

法律声明

□ 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。

□ 课程详情请咨询

■ 微信公众号：小象

■ 新浪微博：ChinaHadoop



数理统计

目录

第一章	统计量与抽样分布
第二章	点估计
第三章	区间估计
第四章	假设检验
第五章	分布的检验
第六章	Bayes统计与统计判决理论
第七章	非参数统计
第八章	抽样调查

目 录

6.1 **Bayes**统计

6.2 先验分布的确定

6.3 Bayes统计推断

6.4 统计判决理论

6.5 Minimax准则

Bayes学派

我们在概率论中学过如下的Bayes公式：

$$P(H_k|A) = \frac{P(H_k)P(A|H_k)}{\sum_i P(H_i)P(A|H_i)}, \quad k = 1, 2, \dots,$$

其中 $\cup H_i = \Omega$ 是必然事件的一个划分. 当时我们称 $P(H_k)$ 为先验(prior), 而 $P(H_k|A)$ 称为后验(posterior). 这个公式是在英国学者Thomas Bayes (1702–1761)于1763年发表的. 从形式上看, 它只是条件概率的简单应用, 但它却包含了归纳推理的思想, 后来的学者把它发展成一种关于统计推断的系统理论与方法, 统称为Bayes统计, 后来发展成了Bayes学派. 当然, Bayes学派在统计界引起了热烈的争论. 有关Bayes方面的内容, 请有兴趣的参见Berger (1985).

Bayes统计

Bayes 学派的观点在统计学界引起了广泛的争论. 频率 (经典) 学派与 Bayes 学派近几十年来的争论推动了数理统计学的发展. 这两大学派之间有共同点, 也有不同点, 为了弄清楚他们的主要差别, 首先介绍统计推断中所使用的三种信息.

(1) 总体信息,即总体分布或总体所属分布族给我们的信息.

(2) 样本信息,即样本提供给我们的信息, 这是最“新鲜”的信息, 并且越多越好, 希望通过样本对总体分布或总体分布的某些特征作出较精准的统计推断.

基于以上两种信息进行统计推断的统计学就称为**经典统计学**.

(3) 先验信息,即抽样之前有关统计问题的一些信息.

基于上述三种信息进行统计推断的统计学称为**贝叶斯统计学**.它与经典统计学的差别就在于是否利用了先验信息.

先验信息的重要作用

例 7.1.1 英国统计学家 Savage 在 1961 年提出一个令人信服的例子说明先验信息有时是很重要的, 且看下面两个试验:

(1) 一位常饮牛奶和茶的女士说她能辨别先倒进杯子里的是茶还是牛奶. 对此做了 10 次试验她都说对了.

(2) 一位音乐家说他能够从一页乐谱辨别是海顿 (Haydn) 还是莫扎特 (Mozart) 的作品, 在 10 次试验中他都说对了.

在上面两个试验中, 如果认为试验者是猜对的, 每次成功概率为 0.5, 则 10 次都猜中的概率为 $(1/2)^{10} = 2^{-10} = 0.0009766$, 这是一个很小的概率, 几乎不可能发生. 故每次猜对的概率为 0.5 的假设被否定. 他们每次说对的概率比 0.5 要大得多, 这不能认为是猜测, 而是经验帮了忙. 可见经验 (先验信息) 在推断中不可忽视, 应当加以利用.

先验信息的重要作用

例 7.1.2 某工厂生产一种产品, 每日抽查一部分 (n 件) 产品以检查废品率 θ , 经过一段时间后获得大量数据, 对 θ 作出估计. 就当日被抽查的那批产品的废品率 θ 而言, 它只是一个固定的数, 并无随机性可言; 但逐日的废品率 θ 受随机因素的影响多少会有些波动. 从长期看, 将“一日废品率 θ ”视为随机变量, 而要估计的某日的废品率是这个随机变量的一个观察值. 根据历史资料可构造一个分布

$$P\left(\theta = \frac{i}{n}\right) = \pi_i, \quad i = 0, 1, 2, \dots, n.$$

这个对先验信息进行整理加工而得到的分布称为先验分布. 先验分布总结了该厂过去产品质量情况. 如果这个分布的概率大多集中在 $\theta = 0$ 附近, 那该产品可以认为是“信得过产品”. 假如以后多次抽样与历史资料提供的先验分布一致, 使用单位就可以作出“免检产品”的决定, 或者每月抽一、两次就足够了, 省去大量人力和物力.

先验分布与后验分布

定义 7.1.1 (先验分布) 参数空间 θ 上的任一概率分布都称为先验分布 (prior distribution).

定义 7.1.2 (后验分布) 在获得样本 X 后, θ 的后验分布 (posterior distribution) 就是给定 $X = x$ 条件下 θ 的条件分布, 后验密度 $\pi(\theta|x)$ 有下列计算公式:

$$\pi(\theta|x) = \frac{f(x, \theta)}{m(x)} = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta}. \quad (7.1.1)$$

(注: 在离散情形下它就转变成 Bayes 公式), 其中

$$f(x, \theta) = f(x|\theta)\pi(\theta)$$

为 X 和 θ 的联合密度, 而

$$m(x) = \int_{\Theta} f(x|\theta)\pi(\theta)d\theta$$

为 X 的边缘密度.

Bayes统计

Bayes统计的基本观点可以用下面三个假设来归纳：

假设I 设随机变量 X 有一个PDF $f(x, \theta)$, 其中 θ 是一个参数, 由于对不同的 θ 值, $f(x, \theta)$ 对应着不同的PDF, 所以, 从Bayes观点看, 它是在给定 θ 值后的一个条件PDF. 因此把 $f(x, \theta)$ 记为 $f(x|\theta)$ 更恰当些, 而 $f(x|\theta)$ 中关于 θ 的信息即为总体信息.

假设II 当给定 θ 后, 从总体 $f(x|\theta)$ 中随机地抽取一组样本 $X_1, \dots, X_n$, 该样本中含有大量的关于 θ 的信息, 这就是样本信息. 于是, 样本分布 $f(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta)$ 即综合了总体与样本信息.

假设III 我们对参数 θ 已经积累了一些资料, 经过分析整理和加工后, 可以获得一些关于 θ 的有用信息, 这种信息称为 θ 的先验信息, 由于 θ 不是永远固定在一个值上, 故从Bayes观点看, 未知参数 θ 也是一个随机变量. 关于 θ 的分布可以从先验信息中归纳出来, 故其分布称为先验分布, 用 $\pi(\theta)$ 表示其PDF.

Bayes统计

这样，综合前面几点，我们有样本 $X_1, \dots, X_n$ 及 θ 的联合PDF为

$$f(\mathbf{x}, \theta) = f(\mathbf{x}|\theta)\pi(\theta). \quad (2.8.1)$$

而我们的目的在于利用上面的联合分布，即三方信息来估计 θ . 当有了样本后，我们可以得到 θ 的条件分布如下：

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}, \theta)}{f(\mathbf{x})} = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int f(\mathbf{x}|\theta)\pi(\theta)d\theta}, \quad (2.8.2)$$

我们称之为 θ 的后验PDF, 而 $f(\mathbf{x}) = \int f(\mathbf{x}|\theta)\pi(\theta)d\theta$ 是 $X_1, \dots, X_n$ 的边际分布或样本的无条件分布.

Bayes统计推断的原则

Bayes 统计推断的原则:对参数 θ 所有的任何推断 (估计、检验等) 都必须基于且只能基于 θ 的后验分布.

目 录

6.1 Bayes统计

6.2 先验分布的确定

6.3 Bayes统计推断

6.4 统计判决理论

6.5 Minimax准则

确定先验分布的方法

用 Bayes 方法处理问题的前提是给出 θ 的先验分布. 因此, 怎样去确定一个先验分布称为一个至关重要的问题. 常用的方法有:

- 1, 客观法
- 2, 主观概率法
- 3, 同等无知原则
- 4, 无信息先验分布
- 5, 共轭先验分布

无信息先验分布

Bayes 分析的一个重要特点就是在统计推断时要利用先验信息. 但常常会出现这样的情况: 没有先验信息或者只有极少的先验信息可利用, 但仍想用 Bayes 方法. 此时所需要的是一种无信息先验 (noninformative prior), 即对参数空间 θ 中的任何一点 θ 没有偏爱、都是同等无知的先验信息. 这就引出了无信息先验分布的概念.

1. 均匀分布与广义先验分布

(1) 若 θ 为有限集, 即 θ 只可能取有限个值, 如 $\theta = \theta_i, i = 1, 2, \dots, n$, 无信息先验给 θ 中的每个元素以概率 $1/n$, 即 $P(\theta = \theta_i) = 1/n, i = 1, 2, \dots, n$.

(2) 若 θ 为 R_1 上的有限区间 $[a, b]$, 则取无信息先验为区间 $[a, b]$ 上的均匀分布 $U(a, b)$ (有时也记为 $R(a, b)$).

(3) 问题是若参数空间 θ 无界, 无信息先验如何选取? 例如, 样本分布为 $N(\theta, \sigma^2)$, σ^2 已知, 此时 $\theta = (-\infty, \infty)$. 若无信息先验取为 $\pi(\theta) \equiv 1$, 则 $\pi(\theta)$ 不是通常的密度, 因为 $\int_{-\infty}^{\infty} \pi(\theta) d\theta = \infty$. 这就引出广义先验分布的概念.

无信息先验分布

定义 7.2.2 设随机变量 $X \sim f(x|\theta)$, $\theta \in \Theta$, 若 θ 的先验密度 $\pi(\theta)$ 满足下列条件:

(1) $\pi(\theta) \geq 0$ 且 $\int_{\Theta} \pi(\theta) d\theta = \infty$;

(2) 后验密度 $\pi(\theta|x)$ 是正常的密度函数,

则称 $\pi(\theta)$ 为 θ 的广义先验密度 (improper prior density).

注 7.2.2 由定义可见广义先验密度 $\pi(\theta)$ 乘以任一给定的常数仍是一个广义先验密度。

无信息先验分布

例 7.2.4 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从 $N(\theta, 1)$ 总体中抽取的随机样本, 设 θ 的先验密度 $\pi(\theta) \equiv 1$, 求 θ 的后验密度.

解 由式 (7.1.1) 可知

$$\begin{aligned}\pi(\theta|\mathbf{x}) &= \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_{-\infty}^{\infty} f(\mathbf{x}|\theta)\pi(\theta)d\theta} = \frac{\exp\left\{-\frac{1}{2}\sum_{i=1}^n(x_i - \theta)^2\right\}}{\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\sum_{i=1}^n(x_i - \theta)^2\right\}d\theta} \\ &= \sqrt{\frac{n}{2\pi}} \exp\left\{-\frac{n}{2}(\theta - \bar{x})^2\right\}.\end{aligned}$$

这是正态分布 $N(\bar{x}, 1/n)$ 的密度函数, 后验分布 $\pi(\theta|\mathbf{x})$ 仍为正常的密度函数, 故 $\pi(\theta) \equiv 1$ 为广义先验密度, 它也是一种无信息先验.

共轭先验分布

另外一种选择先验的方法是从理论角度出发的, 在已知样本分布的情形下, 为了理论上的需要常常选参数的先验分布为共轭先验分布, 其定义如下.

定义 7.2.3 设 θ 是总体分布中的参数, 样本 X 的分布为 $f(x|\theta)$, 如果对任取的先验分布 $\pi(\theta) \in \mathcal{F}$ 及样本 x , 后验分布 $\pi(\theta|x)$ 仍属于 $\mathcal{F}$, 则称 $\mathcal{F}$ 是一个共轭先验分布族 (conjugate prior distribution family).

共轭先验分布

例 7.2.9 设 $X \sim b(n, \theta)$.

(1) 设 $\theta \sim U(0, 1)$, 即 $(0, 1)$ 上的均匀分布, 证明 θ 的后验分布为 Beta 分布;

(2) 若取 θ 的先验分布为 Beta 分布 $\text{Be}(a, b)$, 证明 θ 的后验分布仍为 Beta 分布, 即样本分布如果为二项分布, 则共轭先验分布为 Beta 分布.

共轭先验分布

证 (1) $U(0,1)$ 是 $\text{Be}(1,1)$ 分布, $X \sim b(n, \theta)$, 其概率分布为

$$f(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x},$$

而 θ 的先验分布为 $\pi(\theta) = 1$, 当 $\theta \in (0,1)$, 故有

$$\pi(\theta|x) = \frac{\theta^x (1-\theta)^{n-x}}{\int_0^1 \theta^x (1-\theta)^{n-x} d\theta}. \quad (7.2.1)$$

计算积分得到

$$\int_0^1 \theta^x (1-\theta)^{n-x} d\theta = \frac{\Gamma(n-x+1)\Gamma(x+1)}{\Gamma(n+2)},$$

将上式结果代入式 (7.2.1), 得到后验密度

$$\pi(\theta|x) = \frac{\Gamma(n+2)}{\Gamma(n-x+1)\Gamma(x+1)} \theta^{(x+1)-1} (1-\theta)^{(n-x+1)-1},$$

即 θ 的后验分布是 Beta 分布 $\text{Be}(x+1, n-x+1)$.

共轭先验分布

(2) 又若 $\theta \sim \text{Be}(a, b)$, 则

$$\pi(\theta|x) = \frac{\theta^{x+a-1}(1-\theta)^{n-x+b-1}}{\int_0^1 \theta^x(1-\theta)^{n-x} \theta^{a-1}(1-\theta)^{b-1} d\theta}. \quad (7.2.2)$$

计算积分得到

$$\int_0^1 \theta^{x+a-1}(1-\theta)^{n-x+b-1} d\theta = \frac{\Gamma(x+a)\Gamma(n-x+b)}{\Gamma(n+a+b)}.$$

将上式结果代入式 (7.2.2), 得到后验密度

$$\pi(\theta|x) = \frac{\Gamma(n+a+b)}{\Gamma(x+a)\Gamma(n-x+b)} \theta^{(x+a)-1} (1-\theta)^{(n-x+b)-1},$$

即 θ 的后验分布是 Beta 分布 $\text{Be}(x+a, n-x+b)$. 因此, 样本分布若为二项分布, 其参数 θ 的共轭先验分布族为 Beta 分布族.

共轭先验分布

例 7.2.10 设 $X \sim N(\theta, \sigma^2)$, σ^2 已知而 θ 未知. 令 θ 的先验分布 $\pi(\theta)$ 是 $N(\mu, \tau^2)$, 其中 μ 和 τ^2 已知, 求 θ 的后验分布 $\pi(\theta|x)$.

共轭先验分布

解 给定 θ 时 X 的条件分布记为 $f(x|\theta)$, 令

$$A = \frac{1}{\sigma^2} + \frac{1}{\tau^2}, \quad B = \frac{x}{\sigma^2} + \frac{\mu}{\tau^2}, \quad C = \frac{x^2}{\sigma^2} + \frac{\mu^2}{\tau^2}, \quad (7.2.5)$$

则有

$$\begin{aligned} \pi(\theta|x) &\propto f(x|\theta)\pi(\theta) \propto \exp \left\{ -\frac{1}{2} \left[\frac{(x-\theta)^2}{\sigma^2} + \frac{(\theta-\mu)^2}{\tau^2} \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} [A\theta^2 - 2\theta B + C] \right\} \\ &= \exp \left\{ -\frac{A}{2} \left(\theta - \frac{B}{A} \right)^2 - \frac{1}{2} \left(C - \frac{B^2}{A} \right) \right\} \\ &= \exp \left\{ -\frac{1}{2\eta^2} (\theta - \mu(x))^2 \right\} \cdot \exp \left\{ -\frac{(x-\mu)^2}{2(\sigma^2 + \tau^2)} \right\} \\ &\propto \exp \left\{ -\frac{1}{2\eta^2} (\theta - \mu(x))^2 \right\} \end{aligned} \quad (7.2.6)$$

其中

$$\mu(x) = \frac{B}{A} = \frac{\sigma^2\mu + \tau^2x}{\sigma^2 + \tau^2}, \quad \eta^2 = \frac{1}{A} = \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}, \quad C - \frac{B^2}{A} = \frac{(x-\mu)^2}{\sigma^2 + \tau^2}.$$

共轭先验分布

易见, 式 (7.2.6) 右边是 $N(\mu(x), \eta^2)$ 密度的核, 添加正则化因子, 得到后验密度

$$\pi(\theta|x) = \frac{1}{\sqrt{2\pi\eta}} \exp \left\{ -\frac{1}{2\eta^2} [\theta - \mu(x)]^2 \right\}.$$

若进一步假定 $X_1, \dots, X_n$ i.i.d. $\sim N(\theta, \sigma^2)$, σ^2 已知, $\theta \sim N(\mu, \tau^2)$, 求 θ 的后验密度.

由于样本均值 $\bar{X}$ 为 θ 的充分统计量, 且给定 θ 时 $\bar{X}$ 的分布是 $N(\theta, \sigma^2/n)$, 故将上述结果中的 x 用 $\bar{x}$ 代替, σ^2 用 σ^2/n 代替得到 θ 的后验分布

$$\pi(\theta|\bar{x}) \sim N(\mu_n(\bar{x}), \eta_n^2),$$

其中

$$\mu_n(\bar{x}) = \frac{\sigma^2/n}{\sigma^2/n + \tau^2} \mu + \frac{\tau^2}{\sigma^2/n + \tau^2} \bar{x}, \quad \eta_n^2 = \frac{\tau^2 \cdot \sigma^2/n}{\sigma^2/n + \tau^2} = \frac{\sigma^2 \tau^2}{n\tau^2 + \sigma^2}.$$

此例表明当样本分布为方差已知的正态分布时, 则均值参数 θ 的共轭先验分布族是正态分布族.

共轭先验分布

注 7.2.5 共轭先验分布具有下列优点：① 计算方便；② 后验分布的某些参数常可以得到很好的解释。例如，例 7.2.10 中，后验分布为 $N(\mu_n(\bar{x}), \eta_n^2)$ ，其中

$$\mu_n(\bar{x}) = \frac{\tau^2}{\sigma^2/n + \tau^2} \bar{x} + \frac{\sigma^2/n}{\sigma^2/n + \tau^2} \mu = r\bar{x} + (1 - r)\mu$$

是样本均值 $\bar{x}$ 和先验均值 μ 的加权平均。若 σ^2/n 很小，即样本信息量很大，相对的先验信息很少，则后验均值主要由 $\bar{x}$ 决定，反之若 σ^2/n 很大（即样本信息很少），则后验均值主要由先验分布的均值 μ 来决定。由此可见后验均值是样本均值和先验均值的一个折中。

目 录

6.1 Bayes统计

6.2 先验分布的确定

6.3 Bayes统计推断

6.4 统计判决理论

6.5 Minimax准则

Bayes点估计

从上式可以看出, 参数 θ 的后验分布归纳了 θ 的先验信息、总体中关于 θ 的信息和样本提供的信息, 于是, θ 的估计就应从其后验分布入手. 常用的估计方法为

- (1) 取后验分布 $\pi(\theta|\mathbf{x})$ 的最大值, 即众数作为 θ 的估计, 称之为众数型的Bayes估计.
- (2) 取后验分布 $\pi(\theta|\mathbf{x})$ 的中位数作为 θ 的估计, 称之为中位数型的Bayes估计.
- (3) 取后验分布 $\pi(\theta|\mathbf{x})$ 的期望作为 θ 的估计, 称之为期望型的Bayes估计.

在上述三种常用的估计中, 第三种, 即期望型的Bayes估计最常用. 以后我们将采用后验期望作为 θ 的Bayes估计:

$$\hat{\theta}_B = E(\theta|\mathbf{X}) = \int \theta \pi(\theta|\mathbf{x}) d\theta. \quad (2.8.3)$$

Bayes点估计

例 2.8.1 设某产品的次品率为 θ , 请估计 θ .

解 为估计 θ , 现随机地有返回抽取 n 件产品进行检验, 发现有 X 件次品, 则次品 X 的条件分布为

$$P\{X = x|\theta\} = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad x = 0, 1, \dots, n.$$

如从经典统计来看, 次品数 X 是 θ 的充分统计量, 且其MLE为

$$\hat{\theta}_{ML} = \frac{X}{n} = \frac{\text{样本中次品数}}{\text{样本容量}}.$$

前面我们已经讲过, MLE具有许多优良性质, 也是估计中经常利用的. 对于本例, 假设我们有如表2.4的数据, 则得到的MLE $\hat{\theta}_{ML}$ 也在表中给中.

Bayes点估计

表2.4 MLE与Bayes估计的比较

No.	n	X	$\hat{\theta}_{ML}$	$\hat{\theta}_B$
1	5	5	1	0.875
2	20	20	1	0.955
3	5	0	0	0.143
4	20	0	0	0.045

对于本例，由于我们对过去的信息一无所知，故可采用无信息先验，即采用(0,1)上的均匀分布，即 $\alpha = \beta = 1$ 的 β 分布. 这样，得到的 θ 的Bayes估计为

$$\hat{\theta}_B = \frac{X + 1}{n + 2},$$

这即是我们在前面表中列出的估计值. 比较一下此表中的Bayes估计与MLE，可以看出Bayes估计的合理性.

Bayes点估计

例 7.3.1 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从正态总体 $N(\theta, \sigma^2)$ 中抽取的简单样本, 其中 σ^2 已知. 取 θ 的先验为共轭先验分布 $N(\mu, \tau^2)$, 求 θ 的 Bayes 估计.

解 由例 7.2.10 可知 $\pi(\theta|\mathbf{x})$ 是正态分布 $N(\mu_n(\mathbf{x}), \eta_n^2)$, 其中

$$\mu_n(\mathbf{x}) = \frac{\sigma^2/n}{\sigma^2/n + \tau^2}\mu + \frac{\tau^2}{\sigma^2/n + \tau^2}\bar{x}, \quad \eta_n^2 = \frac{\sigma^2\tau^2}{n\tau^2 + \sigma^2}.$$

由于后验分布 $N(\mu_n(\mathbf{x}), \eta_n^2)$ 为对称分布, 故后验众数估计, 后验中位数估计和后验均值皆相同, 故 θ 的 Bayes 估计为

$$\hat{\theta}_B = \mu_n(\mathbf{x}) = \frac{\sigma^2/n}{\sigma^2/n + \tau^2}\mu + \frac{\tau^2}{\sigma^2/n + \tau^2}\bar{x}.$$

如例 7.2.10 所述, $\hat{\theta}_B$ 是 $\bar{x}$ 和 μ 加权平均, 是对样本信息和先验信息的综合.

区间估计

对于区间估计问题, Bayes 方法具有处理方便和含义清晰的优点. 而经典方法寻求的置信区间常受到批评.

当 θ 的后验分布 $\pi(\theta|x)$ 获得后, 若存在区间 $[a, b]$, 使 θ 落在 $[a, b]$ 内的后验概率为 $1 - \alpha$, 即

$$P(a \leq \theta \leq b|x) = 1 - \alpha,$$

则称 $[a, b]$ 为 θ 的 Bayes 区间估计, 又称为可信区间. 这是对 θ 的后验分布为连续情形, 若 θ 为离散型随机变量, 对给定概率 $1 - \alpha$, 上述的区间 $[a, b]$ 不一定存在, 而要将左边概率适当放大一点使 $P(a \leq \theta \leq b|x) \geq 1 - \alpha$, 这样的区间也是 θ 的 Bayes 可信区间. 定义如下:

区间估计

定义 7.3.2 (可信区间) 设参数 θ 的后验分布为 $\pi(\theta|x)$, 对给定的样本 x 和 $0 < \alpha < 1$ (通常 α 取较小的数), 若存在两个统计量 $\hat{\theta}_1 = \hat{\theta}_1(x)$ 和 $\hat{\theta}_2 = \hat{\theta}_2(x)$, 使得

$$P(\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2 | x) \geq 1 - \alpha,$$

则称 $[\hat{\theta}_1, \hat{\theta}_2]$ 为 θ 的可信水平为 $1 - \alpha$ 的 Bayes可信区间 (Bayesian credible interval). 而满足

$$P(\theta \geq \hat{\theta}_L | x) \geq 1 - \alpha$$

的 $\hat{\theta}_L = \hat{\theta}_L(x)$ 称为 θ 的可信水平为 $1 - \alpha$ 的可信下限. 而满足

$$P(\theta \leq \hat{\theta}_U | x) \geq 1 - \alpha$$

的 $\hat{\theta}_U = \hat{\theta}_U(x)$ 称为 θ 的可信水平为 $1 - \alpha$ 的可信上限.

区间估计

可信水平和可信区间与经典统计方法中的置信水平和置信区间虽是同类概念,但二者存在本质差别,主要表现为:

(1) 基于后验分布 $\pi(\theta|x)$, 在给定 x 和 $1-\alpha$ 后求得的可信区间, 如 $1-\alpha=0.90$ 的可信区间为 $[1.2, 2.0]$, 这时可以写为

$$P(1.2 \leq \theta \leq 2.0|x) = 0.90.$$

既可以说“ θ 属于这个区间的概率为 0.90”, 也可以说“ θ 落入这个区间的概率为 0.90”. 可对置信区间就不能这样说. 因为经典统计方法认为 θ 为常数, 它要么 θ 在 $[1.2, 2.0]$ 之内, 要么在其外, 不能说“ θ 落在 $[1.2, 2.0]$ 中的概率为 0.90”, 而只能说“100 次重复使用这个置信区间, 大约有 90 次能覆盖 θ ”. 这种频率解释对仅使用这个置信区间一次或两次的人来说是毫无意义的. 可见 Bayes 可信区间简单, 自然易被人们接受和理解. 事实上很多实际工作者把求得置信区间当可信区间去用.

区间估计

(2) 经典统计方法寻求置信区间有时是困难的, 要设法构造一个枢轴变量, 要求它的表达式与 θ 有关, 而它的分布与 θ 无关. 这是一项技术性很强的工作, 有时找抽样分布相当困难, 而寻求可信区间只需要利用后验分布, 不需要寻求另外的分布, 与经典统计方法相比要简单得多.

区间估计

例 7.3.7 设 $X = (X_1, \dots, X_n)$ 是从正态总体 $N(\theta, \sigma^2)$ 中抽取的简单样本, 其中 σ^2 已知. θ 的先验分布是 $N(\mu, \tau^2)$, 其中 μ, τ^2 已知. 求 θ 的可信水平 $1 - \alpha$ 的可信区间.

区间估计

解 由例 7.2.10 已知 θ 的后验分布是 $N(\mu_n(\mathbf{x}), \eta_n^2)$, 故有

$$P(\mu_n(\mathbf{x}) - \eta_n u_{\alpha/2} \leq \theta \leq \mu_n(\mathbf{x}) + \eta_n u_{\alpha/2} | \mathbf{x}) = 1 - \alpha,$$

其中 $u_{\alpha/2}$ 为 $N(0, 1)$ 的上侧 $\alpha/2$ 分位数. 因此 $[\mu_n(\mathbf{x}) - \eta_n u_{\alpha/2}, \mu_n(\mathbf{x}) + \eta_n u_{\alpha/2}]$ 就是 θ 的可信水平 $1 - \alpha$ 的可信区间.

看一个数值例子 (其中 $n = 1$). 在儿童智商测验的例子中, 设测验结果 $X \sim N(\theta, 100)$, 其中 θ 为被测验儿童智商 IQ 真值, θ 的先验分布是 $N(100, 225)$, 由前面结果可知后验分布 $\pi(\theta|x)$ 是 $N(\mu(x), \eta^2)$, 当该儿童测验得分 $x = 115$ 时, 有

$$\begin{aligned}\mu(x) &= \frac{\sigma^2}{\sigma^2 + \tau^2} \mu + \frac{\tau^2}{\sigma^2 + \tau^2} x = \frac{4}{13} \times 100 + \frac{9}{13} \times 115 = 110.38, \\ \eta^2 &= \frac{100 \times 225}{100 + 225} = 8.32^2.\end{aligned}$$

故 θ 的后验分布是 $N(110.38, 8.32^2)$, 因此求得 θ 的可信系数为 0.95 的可信区间为

$$[\mu(x) - 1.96\eta, \mu(x) + 1.96\eta] = [94.07, 126.69].$$

区间估计

在这个例子中, 若不用先验信息, 用经典方法 $X \sim N(\theta, 100)$, $x = 115$ 可求 θ 的置信系数 0.95 的置信区间为

$$[115 - 10 \times 1.96, 115 + 10 \times 1.96] = [95.4, 134.6].$$

这两个区间是不同的, 区间长度也不同. 可信区间长度短了一些, 这是由于利用了先验信息. 另一个不同是经典方法不能说“ θ 落入区间 $[95.4, 134.6]$ 中的概率为 0.95”, 也不能说“ θ 属于此区间的概率为 0.95”. 在这一束缚下这个区间还能有什么用? 这就是经典置信区间常受到批评的原因. 但仍有不少人在使用这个区间, 把它当可信区间去用.

假设检验

设检验问题仍如式 (5.1.1) 所述, 即

$$H_0 : \theta \in \Theta_0 \leftrightarrow H_1 : \theta \in \Theta_1.$$

经典的假设检验方法中, 对上述检验问题要通过犯 I, II 型错误的概率的大小来评价检验优劣.

Bayes 方法处理假设检验问题是直截了当的. 在求得 θ 的后验分布 $\pi(\theta|x)$ 后, 计算 Θ_0 和 Θ_1 的后验概率

$$\alpha_0 = P(\theta \in \Theta_0|x), \quad \alpha_1 = P(\theta \in \Theta_1|x).$$

比较 α_0 和 α_1 的大小决定接受 H_0 还是 H_1 . α_0 和 α_1 是综合样本信息和先验信息得出的两个假定实际发生的概率. 检验法则如下:

$$\text{当 } \frac{\alpha_0}{\alpha_1} > 1 \text{ 时接受 } H_0, \text{ 否则拒绝 } H_0. \quad (7.3.1)$$

假设检验

例 7.3.9 考虑对一个儿童作智力测验的情形. 设测验结果 $X \sim N(\theta, 100)$, 其中 θ 为这个孩子测验中的智商 IQ 真值, θ 的先验分布为 $N(100, 225)$. 该儿童测验得分 $x = 115$, 求下列检验问题:

$$H_0 : \theta \leq 100 \leftrightarrow H_1 : \theta > 100.$$

解 由例 7.3.7 可知 θ 的后验分布为 $N(110.38, 8.32^2)$, 查标准正态分布表求得

$$\alpha_0 = P(\theta \leq 100|x) = 0.106, \quad \alpha_1 = P(\theta > 100|x) = 0.894.$$

后验机会比 $\alpha_0/\alpha_1 = 1/8.44$. 这一比值远小于 1, 故否定 H_0 .

目 录

6.1 Bayes统计

6.2 先验分布的确定

6.3 Bayes统计推断

6.4 统计判决理论

6.5 Minimax准则

决策问题

例 7.4.1 一位投资者有一笔资金要进行投资, 有如下几个投资方案供选择:

a_1 : 购买股票, 根据市场情况可净赚 5000 元, 但也可能亏损 10000 元.

a_2 : 购买基金, 根据市场情况可净赚 3000 元, 但也可能亏损 8000 元.

a_3 : 存入银行, 不管市场情况如何, 总可净赚 1000 元.

他应如何决策?

这位投资者在与金融市场博弈. 未来的金融市场也有两种情况: 看涨 (θ_1) 与看跌 (θ_2). 根据上述情况, 可写出投资者的收益矩阵如下:

表 7.4.1

行动	a_1	a_2	a_3
θ_1	5000	3000	1000
θ_2	-10000	-8000	1000

统计决策三要素

(1) 样本空间和样本分布族.

取值于样本空间 $\mathcal{X}$ 内的随机变量 X 及其分布族是 $\{f(x, \theta), \theta \in \Theta\}$ 构成统计决策问题的第一个要素, 其中 $f(x, \theta)$ 是 X 的概率函数, θ 是未知参数, Θ 为参数空间, $\mathbf{X} = (X_1, \dots, X_n)$ 是总体 X 的简单样本.

(2) 行动空间.

决策者或统计工作者对某个统计决策问题可能采取的行动所构成的非空集合, 被称为行动空间 (action space), 记为 $\mathcal{A}$. 在估计问题中, $\mathcal{A}$ 由一切估计量 $\delta(x)$ 构成, 常取 $\mathcal{A} = \Theta$. 在检验问题中 $\mathcal{A}$ 只有两个行动组成, 即 $\mathcal{A} = \{a_0, a_1\}$, 其中 a_0 表示接受原假设 H_0 ; a_1 表示拒绝 H_0 , 接受 H_1 .

(3) 损失函数.

损失函数 (loss function) 是定义于 $\Theta \times \mathcal{A}$ 上的非负函数, 记为 $L(a, \theta)$. 它表示参数为 $\theta \in \Theta$ 时, 采取行动 $a \in \mathcal{A}$ 所蒙受的损失. 损失函数的类型很多, 常用的有“平方损失”、“绝对值损失”和“线性损失”等.

Bayes统计决策第四要素

对 Bayes 统计决策问题, 除上式三要素外还应增加第四个要素:

(4) 先验分布.

定义在参数空间 Θ 上的先验分布函数 $F^\pi(\theta)$ 或概率函数 $\pi(\theta)$.

决策函数和风险函数

定义 7.4.1 定义于样本空间 $\mathcal{X}$ 内而取值于行动空间 $\mathcal{A}$ 内的函数 $\delta = \delta(\mathbf{x})$ 称为决策函数或判决函数 (decision rules).

设 δ 是采取的决策行动, 若参数为 θ , 则由此造成的损失是 $L(\delta, \theta)$, 这个量与样本 $\mathbf{X}$ 有关, 因而是随机的. 故采取行动 δ 的效果用平均损失去度量是相对合理的. 这就引入如下风险函数的概念.

定义 7.4.2 设 δ 是一个决策函数, 称平均损失

$$R(\delta, \theta) = E[L(\delta(\mathbf{X}), \theta)] = \int_{\mathcal{X}} L(\delta(\mathbf{x}), \theta) f(\mathbf{x}, \theta) d\mathbf{x} \quad (7.4.1)$$

为 δ 的风险函数 (risk function).

一致最优决策函数

按照 Wald 的统计决策理论, 评价一个决策函数的唯一依据, 就是其风险函数. 风险函数越小越好. 有了风险函数后可比较不同决策函数的优劣.

定义 7.4.3 (一致最优决策函数) 记 $R(\delta, \theta)$ 为决策函数 δ 的风险函数. 若存在一个决策函数 δ^* , 使得对任一决策函数 δ 有

$$R(\delta^*, \theta) \leq R(\delta, \theta), \quad \text{一切 } \theta \in \Theta, \quad (7.4.2)$$

则称 δ^* 为一致最优解 或一致最优决策函数.

注 7.4.1 若一致最优解 (或一致最优决策函数) 存在, 则毫无疑问应当采用它. 但是除了某些例外情形, 一致最优解通常不存在. 因此必须把标准放宽些, 引进一些比一致最优准则更弱的优良性准则.

Bayes准则

前一段已说过, 一致最优的决策函数通常不存在, 故有必要把标准放宽些, 引进一些条件更弱的优良性准则. 下列定义的 Bayes 风险, 就是制定一种优良性的综合指标, 作为比较决策函数的标准.

定义 7.4.4 设 $R(\delta, \theta)$ 为 δ 的风险函数, $H(\theta)$ 为 θ 的先验分布 (若存在密度, 用 $\pi(\theta)$ 表示), 则称

$$R_H(\delta) = E^\theta[R(\delta, \theta)] = \int_{\Theta} R(\delta, \theta) dH(\theta) \quad (7.4.3)$$

为 δ 的 Bayes 风险 (Bayes risk).

由定义可见 Bayes 风险是将风险函数以 θ 的先验分布 $H(\theta)$ 为权的一种加权平均. 使 Bayes 风险达到最小的决策函数称为决策问题的 Bayes 解, 其定义如下.

定义 7.4.5 设 δ_1 和 δ_2 为 θ 的两个决策函数, 若 $R_H(\delta_1) \leq R_H(\delta_2)$, 则称 δ_1 在 Bayes 风险下优于 δ_2 . 若存在 δ^* , 使得对任一决策函数 δ 有

$$R_H(\delta^*) \leq R_H(\delta), \quad (7.4.4)$$

则称 δ^* 为所考虑的统计决策问题的 Bayes 解.

一般损失下的Bayes估计

1. 平方损失下的 Bayes 估计

定理 7.4.2 设 $a = a(x)$ 为一决策函数, 则在平方损失 $L(a, \theta) = (a - \theta)^2$ 下, θ 的 Bayes 估计为其后验期望. 即

$$\hat{\theta}_B(x) = E(\theta|x). \quad (7.4.8)$$

一般损失下的Bayes估计

例 7.4.2 设 $X \sim N(\theta, \sigma^2)$, 其中 σ^2 已知, θ 的先验分布是 $N(\mu, \tau^2)$. 求平方损失下 θ 的 Bayes 估计.

解 由例 7.2.10 可知 θ 的后验分布是 $N(\mu(x), \eta^2)$, 其中

$$\mu(x) = \frac{\sigma^2}{\sigma^2 + \tau^2} \mu + \frac{\tau^2}{\sigma^2 + \tau^2} x, \quad \eta^2 = \frac{\sigma^2 \tau^2}{\sigma^2 + \tau^2}.$$

由定理 7.4.2 可知, 平方损失下 θ 的 Bayes 估计是

$$\hat{\theta}_B(x) = E(\theta|x) = \mu(x) = \frac{\sigma^2}{\sigma^2 + \tau^2} \mu + \frac{\tau^2}{\sigma^2 + \tau^2} x.$$

一般损失下的Bayes估计

2. 加权平方损失下的 Bayes 估计

定理 7.4.3 设 $a = a(x)$ 为一决策函数, 则在加权平方损失 $L(a, \theta) = w(\theta)(a - \theta)^2$ 下, θ 的 Bayes 估计为

$$\hat{\theta}_B = \frac{E(\theta w(\theta)|x)}{E(w(\theta)|x)}, \quad (7.4.9)$$

其中 $w(\theta)$ 为参数空间上的正值实函数.

一般损失下的Bayes估计

3. 绝对值损失下的 Bayes 解

定理 7.4.4 设 $\delta = \delta(x)$ 为一决策函数, 则在绝对损失 $L(\delta, \theta) = |\delta - \theta|$ 下, θ 的 Bayes 估计为后验中位数.

目 录

6.1 Bayes统计

6.2 先验分布的确定

6.3 Bayes统计推断

6.4 统计判决理论

6.5 Minimax准则

Minimax准则

7.4 节已经说过, 一致最优的决策函数通常不存在, 因此必须把标准放宽些, 引进一些比一致最优准则更弱的优良性准则. 途径之一是用某种方法制定优良性的综合指标, 以之作为比较的标准. Bayes 准则属于这一类. 和 Bayes 风险 $R_{\pi}(\delta)$ 一样, 下列定义的最大风险 $M(\delta)$ 也是一种优良性的综合指标, 用它作为比较决策函数的标准, 称为 Minimax 准则. 因此 Minimax 准则是从综合指标考虑的另一种优良性准则.

设 δ 为一决策函数, $R(\delta, \theta)$ 为其风险函数, 令

$$M(\delta) = \sup_{\theta \in \Theta} R(\delta, \theta).$$

易见 $M(\delta)$ 表示采用决策函数 δ 时所遭受的最大风险. 如果在某项应用中, 使这个最大风险尽可能小是很重要的, 就可以制定如下准则, 通常称为 Minimax 准则或极小极大准则.

Minimax准则

定义 7.5.1 设 δ_1 和 δ_2 为同一个统计决策问题的两个决策函数, 若 $M(\delta_1) < M(\delta_2)$, 则称决策函数 δ_1 优于 δ_2 . 如果存在某个决策函数 δ^* , 对任何决策函数 δ 都有

$$M(\delta^*) \leq M(\delta),$$

则称 δ^* 为该统计决策问题的 Minimax 解, 也称 δ^* 为 Minimax 决策函数. 当统计决策问题为估计或检验时, 也称 δ^* 为 Minimax 估计或 Minimax 检验.

Minimax解的求法

求 Minimax 解通常比较困难, 下列两个定理与其说是求 Minimax 解的方法, 不如说是验证某一特定的解为 Minimax 解的方法.

定理 7.5.1 设 δ^* 为在先验分布 $H(\theta)$ 之下的 Bayes 解, 且 δ^* 的风险函数为常数 c , 即对任何 $\theta \in \Theta$ 有 $R(\delta^*, \theta) = c$, 则 δ^* 为一个 Minimax 解.

Minimax解的求法

定理 7.5.2 设 δ_k 为一个统计决策问题在先验分布 H_k 之下的 Bayes 解, 假定 δ_k 的 Bayes 风险为 r_k , $k = 1, 2, \dots$, 且有

$$\lim_{k \rightarrow \infty} r_k = r < \infty. \quad (7.5.2)$$

又设 δ^* 为同一问题的一个决策函数, 满足条件

$$M(\delta^*) \leq r, \quad (7.5.3)$$

则 δ^* 为此统计决策问题的 Minimax 解.

Minimax解的求法

例 7.5.2 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从正态总体 $N(\theta, 1)$ 中抽取的简单样本, 取损失函数为 $L(d, \theta) = (d - \theta)^2$, 求 θ 的 Minimax 估计.

Minimax解的求法

解 找一串 θ 的先验分布 $\{H_k\}$, H_k 为 $N(0, k^2)$, $k = 1, 2, \dots$. 由例 7.2.10 可知 θ 的后验分布是 $N(\mu_k(\mathbf{x}), \eta^2)$. 再由定理 7.4.2 可知, 平方损失下的 Bayes 解为后验分布的均值, 即

$$\delta_k(\mathbf{x}) = \mu_k(\mathbf{x}) = \frac{nk^2}{nk^2 + 1} \bar{x},$$

其风险函数

$$\begin{aligned} R(\delta_k, \theta) &= E \left(\frac{nk^2 \bar{X}}{nk^2 + 1} - \theta \right)^2 \\ &= D \left(\frac{nk^2 \bar{X}}{nk^2 + 1} \right) + \left(\frac{nk^2}{nk^2 + 1} E(\bar{X}) - \theta \right)^2 \\ &= \frac{nk^4}{(nk^2 + 1)^2} + \frac{\theta^2}{(nk^2 + 1)^2} = \frac{nk^4 + \theta^2}{(nk^2 + 1)^2}, \end{aligned}$$

Minimax解的求法

其中均值 E 和方差 D 是关于“给定 θ 时 X 的分布”计算的. 故 δ_k 的 Bayes 风险是

$$r_k = R_{H_k}(\delta_k) = E^\theta \left[\frac{nk^4 + \theta^2}{(nk^2 + 1)^2} \right] = \frac{nk^4 + k^2}{(nk^2 + 1)^2} = \frac{k^2}{nk^2 + 1},$$

显然有

$$\lim_{k \rightarrow \infty} r_k = \lim_{k \rightarrow \infty} \frac{k^2}{nk^2 + 1} = \frac{1}{n} = r.$$

若取 $\delta^* = \bar{X}$, 其风险函数 $R(\delta^*, \theta) = E(\bar{X} - \theta)^2 = 1/n \leq r$, 故由定理 7.5.2 可知, $\delta^* = \bar{X}$ 为 θ 的 Minimax 估计.

我们在这里

□ 问题答疑：<http://www.xxwenda.com/>

■ 可邀请老师或者其他回答问题

联系我们

小象学院：互联网新技术在线教育领航者

- 微信公众号：小象
- 新浪微博：ChinaHadoop

