第 2 讲：贝叶斯推断简介

张伟平

目录

de Finetti

贝叶斯方法起源于 Bayes 和 Laplace, 现代理论的大部分工作是 de Finetti 在 20 世纪三十年代建立的。贝叶斯方法的基本观点由贝叶斯公式引申而来.

■ **例 1.1.1**(P2) 一种诊断某癌症的试剂, 经临床试验有如下记录: 癌症病人试验结果是阳性的概率为 95%, 非癌症病人试验结果是阴性的概率为 95%. 现用这种试剂在某社区进行癌症普查, 设该社区癌症发病率为 0.5%, 问某人反应为阳性时该如何判断他是否患有癌症.

解: 设 A 表示“反应为阳性”的事件, B 表示“被诊断者患癌症”的事件, 则 $B_1 = B$ 和 $B_2 = \bar{B}$ 构成完备事件群. 由题意

$$P(A|B_1) = 0.95, \quad P(A|B_2) = 1 - 0.95 = 0.05,$$
$$P(B_1) = 0.005, \quad P(B_2) = 0.995.$$

现在要算的是 $P(B_1|A)$ 和 $P(B_2|A)$. 由贝叶斯公式易得

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)} = 0.087 = 8.7\%$$
$$P(B_2|A) = 1 - 0.087 = 0.913 = 91.3\%.$$

1.1 贝叶斯推断的基本概念

- 首先, 贝叶斯推断中所有变量都是随机变量. 贝叶斯推断依赖于主观概率. 每个人对给定事件所赋予的主观概率可以是不同的.

- 其次, 假设参数 θ 的分布为 $\pi(\theta)$(称为先验分布), 则在有了样本 $\mathbf{x}$ 后, 我们对 θ 的信念 (称为后验分布) 可以通过贝叶斯公式来更新:

$$\begin{aligned}\pi(\theta|\mathbf{x}) &= \frac{f(\mathbf{x}|\theta)\pi(\theta)}{f(\mathbf{x})} \\ &\propto f(\mathbf{x}|\theta)\pi(\theta) = L(\theta|\mathbf{x})\pi(\theta)\end{aligned}$$

  其中 $f(\mathbf{x}|\theta)$ 为给定参数 θ 下样本 $\mathbf{x}$ 的模型, 当视为 θ 的函数 $L(\theta) = f(\mathbf{x}|\theta)$ 时候, 称为似然函数.

- 关于参数 θ 的一切推断应该基于其后验分布进行.

贝叶斯推断满足似然原理

证明. 假设 $L(\theta|\mathbf{x}_1) \propto L(\theta|\mathbf{x}_2)$, 以及 θ 的先验分布为 $\pi(\theta)$. 则

$$
\begin{aligned}
\pi(\theta|\mathbf{x}_1) &\propto \pi(\theta)L(\theta|\mathbf{x}_1) \\
&\propto \pi(\theta)L(\theta|\mathbf{x}_2) \\
&\propto \pi(\theta|\mathbf{x}_2) \\
&= \pi(\theta|\mathbf{x}_2)
\end{aligned}
$$

注: 上面我们假设了先验分布为主观先验, 也就是不是通过一些例如总是均匀分布等的规则选取的. 在后者情形下, 是可以违背似然原理的. □

总结如下:

名称	符号	等于
model, likelihood	$f(\mathbf{x}\|\theta)$	
prior	$\pi(\theta)$	
joint	$h(\mathbf{x},\theta)$	$f(\mathbf{x}\|\theta)\pi(\theta)$
marginal	$m(\mathbf{x})$	$\int f(\mathbf{x}\|\theta)\pi(\theta)d\theta$
posterior	$\pi(\theta\|\mathbf{x})$	$f(\mathbf{x}\|\theta)\pi(\theta)/m(\mathbf{x})$
predictive	$f(y\|\mathbf{x})$	$\int_\Theta f(y\|\theta)\pi(\theta\|\mathbf{x})d\theta$

$\uparrow$Example

例2.5.1 (P52) 假设 X 给定 θ 的分布为二项分布 $Bin(n,\theta)$, i.e.,

$$f(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, x = 0, 1, \ldots, n$$

先验分布为 beta 分布, $Be(\alpha,\beta)$, 其中超参数 α 和 β 已知,

$$\pi(\theta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, 0 \le \theta \le 1$$

其中 $B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$. 求联合, 后验, 边际, 预测分布.

$\downarrow$Example

- $h(x,\theta) = \frac{\binom{n}{x}}{B(\alpha,\beta)}\theta^{\alpha+x-1}(1-\theta)^{n-x+\beta-1}, 0 \le \theta \le 1, x = 0, 1, \ldots, n$
- $m(x) = \frac{\binom{n}{x}B(x+\alpha, n-x+\beta)}{B(\alpha,\beta)}, x = 0, 1, \ldots, n$

- $\pi(\theta|x) = \frac{1}{B(x+\alpha, n-x+\beta)}\theta^{\alpha+x-1}(1-\theta)^{n-x+\beta-1}, 0 \leq \theta \leq 1$, 其为 $Be(\alpha+x, n-x+\beta)$
- $f(y|x) = \frac{\binom{n}{x}B(x+y+\alpha, 2n-x-y+\beta)}{B(x+\alpha, n-x+\beta)}, y = 0, 1, \ldots, n$

例2.5.2-3 (P54) 假设 $X|\theta \sim N(\theta, \sigma^2)$, $\theta \sim N(\mu, \tau^2)$, 超参数 σ^2, μ, τ^2 均已知, 求联合, 后验, 边际, 预测分布.

↑Example
↓Example

- $\theta|x \sim N(\frac{\tau^2}{\sigma^2+\tau^2}x + \frac{\sigma^2}{\sigma^2+\tau^2}\mu, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2})$.
- 如果样本 $X_1, \ldots, X_n i.i.d \sim X$, 则

$$\theta|\bar{X} \sim \mathcal{N}\left(\frac{\tau^2}{\frac{\sigma^2}{n}+\tau^2}\bar{X} + \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n}+\tau^2}\mu, \frac{\frac{\sigma^2}{n}\tau^2}{\frac{\sigma^2}{n}+\tau^2}\right)$$

其中 $\bar{X}$ 为样本均值.

- 给定样本 $X_1, \ldots, X_n$, Y 的预测分布为

$$Y|\bar{X} \sim \mathcal{N}\left(\frac{\tau^2}{\frac{\sigma^2}{n}+\tau^2}\bar{X} + \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n}+\tau^2}\mu, \sigma^2 + \frac{\frac{\sigma^2}{n}\tau^2}{\frac{\sigma^2}{n}+\tau^2}\right)$$

后验均值为似然估计和先验均值的加权.

正态-正态模型, $\pi(\theta|X_1,\ldots,X_n) = \pi(\theta|\bar{X})$ 是下述引理的特例.

引理 1. *假设充分统计量 $T(X_1,\ldots,X_n)$ 存在, 则 $\pi(\theta|X_1,\ldots,X_n) = \pi(\theta|T)$.*

证明. 由因子分解定理, $f(x|\theta) = g(t,\theta)h(x)$ 其中 $t = T(x)$, $h(x)$ 无 θ 无关. 因此

$$\begin{aligned}
\pi(\theta|x) &= \frac{\pi(\theta)f(x|\theta)}{\int_\theta \pi(\theta)f(x|\theta)d\theta} = \frac{\pi(\theta)g(t,\theta)h(x)}{\int_\theta \pi(\theta)g(t,\theta)h(x)d\theta} \\
&= \frac{\pi(\theta)g(t,\theta)}{\int_\theta \pi(\theta)g(t,\theta)d\theta} = \frac{\pi(\theta)g(t,\theta)\phi(t)}{\int_\theta \pi(\theta)g(t,\theta)\phi(t)d\theta} = \frac{\pi(\theta)f(t|\theta)}{\int_\theta \pi(\theta)f(t|\theta)d\theta} = \pi(\theta|t)
\end{aligned}$$

其中 $f(t|\theta) = \int_{x:T(x)=t} f(x|\theta)dx = \int_{x:T(x)=t} g(t,\theta)h(x)dx$
$= g(t,\theta)\left[\int_{x:T(x)=t} h(x)dx\right] = g(t,\theta)\phi(t)$. □

统计量 $T = T(X)$ 为充分的（在贝叶斯意义下），当且仅当对任何先验分布，得到的后验分布都满足

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

Definition

定理 1. *统计量 T 在贝叶斯意义下是充分的当且仅当它在经典统计意义下是充分的.*

后验分布是一个试验的终极贝叶斯总结. 后验分布的位置度量 (特别是均值) 非常重要. 后验均值是一种可行的贝叶斯估计. 我们后面会看到, 在决策分析框架下, 后验位置参数 (均值, 众数, 中位数) 为不同损失函数下的贝叶斯估计.

1.2 贝叶斯估计

未知参数 θ 的后验分布 $\pi(\theta)$ 集中了抽样信息和先验信息中关于 θ 的所有信息，所以有关 θ 的点估计，区间估计和假设检验等统计推断方法都是按照一定方式从后验分布中提取信息的，其提取方法与经典统计推断相比要简单明确的多.

- 基于后验分布的统计推断就意味着只考虑已出现的样本观测值，认为未出现的数据与推断无关，这一观点被称为“条件观点”，基于这种观点提出的统计推断方法被称为 **条件方法**.

- 它与我们熟悉的 **频率方法**之间具有很大的差别. 例如，在对估计量的无偏性的认识上，经典统计学认为参数 θ 的无偏估计 $\hat{\theta}(\mathbf{x})$ 应满足 $E[\hat{\theta}(X)]=\theta$，其中平均是对样本空间中所有可能出现的样本而求的，可实际中样本空间中绝大多数样本尚未出现过，甚至重复数百次也不会出现的样本也要在评价估计

量 $\hat{\theta}$ 的好坏中占一席之地, 何况在实际中不少估计量只使用一次或几次，而多数从未出现的样本也要参与平均，是使实际工作者难于理解的，这是“频率方法”的缺点. 这就是贝叶斯方法中的“条件观点”. 故在贝叶斯统计推断中不用无偏性，而条件方法是容易被实际工作者理解和接受的.

- 处理多余参数没有理论问题: 如果 $\theta=(\theta_1,\theta_2)$, θ_2 为多余参数, 则我们可以将先验联合分布表示为 $\pi(\theta)=\pi\left(\theta_1|\theta_2\right)\pi\left(\theta_2\right)$, 从而

$$\pi\left(\theta_1|\mathbf{x}\right)=\int \pi\left(\theta_1|\theta_2,\mathbf{x}\right)\pi\left(\theta_2|\mathbf{x}\right)d\theta_2$$

注意: 如果 $\pi\left(\theta_1|\theta_2,\mathbf{x}\right)$ 对不同的 θ_2 变化较大时候, 则需要考虑进行灵敏度分析. 参见 See Box and Tiao (1992), Section 1.6.

1.2.1 贝叶斯点估计

- 贝叶斯估计问题是被作为一个决策问题处理的．我们应该选择一个可以最小化期望损失的量作为估计量.(将在贝叶斯决策分析内容中详细介绍)

- 有了后验分布后, 可从后验分布出发, 按经典方法 (用后验分布代替通常的样本分布) 求未知参数 θ 的点估计，如后验众数估计、后验中位数估计和后验期望估计等.

$$\hat{\theta}_B = \begin{cases} \hat{\theta}_{MAP} = \arg\max_\theta \pi(\theta|\mathbf{x}), & \text{后验众数} \\ \hat{\theta}_E = E[\theta|\mathbf{x}], & \text{后验期望} \\ \hat{\theta}_{ME} = \arg\max_a E[|\theta - a||\mathbf{x}], & \text{后验中位数,}\theta\text{ 为一维时} \end{cases}$$

可根据问题背景选择其一作为 θ 的估计.

↑Example

例 4.2.6. (P110) 为估计产品不合格品率 θ, 今从一批产品中随机有放回抽取 n 件检查, 检查结果为 $X_1, \cdots, X_n$, 其中 $X_i = 1$ 表示抽出的第 i 件不合格, $X_i = 0$ 表示抽出的第 i 件合格, 不合格品数 $X = \sum_{i=1}^{n} X_i$. 若取 θ 的先验分布为共轭先验 $Be(a, b)$, 求 θ 的后验众数估计和后验期望估计, 并比较这两个估计.

↓Example

易知 θ 的后验众数估计为

$$\hat{\theta}_{MAP} = \frac{(X+a)-1}{(X+a)+(n-X+b)-2} = \frac{X+a-1}{n+a+b-2};$$

后验均值估计为

$$\hat{\theta}_E = \frac{X+a}{n+a+b}.$$

- 当 $a = b = 1$ 时候, θ 的后验众数估计就是经典统计学中的 MLE, 即为无信息先验分布 $U(0, 1)$ 下的贝叶斯估计.

- θ 的后验期望估计要比后验众数更合适一些, 参考课本 111 页表 4.2.1 的对比试验.

例 4.2.3 (P108) 设随机变量 $X \sim f(x|\theta) = e^{-(x-\theta)} \cdot I_{[\theta,\infty)}(x)$, 此处$-\infty < \theta < +\infty$ 为位置参数, 取 θ 的先验分布为柯西分布，即先验密度 $h(\theta) = \frac{1}{\pi(1+\theta^2)}$， $-\infty < \theta < +\infty$, 求 θ 的后验众数估计.

↑Example

↓Example

例 4.2.4 (P108) 设 $X_1, \cdots, X_n$ 为从正态总体 $N(\theta, \sigma^2)$ 中抽取的简单样本, 其中 $\sigma^2 > 0$ 已知, $\theta > 0$ 未知, 取 θ 的先验分布为无信息先验, 即 $\pi(\theta) = I_{(0,\infty)}(\theta)$, 求 θ 的后验期望估计.

例 4.2.7 (P112) 设 $X_1,\cdots,X_n\ iid \sim f(x|\theta)=\theta^{-1}\exp\{-x/\theta\}, x>0$, $\theta>0$, 当观察到前 r 个样本 $X_{(1)}\le X_{(2)}\le\cdots\le X_{(r)}$ 时就停止, 此处 $X_{(1)}\le X_{(2)}\le\cdots\le X_{(n)}$ 为样本 $X_1,\cdots,X_n$ 的次序统计量, 取 θ 的先验分布为无信息先验 $\pi(\theta)=1/\theta$, $\theta>0$, 求可靠性函数 $R(s)=P(X\ge s)=e^{-\frac{s}{\theta}}$, $s>0$ 的后验期望估计.

↑Example

↓Example

1.2.2 贝叶斯估计的误差

- 在经典方法中衡量一个估计量的优劣看其均方误差 (MSE) 大小, 均方误差越小越好. 对贝叶斯估计 $\delta(\mathbf{x})$, 衡量优劣用如下定义的后验均方误差 (简记为 PMSE)

$$PMSE(\delta(\mathbf{x})) = E_{\theta|\mathbf{x}}[(\theta - \delta(\mathbf{x}))^2]$$

来度量估计量的精度, PMSE 越小越好.

- 特别若 $\delta(\mathbf{x}) = E(\theta|\mathbf{x})$ 时, 则 $\delta(\mathbf{x})$ 的 PMSE 即后验方差, 即

$$PMSE(\delta(\mathbf{x})) = E_{\theta|\mathbf{x}}[(\theta - \delta(\mathbf{x}))^2] = D(\theta|\mathbf{x}).$$

- 记 θ 的后验均值 $E(\theta|x) = \mu^\pi(x)$, 后验方差 $V^\pi(x) = D(\theta|x)$, 则 PMSE 与后验方差 $V^\pi(x)$ 的关系如下:

$$\begin{aligned} PMSE(\delta(x)) &= E_{\theta|x}[(\theta - \mu^\pi(x)) + (\mu^\pi(x) - \delta(x))]^2 \\ &= V^\pi(x) + (\mu^\pi(x) - \delta(x))^2 \geq V^\pi(x). \end{aligned}$$

且等号成立的充要条件是 $\delta(x) = \mu^\pi(x)$, 即 θ 的后验均值估计使 PMSE 达到最小.

- 后验期望估计是在 PMSE 准则下的最优估计. 这就是为什么习惯上在后验众数估计、后验中位数估计和后验期望估计这三种估计中常取后验期望 $\mu^\pi(x)$ 作为 θ 的贝叶斯估计.

例4.2.5 (P110) 假设 $X|\theta \sim N(\theta, \sigma^2)$, $\theta \sim N(\mu, \tau^2)$, 超参数 σ^2, μ, τ^2 均已知, 记 $\hat{\theta}_{MAP}$ 为 θ 的后验众数估计, 比较其与 θ 的最大似然估计 $\hat{\theta}_{MLE} = X$ 的 PMSE.

↑Example
↓Example

↑Example

例 4.2.6 (P110) $X \sim B(n,\theta)$，θ 的先验分布是 $Be(a,b)$，(1) 求 θ 后验期望估计与后验方差. (2) 当取 θ 的先验分布是均匀分布 $U(0,1)$ 时，求后验期望估计、后验众数估计及其后验方差和后验均方误差.

↓Example

↑Example

例 4.2.8 (P113) 设一批产品不合格率为 θ, 检查是一个接一个地进行，直到发现第一个不合格产品就停止检查. 设 X 为发现第一个不合格品时已检查的产品数, 则 X 服从几何分布. 假设参数 θ 只能取 1/4, 2/4, 3/4 三个值，且取这三个值的概率相同，如今获得一个样本观测值 $x = 3$, 求 θ 的后验众数估计，并计算它的后验均方误差.

↓Example

1.2.3 贝叶斯可信区间

设参数 θ 的后验分布为 $\pi(\theta|\mathbf{x})$，对给定的样本 $\mathbf{x}$ 和概率 $0 < \alpha < 1$，若存在两个统计量 $\hat{\theta}_1(\mathbf{x})$ 和 $\hat{\theta}_2(\mathbf{x})$，使得

$$P(\hat{\theta}_1 \le \theta \le \hat{\theta}_2|\mathbf{x}) \ge 1 - \alpha,$$

Definition

则称 $[\hat{\theta}_1, \hat{\theta}_2]$ 为 θ 的可信水平为 $1 - \alpha$ 的贝叶斯可信区间 (credible interval).

- 参数 θ 的 1-α 置信区间 (confidence interval) 是指满足下面要求的区间 $[a(\mathbf{x}), b(\mathbf{x})]$:

$$P^{\mathbf{x}|\theta}(a(\mathbf{x}) \le \theta \le b(\mathbf{x})]) = 1 - \alpha, \forall\theta.$$

- 对求出的 90% 贝叶斯可信区间, 例如 [1.2,2.0], 可以说 θ 落在 $[1.2, 2.0]$ 中的概率为 0.90; **频率方法则不能如此解释**. 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此区间估计一次或两次的人来说是毫无意义的.

- 经典统计方法依赖于枢轴变量, 有时会很困难. 贝叶斯方法相比要简单得多.

例4.3.1 (P117) 假设 $X|\theta \sim N\left(\theta, \sigma^2\right)$, $\theta \sim N\left(\mu, \tau^2\right)$, 超参数 σ^2, μ, τ^2 均已知, 求 θ 的 $1-\alpha$ 可信区间.

↑Example
↓Example

↑Example

例 4.3.2 (P118) 设 $X_1, \cdots, X_n$ $iid \sim f(x|\theta) = \theta^{-1}\exp\{-x/\theta\}, x > 0, \theta > 0$, 当观察到前 $r(1 \le r \le n)$ 个样本 $X_{(1)} \le X_{(2)} \le \cdots \le X_{(r)}$ 时就停止, 取 θ 的先验分布为逆伽马分布

$$\pi(\theta; a, b) = \frac{b^a}{\Gamma(a)}\theta^{-(a+1)}e^{-b/\theta}I_{(0,\infty)}(\theta)$$

求 θ 的后验期望和 $1-\alpha$ 可信下限.

↓Example

HPD 可信区间

等尾可信区间在实际中常被用，计算方便，但不是最好的，最好的可信区间应是区间长度最短的．若后验分布是单峰对称，则等尾是最好的．要使可信区间最短，只有把具有最大后验密度的点都包含在区间的，而在区间外的点后验密度的值都不会超过区间内的点后验密度的值，这样的区间称为最大后验密度可信区间，定义如下：

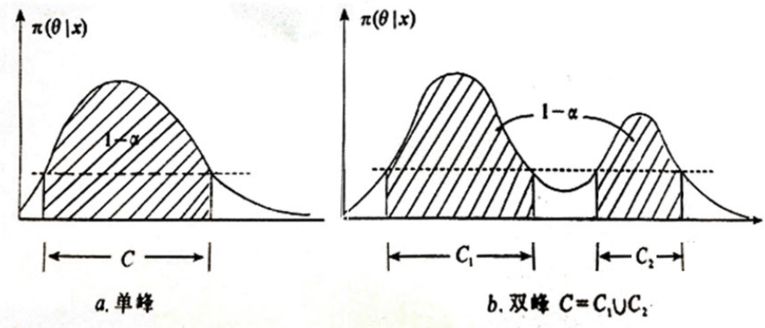
设参数 θ 的后验密度为 $\pi(\theta|\mathbf{x})$，对给定的概率 $1-\alpha$ $(0<\alpha<1)$，集合 C 满足如下条件：

(i) $P(\theta\in C|\mathbf{x})=1-\alpha$,

(ii) 对任给的 $\theta_1 \in C$ 和 $\theta_2 \bar{\in} C$，总有 $\pi(\theta_1|\mathbf{x}) > \pi(\theta_2|\mathbf{x})$，则称 C 为 θ 的可信水平 $1-\alpha$ 最大后验密度可信集，简称为 $1-\alpha$ 的 HPD 可信集（区间）．

Definition

HPD 区间:



1.3 假设检验问题

- 设假设检验问题的一般形式是:

$$H : \theta \in \Theta_H \longleftrightarrow K : \theta \in \Theta_K$$

此处 $\Theta_H \bigcup \Theta_K = \Theta$, 其中 Θ 是参数空间, Θ_H 是 Θ 的非空真子集

- 获得参数 θ 的后验分布后，计算 Θ_H 和 Θ_K 的后验概率

$$p_H(\mathbf{x}) = P(\theta \in \Theta_H|\mathbf{x}), \qquad p_K(\mathbf{x}) = P(\theta \in \Theta_K|\mathbf{x}).$$

若 $p_H(x) > p_K(\mathbf{x})$, 则接受 H，否则拒绝 H.

例 4.4.1 (P127) 设 $X|\theta \sim B(n,\theta)$, θ 的先验分布为 $U(0,1)$, 考虑下述检验

$$H_0 : 0 < \theta \leq 1/2 \leftrightarrow H_1 : 1/2 < \theta < 1$$

↑Example

↓Example

1.3.1 贝叶斯因子

设两个假设 Θ_0 和 Θ_1 的先验概率分别为 π_0 和 π_1, 后验概率分别为 α_0 和 α_1, 比例 α_0/α_1 称为 H_0 对 H_1 的后验机会比, π_0/π_1 被称为先验机会比, 则称

$$BF(\mathbf{x}) = \frac{\text{后验机会比}}{\text{先验机会比}} = \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1} = \frac{\alpha_0\pi_1}{\alpha_1\pi_0}$$

为支持 H_0 的贝叶斯因子. $BF(\mathbf{x})$ 取值越大, 对 H_0 的支持程度越高.

Definition

从贝叶斯因子定义看，它既依赖于数据 $\mathbf{x}$, 又依赖于先验分布 π, 对两种机会比相除，很多人认为，这会减弱先验分布的影响，突出数据的影响. 从这定义上看贝叶斯因子 $BF(\mathbf{x})$ 是反映数据 $\mathbf{x}$ 支持 H_0 的程度.

两点假设情形

- 要解释 $BF(\mathbf{x})$ 的意义，首先看 Θ_0 和 Θ_1 皆为简单假设的情形，即 $H_0 : \Theta_0 = \{\theta_0\} \leftrightarrow H_1 : \Theta_1 = \{\theta_1\}$. 此时

$$\alpha_0 = P(\Theta_0|\mathbf{x}) = \frac{f(\mathbf{x}|\theta_0)\pi_0}{f(\mathbf{x}|\theta_0)\pi_0 + f(\mathbf{x}|\theta_1)\pi_1},$$

$$\alpha_1 = P(\Theta_1|\mathbf{x}) = \frac{f(\mathbf{x}|\theta_1)\pi_1}{f(\mathbf{x}|\theta_0)\pi_0 + f(\mathbf{x}|\theta_1)\pi_1},$$

其中 $f(\mathbf{x}|\theta)$ 为样本的分布，这时后验机会比为

$$\frac{\alpha_0}{\alpha_1} = \frac{\pi_0 f(\mathbf{x}|\theta_0)}{\pi_1 f(\mathbf{x}|\theta_1)}.$$

- 故贝叶斯因子为

$$BF(\mathbf{x}) = \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1} = \frac{f(\mathbf{x}|\theta_0)}{f(\mathbf{x}|\theta_1)}.$$

- 如果要拒绝零假设 H_0, 则要求 $\alpha_0/\alpha_1 < 1$, 故

$$\frac{f(\mathbf{x}|\theta_1)}{f(\mathbf{x}|\theta_0)} > \frac{\pi_0}{\pi_1}.$$

这与著名的 Neyman-Pearson 引理的基本结果类似，从贝叶斯观点看, 这个临界值就是两个先验概率比.

- 因此可见 $BF(\mathbf{x})$ 正是 $\Theta_0 \leftrightarrow \Theta_1$ 的似然比，它通常被认为是由数据给出的 $\Theta_0 \leftrightarrow \Theta_1$ 的机会比. 由于此种情形的贝叶斯因子不依赖于先验分布，仅依赖于样本的似然比，故贝叶斯因子 $BF(\mathbf{x})$ 可视为是数据 $\mathbf{x}$ 支持 Θ_0 的程度.

↑Example

例 4.4.2 (P129) 设 $X_1, \ldots, X_n$ $i.i.d \sim N(\theta, 1)$, 其中 θ 只取 0 和 1 两个值, 考虑假设

$$H_0 : \theta = 0 \leftrightarrow H_1 : \theta = 1$$

↓Example

复杂假设对复杂假设情形

- 考虑下列假设检验问题:

$$H_0 : \theta \in \Theta_0 \longleftrightarrow H_1 : \theta \in \Theta_1,$$

- 将先验分布 $\pi(\theta)$ 写成如下形式:

$$\pi(\theta) = \begin{cases} \pi_0 g_0(\theta), & \text{当 } \theta \in \Theta_0 \\ \pi_1 g_1(\theta), & \text{当 } \theta \in \Theta_1 \end{cases}$$

其中 π_0 和 π_1 分别为 Θ_0 和 Θ_1 上的先验概率, $g_0(\theta)$ 和 $g_1(\theta)$ 分别是 Θ_0 和 Θ_1 上的概率密度函数. 易见

$$\int_\Theta \pi(\theta)d\theta = \int_{\Theta_0} \pi_0 g_0(\theta)d\theta + \int_{\Theta_1} \pi_1 g_1(\theta)d\theta = \pi_0 + \pi_1 = 1\,,$$

即 $\pi(\theta)$ 是先验密度.

- 此时后验概率比为

$$\frac{\alpha_0}{\alpha_1} = \frac{\int_{\Theta_0} f(\mathbf{x}|\theta)\pi_0 g_0(\theta)d\theta}{\int_{\Theta_1} f(\mathbf{x}|\theta)\pi_1 g_1(\theta)d\theta}\ .$$

- 故贝叶斯因子可表示为

$$BF(\mathbf{x}) = \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1} = \frac{\int_{\Theta_0} f(\mathbf{x}|\theta)g_0(\theta)d\theta}{\int_{\Theta_1} f(\mathbf{x}|\theta)g_1(\theta)d\theta} = \frac{m_0(\mathbf{x})}{m_1(\mathbf{x})}\ .$$

- 可见 $BF(\mathbf{x})$ 还依赖于 Θ_0 和 Θ_1 上的先验密度 g_0 和 g_1, 这时贝叶斯因子虽然已不是似然比，但仍可看作 Θ_0 和 Θ_1 上的加权似然比 (权重分别为 g_0 和 g_1). 此时还不能认为对两个假设支持的度量完全由数据 $\mathbf{x}$ 决定，它还和 g_0 和 g_1 有关, 只能说它部分地消除了先验分布的影响.

- 当 $BF(\mathbf{x})$ 对 g_0 和 g_1 的选择相对不敏感时，那么在这一情形下才可以说仅仅由数据来决定上述比值是合理的了.

- 若设 $\hat{\theta}_0$ 与 $\hat{\theta}_1$ 分别是 θ 在 Θ_0 和 Θ_1 的 MLE (极大似然估计), 那么经典统计中使用的似然比统计量

$$\lambda(\mathbf{x}) = \frac{f(\mathbf{x}|\hat{\theta}_0)}{f(x|\hat{\theta}_1)} = \frac{\sup_{\theta\in\Theta_0} f(\mathbf{x}|\theta_0)}{\sup_{\theta\in\Theta_1} f(\mathbf{x}|\theta_1)}$$

是贝叶斯因子 $BF(\mathbf{x})$ 的特殊情况，即认为 $g_0(\theta)$ 和 $g_1(\theta)$ 的质量全部集中在 $\hat{\theta}_0$ 与 $\hat{\theta}_1$ 上.

例 4.4.5 (P131) 设 $X_1, \ldots, X_{10}\ i.i.d \sim N(\theta, 1)$, 样本均值为 1.5, 若 θ 的先验分布为 $N(0.5, 2)$, 考虑假设

$$H_0 : \theta \leq 1 \leftrightarrow H_1 : \theta > 1$$

↑Example

↓Example

简单假设对复杂假设

- 考虑下列假设检验问题:

$$H_0 : \theta = \theta_0 \longleftrightarrow H_1 : \theta \neq \theta_0.$$

这是经典统计中常见的一类检验问题. 当参数 θ 为连续变量时, 用简单假设 $H_0 : \theta = \theta_0$ 是不合理的. 一个合理的假设是改上述检验为

$$H_0 : \ \theta \in [\theta_0 - \varepsilon, \theta_0 + \varepsilon] \longleftrightarrow H_1 : \ \theta \notin \ [\theta_0 - \varepsilon, \theta_0 + \varepsilon].$$

此处 ε 是较小的正数, 可选其为误差范围内一个较小的数.

- 下面考虑 $H_0 : \theta = \theta_0 \longleftrightarrow H_1 : \theta \neq \theta_0$ 的贝叶斯检验如何导出. 对 $H_0 : \theta = \theta_0$, 不能采用连续密度作为先验密度, 一个有效的办法是给 θ_0 一个正概率 π_0, 而对 $\theta \neq \theta_0$, 给一个加权密

度 $\pi_1 g_1(\theta)$ (其中 $\pi_0+\pi_1=1$), 即 θ 的先验密度为

$$\pi(\theta)=\begin{cases}\pi_0, & 当\ \theta=\theta_0\\ \pi_1 g_1(\theta), & 当\ \theta\neq\theta_0\end{cases}$$

- 设样本分布为 $f(\mathbf{x}|\theta)$，则易求边缘分布为

$$m(\mathbf{x})=\int_{\Theta} f(\mathbf{x}|\theta)\pi(\theta)d\theta=\pi_0 f(\mathbf{x}|\theta_0)+\pi_1 m_1(\mathbf{x}).$$

其中

$$m_1(\mathbf{x})=\int_{\{\theta\neq\theta_0\}} f(\mathbf{x}|\theta)g_1(\theta)d\theta.$$

- 故 $\theta=\theta_0$ 的后验概率和 $\theta\neq\theta_0$ 的后验概率分别为

$$\alpha_0=\pi(\Theta_0|\mathbf{x})=\frac{\pi_0 f(\mathbf{x}|\theta_0)}{m(\mathbf{x})},\quad \alpha_1=\pi(\Theta_1|\mathbf{x})=\frac{\pi_1 m_1(\mathbf{x})}{m(\mathbf{x})}.$$

- 后验机会比为:

$$\frac{\alpha_0}{\alpha_1}=\frac{\pi_0 f(\mathbf{x}|\theta_0)}{\pi_1 m_1(\mathbf{x})}.$$

- 于是贝叶斯因子为

$$BF(\mathbf{x}) = \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1} = \frac{f(\mathbf{x}|\theta_0)}{m_1(\mathbf{x})}.$$

可见 Bayes 因子更简单.

- 故实际中常是先计算 $BF(\mathbf{x})$, 后计算 α_0 和 α_1. 因为由贝叶斯因子定义可知 $\alpha_0 + \alpha_1 = 1$, 可推出

$$\alpha_0 = \pi(\Theta_0|\mathbf{x}) = \left[1 + \frac{1-\pi_0}{\pi_0} \cdot \frac{1}{BF(\mathbf{x})}\right]^{-1}.$$

例 4.4.6 (P134) 设 θ 为抛硬币正面朝上的概率, 考虑假设

$$H_0 : \theta = 1/2 \leftrightarrow H_1 : \theta > 1/2$$

为此做了一系列试验, 结果出现 9 次正面和 3 次反面.

↑Example
↓Example

多重假设检验

- 多重假设检验并不比两个假设检验更困难, 即直接计算每一个假设的后验概率, 并比较其大小. 设检验问题是

$$H_i : \theta \in \Theta_i,\ i = 1, \cdots, k;$$

其中 $\Theta_1 \cup \Theta_2 \cup \cdots \cup \Theta_k = \Theta$, 每个 Θ_i 是参数空间 Θ 的非空真子集.

- 计算 Θ_i 的后验概率

$$\alpha_i = P(\Theta_i|x),\ i = 1, \cdots, k$$

取其最大者, 则认为相应的假设成立.

1.4 预测推断

- 预测问题的典型情况是: 若 $X \sim f(x|\theta)$, 获得数据 $Z = z$ 后, 对具有密度为 $g(z|\theta)$ 的随机变量 Z 未来观察值作出预测. 通常假定 Z 和 X 不相关, f 和 g 分别为密度函数.

- 贝叶斯预测的一般想法: 设 $\pi(\theta|\mathbf{x})$ 为 θ 的后验分布, 于是 $g(z|\theta)\pi(\theta|\mathbf{x})$ 为给定 $X = \mathbf{x}$ 的条件下 (Z, θ) 的联合分布, 把它对 θ 积分得到给定 $X = \mathbf{x}$ 时 Z 的边缘分布密度作为预测密度.

Definition

设随机变量 X 的密度函数是 $f(x|\theta)$, θ 的先验密度是 $\pi(\theta)$. 给定 $X = x$, 随机变量 Z 的后验预测密度定义为

$$p(z|x) = \int_\Theta g(z|\theta)\pi(\theta|x)d\theta.$$

- 点预测: 使用后验预测密度 $p(z|x)$ 的期望值, 中位数或众数等作为 z 的预测值

- 区间预测: 可基于预测密度 $p(z|x)$ 的 $1-\alpha$ 可信区间 $[a,b]$:

$$P(a \le Z \le b|x) = \int_a^b p(z|x)dz = 1-\alpha$$

例 4.5.1(P137) 一赌徒在过去 10 次赌博中赢了 3 次，现要对未来 5 次赌博中他赢的次数 Z 做出预测.

↑Example

↓Example

这个问题的一般提法是: 在 n 次独立的 Bernoulli 试验成功了 X 次, $X|\theta \sim B(n,\theta)$, 成功概率 θ 的先验分布为 $Be(a,b)$. 对未来 k 次独立 Bernoulli 试验成功次数 Z 作预测.

在此问题中, $n=10$, $x=3$, $k=5$. 取 $a=b=1$, 即先验分布为 $Be(1,1)=U(0,1)$, 则

$$p(z|x=3)=\binom{5}{z}\frac{\Gamma(12)\Gamma(4+z)\Gamma(13-z)}{\Gamma(4)\Gamma(8)\Gamma(17)}.$$

计算可得 $z=0,1,2,3,4,5$ 时后验预测概率

$$p(0|3)=0.1813,\quad p(1|3)=0.3022,\quad p(2|3)=0.2747,$$
$$p(3|3)=0.1649,\quad p(4|3)=0.0641,\quad p(5|3)=0.01282.$$

由后验预测分布可见, 其概率集中在 $Z=0,1,2,3$ 之间, 即

$$P_{Z|x}\big(0 \le Z \le 3\big) = 0.9231.$$

这表明 $[0,3]$ 是 Z 的 92% 预测区间. 另外分布众数在 $z=1$ 处, 第二大的概率在 $z=2$ 处出现, 可见未来 5 次赌博中胜 1 或 2 次可能性最大.

↑Example

例 4.5.2(P138) 一颗钻石在一架天平上重复称重 n 次，其结果为 $X_1, \cdots, X_n$, 若把钻石放在另一架天平上称重，如何对其称量值作出预测？

↓Example

解: 假设天平称重服从正态分布, 对第一架天平, θ 的后验分布 $\pi(\theta|\bar{x})$ 为 $N(\mu_1, \eta_1^2)$, 其中

$$\mu_1 = \frac{\tau^2}{\sigma_1^2/n + \tau^2}\bar{x} + \frac{\sigma_1^2/n}{\sigma_1^2/n + \tau^2}\mu,$$

$$\eta_1^2 = \frac{\sigma_1^2/n \cdot \tau^2}{\sigma_1^2/n + \tau^2} = \frac{\sigma_1^2\tau^2}{\sigma_1^2 + n\tau^2}.$$

设第二架天平称得钻石的重量为 Z, $Z|\theta \sim N(\theta, \sigma_2^2)$, 则在已知第一架天平的平均称重结果后,$Z$ 的后验预测密度为

$$p(z|\bar{x}) = \int_{-\infty}^{\infty} g(z|\theta)\pi(\theta|\bar{x})d\theta$$

即后验预测分布为 $N(\mu_1, \eta_1^2 + \sigma_2^2)$.

- 后验预测分布的均值、方差分别为

$$E(Z|\bar{x}) = \mu_1 = \frac{\tau^2}{\sigma_1^2/n + \tau^2}\bar{x} + \frac{\sigma_1^2/n}{\sigma_1^2/n + \tau^2}\mu,$$
$$D(Z|\bar{x}) = \eta_1^2 + \sigma_2^2.$$

- 若取预测分布的均值作为 Z 的预测值，则有

$$\hat{Z} = \mu_1 = \frac{\tau^2}{\sigma_1^2/n + \tau^2}\bar{x} + \frac{\sigma_1^2/n}{\sigma_1^2/n + \tau^2}\mu.$$

- 易见 $(Z - \mu_1)/\sqrt{\eta_1^2 + \sigma_2^2} \sim N(0, 1)$. 可见 Z 的 $1 - \alpha$ 的后验预测区间为

$$\left[\mu_1 - u_{\alpha/2}\sqrt{\eta_1^2 + \sigma_2^2},\ \mu_1 + u_{\alpha/2}\sqrt{\eta_1^2 + \sigma_2^2}\ \right],$$

其中 此处 $u_{\alpha/2}$ 为 $N(0, 1)$ 的上侧 $\alpha/2$ 分位数.