

2011年度国家出版基金资助项目



中华医学统计百科全书

◆ 徐天和 / 总主编 ◆

多元统计分册



柳 青 / 主 编

 中国统计出版社
China Statistics Press

R19f.161

20142

2011年度国家出版基金资助项目



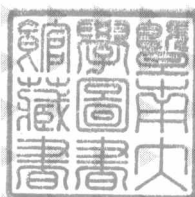
中华医学统计百科全书

◆ 徐天和 / 总主编 ◆

多元统计分册



柳 青 / 主 编



 中国统计出版社
China Statistics Press

图书在版编目(CIP)数据

中华医学统计百科全书. 多元统计分册 / 柳青主编.

-- 北京: 中国统计出版社, 2013. 6

ISBN 978-7-5037-6836-1

I. ①中… II. ①柳… III. ①医学统计—中国—百科全书②医学统计—多元分析 IV. ①R195.1-61

中国版本图书馆 CIP 数据核字(2013)第 126889 号

多元统计分册

作者/柳 青

责任编辑/梁 超

装帧设计/杨 超 李雪燕

出版发行/中国统计出版社

通信地址/北京市西城区月坛南街 57 号 邮政编码/100826

办公地址/北京市丰台区西三环南路甲 6 号 邮政编码/100073

网 址/<http://csp.stats.gov.cn>

电 话/邮购(010)63376907 书店(010)68783172

印 刷/河北天普润印刷厂

经 销/新华书店

开 本/787×1092mm 1/16

字 数/620 千字

印 张/30

版 别/2013 年 6 月第 1 版

版 次/2013 年 6 月第 1 次印刷

定 价/68.00 元

中国统计版图书, 版权所有, 侵权必究。

中国统计版图书, 如有印装错误, 本社发行部负责调换。

《中华医学统计百科全书》

专家指导委员会

主 任	方积乾				
总 主 编	徐天和				
委 员	(以姓氏笔画为序)				
	万崇华	方积乾	王广仪	田小利	田考聪
	苏为华	苏颀龄	周燕荣	柳 青	赵耐青
	饶绍奇	唐 军	徐天和	徐勇勇	徐端正
	景学安	程 琮	颜 虹		

《多元统计分册》

编 委 会

主 编	柳 青				
副 主 编	李 康	林爱华	祁爱琴		
主 审	方积乾				
编 委	(以姓氏笔画为序)				
	于克倩	马 骏	王 彤	王建琼	方 亚
	孙玉文	孙红卫	李 康	刘树仁	刘隆健
	祁爱琴	余松林	陈冠民	郑俊池	张中文
	张岩波	张菊英	杨 珉	林爱华	柳 青
	徐天和	高 永	黄高明	曾 庆	董时富
秘 书 长	高 永				
学术秘书	林爱华	孙红卫			

序 言

国家统计局局长

马建堂

随着时代前进和科学技术的进步,我国的统计科学和医学统计工作的发展进入了一个崭新的阶段。统计科学既是认识社会现象与自然现象数量特征的手段,又是获取信息和进行科学研究的重要工具,历来为人们所重视。自 20 世纪 20 年代起,统计学理论与方法日益广泛地被应用于医学领域。近些年来,随着基因组学、蛋白质组学、药物开发、公共卫生、计算机和信息等学科的迅猛发展,统计学与医学学科的交叉融合不断深入,统计科学在医学领域中的应用与发展提高到了一个新水平。

医学统计是统计科学的重要分支,也是国民经济和社会发展统计的重要组成部分,它关系到人民健康水平的提高和国家的长足发展。医学是强国健民学科,医学研究的对象是人及人群的健康,具有复杂性、特殊性、变异性等特点,这无疑需要全面系统的统计分析方法的支持与帮助。随着统计科学的迅猛发展,一些新的统计方法如遗传统计、多水平模型、结构方程模型、健康量表等不断涌现。一方面这些新的统计方法和理论亟需在医学科学领域内推广应用,为医学发展提供支持和帮助,另一方面,医学科研工作者为了科学研究工作的需要也迫切要求了解和掌握一些最新的、全面系统的统计方法和理论。因此,对当代医学科学研究中的统计分析方法进行全面系统的研究与介绍,是十分重要的一件事情,《中华医学统计百科全书》正是在这样的背景下编纂而成的,它满足了当前医学科学发展的需要,不失为一部好的大型医学统计参考书。

《中华医学统计百科全书》自 2009 年 1 月开始编写,由国内外著名医学院校的统计学教授和专家担任主编和编委,可谓编写力量强大,在编写过程中,他们本着精益求精的精神,精雕细琢,采百家之所长,融国内外华人统计学专家之所成。历时三年,终成其册。本套书内容浩繁,共八个分册,包含描述性统计分册、单变量推断统计分册、多元统计分册、非参数统计分册、管理与健康统计分册、医学研究与临床统计设计分册、健康测量分册和遗传统计分册。各

分册在内容上相互衔接并互为补充,贯穿“从简单到复杂”,“从一般、传统到先进、前沿”的循序渐进的编纂思路,一改目前医学统计著述中普遍存在的方法之间或评价指标之间缺乏相互联系、过于分散和单一的状况,使医学统计理论与方法更加具备了系统性、完整性与时代前沿性。本套书结构严谨,层次分明,科学性强,既突破了传统的辞典式编撰方法,又吸取了辞典的某些特点,在实用性、知识性、可读性、可查性等方面均具独到之处。

《中华医学统计百科全书》适应了我国医学科学研究发展对统计分析方法的需要,本书的出版,势必会大大促进我国现代医学的发展。本书既是我国医学统计工作者、医疗卫生统计信息工作者、高等医学院校师生以及广大医务工作者必备的大型医学统计参考工具书,也适合于医学各不同层次和不同专业的读者阅读。我相信本书的出版,不仅对于促进我国医学统计发展,促进我国与国际生物医学统计间的交流,繁荣社会主义先进文化具有重要意义,而且该书也必定会成为广大医学科学研究工作者的良师益友,故欣然为之作序。

编者的话

近年来,医学统计科学发展迅速,如遗传统计、多水平模型、结构方程模型、健康量表等新的统计理论与方法不断涌现,并被应用到医学科研实践中。这些新的统计理论与方法在医学科学研究中的不断拓展应用,要求广大的医学科技工作者在工作中必须学习和掌握这些新知识。所以,怎样使这些新的统计理论与方法易于被广大的医学科技工作者接受和使用,以提高医疗卫生工作质量,成为统计学专家的首要解决的任务。为此,组织编纂一部适合于广大医学科技工作者学习和使用的工具书,成为当前形势之必需。《中华医学统计百科全书》(下文简称“全书”)正是基于这样的背景而孕育产生的。

编纂“全书”的想法一经提出,就得到了国内高等医学院校和科研院所的统计学专家们的赞同。专家们云集一堂,进行商讨,达成共识——要集全国高等医学院校和科研院所的统计学专家之力,编纂出一部内容全面、概念精确、表述完整、接近世界医学统计学先进水平、编辑形式简洁的大型医学统计学工具书。2008年,“全书”开始酝酿筹备,几经讨论,搭成框架条目,确定编写格式,并开始全面着手编写,终于于2011年初编纂出初稿。值得欣喜的是,在中国统计出版社的大力支持下,“全书”项目先后成功申报了国家出版基金(项目编号2011C₂-003)和全国统计科学研究(计划)课题(立项编号2011LY080),皆荣获批准。有了国家出版基金和全国统计科学研究(计划)课题的支持,“全书”的编纂工作如虎添翼,更上台阶。

通过国内外数十所大学、医学院校与医学科研院所近百位统计学专家教授的共同努力,“全书”终于能够付梓成册,得以与广大读者见面,编者倍感欣慰。“全书”既全面介绍了医学统计学的基本理论、基本知识与方法,又介绍了大量新的统计理论与方法,对生物医学统计的传统方法及最新进展进行了全面梳理,同时还改变了目前医学统计著述中普遍存在的统计方法或指标之间缺乏相互联系,过于分散与单一的现象。这就形成了“全书”的特点:全面、系统、实用、前沿。

“全书”共8个分册:描述性统计分册、单变量推断统计分册、多元统计分册、非参数统计分册、管理与健康统计分册、医学研究与临床统计设计分册、健康测量分册、遗传统计分册,均由著名高校医学统计学教授担纲主编,同

时聘请国内外知名医学统计教授担任顾问。可谓举全国名校之力,集百家精英之长。在编写过程中,专家们严谨认真,精益求精,在注重科学性、知识性、先进性、可读性的前提下,紧紧把握医学科学研究与医疗卫生工作的特殊性和复杂性,精心研究论证各种统计理论与方法在医学领域的适用性与应用条件。为了便于读者学习和理解应用,书中不仅有理论分析,还提供了实例运用,并把计算机软件程序应用于其中,对统计方法或体系的科学性与可行性进行检验,使统计理论与医学实际得到紧密结合。在每一分册的内容安排上,遵循从简单到复杂、从一般到先进、从传统到前沿的原则,使各分册在内容上既相互衔接补充,融为一体,又能各自独立成册。为方便读者查阅,书中各条目层次分明,结构严谨,醒目易读,是广大医学科学工作者学习和使用、必备案头的大型医学统计工具用书。

“全书”在编写过程中,引用了相关专著及教材的部分资料,在此对引用资料的原作者表示衷心感谢!引用资料中多数已在书中注出,也有部分没有一一注出,对于没有注出的部分,在此敬请原作者给予谅解!中国统计出版社教材编辑部和滨州医学院的领导及同仁们为“全书”的编辑和出版付出了大量心血,在此致以诚挚感谢!

由于编者水平有限,书中难免会存在错误和不足之处,恳请广大读者提出宝贵意见。

最后,感谢您学习和使用“全书”,希望它能使您开卷有益。

总主编 徐天和

前 言

随着计算机技术的飞速发展和广泛普及,多元统计方法在医学领域的应用日益广泛。医学研究的对象主要是人,而人的心理、生理和疾病状态是最复杂的自然现象,它与人的内在因素和外部自然环境、社会环境存在密切的联系。因此在医学研究中,不可能只考虑单一因素与单一结果的关系,往往需要研究多个因素与某一结果,甚至多个因素与多种结果的关系。单变量统计方法无法解决多元分析的问题。近年来,在计算机和统计软件的支持下,许多多元统计方法已经成为医学研究资料分析中常规的统计方法。

多元统计分析方法涉及较为复杂的数学理论,计算繁琐。许多多元统计方法无法手工计算完成,必须由计算机和统计软件协助计算。在写作上也没必要将每一步计算步骤表述出来。作为一本百科全书,我们希望尽量将多元统计方法的原理和计算过程介绍清楚,让读者可以根据介绍的计算过程和公式,将计算机程序编制出来。但是,我们认为对于一般的医学工作者,重要的不在于理解多元统计方法的数学原理,也不需要掌握具体的计算步骤,最重要的是了解多元统计方法的分析目的、基本思想、分析逻辑、应用条件和结果解释。所以这部分读者可以忽略有关章节中数学理论和具体计算公式介绍,着重阅读各种方法的应用条件、基本分析思想,实例的具体应用和结果解释。

近年来,多元统计分析方法发展非常快,新的多元统计方法不断涌现,特别是在纵向研究资料分析方法、量表统计分析方法、多水平分析方法等方面发展较快。一些经典的多元统计方法在向多样化,个性化发展,一些多元统计方法随着应用的转变也增加了新的内容。我们尽量将这些新的发展纳入书中,但由于能力和水平所限,该书不可避免地存在错误和不足,尚望广大读者不吝赐教,多提宝贵意见。

柳 青

2012 年 11 月

目 录

多元分布概论	(1)
多变量及其概率分布	(2)
多元正态分布	(5)
多元正态分布参数的最大似然估计	(10)
维希特分布	(13)
检验平均向量的统计量	(15)
总体平均向量的置信区域与联合置信区间	(22)
协方差阵的检验	(25)
总体协方差阵不等时两均数向量的假设检验	(28)
多元方差分析	(30)
完全随机设计的多元方差分析	(31)
随机配伍设计的多元方差分析	(35)
多元方差分析的线性模型形式	(38)
多元方差分析中的多组间两两比较	(46)
重复测定数据的多元方差分析方法	(49)
轮廓分析	(58)
协方差分析	(62)
多元协方差分析	(72)
简单相关系数	(77)
多重线性相关系数	(84)
偏相关系数	(88)
多序列相关系数	(92)
多项相关系数	(96)
典型相关分析	(100)
曲线拟合	(108)
曲线直线化	(110)

迭代配合法	(111)
指数函数曲线配合	(113)
幂函数曲线配合	(116)
多项式曲线配合	(118)
Gompertz 曲线配合	(122)
Logistic 曲线配合	(125)
Logit 曲线配合	(128)
Probit 曲线配合	(130)
样条函数	(133)
简单线性回归	(135)
通过原点的直线回归	(145)
多元线性回归	(147)
确定系数与校正确定系数	(153)
残差分析	(154)
回归诊断	(158)
自变量选择	(162)
影响函数	(170)
多重共线性	(174)
回归模型的失拟合检验	(182)
岭回归	(184)
线性回归模型的加权最小二乘法	(188)
稳健回归	(190)
Tobit 模型	(197)
非参数回归	(202)
广义加法模型	(205)
多项式回归	(211)
正交多项式回归	(216)
趋势面分析	(220)
Logistic 回归分析	(224)
非条件 logistic 回归分析	(237)
条件 logistic 回归分析	(246)
多项分类 logistic 回归模型	(254)
有序 logistic 回归模型	(259)
Probit 回归	(264)
倾向评分	(269)

Poisson 回归	(272)
加速失效时间模型	(276)
指数回归	(278)
Weibull 回归	(283)
比例风险回归	(286)
非比例风险回归	(294)
对数正态回归	(296)
Log-logistic 回归	(299)
多状态 Cox 比例风险回归模型	(301)
对数线性模型	(306)
广义估计方程	(309)
主成分分析	(315)
探索性因子分析	(319)
验证性因子分析	(324)
通径分析	(327)
潜变量分析技术	(333)
对应分析	(335)
结构方程模型	(339)
组内相关系数	(344)
迭代广义最小二乘算法(IGLS)	(346)
期望极大估计算法(EM algorithm)	(348)
随机截距模型	(349)
随机斜率模型	(357)
多水平生长模型	(362)
多水平 logistic 模型	(365)
多水平 Poisson 回归模型	(370)
多水平多项分类 logistic 回归模型	(374)
多水平有序分类 logistic 回归模型	(379)
多元多水平回归模型	(382)
多水平 Meta 分析模型——已知均数和标准差	(387)
多水平 Meta 分析模型——以 OR 为效应尺度	(391)
交叉分类数据模型	(394)
判别分析	(397)
二次型判别	(406)
逐步判别	(408)

聚类分析	(413)
聚类分析的常用统计量	(414)
系统聚类法	(417)
逐步聚类法(动态聚类法)	(425)
有序样品最优分割法	(433)
附录一 统计用表	(438)
附表1 正态分布表	(438)
附表2 t 分布界值表	(441)
附表3 χ^2 分布界值表	(443)
附表4 F 分布界值表	(446)
附表5 相关系数 r 界值表	(456)
附表6 正交多项式系数表	(457)
附录二 英汉医学统计学词汇	(459)
附录三 汉英医学统计学词汇	(461)
本书词条索引	(463)

多元分布概论

生物医学数据中常常包含了每个实验对象的多方面特征,对这些多维数据进行建模和分析,多元分布起了非常关键的作用。多元分布可以通过转换、投影、条件化、极值和截尾等导出各种统计量的分布,以用于统计推断。除此之外,多元分布遍及整个统计学和应用概率论,它们的性质对于理解这些学科和相关领域是非常关键的。关于多元分布的研究最早源于 19 世纪早期对于多元正态分布的研究。

1 随机向量的多元分布函数

假设考虑某地中学生的生长发育情况,由三个指标组成, $X=(X_1, X_2, X_3)'$, 其中 X_1 表示身高, X_2 表示体重, X_3 表示胸围。这三个指标之间都是相关的,这三个变量的联合分布称为多元分布。 X 称为随机向量,是由随机变量 X_1, X_2, X_3 组成的向量。

随机向量 $X=(X_1, X_2, \dots, X_n)'$ 的多元分布函数 $F(x_1, x_2, \dots, x_n)$, 即为随机变量 $X_1, X_2, \dots, X_n$ 的联合分布函数:

$$F(x_1, x_2, \dots, x_n) = P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n)$$

2 分布的类型

对于多元分布的研究主要针对离散型和连续型两种类型。类似于一元分布,离散型多元分布是指像心率、手术人数等多个离散型变量的联合分布,而连续型多元分布主要指像身高、体重、胸围等多个连续型变量的联合分布。

2.1 离散型多元分布

离散型随机向量 $X=(X_1, X_2, \dots, X_n)'$ 的分布函数为 $F(x_1, x_2, \dots, x_n)$, 其概率函数 $p(x_1, x_2, \dots, x_n)$:

$$p(x_1, x_2, \dots, x_n) = P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}$$

很多离散型分布都有其多元时的扩展,比如多元二项分布,多元负二项分布,多元 Poisson 分布等。除此之外,还有多项分布,多元超几何分布等。

2.2 连续型多元分布

连续型随机向量 $X=(X_1, X_2, \dots, X_n)'$ 的分布函数 $F(x_1, x_2, \dots, x_n)$, 其概率密度函数 $f(x_1, x_2, \dots, x_n)$:

$$f(x_1, x_2, \dots, x_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F(x_1, x_2, \dots, x_n)$$

连续型分布中除了多元正态分布外,还有多元 Gamma 分布,多元 χ^2 分布,多元指数分布,多元 t 分布,多元 Weibull 分布等。

由多元分布通过转换,复合等导出的统计量,如维希特分布,赫特林 T^2 分布和维尔克斯分布等是在多元分析中常用的统计量。

参考文献

- [1] Jensen, DR. Multivariate Distributions, Overview. Encyclopedia of Biostatistics. John Wiley & Sons, 2005: 3780—3793.
- [2] 张尧庭. 多元统计分析引论. 北京: 科学出版社, 1982: 362—402.

(孙红卫 祁爱琴)

多变量及其概率分布

1 联合分布函数

如果 $X_1, X_2, \dots, X_n$ 是随机变量, 则其联合分布函数(joint distribution)是

$$F(x_1, x_2, \dots, x_n) = P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n)$$

2 二元分布函数

对于两个随机变量 X 和 Y , 其联合分布函数为

$$F(x, y) = P(X < x, Y < y)$$

联合分布函数的性质:

1) $0 \leq F(x, y) \leq 1$, 且

对任意固定的 y , $F(-\infty, y) = \lim_{x \rightarrow -\infty} F(x, y) = 0$,

对任意固定的 x , $F(x, -\infty) = \lim_{y \rightarrow -\infty} F(x, y) = 0$,

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

2) $F(x, y)$ 关于 x 和 y 均为单调非减函数, 即

对任意固定的 y , 当 $x_2 > x_1$, $F(x_2, y) \geq F(x_1, y)$,

对任意固定的 x , 当 $y_2 > y_1$, $F(x, y_2) \geq F(x, y_1)$;

3) $F(x, y)$ 关于 x 和 y 均为左连续, 即 $F(x, y) = F(x-0, y), F(x, y) = F(x, y-0)$;

4) 对任意四个实数 $x_1 \leq x_2, y_1 \leq y_2$, 有

$$P\{x_1 < x \leq x_2, y_1 < y \leq y_2\} = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1)$$

3 边际分布

如果 $F(x, y)$ 是 X 和 Y 的联合分布函数, 则 X 与 Y 的边际分布 (marginal distribution) 是

$$F_X(x) = \lim_{y \rightarrow +\infty} F(x, y) = F(x, +\infty)$$

$$F_Y(y) = \lim_{x \rightarrow +\infty} F(x, y) = F(+\infty, y)$$

4 离散型二元分布

对于离散型随机变量 (X, Y) ,

1) 联合分布列: $P(X=x, Y=y) = p(x, y)$

2) 概率: $P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = \sum_{x_1 \leq x \leq x_2} \sum_{y_1 \leq y \leq y_2} p(x, y)$

3) 边际分布列: $P(X=x) = \sum_y p(x, y), p(Y=y) = \sum_x p(x, y)$

5 连续型二元分布

对于连续型随机变量 (X, Y) ,

1) 联合密度函数: 一个正的实值函数 $f(x, y)$, 满足

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv \quad \forall x, y \in R$$

2) 概率: $P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f(x, y) dx dy$

3) 边际密度函数: $f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy, f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx$

6 两个随机变量的函数的期望

如果 $g(x, y)$ 是一个实值函数, 当随机变量 (X, Y) 为离散型时, 其联合概率函数为 $p(x, y)$, 当随机变量 (X, Y) 为连续型时, 其密度函数为 $f(x, y)$,

$$E[g(X, Y)] = \begin{cases} \sum_x \sum_y g(x, y) p(x, y) & \text{离散型时} \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy & \text{连续型时} \end{cases}$$

7 多元时的扩展

对于随机变量 $X_1, X_2, \dots, X_n, j=1, 2, \dots, n$

1) 边际分布: $F_{X_j}(x_j) = F_{X_1, X_2, \dots, X_n}(\infty, \dots, \infty, x_j, \infty, \dots, \infty)$

2) 边际概率或密度

$$p_{X_j}(x_j) = \sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) \quad \text{离散型时}$$

$$f_{X_j}(x_j) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) dx_1 \cdots dx_{j-1} dx_{j+1} \cdots dx_n \quad \text{连续型时}$$

3) n 个变量的函数的期望

$$E[g(X_1, X_2, \dots, X_n)]$$

$$= \begin{cases} \sum_{x_1} \cdots \sum_{x_n} g(x_1, x_2, \dots, x_n) p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) & \text{离散型时} \\ \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(x_1, x_2, \dots, x_n) f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) dx_1 \cdots dx_n & \text{连续型时} \end{cases}$$

8 独立性

如果随机变量 $X_1, X_2, \dots, X_n$ 互相独立, 则对于所有的 $x_1, x_2, \dots, x_n$,

$$1) F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n)$$

$$2) f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

$$3) E(X_1 X_2 \cdots X_n) = E(X_1) E(X_2) \cdots E(X_n)$$

4) 对实值函数 $g_1, g_2, \dots, g_n, g_1(X_1), g_2(X_2), \dots, g_n(X_n)$ 是独立的。

9 平均向量和协方差阵

设随机向量 $X = (X_1, X_2, \dots, X_n)'$, $Y = (Y_1, Y_2, \dots, Y_m)'$, $p(x_1, x_2, \dots, x_n)$ 是 X 的联合概率函数(离散型时), $f(x_1, x_2, \dots, x_n)$ 是 X 的联合密度函数(连续型时)。 $p(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m)$ 是 $(X', Y')'$ 的联合概率函数(离散型时), $f(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m)$ 是 $(X', Y')'$ 的联合密度函数(连续型时)。

9.1 平均向量

X 的平均向量为 $EX = (EX_1, EX_2, \dots, EX_n)'$, 其中

$$EX_i = \begin{cases} \sum_{x_1} \cdots \sum_{x_n} x_i p(x_1, x_2, \dots, x_n) & \text{离散型时} \\ \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x_i f(x_1, x_2, \dots, x_n) dx_1 \cdots dx_n & \text{连续型时} \end{cases}$$

9.2 协方差阵

向量 X 的协方差阵为

$$V(X) = E[(X - E(X))(X - E(X))']$$

$$= \begin{bmatrix} V(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & V(X_2) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & V(X_n) \end{bmatrix}$$