贝叶斯动态模型及其预测

山东科学技术出版社

贝叶斯动态模型及其预测

张孝令　刘福升
张承进　葛颜祥　编著

山东科学技术出版社

(鲁)新登字05号

贝叶斯动态模型及其预测

张孝令　刘福升
张承进　葛颜祥　编著

*

山东科学技术出版社出版
(济南市玉函路　邮政编码250002)
山东省新华书店发行
山东新华印刷厂潍坊厂印刷

*

787×1092毫米32开本　13.875印张　290千字
1992年8月第1版　1992年8月第1次印刷
印数：1—1,400

ISBN 7—5331—1077—3/O·47
定价 8.50元

序

过程的贝叶斯预测是英国统计学家P. J. Harrison教授和C. F. Stevens教授在英国帝国化学工业公司工作的时候（1971年），由于要预测突发事件的需要而提倡发展起来的一种有名的预测方法。1976年，他们在英国皇家统计学会上宣读了论文“贝叶斯预测”，引起了人们的重视，此后在英、美等国，这个方法的理论研究和应用迅速地开展起来了。1989年由M. west和J. Harrison合著出版一本《贝叶斯预测和动态模型》（Bayesian Forecasting and Dynamic Models, New york, Springer-verlag），全面地论述了此方法的内容，张孝令老师曾在英国沃力克大学向Harrison教授学习了这个方法，也曾做过这方面的研究工作，并多次讲授这个方法。今天，以他为首写成这本书，把这个方法引进和介绍给我国读者，填补了我国这方面空白，我认为这是很有意义的一件事。我虽然不是这方面的专家，还是大胆接受本书作者的邀请，为此书作序，愿向读者推荐这本书。

本书大量篇幅考虑的是如下一个线性动态模型

观测方程：$\boldsymbol{y}_t = \boldsymbol{F}_t' \boldsymbol{\theta}_t + \nu_t$

状态方程：$\theta_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \omega_t$

这里y_t表示t时刻测量数据，$\boldsymbol{\theta}_t$表示一个未知n维的参数，描述过程在t时刻的状态，是未知而人们想要知道的。$\boldsymbol{F}_t$是已知$n\times 1$矩阵，它描述状态与测量数据间的关系。ν_t是t时刻

的测量误差，或者称为噪声，它服从某个分布。这本书的前七章都是假设v_t服从正态分布$N[0, V_t]$，只在第八章才考虑别的分布，如指数族分布，这一方面因为测量误差往往服从正态分布；另一方面理论推导方便，比较成熟。状态方程事实上是描述t时刻与$t-1$时刻状态的变化关系。G_t是一个$n \times n$已知矩阵，ω_t表示状态带有一定的随机性，服从某个分布。本书第八章也考虑了非线性动态模型，着眼点在于如何化为线性动态模型来处理或近似处理。

本书的特点就是用贝叶斯的观点和方法利用上述模型进行预测，也就是说它不仅仅是依赖于t时刻以往的历史测量的数据，根据模型的知识进行预测，而且包括专家的经验信息，主观的判断来进行预测，这对于预测突发事件特别有用，而历史数据以及预先规定的模型并不能完全反映它们。当发现模型性能不好时，可求助于专家的经验和信息，对模型进行改进。用数学化的语言来说，对模型中包含的未知参数V_t、W_t及初始信息，我们都假设有先验分布。根据已测得的历史数据，以及先验分布的信息，计算后验分布，根据它进行预测。这一步的后验分布又是下一步先验分布的基础，构成一个连环套，易于计算和理解。本书贝叶斯预测另一个重要特点是，重视专家的经验和信息。如年底将建立一个新工厂，届时电力需求将增加，则应事先对电力需求模型进行干预（参考第六章），这样就能提高预测精度。总之，贝叶斯方法是客观数据、模型加进人的主观经验因素来分析处理问题。

这种贝叶斯预测方法，相对 Box—Jenkins传统的时间序列方法而言，有它的优点。它不必假设Box—Jenkins方

法所必须的平稳性假设，如果非平稳，还必须利用差分法转化平稳序列进行研究。而实际问题往往是非平稳时间序列，此时Box—Jenkins 方法要求有足够的数据，但这个要求往往是困难的。而贝叶斯预测方法通过人的主观经验，给出先验分布，使得数据的要求大大减少，而能得到同样精度的预测。本书还论述了一些流行的经典预测模型，可以看作某种贝叶斯模型的特殊情况（如第四章某些部分）。

使用贝叶斯方法有很大争论。国际统计界分为经典学派和贝叶斯学派，争论已上百年了，至今没有结论。但大家公认，这场学术争论给统计学的发展带来了好处，促进了统计的发展。我们不必有畏怕心理，我们学习它、运用它，参加进去讨论，求得在讨论和运用中发展，用实践来检验和发展，既然存在，总有它的道理。经典学派对贝叶斯学派的批评，一是非知参数看作随机变量是否妥当，二是先验分布是否存在。如何选取？这也是两派分歧之点。作为应用角度来看，我看也可以考虑接受把非知参数看作随机变量。例如某产品的废品率 p，如果不全部测量（做不到），p 是不知道的。但凭生产者的经验判断，有多大可能性，p 处在哪个范围之内，这也是人们常常这样做的。这种主观估计即意味着 p 看作随机变量。至于先验分布的存在，就是人们经验的反映，也是可以理解的。这本书对未知参数大都是假设正态分布，因为模型误差服从正态，不是正态时采用共轭分布。最后归结为一个参数选择问题，这多少带有主观成分以及数学上计算简单和漂亮出发的。是否能利用主观经验呢？我想引用我国著名科学家钱学森一段话供读者参考。他说：“处理复杂行为系统的定量学方法学，是科学理论、经验和专家判断力的结

合，这是定量方法学，是半经验半理论的。提出经验性假设（猜想或判断），是建立复杂行为系统数学模型的出发点。这些经验性假设（猜想或判断）不能用严谨的科学方式证明，但需用经验性数据对其确实性进行检测。从经验性假设（猜想或判断）出发，通过定量方法途径获得的结论，仍然具有半经验、半理论的属性。当人们寻求用定量方法处理复杂行为系统时，容易注重于数学模型的逻辑处理，而忽视数学模型微妙的经验含义或解释。要知道，这样的数学模型看来‘理论性’很强，其实不免牵强附会，从而脱离真实。与其如此，反不如从建模一开始就老实承认理论的不足，而求援于经验判断，让定性的方法与定量方法结合起来，最后定量。”（摘自《中国现代科学家传记》中“钱学森”一章）

为了读者深入了解本书的内容，作者引用了一些例子说明概念和用法，这是很好的。由于刚刚在国内介绍这个方法，尚无国内应用实例，希望今后修改时能增加实例，就会更完美了。

成　平

于中国科学院系统科学研究所

1991年12月

前　言

贝叶斯预测系国际上一个最新研究成果。自1976年以来，英、美等国相继开展了对它的研究。近十几年来，贝叶斯预测的理论和应用得到了迅速的发展，并取得了一批重要的研究成果。但在国内至今还没有开展这方面的工作，张孝令于1988年从师于英国统计学家P. J. Harrison教授，并向他学习了这种理论。回国后，作者和他的研究生一起做过这方面的研究工作。编写本书的目的是向理工科和财经类院校高年级学生和研究生以及预测工作者介绍动态模型与贝叶斯预测的一般理论和方法，书中包括了近十几年来国际上有关这方面的新内容和新方法，并结合我们近年来课题研究的经验和体会，进行较为系统全面的介绍。为了帮助读者深入理解贝叶斯预测理论，本书大部分结论都尽可能地给出证明，对于一些较为复杂的理论虽未加证明，但也列出了有关的参考书目，有兴趣的读者可以去查阅。

由于动态模型和贝叶斯预测的研究在国内尚属空白，所以本书出版必将推动国内有更多的读者开展这方面的研究，并为社会主义建设做出贡献。

本书得到了煤炭系统留学回国人员科技基金委员会、山东矿业学院教材建设基金委员会、山东矿业学院应用数学与软件工程系主任吴哲辉教授山东枣庄矿务局王家生副局长以及枣庄市薛城区煤炭工业局胡大国局长的大力支持，还得到

了中国科学院系统科学研究所所长成平研究员的热情指导，提出了宝贵的意见。在此，我们表示衷心的感谢。

由于时间仓促，书中错误和缺点在所难免，敬请广大读者批评指正。

作　者

1992 年 3 月

目　录

第一章　预备知识 ………………………………… (1)

第一节　分布理论 ………………………………… (1)

第二节　正态分布 ………………………………… (10)

第三节　几个重要分布 ……………………………… (19)

第四节　HPD 区域 ………………………………… (25)

第五节　矩阵代数初步 ……………………………… (27)

第二章　动态模型和贝叶斯预测引论 ……………… (38)

第一节　贝叶斯预测基本思想 ……………………… (38)

第二节　常均值模型 ………………………………… (42)

第三章　动态线性模型理论 ………………………… (73)

第一节　动态线性模型 ……………………………… (73)

第二节　单变量时间序列DLM的理论 …………… (96)

第三节　模型设计 ………………………………… (115)

第四章　模型的分类 ……………………………… (136)

第一节　多项式趋势模型 ………………………… (136)

第二节　季节模型 ………………………………… (159)

第三节　回归模型 ………………………………… (190)

第四节　噪声模型 ………………………………… (220)

第五章　标准DLM的扩展与实例 ………………… (230)

第一节　趋势/季节模型的基本分析………………… (231)

第二节　趋势/季节/回归DLM的基本分析……… (241)

第三节　误差分析 …………………………………… (256)
第四节　数据处理 …………………………………… (259)
第六章　模型的干预和监控 ………………………… (265)
第一节　主观干预 …………………………………… (265)
第二节　模型监控 …………………………………… (279)
第七章　多过程模型 ………………………………… (297)
第一节　第一类多过程模型 ………………………… (299)
第二节　第二类多过程模型 ………………………… (317)
第八章　非线性和广义线性动态模型 ……………… (354)
第一节　非线性动态模型 …………………………… (354)
第二节　广义线性动态模型 ………………………… (374)
第九章　多变量动态线性模型 ……………………… (393)
第一节　多变量DLM ………………………………… (393)
第二节　聚集预测 …………………………………… (396)
第三节　矩阵正态DLM ……………………………… (404)

第一章 预备知识

第一节 分布理论

1. 随机向量和分布

定义1.1 由n个随机变量X_1，X_2，…，X_n构成的n维列向量

$$X=(X_1,\ X_2,\ \cdots,\ X_n)'=\begin{pmatrix}X_1\\X_2\\\vdots\\X_n\end{pmatrix}$$

称为n维随机向量，它的概率分布

$$\begin{aligned}F(\boldsymbol{X})&=F(x_1,x_2,\cdots,x_n)\\&=P\{X_1<x_1,\ X_2<x_2,\ \cdots,\ X_n<x_n\}\\&\quad -\infty<x_i<\infty,\ i=1,\ 2,\ \cdots,\ n\end{aligned}$$

称为n元分布函数，也叫做随机变量X_1，X_2，…，X_n的联合分布函数。

分布函数具有如下性质：

(1) $0\leqslant F(\boldsymbol{X})\leqslant 1$；

(2) $F(\boldsymbol{X})$对于每个变元$x_i(i=1,2,\cdots,n)$单调非降并且左连续；

(3) 在变元x_1，x_2，…，x_n中至少有一个趋于$-\infty$时，$F(\boldsymbol{X})\to 0$；

(4) 变元x_1，x_2，…，x_n都趋于$+\infty$时，$F(\mathbf{X})\to 1$，即

$$F(+\infty,+\infty,\cdots,+\infty)=\lim_{\substack{x_i\to\infty\\ i=1,2,\cdots,n}}F(x_1,x_2,\cdots,x_n)=1$$

定义1.2 若$\mathbf{X}$为n维随机向量，由它的$m(m<n)$个分量组成的m维随机向量$\mathbf{X}_1$的分布叫做$\mathbf{X}$的边缘分布。

通过交换$\mathbf{X}$中各个分量的次序，总可以假定$\mathbf{X}_1$正好是$\mathbf{X}$中前m个分量组成的随机向量，再用记号$\mathbf{X}_2$表示$\mathbf{X}$中后$n-m$个分量组成的随机向量，这样可得

$$\mathbf{X}=\begin{pmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{pmatrix}$$

若$\mathbf{X}$的分布函数为$F(\cdot)$，则$\mathbf{X}_1$的分布函数为

$$\begin{aligned}P\{\mathbf{X}_1<u\}&=P\{X_1<u_1,\ \cdots,\ X_m<u_m\}\\ &=P\{X_1<u_1,\ \cdots,X_m<u_m,X_{m+1}<\infty,\cdots,X_n\\ &\quad<\infty\}\\ &=F(u_1,u_2,\cdots,u_m,\infty,\cdots,\infty)\\ &\quad-\infty<u_i<\infty,\ i=1,\ 2,\ \cdots,m\end{aligned}$$

定义1.3 若$\mathbf{X}=\begin{pmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{pmatrix}$，$\mathbf{X}$的分布函数为$F(\mathbf{X})$，$\mathbf{X}_1$和$\mathbf{X}_2$的分布函数分别为$F_1(\mathbf{X}_1)$和$F_2(\mathbf{X}_2)$，当且仅当

$$F(\mathbf{X})=F_1(\mathbf{X}_1)\cdot F_2(\mathbf{X}_2)$$

时，称$\mathbf{X}_1$和$\mathbf{X}_2$相互独立。

定义1.4 设n维随机向量$\mathbf{X}$的分布函数为$F(\mathbf{X})=F(x_1,x_2,\cdots,x_n)$，若存在一个非负的函数$p(\cdot)$，使得

$$F(\mathbf{X})=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_n}p(t_1,t_2,\cdots,t_n)\ dt_1dt_2\cdots dt_n$$

对于一切$\boldsymbol{X}\in R^n$都成立，则称$\boldsymbol{X}$（或$F(\boldsymbol{X})$）有分布密度$p(\cdot)$，并称$\boldsymbol{X}$为连续型随机向量。

一个n个变量的函数$p(\cdot)$能作为$\boldsymbol{R}^n$中某个随机向量的分布密度，当且仅当

(1) $p(\boldsymbol{X}) \geqslant 0$，$\forall \boldsymbol{X}\in \boldsymbol{R}^n$

(2) $\int_{R^n} p(\boldsymbol{X})d\boldsymbol{X} = 1$

若$p(\cdot)$为$\boldsymbol{X}$的分布密度，$\boldsymbol{B}$为$\boldsymbol{R}^n$中任意一个波雷尔可测集，则有

$$P\{\boldsymbol{X}\in \boldsymbol{B}\} = \int_B p(\boldsymbol{X})d\boldsymbol{X}$$

若$\boldsymbol{X}_0$为函数$p(\boldsymbol{X})$的连续点，则

$$p(\boldsymbol{X}_0) = \left.\frac{\partial^n F(x_1, x_2, \cdots, x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n}\right|_{\boldsymbol{X}=\boldsymbol{X}_0}$$

式中$F(\cdot)$为与$p(\cdot)$对应的分布函数。

若$\boldsymbol{X} = \begin{pmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{pmatrix}$，$p(\cdot)$、$p_1(\cdot)$、$p_2(\cdot)$分别为$\boldsymbol{X}$、$\boldsymbol{X}_1$、$\boldsymbol{X}_2$的分布密度，$\boldsymbol{X}_1$和$\boldsymbol{X}_2$的维数分别为$m$和$n-m$，则

$$p_1(\boldsymbol{X}_1) = \int_{R^{n-m}} p(\boldsymbol{X})d\boldsymbol{X}$$

$$p_2(\boldsymbol{X}_2) = \int_{R^m} p(\boldsymbol{X})d\boldsymbol{X}$$

若$\boldsymbol{X}_1$和$\boldsymbol{X}_2$相互独立，则$p(\boldsymbol{X}) = p_1(\boldsymbol{X}_1)\cdot p_2(\boldsymbol{X}_2)$。

定义1.5 若$\boldsymbol{X} = \begin{pmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{pmatrix}$和$\boldsymbol{X}_2$分别有分布密度$p(\boldsymbol{X})$和$g(\boldsymbol{X}_2)$，则称

$$f_1(\mathbf{X}_1|\mathbf{X}_2)=\frac{p(\mathbf{X})}{g(\mathbf{X}_2)}$$

为给定$\mathbf{X}_2$时，$\mathbf{X}_1$的条件分布密度。

类似地，若$\mathbf{X}_1$的分布密度为$h(\mathbf{X}_1)$，则称

$$f_2(\mathbf{X}_2|\mathbf{X}_1)=\frac{p(\mathbf{X})}{h(\mathbf{X}_1)}$$

为给定$\mathbf{X}_1$时，$\mathbf{X}_2$的条件分布密度。

若$\mathbf{X}_1$和$\mathbf{X}_2$独立，则

$$f_1(\mathbf{X}_1|\mathbf{X}_2)=h(\mathbf{X}_1),\quad f_2(\mathbf{X}_2|\mathbf{X}_1)=g(\mathbf{X}_2)$$

由定义1.5，显然有

(1) $p(\mathbf{X})=f_1(\mathbf{X}_1|\mathbf{X}_2)\cdot g(\mathbf{X}_2)=f_2(\mathbf{X}_2|\mathbf{X}_1)\cdot h(\mathbf{X}_1)$

(2) $f_1(\mathbf{X}_1|\mathbf{X}_2)=f_2(\mathbf{X}_2|\mathbf{X}_1)\cdot h(\mathbf{X}_1)/g(\mathbf{X}_2)$

这是多元形式的贝叶斯定理。

2. 贝叶斯定理

若A和B表示两个事件，则

$$P(A|B)=\frac{P(B|A)\cdot P(A)}{P(B|A)\cdot P(A)+P(B|\bar{A})\cdot P(\bar{A})}$$

其中$\bar{A}$为A的逆事件。这是贝叶斯定理的最简单的表达式。更一般的表述是：若A_1，A_2，…，A_m为两两互不相容的m个事件，B是给定的另一个事件，且$B\subset\sum_{i=1}^{m}A_j$，则

$$P(A_j|B)=\frac{P(B|A_j)\cdot P(A_j)}{\sum_{i=1}^{m}P(B|A_i)\cdot P(A_i)},\quad j=1,\ 2,\ \cdots,\ m$$

称$P(A_j)$ $(j=1,\ 2,\ \cdots,m)$为先验概率，它是在试验（抽样）以前，人们对所研究的问题的看法，是主观信息的描述，

称$P(B|A_j)(j=1, 2, \cdots, m)$为似然函数，$P(A_j|B)(j=1, 2, \cdots, m)$为后验概率，它是综合了先验信息和试验（抽样）结果后得到的看法。

若θ是一个连续的随机变量，实际上它是我们所关心的参数，总把它看作是一个随机的量（θ也可以是一个连续的随机向量），设它的分布密度为$f_1(\theta)$，称为先验密度，涉及θ的样本信息记为y，已知θ时y的分布密度为$f_2(y|\theta)$，称为似然函数，则有

$$f(\theta|y)=\frac{f_1(\theta)\cdot f_2(y|\theta)}{\int f_1(\theta)\cdot f_2(y|\theta)d\theta}$$

这里$f(\theta|y)$叫做后验密度，这是参数θ为连续情况下的贝叶斯定理。值得注意的是，在上式中，y既可以是连续的，也可以是离散的。

若H_1和H_2分别是具有先验概率密度$f_1(H_1)$和$f_2(H_2)$的两个模型，满足$f_1(H_1)+f_2(H_2)=1$，且$Y|H_i$的密度函数为$p_i(y|H_i)$，则

$$g_i(H_i|Y=y)=\frac{f_i(H_i)\cdot p_i(y|H_i)}{\sum_{j=1}^{2} f_j(H_j)\cdot p_j(y|H_j)}$$

$$\propto f_i(H_i)\cdot p_i(y|H_i)$$

这里$f(y)=\sum_{j=1}^{2} f_j(H_j)\cdot p_j(y|H_j)$是$Y$的分布，$p_i(y|H_i)$是似然函数，$g_i(H_i|Y=y)$为后验分布。上式即为一般情况下的贝叶斯定理。用文字表述即为

后验概率$\propto$似然函数乘以先验概率

其中符号“$\propto$”表示成比例。

3.数字特征

定义1.6 依次称

$$E(\boldsymbol{X})=\begin{pmatrix}EX_1\\EX_2\\\vdots\\EX_n\end{pmatrix},\quad E(\boldsymbol{Y})=\begin{pmatrix}EY_1\\EY_2\\\vdots\\EY_m\end{pmatrix}$$

为$\boldsymbol{X}=(X_1, X_2, \cdots, X_n)'$和$\boldsymbol{Y}=(Y_1, Y_2,\cdots, Y_m)'$的数学期望（或均值向量）；称

$$E(\boldsymbol{X})=(EX_{ij})_{p\times n}$$

为随机矩阵$\boldsymbol{X}=(X_{ij})_{p\times n}$的均值矩阵；称

$$\begin{aligned}\operatorname{cov}(\boldsymbol{X},\boldsymbol{X})&=E[\boldsymbol{X}-E(\boldsymbol{X})][\boldsymbol{X}-E(\boldsymbol{X})]'\\&=\begin{pmatrix}\operatorname{cov}(X_1,X_1)\cdots\operatorname{cov}(X_1,X_n)\\\operatorname{cov}(X_2,X_2)\cdots\operatorname{cov}(X_2,X_n)\\\cdots\qquad\cdots\qquad\cdots\\\operatorname{cov}(X_n,X_1)\cdots\operatorname{cov}(X_n,X_n)\end{pmatrix}\end{aligned}$$

为n维随机向量$\boldsymbol{X}=(X_1,X_2,\cdots,X_n)'$的自协方差阵，简称为协方差阵或方差阵；称

$$|\operatorname{cov}(\boldsymbol{X}, \boldsymbol{X})|$$

为$\boldsymbol{X}$的广义方差，它是协方差阵的行列式；称

$$\begin{aligned}\operatorname{cov}(\boldsymbol{X}, \boldsymbol{Y})&=E[\boldsymbol{X}-E(\boldsymbol{X})][\boldsymbol{Y}-E(\boldsymbol{Y})]'\\&=\begin{pmatrix}\operatorname{cov}(X_1, Y_1)\cdots\operatorname{cov}(X_1, Y_m)\\\operatorname{cov}(X_2, Y_1)\cdots\operatorname{cov}(X_2, Y_m)\\\cdots\qquad\cdots\qquad\cdots\\\operatorname{cov}(X_n, Y_1)\cdots\operatorname{cov}(X_n, Y_m)\end{pmatrix}\end{aligned}$$

为n维随机向量$\boldsymbol{X}=(X_1, X_2, \cdots, X_n)'$和$m$维随机向量$\boldsymbol{Y}=(Y_1,Y_2,\cdots,Y_m)'$的互协方差阵，简称为$\boldsymbol{X}$与$\boldsymbol{Y}$的协方差阵。