观测过程理论

Observation Process Theory

陈必红 著

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

Observation Process Theory

观测过程理论

陈必红　著

電子工業出版社
Publishing House of Electronics Industry
北京 · BEIJING

内 容 简 介

观测过程理论是作者潜心研究多年提出的有关信息论与统计学的全新的、相对完整的理论体系，是在强干扰和大随机性的情况下，有效提取有用信息的理论。它将现代信息论的基本思想用在各种统计学问题上，认为所有的统计、检测、信号处理、滤波、模式识别等信息处理过程均为观测过程，并以贝叶斯公式递推为基础，提出了主观概率和客观概率的概念，观测主体和观测客体的概念，知识函数或知识分布的概念。为了对这些概念进行数学上的支持，扩充了非标准分析的概念，给出了标准无穷大数和标准无穷小数的定义，使得利用计算机来计算非标准数的四则运算成为可能。全书共分 9 章。

本书可供信息论、统计学、信号处理、语音识别、模式识别、图像分析、系统辩识、检测与估计、时间序列分析等专业领域的研究人员和工程技术人员参考，也可供高等院校相关专业的教师、研究生和高年级学生阅读。

图书在版编目（CIP）数据

观测过程理论/陈必红著．—北京：电子工业出版社，2007.4
ISBN 978-7-121-03662-0

Ⅰ．观…　Ⅱ．陈…　Ⅲ．统计学　Ⅳ．C8

中国版本图书馆 CIP 数据核字（2006）第 155697 号

责任编辑：周　琰
印　　刷：北京智力达印刷有限公司
装　　订：北京中新伟业印刷有限公司
出版发行：电子工业出版社
　　　　　北京市海淀区万寿路 173 信箱　邮编　100036
开　　本：787×980　1/16　印张：9.5　字数：180 千字
印　　次：2007 年 4 月第 1 次印刷
印　　数：3 000 册　　定价：29.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系电话：（010）68279077；邮购电话：（010）88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

前　　言

观测过程理论是作者潜心研究多年提出的有关信息论与统计学的全新的、相对完整的理论体系，是在强干扰和大随机性的情况下，有效提取有用信息的理论。香农（Shannan）于 1948 年创立的信息论应用到通信编码领域，使通信编码技术得到了大发展，但信息论至今仍没在强干扰和大随机性的统计学领域得到应用。本书提出的观测过程理论，就是将现代信息论的基本思想用在各种统计学问题上，认为所有的统计、检测、信号处理、滤波、模式识别等信息处理过程均为观测过程，并以贝叶斯公式递推为基础，提出了主观概率和客观概率的概念，观测主体和观测客体的概念，知识函数或知识分布的概念。为了对这些概念进行数学上的支持，扩充了非标准分析的概念，给出了标准无穷大数和标准无穷小数的定义，使得利用计算机来计算非标准数的四则运算成为可能。

本书共分 9 章。

第 1 章给出了一个观测过程的基本模型，其中包括观测主体、观测客体及观测客体到观测主体的信息通道。观测主体可以是自然人、动物、机器人、雷达、计算机等，但本书中的观测主体实质就是具备计算功能和存储功能的计算机；观测客体是一个随机变量或者随机向量，是观测主体感兴趣、希望知道关于它的一切信息的未知量。观测客体通过信息通道向观测主体发送信息，使观测主体对于观测客体的知识有所变化。

观测过程理论用概率分布（本书中叫知识分布）来描述观测主体关于观测客体的知识多少。这种知识分布在每一次观测前被称为先验分布，在每一次观测后被称为后验分布。而观测主体获得的后验分布，又可以作为进一步观测的先验概率分布。观测主体关于观测客体的知识分布，随着不断接收到观测量而不断地变化，这个变化可通过贝叶斯公式来计算。

第 2 章提出了观测主体的先验分布从何而来的问题，并根据观测主体获得的观测值越多，它关于观测客体的知识分布越集中，或者说知识熵越小的实际情况，反过来顺着时间往回推，得出观测主体对于观测客体的先验分布是广义均匀分布，即取线性空间中任何一点的可能性是一样的。

为了给出广义均匀分布和其他广义分布的概念，本书采用非标准分析体系，并定义了标准无穷大数和标准无穷小数，还重新定义了单位脉冲函数，然后用概率密度函数来描述一切类型的随机变量的分布，包括离散型、连续型及混合型随机变量的分布。

前面两章还给出了观测过程理论的一个重要观点，就是观测主体不需要也不应当进

行所谓检测、估计或模式识别，观测主体只需要如实地记载在接收到观测值之后，自己关于观测客体的知识分布就行了，这个知识分布以概率密度函数的形式存放在观测主体这部计算机中。这是因为在观测完之后，观测主体要么根本不需要做任何决策，在这种情况下，保存知识分布就是最完整地保存观测值中含有的信息；要么必须做决策，最优的决策应当根据知识分布来进行，而不应当根据检测、估计和模式识别的结果来进行。

第 3 章到第 9 章利用前两章的结论去分析各种各样的观测模型。

第 3 章分析了一些特殊的观测器，这些观测器用到了本书的新数学工具，并得出一些以往统计学没有讨论过的结论。

第 4 章用观测过程理论的观点重新论述了有关正态总体的统计问题，虽然这些问题都已经成为国际通用的标准方法，且已经写入了大学的初级概率论与数理统计教程，但这一章仍然用观测过程理论的观点进行了重述。

第 5 章研究一般信号的二元检测问题。当然，观测过程理论并不关心检测，而是研究观测主体关于观测客体的知识分布的变化。这一章最值得注意的是研究了未知噪声与信号幅度的情况，这是以往教科书中没有很好解决、但实际的无线电信号中常见的问题。

第 6 章是将第 5 章的结论简单地从二元推广到多元。

第 7 章用观测过程理论的观点分析了雷达信号，其中的许多观点是第 5 章的延伸，给出了以往所有雷达信号处理教科书所缺少的、在未知信号强度与噪声强度情况下的正确处理方法，本章给出的公式具有相当的价值。

第 8 章用观测过程理论的观点重述了卡尔曼滤波，但不是探讨估计问题，而是简单地描述了知识分布的变化。

第 9 章研究观测过程理论中的卷积相乘算法在图像识别、模式识别及语音识别等方面的妙用，给出了一些有可能在这些领域产生革命性变化的算法。

本书提出的观测过程理论有许多观点与现在广泛流行的统计学观点不尽相同。作为一种尝试，不足之处在所难免。因此，我不敢奢望读者能接受我的所有观点，只希望其中重要章节提出的一些算法在实际应用中会起到好的作用和效果。

作　者

2007 年元月

目　　录

第1章 基本概念

1.1 观测过程的基本模型

1.1.1 简单的观测模型

一个简单的观测模型如图 1-1 所示。

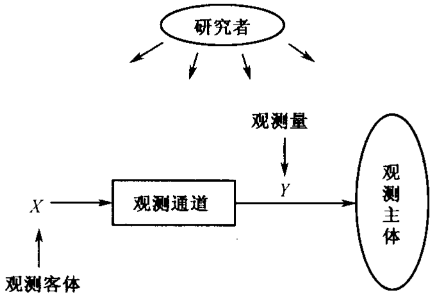

图 1-1 观测模型示意图

观测客体是一个随机变量或者随机向量，当其为随机变量时，用 X 表示；而如果是随机向量，则用黑体 $\mathbf{X}$ 表示，我们先研究观测客体是一随机变量 X 的情况。

观测主体是对观测客体 X 进行认识的主体，可以是一个人，或者是一个机器人，或者是一部雷达，等等。就实质而言，观测主体是一部具有计算能力和存储能力的计算机。这部计算机希望通过观测来获得对观测客体的认识，因此我们研究的也可以说是一个人工智能问题。我们暂时假设这部计算机具有无限的计算能力和存储能力。

观测主体通常不能直接观测到观测客体，X 通常是通过一个**观测通道**或者叫信息通道变换成**观测量** Y 后送往观测主体的。

实际上，在实际搭建的观测系统中，观测主体甚至不能直接观测到观测量 Y，而只是接收到 Y 的一个**样本值** y。观测主体就是根据这个样本值来获得关于观测客体 X 的知识。

图 1-1 还标出了一个**研究者**，这是为以后帮助我们思考问题。研究者就是对整个观

测系统进行研究设计和试验的人，不妨假设研究者是无所不知的人。因此，对于研究者而言，是没有什么随机变量这样的事情的，所有的量在研究者看来都是确定的量，因为他都能够清楚地知道。本文中所谓的“未知”，是指观测主体对观测客体的未知，但研究者总是什么都知道的。

观测客体 X 可以是任何类型的随机变量，可以是连续型的（如一个病人的体温，一只鸡的质量），也可以是离散型的（如病人得了乙肝或者没有），或是混合型的（如一个地区每天的降雨量）。而观测通道则可能是任何的测量仪器，或者是信息通道。这在传统的统计学中，分别称为估计问题、检测问题或者模式识别问题。实际上，观测量 Y 也可以是任何类型的随机变量，连续的、离散的或者混合型的。在下面的分析中，我们暂时将 X 和 Y 看做是连续型的随机变量。

1.1.2 贝叶斯公式

观测过程理论就是要在观测主体获得观测值 $Y=y$ 的条件下，计算出 X 的条件概率密度函数 $f_{X|Y}(x|y)$。这个条件概率密度也称为**后验概率密度**，它正确地反映了观测值中带有的有关观测客体的全部知识或者说全部信息。后验概率密度所对应的分布也称为**后验分布**。

在传统的统计学中，求出后验概率密度后，还要进行检测、估计和模式识别，但观测过程理论只要计算出后验分布就可以了，无须做检测、估计或模式识别。这一点与传统的统计学有所不同。

假设观测主体 X 的概率密度为 $f_X(x)$，观测量 Y 的概率密度为 $f_Y(y)$，它们的联合概率密度为 $f_{X,Y}(x,y)$。因此，$f_X(x)$和 $f_Y(y)$都相当于 $f_{X,Y}(x,y)$的边缘概率密度，满足关系

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\mathrm{d}y \tag{1.1}$$

和

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\mathrm{d}x \tag{1.2}$$

则在 $Y=y$ 条件下的条件概率密度 $f_{X|Y}(x|y)$（即我们关心的后验概率密度）定义为

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \tag{1.3}$$

而在 $X=x$ 条件下的条件概率密度 $f_{Y|X}(y|x)$定义为

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \tag{1.4}$$

上述概率密度反映了整个观测通道的性能，通常可以通过反复试验和统计而获得，即在试验时由研究者固定不同的观测客体的样本值，然后反复进行观测通道的传输试验，以统计观测量 Y 的统计特性，或者通过分析来获得。因此，称条件概率密度 $f_{Y|X}(y|x)$ 为**观测函数**。

由式(1.4)可得

$$f_{X,Y}(x,y) = f_X(x)f_{Y|X}(y|x) \tag{1.5}$$

将式(1.5)代入式(1.2)得

$$f_Y(y) = \int_{-\infty}^{+\infty} f_X(x)f_{Y|X}(y|x)\mathrm{d}x \tag{1.6}$$

将式(1.5)和式(1.6)代入式(1.3)可得

$$f_{X|Y}(x|y) = \frac{f_X(x)f_{Y|X}(y|x)}{\int_{-\infty}^{+\infty} f_X(x)f_{Y|X}(y|x)\mathrm{d}x} \tag{1.7}$$

但分母中的积分变量也用 x，就有可能把人搞糊涂，更严格的写法是分母上的积分变量换用另一个字母，比如说 u，这样将式(1.7)写为

$$f_{X|Y}(x|y) = \frac{f_X(x)f_{Y|X}(y|x)}{\int_{-\infty}^{+\infty} f_X(u)f_{Y|X}(y\mid u)\mathrm{d}u} \tag{1.8}$$

由此看出，本书中的概率密度函数一律以 f 加下标的方法来唯一地确定一个函数，而不是像一些工程学论文中常用的使用自变量的字母来确定一个函数，那种方法容易搞混。因此，在本书中的概率密度函数后面的自变量用什么字母是无关紧要的，将函数 $f_{X|Y}(x|y)$ 写成 $f_{X|Y}(u|v)$ 或者用别的什么字母甚至常量代入都是可以的，符号 $f_{X|Y}$ 就已经规定了此函数。

式(1.8)就是著名的贝叶斯公式，是本书中的核心，可以说本书所有公式的推导都以贝叶斯公式为基础，所有的讨论都围绕贝叶斯公式展开。

1.1.3　正态分布的例子

举一个各种分布都是正态分布的例子。

例 1-1　假设观测客体 X 服从均值为 μ、方差为 σ^2 的正态分布，记为 $X\sim N(\mu,\sigma^2)$，即观测主体关于观测客体 X 的先验分布的概率密度为

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1.9}$$

假设观测量 Y 由观测客体 X 加上一个与之相互独立的误差量 E 构成，$Y=X+E$，$E \sim N(0,\sigma_e^2)$，因此，在给定 $X=x$ 条件下，Y 的条件分布将服从均值为 x、方差为 σ_e^2 的正态分布，即

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}\sigma_e} e^{-\frac{(y-x)^2}{2\sigma_e^2}} \tag{1.10}$$

因为 $Y=X+E$ 是两个相互独立的正态分布的随机变量之和，根据正态分布的特性可知 $Y \sim N(\mu,\sigma^2+\sigma_e^2)$，即

$$f_Y(y) = \frac{1}{\sqrt{2\pi(\sigma^2+\sigma_e^2)}} e^{-\frac{(y-\mu)^2}{2(\sigma^2+\sigma_e^2)}} \tag{1.11}$$

将上面三式代入贝叶斯公式，经过冗繁的推导，可得出在观测主体已经获得观测值 $Y=y$ 条件下，关于 X 的后验概率密度为

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_X(x) f_{Y|X}(y|x)}{f_Y(y)} \\ &= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{1}{\sigma^2}+\frac{1}{\sigma_e^2}} \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_e^2}\right)\left(x-\frac{\mu\sigma_e^2+y\sigma^2}{\sigma_e^2+\sigma^2}\right)^2\right] \end{aligned} \tag{1.12}$$

这个结果后面要经常用到，因此需要详细讨论一下。

我们知道，两个相互独立的正态分布的随机变量相加时，它们的方差也是相加的，学过电学的人如果学过电阻的串并联知识，不妨把方差看做电阻，相互独立的正态分布的随机变量相加时，和的方差与各个变量的方差相当于以串联的形式构成的新的电阻值。

而从式(1.12)可以看出，一个先验正态分布的随机变量在经过一个独立加性正态噪声的测量后，观测主体的后验分布仍然是正态分布，其方差就像是先验方差和测量噪声方差作为电阻并联后获得的阻值；而其均值则是先验均值和测量值之间的加权和，加权的权重与先验方差及测量噪声方差有关系。

因此，考虑到电学在研究电阻并联时用到的电导(电阻的倒数)概念，我们可以定义一个与方差有关的新的随机变量的数字特征。

定义 假设随机变量 X 的方差为 σ^2，则定义此方差的倒数为 X 的**方导**，用字母 γ 表示，也就是

$$\gamma = \frac{1}{\sigma^2}$$

方导越小，方差越大；方导越大，方差越小。而且，一个均值为 μ、方导为 γ 的正态随机变

量的概率密度为

$$f_X(x)=\sqrt{\frac{\gamma}{2\pi}}\mathrm{e}^{-\frac{\gamma(x-\mu)}{2}} \tag{1.13}$$

因此，在例 1-1 中，假设 X 的先验分布的方导为 γ，测量噪声的方导为 γ_e，X 的后验分布的方导为 γ_p，X 的后验均值为 μ_p，则有

$$\begin{cases}\gamma_p=\gamma+\gamma_e\\ \mu_p=\dfrac{\gamma\mu+\gamma_e y}{\gamma+\gamma_e}\end{cases} \tag{1.14}$$

这样的形式看得更清楚，因为在后验均值的加权和中，将和的先验均值 μ 和测量值 y 在公式中捆绑在一起的正是先验方导和测量噪声方导。这时，式(1.12)就可以写成

$$f_{X|Y}(x|y)=\sqrt{\frac{\gamma+\gamma_e}{2\pi}}\exp\left[-\frac{(\gamma+\gamma_e)}{2}\left(x-\frac{\gamma\mu+\gamma_e y}{\gamma+\gamma_e}\right)^2\right] \tag{1.15}$$

当一个随机变量的方导趋向于 0 时，相当于方差趋向于无穷大，即随机变量的取值相当分散。从式(1.15)可以看出，先验方导和测量方导在公式中具有对称性。如果先验方导趋向于 0，则后验方导就趋向于测量误差的方导，且后验均值就取测量值 y。如果测量噪声的方导趋向于 0，说明测量噪声过大，测量结果极不可靠，因此后验方差和后验均值就与先验分布一样，或者说，观测主体关于观测客体的知识没有改变。

例 1-2　设某种鸡的质量服从均值为 1 公斤、标准差为 0.3 公斤的正态分布，一种秤的测量误差服从零均值、标准差为 0.2 公斤的正态分布，用此秤称量其中的一只鸡，秤得的读数为 1.1 公斤，求测量后观测主体对所测量的鸡的质量的后验分布。

解：对于观测主体来说，这只鸡的质量是观测客体 X，它的先验分布服从 $N(1,0.3^2)$，即先验概率密度函数为

$$f_X(x)=\frac{1}{\sqrt{2\pi}\times 0.3}\mathrm{e}^{-\frac{(x-1)^2}{2\times 0.3^2}}=1.33\mathrm{e}^{-5.556(x-1)^2}$$

根据题意，测量误差 E 服从 $N(0,\ 0.2^2)$，此秤的读数 Y 为观测量，满足 $Y=X+E$，由此得到在给定 $X=x$ 条件下，Y 的条件分布是均值为 x、标准差为 0.2 的正态分布，即

$$f_{Y|X}(y|x)=\frac{1}{\sqrt{2\pi}\times 0.2}\mathrm{e}^{-\frac{(y-x)^2}{2\times 0.2^2}}=1.995\mathrm{e}^{-12.5(y-x)^2}$$

因此，我们可以套用例 1-1 的结果，在测量之后，观测主体关于观测客体 X 的后验分布仍为正态分布，方导 γ_p 和均值 μ_p 按式(1.14)算得为

$$\begin{cases}\gamma_p = \dfrac{1}{0.2^2} + \dfrac{1}{0.3^2} = 25 + 11.1 = 36.1 \\ \mu_p = \dfrac{1 \times 11.1 + 1.1 \times 25}{36.1} = 1.069\end{cases}$$

最后获得的观测主体关于观测客体 X 的后验概率密度为

$$f_{X|Y}(x|1.1) = \sqrt{\frac{36.1}{2\pi}}\mathrm{e}^{-\frac{36.1}{2}(x-1.069)^2} = 2.397\mathrm{e}^{-18.05(x-1.069)^2} \tag{1.16}$$

请注意,在此例中,式(1.16)就是我们所要的最终结果,它是一个关于观测客体 X 的概率密度的一元函数。这个一元函数应当存放在观测主体的计算机中,因为观测值代入以后,后验概率密度就只是 x 的一元函数了。

观测过程理论的最终目的就是要获得这样一个概率密度函数,长期存放在观测主体内部的就是函数,而不是一个估计值,这是观测过程理论的要点。

有人认为在计算机中存放一个函数是非常费存储量的。其实不然,尤其是在当代计算机的存储量急剧增加的情况下,保存一条曲线、一个函数存储量并不很大。而在此例中,计算机只需要存储一个函数程序,并保存后验均值和后验方差即可。

1.1.4 知识函数和知识分布

现在我们研究式(1.8)的实现问题,即如何在观测主体的计算机中装入程序,以实现式(1.8)的运算。

从式(1.8)可以看出,要计算出后验概率密度 $f_{X|Y}(x|y)$,观测主体就必须知道 $f_X(x)$ 和 $f_{Y|X}(y|x)$。其中 $f_X(x)$ 为先验概率密度函数,其对应的分布也称为观测主体对观测客体的先验分布。而 $f_{Y|X}(y|x)$ 是观测函数,代表观测通道或者测量仪器的性能,可以通过实验或者分析得到。

因此,作为观测主体的计算机,一开始保存有一个函数 $f_X(x)$,代表在观测前对观测客体的知识。在接收到一个观测值 y 之后,由贝叶斯公式计算出后验概率密度函数 $f_{X|Y}(x|y)$,并存放在内存中。请注意,y 是一个已经得到的具体实数值,对观测主体而言,它是一个具体的常量,因此在固定 y 的情况下,$f_{X|Y}(x|y)$ 也只是 x 的一元函数,观测后在计算机中保存的就是这个一元函数。

因此我们定义

$$f_0(x) = f_X(x)$$

$$f_1(x) = f_{X|Y}(x|y)$$

也就是说，在观测前，观测主体对观测客体的知识是由函数 $f_0(x)$ 来代表的，在接收到观测值 y 之后，经过计算得到的 $f_1(x)$ 代表观测主体对观测客体的新知识。在这里，我们看到的是函数的映射，而我们需要计算的目标是一个函数。由于观测主体已经被假设有无限的计算能力和存储能力，当然能够保存这些函数。

由于先验概率密度 $f_0(x)$ 和后验概率密度 $f_1(x)$ 都反映了观测主体关于观测客体的知识，因此在本书中统称它们为观测主体关于观测客体 X 的**知识函数**，其相应的分布也称为**知识分布**。观测过程理论就是要研究观测主体的知识函数的变化情况和特点。

例 1-3　在例 1-2 中，观测主体用秤对一只鸡的质量进行了测量，在测量前，观测主体关于观测客体（这只鸡的质量）的先验分布是均值为 1 公斤、方差为 0.3 的正态分布，而在测量后，观测主体关于观测客体的后验分布是均值为 1.069 公斤、方导为 36.1 的正态分布。

因此，如果用知识函数 $f_0(x)$ 和 $f_1(x)$ 来表示观测主体关于观测客体的变化，则有

$$f_0(x) = 1.33\mathrm{e}^{-5.556(x-1)^2}$$

$$f_1(x) = 2.397\mathrm{e}^{-18.05(x-1.069)^2}$$

比较上式中的 $f_0(x)$ 和 $f_1(x)$，可以看出观测主体关于观测客体的知识变化情况，这主要体现在方差变小了，或者说方导变大了，或者说观测主体在做了一次测量或者观测之后，对于观测客体的知识变得更为清楚了。

1.1.5　准概率密度函数

我们知道，根据概率论的理论，如果随机变量 X 的概率密度函数为 $f_X(x)$，则它必有

$$\int_{-\infty}^{+\infty} f_X(x)\mathrm{d}x = 1 \tag{1.17}$$

因此，我们将所有与 $f_X(x)$ 成正比的正函数称为 X 的**准概率密度函数**，记为 $g_X(x)$，$g_X(x) = cf_X(x)(c>0)$。因此，X 的准概率密度函数就有无穷多个，但它们之间只差一个常数因子。我们只要知道其中的任何一个准概率密度函数 $g_X(x) = cf_X(x)(c>0)$，就可以通过式(1.17)算出常数 c。由

$$\int_{-\infty}^{+\infty} f_X(x)\mathrm{d}x = \int_{-\infty}^{+\infty} \frac{1}{c} g_X(x)\mathrm{d}x = 1$$

可得

$$c = \int_{-\infty}^{+\infty} g_X(x)\mathrm{d}x \tag{1.18}$$