PROBABILITY, RANDOM VARIABLES AND STOCHASTIC PROCESSES

概率随机变量与随机过程

〔美〕A·帕普里斯 著

保 铮 章潜五 吕胜尚 译

西北电讯工程学院出版社

概率 随机变量与随机过程

（1984年版）

〔美〕A·帕普里斯

保 铮 章潜五 吕胜尚 译

西北电讯工程学院出版社

1 9 8 6

PROBABILITY, RANDOM VARIABLES, AND STOCHASTIC PROCESSES

(Second Edition)

Athanasios Papoulis

McGraw-Hill, Inc. (1984)

内容简介

本书是一本概率论和随机过程的重要著作。全书共分三部分：第一部分为概率论；第二部分论述随机过程；第三部分概要地讨论了某些受到广泛应用和普遍关注的排队论、均方估计、谱估计及熵等专题。书中还包括了一套精选的习题，分附于各章之后。

本书选材广泛，论述精辟。原著第一版在美国曾被推荐为经典名著。本书按作者作了大量修订的1984年第二版译出。

本书主要对象是电子电气工程人员，也适用于物理和应用数学工作者。其中前两部分更宜作工科大学高年级学生和研究生的主要教材。

概率 随机变量与随机过程

[美]A・帕普里斯

保 铮 章潜五 吕胜尚 译

责任编辑 陆光华

西北电讯工程学院出版社出版发行

西北电讯工程学院印刷厂印刷

新华书店经销

开本 787×1092 1/16 印张 29 10/16 字数 721 千字

1986年12月第一版 1986年12月第一次印刷 印数 1—6,000

书号：13322·6 定价：4.95元

译 者 的 话

本书编写的指导思想和内容安排已经在作者的两篇序言里交代了，无需重复。这里只是将本书的作者和本书的情况作一些简单介绍。

A·帕普里斯教授是国际知名学者，他多年致力于使传统的电机工程教育现代化。先后发表过涉及多种学科的近百篇论文，并著有五本教科书。他善于用一般工程人员易于接受的方法和语言介绍最新的科技发展和阐明有关理论，既深入浅出，又不失数学上的严密性，这使得他的著作受到多方面的重视和欢迎。他的著作被先后译成日、俄、波兰、意大利、西班牙等多种语言。在我国，其《信号分析》及本书的第一版也相继被译成中文出版。一九七三年，宾夕法尼亚大学授于他“杰出毕业生”的荣称和金质奖章；一九八四年，由于他“在给人以启迪的讲座和研究中，在信号分析、随机过程和系统方面的教科书中所表现出的灵感四溢的创造才能，以及在革新教育方面的领先地位”，而成为电气电子工程师协会(IEEE)一百周年纪念活动教育金质奖牌的获得者。

《概率 随机变量与随机过程》是作者最成功的著作之一。自一九六五年初版问世后，引起了广泛的兴趣。在美国，它被许多院校用作工科研究生或大学高年级的概率论和随机过程的教材，历时二十年不衰，并于七十年代末被美国科学信息协会指定为“引证经典”*之一。考虑到近二十年来在检测、滤波等方面的巨大进展，作者花费了一年多的时间对本书进行了重大修改，并于一九八四年出版了本书的第二版。在这次修订中，原书经过精炼，成为新版前十章的主要内容，并增加了排队论、最小均方估计、谱估计及熵等专题章节，反映了直至一九八四年的一些新发展。对于电子工程、特别是信号处理工作者，它的确是一本既能提供重要的理论基础，又能架起通向最新发展领域的桥梁的好书。

参加本书翻译的有：保铮(序言、第一、二、三、十三、十四章)，章潜五(第九、十、十一、十二章)吕胜尚(第四、五、六、七、八、十五章)。曹天顺、陆光华对全书译文作了校核，赵玮对部分内容也曾提过某些建议。对于原著中一些明显的印刷错误，在译文中作了订正，但未一一注明。

囿于译校者水平，译文不当之处在所难免，敬请读者赐教。

译者 1986.1.

---

*Citation Classic

再版序言

这是一本反映近廿年来发展的、作了重大修改的修订版。这次修订增加了一些新的专题，增强了某些重要部分，并删去了一些用途很少的章节。不过，所增加的大部分材料是属于应用方面的，前十章的内容基本未变。

在新题材的选择上，我试图着重于某些专题，它们不仅当前有用，而且有助于更好地理解随机过程的基本性质。这些新题材包括：

离散时间过程及其在系统理论中的应用

新息，分解和谱表示

排队论，电平穿越，调频信号的谱，取样定理

均方估计，正交展开，莱文森算法，沃尔德分解，维纳、格型和卡尔曼滤波器

谱估计，窗，外推，伯格法，线谱检测

本书末尾还用了自成系统的一章，从基本原理出发公理化地对"熵"进行了讨论。熵的提出与前几章有连贯性，它包含了参数估计和基本编码理论中的最大熵方法。

如同在第一版里一样，我想特别努力来强调思维结构与物理实在之间在概念上的差别。这一差别可以用本书第一版里的下面一段话来概括：

"科学理论研究的是概念，而不是实在。所有理论的结果，都是用演绎逻辑，从某些公理推导得来的。在自然科学中，说明理论时，都要使它在某种有用的意义上，与现实世界中这样或那样事物相对应。然而，这种对应是近似的，而且所有理论结论的物理证实，都是基于某种形式的归纳推理。"

作为对多年来许多读者评论的答复，我想强调指出，上面一段话决不意味着对自然法则(概型)的存在性有任何怀疑，它只是提醒人们注意在概念和实在之间存在的基本差别。

在原稿的准备过程中，曾得益于和许多同事及朋友的长时间的讨论。我特别要提到格鲁曼的汉斯·斯克雷伯，诺尔顿系统的威廉·沙纳汉，以及我的同事 F·卡萨拉和 B·马格拉里斯，并感谢他们的有价值的建议。此外，N·阿达莫太太娴熟地打印了全部原稿，我也希望在此表示对她的谢意。

A·帕普里斯

第一版序言

早在几年前，我就得出这样的结论：概率论不应再作为“统计学”或“噪声”或其它学科的附属部分，而应当作为一门独立的课程，包括在所有工程师和物理学家的基本训练之中。此后，我对如何教授这样一门课程曾做过许多考虑，我想摘录我早期的部分笔记，或许能使读者了解我酝酿撰写此书的一些指导思想。

“以确定的物理学观点教育出来的大多数学生会觉得这一学科不可靠、不明确，并且难学。这些困难的持续存在，是因为那些基本原理的定义本身就不恰当，因而在假设与逻辑结论之间经常形成混淆。只有以公理化办法发展理论，才能消除这种概念上的模糊。有人说，这样做将要用到测度理论，将使所研究的科目归结为数学的一个分支，并将使学生怀疑他的直觉，以致无所适从。但我并不这么想。我相信应用中所需要的大多数概念能够用简单的数学加以阐明，概率论也象其它任何理论一样，应视为概念的结构，其结论不应依赖直觉，而应基于逻辑。当然，各种概念必须与物理世界相联系，但是这种直观性部分应当与理论的演绎部分区别开来。这样，直观会得到加强，而又不致牺牲逻辑的严密性。

“在一些入门课程里所讲的概率论的基本原理与现代应用所要求的复杂的概念之间，显然缺乏衔接。怎样才能使仅仅懂得朴克和骰子的概率知识的普通学生理解预测理论或调和分析？应用类书籍最多只是简要地讨论一下背景材料，其目的不是借助实际应用来增强学生对基本概念的理解，而仅仅是为了详细讨论专题本身。

“随机变量、变换、数学期望、条件密度和特征函数等，仅仅罗列一下是不可能熟练掌握的。这些概念必须明确定义，且必须逐个加以精心讨论。专题应当用来说明理论，这些专题必须集中在概率的内容上，题外的描述应尽可能少。只有这样，学生才能既经济又正确地学会各种应用。”

我意识到，为了把一门课教得令人信服，所讲的不只是些结论，而是一种连贯的理论，不仅在专题的发展上，而且在结论的证明以及基本原理的引出等方法上必须进行反复斟酌。

“这个理论必须有数学的(演绎的)形式，却不要有数学的普遍性和严密性。概率的哲学意义必须设法讨论。这对消除其神秘性，使学生确信需要公理化方法，需要对假设和逻辑结论之间加以明确的区分等都是必要的。公理化基础不应当只是一本书的附录，它应当贯穿在整个理论之中。

“随机变量必须定义为定义域是实验结果的抽象集合的函数，而不是实线上的一些点。只有这样，才能避免无限维空间的困难，并使向随机过程的扩展得到简化。

“将平均定义为所讨论的空间的值，其不恰当性，在随机过程的处理中是最明显不过了。时间平均必须作为随机积分引入，而它们与过程的统计参数的关系只能用各态历经性的形式建立。

“按照当前的需要，利用学生熟悉系统与变换技术这一特点，强调二阶矩和谱是合

“对均方估计(预测和滤波)这个相当重要的课题，有必要重新做基本的研究。

是避免积分方程和变分法的细节，而把它作为正交性原理(线性回归)的一个应用

量来简要阐明。

“为了保持概念的系统性，我们引入的专题只是用来说明一般的理论，而不必过分注重它们的连贯性。”

这些想法构成了我在布鲁克林(Brooklyn)理工学院教授这一课程的轮廓。在学生反应的鼓舞下，我决定把它撰著成册。应当指出，我并不是把我的工作看作一种和对象无关的对某种完整理论的讲述，而看作是将这种理论的实质向一些特定的学生作介绍的一种尝试。本书既不是为那些满足于查用指导手册的学生写的，也不是为那些能从高级数学教材中深究这一课题的少数人写的。本书的对象主要是那些工程师和物理学家中的大多数，他们相当成熟，能够理解并跟得上逻辑的推理，但由于他们的数学基础有限，用象杜布(Doob)那样的著作作为入门教材是过于困难了。

虽然我已经写入了许多有用的结果，其中有些是新的，但我希望对本书进行评价时，不要着眼于完整性，而要着眼于题材的组织和明确性。在这方面我期待着批评并乐意说明我的想法。有些读者会发现，许多重要定理的证明缺乏严密性。我想强调一下，这不是由于疏忽，而是经过深思熟虑后，我决定在若干情况下只给出似乎合理的论证。“证明就应该是证明，否则就不算”，对这一点我了解得非常清楚。不过，一个严格的证明必须在新概念得到澄清和它的正确性得到基本的阐明后，才能进行。为了达到本书的目的，我觉得重点应放在阐明、实用和经济上。我希望这样做不仅能给读者一种有用的知识，而且还能激起大家对这一引人入胜的课题继续深入学习的兴趣。

尽管我试图在每一个课题里都发挥我个人的观点，但我知道我从其他一些作者的著作受益良多。特别是J·L杜布的“随机过程”和A·布兰克-拉皮尔与R·福太特的“随机函数理论”对我关于随机过程部分的构思有很大影响。

最后，我怀着愉快的敬意，向给予我鼓励，并提出有价值的评论的M·施瓦兹，向提出许多见解与建设性建议的R·皮克霍兹表示诚挚的谢意，同时向促使我并参与了这一尝试的所有同事和学生表示谢意。

A·帕普里斯

目　录

第二部分 随机过程

第三部分 选 题

第一部分

概 率 与 随 机 变 量

第一章 概率的意义

1-1 引 言

概率论研究相继发生或同时发生的大量现象的平均，比如：电子发射，电话呼叫，雷达检测，质量控制，系统故障，机遇游戏，统计力学，扰动，噪声，出生率与死亡率，排队论，以及其它等等。

人们观测到，在这样那样一些领域里，当观测次数增加时，某些量的平均会趋于一常数值；而且即使平均是对实验之前预先规定好的任何子序列求的，其值仍保持不变。例如，在投掷硬币的实验里，正面出现的比例趋于 0.5 或其它某个常数，而且如果每投掷(比如说)三次取一次计数的话，仍然会得到相同的平均值(想找到某种赌博妙法来战胜轮盘赌是徒劳无益的)。

概率论的目的就是用事件的概率来描述和预测这种平均。事件$\mathscr{A}$的概率是赋予这一事件的一个数 $P(\mathscr{A})$，它可以解释为：

如果实验重复进行 n 次，事件$\mathscr{A}$发生 $n_{\mathscr{A}}$次，则当 n 足够大时，$\mathscr{A}$发生的相对频率 $n_{\mathscr{A}}/n$ 以高度的把握性接近 $P(\mathscr{A})$：

$$P(\mathscr{A})\simeq n_{\mathscr{A}}/n \tag{1-1}$$

这种解释是不精确的。术语"以高度的把握性"，"接近"和"足够大"的含义都不明确。但是，这种不精确性是不可避免的。如果我们想用概率论的术语来定义"高度的把握性"，能做到的只不过是延缓得到一个无法避免的结论，即：如同任何自然理论一样，概率论也只能以不准确的形式与物理现象相联系。尽管如此，概率论仍是由一些明确定义的公理出发，逻辑地发展起来的一种确切的规则，而且当把它应用于实际问题时，它行得通。

观测，推理，预测 将概率应用于实际问题时，必须明确区分下列步骤：

步骤 1 (物理的) 我们用一个不准确的过程来确定某一事件$\mathscr{A}_i$的概率 $P(\mathscr{A}_i)$。

这一过程可基于(1-1)式的概率与观测之间的关系：概率数据 $P(\mathscr{A}_i)$等于观测的比值 $n_{\mathscr{A}_i}/n$。它也可基于由某种对称性出发的"推理"：如果总结果有 N 个，其中 $N_{\mathscr{A}}$ 个结果有利于事件$\mathscr{A}$，则 $P(\mathscr{A})=N_{\mathscr{A}}/N$。

例如，如果把一颗偏心骰子投掷 1000 次，有 203 次出现五点，那么五点的概率大约等于 0.2。如果骰子是均匀的，由于它的对称性，出现五点的概率应等于 1/6。

步骤 2 (概念的) 假定概率满足某些公理，通过演绎推理我们可从某些事件$\mathscr{A}_i$的概率 $P(\mathscr{A}_i)$确定另一些事件$\mathscr{B}_i$的概率 $P(\mathscr{B}_i)$。

举个例子，在投掷均匀骰子的游戏里，我们可以推想偶数点出现的概率等于 3/6，我们的推理是这样的：

$$\text{如果}\quad P(1)=\cdots=P(6)=\frac{1}{6}\qquad\text{则}\quad P(\text{偶数点})=\frac{3}{6}$$

步骤 3 (物理的) 用如上所得到的数 $P(\mathscr{B}_i)$ 进行实际的预测。

这一步骤也可根据(1-1)式，将它反过来应用：如果重复试验 n 次，而事件$\mathscr{B}$发生 $n_{\mathscr{B}}$

次，则 $n_{\mathscr{B}} \simeq nP(\mathscr{B})$。

例如，投掷匀称的骰子1000次，我们预计偶数点将出现约500次。

在解决问题时，我们不能过分强调将上述三个步骤分开。但必须明确地区分由经验确定的数据和由逻辑推理所得的结果。

步骤1和3是根据归纳推理的。例如，假定我们希望确定一给定硬币正面的概率。我们应当投掷硬币100次还是1000次？如果投掷了1000次，正面出现的平均数为0.48，基于这个观测我们能作出什么样的预测呢？我们能否推想再投掷1000次时正面的数目将约为480？这些问题只能归纳地作出回答。

在本书中，我们只讨论步骤2，即从某些概率演绎地导出另一些概率。有人可能会争议说：这样的推导只不过是一种同义语的无谓重复，因为其结果已经包含在原假设中，这并没有错，因为在完全相同的意义上，我们也可以说人造卫星运动的复杂的方程式早已包含在牛顿定律里了。

作为结束，我们再重复一遍，一事件$\mathscr{A}$的概率$P(\mathscr{A})$可以解释为赋予这一事件的一个数，就如质量是赋予物体的一个数、电阻是赋予电阻器的一个数一样。在理论的发展过程中我们将不关心这个数的"物理意义"。如同在电路分析、电磁理论、古典力学，或其它一些科学学科中曾做过的那样。当然，除非它能帮助我们解决实际问题，否则理论就没有实际价值。即使仅仅是近似的，我们也必须对实际电阻器给以特定的电阻值，对实际事件给以特定的概率（步骤1）；我们也应当对从理论推导出来的所有结论，给以一定的物理意义（步骤3）。但是概念和观测的这种结合必须和各个理论的纯逻辑结构（步骤2）区别开来。

作为说明，我们在下面的例子里讨论电路理论里电阻意义的解释。

例 1-1 电阻器通常看作为二端器件，它的电压正比于电流

$$R=\frac{v(t)}{i(t)} \tag{1-2}$$

但是，这只是一个简单的抽象。实际电阻器是没有明显的端点、且具有分布电感和电容的复杂器件。(1-2)式的关系，只有在一定频率范围内，加上其它多种限制条件，并容许一定的误差才能成立。尽管如此，在电路理论的发展中，人们并不去管所有这些不确定性，而认为电阻R是满足(1-2)式的精确数值，且以(1-2)式和克希荷夫定律为基础发展了电路理论。我们都会同意，如果在理论研究的每一阶段都要关心R的真正意义，那将是不明智的。

1-2 定　义

在这一节里，我们讨论概率的各种定义和它们在我们的研究中所起的作用。

公理化定义

我们将利用集合论的下列概念（详见第二章）：必然事件$\mathscr{S}$是每次试验均发生的事件。两个事件$\mathscr{A}$和$\mathscr{B}$的并$\mathscr{A}+\mathscr{B}$是一新的事件，它在$\mathscr{A}$发生或$\mathscr{B}$发生，或两者都发生时发生。事件$\mathscr{A}$和$\mathscr{B}$的交$\mathscr{A}\mathscr{B}$是另一事件，它在$\mathscr{A}$和$\mathscr{B}$都发生时发生。如果一事件的发生排斥另一事件的发生，则事件$\mathscr{A}$和$\mathscr{B}$是互斥的（互不相容的）。

我们用投掷骰子实验加以说明：六个面中出现任何一面的事件是必然事件。"偶数点"事

件和"小于三点"事件的并是事件"一点"或"二点"或"四点"或"六点"，而两者的交是事件"二点"。"偶数点"事件和"奇数点"事件是互斥的。

用于概率的公理化方法仅仅以下列三条假定为根据：

一事件$\mathscr{A}$的概率$P(\mathscr{A})$是赋予此事件的一个正数

$$P(\mathscr{A})\geqslant 0 \tag{1-3}$$

必然事件的概率等于1

$$P(\mathscr{S})=1 \tag{1-4}$$

如果两个事件$\mathscr{A}$和$\mathscr{B}$是互斥的，则

$$P(\mathscr{A}+\mathscr{B})=P(\mathscr{A})+P(\mathscr{B}) \tag{1-5}$$

这种方法用于概率的历史并不太久[柯尔莫哥洛夫(A.Kolmogoroff)，1933]*。但是，我们觉得，这是引出概率的最好途径，即使在初等课程里也是如此。它强调理论的演绎特性，避免了概念模糊，也为复杂的应用提供了坚实的基础，而且，至少，它为深入研究这一重要学科提供了一个开端。

概率的公理化展开会用到过多的数学。但是，正如我们希望表明的，情况并非一定如此，理论的基本部分完全可以用基础微积分来说明。

相对频率定义

相对频率方法基于下述定义：一事件$\mathscr{A}$的概率$P(\mathscr{A})$是极限

$$P(\mathscr{A})=\lim_{n\to\infty}\frac{n_{\mathscr{A}}}{n} \tag{1-6}$$

式中$n_{\mathscr{A}}$是$\mathscr{A}$的发生次数，n是试验次数。

这个定义看起来是合理的。由于概率用以描述相对频率，用这一频率的极限来定义它是很自然的。这样，与先验定义联系在一起的那些麻烦被消除了，人们可能觉得，这一理论是建筑在观测的基础上的。

但是，尽管相对频率概念是概率应用的基础(步骤1和3)，但对于用它作为演绎论的基础(步骤2)却必须问几个为什么。事实上，在实际实验中，$n_{\mathscr{A}}$和n虽然可以很大，但它们终究是有限的；所以它们的比值，不能与极限等同，甚至也不能近似地等同。如果用(1-6)式定义$P(\mathscr{A})$，这个极限只能作为一种假说来接受，而不是一个可以用实验确定的数。

本世纪早期，V·密塞思(Von Mises)**用(1-6)式作为新理论的基础。在那个年代里，盛行的观点仍然是古典的。他的工作对于概率的先验概念提供了一个极好的对立面，使得后者不得不审视自己抽象推理的真正含意，从而表明它之所以能得出一些有用的结论，主要是由于隐含地使用了基于我们大量经验的相对频率。虽然(1-6)式将$P(\mathscr{A})$和观测的频率联系起来，但用(1-6)式作为演绎论的基础却并非人人乐于接受的。一般认为公理化方法(柯尔莫哥洛夫)更为优越。

我们仍用理想电阻R的定义作例，并冒昧将两种方法作一比较。定义R为一极限

* A. Kolmogoroff: Grundbegriffe der Wahrscheinlichkeits Rechnung, *Ergeb. Math und ihrer Grensg.* Vol. 2, 1933.

** Richard Von Mises: *Probability, statistics and Truth*, English edition, H. Geiringer, ed, G. Allen and Unwin Ltd., London, 1957.

$$R=\lim_{n\to\infty}\frac{e(t)}{i_n(t)}$$

式中 $e(t)$ 是电压源，$i_n(t)$是一系列实际电阻器里的电流，而这些电阻器在某种意义上趋于一理想的二端元件。这个定义可以表示出实际电阻器和理想元件之间的关系，但所得的理论是复杂的。显然，基于克希荷夫定律的 R 的公理化定义要好一些。

古典定义

在好几个世纪的时间里，概率论都是建立在古典定义的基础上的。直到今天，这一概念还用来确定概率性数据，并作为行之有效的假说，下面我们来看看它的意义。

按照古典定义，一个事件$\mathscr{A}$的概率 $P(\mathscr{A})$ 可以不经实际实验而先验确定：它的值由

$$P(\mathscr{A})=\frac{N_{\mathscr{A}}}{N} \tag{1-7}$$

给出，式中 N 是可能的结果的总数，而 $N_{\mathscr{A}}$是有利于事件$\mathscr{A}$的结果数。

在掷骰子实验中，可能的结果为六个，而有利于偶数点这一事件的结果有三个，所以 P(偶数点)$=3/6$。

但是，应该注意到，N 和 $N_{\mathscr{A}}$ 这些数的意义并不总是明确的。下面的问题可以使我们看到这种内在含糊。

例 1-2 投掷两颗骰子，我们需要求出所出现的点数之和等于 7 的概率。

用(1-7)式来解决这一问题时，首先需确定 N 和 $N_{\mathscr{A}}$ 这两个数。

(a) 我们可以认为可能的结果有 11 种，即其和值为 2,3, …, 12, 而和值为 7 的只是其中的 1 个，这是有利的结果，所以 $P=1/11$。这个结果显然是错误的。

(b) 我们可以把所有的点数对作为可能的结果，而对两颗骰子不加区分。这样我们有 21 个可能的结果，其中只有(3, 4)，(5, 2)和(6, 1)是有利的结果。在这种情况下 $N=21$ 和 $N_{\mathscr{A}}=3$，所以 $P=3/21$。这一结果也是错误的。

(c) 我们看出，上述两个解之所以错误，是由于(a)和(b)的各种结果并不是等可能的。要“正确地”解决这个问题，必须在区分第一颗和第二颗骰子的条件下计算所有的点数对。这时结果的总数为 36，而有利的结果为六对，即(3, 4)，(4, 3)，(5, 2)，(2, 5)，(6,1)，和(1, 6)，所以 $P=6/36$。

上述例子表明，有必要改进定义(1-7)式。改进后的形式如下：

如果所有结果是等可能的，一事件的概率等于有利于它的结果数与总结果数的比。

下面很快将会看到，这种改进并不能消除古典定义存在的问题。

注 1. 古典定义是由不充分推理原理的结果引出的*：“当没有先验知识时，我们只有假定事件 $\mathscr{A}_i$ 具有等概率性”。这种结果实际上认为概率只是我们自己关于事件 $\mathscr{A}_i$ 的知识状态的量度。实际上，如果这些事件 $\mathscr{A}_i$ 不真正是等概率的，我们只要改变它们的指标，就能够得到不同的概率，而无须改变我们的知识状态。

2. 在最后一章我们将说明，不充分推理原理等价于最大熵原理。

评论 古典定义在下列几个方面是有问题的：

A. 在改进形式的(1-7)式中所用的术语“等可能”，实际上意味着“等或然”。因此，在

* H. Bernoulli, Arts Conjectandi, 1713,