

## 谈谈概率论与其他学科的若干交叉

陈木法 \*

(北京师范大学数学科学学院, 北京, 100875)

**摘要:** 近一二十年以来, 概率论获得了很大发展, 特别是与其他学科交叉融合, 形成了一些新的学科分支和学科生长点. 我们首先从 2002 年国际数学家大会 (ICM2002) 所反映的情况予以说明. 作为这种交融的一个侧面, 也概述我们研究群体的三项成果. 最后介绍取得这些成果的一种数学工具及其与线性规划和非线性偏微分方程等学科的联系.

**关键词:** 概率论; ICM2002; 随机算法; 自由概率; 渗流; 特征值; 遍历理论; 最优运输

**MR(1991) 主题分类:** 60-XX / 中图分类号: O211

**文献标识码:** A      **文章编号:** 1000-0917(2005)06-661-12

### 1 从 ICM2002 看概率论

#### 1.1 概述

在 ICM2002 的开幕式上, 就可以感受到概率论的气息. 作为 Nevalinna 奖的获奖人, M. Sudan 的两项主要贡献中的第一项就是关于 NP 类的概率特征的刻画.

在 20 个一小时报告中, 有六个涉及概率论的. 名单及所代表的领域列表如下:

A. Alon	离散数学
L. A. Caffarelli	偏微分方程
U. Haagerup	算子代数
S. Goldwasser	计算机科学
H. Kesten	概率
D. Mumford	认知科学

第一个报告介绍两种方法, 其一是概率方法 (随机图论), 另一是代数方法. 最后一个报告贯穿了概率方法. 我们将在本文之末说明第二个报告与概率论之间的联系. 对于其他三个报告, 将在这一部分的随后几节中予以说明.

在 19 组 45 分钟报告中, 除数学教育、数学普及和概率统计 3 组外, 有 6 组都涉及概率论. 其中“计算机科学中的数学”一组, 六个报告中有五个与概率论有关, 而“算子代数与泛函分析”一组, 六个报告中两个“自由概率”, 一个“高斯测度不等式”, 即有三个报告是概率论的交叉学科方向.

#### 1.2 概率与随机算法和计算复杂性

收稿日期: 2004-08-03.

基金项目: 国家自然科学基金委员会创新研究群体项目 (No. 10121101)、973 项目和教育部博士点基金项目.

E-mail: mfchen@bnu.edu.cn Home page: <http://math.bnu.edu.cn/~chenmf>

\* 本文是第十二次院士大会、数学物理学部的学术报告 (2004 年 6 月 5 日).

NP 问题的典型例子是

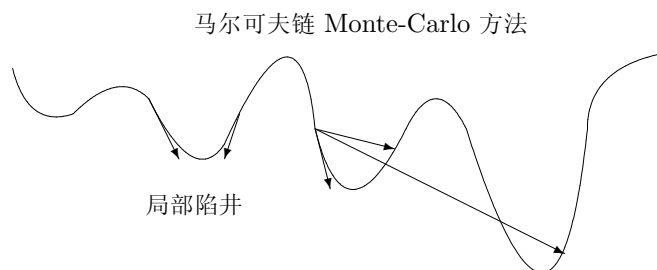
**货郎担问题:** 给定全国 144 个城市, 找出一条经过所有城市而又不迂回的最短闭路.

总共 144 个城市, 这个数字并不大, 但它经组合起来的闭路则有  $143!$  条. 即使是每秒计算一亿亿 ( $100^9$ ) 条路的计算机, 也需要  $100^{111}$  年, 因而是一个典型的 NP 问题. 在组合最优化领域里, 存在大量的这类问题.

如何处理 NP 问题, 自然是一个严峻的挑战! 似乎无路可循. 正是在这个人们以为“山穷水尽疑无路”的地方, 随机的思想给我们带来了“柳暗花明又一村”. 想法是: 如果允许算法以小概率犯错误, 则可将一些 NP 问题转化为 P 多项式问题. 针对货郎担问题, 有一种模拟退火算法 (又称为马尔可夫链 Monte-Carlo 方法). 目标是求一个函数的最小值. 其原理为

- (1) 依函数值的大小确定一个概率分布  $\mu$ : 函数值越小, 取值越大. 此即是 Gibbs 分布原理.
- (2) 构造一马尔可夫链, 以  $\mu$  为极限分布, 即当时间趋于无穷大时, 这个马尔可夫链趋于取值为  $\mu$  的分布.

留心通常的算法是“那里小就往那里走”, 因而容易掉进局部陷阱. 此法的特点是要到处看看.



当然, 仍然有一些细节需要处理. 例如, 需要“退火”, 即随时间的发展,  $\mu$  越来越集中于整体最小值等等. 对于上述货郎担问题, 利用这种方法找到一条长为 30421 公里的闭路. 这与目前所知的最好结果 30380 公里相差无几. 而真正的全局最优解依然是人们力所不能及的<sup>[4]</sup>.

我们指出, 这种算法的有效性取决于马尔可夫链收敛于平稳分布  $\mu$  的速度. 这个速度由马尔可夫链转移概率矩阵的第一个非平凡特征值所决定. 更详细的内容可参考 R. Kannan 的 45 分钟报告.

人们常说, 概率论是研究大量偶然现象中的必然性规律. 然而, 这里的研究对象却完全是确定性的, 毫无随机性可言. “随机性”思想的主动出击, 在这里得以充分体现. 这也是现代概率论研究的典型特征之一.

### 1.3 自由概率论 (Free Probability)

U. Haagerup 的一小时报告的题目是“Random matrices, free probability and the invariant subspace problem relative to a non Neumann algebra”. 我们先从随机矩阵谈起. 设  $A_N =$

$(a_{ij}^{(N)})$  是  $N$  阶 Hermite 方阵, 假定

$$\left(a_{ii}^{(N)}\right)_i, \quad \left(\sqrt{2} \operatorname{Re} a_{ij}^{(N)}\right)_{i<j}, \quad \left(\sqrt{2} \operatorname{Im} a_{ij}^{(N)}\right)_{i<j}$$

为独立同分布随机变量, 服从均值为零、方差为  $\frac{1}{\sqrt{N}}$  的正态分布. 命

$$\sigma(x, A_N) = \frac{1}{N} \#\{\text{特征值 (可重复)} \leq x\}$$

则有优美的

**E. Wigner 半圆律** (1955-1965): 随机变量  $\sigma(x, A_N)$  弱收敛于一个非随机的函数  $\sigma_W(x)$ , 其密度为

$$\sigma'_W(x) = \begin{cases} \frac{1}{2\pi} \sqrt{4-x^2}, & \text{若 } |x| \leq 2, \\ 0, & \text{若 } |x| > 2. \end{cases}$$

随机矩阵理论有两个重要来源, 一是矩阵力学, 另一是多元统计. 我国概率统计的前辈许宝騵先生是这一理论的早期开拓者之一 (1939). 这个理论甚至紧密联系于 Riemann 猜想.

若将复共轭视为  $*$  运算, 则随机矩阵自然构成典型的  $C^*$  代数. 欲将上述半圆律拓广到  $C^*$  代数上, 首先需定义  $C^*$  代数上的独立性. 这就引进了自由概率的概念: 换言之, 这里的“自由”即是通常概率论中的“独立性”. 这个概念及“自由熵”等首先由 D. Voiculescu (1985) 引入 (获 2004 年美国国家科学院奖), 并成功地应用于 von Neumann 代数的分类问题, 导出了“若干革命性成果”, 引发了大量的研究. 矩阵论作为非交换数学的基本工具是天然的. 然而随机矩阵论作为算子代数的基本工具则多少令人吃惊.

#### 1.4 概率论与物理

40 年来, 概率论与物理 (特别是统计物理) 的交融汇合, 产生出若干新的分支学科. 最具代表性的有随机游、相互作用粒子系统、渗流理论和测度值随机过程.

##### (1) 渗流理论 (Percolation Theory)

1982 年, H. Kesten 出版了专著 “Percolation Theory for Mathematicians”, 从数学上系统总结了已有的 (特别是物理学家) 所取得的成果. 自此以后, 渗流理论成为概率学家的一个专门的发展领域. 上世纪末, 渗流理论作为一个基本工具, 解决了相互作用粒子系统的一个著名难题.

渗流理论特别象数论, 问题很好懂、但却很难做. 还是从一个基本模型开始.

考虑  $d$  维的格子图. 给定每条边开的概率为  $p$  (闭的概率为  $1-p$ ). 各边开或闭相互独立. 如果一条路的依次相互连接的边都是开的, 则称为一个开串. 显然, 若  $p=1$ , 则所有的边都是开的, 因而存在无限长的开串. 反之, 若  $p=0$ , 则不存在开串. 这就引出临界值  $p_c$  的定义

$$p_c = \inf\{p : \text{存在包含原点的无穷开串的概率大于零}\}.$$

对于二维 ( $d=2$ ) 情形, 已知  $p_c = \frac{1}{2}$ . 但当  $d \geq 3$  时,  $p_c$  却是至今无人能够确定的. 若把边的开、闭换成格点的开、闭, 则上述边模型就变成点模型. 此时, 仅当两顶点均开时, 所联结的边才是开的. 对于二维三角形点渗流, 已知  $p_c$  也等于  $\frac{1}{2}$ . 通常, 物理学家知道得更多. 例

如, 他们不仅知道三角形点渗流的  $p_c = \frac{1}{2}$ , 而且还知道下式

$$\text{当 } p \downarrow p_c \text{ 时, 原点属于无穷开串的概率} = (p - p_c)^\alpha$$

中的临界指数  $\alpha = \frac{5}{36} + o(1)$ . 这是一种统计物理所研究的普通常数. 然而, 长时期以来, 数学家对普通常数束手无策, 研究状况处于完全真空的状态. 直到 2001 年, 才由 S. Smirnov 取得突破 (解决了物理学家 J. L. Cardy (1992) 基于共形场论的猜想). 他于同年荣获 Clay 研究奖 (颁布百万美元奖金的 7 大数学难题的研究所). 所使用的工具是布朗运动与共形映照. 这一点又很象解析数论, 使用复分析 (连续) 来处理数论问题 (离散). 这是 H. Kesten 的一小时报告和 G. F. Lawler 的 45 分钟报告的主题.

## (2) 研究相变的一种新方法

相变现象是统计物理的中心课题之一. 作为无穷维的数学, 研究相变现象的数学工具并不多. 十几年来, 逐步形成了一种新方法, 即以第一 (非平凡) 特征值来刻画相变. 例如对格子模型 (Ising 模型), 一维情形无相变. 事实上, 对于算子

$$Lf(\sigma) = \sum_{x \in \mathbb{Z}^d} c(x, \sigma) [f(\sigma^x) - f(\sigma)], \quad \sigma \in \{-1, 1\}^{\mathbb{Z}^d},$$

其中  $\sigma^x$  表示在  $x$  处的自旋,

$$c(x, \sigma) = \frac{1}{\left\{ 1 + \exp \left[ \beta \sigma(x) \sum_{|x-y|=1} \sigma(y) \right] \right\}}, \quad x \in \mathbb{Z}^d, \sigma \in \{-1, 1\}^{\mathbb{Z}^d},$$

而  $\beta > 0$  为反温度, R. A. Minlos 和 A. G. Trishch (1994) 甚至算出了第一特征值的精确值  $1 - \tanh \beta > 0$ . 文章只有两页, 但使用了第二量子化的漂亮技术, 构造此模型的  $L^2$  空间与圆周上的  $L^2$  空间所生成的反对称 Fock 空间之间的酉同构.

对于高维情形, 随着温度的下降, 人们普遍认为第一特征值应由正变为零. 目前已证出: 当温度足够高时, 第一特征值为正; 而当温度足够低时, 第一特征值为零. 关于后者, 事实上知道得更多. 当温度充分低时, 边长为  $L$  的正方体的格子区域内的 Ising 模型, 当  $L$  趋于无穷时, 其第一特征值有渐近式  $\exp[-c(\beta)L^{d-1}]$ , 其中  $c(\beta)$  是与维数  $d$  无关的常数 [5, 7, 8].

从这一节可以看出物理学对于当代概率论的深刻影响.

## 2 特征值估计与遍历性

众所周知, 谱理论在数学各分支和物理学中均有重要地位. 第一特征值乃是谱的主阶, 因而无疑有重要价值. 上述的随机算法和相变刻画, 也已显示其重要性.

### 2.1 问题与难度

考虑如下无限矩阵

$$Q = (q_{ij}) = \begin{pmatrix} -b_0 & b_0 & 0 & 0 & \dots \\ a_1 & -(a_1 + b_1) & b_1 & 0 & \dots \\ 0 & a_2 & -(a_2 + b_2) & b_2 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

其中  $a_k, b_k > 0$  ( $k = 0, 1, 2, \dots$ ). 因为仅有处于对角线附近的三条线的元素非零, 故称为三对角阵. 有限情形是计算数学处理矩阵特征值计算的最主要对象. 注意矩阵  $Q$  的每一行和为零, 因此它与元素恒为 1 的常数列向量  $\mathbf{1}$  的乘积为元素恒为零的列向量  $\mathbf{0}$ :  $Q\mathbf{1} = \mathbf{0} = \mathbf{0} \cdot \mathbf{1}$ . 即矩阵  $Q$  有平凡特征值  $\lambda_0 = 0$ . 其次, 若考虑它的前  $n$  阶子矩阵  $Q_n$ , 则  $-Q_n$  有  $n$  个特征值:  $0 = \lambda_0 < \lambda_1 \leq \dots \leq \lambda_{n-1}$ . 我们所关心的是  $\lambda_1$ , 即第一个非平凡特征值.

为使大家对于此问题的难度有点具体的感受, 让我们看看一些简单例子. 先看四阶情形 ( $Q_3$ ), 此时有 6 个参数:  $b_0, b_1, b_2, a_1, a_2, a_3$ . 第一特征值是

$$\lambda_1 = \frac{D}{3} - \frac{C}{3 \times 2^{\frac{1}{3}}} + \frac{2^{\frac{1}{3}}(3B - D^2)}{3C},$$

其中  $D, B, C$  三个量的表达式并不复杂:

$$D = a_1 + a_2 + a_3 + b_0 + b_1 + b_2,$$

$$B = a_3 b_0 + a_2 (a_3 + b_0) + a_3 b_1 + b_0 b_1 + b_0 b_2 + b_1 b_2 + a_1 (a_2 + a_3 + b_2),$$

$$C = \left( A + \sqrt{4(3B - D^2)^3 + A^2} \right)^{\frac{1}{3}},$$

但

$$\begin{aligned} A = & -2a_1^3 - 2a_2^3 - 2a_3^3 + 3a_3^2 b_0 + 3a_3 b_0^2 - 2b_0^3 + 3a_3^2 b_1 - 12a_3 b_0 b_1 + 3b_0^2 b_1 \\ & + 3a_3 b_1^2 + 3b_0 b_1^2 - 2b_1^3 - 6a_3^2 b_2 + 6a_3 b_0 b_2 + 3b_0^2 b_2 + 6a_3 b_1 b_2 - 12b_0 b_1 b_2 \\ & + 3b_1^2 b_2 - 6a_3 b_2^2 + 3b_0 b_2^2 + 3b_1 b_2^2 - 2b_2^3 + 3a_1^2 (a_2 + a_3 - 2b_0 - 2b_1 + b_2) \\ & + 3a_2^2 [a_3 + b_0 - 2(b_1 + b_2)] \\ & + 3a_2 [a_3^2 + b_0^2 - 2b_1^2 - b_1 b_2 - 2b_2^2 - a_3(4b_0 - 2b_1 + b_2) + 2b_0(b_1 + b_2)] \\ & + 3a_1 [a_2^2 + a_3^2 - 2b_0^2 - b_0 b_1 - 2b_1^2 - a_2(4a_3 - 2b_0 + b_1 - 2b_2) \\ & + 2b_0 b_2 + 2b_1 b_2 + b_2^2 + 2a_3(b_0 + b_1 + b_2)]. \end{aligned}$$

这样, 诸参数对于  $\lambda_1$  的贡献就完全糊涂了. 当然, 对于 6 阶或 6 阶以上的情形 (因  $\lambda_0 = 0$  而多一阶), 根据伽罗瓦理论, 根本不可能写出显式解. 因此, 不可能指望把  $\lambda_1$  准确地算出来.

既然如此, 我们退而求其次, 即尝试估计  $\lambda_1$ . 现在考虑无限矩阵. 以  $\text{Degree}(g)$  表示  $\lambda_1$  的特征向量  $g$  的主阶 (如  $g$  为多项式). 下表的三个例子显示了  $\lambda_1$  和  $\text{Degree}(g)$  的摄动情况.

$b_i$ ( $i \geq 0$ )	$a_i$ ( $i \geq 1$ )	$\lambda_1$	$\text{Degree}(g)$
$i + c$ ( $c > 0$ )	$2i$	1	1
$i + 1$	$2i + 3$	2	2
$i + 1$	$2i + (4 + \sqrt{2})$	3	3

表中的第一行是著名的线性模型,  $\lambda_1 = 1$  而与常数  $c > 0$  无关, 相应的特征向量  $g$  是一次多项式函数. 其次, 保持  $b_i = i + 1$  不变. 那么, 当  $a_i$  从  $2i$  变到  $2i + 3$  再变到  $2i + 4 + \sqrt{2}$  时,  $\lambda_1$  依次从 1 跳到 2 再跳到 3. 更奇妙的是特征向量依次从一次跳到二次再跳到三次多项式. 至于  $a_i$  取值介于  $2i, 2i + 3$  和  $2i + 4 + \sqrt{2}$  之间时, 情况更糟, 我们根本不知道  $\lambda_1$  为何值, 因

为此时特征向量  $g$  并非多项式而不知如何计算. 这样, 在一般情况下,  $\lambda_1$  及其特征向量都是极为敏感的, 要估计  $\lambda_1$  也是极端艰难的.

## 2.2 特征值估计

很幸运, 我们能够在此专题上取得一些进展. 为陈述三对角矩阵情形的主要结果, 需引进几个记号. 命

$$\mu_0 = 1, \quad \mu_i = \frac{b_0 \cdots b_{i-1}}{a_1 \cdots a_i}, \quad i \geq 1, \quad Z := \sum_{i \geq 0} \mu_i,$$

$$\mathscr{W}'' = \{w : w_0 = 0, w_i \text{ 为 } i \text{ 的严格增函数}\},$$

$$\mathscr{W}' = \{w : w_0 = 0, \text{ 存在 } k : 1 \leq k \leq \infty \text{ 使得 } w_i = w_{\min\{i, k\}} \text{ 且 } w \text{ 在 } [0, k] \text{ 上严格增}\},$$

$$I_i(w) = \frac{1}{\mu_i b_i (w_{i+1} - w_i)} \sum_{j=i+1}^{\infty} \mu_j w_j,$$

这里本质上只有两个记号  $\mathscr{W}''$  和  $I(w)$ ,  $\mathscr{W}'$  只是将  $\mathscr{W}''$  中的函数 (数列) 从后面拉平. 记  $\bar{w}_i = w_i - \sum_{i \geq 0} \mu_i w_i / Z, i \geq 0$ . 那么, 我们有如下结果:

**定理** (陈:1996-2001) 假定半群唯一:

$$\sum_{k=0}^{\infty} \frac{1}{b_k \mu_k} \sum_{i=0}^k \mu_i = \infty \quad \text{以及} \quad Z = \sum_{i=0}^{\infty} \mu_i < \infty.$$

则有

$$(1) \text{ 对偶变分公式: } \inf_{w \in \mathscr{W}'} \sup_{i \geq 1} I_i(\bar{w})^{-1} = \lambda_1 = \sup_{w \in \mathscr{W}''} \inf_{i \geq 0} I_i(\bar{w})^{-1}.$$

$$(2) \text{ 显式估计: } Z\delta^{-1} \geq \lambda_1 \geq (4\delta)^{-1}, \text{ 其中 } \delta = \sup_{i \geq 1} \sum_{j \leq i-1} (\mu_j b_j)^{-1} \sum_{j \geq i} \mu_j.$$

$$(3) \text{ 逼近程序: 可构造出显式序列 } \{\eta'_n\} \text{ 和 } \{\eta''_n\} \text{ 使得 } \eta'_n \downarrow \geq \lambda_1 \geq \eta''_n \uparrow \geq (4\delta)^{-1}.$$

容易看出, 第一条中  $\lambda_1$  的左、右两端分别用于上、下界估计: 对于每一个严格增的正数列  $w_i (w_0 = 0)$ , 代入 (1) 式的右端, 便可得出  $\lambda_1$  的一个下界估计  $\inf_{i \geq 0} I_i(\bar{w})^{-1}$ . 这就是“变分”一词的含义. “对偶”一词意指: 若交换“sup”和“inf”, 则上、下界的表达式互换, 只是  $\mathscr{W}'$  和  $\mathscr{W}''$  略有差别. 留意此公式与古典变分公式

$$\lambda_1 = \inf \left\{ \sum_{i \geq 0} \mu_i b_i (f_{i+1} - f_i)^2 : \sum_{i \geq 0} \mu_i f_i = 0, \sum_{i \geq 0} \mu_i f_i^2 = 1 \right\}$$

完全不同, 因而先前从未出现过. “显式”一词意指表达式只依赖于系数  $a_k$  和  $b_k$ . 将此定理与上面所论的例子作一对比, 很难想象能够得到这样简洁和彻底的解答.

对于直线上的椭圆算子, 结果是平行的. 下界变分公式也适用于高维情形. 作为一个代表, 这里陈述紧黎曼流形情形的一个结果. 分别以  $d, D, K$  表流形的维数、直径和 Ricci 曲率的下界, 则 Laplace 算子的第一特征值的下界有如下

**变分公式** (陈、王风雨:1997)

$$\lambda_1 \geq \frac{4 \sup_{f \in \mathscr{F}} \inf_{r \in (0, D)} f(r)}{\int_0^T C(s)^{-1} ds \int_s^D f(u) C(u) du},$$

这里用到两个记号:  $C(r) = \left( \cosh \left[ \frac{r}{2} \sqrt{\frac{-K}{d-1}} \right] \right)^{d-1}$ ,  $\mathcal{F}$  为  $[0, D]$  上正的连续函数的全体.

此公式不仅统一了、而且把几何学家 (包括 A. Lichnerowicz 和丘成桐等在内) 40 年来所得到的八种著名估计 (五种是最优的) 全部改进. 例如取  $f(r) = \sin \left( \frac{r\pi}{2D} \right)$  便可改进 Li-Yau 估计和钟家庆 - 杨洪苍最佳估计  $\frac{\pi^2}{D^2}$ <sup>1</sup>. 我们还确定了最佳线性下界:  $\frac{\pi^2}{D^2} + \frac{K}{2}$  ( $K$  任意)<sup>2</sup>. 我们的所有结果都适用于带边界流形. 但对于后者, 即使是钟 - 杨估计也未被证明, 更不用说新估计了.

这个变分公式是使用概率方法证明的. 这可视为概率论对于谱理论的应用. 我们将在第三部分以分析的语言介绍所使用的数学工具. 在以下两小节, 先简要介绍一下两个进一步的研究课题和进展.

### 2.3 遍历性关系图

考虑一般的概率空间  $(E, \mathcal{E}, \mu)$ , 以  $L^p(\mu)$  表通常的实  $L^p$  空间, 其范数记为  $\|\cdot\|_p$ . 我们需要用到对称型  $(D(f), \mathcal{D}(D))$ . 对于流形  $M$  上的拉氏算子, 对称型是

$$D(f) := D(f, f) = \int_M \|\nabla f\|^2, \quad f \in C^\infty(M).$$

一般的  $D(f, g)$  由四边形法则给出:  $D(f, g) = \frac{D(f+g) - D(f-g)}{4}$ . 对于欧氏空间  $\mathbb{R}^d$  中的自共轭二阶椭圆算子 (二阶系数为  $a(x)$ ), 对称型是

$$D(f) = \int_{\mathbb{R}^d} a|\nabla f|^2 d\mu, \quad f \in C_K^\infty(\mathbb{R}^d).$$

而对于对称测度  $J(dx, dy)$ , 我们有积分算子的对称型

$$D(f) = \frac{1}{2} \int_{E \times E} J(dx, dy) (f(y) - f(x))^2, \quad f \in L^2(\mu).$$

第一特征值  $\lambda_1$  的古典变分公式可改写成下述的 Poincaré 不等式,

$$\|f - \mu(f)\|_2^2 \leq CD(f), \quad f \in L^2(\mu),$$

其中  $\mu(f) = \int f d\mu$ , 最佳常数  $C_{\min} = \frac{1}{\lambda_1}$ . 这样, 自然要研究其他重要不等式. 于是我们就进入了一个范围更广的研究层次. 首先是比上式更强些的 L. Gross (1976) 的对数 Sobolev 不等式

$$\int_E f^2 \log \frac{f^2}{\|f\|_2^2} d\mu \leq CD(f), \quad f \in L^2(\mu).$$

1. 丘成桐获 Fields 奖的六项成就中的第四项是给出了第一特征值的 Li-Yau 下界估计  $\frac{\pi^2}{2D^2}$ , 后经钟家庆和杨洪苍改进为  $\frac{\pi^2}{D^2}$  (当曲率  $K = 0$  时达到最优). 后者是公认的精深结果, 也是钟获首届陈省身奖的两项主要成果之一. 使用变分公式再改进为  $\frac{\pi^2}{D^2} + \max \left\{ \frac{\pi}{4d}, 1 - \frac{2}{\pi} \right\} K$  (均指  $K \geq 0$  情形).

2. 陈、E. Scacciatelli 和姚亮: 2001.

这是通常的 Sobolev 不等式在无穷维空间的替代物, G. Perelman 在他的著名论文 (2002, arXiv: math.DG/0211159) 中也用到. 其次是著名的 J. Nash (1958) 不等式

$$\|f - \mu(f)\|_2^2 \leq CD(f)^{\frac{1}{p}} \|f\|_1^{\frac{1}{q}}, \quad \frac{1}{p} + \frac{1}{q} = 1, \quad f \in L^2(\mu).$$

我们使用几何方法 (拓广的 Cheeger 不等式、等周不等式等) 研究了这些不等式的最佳常数的估计.

这些不等式的重要性在于: 它们刻画了相应的马尔可夫半群的某种遍历性. 例如, 若对称型对应于半群  $\{T_t\}_{t \geq 0}$ , 则 Poincaré 不等式等价于  $L^2$  指数式收敛性 ( $\varepsilon_2 = \lambda_1, C_2 = 1$ ):

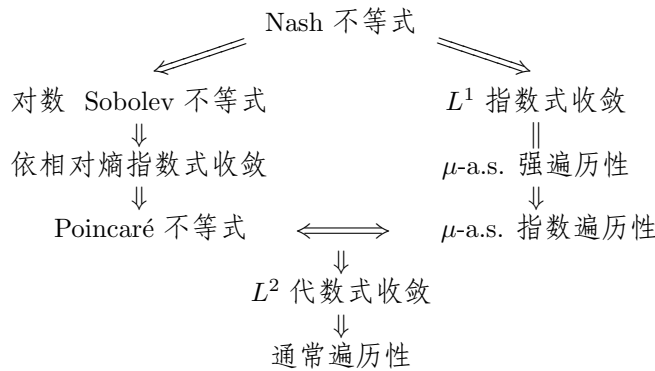
$$L^p \text{ 指数式收敛性: } \|T_t f - \mu(f)\|_p \leq C_p \|f - \mu(f)\|_p e^{-\varepsilon_p t}, \quad t \geq 0, f \in L^p(\mu).$$

由此进入更广的研究层次, 即研究马尔可夫半群 (过程) 的各种遍历性. 在传统的马尔可夫过程理论的研究中, 有以下三种遍历性.

$$\begin{aligned}
\text{通常遍历性: } & \lim_{t \rightarrow \infty} \|p_t(x, \cdot) - \pi\|_{\text{var}} = 0, \\
\text{指数遍历性: } & \|p_t(x, \cdot) - \pi\|_{\text{var}} \leq C(x)e^{-\alpha t}, \quad \alpha > 0, \\
\text{强遍历性: } & \limsup_{t \rightarrow \infty} \sup_x \|p_t(x, \cdot) - \pi\|_{\text{var}} = 0 \\
& \iff \lim_{t \rightarrow \infty} e^{\beta t} \sup_x \|p_t(x, \cdot) - \pi\|_{\text{var}} = 0, \quad \beta > 0,
\end{aligned}$$

此处  $p_t(x, dy)$  是马尔可夫过程的转移概率函数, 而  $\|\cdot\|_{\text{var}}$  是全变差范数. 问题是: 所有这些收敛性和遍历性之间有何联系? 下述结果给出了完整的解答.

**定理** (陈: 1999, 2002; 毛永华: 2002) 考虑对称 (细致平衡) 马尔可夫过程, 若其转移概率函数关于对称概率分布有密度, 则下述蕴涵关系成立.



此图是完备的. 有反例表明, 所有单方向的蕴涵关系都不可逆, 而无蕴涵关系者不可比较. 此图的价值是显然的. 例如利用等价关系, 可由 Poincaré 不等式的判别准则得出指数遍历性准则, 还可以得出遍历性速度估计, 后者先前所知的结果甚少. 另一方面, 利用等同性, 可由强遍历性推出关于  $L^1$  指数式收敛的各种性质, 后者是 Banach 空间, 其谱性质不易直接处理.

### 2.4 判别准则 (与显式估计)



接下来的问题是: 对于这些遍历性 (不等式), 能否给出判别准则? 下面仅就三对角阵情形, 将 10 个显式判别准则列表如下. 记  $\mu[i, k] = \sum_{i \leq j \leq k} \mu_j$ .

性质	判别准则
唯一性	$\sum_{n \geq 0} \frac{1}{\mu_n b_n} \mu[0, n] = \infty \quad (*)$
常返性	$\sum_{n \geq 0} \frac{1}{\mu_n b_n} = \infty$
遍历性	$(*) \& \mu[0, \infty) < \infty$
指数遍历 $L^2$ 指数式收敛	$(*) \& \sup_{n \geq 1} \mu[n, \infty) \sum_{j \leq n-1} \frac{1}{\mu_j b_j} < \infty$
离散谱	$(*) \& \lim_{n \rightarrow \infty} \sup_{k \geq n+1} \mu[k, \infty) \sum_{n \leq j \leq k-1} \frac{1}{\mu_j b_j} = 0$
对数 Sobolev 不等式	$(*) \& \sup_{n \geq 1} \mu[n, \infty) \log[\mu[n, \infty)^{-1}] \sum_{j \leq n-1} \frac{1}{\mu_j b_j} < \infty$
强遍历 $L^1$ 指数式收敛	$(*) \& \sum_{n \geq 0} \frac{1}{\mu_n b_n} \mu[n+1, \infty) = \sum_{n \geq 1} \mu_n \sum_{j \leq n-1} \frac{1}{\mu_j b_j} < \infty$
Nash 不等式	$(*) \& \sup_{n \geq 1} \mu[n, \infty)^{(q-2)/(q-1)} \sum_{j \leq n-1} \frac{1}{\mu_j b_j} < \infty \quad (\varepsilon)$

表中的“(\*) & ...”表示第一行中的唯一性条件加上条件“...”. 而最后一行中的“(ε)”表示仍有小空隙有待解决.

此表的前三个结果是经典的, 其余都是新的 (陈:2000–2002; S.G. Bobkov 和 F. Götze: 1999; 毛永华:2002; L. Miclo: 1999; 张余辉:2001), 解决了长期未果难题. 这里的“离散谱”意指谱集仅有有限重的特征值. 此表得益于调和分析中的加权 Hardy 不等式. 类似结果适用于一维椭圆算子. 使用容度理论, 我们也给出了高维情形的判别准则 (可惜常常非显式). 使用 Orlicz 空间理论, 还可给出更为广泛的、统一的判别准则.

### 3 一个数学工具: 概率距离与耦合方法

这一部分介绍得出上述第一特征值变分公式的数学工具以及与线性规划、偏微分方程等学科的联系.

#### 3.1 定义

给定距离空间  $(E, \rho)$  上的两个概率测度  $\mu_1$  和  $\mu_2$ , 可构造一个概率空间上的两个随机变量  $\xi_1$  和  $\xi_2$  使得  $\xi_i$  的分布为  $\mu_i (i = 1, 2)$ . 于是可定义  $\xi_1$  和  $\xi_2$  的通常的  $L^p (p \geq 1)$  距离:  $(\mathbb{E} \rho(\xi_1, \xi_2)^p)^{\frac{1}{p}}$ . 上述构造概率空间的手续实质上就是选择  $\xi_1$  和  $\xi_2$  的一个联合分布, 即  $\mu_1$  和  $\mu_2$  的一种耦合  $\tilde{\mu}$ , 它是乘积空间上的概率测度, 两个边缘分别是  $\mu_1$  和  $\mu_2$ : 对于一切可测集  $B$ , 有  $\tilde{\mu}(B \times E) = \mu_1(B)$ ,  $\tilde{\mu}(E \times B) = \mu_2(B)$  (最简单的是独立乘积  $\tilde{\mu} = \mu_1 \times \mu_2$ ). 为脱离参

考标架 (概率空间), 自然取  $\inf$ , 即定义

$$W_p(\mu_1, \mu_2) = \inf_{\tilde{\mu}} \left( \int_{E \times E} \rho(x_1, x_2)^p \tilde{\mu}(dx_1, dx_2) \right)^{\frac{1}{p}},$$

其中  $\tilde{\mu}$  跑遍  $\mu_1$  和  $\mu_2$  的耦合, 当  $p = 1$  时, 称  $W_1$  为 Wasserstein 距离. 这是 L. N. Wasserstein 在研究随机场时提出来的 (1969).

### 3.2 概率论

上述距离是 R. L. Dobrushin 于 1970 年命名的. 他还详细研究了这种距离的拓扑性质 (完备性、可分性和紧性) 及与弱收敛拓扑之间的关系等. 对于概率学家, 这个概率距离已经使用了 35 年.

这个距离拥有内在的几何特征. 例如在欧氏空间  $\mathbb{R}^d$  中, 取  $\xi_2$  为  $\xi_1$  的平移:  $\xi_2 = \xi_1 + a$ , 那么两者分布的  $W_p$  距离恰好等于平移的长度  $|a|$ .

### 3.3 线性规划

大约 10 年之后 (1980 前后), 人们发现若将  $\rho$  改为费用函数, 那么这个距离就变成 L. Kantorovich (1942) 所提出的最优运输问题, 他给出了强有力的对偶表示并和 G. Sh. Rubinshtein (1957) 作过深入研究, 只是限于紧空间情形. 因此, 有时也将上述距离称为 Kantorovich-Rubinshtein-Wasserstein 距离. 大家知道, 基于对线性规划所作出的贡献, L. Kantorovich 获 1989 年度 Nobel 经济学奖. 其实, Kantorovich 本人 (1948) 已经注意到, 他的问题可以追溯到 G. Monge (1781). Monge 的目标是寻求运输映射  $\Phi: \mathbb{R}^d \rightarrow \mathbb{R}^d$ , 使得  $\mu_2 = \mu_1 \circ \Phi^{-1}$  及运费达到最小

$$\inf_{\Phi} \int_{\mathbb{R}^d} |x - \Phi(x)| \mu_1(dx).$$

这样的  $\Phi$  若存在, 此即化为 Wasserstein 距离. 但一般情况下可能不存在. 换言之, Monge 问题乃是 Kantorovich 问题的加强形式. 百年之后 (1885), 巴黎科学院曾设奖征求 Monge 问题的解答. 特殊情形由 P. Appell (1887) 解决. 然而, 即使对于  $W_1$ , 其解答也只是在 200 年之后由 V. N. Sudakov (1979) 完成 (长达 178 页的论文).

### 3.4 偏微分方程

1990 年前后, 对于  $E = \mathbb{R}^d$ ,  $p = 2$  情形, 下述数学家  
Y. Brenier (1987, 1991),  
S. T. Rachev 和 L. Rüschendorf (1990, 1995),  
C. Smith 和 M. Knott (1987).

独立地证明了在适当的条件下, 存在凸函数  $\Psi$  (常常非唯一), 使得运输映射  $\Phi = D\Psi$  (常常几乎处处唯一).

更进一步, L. C. Evans 和 W. Gangbo (1999) 证明了  $\Phi$  满足下述的非线性 Monge-Ampère 方程 (弱解)  $f_2(D\Psi) \det D^2\Psi = f_1$ , 其中  $f_i$  为  $\mu_i$  关于 Lebesgue 测度的分布密度函数 (假定存在). 等价地,  $f_2(\Phi) \det D\Phi = f_1$ . 形式上看, 这只是“体积元”的变量替换 ( $x \rightarrow \Phi(x)$ ) 公式. 我们回到了原始的 Monge 问题. 这些新发展引发了大量文献, 也构成了 L. A. Caffarelli 的一小时报告和 L. Ambrosio 的 45 分钟报告的主要内容.

现在, 人们常常把这一研究专题统称为 Monge-Kantorovich (运输) 问题, 而依然把  $W_1$  称为 Wasserstein 距离.

### 3.5 统计

耦合 (coupling) 方法对于统计也有广泛的应用. 已出版专著 R. B. Nelssen (1999). 统计学家也常常使用 “copulas” 代替 “coupling”.

### 3.6 动态系统

至今为止, 所讨论的都是静态系统, 即不含时间  $t$ . 然而, 我们所走的是另一条路线, 研究动态系统, 即由某种算子生成的动态系统. 这里, 已知的是所给定的算子 (例如行和为零、非对角线元素非负的矩阵或二阶椭圆算子等). 问题是: 对于给定的两个算子  $L_1$  和  $L_2$ , 如何定义它们之间的 Wasserstein 距离? 结果是, “距离” 失去意义, 但却可以定义最优耦合. 设  $L_k$  为  $E_k$  上的算子 ( $k = 1, 2$ ). 一个乘积空间  $E_1 \times E_2$  上的算子  $\tilde{L}$  称为  $L_1$  和  $L_2$  的耦合, 如果将单变量 (有界) 函数  $f(x_1)$  视为双变量函数  $\tilde{f}(x_1, x_2) = f(x_1)$ , 则对于一切  $x_1 \in E_1$  和  $x_2 \in E_2$ , 有  $\tilde{L}\tilde{f}(x_1, x_2) = L_1f(x_1)$ . 简记为  $\tilde{L}\tilde{f} = L_1f$ . 类似地,  $\tilde{L}\tilde{f} = L_2f$ . 一个耦合算子  $\tilde{L}$  称为  $\rho$  最优 (陈:1994), 如果  $\tilde{L}\rho = \inf_{\tilde{L}} \tilde{L}\rho$ , 此处  $\tilde{L}$  跑遍一切  $L_1$  和  $L_2$  的耦合.

有了这一概念之后, 便可陈述第一特征值下界估计的耦合方法: 若存在耦合算子  $\tilde{L}$ , 距离  $\tilde{\rho}$  和常数  $\alpha \geq 0$  使得  $\tilde{L}\tilde{\rho} \leq -\alpha\tilde{\rho}$ , 则第一特征值  $\geq \alpha$  (陈、王风雨:1994). 由此可见  $\rho$  最优耦合算子  $\tilde{L}$  给出相对于  $\rho$  的最优估计. 我们完成了若干最优耦合算子的具体构造. 这是第一特征值变分公式证明的第一个要点. 另一个要点是需要变换距离  $\tilde{\rho}$  (事实上使用了一族距离), 因为收敛速度和 Wasserstein 距离都不是拓扑概念, 自然非常依赖于距离的选取.

**结束语** 一方面, 随机数学是在其他数学分支 (特别是分析) 的哺育下成长壮大的. 最近一个历史时期的代表性工作有 P. Malliavin (1977) 的 Malliavin 分析, S. Smale (1981) 等的概率计算复杂性和上述 D. Voiculescu (1985) 的自由概率论等. 这些原创者均非概率论出身. 另一方面, 人们常常误认为只是在说不清楚的地方才需要使用随机数学. 上面所介绍的几个方面, 展示了随机数学有时比决定性数学更精细, 也显示出随机性思想的重要性和威力. 概率论对于其他领域有重要影响的工作有大家比较熟悉的狄氏型理论, 还有关于完全非线性方程的 N. V. Krylov 和 M. V. Safonov (1979) 估计, 关于期权定价的 F. Black 和 M. Scholes (1973) 公式, 等等. 随机数学与其他学科之间的广泛的交叉渗透, 是一种很健康的现象, 不足为怪. 因为从理论上讲, 哲学的三大要素 (对立统一, 量变质变, 偶然与必然) 之一、物理学的两大理论 (相对论和量子论) 之一都包含随机性; 从实践中看, 如同许多人都深有感触地说, 数学的各分支和理论物理乃是一个统一的整体.

### 参考文献

为节省篇幅, 凡文集 [1] 已包含的文献不再列入.

- [1] Proceedings of “ICM 2002” I, II, III [M]. Higher Education Press, Beijing: 2002.
- [2] Chen Mufa(陈木法). Eigenvalues, Inequalities, and Ergodic Theory [M]. Springer, London: 2005.
- [3] Conrey J B. The Riemann hypothesis [J]. *Notices of AMS*, 2003, 50(3): 341–353.
- [4] 康立山等著. 非数值并行算法 (第一册) [M]. 北京: 科学出版社, 1994.
- [5] Martinelli F. Lectures on Glauber dynamics for discrete spin models [J]. *LNM*, 1999, 1717: 93–191. (Springer-Valerg)
- [6] Minlos R A, Trishch A G. Complete spectral decomposition of the generator for one-dimensional Glauber dynamics (in Russian) [J]. *Uspekhi Matem. Nauk*, 1994, 49: 209–211.

- [7] Schonmann R H. Slow drop-driven relaxation of stochastic Ising models in the vicinity of the phase coexistence region [J]. *Commun. Math. Phys.*, 1994, 161: 1-49.
- [8] Sokal A D, Thomas L E. Absence of mass gap for a class of stochastic contour models [J]. *J. Statis. Phys.*, 1988, 51(5/6): 907-947.
- [9] Wang Fengyu(王风雨). *Functional Inequalities, Markov Processes, and Spectral Theory* [M]. 北京: 科学出版社, 2005.

## Some Interaction of Probability Theory and Other Subjects

CHEN Mu-fa

(*Department of Mathematics, Beijing Normal University, Beijing, 100875, P. R. China*)

**Abstract:** Remarkable progress in probability theory has been made in the past couple decades. Especially, the interaction of probability theory and other subjects has led to several new branches in mathematics. This point of view was set out by the International Congress of Mathematicians in 2002 (ICM2002). As an additional illustration, three results obtained by our group are surveyed in this paper. Finally, the mathematical tool, used in proving the above results, are introduced and related to other branch of mathematics, such as mathematical programming, partial differential equations, and so on.

**Key word:** Probability theory; ICM2002; random algorithm, free probability; percolation theory; eigenvalue; ergodic theory; optimal transportation.