清華 汇智文库
QINGHUA HUIZHI WENKU

隐马尔可夫链、马尔可夫状态转换模型及在量化投资中的应用

HIDDEN MARKOV CHAIN, MARKOV SWITCHING MODEL, AND THE APPLICATIONS IN QUANTITATIVE FINANCE

王 犇◎著

清華大学出版社

清華汇智文库
QINGHUA HUIZHI WENKU

隐马尔可夫链、马尔可夫状态转换模型及在量化投资中的应用

HIDDEN MARKOV CHAIN, MARKOV SWITCHING MODEL, AND THE APPLICATIONS IN QUANTITATIVE FINANCE

王　犇◎著

清華大學出版社
北　京

内 容 简 介

本书属于数理金融(量化投资)的范畴，论述了隐马尔可夫链和马尔可夫状态转换模型的数学原理、数值算法及在量化投资中的应用。出于完备性的考虑，本书的前两章回顾了理解模型所必需的概率统计、随机过程的相关知识。本书虽涉及一定程度的数学知识，但总体而言仍定位于业界，也就是金融领域尤其是量化投资领域的实务工作者，同时对该领域的学术研究者也有一定的参考价值。

图书在版编目（CIP）数据

隐马尔可夫链、马尔可夫状态转换模型及在量化投资中的应用/王犇著. —北京：清华大学出版社, 2017

（清华汇智文库）

ISBN 978-7-302-45324-6

Ⅰ. ①隐…　Ⅱ. ①王…　Ⅲ. ①投资-研究　Ⅳ. ①F830.59

中国版本图书馆 CIP 数据核字(2016)第 260837 号

责任编辑：张　伟
封面设计：汉风唐韵
责任校对：王凤芝
责任印制：杨　艳

出版发行：清华大学出版社
　　网　址：http://www.tup.com.cn，http://www.wqbook.com
　　地　址：北京清华大学学研大厦 A 座　　**邮　编**：100084
　　社总机：010-62770175　　**邮　购**：010-62786544
　　投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
　　质 量 反 馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：三河市吉祥印务有限公司
经　　销：全国新华书店
开　　本：170mm×230mm　　**印　张**：12.75　　**字　数**：175 千字
版　　次：2017 年 1 月第 1 版　　**印　次**：2017 年 1 月第1 次印刷
定　　价：69.00 元

产品编号：071891-01

前言

Foreword

金融学及金融投资的本质是在不确定的环境下进行稀缺资源的跨期优化配置。近几十年来，随着金融不断与数学、统计、计算机等学科交叉融合，金融学已呈现出定量化、精确化和模型化的大趋势。与此同时，伴随着数学科学、统计分析、计算机技术的不断进步，一股量化投资的热潮也扑面而来。可以预见，基于数理模型和计算机程序的量化投资，将与基本面分析、技术分析“三分天下”“鼎足而立”，从根本上塑造金融投资的新格局。

本书属于数理金融（量化投资）的范畴，论述了隐马尔可夫链和马尔可夫状态转换模型的数学原理、数值算法以及在投资实务中的应用。全书分为5个部分。为了满足自封性 (self-contained) 的要求，第 1 章简单回顾了模型必需的概率统计知识。第 2 章紧随其后，介绍了一类非常重要的随机过程——马尔可夫链，重点在 C–K 方程、马尔可夫链的若干重要性质，以及马氏链的极大似然估计。前两章是后续部分的基础。第 3 章介绍了状态独立混合分布模型，这可以看作第 4 章的一个“预演”，因为隐马尔可夫链从本质上说，是一个状态不独立（相关）的混合分布模型。第 4 章和第 5 章是全书的核心，也是最具有实战价值的部分。这两章分别论述了隐马尔可夫链和马尔可夫状态转换模型，两个模型最重要的实务价值在于从数据中找出判断金融市场上升或者下降的信息。这两章均遵循从数学原理到计算机算法再到金融市场实证分析的研究框架，各个算法（伪代码）在正文相应部分给出，对应的 Matlab 和 R 程序统一包含在附录里。

本书虽涉及一定程度的数学知识，但总体而言仍定位于业界，也就是金融领域的实务工作者。在许多情况下，不过多纠缠过于复杂的数学细节，而是直击数学模型背后朴素的“物理”意义，尽量用工科的直观代

替理科的严谨。

本书的顺利出版，要感谢清华大学出版社的大力支持和张伟编辑的辛勤帮助。作者才疏学浅，书中定有疏漏之处，还望各位师长、广大同仁不断批评指正。作者邮箱：benben_england@163.com。

作　者

2016 年 9 月

目录

Contents

第 1 章

概率统计必要知识回顾

概率论是金融数学的基石。一方面，概率论本身是研究不确定性强有力的数学工具；另一方面，概率论是其他几个学科如数理统计、随机过程、随机分析的基础。概率论和数理统计本身是一个宏大的话题，但本着“学以致用”的原则，在本章，我们只聚焦于理解隐马尔可夫链和马尔可夫状态转换模型所必须具备的概率统计相关知识。

1.1 条件概率与条件期望的两个引理

条件概率与条件期望在概率论中具有非常高的地位，“取条件”这个思想贯穿全书始终。一方面，在很多情况下，我们对给定部分信息的概率或者期望感兴趣；另一方面，通过适当地“取条件”，我们可以使问题简化。本节的核心是两个引理，在论述它们之前，先回顾一下“划分”“全概率公式”和“贝叶斯公式”等基本概念。

1.1.1 划分、全概率公式与贝叶斯公式

定义 1.1 [划分 (Partition)]

$A_1, A_2, \cdots, A_n$ 是样本空间 Ω 的一个划分，如果满足

$$\begin{cases} A_i \bigcap A_j = \varnothing \quad (i \neq j) \\ \bigcup_{i=1}^{n} A_i = \Omega \end{cases}$$

直观地理解,“划分”指的就是样本空间的一组子集，这些子集彼此之间没有交集，拼在一起构成全集。

定理 1.1 [全概率公式 (Total Probability Formula)]

若 $A_1, A_2, \cdots, A_n$ 是样本空间 Ω 的一个划分，B 为任意一个事件，则

$$P(B) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)$$

证明：

因为 $A_1, A_2, \cdots, A_n$ 是 Ω 的一个划分，故由定义 1.1 可得

(1) $$(B \bigcap A_i) \bigcap (B \bigcap A_j) = \varnothing \quad (i \neq j)$$

即 $\{B \bigcap A_i\}$ 两两互斥

(2) $$\bigcup_{i=1}^{n} (B \bigcap A_i) = B$$

因此

$$\begin{aligned} P(B) &= P\left(\bigcup_{i=1}^{n} (B \bigcap A_i)\right) \\ &= \sum_{i=1}^{n} P(B \bigcap A_i) \\ &= \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i) \end{aligned}$$

证毕

定理 1.2 [贝叶斯公式 (Bayes Formula)]

若 $A_1, A_2, \cdots, A_n$ 是样本空间 Ω 的一个划分，B 是一个事件，则

$$P(A_k|B) = \frac{P(B|A_k) \cdot P(A_k)}{\sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)} \quad (k = 1, 2, \cdots, n)$$

证明：

$$\begin{aligned} P(A_k|B) &= \frac{P(A_k \bigcap B)}{P(B)} \\ &= \frac{P(B|A_k) \cdot P(A_k)}{\sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)} \end{aligned}$$

(分母基于全概率公式，定理 1.1)

证毕

特别的，$\{A, \bar{A}\}$ 这两个集合是样本空间 Ω 的一个划分，故由定理 1.2 可得

$$\begin{aligned} P(A|B) &= \frac{P(A \bigcap B)}{P(B)} \\ &= \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})} \end{aligned}$$

在连续情形下，概率质量函数由概率密度函数替代，求和由积分替代，可得连续情形下的贝叶斯公式为

$$\begin{aligned} f(\theta|x) &= \frac{f(x,\theta)}{f(x)} \\ &= \frac{f(x|\theta)f(\theta)}{\int_\theta f(x,\theta)\mathrm{d}\theta} \\ &= \frac{f(x|\theta)f(\theta)}{\int_\theta f(x|\theta)f(\theta)\mathrm{d}\theta} \end{aligned}$$

1.1.2 两个引理

引理 1

$$P(AB|C) = P(A|C) \cdot P(B|AC)$$

证明：

$$\begin{aligned}
\text{LHS} &= \frac{P(ABC)}{P(C)} \\
\text{RHS} &= \frac{P(AC)}{P(C)} \cdot \frac{P(ABC)}{P(AC)} = \frac{P(ABC)}{P(C)}
\end{aligned}$$

故很明显，LHS = RHS。这里 LHS 是 Left-Hand Side（左边）的简写，RHS 是 Right-Hand Side（右边）的简写。

证毕

从证明上看，引理 1 虽然仅仅是条件概率公式的简单应用，但在本书中，我们将会非常频繁地使用这个公式，同时，它也是许多复杂概率模型的基础。因此，我们把它单独作为一个引理给出。

引理 2

$$E[E(X|Y)] = E(X)$$

证明：

这里我们仅给出离散情况下的证明，连续情形与此类似。

$$\begin{aligned}
E(X|Y) &= E(X|Y=y) \\
&= \sum_x x \cdot P(X=x|Y=y)
\end{aligned}$$

它是 y 的一个函数，因此，

$$\begin{aligned}
E[E(X|Y)] &= \sum_y E(X|Y) \cdot P(Y=y) \\
&= \sum_y \sum_x x \cdot P(X=x|Y=y) \cdot P(Y=y) \\
&= \sum_y \sum_x x \cdot P(X=x, Y=y)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_x x \cdot \overbrace{\sum_y P(X=x, Y=y)}^{X\text{ 的边际密度}} \\
&= \sum_x x \cdot P(X=x) \\
&= E(X)
\end{aligned}
$$

证毕

引理 2 也被称为重期望公式或全期望公式，它是条件期望的一个重要性质。正如期望是概率的推广一样，全期望公式（引理 2）也可以被视为全概率公式（定理 1.1）的推广。直观的理解，引理 2 告诉我们，计算 $E(X)$ 等同于计算 X 的条件期望的加权平均，条件期望取在事件 $\{Y=y\}$ 给定时，而权重是事件 $\{Y=y\}$ 发生的概率。

1.2 极大似然估计

1.2.1 极大似然估计的基本思想

极大似然估计 (maximum likelihood estimation, MLE) 是求统计模型参数值的强有力工具，其核心思想是求使得样本数据出现的联合概率最大的参数值是多少。极大似然估计法本身在后续章节中有很多应用，同时，它的基本思想也是后面要介绍的期望最大化算法 (EM Method) 的基础。

对于一个离散（连续）的随机变量 X，其概率质量（密度）函数为 $p(x|\theta)(f(x|\theta))$；此外，有大小为 n 的样本，这些样本点 $x_1, x_2, \cdots, x_n$ 是独立同分布的 (i.i.d., independent & identically distributed)。这样，样本数据的联合概率密度为

$$p(x_1, x_2, \cdots, x_n|\theta) = \prod_{i=1}^{n} p(x_i|\theta)$$

这里 $x_1, x_2, \cdots, x_n$ 都是已知的，所以上式是 θ 的一个函数，称为似然函数 (likelihood function)

$$\begin{aligned} L(\theta|x_1, x_2, \cdots, x_n) &= p(x_1, x_2, \cdots, x_n|\theta) \\ &= \prod_{i=1}^{n} p(x_i|\theta) \end{aligned}$$

极大似然估计的目标是找出 $\hat{\theta}$ ，从而最大化 L（或 $\log L$），用数学语言表示就是

$$\hat{\theta} = \arg\max_{\theta} L(\theta|x_1, x_2, \cdots, x_n)$$

$\hat{\theta}$ 被成为 θ 的极大似然估计量 (maximum likelihood estimator)

1.2.2 三个实例

本小节我们将给出用极大似然估计法估计模型参数的三个实例。

例 1-1 指数分布

$X \sim \mathrm{Exp}(\lambda)$，其密度函数为

$$f(x|\lambda) = \begin{cases} \lambda \mathrm{e}^{-\lambda x} & (x > 0) \\ 0 & (\text{其他情况}) \end{cases}$$

解:

似然函数为

$$\begin{aligned} & L(\lambda|x_1, x_2, \cdots, x_n) \\ &= f(x_1, x_2, \cdots, x_n|\lambda) \\ &= \prod_{i=1}^{n} f(x_i|\lambda) \\ &= \prod_{i=1}^{n} \lambda \mathrm{e}^{-\lambda x_i} \\ &= \lambda^n \mathrm{e}^{-\lambda \sum\limits_{i=1}^{n} x_i} \end{aligned}$$

对数似然函数为

$$\log L = n\log\lambda - \lambda\sum_{i=1}^{n} x_i$$

因此

$$\begin{aligned}
&\frac{\partial \log L}{\partial \lambda} = 0\\
\Rightarrow &\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0\\
\Rightarrow &\hat{\lambda} = \frac{n}{\sum\limits_{i=1}^{n} x_i} = \frac{1}{\sum\limits_{i=1}^{n} x_i/n} = \frac{1}{\bar{x}}
\end{aligned}$$

例 1-2　正态分布

$X \sim N(\mu, \sigma^2)$，其密度函数为

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x-\mu)^2/2\sigma^2}$$

解：

似然函数为

$$\begin{aligned}
&L(\mu,\sigma^2|x_1,x_2,\cdots,x_n)\\
=&f(x_1,x_2,\cdots,x_n|\mu,\sigma^2)\\
=&\prod_{i=1}^{n} f(x_i|\mu,\sigma^2)\\
=&\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x_i-\mu)^2/2\sigma^2}\\
=&(2\pi)^{-\frac{n}{2}}\cdot(\sigma^2)^{-\frac{n}{2}}\cdot\mathrm{e}^{-\sum\limits_{i=1}^{n}(x_i-\mu)^2/2\sigma^2}
\end{aligned}$$

(将 σ^2 作为一个整体对待)

对数似然函数为

$$\log L = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \sum_{i=1}^{n}(x_i-\mu)^2/2\sigma^2$$

分别对 μ 和 σ^2 求偏导并令偏导数为 0，可得

- $$\frac{\partial \log L}{\partial \mu} = 0$$
$$\Rightarrow -\frac{1}{2\sigma^2}\sum_{i=1}^{n}2(x_i-\mu)(-1) = 0$$
$$\Rightarrow \sum_{i=1}^{n}(x_i-\mu) = 0$$
$$\Rightarrow n\mu = \sum_{i=1}^{n}x_i$$
$$\Rightarrow \hat{\mu} = \sum_{i=1}^{n}x_i/n = \bar{x}$$

- $$\frac{\partial \log L}{\partial \sigma^2} = 0$$
$$\Rightarrow -\frac{n}{2}\frac{1}{\sigma^2} - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\cdot(-1)\cdot\frac{1}{(\sigma^2)^2} = 0$$
$$\Rightarrow \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2} = n$$
$$\Rightarrow \hat{\sigma^2} = \frac{\sum_{i=1}^{n}(x_i-\hat{\mu})^2}{n} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$$

注： 这里用极大似然估计法求得的方差估计量不是无偏估计量，σ^2 的无偏估计量是

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$$

例 1-3　线性回归模型

经典的（一元）线性回归模型可表示为

$$y_t = \beta_1 + \beta_2 x_t + u_t$$

其中 u_t 是 i.i.d.，并且 $u_t \sim N(0, \sigma^2)$

因此，y_t 也是 i.i.d.，并且 $y_t \sim N(\beta_1 + \beta_2 x_t, \sigma^2)$

现在的问题是，给定观测值 $\{x_1, x_2, \cdots, x_T\}$ 和 $\{y_1, y_2, \cdots, y_T\}$，用极大似然估计法求出 $\beta_1, \beta_2, \sigma^2$。

解：

y_t 满足正态分布，其密度函数为

$$f(y_t|x_t, \beta_1, \beta_2, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(y_t-\beta_1-\beta_2 x_t)^2/2\sigma^2}$$

故 $y_1, y_2, \cdots, y_T$ 的联合密度函数为

$$\begin{aligned} & f(y_1, y_2, \cdots, y_T|x_1, x_2, \cdots, x_T, \beta_1, \beta_2, \sigma^2) \\ & = \prod_{t=1}^{T} f(y_t|x_t, \beta_1, \beta_2, \sigma^2) \\ & = (2\pi)^{-\frac{T}{2}} \cdot (\sigma^2)^{-\frac{T}{2}} \cdot \mathrm{e}^{-\sum\limits_{t=1}^{T}(y_t-\beta_1-\beta_2 x_t)^2/2\sigma^2} \end{aligned}$$

（仍将 σ^2 作为一个整体对待）

因为 $x_1, x_2, \cdots, x_T$ 与 $y_1, y_2, \cdots, y_T$ 都已经给定，所以上式仅为 $\beta_1, \beta_2, \sigma^2$ 的函数，也即为似然函数

$$\begin{aligned} & L(\beta_1, \beta_2, \sigma^2|x_1, x_2, \cdots, x_T; y_1, y_2, \cdots, y_T) \\ & = (2\pi)^{-\frac{T}{2}} \cdot (\sigma^2)^{-\frac{T}{2}} \cdot \mathrm{e}^{-\sum\limits_{t=1}^{T}(y_t-\beta_1-\beta_2 x_t)^2/2\sigma^2} \end{aligned}$$

对数似然函数为

$$\log L = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \sum_{t=1}^{T}(y_t - \beta_1 - \beta_2 x_t)^2/2\sigma^2$$