

TURING

图灵数学·统计学丛书

Probability in Electrical Engineering  
and Computer Science  
An Application-Driven Course

# EECS 应用概率论

[美] Jean Walrand 著

黄隆波 译

以PageRank等热门的计算机和通信算法为实例，  
带你进入概率论的奇妙世界



中国工信出版集团



人民邮电出版社  
POSTS & TELECOM PRESS

TURING

图灵数学·统计学丛书

Probability in Electrical Engineering  
and Computer Science  
An Application-Driven Course

EECS★  
应用概率论

[美] Jean Walrand 著

黄隆波 译

人民邮电出版社  
北京

## 图书在版编目 (C I P) 数据

EECS应用概率论 / (美) 瓦尔朗 (Walrand, J.) 著 ;  
黄隆波译. — 北京 : 人民邮电出版社, 2015.9  
(图灵数学·统计学丛书)  
ISBN 978-7-115-39896-3

I. ①E… II. ①瓦… ②黄… III. ①概率论 IV.  
①0211

中国版本图书馆CIP数据核字(2015)第162047号

## 内 容 提 要

本书精心选取了6个当前热门的科技应用:谷歌 PageRank 算法、链路复用技术、数字链路通信、追踪预测、语音识别和路线规划,并通过讲述概率论在不同应用中的作用来详细介绍基础的概率知识以及概率论中的重要概念,包括马尔可夫链、大数定律、中心极限定理、假设检验、最小方差预测等。

读者对象包括高等院校电子、计算机和统计类专业的高年级本科生及研究生,也包括从事算法、语音识别和机器学习等工作的软件工程师和通信工程师等。

---

◆ 著 [美] Jean Walrand  
译 黄隆波  
责任编辑 朱 巍  
执行编辑 杨 琳  
责任印制 杨林杰

◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号  
邮编 100164 电子邮件 315@ptpress.com.cn  
网址 <http://www.ptpress.com.cn>  
北京隆昌伟业印刷有限公司印刷

◆ 开本: 700×1000 1/16  
印张: 18 彩插: 4  
字数: 352千字 2015年9月第1版  
印数: 1-4 000册 2015年9月北京第1次印刷

著作权合同登记号 图字: 01-2014-4186号

---

定价: 69.00元

读者服务热线: (010)51095186转600 印装质量热线: (010)81055316

反盗版热线: (010)81055315

广告经营许可证: 京崇工商广字第0021号

# 版 权 声 明

Authorized translation from the English language edition, entitled *Probability in Electrical Engineering and Computer Science: An Application-Driven Course* by Jean Walrand, Copyright © 2014 by Jean Walrand.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from the author.

CHINESE language edition published by Posts & Telecom Press, Copyright © 2015.

本书简体中文版由 Jean Walrand 授权人民邮电出版社独家出版。未经出版者许可，不得以任何方式复制或抄袭本书内容。

版权所有，侵权必究。

# 译者序

你现在看到的这本概率书，和市面上绝大部分的概率书籍都非常不同。

尽管概率论在工程与自然科学中已经有了广泛的应用，但现有的绝大部分概率书籍往往沿用数学教材的经典方式，即从定义出发、列举重要结果定理证明，并给出相关例子。虽然这一模式极为有效，但是数学教材中的例子以说明概念为主，鲜少阐述概率论如何在实际的工程问题中发挥作用。因此，许多国内外 EECS (Electrical Engineering and Computer Science, 电子工程与计算机科学系) 的同学在学习概率论的时候，除了觉得好玩有趣之外，往往容易感到困惑：不知道自己为什么需要学习概率论，也不知道概率论究竟能够应用在哪些地方。正因如此，不少同学认为概率论是一门只有做学术研究的人才需要的课程。

本书在内容编撰上另辟蹊径，从应用出发，通过精心选取 EECS 领域核心方向中的几个基本问题，深入浅出却不失系统性地介绍概率论的基础及其在实际问题中的应用。这其中包括：

- (1) PageRank 与马尔可夫链
- (2) 网络多路复用与大数定律和中心极限定理
- (3) 数字链路与时钟和假设检验
- (4) 追踪定位与最小方差估计和卡尔曼滤波器
- (5) 语音识别、大数据与隐马尔可夫链和期望最大化
- (6) 路线规划与马尔可夫决策问题和线性二次型高斯问题

以上列举的应用问题通常很少与书中涵盖的概率论知识一同出现。本书作者这样安排，正是希望通过一种以应用为主的方式，解答读者在概率论学习过程中对其实际效用的疑惑；也希望通过这种讲授方式，使读者对概率论的重要性有更好的认识；并且希望读者在学习工作中遇到问题时，能从概率论的角度出发进行思考。

本书作者 Jean Walrand 教授在加州大学伯克利分校的电子计算机系从教 30 余年，是计算机网络、随机过程与控制研究的大师。他不仅在概率论与工程研究上有极高的造诣，而且深谙概率工具在 EECS 领域中应用的关键。在此书中，Walrand 教授无疑成功地将这两者进行了结合。

在著书之前，Walrand 教授就已在加州大学伯克利分校以基于应用模式讲授概率论的方式积累了多个学期的课堂讲学经验，并根据学生与教师的反馈不断改进此书的内容选取与编排。因此，这无疑是 Walrand 教授精雕细琢的倾力之作。译者从事随机建模与分析多年并曾师从 Walrand 教授，但在翻译的过程中，亦不时被作者精彩新颖的观点与讲解打动，受益匪浅。译者希望，译本也能为读者带来同样的感受。

最后，特别致谢清华大学电子系的连婧与别致两位同学。她们在本书的翻译、排版与审校过程中给予译者鼎力协助。同时也特别感谢图灵文化的朱巍女士和杨琳女士在本书的编辑与审校中给予的大力支持。正因有她们的努力，本书的翻译工作才得以圆满完成。

黄隆波

2015年6月29日

# 前 言

这是一本介绍概率论在电子工程和计算机科学领域应用的书。本书关注一些具有代表性并用到了大量概率论知识的重要实际科技应用。我们无意涵盖所有的应用，一来内容过于宽泛，二来工程量过于浩大。

如同微积分和离散数学一样，掌握概率建模和分析对于计算机科学家和电子工程师来说至关重要。由于这些科学家和工程师应用和设计的复杂系统通常在动态的环境中运作，理解和量化不确定性对系统产生的影响已成为系统设计过程中极为重要的一环。

本书是为加州大学伯克利分校电子工程和计算机科学系的 EECS126 概率论课程而编写的。该课程是为大三、大四学生开设的进阶课程。这门课的学生大多已选修过基础的概率论课程。他们了解事件、概率、条件概率、贝叶斯公式、离散随机变量及期望等概念，并且对矩阵运算有基本的了解（这些概念在附录中都有所回顾）。这些学生都非常聪明、勤奋并对复杂的新知识具有很大兴趣。在这门课程上，学生会学习并了解马尔可夫链、随机动态规划、检验和估计等知识。他们不仅能直观理解这些概念和方法，同时也能熟知如何运用它们。

目前绝大部分概率论的入门书籍在内容介绍上仍然延用了概率空间、随机变量、数学期望、检验、估计和马尔可夫链的顺序。相比之下，本书内容编排的一大特点是以应用作为导引。对于概率论中的每一个知识点，本书均以电子及计算机领域相应的重要实际应用作为媒介，详细讲述该知识点背后的理论及其如何在应用中发挥功用。我们认为，阐述理论在实际应用中功用的讲授方式能让人们更容易了解到概率论的巨大价值。不仅如此，本书同时强调让学生独立应用 Matlab 和 Simulink 数学软件完成对课程项目的仿真和计算。这些题目经过精心设计，旨在加强学生对概念的直观理解，并为其日后进行独立探索打好基础。除了最后一章和附录之外，本书中其余的章节均分为 A 和 B 两部分。A 部分介绍主要的概念，而 B 部分则涵盖知识点中更为深奥的部分。因此，本书的一种使用方式是给大三的学生讲授 A 部分和附录中的内容；另一种授课方式是开设一门两学期的课程，在第一阶段课程中讲授 A 部分，将 B 部分纳入第二阶段的学习中。如果想要提升课程难度，也可以在一门课程中先教授 A 部分，再教授 B 部分。按章节顺序授课也不失为一种不错的选择。本书的最后一章收录

了一些前沿课题，读者和教师可以有选择地使用。

附录部分为大多数读者提供了有用的背景知识。附录 A 回顾了概率论的基本概念。根据学生背景知识的不同，教师可以选择从附录 A 的综述讲起。附录 B 复习了一些线性代数的基本知识，而附录 C 提供了一些 Matlab 的实例。采用 Matlab 是因为现在许多大学都为学生提供了购买 Matlab 的许可。即使学校没有获得授权，学生也能负担起该软件的学生版。学生通过在 Matlab 中尝试一些算法并模拟一些系统，能更好地理解相关知识点和认知系统的具体运作方式。不仅如此，让学生轻松实现一些看起来复杂的算法会让他们在获得更多的成就感和乐趣。

书中的理论从基于未知量的模型入手。我们将这样的未知量记为  $X$  和  $Y$ ，并考虑一个能用于计算关于函数  $h(X)$  期望值  $E(h(X))$  的模型。举例来说， $X$  可表示一块太阳能电池板在一个月中每天产生的能量， $h(X)$  为这块电池板一个月产生的总能量。那么， $E(h(X))$  就是电池板每月平均产生的能量。其他例子还包括通信网络中的平均数据包延时，数据中心完成一件工作的平均耗时，等等。

估计  $E(h(X))$  的过程称为性能评估。在许多情况下，在需要处理未知量的系统中都存在一些参数  $\theta$ 。人们可以通过调节这些参数的值来调整系统作业。比如说，太阳能电池板的朝向可以进行调整，数据中心的运作也能调整。在这种情况下，我们可以通过将系统性能描述成未知量  $X$  和可变参数  $\theta$  的函数  $h(X, \theta)$  来衡量参数取值对系统的影响。如此一来，优化系统这一问题就等同于寻找能最大化  $E(h(X, \theta))$  的参数值  $\theta$ 。这个问题通常并不简单，尤其是在  $E(h(X, \theta))$  没有解析表达式时。本书会讲解这样的优化问题。

$$\begin{array}{ccc} X & \xrightarrow{?} & E(h(X)) & \quad & \max_{\theta} E(h(X, \theta)) \\ & & \text{评估} & & \text{优化} \end{array}$$

在许多问题中，我们能观测到  $Y$ ，但是希望估计未被观测的  $X$  值。举个例子， $X$  可以是发射机发出的信号，而  $Y$  是接收机收到的信号。根据  $Y$  估计  $X$  值的问题是一个推断问题。这方面的例子包括探测问题（附近有没有火源，你有没有感冒）和估计问题（发出 GPS 信号的 iPhone 在哪里）。

$$\begin{array}{ccc} & & ? \\ Y & \rightarrow & X \\ & & \text{推断} \end{array}$$

最后，有一类问题关注基于观测的动态系统控制。比如说，一辆利用激光测距仪、GPS 和相机进行观测的无人驾驶汽车。我们称之为控制问题。

$$\begin{array}{ccc} X & \rightarrow & Y \\ & \uparrow & | \\ & \text{控制} & \end{array}$$



本课程讨论性能评估、优化、推断和控制问题。这其中的一些问题在计算机科学中被称为人工智能，而在电子工程中则被称作统计信号处理。概率学家称之为样本，而数学家将其称为特例。我们通过讲述具体的实际应用，如网络搜索、多路传输、数字通信、语音识别、图像跟踪、路径规划和推荐系统等来介绍这些技术。在学习和了解这些概念的过程中，我们也会为大家介绍这个领域的相关巨擘。

本书的内容十分有趣、能引人思考。希望你们能分享我对这些想法的热情。

十分感谢在本书撰写过程中给予我大力支持的各位同事和学生们。我要特别感谢仔细阅读了手稿的黄隆波博士和 Ramtin Pedarsani，感谢提供了宝贵意见的 Abhay Parekh 博士、David Aldous 教授、Venkat Anantharam 教授、Tom Courtade 教授、Michael Lustig 教授、John Musacchio 教授、Kannan Ramchandran 教授、Anant Sahai 教授、David Tse 教授、Martin Wainwright 教授和 Avidoh Zakhor 教授。感谢担任课程助教并设计了课程作业的 Stephan Adams、Vijay Kamble、Shiang Jiang 博士、Sudeep Kamath 博士、Jerome Thai、Baosen Zhang 博士和 Antonis Dimakis 教授。感谢教授我概率论的 Pravin Varaiya 教授和 Eugene Wong 教授，以及支持我的 Tsu-Jae King Liu 教授。最后，感谢提供反馈的 EECS126 课程的学生们。

网站 <https://sites.google.com/site/walrandpeecs/home> 为本书提供了额外的资源，如勘误、附加题、评论和幻灯片等。教师也可以使用该网站来寻求问题的解决方案。

# 目 录

第 1 章 PageRank—A	1
1.1 模型	1
1.2 马尔可夫链	3
1.2.1 定义	3
1.2.2 $n$ 步后的分布和稳态分布	4
1.3 分析	5
1.3.1 不可约性和非周期性	5
1.3.2 大数定律	5
1.3.3 长期时间比例	6
1.4 击中时间	7
1.4.1 平均击中时间	7
1.4.2 击中另一状态之前命中某一状态的概率	8
1.4.3 马尔可夫链的首步方程	9
1.5 小结	10
1.6 参考资料	10
1.7 练习	11
第 2 章 PageRank—B	15
2.1 样本空间	15
2.2 投掷硬币的大数定律	17
2.2.1 依概率收敛	17
2.2.2 几乎处处收敛	18
2.3 独立同分布随机变量的大数定律	20
2.3.1 弱大数定律	20
2.3.2 强大数定律	21
2.4 马尔可夫链的大数定律	22
2.5 期望的收敛	23
2.6 大定理的证明	25
2.6.1 定理 1.2(a)的证明	25
2.6.2 定理 1.2(b)的证明	26

2.6.3 周期性	27
2.7 小结	29
2.8 参考资料	29
2.9 练习	30
<b>第 3 章 多路复用—A</b>	31
3.1 链路共享	32
3.2 高斯随机变量与中心极限定理	34
3.3 多路复用与高斯分布	37
3.4 置信区间	37
3.5 缓冲器	39
3.6 多址访问	43
3.7 小结	44
3.8 参考资料	45
3.9 练习	45
<b>第 4 章 多路复用—B</b>	47
4.1 特征方程	47
4.2 中心极限定理的证明 (概要)	48
4.3 $\mathcal{N}(0,1)$ 的高阶矩	49
4.4 两个独立同分布于 $\mathcal{N}(0,1)$ 的随机变量平方和	50
4.5 特征函数的两个应用	51
4.5.1 泊松分布作为二项分布的近似	51
4.5.2 指数分布作为几何分布的近似	51
4.6 误差函数	52
4.7 自适应多址访问	53
4.8 小结	55
4.9 参考资料	55
4.10 练习	55
<b>第 5 章 数字链路—A</b>	57
5.1 检测与贝叶斯准则	58
5.1.1 贝叶斯准则	58
5.1.2 最大后验概率 (MAP) 与最大似然估计 (MLE)	59
5.1.3 二元对称信道	60
5.2 霍夫曼编码	62
5.3 高斯信道	64
5.4 多维高斯信道	66
5.5 假设检验	67

5.5.1	规范化问题	68
5.5.2	解答	68
5.5.3	示例	69
5.6	小结	75
5.7	参考资料	76
5.8	练习	76
<b>第 6 章</b>	<b>数字链路—B</b>	<b>79</b>
6.1	霍夫曼编码最优性的证明	79
6.2	低密度奇偶校验码 (LDPC 码)	80
6.3	联合高斯分布随机变量	85
6.4	联合高斯分布随机变量的密度函数	86
6.5	奈曼-皮尔逊定理 5.6 的证明	88
6.6	小结	89
6.7	参考资料	90
6.8	练习	90
<b>第 7 章</b>	<b>追踪定位—A</b>	<b>91</b>
7.1	估计问题	92
7.2	线性最小平方估计 (LLSE)	93
7.3	线性回归	97
7.4	最小均方估计 (MMSE)	98
7.5	随机向量的情况	104
7.6	卡尔曼滤波器	106
7.6.1	滤波器	106
7.6.2	示例	107
7.7	小结	110
7.8	参考资料	110
7.9	练习	111
<b>第 8 章</b>	<b>追踪定位—B</b>	<b>115</b>
8.1	LLSE 的更新	115
8.2	卡尔曼滤波器的推导	116
8.3	卡尔曼滤波器的特性	118
8.3.1	可观测性	119
8.3.2	可达性	120
8.4	扩展卡尔曼滤波器	121
8.5	小结	124
8.6	参考资料	124

<b>第 9 章</b>	<b>语音识别—A</b> .....	125
9.1	学习：概念和示例.....	125
9.2	隐马尔可夫链.....	126
9.3	期望最大化和聚类.....	129
9.3.1	一个简单的聚类问题.....	129
9.3.2	回首再探.....	130
9.4	学习：隐马尔可夫链.....	132
9.4.1	硬期望最大化.....	132
9.4.2	训练维特比算法.....	132
9.5	小结.....	132
9.6	参考资料.....	133
9.7	练习.....	133
<b>第 10 章</b>	<b>语音识别—B</b> .....	135
10.1	在线线性回归.....	135
10.2	随机梯度投影理论.....	136
10.2.1	梯度投影.....	137
10.2.2	随机梯度投影算法.....	140
10.2.3	鞅收敛定理.....	142
10.3	大数据.....	143
10.3.1	相关数据.....	143
10.3.2	压缩感知.....	147
10.3.3	推荐系统.....	150
10.4	小结.....	151
10.5	参考资料.....	151
10.6	练习.....	151
<b>第 11 章</b>	<b>路线规划—A</b> .....	153
11.1	系统建模.....	153
11.2	方法 1：提前规划.....	154
11.3	方法 2：适应性算法.....	155
11.4	马尔可夫决策问题.....	156
11.5	无限时域问题.....	161
11.6	小结.....	162
11.7	参考资料.....	162
11.8	练习.....	163
<b>第 12 章</b>	<b>路线规划—B</b> .....	166
12.1	线性二次型高斯问题.....	166

12.2	有噪声观测时的线性二次型高斯问题	169
12.3	部分可观测的马尔可夫决策问题	171
12.4	小结	173
12.5	参考资料	174
12.6	练习	174
<b>第 13 章</b>	<b>视野拓展和补充</b>	<b>176</b>
13.1	推断问题	176
13.2	充分统计量	177
13.3	无限马尔可夫链	179
13.4	泊松过程	181
13.4.1	定义	181
13.4.2	独立自增量	182
13.4.3	跳跃次数	183
13.5	连续时间马尔可夫链	184
13.6	二元对称信道的容量	186
13.7	概率界	190
13.8	鞅	194
13.8.1	定义	194
13.8.2	示例	195
13.8.3	大数定律	199
13.8.4	沃尔德等式	200
13.9	小结	201
13.10	参考资料	201
13.11	练习	202
附录 A	概率论基础知识	206
附录 B	线性代数基本知识	240
附录 C	Matlab	253
	参考文献	273

# 第1章

## PageRank—A

应用：在网页搜索中，按网页相关度由高到低进行排序

主题：有限离散时间马尔可夫链，强大数定律

背景知识：附录 A.1 ~ A.2

搜索引擎采用不同的算法将网页按给定的关键字以相关度递减的方式排序。其中一种算法的思想是计算马尔可夫链的稳态分布。本章讨论这种分布的存在性与唯一性，以及在随机浏览时找到一个特定网页所需的平均时间。我们将采用强大数定律证明马尔可夫链处于每个特定状态的时间比例是收敛的。

### 1.1 模型

互联网由一系列相互链接的网页组成。这些网页和它们之间的链接关系构成了一张图。如图 1-1 所示，图中的节点是所有的网页  $\mathcal{X}$ 。若网页  $i$  有一个到网页  $j$  的链接，则图中有一条由  $i$  到  $j$  的弧（有向边）。

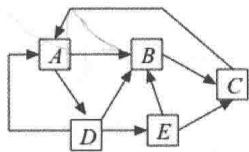


图 1-1 在网络中，网页总是指向其他的网页。在本图中  $P(A, B) = 1/2$ ， $P(D, E) = 1/3$

直观上看，一个高级别网页所指向的网页也应具有较高的级别（在实际中，除了我们在此讨论的网页级别度量方式，搜索引擎的排序结果还取决于网页中关键字是否出现以及其他许多因素）。因此，我们定义网页  $i$  的级别  $\pi(i)$  为一个正数，并且满足

$$\pi(i) = \sum_{j \in \mathcal{X}} \pi(j) P(j, i), i \in \mathcal{X}.$$

$P(j, i)$ 表示所有网页  $j$  的外向链接中指向  $i$  的链接所占的比例. 如果  $j$  没有指向  $i$  的链接, 则  $P(j, i) = 0$ . 在上述例子中, 我们有  $P(A, B) = 1/2$ ,  $P(D, E) = 1/3$ ,  $P(B, A) = 0$  等. 这一算法的思想源自拉里·佩奇 (Larry Page, 见图 1-2), 这也是 PageRank (佩奇排序) 这一名字的由来.



图 1-2 谷歌公司创始人之一拉里·佩奇

可以将这些等式以矩阵的形式记作

$$\pi = \pi P. \quad (1.1)$$

这里,  $\pi$  是一个以  $\pi(i)$  为分量的行向量, 而  $P$  是一个以  $P(i, j)$  为元素的方阵.

式 (1.1) 称为平衡方程. 这里我们注意到, 如果一个向量  $\pi$  是这个方程的解, 那么  $\pi$  的任意倍数也是这个方程的解. 为方便起见, 我们将解归一化, 使得所有网页的级别之和为 1:

$$\sum_{i \in \mathcal{X}} \pi(i) = 1. \quad (1.2)$$

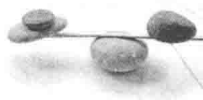


图 1-3 这是平衡方程吗

对于图 1-1 中的例子, 平衡方程为

$$\begin{aligned} \pi(A) &= \pi(C) + \pi(D)(1/3) \\ \pi(B) &= \pi(A)(1/2) + \pi(D)(1/3) + \pi(E)(1/2) \\ \pi(C) &= \pi(B) + \pi(E)(1/2) \\ \pi(D) &= \pi(A)(1/2) \\ \pi(E) &= \pi(D)(1/3). \end{aligned}$$

再加上  $\pi(i)$  的和为 1 这一条件, 可以得到



$$\pi = [\pi(A), \pi(B), \pi(C), \pi(D), \pi(E)] = \frac{1}{39} [12, 9, 10, 6, 2].$$

由此可以看到，网页  $A$  具有最高的级别，网页  $E$  具有最低的级别。运用这一方法的搜索引擎会将这些网页级别与其他因素相结合来进行网页排序。一些搜索引擎也会采用这一度量的变种来对网页进行排序。

## 1.2 马尔可夫链

想象一下你正在浏览网页：假设你在网页  $i$  上浏览了一个单位时间，然后随机点击进入了网页  $i$  指向的一个网页。在这个过程中，从网页  $i$  到网页  $j$  的概率正好为  $P(i, j)$ ，与前面的例子相同。

### 1.2.1 定义

考虑一个包含节点  $\mathcal{N} = \{1, 2, \dots, N\}$  和有向边的有限图。假设其中有些节点具有指向自己的边。图中每条边  $(i, j)$ ，都有一个权重  $P(i, j) > 0$ 。这些权值使得每个点外向边的权和为 1。根据习惯，如果图中没有从  $i$  到  $j$  的边，则  $P(i, j)$  为 0。

以上述方式得到的矩阵  $P = P[(i, j)]$  叫作随机矩阵。这种矩阵的每个元素均为非负，并且每行的和为 1。现在，我们定义以下过程  $\{X(n), n \geq 0\}$ ：在时刻 0 的时候用  $X(0)$  表示系统所处的状态；随后的每一个时刻  $n$ ，系统由状态  $X(n-1) = i$  跳到状态  $X(n) = j$  的概率为  $P(i, j)$ ，即系统所处的状态仅由  $X(n-1)$  与  $P$  决定。这样定义的过程  $\{X(n), n \geq 0\}$  被称为马尔可夫链。该有限图则被称为马尔可夫链的状态转移图。



图 1-4 安德雷·马尔可夫 (1856—1922)

图 1-5 展示了 3 个马尔可夫链的状态转移图。