7.4 贝叶斯估计

统计学中有两个学派：频率学派（又叫经典学派）和Bayes学派。他们之间有共同点，但又具有基本区别.Bayes方法在某些方面对统计学还是有相当帮助的.并且有许多成功的应用.近年来贝叶斯统计发展非常迅猛，影响日益扩大，应用越来越广泛。

1.6.1 三种信息

（1）总体信息，即总体分布族提供的信息，也叫做模型信息。比如，“总体是正态分布”这句话就给我们提供了很多信息。

（2）样本信息，这是最“新鲜”的信息。

基于以上两种信息进行统计推断的统计学就是经典统计学。然而还有第三种信息—先验信息，它也可以用于统计推断.

（3）先验信息，即在抽样之前的有关统计问题的信息。

基于以上三种信息进行统计推断的统计学就是贝叶斯统计学。它与经典统计学的差别就在于是否利用先验信息。

在经典统计学中,参数被认为是一个未知,但固定的量(不是随机变量).基于样本观察值获得的知识,统计推断在设定的统计模型下基于样本进行.经典统计学中的信息来源有两个方面:(1)结构信息（模型信息），（2）样本信息.

在Bayes统计中,参数被考虑成一个随机变量,其不确定性可被一个概率分布描述,该分布叫做先验分布.这是一个主观的分布,建立在试验者的信念之上,并且在得到观察数据之前就已经用公式制定好了.然后从以为指标的总体中抽取样本,先验分布通过样本信息得到校正,这个被校正的分布叫做后验分布,再利用后验分布进行推断.由于这个校正工作是通过Bayes法则完成的,因而称为Bayes统计.Bayes统计除了利用结构信息和样本信息外,还要利用先验信息.

贝叶斯统计学起源于英国学者贝叶斯（Bayes.T.R）去世后发表的一篇论文《论有关机遇问题求解》，此文中提出了著名的贝叶斯公式，之后被一些统计学家发展成一种系统的统计推断方法。

1.6.2 后验分布的推导

下面来推导后验分布,这里都以连续情况来说明,对于离散情况可类似进行.

设的先验密度为(这里，随机变量,密度函数中的自变量用同一个字母),样本的密度记为,其中,注意此处样本的密度与经典统计学中所说的样本的密度有所区别.Bayes统计学中所说的样本的密度是指在参数(它是一个随机变量)取定(它是一个值)的条件下,样本的条件密度.也就是说样本的产生分两步进行,首先设想从先验分布中产生一参数值(怎么产生的?不知道,“老天爷”产生的)。然后从条件分布中产生样本。这样，样本与参数的联合密度为

,

我们的任务就是对参数作推断.在没有样本时，只能依据先验分布.在有了样本观察值后,我们应依据作推断,更具体地,应依据的条件密度对作推断.

由样本与参数的联合密度，可得样本的边缘密度为

，

从而在给定样本的条件下,的条件密度为

，

这个密度称为参数的后验密度，更一般地叫做后验分布,后验分布是在给定样本观察值的条件下参数的条件分布.它集中了模型信息、样本信息(这两种信息综合在之中,在经典统计学中称它为似然函数,在贝叶斯统计学中也叫似然函数)以及先验信息(指).

1.6.3 贝叶斯估计

关于参数的所有信息都包含在后验分布中.涉及参数的推断都基于这个后验分布. 参数的推断同样也有估计和检验问题.参数的一个自然的估计是这个后验分布的均值，称为后验均值估计.有时也会用后验中位数或后验众数作为参数的估计.

作为的估计的均方误差是

,

这里要注意的是,上面的期望是对后验分布取的期望(再次记住:是随机变量,具有分布,而是样本的函数，不是随机变量).即

.

根据概率论的知识知时, 取得最小值,且最小值就是的后验方差. 可见参数的后验均值是使得均方误差最小的估计.如果用后验均值估计参数,那么后验方差就可刻画该估计的精度.后验均值估计常称为贝叶斯估计.

例1.6.1设总体～，来自总体的简单随机样本，又设的先验密度为，

（1）求的后验分布;

(2) 设的先验分布为(这里我们假定都已知),求的Bayes估计;

(3) 设的先验分布为(这里我们假定已知),求的Bayes估计.

解：(1)给定的条件下，样本的条件密度（概率质量密度）函数为



从而样本与的联合密度为



那么样本的边缘密度函数为

，

从而得的后验分布为

.

(2) 的后验分布为

,

注意到后验分布中的分母与无关,是一个常数.而分子是,它正是分布的核,可记为

∝

由此可以看出的后验分布为.从而得的后验均值估计为

.

它是先验均值与样本均值的加权平均.

后验方差为

.

的后验众数估计为

.

考虑网格的纤维数数据,假设的先验分布为,由数据得样本均值为,从而得的后验均值估计值为



后验方差为



后验众数估计值为

(3) 的后验分布为

,,

为求的后验均值与后验众数,我们不得不采取数值计算的方法.

考虑网格的纤维数数据,假设的先验分布为,通过数值计算得 ,得的后验均值估计值为,后验众数估计值为.

把以上计算结果及最大似然估计结果作对比,见下表:

|  |  |  |  |
| --- | --- | --- | --- |
| 估计 | 贝叶斯1 | 贝叶斯2 | 最大似然 |
| 众数  均值  标准差 | 24.6  24.7  1.02 | 24.9  25.0  1.04 | 24.9  1.04 |

例1.6.2 (二项分布的Bayes的估计) 设是来自总体的样本,则～.假定的先验分布是贝塔分布, 已知.求的贝叶斯估计.

解 与的联合分布为



的边缘分布密度为

，

给定条件下,的后验分布为



即得后验分布是,的贝叶斯估计为



该估计可分成两部:



这样，就表示成样本均值和先验均值的加权平均，其权重由确定.特别时，即的先验分布是均匀分布时，的Bayes估计为

.

这个估计与经典统计中的估计有些差别,当很大时两者差别不大,而当很小时两者差别明显. 当很小时,Bayes估计更合理些.

在具体问题中，两个超参数(先验分布的参数叫做超参数)是需具体确立下来的。下面分几种情况讨论。

（1） 如果对先验信息很缺乏，那么设定的先验分布是均匀分布（即）是可以接受的。这是一种无信息先验，也称为贝叶斯假设。

（2）如根据先验信息，比如有一些历史数据，经加工整理可得到的先验均值和先验方差，由贝塔分布的期望和方差，建立如下方程组：



解此方程组，确定超参数.

(3)假如根据先验信息只能获得先验均值,可建立如下一个方程



但一个方程不能确定两个超参数.如果能指定就能得到两个超参数的值. 由于的先验方差随增大而变小,比如,选择的问题就转化为决策人对的确信程度.若对很确信,那么就选得大些,否则就选得小些.

(4)用分位数来确定超参数.比如，利用先验信息可以确定的先验下四分位数和上四分位数,那么可建立如下方程组



由此确定超参数(需要用到数值计算方法).

1.6.4 共轭先验分布

先验分布的确定在贝叶斯统计中是关键的一步,它会影响最后的推断结果.先验分布确定的原则有两个:一是要根据先验信息(经验和历史资料等);二是要使用方便.当然第二个原则要服从第一个原则,但如果可利用的先验信息很少甚至没有时，则应优先考虑 “使用方便”的原则，并且还应作灵敏性分析。确定先验已有一些较为成熟的方法，具体有

（1）共轭先验；

（2）无信息先验；

（3）多层先验等。

在例1.6.1中，我们发现若先验分布为Gamma分布,则后验分布也为Gamma分布(这与总体分布有关),只是Gamma分布中的参数有变化,这样的先验就称为共轭先验.如果使用共轭先验,那后验分布就可容易地导出,这就对数学上的处理带来了很大方便. 共轭先验的一般定义如下.

定义 设是某分布中一个参数，其先验分布为,假如结合样本导出的后验分布与先验分布为同属于一个分布族,则称先验为共轭先验.

从定义可以看出,共轭先验是针对某一分布中的参数而言的,具体说就是一个先验是否为共轭先验与总体分布族(或统计模型,即给定后样本的条件分布)有关.在例1.6.1中,若总体分布不是泊松分布,Gamma分布未必是共轭先验.下表列出了一些常用的共轭先验.

|  |
| --- |
| 总体分布 参数 共轭先验  二项分布 成功概率 贝塔分布  泊松分布 均值 伽玛分布  指数分布 均值倒数 伽玛分布  正态分布(方差已知) 均值 正态分布  正态分布(均值已知) 方差 倒伽玛分布 |

例1.6.3 若总体分布为正态分布,其中已知, 为未知参数,证明正态分布是的共轭先验,并求的贝叶斯估计及后验方差.

证明 与的联合密度为



所以的后验分布密度为

∝

由于







这里与无关.所以

∝

即可得的后验分布为正态分布



从而得的后验均值估计为



,

后验方差为

.

在共轭先验布中常含有待定的超参数.比如在上例1.6.3中均值的先验分布中有两个超参数,这是需要根据先验信息确定的。在先验分布类型已定，但其中含有超参数，确定先验分布的问题就转化为估计超参数的问题。

1.6.5 后验分布的计算

1.利用分布的核简化后验分布的计算

在给定样本分布和先验分布后,后验分布

,

其中是样本的边际密度,它与无关,在后验分布计算中仅起到一个正则化因子的作用.这时可把称为后验分布的核,假如中还含不依赖于因子,还可剔去这个因子,使核进一步精炼.这种处理方法其实在例1.6.1中已经使用过了.

2.利用充分统计量简化后验分布的计算

在经典统计学中,由因子分解定理,统计量是参数的充分统计量的充要条件是样本的密度可表示为

,

在贝叶斯统计学中,对充分统计量也有一个充要条件: 统计量是参数的充分统计量的充要条件是参数的后验分布可表示为和的函数:

.

可以证明,以上两个充要条件是等价的.这样在推导后验分布时,可以利用充分统计量的分布,具体做法是这样的.

设是充分统计量,其密度为,则与的联合密度为



从而得的后验密度为



其中.

充分统计量的概念及统计推断的充分性原则是两个学派相一致的少数几个论点之一.另一个相一致的论点是:似然原理.

例1.6.4 (正态总体均值的Bayes的估计)设总体～,为来自总体的简单随机样本,设的先验分布(这里我们假定都已知). 求的Bayes估计.

解 由～,可得与的联合密度为



经过推导,可得后验分布

∝.

所以的后验分布为.

例1.6.5经过早期筛选后的彩色电视接收机的寿命服从指数分布，它的密度为

，

其中是彩电的平均寿命。现从一批彩电中随机取台作定数截尾寿命试验.试验到第台失效为止，其失效时间分别为，另外还未失效，但也停止试验。所得样本为截尾样本。试求的贝叶斯估计。

解 截尾样本的联合分布为







其中称为总试验试时间。

为寻求的贝叶斯估计，我们给出的先给分布。根据国内外的经验，选用倒伽玛分布作为先验分布是合适的，其密度函数为



其中是两个待定的超参数，倒伽玛分布的期望为。

利用贝叶斯公式可得的后验密度为



可见后验分布是倒伽玛分布，后验期望为，故的贝叶斯估计为

。

为了确定这个估计，还须确定两个超参数。我们收集了大量的先验信息。我国彩电生产厂做了大量的寿命试验，仅15个工厂实验室和一些独立实验室就对13412台彩电进行了共计5369812台时试验，而且还对9240台彩电进行了三年现场跟踪试验，总共进行了5547810台时试验.这两类试验总共失效台数不超过250台。对如此大量先验信息加工整理后，确认我们彩电平均寿命不低于30000小时，它的10%的分位数大行为小时，专家认为这两个数据是符合我国前几年彩电的实际情况，也是留有余地的。

由此可列出如下两个方程：



利用计算机解此方程组，得

，

这样，我们就完全确定了先验分布.假如随机抽取100台彩电进行400小时试验,没有一台失效.这时总试验时间小时，，于是彩电平均寿命的贝叶斯估计为小时。

1.6.6 贝叶斯区间估计

1 可信区间

对区间估计，贝叶斯方法具有处理方便且含义清晰的优点。

当参数的后验分布获得后，立即可计算落入任一区间的概率。反过来，对于给定的概率，如果区间满足



我们就称为的的可信区间。当参数的后验分布是连续分布时，这样的区间总是找得到的。但若参数的后验分布是离散分布时，这样的区间不一定存在，此时把上面的概率等式改为不等式

,

我们也称为的的可信区间。一般定义如下。

定义 设参数的后验分布为，对于给定的样本和概率，若存在两个统计量和，使得

，

则称区间为的可信水平为的可信区间(Credible Interval)，或简称为的的可信区间。

若统计量满足

，

则称为的可信水平为的单侧可信下限，简称为的的可信下限。

若统计量满足

，

则称为的可信水平为的单侧可信上限，简称为的的可信上限。

例1.6.6 是来自正态总体的样本，其中已知。设的先验分布为，其中已知。那么的的后验分布为，其中

，。

由此可很容易地得到的的可信区间，只要区间满足



那么区间就是的的可信区间。自然地我们希望找到最短的的可信区间。由正态分布密度的性质，容易看出最短的可信区间是关于后验均值对称的区间，该区间就是

例如，在儿童智商测验中设某儿童测验得分为，其中智商的先验分布为，在仅有一个样品（）情况下，算得的后验分布为，其中

，。

该儿童在一次智商测验中的得分为，立即可得的后验分布为，的的可信区间为。

在比例子中，若不用先验信息仅使用样本信息，则按经典方法可得的的置信区间为。

例1.6.7 经过早期筛选后的彩色电视接收机的寿命服从指数分布，它的密度为

，

其中是彩电的平均寿命。在例2.6.5中曾选用的共轭先验分布—倒伽玛分布，并利用先验信息确定其中两个参数：，。后又利用样本信息（100台彩电进行400小时试验，无一台失效，即，）。最后得到后验分布，在那里还获得了的贝叶斯估计

（小时）

现在来求平均寿命的0.90可信下限，即找出，使得，或。

这需要用的后验分布去计算，但没有现成的倒伽玛分布的分位数表，但我们可利用伽玛分布与分布的关系将倒伽玛分布的分位数的计算转换成分布的分位数的计算：因为，从而，所以



这样



故有

，

即



将，， ，代入上式可得

。

2 最大后验密度（HDP）可信区间

对于给定的可信水平,从后验分布获得的可信区间可以有许多个,常用的方法是把平分,用和分位数来获得区间估计.

等尾可信区间使用简便,在实际应用也经常使用,但不是最理想的,最理想的可信区间应是最短的那个可信区间.容易达到，最短的可信区间应该是这样的区间:具有高后验密度的点组成的区间,此区间外的点的后验密度函数值都不会超过区间内的点的密度值,这样的区间称为最大后验密度可信区间,一般定义如下.

定义 设参数的后给密度,对于给定的,若在直线上存在一个子集,满足

(1) ;

(2)对于任意的和,都有,

则称为的的最大后验密度可信集,简称为的的HDP可信集。如果是一个区间，则称它为的HDP可信区间。

这个定义仅对后验分布为连续分布而给，这是因为当后验分布是离散分布时，HDP可信集很难实现。从这个定义可以看出，当后验密度函数是单峰时，一般总可找到HDP可信区间，而当后验密度函数是多峰时，可能得到几个互不连接的区间组成的HDP可信集,此时很多统计学家建议放弃HDP准则,采用等尾可信区间为宜。顺便指出,当后验密度出现多峰时，常常是由于先验信息与抽样信息不一致引起的,认识或研究此种不一致往往是重要的.共轭先验分布大多是单峰的,这导致后验分布也是单峰的,它可能会掩盖这种抵触,这种掩盖有时是不好的,这就告诉我们,要慎重对待和使用共轭先验分布.

当后验密度是单峰对称时，寻求参数的的HDP可信区间较为容易，它就是等尾可信区间。当后验密度是单峰但不对称时，寻求HDP可信区间并不容易，这时可借助计算机。