



概率统计的研究与发展*

文/巩馥洲

中国科学院数学与系统科学研究院 北京 100190

【摘要】 文章简要介绍了概率统计学的战略地位、主要的研究方向、方法以及发展状况及趋势。

【关键词】 概率统计, 研究现状, 发展趋势

DOI :10.3969/j.issn.1000-3045.2012.02.006



中国科学院

1 概率统计的战略地位与发展趋势

概率统计是研究随机现象数量规律的数学学科,包括随机现象的数学理论,以及如何有效地收集、整理和分析相关数据,并对所考察问题做出推断或预测^[1]。

随着计算机技术等高技术的迅速发展,各种高性能数据的收集设备和手段不断更新,使人们可以针对许多复杂事物和现象直接获得形式多样的海量复杂数据。这也使得数据驱动式的科学研究成为时代的潮流,即直接利用海量复杂数据,研究传统方法和手段失效的复杂事物和现象,发现隐藏其中的复杂结构和科学规律^[2]。值得指出的是,一方面这些形式多样的海量复杂数据之间存在着很高的相依性和聚集性,使其整体呈现出与随机性数据类似的性质;另一方面,由于各种未知的不可控因素对复杂事物和现象以及相关的数据收集设备的影响,使

得绝大多数的海量复杂数据本身就是随机性数据。因此,对海量复杂数据进行数据驱动式科学研究,离不开概率统计的理论与方法。

概率统计中的概率论主要研究和关于随机现象的数学理论,其主要目的是揭示蕴藏在随机现象中的结构和规律,同时也为统计理论与方法提供理论基础^[1]。

与大部分数学分支的起源不同,概率论起源于对赌博问题的研究。概率论的创始人分别是16世纪末的意大利数学家卡丹诺(G. Cardano)和17世纪的法国数学家帕斯卡(B. Pascal)、费马(P. Fermat)以及瑞士数学家雅各-伯努利(J. Bernoulli)。人们逐渐发现,许多领域都存在与赌博中出现的输赢现象相似的重要随机现象,概率论也就被应用到更多领域中,从而极大扩展了其应用范围。19世纪末,波尔茨曼(L. Boltzmann)和吉布斯(J.W. Gibbs)建立的统计力学,运用大量随机运动的粒子解释了气体的性质,成

* 收稿日期 2012年3月13日

为用概率论解释自然现象的一个巨大成功。20世纪初,受统计力学研究的刺激,人们开始研究随时间变化的随机现象,形成了有广泛应用的随机过程理论。1900年庞加莱的学生巴夏里埃(L. Bachelier)首先研究了投机理论,并建立了布朗运动的模型,其后爱因斯坦以及控制论创始人维纳(N. Wiener)分别建立了布朗运动的物理模型与数学模型。1906年俄罗斯数学家马尔可夫(A. Markov)提出了马尔可夫链的概念。1933年苏联数学家柯尔莫哥洛夫为概率论奠定了严密的数学基础,给出了概率空间的公理化定义。这是概率论发展史上的一个重要里程碑,为概率论的迅速发展铺平了道路^[3]。柯尔莫格罗夫也因包括这一工作在内的一系列杰出贡献于1980年获得沃尔夫数学奖。1939年法国数学家温勒(J. Ville)提出了鞅的概念。1942年日本数学家伊藤清开创了随机分析。

概率论不仅在后来的发展中越来越显示出它在众多领域的应用性和实用性,而且对很多基本应用问题的研究也推动了概率论的迅速发展。例如,2006年菲尔兹奖得主奥克恩科夫(A. Okounkov)的获奖工作解决了与弦物理学有关的一个重要数学问题。在此过程中,他建立了概率论、表示论和代数几何之间的桥梁。该工作不仅揭示了概率论与数学中多个表面上不相关分支之间的深刻联系,而且也为弦物理学中一些重要问题的解决提供了新的思想。同年另一个菲尔兹奖得主陶哲轩(T. Tao)的获奖工作解决了一个长期悬而未决的著名数论难题^[4]。在此过程中,他利用了概率论的思想和动力系统的多重遍历理论。同时,他的工作也刺激了概率论学者进一步研究马尔可夫过程的多重遍历理论。利用动力系统理论和偏微分方程理论,维拉尼(C. Villani)解决了流体力学中的朗道阻尼和非平衡统计力学中波尔兹曼方程的长时间行为这两个难题,并因此获2010年菲尔兹奖^[5]。他的研究大量借鉴了概率论中泛函不等式和最优输运理论的思想和方法。

概率论和社会经济相互促进发展的经典范例

之一就是资产定价与风险度量理论的建立与广泛应用。资产定价与风险度量问题是经济金融领域长期以来受到普遍关注的重大问题,1971年以前的近百年一直未得到很好的解决。1972年布莱克(F. Black)和斯科尔斯(M. Scholes)利用随机微分方程来描述资产价格的变动,并基于此给出了资产定价与风险度量的内在形成机制,提出了著名的Black-Schole公式,创造性地解决了资产定价和风险度量问题,极大地促进了现代金融业的发展壮大。斯科尔斯因此获得了1997年的诺贝尔经济学奖。同时,由于资产定价与风险度量理论使用伊藤清建立的随机积分理论,伊藤清被称为“华尔街最有名的日本人”,他因此获得了1987年的沃尔夫数学奖及2006年国际数学家大会新设的高斯奖^[4]。另外,资产定价与风险度量理论的深入研究也推动了概率论的进一步发展。例如,山东大学彭实戈院士因发展了用于刻画动态风险度量的非线性数学期望及相应的随机分析理论,应邀在2010年国际数学家大会做大会报告^[5],成为首位获此殊荣的国内数学家。概率论和工程技术等领域也相互促进发展,例如自动化领域著名的卡尔曼滤波器设计的理论基础之一就是随机过程理论,而滤波理论也推动了马尔可夫过程理论中布莱克威尔(A. Blackwell)猜测和库尼塔(H. Kunita)问题的提出与研究。信息技术领域的复杂网络理论与概率论极为密切。互联网、无线通讯网以及物联网甚至包括基因调控网和蛋白质相互作用网等大规模网络及其上的信息与物质流动,都是不断发生变化的且在整体上呈现出与随机变动类似的行为。因此,建立相应的随机过程模型并利用概率方法进行分析,并基于此开展复杂网络工程的设计成为研究网络最好的模式^[6]。

从上述事实可以看出,概率论进一步与数学其他分支、自然科学、工程技术以及社会经济等领域相互促进发展仍是其主要发展趋势。特别是,随着这些领域“数据驱动”式的科学研究成为时代的潮流,将越来越需要概率论为这类研究提供理论基

基础,其研究重点集中在与海量复杂数据处理与分析有关的概率问题上。

概率统计中的统计学主要研究和发展有效地收集、整理和分析相关数据的理论与方法,并对所考察问题做出推断或预测,其理论基础是概率论^[1]。它的主要目的是提取和挖掘隐藏于数据背后的结构和科学规律,并对相关的过去或未来进行推断。因此,统计学需要关注推断方法和所得结论与实际吻合程度。

近代统计学的产生和发展与生物学研究密切相关。英国遗传学家高尔顿(F. Galton)受达尔文进化论的影响,从1860年开始首先研究甜豆的遗传,他发现,甜豆子代直径的平均会趋向于全部甜豆直径的平均。这不仅很好地解释了经历繁衍甜豆的总体不发生变化的原因,而且也成了线性回归方法研究的起源。他在1884年的伦敦世博会上建立了测量站,测量了数千人的身体数据,并于1886年发表的文章中给出了一张相当于正态相关曲面的表,用来比较父母双方平均身高与子女身高的关系。在他的促使下,迪克逊(H. Dickson)对这些数据开展了最早的相关性研究。高尔顿因此被公认为生物计量学的奠基人。接着,卡尔·皮尔逊(K. Pearson)以讨论正态曲线和相关性为起点,定义了标准差、直方图、众数和相关系数等统计学的基本概念。他还提出了卡方检验的方法,用来检验观测数据与正态曲线吻合的程度。20世纪20年代,卡尔·皮尔逊的儿子埃贡·皮尔逊(Egon Pearson)和纽曼(J. Neyman)合作,利用概率论建立了一般假设检验的数学理论,使得人们可以利用观测数据来判断它们是否服从某种科学规律。这极大地改变了统计学的面貌,使统计学完成了从描述统计到统计推断的革命性转变^[3]。

从历史上看,统计学中很多的开创性工

作都是从研究非常具体的实际问题开始的。例如,英国统计学家戈塞特(W. Gosset)开创的小样本统计理论,就始于他对酿酒中使用的麦子质量的研究。另外,费歇尔开创性地提出的试验设计之随机化原则,就始于他对农业试验的研究。试验设计随机化原则后来的应用具有十分重要的影响。

到了现代,随着人们研究的事物和现象越来越复杂,要求也越来越精细,描述数据内在规律的概率统计模型也越来越复杂。事实上,参数估计已发展为半参数与非参数估计,模型已由线性扩展到非线性,数据也由独立改变为相依。这方面一个很好的例子就是,资产定价与风险度量中Black-Schole公式里刻画价格风险因子的识别问题。然而,解决这一问题就需要对价格变动过程中积累的大量的条件异方差的与相依的复杂数据之间的相互关系进行识别与分析。由于处理这个问题需要的统计学理论与方法超出了当时统计学的范围,致使10年间未有进展。直到1982年恩格尔(R. Engle)创造性地提出了统计学中时间序列分析的新模型

ARCH模型,才成功地解决了资产定价与风险度量中风险因子的识别与估计问题,他因此获2003年诺贝尔经济学奖。

另外,近些年来一个值得注意的事实是,自然科学、社会经济和工程技术领域获得的数据在规模和类型上都发生了巨大变化,统计学面临着与以往完全不同的新问题。例如,欧洲核子研究中心(CERN)的大型强子对撞机(LHC)平均每天产生数据约4万亿个字节^[7]。同时,美国国家生物技术和信息中心(NCBI)的基因库早已收录了1亿多条记录,1千多亿个碱基^[8]。由于人们缺乏对这类海量复杂数据背后复杂现象的理解而不具备提出合理科学假设的能力,所以,在面对这类海量复杂数据时,人们无法事先



中国科学院

假设隐藏其中的科学规律。这就需要人们开展数据驱动式研究,即从这类海量复杂数据中直接发现新的结构和科学规律,而不是先对隐藏在这类海量复杂数据中的结构和科学规律提出合理假设,然后再利用这类数据进行假设检验。这种情况未曾有过,其对统计学的已有理论与方法提出了严峻的挑战。

2 现代概率论研究的主要方向与发展状况

自从柯尔莫哥洛夫为概率论奠定了严密数学基础后,不但其理论迅速发展,而且研究方向也快速增加。一些方向继续为统计学提供所需的理论基础,另一些则大量使用数学其他分支的思想方法,侧重于直接研究许多领域中可转化为随机结构的问题,形成了现代概率论丰富多彩的研究局面。

2.1 随机分析

20世纪40年代,从美国数学家杜布系统地研究一般的鞅以及日本数学家伊藤清引入随机积分概念开始,经过几代学者的努力形成了现代概率论的随机分析研究方向。随机分析借鉴数学中分析学的思想方法研究随机过程的局部结构和整体特征,被称为是具有随机趣味的分析学。随机分析主要包括一般半鞅理论、随机微分方程、随机偏微分方程、马列奥万(P. Malliavin)分析与随机微分几何以及狄氏型理论等^[9]。其在数学其他分支、力学、物理、化学、生物与医学、经济金融、管理科学以及工程技术等领域有广泛应用。

一般半鞅理论研究,目前集中在3个方面:(1)利用半鞅理论中扩大过程信息流的方法,研究含有信用风险等道德风险的资产定价与风险度量问题;(2)依据行为金融学中的情景理论,重新研究资产定价和风险度量问题;(3)对长记忆与强关联的随机过程类,包括可描述复杂网络上信息传输过程的分数维布朗运动等,发展它们的随机积分理论。

随机微分方程研究,目前集中在两个方面:(1)在系数仅满足某种非李普希茨条件时,研究随机微分方程解的基本性质,以及利用随机分析方法,研究系数仅具有弱正则性时常微分方程解的基本性

质;(2)动力系统与不确定环境耦合后动力学行为变化特征的研究。这其中包括了随机动力系统以及目前最受国际数学界关注的随机劳威纳演化(Stochastic Loewner Evolution, SLE)研究。我们稍后介绍随机劳威纳演化情况。值得特别强调的是研究如下情况的重要性。目前,我们可以大量获得关于宇宙深处天体的海量复杂数据,然而,对该天体周围环境的物理力学状况几乎一无所知,仅根据宇宙学知道在不与环境耦合时它的动力系统描述方式。事实上,类似情况在自然科学领域广泛存在。这与以往动力系统研究的状况完全不同,需要结合获得的海量复杂数据,借鉴随机分析、动力系统和统计学的思想方法,发展适合的理论与方法,研究这些动力系统与不确定环境耦合后动力学行为的变化特征。

随机偏微分方程的研究,目前相对集中在这样的方程,它们是力学、物理以及工程技术等领域的重要偏微分方程,但带有随机驱动噪声。(1)研究这类方程解的存在唯一性以及遍历性等基本性质。这其中具有挑战性的是随机噪声退化的情况,目前仅有不多的研究成果,需要发展新的随机分析理论;(2)研究与这些方程有关的控制问题,其中一些问题具有很强的挑战性。比如,处于随机复杂环境中的传感器网络与智能机械等柔性和弹性系统都是典型的无穷维随机系统,而对这些系统的分析与控制是必不可少的。虽然,实际工程控制都是通过有限维离散模型的控制来实现的,然而,为了给出精度更高的实际控制模型和控制器设计方法,需要对描述这些系统的随机偏微分方程组进行分析。经典的分析与控制理论在柔性和弹性系统的分析与控制方面有很大的局限性,因此发展新的研究框架、思路和方法刻不容缓。特别地,需要研究随机的热传导与非线性双曲守恒律系统描述的波传播过程、随机的牛顿与非牛顿流体和电磁流体的分析与控制,以及随机复杂环境中非线性弹性模型和流体与弹性体耦合模型的控制及其应用。在上述问题中,仅限于所涉及随机偏微分方程组的可控

性问题在数学理论上就极具挑战性。

马列奥万分析与随机微分几何研究,目前相对集中在两方面:(1)运用马列奥万分析与随机微分几何的基本方法研究数理金融学以及数学中分析学、有限维黎曼流形、有限维复流形等领域中的一些困难问题,比如,基础资产价格或交易策略带有强非线性约束的资产定价和风险度量,奇异扩散半群的Harnack不等式等;(2)研究无穷维线性空间和无穷维流形上的马列奥万分析与随机微分几何问题。例如,黎曼流形的Wasserstein空间上梯度流的研究。在这方面的研究中,具有挑战性的是马列奥万矩阵算子非退化但其伪逆无界的情况,这是以前马列奥万分析与随机微分几何研究中几乎未曾遇到的,需要开创性的研究思路与方法。

狄氏型理论的研究,目前相对集中在如下几个方面:利用狄氏型理论,研究区域上Levy过程等跳跃型马尔可夫过程的内在超压缩性、热核估计、边界附近行为与相应的位势理论,以及进一步深入研究非对称拟正则狄氏型的Beurling-Deny型分解理论及其应用等。

现在,我们介绍关于随机劳威纳演化的情况。

为了研究复分析里著名的比伯巴赫(L. Bieberbach)猜想,劳威纳(C. Loewner)于1923年引进了描述单参数共形映射形变的劳威纳演化(Loewner Evolution, LE)。自20世纪80年代以来,共形场论(CFT)已成为理论物理与数学相互作用的主要物理分支。至今,它仍然在弦物理理论的发展中起相当重要的作用。从共形场论角度出发,许多物理学家预言或猜测,诸多来自统计物理的两维模型其标度(Scaling)极限是共形不变的。这些模型包括渗流模型、自回避游走、Ising模型、FK模型与 $O(n)$ 圈模型、去圈随

机游动(LERW)、均匀展开树(UST)等。从1994年开始,一些数学家如Wolf奖得主R. Langlands等就试图从数学角度理解渗流模型标度极限的共形不变性。为了研究上述模型共形不变的标度极限,首先必须回答,这些极限是怎样的数学对象。2000年,微软研究院的数学家施拉姆(O. Schramm)引进随机劳威纳演化,并推测这些极限可能是某个随机劳威纳演化。通俗地讲,随机劳威纳演化是平面上具有共形不变性和一种马尔可夫性的单参数随机曲线族。引进随机劳威纳演化的原始动机,来自于寻找劳勒(G. F. Lawler)于1980年引进的去圈随机游动的连续版本模型,而平面上共形不变随机曲线的研究,则可以回溯到莱维1946年关于复布朗运动共形不变性的发现。沃纳(W. Werner)、劳勒和施拉姆一起发展了随机劳威纳演化理论,深化了关于平面上共形不变的随机曲线的理解和认识。他们严格证明了一些两维空间上统计力学模型的标度极限具有共形不变性。自此随机劳威纳演化理论吸引了大批数学家和物理学家的兴趣。这是因为,共形场论的出发点是假定两维空间上离散的统计物理模型的标度极限具有共形不变性。因而,证明两维空间上统计物理模型标度极限的共形不变性就变得意义重大而具有挑战性。沃纳因随机劳威纳演化的研究因此获得了2006年菲尔茨奖^[5]。同样,因为随机劳威纳演化方面的杰出成就,斯米尔诺夫(S. Smirnov)获得了2010年的菲尔茨奖^[5]。目前,下述模型标度极限的共形不变性已被证明,它们是三角形格点上的渗流、去圈随机游动、均匀展开、Harmonic Explorer与高斯自由场、Ising模型以及Ising random cluster模型。但是,大量两维空间上统计物理模型标度极限的共形不变性还缺乏严格的数学证明。毫无疑问,继续证明两



中国科学院

维空间上一些统计物理模型标度极限的共形不变性仍将是随机劳威纳演化理论的核心问题。另外,利用随机劳威纳演化理论理解或重建共形场论也会是一个重要的研究内容。最后,把随机劳威纳演化理论进行拓展,并用来理解二维空间上统计物理模型在 off-critical 情形下的性质也将是一个重要的发展趋势^[10]。

2.2 马尔可夫过程

自从1906年马尔可夫提出马尔可夫链的概念起,历经柯尔莫哥洛夫、莱维等一大批概率学者的努力,到60年代已形成比较完备的马尔可夫过程理论体系。马尔可夫过程的特征是它的无后效性,即已知现在的状态则未来与历史无关。所有具有短记忆特征的变化过程都可用马尔可夫过程来描述^[11]。因此,它是随机过程中研究最多的与应用最广的过程类之一。马尔可夫过程理论主要的研究方法是利用其生成元或转移概率来研究它们的概率性质。由于马尔可夫过程理论的研究内容极其丰富,这里仅着重介绍几个方面。

马尔可夫过程遍历性研究,目前重点集中在具有重要实际背景而且研究难度较大的不可逆马尔可夫过程上,特别是对一些不可逆典型过程的谱空隙进行研究和估计。例如,某些不可逆排队论模型的遍历性问题,特别是 M/PH/1, PH/M/1 等模型在位相无限时的一般遍历性、指数遍历性和强遍历性;多物种模型的过程存在性、唯一性和相应的遍历理论等。

马尔可夫过程的衰减性与泛函不等式研究,其包括两方面:(1)借鉴其遍历性研究的丰富成果,建立衰减性的一套比较完整的理论,这也包括一些相应的泛函不等式;(2)研究具有非线性和非对称性的随机偏微分方程的泛函不等式、运费不等式和最优传输不等式;某些条件下生灭过程或一般 Q 过程的泛函不等式、修正 Log Sobolev 不等式和相应的显式判别准则;以及粒子之间相互排斥的连续型气体模型的泛函不等式,比如修正的 Log Sobolev 不等式和协方差估计等。

粒子系统与测度值过程及其遍历性研究,包括:(1)利用拟正则狄氏型理论构造一般 Polish 空间上的粒子系统,并构造合理的测度值过程使其关于相对熵指数收敛到 Poisson-Dirichlet 分布;研究随机流产生的超过程以及包括相关的 Konno-Shiga 方程在内的一类随机偏微分方程解的轨道唯一性;(2)研究 stepping-stone 模型的遍历行为,包括其泛函不等式和可逆性;研究有选择的 FV 超过程的遍历性和有移民分支过程的遍历性和可逆性,以及选择和重组模型与非平衡动力模型的极限行为;研究测度值过程和分支过程的大偏差等极限性质,利用 Stein 方法和耦合方法等,研究测度值过程和分支过程的极限和收敛性的上界估计。

随机环境模型的研究,目前有两个方面:(1)研究具有随机移民的超布朗运动在 Quenched 概率下的中心极限定理和大偏差等性质,以及大偏差理论在通讯、排队论中的应用;(2)研究随机环境中随机游动的一些挑战性的基本问题,例如,在二维情形以及环境平稳、遍历、一致椭圆、混合时随机游动沿某方向运动的 0-1 率。目前,关于随机环境中的随机游动的研究在概率界是很活跃的。

马尔可夫骨架过程研究,其是马尔可夫过程的一种重要推广,仍借鉴马尔可夫过程的思想与方法。它的研究分两个方面:(1)包括了排队论中各种排队过程模型在内的大量混杂随机模型的马尔可夫骨架过程的研究。这类马尔可夫骨架过程由我国中南大学的侯振挺教授和合作者于1997年引入,目前的研究集中于该理论在金融保险和排队理论中的应用。(2)大规模网络上 Web 马尔可夫骨架过程的理论及其应用的研究。Web 马尔可夫骨架过程是马志明院士和合作者于2009年从万维网搜索引擎设计的研究中提炼出来的一类新的重要随机过程,囊括了离散时间马氏链、Q-过程、更新过程等经典随机过程以及一类新的重要随机过程——镜面半马氏过程。它概括了万维网中信息的输入、传输和使用的基本特征,可用来描述互联网上用户的浏览过程,计算机病毒在互联网中的传播过程,

以及计算机黑客对计算机网站进攻的演化过程等。除此之外,这类过程也可以用来描述宽带无线移动通信网络、传感器网络、物联网和生物分子网络等大规模网络中信息的输入、传输和使用过程,以及这些网络上故障的传播与扩散过程等。目前的研究相对集中在丰富和发展大规模网络上 Web 马尔可夫骨架过程的理论与方法,解决若干大规模网络上搜索引擎设计、热点信息挖掘以及故障的传播与扩散现象等问题。上述的两类马尔可夫骨架过程互不包含,既有重叠又有区别。

2.3 非线性数学期望及相应的随机分析

非线性数学期望的理论研究可追溯到上世纪 50 年代。1954 年,法国数学家邵盖(G. Choquet)提出了容度概念,并利用容度定义了一种现在被称为 Choquet-积分的非线性积分,它定义的期望就是一种非线性数学期望。这一理论目前在金融和经济学中得到广泛应用。70 年代以来,这一理论被推广为非可加测度和非线性数学期望理论。特别是近 10 多年,非可加测度和非线性积分理论的研究取得了一系列成果。1997 年,山东大学彭实戈院士给出了定义非线性数学期望的新思路,即通过倒向随机微分方程引入了非线性数学期望 G -期望与相应的条件 G -期望的概念,在一定范围内建立了动态非线性数学期望理论的基础。近些年发现, G -期望是研究递归效用理论与金融风险度量的有力工具。 G -期望保持了经典数学期望除线性性质之外的所有基本性质。但 G -期望的研究仍然是基于给定概率空间的,并且关于维纳测度由 G -期望确定的概率族是绝对连续的。换言之, G -期望的研究没有脱离概率空间。鉴于此,2005 年彭实戈又提出了一般的非概率框架下由马尔可夫链所产生的非线性数学期望理论。接着,他在 2006

年提出了 G -正态分布和 G -期望的概念。关于维纳测度由 G -期望确定的概率族一般并不绝对连续,但仍具有动态相容性。在这一框架下,他定义了 G -期望下的 G -布朗运动,建立了 G -布朗运动的随机积分,得到了相应的 Ito-公式。这一概念受到数学家和金融学家的广泛关注,可用它来研究证券的收益率和波动率在不确定情形下金融风险的度量问题和随机波动问题。在理论方面,它是柯尔莫哥洛夫概率公理体系的推广。最近,彭实戈还证明了大数定律和中心极限定律,并给出了这些定律令人满意的金融解释。

关于非线性数学期望及相应的随机分析研究目前相对集中在:进一步加强倒向随机微分方程理论的研究,研究 G -风险度量与其他风险度量之间的关系,为金融业选择合适的风险度量工具提供理论依据;开展倒向重随机控制系统的理论研究,以及各类反射倒向重随机微分方程及相关的随机偏微分方程理论研究,进一步刻画重随机干扰下随机控制问题,包括最大值原理及动态规划原理的研究,利用随机分析的核心工具研究倒向随机系统的滤波理论,进一步开展部分信息下随机控制问题的研究,包括部分信息下随机递归和随机控制问题的最大值原理以及线性二次控制问题等。另外, G -期望是一个新兴的研究方向,还处在初创阶段,还有许多基础理论问题尚未解决,需要进一步发展和完善 G -期望下的随机分析理论。

2.4 应用概率

应用概率研究的内涵和外延,一直随着历史和概率理论应用的发展而在不断地变化和发展。下面将集中介绍近期受到概率界广泛关注的几方面研究,即随机矩阵、随机复杂网络、渗流模型(Percolation Theory)和保险数学理论。

随机矩阵研究始于 1928 年维希特(J.



中国科学院

Wishart)关于多元统计分析的工作。上世纪30年代,中国数学家许宝騄在多元统计分析研究中获得了许多重要结果,对随机矩阵理论发展做出了重要贡献。50年代,受到核物理和量子力学研究的重大推动,随机矩阵理论在数学和理论物理两个领域中都获得了迅速发展^[12]。随机矩阵理论关注的核心问题是刻画满足某种对称性条件的高维随机矩阵如下相关量的渐近分布,它们是:特征值和特征向量、最大特征值、特征值之间的间隙等。特别是它们的普适性问题,即不同类型的随机矩阵模型的特征值和特征向量是否具有某种共同的渐近分布。1952年,物理学家维格纳证明了高斯型厄米随机矩阵模型(Gaussian Unitary Model)的特征谱测度弱收敛到半圆周律。随后,又有一些其他类型随机矩阵模型的特征谱和最大特征值的渐近分布被确定。非常有趣的是,蒙哥马利(H. Montgomery)关于黎曼- ζ 函数非平凡零点对分布规律的关联猜想,被物理学家和随机矩阵专家弗里曼·戴森(Freeman Dyson)指出,其密度函数正好是高斯型厄米随机矩阵模型特征值对的关联密度函数。后来,奥德里兹科(A.M. Odlyzko)做了大量的数值计算,经过适当的归一化后,验证了黎曼- ζ 函数非平凡零点的间距分布与高斯型厄米随机矩阵模型特征值间距分布确实相同。这为蒙哥马利所猜测的零点分布与随机矩阵理论间的联系提供了大量的数值证据,其被称为蒙哥马利-奥德里兹科定律(Montgomery-Odlyzko Law)。该定律说明,黎曼- ζ 函数非平凡零点与一个典型随机矩阵的特征值相对应。这与希尔伯特-波利亚猜想(Hilbert-Polya conjecture)很相似。这一猜想预言,黎曼- ζ 函数非平凡零点与某个厄米算符(Hermitian operator)的特征值相对应。这促使一些学者利用随机矩阵理论去研究著名的黎曼- ζ 函数猜想。随机矩阵理论目前的研究包括:随机多项式、随机全纯函数及全纯截面的零点分布,随机矩阵与典型域上的积分理论,酉群上的Harish-Chandra-Itzkson-Zuber积分,以及随机矩阵与曲线模空间上的相交理论等。

随机复杂网络理论始于1998年沃兹(D.J. Watts)和斯道格兹(S.H. Strogatz)在《自然》的一篇文章,在文章中引入了一个“小世界”网络模型,用于描写人际关系网络。1999年,鲍劳巴希(A-L. Barabasi)和艾伯特(R. Albert)在《科学》上又发表了一篇文章,引入了无标度网络的模型,用于刻画互联网和科学家之间合作关系的网络。随后,随机复杂网络研究便成为热点。许多领域中都发现了这种网络结构,比如基因调控网就是无标度网络等^[6]。这些大规模网络是不断地发生变化的,且从对它们的统计研究结果来看,它们整体上呈现出类似于随机变动的行为。因此,可以选择某个典型网络的随机模型来研究它们。随机复杂网络研究目前集中在3个方面:(1)利用现代概率论思想与方法研究小世界网络、无标度随机网络和演化网络等随机复杂网络的构造及特征性质;寻求新的方法来研究小世界网络、无标度随机网络和演化网络等随机复杂网络若干重要特征刻画的解析表达式,探索构造演化网络的新机制,建立更符合现实世界的网络模型。(2)构建群体遗传中新的随机复杂网络模型及其统计推断方法,并应用于同物种内不同亚种生物进化历史研究;利用随机复杂网络的概率特征,研究不同类型生物基因组和蛋白质组的差异和功能预测;利用随机复杂网络方法研究基因动态调控模型、非编码区的遗传功能及其与编码基因的关系;利用随机复杂网络的形成机制与结构稳定性,构建干细胞等重要生物体的演化模型并研究它们的生物功能特征。(3)考察计算机病毒在互联网或邮件网络中的传播与计算机黑客对计算机网站进攻的演化过程,SARS、禽流感和艾滋病等恶性传染病在人群构成的复杂网络中的传播,以及信用风险与非法资金在金融机构形成的复杂网络中的传播和扩散等;建立描述复杂网络中随机过程的数学模型并研究其基本性质,要求其涵盖上述现实网络中重点关注信息的传播和扩散现象。

渗流模型具有十分明确的应用背景,用于描述地下岩层中石油流动的不规则通道,在如何轰炸和

封锁跑道等军事问题的研究中也应用。在理论上,渗流模型展示相变现象。相变是统计物理中的一个重要概念,它是指当某些参数连续变化时,系统骤然间发生了巨大变化,表现为某个宏观观察量的不连续性。在渗流模型中这个参数就是每条边为开通的概率 p 。当 p 很小时,开通的边所连接而成的连通分支(open cluster)是有限的;而当 p 上升且超过临界值时,该连通分支包含了无穷多个顶点。渗流模型主要研究临界值和连通分支的大小以及一系列相关量的估计。自上世纪90年代以来,渗流模型的研究范围被大大扩展。人们发现渗流模型在许多其他学科的应用,可由此挖掘出更加精细且丰富的性质。渗流模型的研究相对集中在两方面:(1)研究与验证随机劳威纳演化有关的渗流模型的精准的估计。渗流模型中与随机劳威纳演化紧密相关,例如,三角形格点上渗流模型的标度极限已被证明是随机劳威纳演化。目前人们还只能对几个特殊模型验证随机劳威纳演化理论成立。验证随机劳威纳演化通常需要4个步骤,其中一步就是要建立非常精准的估计,而这需要对渗流模型有深刻的认识。(2)研究渗流模型中无穷连通分支上的随机游动以及其他随机过程。无穷连通分支虽然很不规则,人们却相信,它与原图具有相同或相近的性质,其上的随机过程也应该具有相同或相近的性质。例如,三维欧氏格点上无穷连通分支的随机游动是非常返的,随机游动的热核估计也与普通欧氏格点上随机游动具有相同的形式,以及不变原理依然成立。换言之,经过适当的标度变化,无穷连通分支上的随机游动非常接近布朗运动。如果把随机运动替换为其他随机过程,如无穷粒子系统,则许多问题有待研究,例如不变分布的存在唯一性,其他分布收敛到不变

分布的速度等。

保险数学理论是保险公司对其风险进行定量分析和预测并根据这些结果管理与控制风险的一般理论。它的主要研究内容包括,保险模型的建立、破产理论、分红理论、风险分析以及决策与风险控制等。破产理论是经典保险理论的核心内容。瑞典精算师林德伯格(F. Lundberg)1903年的博士论文是破产理论研究的源头。利用随机过程理论,1955年瑞典概率论学家克莱默(H. Cramer)建立了林德伯格破产理论的严格数学基础。后来,利用随机过程与分析理论,破产理论得到突飞猛进的发展。通俗地讲,在忽略投资回报、利率和通货膨胀等金融因素影响的前提下,破产理论从理赔的角度研究了破产风险对保险公司偿付能力的影响。它主要研究保险公司所关心的如下几个精算量:破产概率、破产时、破产前余额以及破产赤字等。然而,随着金融保险业的空前发展和金融业一体化进程的加快,保险公司面临着如何动态地采取更为有效的方式,来规避风险以达到自身效益的最大化^[13]。因此,破产理论中原有的几个静态保险精算量的刻画,已经不再能满足保险公司的需求。保险公司需要进行风险决策与控制。事实上,许多金融保险公司所面临的风险决策与控制问题,均可转化为相应的随机控制问题。离散时间的随机控制理论称为马尔可夫决策理论,主要研究受控制马尔可夫链的优化问题。连续时间的随机控制理论,主要利用马尔可夫过程控制理论和粘性解理论进行研究。保险数学理论目前的研究包括,多维风险模型的研究,博弈论应用于再保险市场的研究,以及对投资连结保险中最优控制问题的研究等。

2.5 极限理论

概率论中的极限理论从概率论诞生起



中国科学院

就一直是一个重要研究方向。它直接为统计学中的大样本理论和统计推断理论等提供理论支撑,许多研究内容来自于统计学的需求。因此,研究内容很丰富,我们仅介绍较熟悉的两个方面。

斯坦(C. Stein)方法和自正则化极限理论。利用正态分布的分部积分公式,斯坦于1972年精确估计了不同概率逼近的差异。特别地,他估计了独立随机变量和与正态分布的差异,得到了著名的中心极限定理以及一致与非一致的Berry-Essen界。斯坦方法是一种精确估计各种概率逼近的有力工具,无论对独立与相依随机变量列还是正态逼近与非正态逼近都适用。同时,利用斯坦方法可以研究概率逼近的绝对误差和相对误差。自正则化方法则是用随机变量列部分和的经验方差取代理论方差,来正则化随机变量列的部分和。自正则化极限定理的主要优点是,不像经典的极限定理,对随机变量矩条件的假设无要求或要求很弱。将斯坦方法和自正则化方法结合,可以获得许多假设条件很弱的概率逼近的绝对误差和相对误差估计。很多时候,斯坦方法、自正则化方法再结合正态随机变量的小概率估计方法,可以解决组合学和离散数学等领域中难度很大的问题。目前,这一方面相对集中的研究,是对来自于海量复杂数据统计理论的问题的研究,以及随机组合概率模型和统计力学模型标度极限理论中的相关问题的研究。例如,随机配置中最优配置方差的精确渐近界与中心极限定理,欧氏空间上极小生成树和最短路径的统计量概率极限性质,不同概率测度下随机划分的渐近性质,随机划分和随机矩阵之间内在联系的刻画,以及随机划分所生成的点过程及其极限点过程和极限过程分布性质的刻画等。

大偏差理论。大偏差理论的研究,可以追溯到20世纪30年代初关于随机变量尾概率的精确渐近性的研究。在克莱默、萨诺夫(I. N. Sanov)和斯莱德(M. Schilder)等人工作的基础上,瓦拉德汗(S. R.S Varadhan)于1966年提出了大偏差原理的现代定义。70—80年代,唐斯克(M.D. Donsker)和瓦拉

德汗为了认识薛定谔算子第一本征值的变分公式与大偏差的关系,建立了马尔可夫过程大时间渐近行为的大偏差原理。为了研究狄利克雷边值问题基本解的小时间渐近性质,瓦拉德汗建立了扩散过程的小时间大偏差原理^[15]。马尔可夫过程的大偏差理论已被广泛应用于许多相关问题,例如,维纳Sausage问题、极问题与流体力学极限等。马尔可夫过程的大偏差理论还被推广到平稳过程、动力系统、吉布斯测度等。对一些随机环境中随机游动和扩散过程、随机矩阵以及无穷交互粒子系统模型也建立了大偏差和中偏差原理。70年代,为了研究动力系统的随机扰动问题,在斯莱德的工作基础上,弗瑞德林(M.I. Freidlin)和温茨尔(A.D. Wentzell)系统地发展了随机扰动的大偏差理论。这方面的结果已经成为研究动力系统随机扰动稳定性、排队网络等相关问题的一个强有力工具。20世纪末和本世纪初,为了研究随机过程的轨道大偏差,随机过程大偏差的弱收敛方法已被提出,并对一些随机偏微分方程建立了Freidlin-Wentzell型的大偏差估计。目前,大偏差理论相对集中于研究来自海量复杂数据统计理论中相关精确估计的问题,以及统计力学模型标度极限研究中相关精确渐近估计问题等。

3 现代统计研究的主要方法与发展状况

统计学在埃贡·皮尔逊和纽曼建立了一般假设检验的数学理论后迅速发展,逐步突破了经典理论的标准假设。例如,处理的数据由数值向多维向量甚至无穷维向量和各种集合值发展,以及由静态数据向时间序列数据发展,如时空数据、函数型数据、多尺度数据、纵向数据、区间数据、图像数据和文本数据等;估计的类型由点估计向半参数估计和非参数估计发展;数据与残差的性质由独立性与正态性向各种相依性和非正态性发展,以及由简单性向各种复杂性发展;数据的规模由适当规模向小子样和海量发展;数据的完整性由完全数据向各种非完全数据发展;以及模型假设由线性假设向非线性假设甚至无模型假设发展等。从历史角度看,统计学理

论和方法的研究热点,与自然科学、工程技术和 社会经济等领域的热点需求是一致的,而且统计学的研究重点是统计方法,而统计理论对一般统计方法起支撑作用。这与数学的其他分支有所不同。所以,在此仅介绍现代统计学的主要方法,而且选择我们比较了解的主要方法加以介绍。

3.1 大维数据分析与随机矩阵的谱理论

由于当代计算技术的迅速发展,可以搜集、存储和处理海量的高维数据,而且数据的维数可以随着样本量的增加而迅速增加。这类数据被称为大维数据。通俗地讲,对大维数数据,数据的维数上界与样本量之比趋于无穷。这与经典统计的假设完全不同,也使许多经典的统计方法失效。在这种情况下,大维随机矩阵的谱分析理论成为目前唯一可用于大维数据分析的极限理论。通过发展大维随机矩阵的谱分析理论,并将其用于大维数据的样本协方差阵分析,人们发现,像均值与协方差阵的似然比检验都需要修正。不仅如此,实际中样本量永远不会是无穷,人们还需要发展对数据大维性的判别方法。通过大维随机矩阵的谱分析理论,人们可找到适当的统计方法来完成此事。为了扩展大维数据分析的应用范围,就需要逐步放宽对样本协方差阵的限制条件,这导致一些已有的大维随机矩阵的谱分析理论不适用于这样的大维数据分析。因此,需要进一步发展相应的大维随机矩阵的谱分析理论^[16]。下面的问题需要重点进行研究,即样本协方差阵的弱 Haar 猜测,以及 Tracy-Widom 律和特征根间隙极限的普适性。

3.2 大规模数据分析与降维技术

目前,几乎各个科学与工程领域都普遍存在如下的数据,即表面上数据量很大,但由于对可以唯一确定所考察对象状态的适当特征数目几乎一无所知,而只知道几

乎所有可能来确定该考察对象状态的特征;然而,这些几乎所有可能的特征数目巨大,平均到每个特征的样本量很少,从而导致经典的统计方法失效。通俗地,这种类型的数据被称为大规模数据。为分析这样的数据,人们需要发展统计方法,去估计可唯一确定所考察对象状态的适当特征数目以及可包含所给数据的适当空间,使得可对这些数据进行可靠的统计分析。粗略讲,用来达到上述目标的统计方法被称为降维技术,可包含所给数据的适当空间被称为中心降维子空间,其维数被称为结构维数。人们已提出有效的统计方法,处理相关问题,如,识别中心降维子空间的切片逆回归方法,估计结构维数的贝叶斯(T. Bayes)信息型准则等^[16]。目前,需要进一步研究的问题很多。例如,当数据的表现维数相对于样本量非常大甚至大于样本数时,如何进行降维。

3.3 变系数模型

变系数模型是近年来受到重视的高维数据回归分析的一个新的建模技术,是线性模型的推广。其特点是,模型的回归系数是某些因子的函数。这一模型既保留了参数模型容易解释的优点,又保留了非参数回归模型的灵活性与稳健性,还可以减少建模偏差和避免维数祸根。另外,它还包含了可加模型、部分线性模型、单指标系数回归模型以及自适应变系数线性模型等。已有成果包括了独立数据和纵向数据,并对回归系数估计使用了许多有效的非参数估计方法^[16]。值得研究的问题是,发展一些复杂数据下变系数模型和半参数变系数模型的统计方法。

3.4 纵向数据模型的稳健推断

对一系列不同的个体在不同时间点上重复观察获得的数据称为纵向数据或面板数据。纵向数据在生物医学、临床实验和社会经济等领域大量出现,需要对其进行统计



中国科学院

分析。分析纵向数据的目的在于,针对单一个体获得其随时间变化的规律;研究个体间的差异;确定关注变量与其他变量的关系。若把纵向数据按个体顺序排成列,再按时间顺序排成矩阵,则它是列相关且行独立的随机矩阵。因此,不同于一般的时间序列数据,需要专门的统计分析方法。根据上述特点可知,纵向数据的内部相关性,一个来自于个体间的差异,一个来自于单一个体的不同时间的差异。前一个用回归系数的随机性来刻画,后一个用测量误差来反映。根据不同要求,可用很多方法对纵向数据进行建模和推断^[16]。从纵向数据出现的领域及要达到的目的看,所建模型要对数据的分布律有一定的稳定性。因此,要求所建模型的推断具有稳健性。目前,已有一些建模方法符合上述要求。进一步研究的问题有,不完全纵向数据的分析以及纵向数据模型的模型选择。

3.5 测量误差模型及其统计推断方法

在对经济学、流行病学、化学与工程等领域的许多数据进行分析时,常出现以下情况:理论或经验知识告诉人们,可以用回归方法进行分析,但结果总不能令人满意。出现上述情况的一个较大的可能性是,获得的数据带有一定的误差,未能较真实地反映所考察对象的信息。这就是所谓的带测量误差的数据,对其所建的模型就称为测量误差模型。对测量误差的几种较简单类型,虽然问题具有难度,但仍给出了有效的统计方法去处理它们^[16]。目前,所研究的测量误差的类型在不断变复杂,值得深入研究。特别是,当测量误差是非线性复杂数据时,对它的各种统计推断方法和稳健统计方法的研究是很值得深入研究的。

3.6 缺失数据的回归分析

当多维数据的部分分量因某种原因无法完全获得时,这类数据称为缺失数据,其在医学和社会经济等领域普遍出现。由于删除这些无法完全获得的数据会造成信息损失,人们需要根据不同的缺失机制来研究此问题。通常考虑如下3种缺失机制:(1)完全随机缺失(MCAR)机制,指的是,数据

缺失不依赖其他任何变量;(2)随机缺失(MAR)机制,指的是,数据缺失仅依赖被观察到的部分而不依赖可能缺失的部分。这种情况在现实中更常见。最后是,不可忽略缺失或非随机机制,指的是,数据缺失依赖可能缺失的部分。一般,人们考虑随机缺失机制,可以根据情况采用似然方法、插补方法、逆概率加权法以及经验似然方法来分析^[16]。

3.7 复发事件数据的统计分析

复发事件数据是指,对一些个体进行观察时,一些令人感兴趣的事件重复发生的时间组成的数据。这类数据经常出现在生物和社会经济等领域。由于这类数据有次序和相依性,同时由于删失了很多时间且这些删失的时间可能与该事件发生的时间具有相依性,使得复发事件数据的分析、建模和统计推断变得困难。然而,复发事件数据既具有重要特点又有广泛应用,对它的研究受到高度重视。近十几年来,随着生物和医学统计的进步,复发事件数据的研究已取得快速发展。这些研究,主要根据对时间间隔的假设以及事件发生的时间与删失时间相依程度由简单到复杂的顺序开展^[16]。

3.8 因果推断

寻找事物之间因果关系,几乎是所有科学研究的最重要目的之一。这是因为,当作为原因的事件发生时,利用因果关系我们可以断言或预测作为结果的事件必然发生。然而,当事件之间仅知道存在相关和关联关系时,人们不好做出上述的断言或预测。根据费歇尔提出的试验设计的随机化原则,我们知道,只要可以控制试验使其完全随机化,人们仍可以利用试验获得的数据推断事件之间的因果关系。事实上,由于很多原因人们无法使试验完全随机化,只能获得观测的数据。对这类观测数据寻找统计方法,努力对事件之间的因果关系作出推断,就是因果推断。统计推断的目的是研究变量之间的相关和关联关系,而因果推断则是研究变量之间的因果机制。关于因果推断的模型主要有3个:(1)潜在结果模型,通过引入潜在变量推断因果关系;(2)因果网络模型,利用有向非循环图来描述因

果关系 ;(3) 格兰杰(C.W. Granger)因果模型,利用对两个事件观察获得的时间序列数据之间的预测关系,来推断这两个事件之间的因果关系。除了因果模型的研究,还需要深入研究现实中出现的替代指标问题和混杂因素的判断问题^[16]。

3.9 时间序列分析

一般根据时间顺序获得的数据称为时间序列数据,对它的统计分析就是时间序列分析。勿用置疑,时间序列数据遍布各个领域,它是人们经常遇到的数据^[17],长期以来在金融领域尤其受关注。受资产定价和风险度量理论的影响和推动,各种自回归条件异方差时间序列模型被提出,并被广泛研究。原因是,这类模型可以用来揭示和估计蕴藏在金融资产价格过程中的风险变动情况^[18]。这其中比较著名和常用的是ARCH模型和GARCH模型。事实上,在很合理的经济假设下,可以从理论上证明,资产的价格序列符合GARCH模型。对这类事件序列的理论研究需要运用现代马尔科夫过程理论和遍历理论。另外,海量复杂的金融资产价格序列在不同时间尺度上表现出不同的高频震荡特征,也向这类数据的时间序列分析与建模提出了挑战,需要发展新的时间序列模型去研究分析它。

3.10 小子样数据的可靠性推断

在实际中,有许多产品不可能通过大量的试验来验证其可靠性,只能进行很少的试验。如何从这些少量的试验数据来推断该产品的可靠性?研究这一问题的统计理论与方法就是小子样数据可靠性推断的理论与方法。这一理论与方法的基本思想是:所考察的产品一般有许多部件,这些部件和它们之间的相互关系形成了一个网络,通过分析每个部件生产过程中积累的历史数据,或对一些关键部件进行适量的试验,就可获得

各个部件的可靠程度与寿命分布。然后,利用概率统计知识和部件的网络关系图,就可以得到该产品的可靠程度与寿命分布。最后,结合这些少量的试验数据就可以对该产品的可靠性做出推断。在部件的网络关系图不太复杂时,上述思想行之有效,但当比较复杂且部件数量巨大时,如何实施上述思想需要在理论和方法上进一步研究。

4 我国概率统计研究队伍现状及建议

我国现代概率统计的教学科研始于许宝录,1940年底从英国回到西南联大执教。在1956年制定的《全国科学发展十二年远景规划》中把概率统计列为数学的3个重点发展方向之一。随后,许宝录在上级和同行专家大力支持下,招收来自全国的进修教师和学生在北京大学集中开班,延聘外国知名学者来华讲学,开设现代概率统计课程,组织概率统计教材建设和学术讨论班,带领青年教师和学生开展概率统计研究^[19]。目前,一支活跃在教学科研一线的高质量的概率统计教学科研队伍已经形成。当选为中科院院士的概率统计学者有6位,获国家自然科学基金委创新群体资助的概率统计研究团队有3个,获国家自然科学基金委杰出青年资助的概率统计学者有17位。还有4个概率统计的省部级重点实验室,4人次担任科技部973项目首席科学家。中国的概率统计研究处于国际学术前沿,部分研究工作处于国际领先地位。

然而,无论是从国际概率统计的发展趋势还是从我国建设创新型国家的需求来看,我国的概率统计研究还有很大的提升空间,鉴于概率统计在数学领域的特殊地位及其重大的应用价值,建议国家有关部门应着手筹建概率统计国家重点实验室,以更有效地组织队伍为我国的概率统计学科的崛起奠基,为我国经济社会发展做出更大贡献。



中国科学院

主要参考文献

- 1 数学手册编写组. 数学手册. 北京:高等教育出版社,1979,782-920.
- 2 Hey T, Tansley S, Tolle K. The Fourth Paradigm,2009.
- 3 约翰·塔巴克. 杨静译. 概率论和统计学 不确定性科学.北京:商务印书馆,2008.
- 4 Proceedings of ICM 2006,. Madrid,.I.. European Mathematical Society, 2007.
- 5 Proceedings of ICM 2010, Hyderabad, I. Hindstan Book Agency, 2010.
- 6 National Research Council. Network Science,2006.
- 7 LHC 网站.http://lhc.web.cern.ch/lhc/
- 8 GenBank 网站http://www.ncbi.nlm.nih.gov/genbank/
- 9 黄志远. 随机分析学基础(第二版). 北京: 科学出版社,2001.
- 10 Schramm O. Conformally invariant scaling limits: an overview and a collection of problems. Proceedings of ICM 2006 Madrid, 2006, I: 513-543. European Mathematical Society, 2007.
- 11 Kallenberg O. 现代概率论基础(Foundations of modern Probability)(影印版). 北京:科学出版社, Springer,2001.
- 12 Guionnet A.. Large deviations and stochastic calculus for large random matrices Probability Surveys, 2004, 1: 72-172.
- 13 Aase, K K. Perspectives of risk sharing. Scand. Actuarial J.2002, 2: 73-128.
- 14 Rajendra Bhatia. Abstracts, Plenary Lectures, Invited Lectures, Panel Discussions. Hindstan Book Agency, 2010.
- 15 Eills R S. Entropy, Large Deviations and statistical Mechanics. New York: Springer-Verlag,1985.
- 16 王启华,史宁中,耿直主编. 现代统计学基础. 北京:科学出版社,2010.
- 17 安鸿志,陈敏. 非线性时间序列分析. 上海:上海科学技术出版社,1998.
- 18 Levine A G, Finance's Quant(um) Mechanics. Science, 2008,322.
- 19 许宝騄先生纪念文集编委会. 道德文章垂范人间. 北京:北京大学出版社,2010,286-293.

The Research and Development of Probability and Statistics

Gong Fuzhou

(Academy of Mathematics and Systems Science, CAS 100190 Beijing)

Abstract This paper introduced strategic status and developing trends of Probability and Statistics, as well as research directions, research methods, and the state of the art of Probability and Statistics. Furthermore, some suggestions to develop Probability and Statistics of China were given.

Keywords probability and statistics, the state of the art, developing trends

巩馥洲 中科院数学与系统科学院应用数学所研究员,所长。中科院随机复杂结构与数据科学重点实验室主任,中科院国家数学与交叉科学中心数学与生物/医学交叉研究部主任,国家自然科学基金委杰出青年基金获得者。主要研究领域为随机分析及其应用,关注随机分析与马尔可夫过程的遍历理论及其在量子场论、统计力学、计算生物学、信息科学、经济学以及金融学中的应用。发表学术论文30余篇。E-mail: fzgong@amt.ac.cn