

统计决策理论与贝叶斯分析

引言

两个学派：

- 频率学派：

建立在**测度论**基础上的**公理化**体系

基本观点：数据看成是一定概率分布的总体，研究对象是这个**总体**而不局限于数据本身

依据：总体信息（模型）和样本信息（数据）

- 贝叶斯学派

起源于贝叶斯的两项工作

- 贝叶斯定理
- 贝叶斯假设

基本观点：任何一个未知量 θ 都被看作为随机变量，应用一个概率分布去描述对 θ 的未知情况。在抽样前就具有先验分布

依据：总体信息（模型）和样本信息（数据）和先验信息

两个学派之间的差异：

1. 对概率的概念理解有差别：

- 频率学派：坚持概率的频率解释
- 贝叶斯学：派赞成主观概率

2. 统计推断理念的差异：

- 频率学派：选定模型、确定统计量、决定统计量的分布
- 贝叶斯学派：先验信息+样本信息->后验信息

两个学派之间的批评：

- 对贝叶斯学派的批评：

1. 参数看成是随机变量是否妥当？
2. 先验分布是否存在？如何选取？

- 对频率学派的批评：

1. 问题的提法不妥当
2. 判断统计方法好坏的标准不妥当

贝叶斯假设：

无信息先验分布（Non-informative Priors）应该选取在 θ **取值范围内的均匀分布**

现代贝叶斯统计理论研究现状

1. 先验分布理论的研究

可以分为两类

- 无信息先验分布

一般采用取值范围内的均匀分布，贝叶斯假设

- 共轭先验分布

如果先验分布和似然函数可以使得先验分布和后验分布有相同的形式，则先验分布和似然函数是共轭的

对于正态分布 $N(\mu, \sigma^2)$ 而言，

1. μ 已知， σ 未知， σ 的共轭分布族为逆伽马分布
2. μ 未知， σ 未知，参数的共轭分布族为正太-逆伽马分布

2. 后验分布的统计推断

- 蒙特卡洛方法
- 利用共轭先验分布的特性

先验分布和后验分布

统计推断中可用的三种信息

1. 总体信息
2. 样本信息
3. 先验信息

贝叶斯公式的三种形式

1. 事件形式：

$$P(A_i|B) = \frac{P(A_i, B)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^k P(B|A_j)P(A_j)}$$

2. 密度函数形式：

- 先验分布 $\pi(\theta)$
- 后验分布 $h(x_1, \dots, x_n, \theta) = p(x_1, \dots, x_n|\theta)\pi(\theta)$

$$\pi(\theta|x_1, \dots, x_n) = \frac{h(x_1, \dots, x_n, \theta)}{m(x_1, \dots, x_n)} = \frac{p(x_1, \dots, x_n|\theta)\pi(\theta)}{\int_{\Theta} p(x_1, \dots, x_n|\theta)\pi(\theta)d\theta}$$

3. 离散形式：

$$\pi(\theta_i|x) = \frac{p(x|\theta_i)\pi(\theta_i)}{\sum_j p(x|\theta_j)\pi(\theta_j)}$$

后验分布是三种信息的综合

例1.4. 设事件A发生的概率为 θ , 即 $P(A)=\theta$. 为了估计 θ 而作 n 次独立观察, 其中事件A出现的次数为 X , 则有 X 服从二项分布 $b(n, \theta)$, 即 $P(X=x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$, $x=0, 1, \dots, n$ 如何求 θ 的后验分布?

A的发生概率 $\theta \in [0, 1]$. 根据贝叶斯假设, θ 为 $[0, 1]$ 上的均匀分布

$$\pi(\theta) = \begin{cases} 1, & 0 < \theta < 1 \\ 0, & \text{others} \end{cases} \quad h(x, \theta) = P(X=x|\theta)\pi(\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x=0, 1, \dots, n, 0 < \theta < 1$$

$$\pi(\theta|X) = \frac{h(x, \theta)}{m(x)}$$

$$m(x) = \int_0^1 \binom{n}{x} \theta^x (1-\theta)^{n-x} d\theta \quad x=0, 1, \dots, n$$

$$= \binom{n}{x} \int_0^1 \theta^x (1-\theta)^{n-x} d\theta = \binom{n}{x} \frac{\Gamma(x+1) \Gamma(n-x+1)}{\Gamma(n+2)} \quad x=0, 1, \dots, n$$

$$= \frac{\Gamma(n+2)}{\Gamma(n+1) \Gamma(n-x+1)} \theta^x (1-\theta)^{n-x}, \quad 0 < \theta < 1$$

具体例子:

男婴为251527个, 女婴为241945个, $U(0, 1)$ 的均匀分布, 其中 $X=251527$, $n=493472$

$$P(\theta \leq 0.5 | X) = \frac{\Gamma(n+2)}{\Gamma(n+1) \Gamma(n-X+1)} \int_0^{0.5} \theta^X (1-\theta)^{n-X} d\theta = 1.15 \times 10^{-42}$$

故他断言男婴出生的概率大于0.5

伽马分布和贝塔分布:

- 伽马分布

$$\Gamma(s) = \int_0^\infty x^{s-1} e^{-x} dx, \quad s > 0$$

$$\Gamma(s+1) = s\Gamma(s), \quad \Gamma(n+1) = n!$$

- 贝塔分布

$$B(p, q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx, \quad p > 0, q > 0$$

$$B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$$

$$\beta_I(p, q) = \beta_e(p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \theta^{p-1} (1-\theta)^{q-1}$$

特例:

- $p = q = 1$ 时, $\beta_I(1, 1)$ 为区间 $[0, 1]$ 上的均匀分布
- $p = q = 1/2$ 时, $\beta_I(1/2, 1/2)$ 为反余弦函数

投资决策问题:

为了提高某产品的质量，公司经理考虑增加投资来改进生产设备，预计需要投资90万元，但从投资效果来看，下属部门有两种意见：

- θ_1 ：改进生成设备后，高质量产品可占比90%，可信度40%
- θ_2 ：改进生成设备后，高质量产品可占比70%，可信度60%

加了一个实验A：试制了5个产品，全为高质量产品

实验信息
投资决策： $\pi(\theta_1) = 40\% = 0.4$ $\pi(\theta_2) = 60\% = 0.6$

样本信息 A，求 $\pi(\theta_1|A)$ $\pi(\theta_2|A)$

$$\pi(\theta_1|A) = \frac{h(A|\theta_1)\pi(\theta_1)}{m(A)} = \frac{P(A|\theta_1)\pi(\theta_1)}{P(A|\theta_1)\pi(\theta_1) + P(A|\theta_2)\pi(\theta_2)}$$

$$= \frac{0.9^5 \cdot 0.4}{0.9^5 \cdot 0.4 + 0.7^5 \cdot 0.6} = \frac{0.231}{0.337} = 0.70$$

$\pi(\theta_2|A) = 0.300$

又增加了一个实验B：试制了10个产品，9个高质量产品

$$\pi(\theta_1) = 0.7 \quad \pi(\theta_2) = 0.3$$

$$\pi(\theta_1|B) = \frac{h(B|\theta_1)\pi(\theta_1)}{P(B|\theta_1)\pi(\theta_1) + P(B|\theta_2)\pi(\theta_2)} = \frac{(0.9)^9 \cdot 0.7}{(0.9)^9 \cdot 0.7 + (0.7)^9 \cdot 0.3}$$

$$= \frac{0.027}{0.027 + 0.036} = 0.883$$

$\pi(\theta_2|B) = 0.117$

\therefore 选 θ_1 (高质量 90% 以上) 概率 0.883

问：如何决策

共轭先验分布

共轭先验分布是对某一分布中的参数而言的

θ 是总体分布的参数， $\pi(\theta)$ 是先验密度函数，假设由样本信息得到后验密度函数与 $\pi(\theta)$ 具有相同的形式，则称 $\pi(\theta)$ 是 θ 的共轭先验分布

正态均值的共轭先验分布是正态分布

某一正态分布 $N(\theta, \sigma^2)$ ，其均值 θ 服从一个正态分布 $\pi(\theta) = N(\mu, \tau)$ (先验分布)，则后验分布也服从一个正态分布 $\pi(\theta|x) = N(\mu_1, \tau_1)$

$$\begin{aligned} \mu_1 &= \frac{\hat{x}\sigma_0^{-2} + \mu\tau^{-2}}{\sigma_0^{-2} + \tau^{-2}} \\ \tau_1^{-2} &= \sigma_0^{-2} + \tau^{-2} \\ \sigma_0^2 &= \frac{\sigma^2}{n} \end{aligned}$$

例1.6: 证明正态均值 (方差已知) 的共轭先验分布是正态分布

正态分布的PDF = $\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\theta)^2}{2\sigma^2}}$

例1.6. 证明正态均值 (方差已知) 的共轭先验分布是正态分布

证明: 设 x_1, \dots, x_n 是来自正态分布 $N(\theta, \sigma^2)$ 的一组样本观测值, $\frac{(x-\mu)^2}{2\sigma^2}$

① 其中 σ^2 已知, 则似然函数为:

正态 PDF = $\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$$P(x|\theta) = \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\theta)^2}{2\sigma^2}} \right)^n$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2 \right\} \quad \text{①} \quad -\infty < x_1, \dots, x_n < +\infty$$

② 确定先验分布, 取另一个正态分布 $N(\mu, \tau^2)$ 作为正态均值 θ 的先验分布

$\pi(\theta) = \frac{1}{\sqrt{2\pi}\tau} e^{-\frac{(\theta-\mu)^2}{2\tau^2}}, -\infty < \theta < +\infty$, 其中 μ 与 τ^2 已知 ②

③ 计算后验分布 根据①②

$$\pi(\theta|x) = \frac{h(x, \theta)}{m(x)}$$

$$= \frac{P(x|\theta)\pi(\theta)}{\int_{-\infty}^{+\infty} P(x|\theta)\pi(\theta)d\theta}$$

$$= \left(\frac{2\pi}{A} \right)^{-\frac{1}{2}} \exp \left\{ -\frac{(\theta-B/A)^2}{2/A} \right\}$$

根据①② $h(x, \theta) =$

$$\frac{(2\pi)^{-n/2}}{\sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \left[n\theta^2 - 2n\theta\bar{x} + \sum_{i=1}^n x_i^2 \right] - \frac{(\theta-\mu)^2}{2\tau^2} \right\}$$

$$= K_1 \exp \left\{ -\frac{1}{2} [A\theta^2 - 2B\theta + C] \right\} = K_2 \exp \left\{ -\frac{(A\theta - B/A)^2}{2/A} \right\} \quad \text{③}$$

其中 $A = \frac{1}{\sigma_0^2} + \frac{1}{\tau^2}$, $\sigma_0^2 = \frac{\sigma^2}{n}$, $B = \frac{\bar{x}}{\sigma_0^2} + \frac{\mu}{\tau^2}$, $C = \frac{1}{\sigma_0^2} \sum_{i=1}^n x_i^2 + \frac{\mu^2}{\tau^2}$

$$K_2 = K_1 \exp \left\{ -\frac{1}{2} (C - B^2/A) \right\}$$

$$\therefore m(x) = \int_{-\infty}^{+\infty} h(x, \theta) d\theta = K_2 \left(\frac{2\pi}{A} \right)^{-\frac{1}{2}} \quad \text{④}$$

$$\theta \text{ 的后验分布为 } \pi(\theta|x) = \left(\frac{2\pi}{A}\right)^{-\frac{1}{2}} \exp\left(-\frac{(\theta-\mu_1)^2}{2A}\right)$$

这是一个正态分布！

$$\text{均值 } \mu_1 = B/A = \frac{\bar{x}\sigma_0^{-2} + M\tau^{-2}}{\sigma_0^{-2} + \tau^{-2}}, \quad \text{方差 } \tau_1^2: \frac{1}{\tau_1^2} = \frac{1}{\sigma_0^2} + \frac{1}{\tau^2}$$

∴ 正态分布的均值 (方差已知) 的共轭先验分布是正态分布。

例如 $X \sim N(\theta, \sigma^2)$, $\theta \sim N(\mu, \tau^2)$

例如 $X \sim N(\theta, \sigma^2)$, $\theta \sim N(\mu, \tau^2)$, 则 $\pi(\theta|x)$ 是正态分布

$$\pi(\theta|x) \sim N(\mu_1, \tau_1^2)$$

$$\mu_1 = \frac{\bar{x}\sigma_0^{-2} + M\tau^{-2}}{\sigma_0^{-2} + \tau^{-2}} \quad \sigma_0^2 = \frac{\sigma^2}{n}$$

$$\frac{1}{\tau_1^2} = \frac{1}{\sigma_0^2} + \frac{1}{\tau^2} \quad \sigma_0 = \frac{\sigma^2}{n}$$

例) $X \sim N(\theta, \sigma^2)$, $\theta \sim N(10, 3^2)$ X 有 25 个, $\bar{x} = 12.1$

$$\mu_1 = \frac{\bar{x}\sigma_0^{-2} + M\tau^{-2}}{\sigma_0^{-2} + \tau^{-2}} = \frac{12.1 \cdot \frac{25}{16} + \frac{10}{9}}{\frac{25}{16} + \frac{1}{9}} = \frac{288.25}{241} = 11.96$$

$$\frac{1}{\tau_1^2} = \frac{1}{\sigma_0^2} + \frac{1}{\tau^2} = \frac{25}{16} + \frac{1}{9} = \frac{49}{36}$$

$$\tau_1 = \frac{7}{6}$$

$$\therefore \pi(\theta|x) \sim N(11.96, \frac{7}{6})$$

补充例题:

设 X 表示人的胸围, 根据经验, 胸围是近似服从正态分布的。现测量了 $n=10000$ 个人的胸围, 得到样本均值为 39.8cm , 样本方差为 4 , 假设 θ 的先验分布为 $N(38, 9)$, 求 θ 的后验分布

$$\begin{aligned}
 X &\sim N(\theta, \frac{4}{N}) & \theta &\sim \pi(\theta) = N(38, 9) & \bar{X} &= 39.8 & N &= 10000 \\
 \sigma_0^2 &= \frac{4}{10000} & \mu_1 &= \frac{39.8 \cdot \frac{10000}{4} + 38 \cdot \frac{1}{9}}{\frac{10000}{4} + \frac{1}{9}} & &= \cancel{39.8} 39.79992 \\
 & & & \approx 39.8 \\
 \frac{1}{\sigma_1^2} &= \frac{1}{\sigma_0^2} + \frac{1}{\sigma^2} = \frac{1}{9} + \frac{1}{\frac{4}{10000}} \\
 & & &= 112500
 \end{aligned}$$

简化后验分布的计算

省略常数因子

$$\pi(\theta|x) = \frac{h(x, \theta)}{m(x)} = \frac{p(x|\theta)\pi(\theta)}{m(x)}$$

$m(x)$ 不依赖于 θ ，作为常数因子，把他省略

$$\pi(\theta|x) = p(x|\theta)\pi(\theta)$$

右端为后验分布的核

简化例1.3的计算，得到后验分布的均值为 B/A ，方差为 A^{-1}

二项分布的成功概率 θ 的共轭先验分布是贝塔分布

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

证明：

二项分布的成功概率 θ 的共轭先验分布是贝塔分布。

证明 $X \sim b(n, \theta)$, 则 $b(n, \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \propto \theta^x (1-\theta)^{n-x}$

设 θ 的先验分布为贝塔分布, $Be(\alpha, \beta) \propto \theta^{\alpha-1} (1-\theta)^{\beta-1}$

$$\pi(\theta|x) \propto \theta^x (1-\theta)^{n-x} \cdot \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

$$= \frac{\theta^{\alpha+x-1} (1-\theta)^{n+\beta-x-1}}{0 < \theta < 1}$$

贝塔分布的核

$$\pi(\theta|x) = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)} \theta^{\alpha+x-1} (1-\theta)^{n+\beta-x-1} \quad 0 < \theta < 1$$

共轭先验分布的优缺点

优点:

1. 计算方便

2. 后验分布的一些参数可以得到很好的解释:

◦ 正态均值的共轭先验分布是正态分布

■ 均值

$$\sigma_0^2 = \sigma^2 / n$$

$$\mu_1 = \frac{\hat{x}\sigma_0^{-2} + \mu\tau^{-2}}{\sigma_0^{-2} + \tau^{-2}} = \gamma\hat{x} + (1-\gamma)\mu$$

γ 是用方差倒数组成的权

所以后验均值 μ_1 是样本均值和先验均值的加权平均

■ 方差

$$\frac{1}{\tau_1^2} = \frac{1}{\sigma_0^2} + \frac{1}{\tau^2} = \frac{n}{\sigma_0^2} + \frac{1}{\tau^2}$$

后验分布的精度是样本均值分布的精度与先验分布精度之和

增加样本量 n 或减少先验分布方差都有利于提高后验分布的精度

◦ 二项分布成功概率 θ 的共轭先验分布是贝塔分布

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

■ 均值

后验均值介于样本均值和先验均值之间

■ 方差

样本增大，方差越来越小

缺点：如何找到合适的先验分布

常用的一些共轭先验分布

共轭先验分布选取的一般原则：

由似然函数 $L(\theta) = p(x|\theta)$ 中所含的因式所决定的，即选与似然函数具有**相同核的分布**作为先验分布

例1.10

正态方差的共轭先验分布

为倒伽马分布

共轭先验分布

- 正态分布均值

正态分布 $N(\mu, \tau^2)$

后验分布期望 $\mu_1 = \frac{\hat{x}\sigma_0^2 + \mu\tau^2}{\sigma_0^2 + \tau^2}$

后验分布的方差 $\frac{1}{\tau_1^2} = \frac{1}{\sigma_0^2} + \frac{1}{\tau^2}$

- 正态分布方差

倒伽马分布 $IGa(a, b)$

- 二项分布成功的概率

贝塔分布

后验分布的期望 $\frac{a+x}{a+b+n}$

- Poisson分布均值

伽马分布

$$\frac{a+x}{b+1}$$

- 指数分布均值的倒数

伽马分布

超参数及其确定

先验分布中所含的未知参数被称为超参数

确定超参数的方法

1. 先验矩
2. 先验分位数
3. 先验矩和先验分位数
4. 其他方法

例1.4.1:

二项分布中成功概率 θ 的共轭先验分布式贝塔分布 $Be(\alpha, \beta)$, 具有两个超参数

1. 先验数据

先取得 θ 的均值与方差

$$\bar{\theta} = \frac{1}{K} \sum_{i=1}^K \theta_i, \quad S_{\theta}^2 = \frac{1}{K-1} \sum_{i=1}^K (\theta_i - \bar{\theta})^2$$

令其分别等于贝塔分布 $Be(\alpha, \beta)$ 的期望与方差

$$\begin{cases} \frac{\alpha}{\alpha + \beta} = \bar{\theta} \\ \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = S_{\theta}^2 \end{cases} \Rightarrow \begin{aligned} \hat{\alpha} &= \bar{\theta} \left(\frac{(1 - \bar{\theta})\bar{\theta}}{S_{\theta}^2} - 1 \right) \\ \hat{\beta} &= (1 - \bar{\theta}) \left(\frac{(1 - \bar{\theta})\bar{\theta}}{S_{\theta}^2} - 1 \right) \end{aligned}$$

2. 先验分布函数:

$$\begin{aligned} \int_0^1 \frac{r(\alpha + \beta)}{r(\alpha)r(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} d\theta &= 0.75 \\ \int_0^1 \frac{r(\alpha + \beta)}{r(\alpha)r(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} d\theta &= 0.25 \end{aligned} \Rightarrow$$

3. 先验数据 + 先验分布函数:

$$\begin{cases} \frac{\alpha}{\alpha + \beta} = \bar{\theta} \\ \int_0^p \frac{r(\alpha + \beta)}{r(\alpha)r(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} d\theta = p \end{cases} \Rightarrow$$

多参数模型

求某一个参数的后验分布的基本思想：

先根据先验信息给出参数的先验分布，然后按贝叶斯公式算得后验分布，即：

1. 总体的密度函数 $p(x|\theta)$
2. 先验密度 $\pi(\theta)$
3. $\pi(\theta|x) = p(x|\theta)\pi(\theta)$

但是有多个参数的情况？

充分统计量

- 充分性：**不损失信息的统计量**
- 定义：设 $x = (x_1, \dots, x_n)$ 是来自于分布函数 $F(x|\theta)$ 的一个样本， $T = T(x)$ 是统计量，假设在给定 $T(x) = t$ 的条件下， x 的条件分布与 θ 无关的话，则称该统计量为 θ 的充分统计量。
- 特性：当得到充分统计量 T 的某个取值 t 之后，**而失去原样本的观察值也没有关系。**
- 充要条件：

因子分解定理

存在一个 t 与 θ 的函数 $g(t, \theta)$ 和一个样本 x 的函数 $h(x)$ ，使得对任一样本 x 和任意 θ ，样本的联合密度 $p(x|\theta)$ 可表示为它们的乘积，即 $p(x|\theta) = g(T(x), \theta)h(x)$

样本分布可以被拆开为两个因子的乘积：一个是和 θ 无关仅和 x 有关；另一个是可以与 θ 有关，但与 x 的关系仅仅通过充分统计量 $T(x)$ 表现出来

贝叶斯与充分统计量

设 $x = (x_1, \dots, x_n)$ 是来自于密度函数 $p(x|\theta)$ 的一个样本， $T = T(x)$ 是统计量，它的密度函数为 $p(t|\theta)$ ，又设 $H = \pi(\theta)$ 是它的某个先验分布族，则 $T(x)$ 为 θ 的充分统计量的充要条件是对任意一个先验分布，有

$$\pi(\theta|T(x)) = \pi(\theta|x)$$

即用样本分布计算得到的后验分布和用统计量计算的后验分布相同的

条件方法

未知参数的后验分布：总体、样本和先验

条件方法的基本思想：基于后验分布的统计推断实际上只考虑了**已出现的数据而认为未出现的数据与推断无关**

和频率方法的区别：

频率采用的是对样本空间中所有可能出现的样本而求的

贝叶斯估计

后验密度 $\pi(\theta|x)$

- 到达最大值 θ_{MD} : 最大后验估计
- 中位数 $\hat{\theta}_{Me}$: 后验中位数估计
- 期望值: 后验期望值估计

这三个估计都称为贝叶斯估计 $\hat{\theta}_B$