信息论学术文库·信息处理

观测过程理论

（修订版）

陈必红　著

Observation Process Theory

華中科技大學出版社
http://www.hustp.com

信息论学术文库·信息处理

观测过程理论

Observation Process Theory

（修订版）

陈必红 著

華中科技大學出版社
http://www.hustp.com
中国·武汉

图书在版编目(CIP)数据

观测过程理论(修订版)/陈必红著.—武汉：华中科技大学出版社,2013.8

ISBN 978-7-5609-9379-9

Ⅰ.①观… Ⅱ.①陈… Ⅲ.信息论-应用-统计学 Ⅳ.①C8

中国版本图书馆 CIP 数据核字(2013)第 213787 号

观测过程理论(修订版) 陈必红 著

责任编辑：王汉江

封面设计：范翠璇

责任校对：李 琴

责任监印：周治超

出版发行：华中科技大学出版社(中国·武汉)

武昌喻家山 邮编：430074 电话：(027)81321915

录 排：武汉市洪山区佳年华文印部

印 刷：华中理工大学印刷厂

开 本：850mm×1168mm 1/32

印 张：6.25

字 数：170 千字

版 次：2013 年 11 月第 2 版第 1 次印刷

定 价：20.00 元

华中出版

本书若有印装质量问题,请向出版社营销中心调换

全国免费服务热线：400-6679-118 竭诚为您服务

版权所有 侵权必究

内 容 简 介

本书试图给出有关统计学的新理论——观测过程理论，在此理论中，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。

本书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，给出了标准的无穷大数和无穷小数的定义，使得利用计算机来计算非标准数的四则运算成为可能。

本书可供从事信息论、统计学、随机信号处理、模式识别、检测与估计、时间序列分析等科研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、研究生和高年级本科生阅读。

关键词：信息论、数理统计、模式识别、检测与估计、随机信号处理、时间序列分析。

目　　录

此为试读，需要完整PDF请访问：www.ertongbook.com

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。

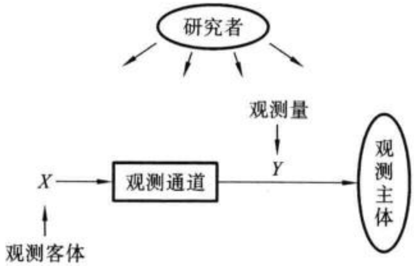

图1-1 观测模型示意图

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，用白体 X 表示，而如果是随机向量，则用黑体 $\mathbf{X}$ 表示。下面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，或者一个机器人、一部雷达等等。就实质而言，观测主体是一台具有计算能力和存储能力的计算机。这台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。我们暂时假设这台计算机具有无限的计算能力和存储能力。

观测主体通常不能够直接观测到观测客体 X,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 y 来获得关于观测客体 X 的知识。

在图 1-1 中还标出了一个**研究者**,其目的是为了在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,不妨假设研究者是一个无所不知的人。因此,对于研究者而言,是没有什么随机变量可言的,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。本书中所谓的"未知",是指的观测主体对观测客体的未知,但同时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、一只鸡的重量;也可以是离散型的,如病人得了乙肝或者没有得乙肝;也可以是混合型的,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,它可以是连续型的、离散型的,或者是混合型的。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

观测过程理论的全部目的,就是要让观测主体在获得观测值 $Y=y$ 的条件下,计算出 X 的条件概率密度函数 $f_{X|Y}(x|y)$。这个

此为试读,需要完整PDF请访问: www.ertongbook.com

条件概率密度，也称作**后验概率密度**，被认为是正确地反映了观测值中带有的有关观测客体的全部知识，或者说是全部信息。后验概率密度所对应的分布，也称作**后验分布**。

有人也许会问："这就是全部目的么？就不作检测和估计么？就不搞模式识别么？"我们的回答是肯定的。只要计算出后验分布，对于观测过程理论来讲，就万事大吉，就不需要再作检测、估计或者模式识别。也就是说，笔者在原则上不赞成任何检测、估计或者模式识别。这一点许多传统的统计学家可能不认同，但下面笔者将进一步说明理由。

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 $f_{X|Y}(x|y)$即是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

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

此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来获得，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

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

将式(1.5)和式(1.6)代入式(1.3),得

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

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,比如用 u 表示,这样将式(1.7)可表示为

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

由此看出,本书中的概率密度函数一律以 f 加下标的办法来唯一地确定一函数,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,那种办法是经常容易搞混的。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或者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**,是本书的核心,可以说本书中的所有公式的推导都以贝叶斯公式为基础,所有的讨论都是以贝叶斯公式展开的。

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

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

假设观测量 Y 为观测客体 X 加上了一个与之相互独立的误差量 E 构成，即

$$Y=X+E,\quad E\sim N(0,\sigma_e^2)$$

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

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

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

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

将上面三式代入贝叶斯公式(1.8)，经过推导可得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

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

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，学过电学的人都知道电阻的串并联知识，若把方差看做是电阻，则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，相当于串联构成的新的电阻值。

而从式(1.12)可以看出，当一个先验正态分布的随机变量，在经过了一个**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。

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)的概念，现在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　假设一个随机变量 X 的方差存在，且为 σ^2，则定义此方差的倒数为 X 的**方导**，用字母 γ 表示，即

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

由此可知，方导越小，则方差越大，方导越大，则方差越小。而且，一个均值为 μ、方导为 γ 的正态随机变量，它的概率密度为

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

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

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

这样的形式显得更清楚，因为在后验均值的加权和中，与先验均值 μ 和测量值 y 在公式中捆绑在一起的正是先验的方导和测量噪声的方导。这时，式(1.12)就可以写成

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

当一个随机变量的方导趋于 0 时，相当于方差趋于无穷大，即

随机变量的取值相当分散。由式(1.15)可以看出,先验的方导和测量方导在公式中具有对称性。如果先验方导趋于0,则后验分布的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保持与先验分布的一样,或者说,观测主体关于观测客体的知识没有改变。

例 1-2 设某一种鸡的重量服从均值为 1 kg,标准差为 0.3 kg 的正态分布,一种秤的测量误差服从零均值,标准差为 0.2 kg 的正态分布,用此秤称量其中的一只鸡,秤得的读数为 1.1 kg,求测量后观测主体对所测量的鸡的重量的后验分布。

解 这只鸡的重量对于观测主体来说,是观测客体 X,它的先验分布服从 $N(1,0.3^2)$,则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 $X=x$ 的条件下,Y 的条件分布为均值是 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,可知在测量之后,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)计算为

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

最后获得的观测主体关于观测客体 X 的后验概率密度为

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

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。这个一元函数就应当存放在观测主体这台计算机中。这是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。

而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。有人问,难道就不估值了吗?是的。长期存放在观测主体内部的就是函数,而不是估计值,这是观测过程理论的要点。

有人会说,计算机存放一个函数是非常浪费计算机存储空间的。其实不然,尤其是在当代计算机的存储空间急剧增加的情况下,保存一条曲线、一个函数,对计算机来说,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。

由式(1.8)可以看出,观测主体若要计算出后验概率密度 $f_{X|Y}(x|y)$,则必须知道 $f_X(x)$ 和 $f_{Y|X}(y|x)$,其中,$f_X(x)$ 称为先验概率密度函数,其对应的分布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而得到。

因此,作为观测主体的计算机,一开始是保存有一个函数 $f_X(x)$,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 $f_{X|Y}(x|y)$ 存放在

此为试读,需要完整PDF请访问: www.ertongbook.com