函数)，$X_1, X_2, \cdots, X_n$ 是其样本，$x_1, x_2, \cdots, x_n$ 为样本值.取区间$[a,b]$，使 $x_i \in (a,b)$，$i = 1,2,\cdots,n$，再等分$[a,b]$，得(要求 $m < n$)：

$$a = a_0 < a_1 < \cdots < a_m = b,$$

其中 $\Delta a_j = a_j - a_{j-1} = \dfrac{b-a}{m}$，$j = 1,2,\cdots,m$.

数出样本值 $x_i \in (a_{j-1}, a_j]$中的个数，记为 n_j，且记

$$f_j = \frac{n_j}{n}, \quad y_j = \frac{f_j}{\Delta a_j}, \quad j = 1,2,\cdots,m.$$

在 xOy 平面的横坐标 Ox 轴上，以区间$(a_{j-1}, a_j]$为底、y_j 为高作一排长方形，称为**直方图**.再在直方图上方大致画一条光滑曲线，则当样本容量 n 很大时，这条光滑曲线近似等于 $f(x)$.

11) 对于二维总体(X,Y)，样本为$(X_1,Y_1),(X_2,Y_2),\cdots,(X_n,Y_n)$，以 $\bar{X}, S_1^2$ 表示总体 X 的样本均值和样本方差；$\bar{Y}, S_2^2$ 表示总体 Y 的样本均值和样本方差.有关的统计量有：

样本协方差：$S_{12} \triangleq \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$；

样本相关系数：$R \triangleq \dfrac{S_{12}}{S_1 S_2}$.

3. 主要结论

1) 记总体 X 的分布函数为 $F(x)(= P\{X \leqslant x\})$，密度函数或分布律为 $f(x)$，$X_1, X_2, \cdots, X_n$ 为 X 的样本.则样本 $X_1, X_2, \cdots, X_n$ 的联合分布函数、联合密度函数(或联合分布律)分别为

$$F^*(x_1, x_2, \cdots, x_n) \triangleq \prod_{i=1}^{n} F(x_i);$$

$$f^*(x_1, x_2, \cdots, x_n) \triangleq \prod_{i=1}^{n} f(x_i).$$

2) $S^2=\frac{1}{n}(\sum_{i=1}^{n}X_i^2-n\bar{X}^2)$.

3) 设总体 X 的 $E(X)=\mu, D(X)=\sigma^2, X_1, X_2, \cdots, X_n$ 为 X 的样本,则

$$E(\bar{X})=\mu,\quad D(\bar{X})=\frac{\sigma^2}{n};$$

$$E(S^2)=\frac{n-1}{n}\cdot\sigma^2,\ E(\tilde{S}^2)=\sigma^2.$$

4) 设总体 X 的分布函数为 $F(x)$,密度函数为 $f(x)$,X 的样本为 $X_1, X_2, \cdots, X_n$,则

$X_{(n)}$ 的分布函数为 $F_{(n)}(x)=[F(x)]^n$;

$X_{(n)}$ 的密度函数为 $f_{(n)}(x)=n\cdot[F(x)]^{n-1}\cdot f(x)$;

$X_{(1)}$ 的分布函数为 $F_{(1)}(x)=1-[1-F(x)]^n$;

$X_{(1)}$ 的密度函数为 $f_{(1)}(x)=n\cdot[1-F(x)]^{n-1}\cdot f(x)$.

4. 补充结论

1) 来自总体 X 的样本 $X_1, X_2, \cdots, X_n$ 具有如下特性:

(ⅰ) $X_1, X_2, \cdots, X_n$ 是相互独立的随机变量(即各次试验之间互不影响);

(ⅱ) 每个 X_i 与总体 X 同分布(有代表性,代表总体 X);

(ⅲ) 当 n 次观测一经完成,则得到一组实数值:$x_1, x_2, \cdots, x_n$,即得 $X_1, X_2, \cdots, X_n$ 的一组观测值,称为**样本值**.

2) $\tilde{S}^2=\frac{n}{n-1}\cdot S^2$.

3) 对于总体 X,样本为 $X_1, X_2, \cdots, X_n$,则当 $n\to\infty$ 时,有

$$\bar{X}\xrightarrow{P}E(X),$$

$$S^2\xrightarrow{P}D(X),$$

$$A_k\xrightarrow{P}E(X^k),\quad k=1,2,\cdots$$

4) 利用顺序统计量 $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ 的观测值 $x_{(1)}, x_{(2)}, \cdots, x_{(n)}$,则经验分布函数可化为

$$F_n(x)=\begin{cases}0, & x<x_{(1)}\\ \frac{k}{n}, & x_{(k)}\leqslant x<x_{(k+1)}, \quad k=1,2,\cdots,n-1.\\ 1, & x\geqslant x_{(n)}\end{cases}$$

也可直接利用顺序统计量 $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$,则经验分布函数可表示为

$$F_n(x)=\begin{cases}0, & x<X_{(1)}\\ \dfrac{k}{n}, & X_{(k)}\leqslant x<X_{(k+1)}, \quad k=1,2,\cdots,n-1.\\ 1, & x\geqslant X_{(n)}\end{cases}$$

5) $V_n(x)\sim b(n,F(x))$(二项分布), $E(F_n(x))=F(x)$.

6) 当 $n\to\infty$ 时, 有 $F_n(x)\xrightarrow{\text{P}}F(x)$.

7)* 当 $n\to\infty$ 时,有 $F_n(x)\xrightarrow{\text{a.s.}}F(x)$.

8)* **格里文科定理**(于 1933 年证得): 若记 $D_n=\sup\limits_{-\infty<x<+\infty}|F_n(x)-F(x)|$, 则

$$P\{\lim_{n\to\infty}D_n=0\}=1.$$

9) 当样本容量 n 很大时,有 $F_n(x)\approx F(x)$.

注: “$\xrightarrow{\text{a.s.}}$”表示几乎处处收敛.

5. 注释

1) 概率论和数理统计俗称概率统计,它是研究大量随机现象的统计规律性的数学学科. 概率论的特点是首先提出随机现象的数学模型并建立相应的数学理论,然后去研究这种模型的性质、特点、规律性,如概率分布、数字特征等. 然而在解决实际问题中,人们并没有掌握随机现象(或随机变量)的概率分布和数字特征.如:人口调查中,某地区每个人的身高这一随机变量(或随机现象)的分布情况是未知的,要经过抽样、观测,进行分析(统计推断)才能得到有关结论.解决这些问题就是数理统计的任务了.

2) 数理统计可描述为: 以概率论为理论基础、以直观背景为向导、以观测数据(对随机现象的观测)为研究对象的一门实用性学科.其最终目的是研究如何有效地收集、整理和分析受随机性影响的数据,并对所考察的问题(随机现象)作出种种合理的估计和判断,为采取决策和行动提供依据和建议.

3) 数理统计的研究内容随着科学技术和生产实践的不断发展而逐步扩大,概括起来大致可分为两大类:

(ⅰ)**数据的收集**,包括抽样技术及试验设计的理论和方法的研究,即研究如何对随机现象进行科学的观测和试验,使获得的数据资料既真实又具有代表性;

(ⅱ)**统计推断**,即研究如何对已取得的观测值进行整理、分析并作出决策的方法,以推断总体的规律性.

限于篇幅,本书只总结统计推断这一类问题.

4) 样本是进行统计推断的依据,而样本本身是一堆“杂乱无章”的原始数据,必须进行处理、加工等方法才能提取有用的信息.在应用中,往往是针对不同的问题,构造样本的适当的函数(在一定条件下,称为统计量),利用这些函数进行统计推断.

5) 统计推断:根据样本值(抽样、观测或试验获得的数据)对所提问题作出估计或判断.基本方法分为两类:参数估计方法和假设检验方法.而参数估计方法又分为参数的点估计方法和区间估计方法;假设检验方法又分为参数的假设检验方法和非参数的假设检验方法.

6) 在实际问题中,对所研究的总体,从了解的程度来说,可视为二类:

(ⅰ)若对总体 X 的分布函数一无所知,或知之甚少,则可用经验分布函数或直方图去近似,或假设检验方法去判断.

(ⅱ)已知总体 X 的分布函数的形式,即已知 $X\sim F(x;\theta)$,但含未知参数 θ.例如:已知总体(身高)$X\sim N(\mu,\sigma^2)$,但 μ,σ^2 是未知参数,要求对未知参数 μ,σ^2 进行统计推断.

7) 样本均值、样本中位数是反映总体均值 $E(X)$的统计量,但样本中位数的理论性质偏难;样本方差、样本标准差、样本极差是反映总体方差 $D(X)$或总体标准差$\sqrt{D(X)}$的统计量,但样本极差的理论性质偏难.

8) 本节的重点是理解和熟练掌握样本、统计量、常用统计量(样本均值、样本方差、样本矩、最小顺序统计量、最大顺序统计量);难点是顺序统计量和经验分布函数的有关理论性质.

1.1.2 例题解析

例1 从下列总体 X:

1) 总体 $X\sim N(\mu,\sigma^2)$;　　2) 总体 $X\sim b(1,p)$(两点分布)

中抽取容量为 n 的样本 $X_1,X_2,\cdots,X_n$.试写出样本的联合密度函数或联合分布律.

解:1) 样本 $X_1,X_2,\cdots,X_n$ 的联合密度函数为

$$f^*(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}\left\{\frac{1}{\sqrt{2\pi}\sigma}\cdot\exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]\right\}$$

$$=(2\pi)^{-\frac{n}{2}}\cdot(\sigma^2)^{-\frac{n}{2}}\cdot\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right],$$

$$-\infty<x_i<+\infty,\ i=1,2,\cdots,n.$$

2) 样本 $X_1,X_2,\cdots,X_n$ 的联合分布律为

$$P\{X_1=x_1,X_2=x_2,\cdots,X_n=x_n\}=\prod_{i=1}^{n}[p^{x_i}(1-p)^{1-x_i}]$$

$$=p^{\sum_{i=1}^{n}x_i}\cdot(1-p)^{n-\sum_{i=1}^{n}x_i},\quad x_i=0\text{ 或 }1,\ i=1,2,\cdots,n.$$

例 2　设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本，$\bar{X}$ 为样本均值，S^2 为样本方差.求证:

$$S^2=\frac{1}{n}\Big(\sum_{i=1}^{n}X_i^2-n\bar{X}^2\Big).$$

证: 由 $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,S^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$，则

$$S^2=\frac{1}{n}\sum_{i=1}^{n}(X_i^2-2X_i\bar{X}+\bar{X}^2)=\frac{1}{n}\Big(\sum_{i=1}^{n}X_i^2-2\bar{X}\cdot\sum_{i=1}^{n}X_i+\sum_{i=1}^{n}\bar{X}^2\Big)$$

$$=\frac{1}{n}\Big(\sum_{i=1}^{n}X_i^2-2n\bar{X}\cdot\frac{1}{n}\sum_{i=1}^{n}X_i+n\bar{X}^2\Big)$$

$$=\frac{1}{n}\Big(\sum_{i=1}^{n}X_i^2-2n\bar{X}^2+n\bar{X}^2\Big)$$

$$=\frac{1}{n}\Big(\sum_{i=1}^{n}X_i^2-n\bar{X}^2\Big).$$

例 3　设总体 X 的 $E(X)=\mu,D(X)=\sigma^2,X_1,X_2,\cdots,X_n$ 为 X 的样本，求证:

1) $E(\bar{X})=\mu,\ D(\bar{X})=\frac{\sigma^2}{n}$;

2) $E(S^2)=\frac{n-1}{n}\cdot\sigma^2,\ E(\tilde{S}^2)=\sigma^2$.

证: 1) $E(\bar{X})=E\Big(\frac{1}{n}\sum_{i=1}^{n}X_i\Big)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=E(X)=\mu,$

$$D(\bar{X})=D\Big(\frac{1}{n}\sum_{i=1}^{n}X_i\Big)=\Big(\frac{1}{n}\Big)^2\sum_{i=1}^{n}D(X_i)$$

$$=\frac{1}{n}D(X)=\frac{\sigma^2}{n}.$$

2) 由 $S^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{1}{n}\Big(\sum_{i=1}^{n}X_i^2-n\bar{X}^2\Big)$,

且
$$E(X_i^2)=D(X_i)+[E(X_i)]^2=\sigma^2+\mu^2,$$

$$E(\bar{X}^2)=D(\bar{X})+[E(\bar{X})]^2=\frac{\sigma^2}{n}+\mu^2,$$

所以
$$E(S^2)=\frac{1}{n}\Big[\sum_{i=1}^{n}E(X_i^2)-nE(\bar{X}^2)\Big]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n}(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right]$$
$$= \frac{n-1}{n} \cdot \sigma^2,$$

又由 $\tilde{S}^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n}{n-1} \cdot \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n}{n-1}S^2$，则

$$E(\tilde{S}^2) = \frac{n}{n-1} \cdot E(S^2) = \frac{n}{n-1} \cdot \frac{n-1}{n} \cdot \sigma^2 = \sigma^2.$$

注：此例说明，样本均值 $\bar{X}$ 的数学期望等于总体 X 的期望；样本均值 $\bar{X}$ 的方差等于总体 X 的方差乘 $\frac{1}{n}$；修正的样本方差 $\tilde{S}^2$ 的数学期望等于总体 X 的方差.

例 4 设总体 $X \sim b(1,p)$（两点分布），$X_1, X_2, \cdots, X_n$ 为 X 的样本，$\bar{X}$ 为样本均值，S^2 为样本方差，$\tilde{S}^2$ 为修正的样本方差. 试求：$E(\bar{X})$，$D(\bar{X})$，$E(S^2)$ 及 $E(\tilde{S}^2)$.

解：已知总体 $X \sim b(1,p)$，则

$$E(X) = p,\ D(X) = p(1-p).$$

由上例知

$$E(\bar{X}) = E(X) = p,$$
$$D(\bar{X}) = \frac{1}{n} \cdot D(X) = \frac{1}{n} \cdot p(1-p),$$
$$E(S^2) = \frac{n-1}{n} \cdot D(X) = \frac{n-1}{n} \cdot p(1-p),$$
$$E(\tilde{S}^2) = D(X) = p(1-p).$$

例 5 设总体 X 的概率密度为

$$f(x) = \frac{1}{2} \cdot e^{-|x|}, \quad -\infty < x < +\infty.$$

$X_1, X_2, \cdots, X_n$ 为总体 X 的简单随机样本，其样本方差为 S^2，则 $E(S^2) =$ ______.

答案：$\frac{2(n-1)}{n}$.

解：由 $E(X) = \int_{-\infty}^{+\infty} x \cdot f(x)\mathrm{d}x = \int_{-\infty}^{+\infty} x \cdot \frac{1}{2} \cdot e^{-|x|}\mathrm{d}x = 0$,

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 \cdot \frac{1}{2} \cdot e^{-|x|}\mathrm{d}x = \int_{0}^{+\infty} x^2 \cdot e^{-x}\mathrm{d}x = 2,$$

则　$D(X)=E(X^2)-[E(X)]^2=2.$

又由　$E(S^2)=\frac{n-1}{n}\cdot D(X)$,

$\Rightarrow$　$E(S^2)=\frac{2(n-1)}{n}.$

例 6　设 $X_1, X_2, \cdots, X_n$ 为总体 X 的样本，$\bar{X}$ 为样本均值，若 X 的二阶矩存在，求证：$X_i-\bar{X}$ 与 $X_j-\bar{X}$ 的相关系数 $\rho=-\frac{1}{n-1}$ $(i\neq j, i, j=1,2,\cdots,n)$.

证：对于 $i\neq j, i, j=1,2,\cdots,n$，$X_i-\bar{X}$ 与 $X_j-\bar{X}$ 的相关系数为

$$\rho=\frac{\operatorname{cov}(X_i-\bar{X}, X_j-\bar{X})}{\sqrt{D(X_i-\bar{X})\cdot D(X_j-\bar{X})}}.$$

记 $D(X)=\sigma^2$，则 $D(X_i)=\sigma^2, i=1,2,\cdots,n$，且

$$\begin{aligned}\operatorname{cov}(X_i, \bar{X}) &= \operatorname{cov}\left(X_i, \frac{1}{n}\sum_{k=1}^{n}X_k\right) = \frac{1}{n}\sum_{k=1}^{n}\operatorname{cov}(X_i, X_k)\\ &= \frac{1}{n}\cdot\operatorname{cov}(X_i, X_i)\quad\text{（用到样本 }X_1, X_2, \cdots, X_n\text{ 的独立性）}\\ &= \frac{1}{n}D(X_i) = \frac{\sigma^2}{n},\end{aligned}$$

$$\begin{aligned}D(X_i-\bar{X}) &= D(X_i)+D(\bar{X})-2\cdot\operatorname{cov}(X_i, \bar{X})\\ &= \sigma^2+\frac{\sigma^2}{n}-\frac{2}{n}\cdot\sigma^2=\frac{n-1}{n}\cdot\sigma^2,\end{aligned}$$

同理，得 $D(X_j-\bar{X})=\frac{n-1}{n}\cdot\sigma^2.$

又由

$$\begin{aligned}\operatorname{cov}(X_i-\bar{X}, X_j-\bar{X}) &= \operatorname{cov}(X_i, X_j-\bar{X})-\operatorname{cov}(\bar{X}, X_j-\bar{X})\\ &= \operatorname{cov}(X_i, X_j)-\operatorname{cov}(X_i, \bar{X})-\operatorname{cov}(\bar{X}, X_j)+\operatorname{cov}(\bar{X}, \bar{X})\\ &= 0-\frac{\sigma^2}{n}-\frac{\sigma^2}{n}+D(\bar{X})\\ &= -\frac{\sigma^2}{n}-\frac{\sigma^2}{n}+\frac{\sigma^2}{n}=-\frac{\sigma^2}{n}.\end{aligned}$$

所以　$$\rho=\frac{-\frac{1}{n}\cdot\sigma^2}{\sqrt{\frac{n-1}{n}\sigma^2\cdot\frac{n-1}{n}\sigma^2}}=-\frac{1}{n-1}.$$

例 7　设总体 X 的分布函数为 $F(x)$，密度函数为 $f(x)$，X 的样本为 $X_1, X_2, \cdots, X_n$，求证：$X_{(n)}$ 的分布函数 $F_{(n)}(x)$ 和密度函数 $f_{(n)}(x)$ 分别为

$$F_{(n)}(x)=[F(x)]^n;$$

$$f_{(n)}(x)=n\cdot[F(x)]^{n-1}\cdot f(x).$$

证：由 $F_{(n)}(x)\triangleq P\{X_{(n)}\leqslant x\}=P\{\max\{X_1,X_2,\cdots,X_n\}\leqslant x\}$

$$=P\{X_1\leqslant x,X_2\leqslant x,\cdots,X_n\leqslant x\}$$

$$=\prod_{i=1}^{n}P\{X_i\leqslant x\}\quad(\text{由样本 } X_1,X_2,\cdots,X_n \text{ 的独立性})$$

$$=[P\{X_1\leqslant x\}]^n\quad(\text{由样本 } X_1,X_2,\cdots,X_n \text{ 是同分布})$$

$$=[F(x)]^n.\ (\text{由每个 } X_i \text{ 与总体 } X \text{ 同分布})$$

则

$$f_{(n)}(x)=[F_{(n)}(x)]'_x=n\cdot[F(x)]^{n-1}\cdot f(x).$$

例 8　假设总体 X 的各阶矩都存在，$X_1,X_2,\cdots,X_n$ 为 X 的样本. 当 $n\to\infty$ 时，求证：

$$\bar{X}\xrightarrow{P}E(X),$$

$$A_k\xrightarrow{P}E(X^k)\quad(k=1,2,\cdots),$$

$$S^2\xrightarrow{P}D(X).$$

证：由于样本 $X_1,X_2,\cdots,X_n$ 是相互独立的，且每个 X_i 与总体 X 同分布. 当 $n\to\infty$ 时，由辛钦大数定律知

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{P}E(X),$$

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k\xrightarrow{P}E(X^k),\ k=1,2,\cdots$$

又由 $S^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2$，

由依概率收敛的性质知，则

$$S^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2\xrightarrow{P}E(X^2)-[E(X)]^2=D(X).$$

例 9　记总体 X 的分布函数为 $F(x)(=P\{X\leqslant x\})$，$X_1,X_2,\cdots,X_n$ 是 X 的样本，对任给定实数 x，$V_n(x)$ 表示在 n 次重复独立观测中事件 $\{X\leqslant x\}$ 出现的次数，$F_n(x)$ 为经验分布函数. 求证：

1) $V_n(x)\sim b(n,F(x))$(二项分布)；

2) $E(F_n(x))=F(x)$；

3) 当 $n\to\infty$ 时，有 $F_n(x)\xrightarrow{P}F(x)$.