

# 贝 叶 斯 统 计

韦 来 生 编 著

高 等 教 育 出 版 社

第一版, 2016

# 贝 叶 斯 统 计

韦 来 生 编 著

高 等 教 育 出 版 社

第一版, 2016

## 1 4.1 贝叶斯点估计

- 4.1.1 条件方法
- 4.1.2 贝叶斯点估计
- 4.1.3 贝叶斯点估计的精度-估计的误差
- 4.1.4 贝叶斯点估计的几个例子
- 4.1.5 多参数情形

## 2 4.2 贝叶斯区间估计

- 4.2.1 可信区间定义及例子
- 4.2.2 最大后验密度 (HPD) 可信区间
- 4.2.3 大样本方法

## 3 4.3 假设检验

- 4.3.1 一般方法
- 4.3.2 贝叶斯因子
- 4.3.3 简单假设对简单假设情形
- 4.3.4 复杂假设对复杂假设情形
- 4.3.5 简单假设对复杂假设
- 4.3.6 多重假设检验

## 条件方法

- 后验分布  $\pi(\theta|\mathbf{x})$  是在样本  $\mathbf{x}$  给定下  $\theta$  的条件分布，基于后验分布的统计推断就意味着只考虑已出现的数据 (样本观测值)，而认为未出现的数据与推断无关，这一重要的观点被称为“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  $\theta$  的无偏估计  $\hat{\theta}(\mathbf{X})$  应满足  $E[\hat{\theta}(\mathbf{X})] = \int_{\mathcal{X}} \hat{\theta}(\mathbf{x})p(\mathbf{x}|\theta)d\mathbf{x} = \theta$ ，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”。

## 条件方法

- 后验分布  $\pi(\theta|\mathbf{x})$  是在样本  $\mathbf{x}$  给定下  $\theta$  的条件分布，基于后验分布的统计推断就意味着只考虑已出现的数据 (样本观测值)，而认为未出现的数据与推断无关，这一重要的观点被称为“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  $\theta$  的无偏估计  $\hat{\theta}(\mathbf{X})$  应满足  $E[\hat{\theta}(\mathbf{X})] = \int_{\mathcal{X}} \hat{\theta}(\mathbf{x})p(\mathbf{x}|\theta)d\mathbf{x} = \theta$ ，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”。

# 贝叶斯点估计方法

- 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  $\theta$  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, 其定义如下:
- **定义 4.1.1** 使后验密度  $\pi(\theta|x)$  达到最大时  $\theta$  的值作为  $\theta$  的估计量, 称为后验众数估计, 也称为广义极大似然估计或后验极大似然估计, 记为  $\hat{\theta}_{MD}$ ; 用后验分布的中位数作为  $\theta$  的估计量, 称为  $\theta$  的后验中位数估计, 记为  $\hat{\theta}_{ME}$ ; 用后验分布的期望作为  $\theta$  的估计量, 称为后验期望估计, 记为  $\hat{\theta}_E$ . 三个估计都称为贝叶斯估计, 在不会引起混淆的情况下上述三个估计皆用  $\hat{\theta}_B$  来记.
- **注 4.1.1** 一般场合下这三种估计是不同的, 但当后验密度为单峰对称时,  $\theta$  的上述三种贝叶斯估计重合. 使用时可根据需要选用其中的一种. 一般来说, 当先验分布为共轭先验时, 求上述三种估计比较容易.

## 贝叶斯点估计方法

- 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  $\theta$  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, 其定义如下:
- **定义 4.1.1** 使后验密度  $\pi(\theta|x)$  达到最大时  $\theta$  的值作为  $\theta$  的估计量, 称为后验众数估计, 也称为广义极大似然估计或后验极大似然估计, 记为  $\hat{\theta}_{MD}$ ; 用后验分布的中位数作为  $\theta$  的估计量, 称为  $\theta$  的后验中位数估计, 记为  $\hat{\theta}_{ME}$ ; 用后验分布的期望作为  $\theta$  的估计量, 称为后验期望估计, 记为  $\hat{\theta}_E$ . 三个估计都称为贝叶斯估计, 在不会引起混淆的情况下上述三个估计皆用  $\hat{\theta}_B$  来记.
- **注 4.1.1** 一般场合下这三种估计是不同的, 但当后验密度为单峰对称时,  $\theta$  的上述三种贝叶斯估计重合. 使用时可根据需要选用其中的一种. 一般来说, 当先验分布为共轭先验时, 求上述三种估计比较容易.

## 贝叶斯点估计方法

- 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  $\theta$  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, 其定义如下:
- **定义 4.1.1** 使后验密度  $\pi(\theta|x)$  达到最大时  $\theta$  的值作为  $\theta$  的估计量, 称为后验众数估计, 也称为广义极大似然估计或后验极大似然估计, 记为  $\hat{\theta}_{MD}$ ; 用后验分布的中位数作为  $\theta$  的估计量, 称为  $\theta$  的后验中位数估计, 记为  $\hat{\theta}_{ME}$ ; 用后验分布的期望作为  $\theta$  的估计量, 称为后验期望估计, 记为  $\hat{\theta}_E$ . 三个估计都称为贝叶斯估计, 在不会引起混淆的情况下上述三个估计皆用  $\hat{\theta}_B$  来记.
- **注 4.1.1** 一般场合下这三种估计是不同的, 但当后验密度为单峰对称时,  $\theta$  的上述三种贝叶斯估计重合. 使用时可根据需要选用其中的一种. 一般来说, 当先验分布为共轭先验时, 求上述三种估计比较容易.



## 例子

- **例 4.1.1** 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  $\sigma^2$  已知. 取  $\theta$  的先验为共轭先验分布  $N(\mu, \tau^2)$ , 求  $\theta$  的贝叶斯估计.
- **例 4.1.2** 为估计产品不合格品率  $\theta$ , 今从一批产品中随机抽取  $n$  件检查, 其中不合格品数  $X \sim B(n, \theta)$ , 若取  $\theta$  的先验分布为  $Be(1, 1)$ , 即  $\theta$  的先验分布为  $(0, 1)$  上的均匀分布, 并假定  $x = 0$ , 求  $\theta$  的后验中位数估计.
- **例 4.1.3** 设随机变量  $X \sim f(x|\theta) = e^{-(x-\theta)}$ ,  $\theta < x < \infty$ , 此处  $-\infty < \theta < +\infty$  为位置参数, 取  $\theta$  的先验分布为柯西分布  $C(0, 1)$ , 即先验密度  $\pi(\theta) = \frac{1}{\pi(1+\theta^2)}$ ,  $-\infty < \theta < +\infty$ , 求  $\theta$  的后验众数估计.

## 例子

- **例 4.1.1** 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  $\sigma^2$  已知. 取  $\theta$  的先验为共轭先验分布  $N(\mu, \tau^2)$ , 求  $\theta$  的贝叶斯估计.
- **例 4.1.2** 为估计产品不合格品率  $\theta$ , 今从一批产品中随机抽取  $n$  件检查, 其中不合格品数  $X \sim B(n, \theta)$ , 若取  $\theta$  的先验分布为  $Be(1, 1)$ , 即  $\theta$  的先验分布为  $(0, 1)$  上的均匀分布, 并假定  $x = 0$ , 求  $\theta$  的后验中位数估计.
- **例 4.1.3** 设随机变量  $X \sim f(x|\theta) = e^{-(x-\theta)}$ ,  $\theta < x < \infty$ , 此处  $-\infty < \theta < +\infty$  为位置参数, 取  $\theta$  的先验分布为柯西分布  $C(0, 1)$ , 即先验密度  $\pi(\theta) = \frac{1}{\pi(1+\theta^2)}$ ,  $-\infty < \theta < +\infty$ , 求  $\theta$  的后验众数估计.

## 例子

- **例 4.1.1** 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  $\sigma^2$  已知. 取  $\theta$  的先验为共轭先验分布  $N(\mu, \tau^2)$ , 求  $\theta$  的贝叶斯估计.
- **例 4.1.2** 为估计产品不合格品率  $\theta$ , 今从一批产品中随机抽取  $n$  件检查, 其中不合格品数  $X \sim B(n, \theta)$ , 若取  $\theta$  的先验分布为  $Be(1, 1)$ , 即  $\theta$  的先验分布为  $(0, 1)$  上的均匀分布, 并假定  $x = 0$ , 求  $\theta$  的后验中位数估计.
- **例 4.1.3** 设随机变量  $X \sim f(x|\theta) = e^{-(x-\theta)}$ ,  $\theta < x < \infty$ , 此处  $-\infty < \theta < +\infty$  为位置参数, 取  $\theta$  的先验分布为柯西分布  $C(0, 1)$ , 即先验密度  $\pi(\theta) = \frac{1}{\pi(1+\theta^2)}$ ,  $-\infty < \theta < +\infty$ , 求  $\theta$  的后验众数估计.

## 贝叶斯点估计的精度-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) 的大小, 均方误差越小越好. 对贝叶斯估计  $\delta(x)$ , 衡量优劣用如下定义的后验均方误差 (简记为PMSE)

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

来度量估计量的精度, PMSE 越小越好.

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

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

- 记  $\theta$  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  $\theta$  的后验期望估计使 PMSE 达到最小.

## 贝叶斯点估计的精度-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) 的大小, 均方误差越小越好. 对贝叶斯估计  $\delta(x)$ , 衡量优劣用如下定义的后验均方误差 (简记为PMSE)

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

来度量估计量的精度, PMSE 越小越好.

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

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

- 记  $\theta$  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  $\theta$  的后验期望估计使 PMSE 达到最小.

## 贝叶斯点估计的精度-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) 的大小, 均方误差越小越好. 对贝叶斯估计  $\delta(x)$ , 衡量优劣用如下定义的后验均方误差 (简记为PMSE)

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

来度量估计量的精度, PMSE 越小越好.

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

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

- 记  $\theta$  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  $\theta$  的后验期望估计使 PMSE 达到最小.

# 例子

- 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  $\mu^\pi(\mathbf{x})$  作为  $\theta$  的贝叶斯估计.
- 例 4.1.4 设  $X \sim N(\theta, \sigma^2)$ ,  $\sigma^2$  已知, 设  $\mathbf{X} = (X_1, \dots, X_n)$  是从正态总体  $X$  中抽取的随机样本.
  - (1) 若  $\theta$  的先验分布为无信息先验  $\pi(\theta) \equiv 1$ , 求  $\theta$  的后验期望估计及其后验方差.
  - (2) 若取  $\theta$  的先验分布为共轭先验, 即  $\pi(\theta) \sim N(\mu, \tau^2)$ , 求  $\theta$  的后验期望估计  $\hat{\theta}_E$  的方差  $V^\pi(\mathbf{x})$ , 并将其与  $\theta$  的经典估计  $\delta(\mathbf{x}) = \bar{x}$  的 PMSE 进行比较.

## 例子

- 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  $\mu^\pi(\mathbf{x})$  作为  $\theta$  的贝叶斯估计.
- **例 4.1.4** 设  $X \sim N(\theta, \sigma^2)$ ,  $\sigma^2$  已知, 设  $\mathbf{X} = (X_1, \dots, X_n)$  是从正态总体  $X$  中抽取的随机样本.
  - (1) 若  $\theta$  的先验分布为无信息先验  $\pi(\theta) \equiv 1$ , 求  $\theta$  的后验期望估计及其后验方差.
  - (2) 若取  $\theta$  的先验分布为共轭先验, 即  $\pi(\theta) \sim N(\mu, \tau^2)$ , 求  $\theta$  的后验期望估计  $\hat{\theta}_E$  的方差  $V^\pi(\mathbf{x})$ , 并将其与  $\theta$  的经典估计  $\delta(\mathbf{x}) = \bar{x}$  的 PMSE 进行比较.



## 例子

- 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  $\mu^\pi(\mathbf{x})$  作为  $\theta$  的贝叶斯估计.
- **例 4.1.4** 设  $X \sim N(\theta, \sigma^2)$ ,  $\sigma^2$  已知, 设  $\mathbf{X} = (X_1, \dots, X_n)$  是从正态总体  $X$  中抽取的随机样本.
  - (1) 若  $\theta$  的先验分布为无信息先验  $\pi(\theta) \equiv 1$ , 求  $\theta$  的后验期望估计及其后验方差.
  - (2) 若取  $\theta$  的先验分布为共轭先验, 即  $\pi(\theta) \sim N(\mu, \tau^2)$ , 求  $\theta$  的后验期望估计  $\hat{\theta}_E$  的方差  $V^\pi(\mathbf{x})$ , 并将其与  $\theta$  的经典估计  $\delta(\mathbf{x}) = \bar{x}$  的 PMSE 进行比较.

# 例子

- 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  $\mu^\pi(\mathbf{x})$  作为  $\theta$  的贝叶斯估计.
- 例 4.1.4 设  $X \sim N(\theta, \sigma^2)$ ,  $\sigma^2$  已知, 设  $\mathbf{X} = (X_1, \dots, X_n)$  是从正态总体  $X$  中抽取的随机样本.
  - (1) 若  $\theta$  的先验分布为无信息先验  $\pi(\theta) \equiv 1$ , 求  $\theta$  的后验期望估计及其后验方差.
  - (2) 若取  $\theta$  的先验分布为共轭先验, 即  $\pi(\theta) \sim N(\mu, \tau^2)$ , 求  $\theta$  的后验期望估计  $\hat{\theta}_E$  的方差  $V^\pi(\mathbf{x})$ , 并将其与  $\theta$  的经典估计  $\delta(\mathbf{x}) = \bar{x}$  的 PMSE 进行比较.

## 贝叶斯点估计的几个例子

- **例 4.1.5** 为估计产品不合格品率  $\theta$ , 今从一批产品中随机抽取  $n$  件检查, 检查结果为  $X_1, \dots, X_n$ , 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  $X = \sum_{i=1}^n X_i \sim B(n, \theta)$ . 若取  $\theta$  的先验分布为  $Be(\alpha, \beta)$ .

(1) 求的几个例子  $\theta$  的后验期望估计及其后验方差.

(2) 当  $\alpha = \beta = 1$  时, 求  $\theta$  的后验期望估计和后验众数估计, 并将这两个估计进行比较.

(3) 当  $\alpha = \beta = 1$  时, 求  $\theta$  的后验期望估计和后验众数估计的后验方差或后验均方误差, 并将其进行比较.

- 易知  $\theta$  的后验众数估计为

$$\hat{\theta}_{MD} = \frac{(X + \alpha) - 1}{(X + \alpha) + (n - X + \beta) - 2} = \frac{X + \alpha - 1}{n + \alpha + \beta - 2};$$

## 贝叶斯点估计的几个例子

- **例 4.1.5** 为估计产品不合格品率  $\theta$ , 今从一批产品中随机抽取  $n$  件检查, 检查结果为  $X_1, \dots, X_n$ , 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  $X = \sum_{i=1}^n X_i \sim B(n, \theta)$ . 若取  $\theta$  的先验分布为  $Be(\alpha, \beta)$ .

(1) 求的几个例子  $\theta$  的后验期望估计及其后验方差.

(2) 当  $\alpha = \beta = 1$  时, 求  $\theta$  的后验期望估计和后验众数估计, 并将这两个估计进行比较.

(3) 当  $\alpha = \beta = 1$  时, 求  $\theta$  的后验期望估计和后验众数估计的后验方差或后验均方误差, 并将其进行比较.

- 易知  $\theta$  的后验众数估计为

$$\hat{\theta}_{MD} = \frac{(X + \alpha) - 1}{(X + \alpha) + (n - X + \beta) - 2} = \frac{X + \alpha - 1}{n + \alpha + \beta - 2};$$

## 贝叶斯点估计的几个例子

- **例 4.1.5** 为估计产品不合格品率  $\theta$ , 今从一批产品中随机抽取  $n$  件检查, 检查结果为  $X_1, \dots, X_n$ , 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  $X = \sum_{i=1}^n X_i \sim B(n, \theta)$ . 若取  $\theta$  的先验分布为  $Be(\alpha, \beta)$ .

(1) 求的几个例子  $\theta$  的后验期望估计及其后验方差.

(2) 当  $\alpha = \beta = 1$  时, 求  $\theta$  的后验期望估计和后验众数估计, 并将这两个估计进行比较.

(3) 当  $\alpha = \beta = 1$  时, 求  $\theta$  的后验期望估计和后验众数估计的后验方差或后验均方误差, 并将其进行比较.

- 易知  $\theta$  的后验众数估计为

$$\hat{\theta}_{MD} = \frac{(X + \alpha) - 1}{(X + \alpha) + (n - X + \beta) - 2} = \frac{X + \alpha - 1}{n + \alpha + \beta - 2};$$

## 贝叶斯点估计的几个例子

- **例 4.1.5** 为估计产品不合格品率  $\theta$ , 今从一批产品中随机抽取  $n$  件检查, 检查结果为  $X_1, \dots, X_n$ , 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  $X = \sum_{i=1}^n X_i \sim B(n, \theta)$ . 若取  $\theta$  的先验分布为  $Be(\alpha, \beta)$ .

- (1) 求的几个例子  $\theta$  的后验期望估计及其后验方差.
- (2) 当  $\alpha = \beta = 1$  时, 求  $\theta$  的后验期望估计和后验众数估计, 并将这两个估计进行比较.
- (3) 当  $\alpha = \beta = 1$  时, 求  $\theta$  的后验期望估计和后验众数估计的后验方差或后验均方误差, 并将其进行比较.

- 易知  $\theta$  的后验众数估计为

$$\hat{\theta}_{MD} = \frac{(X + \alpha) - 1}{(X + \alpha) + (n - X + \beta) - 2} = \frac{X + \alpha - 1}{n + \alpha + \beta - 2};$$

## 贝叶斯点估计的几个例子

- **例 4.1.5** 为估计产品不合格品率  $\theta$ , 今从一批产品中随机抽取  $n$  件检查, 检查结果为  $X_1, \dots, X_n$ , 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  $X = \sum_{i=1}^n X_i \sim B(n, \theta)$ . 若取  $\theta$  的先验分布为  $Be(\alpha, \beta)$ .

(1) 求的几个例子  $\theta$  的后验期望估计及其后验方差.

(2) 当  $\alpha = \beta = 1$  时, 求  $\theta$  的后验期望估计和后验众数估计, 并将这两个估计进行比较.

(3) 当  $\alpha = \beta = 1$  时, 求  $\theta$  的后验期望估计和后验众数估计的后验方差或后验均方误差, 并将其进行比较.

- 易知  $\theta$  的后验众数估计为

$$\hat{\theta}_{MD} = \frac{(X + \alpha) - 1}{(X + \alpha) + (n - X + \beta) - 2} = \frac{X + \alpha - 1}{n + \alpha + \beta - 2};$$

## 贝叶斯点估计的几个例子

- $\theta$  的后验期望估计为

$$\hat{\theta}_E = \frac{X + \alpha}{n + \alpha + \beta}.$$

- 当先验分布取  $\alpha = 1, \beta = 1$ , 则  $\pi(\theta)$  是均匀分布  $U(0, 1)$  时有

$$\hat{\theta}_{MD} = X/n, \quad \hat{\theta}_E = (X + 1)/(n + 2).$$

对两个估计作如下说明: (i)  $\theta$  的后验众数估计 (即广义 MLE) 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. 贝叶斯学派对这种现象的看法是: 任何使用经典统计方法的人都自觉或不自觉地使用贝叶斯方法, 即经典统计方法下的许多估计量是特殊先验分布下的贝叶斯估计. (ii)  $\theta$  的后验期望估计要比后验众数更合适一些, 下面列出四个特殊试验结果:



## 贝叶斯点估计的几个例子

- $\theta$  的后验期望估计为

$$\hat{\theta}_E = \frac{X + \alpha}{n + \alpha + \beta}.$$

- 当先验分布取  $\alpha = 1, \beta = 1$ , 则  $\pi(\theta)$  是均匀分布  $U(0, 1)$  时有

$$\hat{\theta}_{MD} = X/n, \quad \hat{\theta}_E = (X + 1)/(n + 2).$$

对两个估计作如下说明: (i)  $\theta$  的后验众数估计 (即广义 MLE) 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. 贝叶斯学派对这种现象的看法是: 任何使用经典统计方法的人都自觉或不自觉地使用贝叶斯方法, 即经典统计方法下的许多估计量是特殊先验分布下的贝叶斯估计. (ii)  $\theta$  的后验期望估计要比后验众数更合适一些, 下面列出四个特殊试验结果:

# 贝叶斯点估计的几个例子

● 表 4.1.1 不合格品率  $\theta$  的两种贝叶斯估计的比较表

试验号	样本量 $n$	不合格品数 $x$	$\hat{\theta}_{MD} = x/n$	$\hat{\theta}_E = (x+1)/(n+2)$
1	3	0	0	0.200
2	10	0	0	0.083
3	3	3	1	0.800
4	10	10	1	0.917

- 1 号试验与 2 号试验各抽 3 个与 10 个, 其中没有一个合格, 这两件事在人们心目中留下的印象是不同的, 后者的质量要比前者更信得过. 但  $\hat{\theta}_{MD}$  皆为 0, 显示不出二者的差别, 而  $\hat{\theta}_E$  可显示出二者的差别.

# 贝叶斯点估计的几个例子

● 表 4.1.1 不合格品率  $\theta$  的两种贝叶斯估计的比较表

试验号	样本量 $n$	不合格品数 $x$	$\hat{\theta}_{MD} = x/n$	$\hat{\theta}_E = (x+1)/(n+2)$
1	3	0	0	0.200
2	10	0	0	0.083
3	3	3	1	0.800
4	10	10	1	0.917

- 1 号试验与 2 号试验各抽 3 个与 10 个, 其中没有一个合格, 这两件事在人们心目中留下的印象是不同的, 后者的质量要比前者更信得过. 但  $\hat{\theta}_{MD}$  皆为 0, 显示不出二者的差别, 而  $\hat{\theta}_E$  可显示出二者的差别.

## 贝叶斯点估计的几个例子

- 对 3 号和 4 号试验, 也有类似问题. 显然我们看到, 当  $x = 0$  或  $x = n$  时,  $\hat{\theta}_{MD} = 0$  或  $1$ , 这种估计未免太极端一些, 而  $\hat{\theta}_E$  可以看作是对它的修正, 更合理些. 所以在实际中人们更愿意选用后验期望估计作为贝叶斯估计. 由于  $\hat{\theta}_{MD}$  与经典估计相同, 故贝叶斯估计  $\hat{\theta}_E$  显示出相对于经典估计的优点.
- 故可见: 作为  $\theta$  的估计, 后验期望估计常优于后验众数估计.
- 若取先验分布  $\pi(\theta)$  为  $U(0, 1)$ , 则  $\theta$  的后验分布为贝塔分布  $Be(x + 1, n - x + 1)$ , 易知  $\hat{\theta}_E$  的后验方差为

$$V^{\pi}(x) = \frac{(x+1)(n-x+1)}{(n+2)^2(n+3)},$$

## 贝叶斯点估计的几个例子

- 对 3 号和 4 号试验, 也有类似问题. 显然我们看到, 当  $x = 0$  或  $x = n$  时,  $\hat{\theta}_{MD} = 0$  或  $1$ , 这种估计未免太极端一些, 而  $\hat{\theta}_E$  可以看作是对它的修正, 更合理些. 所以在实际中人们更愿意选用后验期望估计作为贝叶斯估计. 由于  $\hat{\theta}_{MD}$  与经典估计相同, 故贝叶斯估计  $\hat{\theta}_E$  显示出相对于经典估计的优点.
- 故可见: 作为  $\theta$  的估计, 后验期望估计常优于后验众数估计.
- 若取先验分布  $\pi(\theta)$  为  $U(0, 1)$ , 则  $\theta$  的后验分布为贝塔分布  $Be(x + 1, n - x + 1)$ , 易知  $\hat{\theta}_E$  的后验方差为

$$V^{\pi}(x) = \frac{(x+1)(n-x+1)}{(n+2)^2(n+3)},$$

## 贝叶斯点估计的几个例子

- 对 3 号和 4 号试验, 也有类似问题. 显然我们看到, 当  $x = 0$  或  $x = n$  时,  $\hat{\theta}_{MD} = 0$  或  $1$ , 这种估计未免太极端一些, 而  $\hat{\theta}_E$  可以看作是对它的修正, 更合理些. 所以在实际中人们更愿意选用后验期望估计作为贝叶斯估计. 由于  $\hat{\theta}_{MD}$  与经典估计相同, 故贝叶斯估计  $\hat{\theta}_E$  显示出相对于经典估计的优点.
- 故可见: 作为  $\theta$  的估计, 后验期望估计常优于后验众数估计.
- 若取先验分布  $\pi(\theta)$  为  $U(0, 1)$ , 则  $\theta$  的后验分布为贝塔分布  $Be(x + 1, n - x + 1)$ , 易知  $\hat{\theta}_E$  的后验方差为

$$V^{\pi}(x) = \frac{(x + 1)(n - x + 1)}{(n + 2)^2(n + 3)},$$

## 贝叶斯点估计的几个例子

- 对 3 号和 4 号试验, 也有类似问题. 显然我们看到, 当  $x = 0$  或  $x = n$  时,  $\hat{\theta}_{MD} = 0$  或  $1$ , 这种估计未免太极端一些, 而  $\hat{\theta}_E$  可以看作是对它的修正, 更合理些. 所以在实际中人们更愿意选用后验期望估计作为贝叶斯估计. 由于  $\hat{\theta}_{MD}$  与经典估计相同, 故贝叶斯估计  $\hat{\theta}_E$  显示出相对于经典估计的优点.
- 故可见: 作为  $\theta$  的估计, 后验期望估计常优于后验众数估计.
- 若取先验分布  $\pi(\theta)$  为  $U(0, 1)$ , 则  $\theta$  的后验分布为贝塔分布  $Be(x + 1, n - x + 1)$ , 易知  $\hat{\theta}_E$  的后验方差为

$$V^{\pi}(x) = \frac{(x + 1)(n - x + 1)}{(n + 2)^2(n + 3)},$$

# 贝叶斯点估计的几个例子

- $\hat{\theta}_{MD}$  的后验均方误差为

$$\begin{aligned} PMSE(\hat{\theta}_{MD}) &= V^{\pi}(x) + (\hat{\theta}_{MD} - \hat{\theta}_E)^2 \\ &= \frac{(x+1)(n-x+1)}{(n+2)^2(n+3)} + \left( \frac{x+1}{n+2} - \frac{x}{n} \right)^2 \geq V^{\pi}(x). \end{aligned}$$

- 表 4.1.2 是根据 4 对  $(n, x)$  的值算得的  $\hat{\theta}_E$  和  $\hat{\theta}_{MD}$  的后验方差和后验均方误差的值:

表 4.1.2  $\hat{\theta}_E$  和  $\hat{\theta}_{MD}$  的后验方差和后验均方误差

n	x	$\hat{\theta}_E$	$V^{\pi}(x)$	$\hat{\theta}_{MD}$	PMSE
3	0	1/5	0.02667	0	0.06667
10	0	1/12	0.00588	0	0.01282
10	1	2/12	0.01068	1/10	0.01512
20	1	2/22	0.00359	1/20	0.00527



## 贝叶斯点估计的几个例子

- 从表 4.1.2 中可见后验期望估计的  $V^\pi(x)$  小于后验众数估计的 PMSE, 由此表还可见样本量的增加有利于后验方差和后验均方误差的减小.
- 例 4.1.6 设  $X_1, \dots, X_n \text{ iid} \sim f(x|\theta) = \theta^{-1} \exp\{-x/\theta\}$ ,  $x > 0$ ,  $\theta > 0$ , 当观察到前  $r$  个样本  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(r)}$  时就停止, 此处  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  为样本  $X_1, \dots, X_n$  的次序统计量, 取  $\theta$  的先验分布为无信息先验  $\pi(\theta) = 1/\theta$ ,  $\theta > 0$ , 求可靠性函数  $R(s) = P(X \geq s) = e^{-\frac{s}{\theta}}$ ,  $s > 0$  的后验期望估计及其后验方差.
- 下面的例子给出了样本分布和先验分布都是离散分布时, 如何求贝叶斯估计.

## 贝叶斯点估计的几个例子

- 从表 4.1.2 中可见后验期望估计的  $V^\pi(x)$  小于后验众数估计的 PMSE, 由此表还可见样本量的增加有利于后验方差和后验均方误差的减小.
- 例 4.1.6** 设  $X_1, \dots, X_n \text{ iid} \sim f(x|\theta) = \theta^{-1} \exp\{-x/\theta\}$ ,  $x > 0$ ,  $\theta > 0$ , 当观察到前  $r$  个样本  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(r)}$  时就停止, 此处  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  为样本  $X_1, \dots, X_n$  的次序统计量, 取  $\theta$  的先验分布为无信息先验  $\pi(\theta) = 1/\theta$ ,  $\theta > 0$ , 求可靠性函数  $R(s) = P(X \geq s) = e^{-\frac{s}{\theta}}$ ,  $s > 0$  的后验期望估计及其后验方差.
- 下面的例子给出了样本分布和先验分布都是离散分布时, 如何求贝叶斯估计.

## 贝叶斯点估计的几个例子

- 从表 4.1.2 中可见后验期望估计的  $V^\pi(x)$  小于后验众数估计的 PMSE, 由此表还可见样本量的增加有利于后验方差和后验均方误差的减小.
- 例 4.1.6** 设  $X_1, \dots, X_n \text{ iid} \sim f(x|\theta) = \theta^{-1} \exp\{-x/\theta\}$ ,  $x > 0$ ,  $\theta > 0$ , 当观察到前  $r$  个样本  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(r)}$  时就停止, 此处  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  为样本  $X_1, \dots, X_n$  的次序统计量, 取  $\theta$  的先验分布为无信息先验  $\pi(\theta) = 1/\theta$ ,  $\theta > 0$ , 求可靠性函数  $R(s) = P(X \geq s) = e^{-\frac{s}{\theta}}$ ,  $s > 0$  的后验期望估计及其后验方差.
- 下面的例子给出了样本分布和先验分布都是离散分布时, 如何求贝叶斯估计.

## 离散情形的例子

- **例 4.1.7** 设一批产品不合格率为  $\theta$ , 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  $\theta$  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  $\theta$  的后验众数估计, 并计算它的后验均方误差.
- 将后验分布列表如下:

$\theta$	$1/4$	$2/4$	$3/4$
$P(\theta = i/4   X = 3)$	$9/20$	$8/20$	$3/20$

## 离散情形的例子

- **例 4.1.7** 设一批产品不合格率为  $\theta$ , 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  $\theta$  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  $\theta$  的后验众数估计, 并计算它的后验均方误差.
- 将后验分布列表如下:

$\theta$	$1/4$	$2/4$	$3/4$
$P(\theta = i/4   X = 3)$	$9/20$	$8/20$	$3/20$

## 一个应用实例

**例 4.1.8** 保险公司原本没有设立车祸保险项目. 现考虑增设此项目, 需要了解车祸发生情况. 设一年中每千人开车者发生车祸的次数  $X|\lambda \sim P(\lambda)$ , 即参数为  $\lambda$  的泊松分布. 保险公司认为  $\lambda$  的先验分布为伽玛分布  $\Gamma(35, 1)$ . 为慎重起见, 通过对愿意购买车祸保险的 2000 人进行调查, 发现他们一年中有 85 人次的车祸发生. 假定每次车祸, 保险公司平均支付 1000 元人民币. 保险公司增设广告宣传此项保险, 一年费用 50 万人民币. 若每张保单收保险费 50 元人民币, 一年可卖出保单 10 万张, 按这样收费, 保险公司一年能盈利多少?

## 多参数情形的后验分布

- 若  $\theta = (\theta_1, \dots, \theta_p)^\tau$  是向量,  $\theta$  的后验分布为  $\pi(\theta|x)$ , 估计  $\theta$  的方法如下:
  - (1) 后验众数估计: 即从后验分布出发用广义极大似然估计获得使后验密度达到极大时  $\theta$  的值作为后验众数估计.
  - (2) 后验期望估计:  $\mu^\pi(x) = E^{\theta|x}(\theta) = (\mu_1^\pi(x), \dots, \mu_p^\pi(x))^\tau$ . 估计量的精度用后验协方差阵 (记为  $\text{Cov}^\pi(x)$ ) 来衡量

$$\text{Cov}^\pi(x) = E^{\theta|x}[(\theta - \mu^\pi(x))(\theta - \mu^\pi(x))^\tau].$$

- 对  $\theta$  的任一估计  $\delta(x)$ , 其后验协方差阵可分解为

$$\begin{aligned}\text{Cov}(\delta) &= E^{\theta|x}[(\theta - \delta(x))(\theta - \delta(x))^\tau] \\ &= \text{Cov}^\pi(x) + (\mu^\pi(x) - \delta(x))(\mu^\pi(x) - \delta(x))^\tau \geq \text{Cov}^\pi(x).\end{aligned}$$

可见后验均值估计仍使后验协方差矩阵达到最小.

## 多参数情形的后验分布

- 若  $\theta = (\theta_1, \dots, \theta_p)^\tau$  是向量,  $\theta$  的后验分布为  $\pi(\theta|x)$ , 估计  $\theta$  的方法如下:
- (1) 后验众数估计: 即从后验分布出发用广义极大似然估计获得使后验密度达到极大时  $\theta$  的值作为后验众数估计.
- (2) 后验期望估计:  $\mu^\pi(x) = E^{\theta|x}(\theta) = (\mu_1^\pi(x), \dots, \mu_p^\pi(x))^\tau$ . 估计量的精度用后验协方差阵 (记为  $Cov^\pi(x)$ ) 来衡量

$$Cov^\pi(x) = E^{\theta|x}[(\theta - \mu^\pi(x))(\theta - \mu^\pi(x))^\tau].$$

- 对  $\theta$  的任一估计  $\delta(x)$ , 其后验协方差阵可分解为

$$\begin{aligned} Cov(\delta) &= E^{\theta|x}[(\theta - \delta(x))(\theta - \delta(x))^\tau] \\ &= Cov^\pi(x) + (\mu^\pi(x) - \delta(x))(\mu^\pi(x) - \delta(x))^\tau \geq Cov^\pi(x). \end{aligned}$$

可见后验均值估计仍使后验协方差矩阵达到最小.



## 多参数情形的后验分布

- 若  $\theta = (\theta_1, \dots, \theta_p)^\tau$  是向量,  $\theta$  的后验分布为  $\pi(\theta|x)$ , 估计  $\theta$  的方法如下:
- (1) 后验众数估计: 即从后验分布出发用广义极大似然估计获得使后验密度达到极大时  $\theta$  的值作为后验众数估计.
- (2) 后验期望估计:  $\mu^\pi(x) = E^{\theta|x}(\theta) = (\mu_1^\pi(x), \dots, \mu_p^\pi(x))^\tau$ . 估计量的精度用后验协方差阵 (记为  $Cov^\pi(x)$ ) 来衡量

$$Cov^\pi(x) = E^{\theta|x}[(\theta - \mu^\pi(x))(\theta - \mu^\pi(x))^\tau].$$

- 对  $\theta$  的任一估计  $\delta(x)$ , 其后验协方差阵可分解为

$$\begin{aligned} Cov(\delta) &= E^{\theta|x}[(\theta - \delta(x))(\theta - \delta(x))^\tau] \\ &= Cov^\pi(x) + (\mu^\pi(x) - \delta(x))(\mu^\pi(x) - \delta(x))^\tau \geq Cov^\pi(x). \end{aligned}$$

可见后验均值估计仍使后验协方差矩阵达到最小.

## 多参数情形的后验分布

- 若  $\theta = (\theta_1, \dots, \theta_p)^\tau$  是向量,  $\theta$  的后验分布为  $\pi(\theta|x)$ , 估计  $\theta$  的方法如下:
- (1) 后验众数估计: 即从后验分布出发用广义极大似然估计获得使后验密度达到极大时  $\theta$  的值作为后验众数估计.
- (2) 后验期望估计:  $\mu^\pi(x) = E^{\theta|x}(\theta) = (\mu_1^\pi(x), \dots, \mu_p^\pi(x))^\tau$ . 估计量的精度用后验协方差阵 (记为  $Cov^\pi(x)$ ) 来衡量

$$Cov^\pi(x) = E^{\theta|x}[(\theta - \mu^\pi(x))(\theta - \mu^\pi(x))^\tau].$$

- 对  $\theta$  的任一估计  $\delta(x)$ , 其后验协方差阵可分解为

$$\begin{aligned} Cov(\delta) &= E^{\theta|x}[(\theta - \delta(x))(\theta - \delta(x))^\tau] \\ &= Cov^\pi(x) + (\mu^\pi(x) - \delta(x))(\mu^\pi(x) - \delta(x))^\tau \geq Cov^\pi(x). \end{aligned}$$

可见后验均值估计仍使后验协方差矩阵达到最小.

1

## 4.1 贝叶斯点估计

- 4.1.1 条件方法
- 4.1.2 贝叶斯点估计
- 4.1.3 贝叶斯点估计的精度-估计的误差
- 4.1.4 贝叶斯点估计的几个例子
- 4.1.5 多参数情形

2

## 4.2 贝叶斯区间估计

- 4.2.1 可信区间定义及例子
- 4.2.2 最大后验密度 (HPD) 可信区间
- 4.2.3 大样本方法

3

## 4.3 假设检验

- 4.3.1 一般方法
- 4.3.2 贝叶斯因子
- 4.3.3 简单假设对简单假设情形
- 4.3.4 复杂假设对复杂假设情形
- 4.3.5 简单假设对复杂假设
- 4.3.6 多重假设检验

# 贝叶斯可信区间的定义

- **定义 4.2.1** (可信区间) 设参数  $\theta$  的后验分布为  $\pi(\theta|x)$ , 对给定的样本  $x$  和概率  $0 < \alpha < 1$  (通常  $\alpha$  取较小的数), 若存在两个统计量  $\hat{\theta}_1(x)$  和  $\hat{\theta}_2(x)$ , 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  $\theta$  的可信水平为  $1 - \alpha$  的贝叶斯可信区间.

- 而满足

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

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

# 贝叶斯可信区间的定义

- **定义 4.2.1** (可信区间) 设参数  $\theta$  的后验分布为  $\pi(\theta|x)$ , 对给定的样本  $x$  和概率  $0 < \alpha < 1$  (通常  $\alpha$  取较小的数), 若存在两个统计量  $\hat{\theta}_1(x)$  和  $\hat{\theta}_2(x)$ , 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  $\theta$  的可信水平为  $1 - \alpha$  的贝叶斯可信区间.

- 而满足

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

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

## 贝叶斯可信区间的定义

- **定义 4.2.1** (可信区间) 设参数  $\theta$  的后验分布为  $\pi(\theta|x)$ , 对给定的样本  $x$  和概率  $0 < \alpha < 1$  (通常  $\alpha$  取较小的数), 若存在两个统计量  $\hat{\theta}_1(x)$  和  $\hat{\theta}_2(x)$ , 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  $\theta$  的可信水平为  $1 - \alpha$  的贝叶斯可信区间.

- 而满足

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

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

## 贝叶斯可信区间与置信区间的本质差别

- 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  $1 - \alpha = 0.9$  的可信区间为  $[1.2, 2.0]$ , 这时我们可以写为

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

既可以说“ $\theta$  属于这个区间的概率为 0.9”, 也可以说“ $\theta$  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  $\theta$  为常数, 它要么  $\theta$  在  $[1.2, 2.0]$  之内, 要么在其外, 不能说“ $\theta$  落在  $[1.2, 2.0]$  中的概率为 0.90”, 而只能说“100 次重复使用此置信区间, 大约有 90 次能覆盖  $\theta$ ”. 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. 故贝叶斯可信区间简单、自然易被人们接受.

- 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, 这是一项技术性很强的工作, 有时找枢轴变量的分布相当困难. 而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.

## 贝叶斯可信区间与置信区间的本质差别

- 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  $1 - \alpha = 0.9$  的可信区间为  $[1.2, 2.0]$ , 这时我们可以写为

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

既可以说“ $\theta$  属于这个区间的概率为 0.9”, 也可以说“ $\theta$  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  $\theta$  为常数, 它要么  $\theta$  在  $[1.2, 2.0]$  之内, 要么在其外, 不能说“ $\theta$  落在  $[1.2, 2.0]$  中的概率为 0.90”, 而只能说“100 次重复使用此置信区间, 大约有 90 次能覆盖  $\theta$ ”. 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. 故贝叶斯可信区间简单、自然易被人们接受.

- 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, 这是一项技术性很强的工作, 有时找枢轴变量的分布相当困难. 而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.



# 例子

- **例 4.2.1** 设  $X_1, \dots, X_n$  是从正态总体  $N(\theta, \sigma^2)$  中抽取的简单样本, 其中  $\sigma^2$  已知.  $\theta$  的先验分布是  $N(\mu, \tau^2)$ , 其中  $\mu, \tau^2$  已知. 求  $\theta$  的可信水平  $1 - \alpha$  的可信区间.
- 看一个数值例子. 在儿童智商测验的例子中, 设测验结果  $X \sim N(\theta, 100)$ , 其中  $\theta$  为被测验儿童智商 IQ 真值,  $\theta$  的先验分布是  $N(100, 225)$ , 当该儿童测验得分  $x = 115$  时, 由例 4.1.1 的结果可知  $\theta$  的后验分布是  $N(110.38, 8.32^2)$ , 因此求得  $\theta$  的可信系数 0.95 的可信区间为

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

- 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  $\theta$  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.

# 例子

- **例 4.2.1** 设  $X_1, \dots, X_n$  是从正态总体  $N(\theta, \sigma^2)$  中抽取的简单样本, 其中  $\sigma^2$  已知.  $\theta$  的先验分布是  $N(\mu, \tau^2)$ , 其中  $\mu, \tau^2$  已知. 求  $\theta$  的可信水平  $1 - \alpha$  的可信区间.
- 看一个数值例子. 在儿童智商测验的例子中, 设测验结果  $X \sim N(\theta, 100)$ , 其中  $\theta$  为被测验儿童智商 IQ 真值,  $\theta$  的先验分布是  $N(100, 225)$ , 当该儿童测验得分  $x = 115$  时, 由例 4.1.1 的结果可知  $\theta$  的后验分布是  $N(110.38, 8.32^2)$ , 因此求得  $\theta$  的可信系数 0.95 的可信区间为

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

- 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  $\theta$  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.

## 例子

- **例 4.2.1** 设  $X_1, \dots, X_n$  是从正态总体  $N(\theta, \sigma^2)$  中抽取的简单样本, 其中  $\sigma^2$  已知.  $\theta$  的先验分布是  $N(\mu, \tau^2)$ , 其中  $\mu, \tau^2$  已知. 求  $\theta$  的可信水平  $1 - \alpha$  的可信区间.
- 看一个数值例子. 在儿童智商测验的例子中, 设测验结果  $X \sim N(\theta, 100)$ , 其中  $\theta$  为被测验儿童智商 IQ 真值,  $\theta$  的先验分布是  $N(100, 225)$ , 当该儿童测验得分  $x = 115$  时, 由例 4.1.1 的结果可知  $\theta$  的后验分布是  $N(110.38, 8.32^2)$ , 因此求得  $\theta$  的可信系数 0.95 的可信区间为

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

- 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  $\theta$  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.

# 例子

- **例 4.2.2** 彩色电视机的寿命服从指数分布  $Exp\{1/\theta\}$ , 其密度函数为

$$f(x|\theta) = \theta^{-1} \exp\{-x/\theta\} \cdot I_{(0,\infty)}(x),$$

其中  $\theta > 0$  为彩电的平均寿命. 现从一批彩电中随机抽取  $n$  台进行寿命试验, 试验进行到第  $r$  ( $1 \leq r \leq n$ ) 台失效为止, 其失效时间为  $t_1 \leq t_2 \leq \cdots \leq t_r$ , 这样的试验称为定数截尾寿命试验, 所得样本  $(t_1, t_2, \cdots, t_r)$  称为定数截尾样本, 假定  $\theta$  的先验分布为逆伽玛分布  $\Gamma^{-1}(\alpha, \beta)$ , 求此批彩电的平均寿命和平均寿命的可信下限.

- 下面是一个具体数字的例子. 为求出彩电平均寿命  $\theta$  的贝叶斯估计的具体值, 需要确定先验分布中超参数  $\alpha$  和  $\beta$  的值. 我国彩电生产实验室和一些独立的实验室收集到 13142 台彩电寿命试验的数据, 共计 5369812 台时, 此外还有 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, 试验中总共失效台数不超过 250 台.

## 例子

- **例 4.2.2** 彩色电视机的寿命服从指数分布  $Exp\{1/\theta\}$ , 其密度函数为

$$f(x|\theta) = \theta^{-1} \exp\{-x/\theta\} \cdot I_{(0,\infty)}(x),$$

其中  $\theta > 0$  为彩电的平均寿命. 现从一批彩电中随机抽取  $n$  台进行寿命试验, 试验进行到第  $r$  ( $1 \leq r \leq n$ ) 台失效为止, 其失效时间为  $t_1 \leq t_2 \leq \cdots \leq t_r$ , 这样的试验称为定数截尾寿命试验, 所得样本  $(t_1, t_2, \cdots, t_r)$  称为定数截尾样本, 假定  $\theta$  的先验分布为逆伽玛分布  $\Gamma^{-1}(\alpha, \beta)$ , 求此批彩电的平均寿命和平均寿命的可信下限.

- 下面是一个具体数字的例子. 为求出彩电平均寿命  $\theta$  的贝叶斯估计的具体值, 需要确定先验分布中超参数  $\alpha$  和  $\beta$  的值. 我国彩电生产实验室和一些独立的实验室收集到 13142 台彩电寿命试验的数据, 共计 5369812 台时, 此外还有 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, 试验中总共失效台数不超过 250 台.

## 例子

- 对如此大量先验信息加工整理后，确认我国彩电平均寿命不低于 30000 小时，它的 10% 的分位数  $\theta_{0.1}$  大约为 11250 小时，经过一些专家认定，这两个数据符合我国八十年代彩电寿命的实际情况，也是留有余地的。
- 由此可列出如下两个方程

$$\begin{cases} \frac{\beta}{\alpha - 1} = 30000; \\ \int_0^{11250} \pi(\theta) d\theta = 0.1, \end{cases}$$

其中第一个方程式由先验分布为逆伽玛分布  $\Gamma^{-1}(\alpha, \beta)$  的数学期望  $E(\theta) = \beta/(\alpha - 1)$  确定的。在计算机上解此方程组得

$$\alpha = 2.9914, \quad \beta = 59742.$$

- 由超参数  $\alpha$  和  $\beta$  的确定，得到  $\theta$  的先验分布  $\Gamma^{-1}(2.9914, 59742)$ ，和后验分布  $\theta|t \sim \Gamma^{-1}(r + 2.9914, t + 59742)$ 。

## 例子

- 对如此大量先验信息加工整理后，确认我国彩电平均寿命不低于 30000 小时，它的 10% 的分位数  $\theta_{0.1}$  大约为 11250 小时，经过一些专家认定，这两个数据符合我国八十年代彩电寿命的实际情况，也是留有余地的。
- 由此可列出如下两个方程

$$\begin{cases} \frac{\beta}{\alpha - 1} = 30000; \\ \int_0^{11250} \pi(\theta) d\theta = 0.1, \end{cases}$$

其中第一个方程式由先验分布为逆伽玛分布  $\Gamma^{-1}(\alpha, \beta)$  的数学期望  $E(\theta) = \beta/(\alpha - 1)$  确定的。在计算机上解此方程组得

$$\alpha = 2.9914, \quad \beta = 59742.$$

- 由超参数  $\alpha$  和  $\beta$  的确定，得到  $\theta$  的先验分布  $\Gamma^{-1}(2.9914, 59742)$ ，和后验分布  $\theta|t \sim \Gamma^{-1}(r + 2.9914, t + 59742)$ 。

## 例子

- 对如此大量先验信息加工整理后，确认我国彩电平均寿命不低于 30000 小时，它的 10% 的分位数  $\theta_{0.1}$  大约为 11250 小时，经过一些专家认定，这两个数据符合我国八十年代彩电寿命的实际情况，也是留有余地的。
- 由此可列出如下两个方程

$$\begin{cases} \frac{\beta}{\alpha - 1} = 30000; \\ \int_0^{11250} \pi(\theta) d\theta = 0.1, \end{cases}$$

其中第一个方程式由先验分布为逆伽玛分布  $\Gamma^{-1}(\alpha, \beta)$  的数学期望  $E(\theta) = \beta/(\alpha - 1)$  确定的。在计算机上解此方程组得

$$\alpha = 2.9914, \quad \beta = 59742.$$

- 由超参数  $\alpha$  和  $\beta$  的确定，得到  $\theta$  的先验分布  $\Gamma^{-1}(2.9914, 59742)$ ，和后验分布  $\theta|t \sim \Gamma^{-1}(r + 2.9914, t + 59742)$ 。



## HPD 可信区间的定义

- 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：
- 定义 4.2.2 设参数  $\theta$  的后验密度为  $\pi(\theta|x)$ ，对给定的概率  $1 - \alpha$  ( $0 < \alpha < 1$ )，集合  $C$  满足如下条件：
  - (i)  $P(\theta \in C | x) = 1 - \alpha$ ,
  - (ii) 对任给的  $\theta_1 \in C$  和  $\theta_2 \notin C$ ，总有  $\pi(\theta_1 | x) > \pi(\theta_2 | x)$ ，则称  $C$  为  $\theta$  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)。

## HPD 可信区间的定义

- 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：
- **定义 4.2.2** 设参数  $\theta$  的后验密度为  $\pi(\theta|x)$ ，对给定的概率  $1 - \alpha$  ( $0 < \alpha < 1$ )，集合  $C$  满足如下条件：
  - (i)  $P(\theta \in C | x) = 1 - \alpha$ ,
  - (ii) 对任给的  $\theta_1 \in C$  和  $\theta_2 \notin C$ ，总有  $\pi(\theta_1 | x) > \pi(\theta_2 | x)$ ，则称  $C$  为  $\theta$  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)。

## HPD 可信区间的定义

- 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：
- **定义 4.2.2** 设参数  $\theta$  的后验密度为  $\pi(\theta|x)$ ，对给定的概率  $1 - \alpha$  ( $0 < \alpha < 1$ )，集合  $C$  满足如下条件：
  - (i)  $P(\theta \in C | x) = 1 - \alpha$ ,
  - (ii) 对任给的  $\theta_1 \in C$  和  $\theta_2 \notin C$ ，总有  $\pi(\theta_1 | x) > \pi(\theta_2 | x)$ ，则称  $C$  为  $\theta$  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)。

## HPD 可信区间的定义

- 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：
- **定义 4.2.2** 设参数  $\theta$  的后验密度为  $\pi(\theta|x)$ ，对给定的概率  $1 - \alpha$  ( $0 < \alpha < 1$ )，集合  $C$  满足如下条件：
  - (i)  $P(\theta \in C | x) = 1 - \alpha$ ,
  - (ii) 对任给的  $\theta_1 \in C$  和  $\theta_2 \notin C$ ，总有  $\pi(\theta_1 | x) > \pi(\theta_2 | x)$ ，则称  $C$  为  $\theta$  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)。

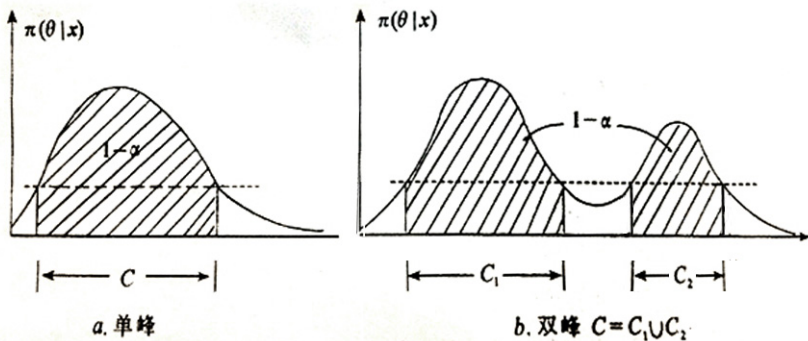
## HPD 可信区间的定义

- 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：
- **定义 4.2.2** 设参数  $\theta$  的后验密度为  $\pi(\theta|x)$ ，对给定的概率  $1 - \alpha$  ( $0 < \alpha < 1$ )，集合  $C$  满足如下条件：
  - (i)  $P(\theta \in C | x) = 1 - \alpha$ ,
  - (ii) 对任给的  $\theta_1 \in C$  和  $\theta_2 \notin C$ ，总有  $\pi(\theta_1 | x) > \pi(\theta_2 | x)$ ，则称  $C$  为  $\theta$  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)。

## 4.2 贝叶斯区间估计

## 4.2.2 最大后验密度 (HPD) 可信区间

图 4.2.1



# HPD 可信区间的定义

- **注 4.2.1** 当后验密度  $\pi(\theta|x)$  为单峰时, HPD 可信区间总存在; 当后验密度为多峰时, 可能得到几个互不连接的区间组成 HPD 可信区间 (见图 4.2.1).
- 当后验分布为单峰对称时,  $1 - \alpha$  的 HPD 可信区间易求, 它就是等尾的区间. 当后验密度虽为单峰, 但不对称时, 寻求 HPD 可信区间并不容易, 可用计算机进行数值计算. 例如, 当后验密度  $\pi(\theta|x)$  为单峰连续函数时, 可按下述方法逐步逼近, 获得  $\theta$  的  $1 - \alpha$  的 HPD 可信区间.

## HPD 可信区间的定义

- **注 4.2.1** 当后验密度  $\pi(\theta|x)$  为单峰时, HPD 可信区间总存在; 当后验密度为多峰时, 可能得到几个互不连接的区间组成 HPD 可信区间 (见图 4.2.1).
- 当后验分布为单峰对称时,  $1 - \alpha$  的 HPD 可信区间易求, 它就是等尾的区间. 当后验密度虽为单峰, 但不对称时, 寻求 HPD 可信区间并不容易, 可用计算机进行数值计算. 例如, 当后验密度  $\pi(\theta|x)$  为单峰连续函数时, 可按下述方法逐步逼近, 获得  $\theta$  的  $1 - \alpha$  的 HPD 可信区间.



## HPD 可信区间数值计算的程序

- (1) 对给定的  $k$  建立子程序：解方程  $\pi(\theta|x) = k$ , 解得  $\theta_1(k)$  和  $\theta_2(k)$ , 从而组成一个区间

$$C(k) = [\theta_1(k), \theta_2(k)] = \{\theta : \pi(\theta|x) \geq k\}.$$

- (2) 建立第二个子程序, 用来计算概率:

$$P(\theta \in C(k) | x) = \int_{C(k)} \pi(\theta|x) d\theta.$$

- (3) 对给定的  $k$  若  $P(\theta \in C(k) | x) \approx 1 - \alpha$ , 则  $C(k)$  即为所求的 HPD 可信区间.
- 若  $P(\theta \in C(k) | x) > 1 - \alpha$ , 则增大  $k$ , 再转入 (1) 与 (2);  
若  $P(\theta \in C(k) | x) < 1 - \alpha$ , 则减小  $k$ , 再转入 (1) 与 (2).

## HPD 可信区间数值计算的程序

- (1) 对给定的  $k$  建立子程序：解方程  $\pi(\theta|x) = k$ , 解得  $\theta_1(k)$  和  $\theta_2(k)$ , 从而组成一个区间

$$C(k) = [\theta_1(k), \theta_2(k)] = \{\theta : \pi(\theta|x) \geq k\}.$$

- (2) 建立第二个子程序，用来计算概率：

$$P(\theta \in C(k) | x) = \int_{C(k)} \pi(\theta|x) d\theta.$$

- (3) 对给定的  $k$  若  $P(\theta \in C(k) | x) \approx 1 - \alpha$ , 则  $C(k)$  即为所求的 HPD 可信区间.
- 若  $P(\theta \in C(k) | x) > 1 - \alpha$ , 则增大  $k$ , 再转入 (1) 与 (2);  
若  $P(\theta \in C(k) | x) < 1 - \alpha$ , 则减小  $k$ , 再转入 (1) 与 (2).

## HPD 可信区间数值计算的程序

- (1) 对给定的  $k$  建立子程序：解方程  $\pi(\theta|x) = k$ , 解得  $\theta_1(k)$  和  $\theta_2(k)$ , 从而组成一个区间

$$C(k) = [\theta_1(k), \theta_2(k)] = \{\theta : \pi(\theta|x) \geq k\}.$$

- (2) 建立第二个子程序，用来计算概率：

$$P(\theta \in C(k) | x) = \int_{C(k)} \pi(\theta|x) d\theta.$$

- (3) 对给定的  $k$  若  $P(\theta \in C(k) | x) \approx 1 - \alpha$ , 则  $C(k)$  即为所求的 HPD 可信区间.
- 若  $P(\theta \in C(k) | x) > 1 - \alpha$ , 则增大  $k$ , 再转入 (1) 与 (2);  
若  $P(\theta \in C(k) | x) < 1 - \alpha$ , 则减小  $k$ , 再转入 (1) 与 (2).

## HPD 可信区间数值计算的程序

- (1) 对给定的  $k$  建立子程序：解方程  $\pi(\theta|x) = k$ , 解得  $\theta_1(k)$  和  $\theta_2(k)$ , 从而组成一个区间

$$C(k) = [\theta_1(k), \theta_2(k)] = \{\theta : \pi(\theta|x) \geq k\}.$$

- (2) 建立第二个子程序，用来计算概率：

$$P(\theta \in C(k) | x) = \int_{C(k)} \pi(\theta|x) d\theta.$$

- (3) 对给定的  $k$  若  $P(\theta \in C(k) | x) \approx 1 - \alpha$ , 则  $C(k)$  即为所求的 HPD 可信区间.
- 若  $P(\theta \in C(k) | x) > 1 - \alpha$ , 则增大  $k$ , 再转入 (1) 与 (2);  
若  $P(\theta \in C(k) | x) < 1 - \alpha$ , 则减小  $k$ , 再转入 (1) 与 (2).

# 例子

- **例 4.2.3** 设  $X_1, \dots, X_n \text{ iid} \sim N(\theta, \sigma^2)$ ,  $\sigma^2$  已知,  $\theta$  的先验分布为无信息先验, 即  $\pi(\theta) \equiv 1$ ,  $\theta \in R_1$ , 求  $\theta$  的  $1-\alpha$  的 HPD 可信区间.
- **例 4.2.4** 设  $X_1, \dots, X_n$  为从柯西分布  $C(\theta, 1)$ ,  $\theta > 0$  中抽取的简单样本, 取先验  $\pi(\theta) \equiv 1$ ,  $\theta \in R_1$ , 求  $\theta$  的  $1-\alpha$  的 HPD 可信区间.
- 这是一个很不容易计算的后验分布, 找  $1-\alpha$  的 HPD 可信区间由计算机进行很容易. 如  $n = 5$ ,  $\mathbf{x} = (4.0, 5.5, 7.5, 4.5, 3.0)$  则 95% HPD 可信区间为 (3.10, 6.06). 相反, 对此问题如何求得一个经典置信区间还不很清楚, 因求经典置信区间需要构造一个枢轴变量, 找出枢轴变量的分布是一项技术性很强的工作, 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, 相对说来要简单得多.

## 例子

- **例 4.2.3** 设  $X_1, \dots, X_n \text{ iid} \sim N(\theta, \sigma^2)$ ,  $\sigma^2$  已知,  $\theta$  的先验分布为无信息先验, 即  $\pi(\theta) \equiv 1$ ,  $\theta \in R_1$ , 求  $\theta$  的  $1-\alpha$  的 HPD 可信区间.
- **例 4.2.4** 设  $X_1, \dots, X_n$  为从柯西分布  $C(\theta, 1)$ ,  $\theta > 0$  中抽取的简单样本, 取先验  $\pi(\theta) \equiv 1$ ,  $\theta \in R_1$ , 求  $\theta$  的  $1-\alpha$  的 HPD 可信区间.
- 这是一个很不容易计算的后验分布, 找  $1-\alpha$  的 HPD 可信区间由计算机进行很容易. 如  $n = 5$ ,  $\mathbf{x} = (4.0, 5.5, 7.5, 4.5, 3.0)$  则 95% HPD 可信区间为 (3.10, 6.06). 相反, 对此问题如何求得一个经典置信区间还不很清楚, 因求经典置信区间需要构造一个枢轴变量, 找出枢轴变量的分布是一项技术性很强的工作, 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, 相对说来要简单得多.

# 例子

- **例 4.2.3** 设  $X_1, \dots, X_n \text{ iid} \sim N(\theta, \sigma^2)$ ,  $\sigma^2$  已知,  $\theta$  的先验分布为无信息先验, 即  $\pi(\theta) \equiv 1$ ,  $\theta \in R_1$ , 求  $\theta$  的  $1-\alpha$  的 HPD 可信区间.
- **例 4.2.4** 设  $X_1, \dots, X_n$  为从柯西分布  $C(\theta, 1)$ ,  $\theta > 0$  中抽取的简单样本, 取先验  $\pi(\theta) \equiv 1$ ,  $\theta \in R_1$ , 求  $\theta$  的  $1-\alpha$  的 HPD 可信区间.
- 这是一个很不容易计算的后验分布, 找  $1-\alpha$  的 HPD 可信区间由计算机进行很容易. 如  $n = 5$ ,  $\mathbf{x} = (4.0, 5.5, 7.5, 4.5, 3.0)$  则 95% HPD 可信区间为 (3.10, 6.06). 相反, 对此问题如何求得一个经典置信区间还不很清楚, 因求经典置信区间需要构造一个枢轴变量, 找出枢轴变量的分布是一项技术性很强的工作, 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, 相对说来要简单得多.

# 1. 一元情形

- 设  $X_1, \dots, X_n \text{ iid} \sim f(x|\theta)$ , 记  $\mathbf{X} = (X_1, \dots, X_n)$ ,  $f_n(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta)$ ,  $\pi(\theta)$  为先验分布, 则

$$\pi_n(\theta|\mathbf{x}) = \frac{f_n(\mathbf{x}|\theta)\pi(\theta)}{m_n(\mathbf{x})}.$$

此处  $m_n(\mathbf{x})$  为边缘分布.

- 在适当条件下可证明: 当  $n$  充分大时, 后验分布  $\pi_n(\theta|\mathbf{x})$  近似服从  $N(\mu^\pi(\mathbf{x}), V^\pi(\mathbf{x}))$ , 此处  $\mu^\pi(\mathbf{x})$  和  $V^\pi(\mathbf{x})$  分别为后验期望和后验方差.
- 此时  $\theta$  的可信水平近似为  $1 - \alpha$  的 HPD 可信区间为

$$\left( \mu^\pi(\mathbf{x}) - u_{\alpha/2} \sqrt{V^\pi(\mathbf{x})}, \mu^\pi(\mathbf{x}) + u_{\alpha/2} \sqrt{V^\pi(\mathbf{x})} \right).$$



# 1. 一元情形

- 设  $X_1, \dots, X_n \text{ iid} \sim f(x|\theta)$ , 记  $\mathbf{X} = (X_1, \dots, X_n)$ ,  $f_n(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta)$ ,  $\pi(\theta)$  为先验分布, 则

$$\pi_n(\theta|\mathbf{x}) = \frac{f_n(\mathbf{x}|\theta)\pi(\theta)}{m_n(\mathbf{x})}.$$

此处  $m_n(\mathbf{x})$  为边缘分布.

- 在适当条件下可证明: 当  $n$  充分大时, 后验分布  $\pi_n(\theta|\mathbf{x})$  近似服从  $N(\mu^\pi(\mathbf{x}), V^\pi(\mathbf{x}))$ , 此处  $\mu^\pi(\mathbf{x})$  和  $V^\pi(\mathbf{x})$  分别为后验期望和后验方差.
- 此时  $\theta$  的可信水平近似为  $1 - \alpha$  的 HPD 可信区间为

$$\left( \mu^\pi(\mathbf{x}) - u_{\alpha/2} \sqrt{V^\pi(\mathbf{x})}, \mu^\pi(\mathbf{x}) + u_{\alpha/2} \sqrt{V^\pi(\mathbf{x})} \right).$$

# 1. 一元情形

- 设  $X_1, \dots, X_n \text{ iid} \sim f(x|\theta)$ , 记  $\mathbf{X} = (X_1, \dots, X_n)$ ,  $f_n(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta)$ ,  $\pi(\theta)$  为先验分布, 则

$$\pi_n(\theta|\mathbf{x}) = \frac{f_n(\mathbf{x}|\theta)\pi(\theta)}{m_n(\mathbf{x})}.$$

此处  $m_n(\mathbf{x})$  为边缘分布.

- 在适当条件下可证明: 当  $n$  充分大时, 后验分布  $\pi_n(\theta|\mathbf{x})$  近似服从  $N(\mu^\pi(\mathbf{x}), V^\pi(\mathbf{x}))$ , 此处  $\mu^\pi(\mathbf{x})$  和  $V^\pi(\mathbf{x})$  分别为后验期望和后验方差.
- 此时  $\theta$  的可信水平近似为  $1 - \alpha$  的 HPD 可信区间为

$$\left( \mu^\pi(\mathbf{x}) - u_{\alpha/2} \sqrt{V^\pi(\mathbf{x})}, \mu^\pi(\mathbf{x}) + u_{\alpha/2} \sqrt{V^\pi(\mathbf{x})} \right).$$

# 例子

- **例 4.2.5** (续例 4.2.4) 由例 4.2.4 中给出的后验密度显然不是正态的, 用后验分布的渐近正态分布求  $\theta$  的可信水平近似为  $1 - \alpha$  的 HPD 可信区间.

- 表 4.2.1 实际的和近似的后验分位数

$\alpha$ (%)	2.5	25	50	75	97.5
$\pi(\theta \mathbf{x})$ 的 $\alpha$ 分位数	3.17	4.07	4.52	5.00	6.15
$N(\mu^\pi, V^\pi)$ 的 $\alpha$ 分位数	3.08	4.05	4.55	5.06	6.02

- 由后验分布的渐近正态分布得到  $\theta$  的可信水平近似 95% 的 HPD 可信区间为  $C = (3.08, 6.02)$ , 它与在例 4.2.4 中经由计算机计算获得的真实可信水平为 95% 的 HPD 可信区间  $(3.10, 6.06)$  非常接近, 这个近似可信区间  $C$  具有实际的后验概率为 0.948, 与 0.95 非常接近.

# 例子

- **例 4.2.5** (续例 4.2.4) 由例 4.2.4 中给出的后验密度显然不是正态的, 用后验分布的渐近正态分布求  $\theta$  的可信水平近似为  $1 - \alpha$  的 HPD 可信区间.

- **表 4.2.1 实际的和近似的后验分位数**

$\alpha$ (%)	2.5	25	50	75	97.5
$\pi(\theta \mathbf{x})$ 的 $\alpha$ 分位数	3.17	4.07	4.52	5.00	6.15
$N(\mu^\pi, V^\pi)$ 的 $\alpha$ 分位数	3.08	4.05	4.55	5.06	6.02

- 由后验分布的渐近正态分布得到  $\theta$  的可信水平近似 95% 的 HPD 可信区间为  $C = (3.08, 6.02)$ , 它与在例 4.2.4 中经由计算机计算获得的真实可信水平为 95% 的 HPD 可信区间  $(3.10, 6.06)$  非常接近, 这个近似可信区间  $C$  具有实际的后验概率为 0.948, 与 0.95 非常接近.

# 例子

- **例 4.2.5** (续例 4.2.4) 由例 4.2.4 中给出的后验密度显然不是正态的, 用后验分布的渐近正态分布求  $\theta$  的可信水平近似为  $1 - \alpha$  的 HPD 可信区间.

- **表 4.2.1 实际的和近似的后验分位数**

$\alpha$ (%)	2.5	25	50	75	97.5
$\pi(\theta \mathbf{x})$ 的 $\alpha$ 分位数	3.17	4.07	4.52	5.00	6.15
$N(\mu^\pi, V^\pi)$ 的 $\alpha$ 分位数	3.08	4.05	4.55	5.06	6.02

- 由后验分布的渐近正态分布得到  $\theta$  的可信水平近似 95% 的 HPD 可信区间为  $C = (3.08, 6.02)$ , 它与在例 4.2.4 中经由计算机计算获得的真实可信水平为 95% 的 HPD 可信区间  $(3.10, 6.06)$  非常接近, 这个近似可信区间  $C$  具有实际的后验概率为 0.948, 与 0.95 非常接近.

1

## 4.1 贝叶斯点估计

- 4.1.1 条件方法
- 4.1.2 贝叶斯点估计
- 4.1.3 贝叶斯点估计的精度-估计的误差
- 4.1.4 贝叶斯点估计的几个例子
- 4.1.5 多参数情形

2

## 4.2 贝叶斯区间估计

- 4.2.1 可信区间定义及例子
- 4.2.2 最大后验密度 (HPD) 可信区间
- 4.2.3 大样本方法

3

## 4.3 假设检验

- 4.3.1 一般方法
- 4.3.2 贝叶斯因子
- 4.3.3 简单假设对简单假设情形
- 4.3.4 复杂假设对复杂假设情形
- 4.3.5 简单假设对复杂假设
- 4.3.6 多重假设检验

# 假设检验的一般方法

- 设检验问题仍如 § 1.2.3 节所述, 可写成如下形式:

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

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

- 贝叶斯方法处理假设检验问题是直接了当的. 在求得  $\theta$  的后验分布  $\pi(\theta|x)$  后, 计算  $\Theta_0$  和  $\Theta_1$  的后验概率

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

比较  $\alpha_0$  和  $\alpha_1$  的大小决定接受  $H_0$  还是  $H_1$ .  $\alpha_0$  和  $\alpha_1$  是综合了样本信息和先验信息得出两个假设实际发生的后验概率. 得检验如下

当  $\alpha_0/\alpha_1 > 1$  时接受  $H_0$ , 否则拒绝  $H_0$ .

# 假设检验的一般方法

- 设检验问题仍如 § 1.2.3 节所述, 可写成如下形式:

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

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

- 贝叶斯方法处理假设检验问题是直接了当的. 在求得  $\theta$  的后验分布  $\pi(\theta|x)$  后, 计算  $\Theta_0$  和  $\Theta_1$  的后验概率

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

比较  $\alpha_0$  和  $\alpha_1$  的大小决定接受  $H_0$  还是  $H_1$ .  $\alpha_0$  和  $\alpha_1$  是综合了样本信息和先验信息得出两个假设实际发生的后验概率. 得检验如下

当  $\alpha_0/\alpha_1 > 1$  时接受  $H_0$ , 否则拒绝  $H_0$ .



## 假设检验的例子

- 由此可见: 贝叶斯假设检验方法是简单的. 与频率方法相比, 它无需选择检验统计量、确定抽样分布; 无需给出显著性水平, 确定否定域; 而且容易推广到多重假设检验情形.
- 例 4.3.1 设随机变量  $X$  是从二项分布  $b(n, \theta)$  中抽取的一个样本, 取  $\theta$  的先验分布为均匀分布  $U(0, 1)$ , 求下列检验问题:

$$H_0: \theta \leq 1/2 \leftrightarrow H_1: \theta > 1/2.$$

- 取  $n = 5$ , 可算得各种  $x$  下后验概率及后验机会比  $\frac{\alpha_0}{\alpha_1}$  如下

表 4.3.1  $\theta$  的后验机会比

$x$	0	1	2	3	4	5
$\alpha_0$	63/64	57/64	42/64	22/64	7/64	1/64
$\alpha_1$	1/64	7/64	22/64	42/64	57/64	63/64
$\alpha_0/\alpha_1$	63.0	8.14	1.91	0.52	0.12	0.016

可见当  $x = 0, 1, 2$  时接受  $H_0$ , 当  $x = 3, 4, 5$  时拒绝  $H_0$ .

# 假设检验的例子

- 由此可见: 贝叶斯假设检验方法是简单的. 与频率方法相比, 它无需选择检验统计量、确定抽样分布; 无需给出显著性水平, 确定否定域; 而且容易推广到多重假设检验情形.
- 例 4.3.1** 设随机变量  $X$  是从二项分布  $b(n, \theta)$  中抽取的一个样本, 取  $\theta$  的先验分布为均匀分布  $U(0, 1)$ , 求下列检验问题:

$$H_0: \theta \leq 1/2 \leftrightarrow H_1: \theta > 1/2.$$

- 取  $n = 5$ , 可算得各种  $x$  下后验概率及后验机会比  $\frac{\alpha_0}{\alpha_1}$  如下

表 4.3.1  $\theta$  的后验机会比

$x$	0	1	2	3	4	5
$\alpha_0$	63/64	57/64	42/64	22/64	7/64	1/64
$\alpha_1$	1/64	7/64	22/64	42/64	57/64	63/64
$\alpha_0/\alpha_1$	63.0	8.14	1.91	0.52	0.12	0.016

可见当  $x = 0, 1, 2$  时接受  $H_0$ , 当  $x = 3, 4, 5$  时拒绝  $H_0$ .

## 假设检验的例子

- 由此可见: 贝叶斯假设检验方法是简单的. 与频率方法相比, 它无需选择检验统计量、确定抽样分布; 无需给出显著性水平, 确定否定域; 而且容易推广到多重假设检验情形.
- 例 4.3.1** 设随机变量  $X$  是从二项分布  $b(n, \theta)$  中抽取的一个样本, 取  $\theta$  的先验分布为均匀分布  $U(0, 1)$ , 求下列检验问题:

$$H_0: \theta \leq 1/2 \leftrightarrow H_1: \theta > 1/2.$$

- 取  $n = 5$ , 可算得各种  $x$  下后验概率及后验机会比  $\frac{\alpha_0}{\alpha_1}$  如下

表 4.3.1  $\theta$  的后验机会比

$x$	0	1	2	3	4	5
$\alpha_0$	63/64	57/64	42/64	22/64	7/64	1/64
$\alpha_1$	1/64	7/64	22/64	42/64	57/64	63/64
$\alpha_0/\alpha_1$	63.0	8.14	1.91	0.52	0.12	0.016

可见当  $x = 0, 1, 2$  时接受  $H_0$ , 当  $x = 3, 4, 5$  时拒绝  $H_0$ .

## 贝叶斯因子的定义

- **定义 4.3.1** 设两个假设  $\Theta_0$  和  $\Theta_1$  的先验概率分别为  $\pi_0$  和  $\pi_1$ , 后验概率分别为  $\alpha_0$  和  $\alpha_1$ , 比例  $\alpha_0/\alpha_1$  称为  $H_0$  对  $H_1$  的后验机会比,  $\pi_0/\pi_1$  被称为先验机会比, 则称

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

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

- **注 4.3.1** 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  $\pi$ , 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  $B^\pi(\mathbf{x})$  是反映数据  $\mathbf{x}$  支持  $H_0$  的程度.

## 贝叶斯因子的定义

- **定义 4.3.1** 设两个假设  $\Theta_0$  和  $\Theta_1$  的先验概率分别为  $\pi_0$  和  $\pi_1$ , 后验概率分别为  $\alpha_0$  和  $\alpha_1$ , 比例  $\alpha_0/\alpha_1$  称为  $H_0$  对  $H_1$  的后验机会比,  $\pi_0/\pi_1$  被称为先验机会比, 则称

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

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

- **注 4.3.1** 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  $\pi$ , 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  $B^\pi(\mathbf{x})$  是反映数据  $\mathbf{x}$  支持  $H_0$  的程度.

## 简单假设对简单假设

- 要解释  $B^\pi(\mathbf{x})$  的意义, 首先看  $\Theta_0$  和  $\Theta_1$  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

$$\frac{\alpha_0}{\alpha_1} = \frac{\pi_0 f(\mathbf{x}|\theta_0)}{\pi_1 f(\mathbf{x}|\theta_1)}. \quad (4.3.4)$$

- 故贝叶斯因子为

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

## 简单假设对简单假设

- 要解释  $B^\pi(\mathbf{x})$  的意义, 首先看  $\Theta_0$  和  $\Theta_1$  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

$$\frac{\alpha_0}{\alpha_1} = \frac{\pi_0 f(\mathbf{x}|\theta_0)}{\pi_1 f(\mathbf{x}|\theta_1)}. \quad (4.3.4)$$

- 故贝叶斯因子为

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

## 简单假设对简单假设

- 要解释  $B^\pi(\mathbf{x})$  的意义, 首先看  $\Theta_0$  和  $\Theta_1$  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

$$\frac{\alpha_0}{\alpha_1} = \frac{\pi_0 f(\mathbf{x}|\theta_0)}{\pi_1 f(\mathbf{x}|\theta_1)}. \quad (4.3.4)$$

- 故贝叶斯因子为

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



## 简单假设对简单假设

- 如果要拒绝零假设  $H_0$ , 则要求  $\alpha_0/\alpha_1 < 1$ , 由 (4.3.4) 可见其等价于

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

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  $B^\pi(\mathbf{x})$  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  $B^\pi(\mathbf{x})$  可视为是数据  $\mathbf{x}$  支持  $\Theta_0$  的程度.
- 例 4.3.2 设  $X \sim N(\theta, 1)$ , 其中  $\theta$  的取值有两种可能, 非 0 即 1, 令  $\mathbf{X} = (X_1, \dots, X_n)$  为从总体  $X$  中抽取的简单样本. 要检验假设

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

## 简单假设对简单假设

- 如果要拒绝零假设  $H_0$ , 则要求  $\alpha_0/\alpha_1 < 1$ , 由 (4.3.4) 可见其等价于

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

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  $B^\pi(\mathbf{x})$  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  $B^\pi(\mathbf{x})$  可视为是数据  $\mathbf{x}$  支持  $\Theta_0$  的程度.
- 例 4.3.2 设  $X \sim N(\theta, 1)$ , 其中  $\theta$  的取值有两种可能, 非 0 即 1, 令  $\mathbf{X} = (X_1, \dots, X_n)$  为从总体  $X$  中抽取的简单样本. 要检验假设

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

## 简单假设对简单假设

- 如果要拒绝零假设  $H_0$ , 则要求  $\alpha_0/\alpha_1 < 1$ , 由 (4.3.4) 可见其等价于

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

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  $B^\pi(\mathbf{x})$  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  $B^\pi(\mathbf{x})$  可视为是数据  $\mathbf{x}$  支持  $\Theta_0$  的程度.
- **例 4.3.2** 设  $X \sim N(\theta, 1)$ , 其中  $\theta$  的取值有两种可能, 非 0 即 1, 令  $\mathbf{X} = (X_1, \dots, X_n)$  为从总体  $X$  中抽取的简单样本. 要检验假设

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

## 复杂假设对复杂假设

- 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  $\pi_0$  和  $\pi_1$  分别为  $\Theta_0$  和  $\Theta_1$  上的先验概率,  $g_0(\theta)$  和  $g_1(\theta)$  分别是  $\Theta_0$  和  $\Theta_1$  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.

## 复杂假设对复杂假设

- 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  $\pi_0$  和  $\pi_1$  分别为  $\Theta_0$  和  $\Theta_1$  上的先验概率,  $g_0(\theta)$  和  $g_1(\theta)$  分别是  $\Theta_0$  和  $\Theta_1$  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.

## 复杂假设对复杂假设

- 在上述记号下，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

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

## 复杂假设对复杂假设

- 在上述记号下，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

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

## 复杂假设对复杂假设

- 可见  $B^\pi(\mathbf{x})$  还依赖于  $\Theta_0$  和  $\Theta_1$  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  $\Theta_0$  和  $\Theta_1$  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  $B^\pi(\mathbf{x})$  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  $\theta$  在  $\Theta_0$  和  $\Theta_1$  的 MLE (极大似然估计), 那么经典统计中使用的似然比统计量

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

是贝叶斯因子  $B^\pi(\mathbf{x})$  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.



## 复杂假设对复杂假设

- 可见  $B^\pi(\mathbf{x})$  还依赖于  $\Theta_0$  和  $\Theta_1$  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  $\Theta_0$  和  $\Theta_1$  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  $B^\pi(\mathbf{x})$  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  $\theta$  在  $\Theta_0$  和  $\Theta_1$  的 MLE (极大似然估计), 那么经典统计中使用的似然比统计量

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

是贝叶斯因子  $B^\pi(\mathbf{x})$  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.

## 复杂假设对复杂假设

- **例 4.3.3** 在一个儿童做智力测验问题中, 设测验结果  $X \sim N(\theta, 100)$ ,  $\theta$  为这个孩子测验中的智商真值,  $\theta$  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:

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

- **例 4.3.4** 当  $X \sim N(\theta, \sigma^2)$ ,  $\sigma^2$  已知. 设  $\theta$  的先验分布为无信息先验, 即  $\pi(\theta) \equiv 1$ ,  $\theta \in R_1$ , 求检验问题:

$$H_0 : \theta \leq \theta_0 \longleftrightarrow H_1 : \theta > \theta_0.$$

- **例 4.3.5** 设从正态总体  $N(\theta, 1)$  中随机抽取一个容量为 10 的样本  $\mathbf{X} = (X_1, \dots, X_{10})$ , 算得样本均值  $\bar{X} = 1.5$ , 令  $\theta$  的先验分布为共轭先验分布  $N(0.5, 2)$ , 求下列检验问题:

$$H_0 : \theta \leq 1 \longleftrightarrow H_1 : \theta > 1.$$

## 复杂假设对复杂假设

- **例 4.3.3** 在一个儿童做智力测验问题中, 设测验结果  $X \sim N(\theta, 100)$ ,  $\theta$  为这个孩子测验中的智商真值,  $\theta$  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:

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

- **例 4.3.4** 当  $X \sim N(\theta, \sigma^2)$ ,  $\sigma^2$  已知. 设  $\theta$  的先验分布为无信息先验, 即  $\pi(\theta) \equiv 1$ ,  $\theta \in R_1$ , 求检验问题:

$$H_0 : \theta \leq \theta_0 \longleftrightarrow H_1 : \theta > \theta_0.$$

- **例 4.3.5** 设从正态总体  $N(\theta, 1)$  中随机抽取一个容量为 10 的样本  $\mathbf{X} = (X_1, \dots, X_{10})$ , 算得样本均值  $\bar{X} = 1.5$ , 令  $\theta$  的先验分布为共轭先验分布  $N(0.5, 2)$ , 求下列检验问题:

$$H_0 : \theta \leq 1 \longleftrightarrow H_1 : \theta > 1.$$

## 复杂假设对复杂假设

- **例 4.3.3** 在一个儿童做智力测验问题中, 设测验结果  $X \sim N(\theta, 100)$ ,  $\theta$  为这个孩子测验中的智商真值,  $\theta$  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:

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

- **例 4.3.4** 当  $X \sim N(\theta, \sigma^2)$ ,  $\sigma^2$  已知. 设  $\theta$  的先验分布为无信息先验, 即  $\pi(\theta) \equiv 1$ ,  $\theta \in R_1$ , 求检验问题:

$$H_0 : \theta \leq \theta_0 \longleftrightarrow H_1 : \theta > \theta_0.$$

- **例 4.3.5** 设从正态总体  $N(\theta, 1)$  中随机抽取一个容量为 10 的样本  $\mathbf{X} = (X_1, \dots, X_{10})$ , 算得样本均值  $\bar{X} = 1.5$ , 令  $\theta$  的先验分布为共轭先验分布  $N(0.5, 2)$ , 求下列检验问题:

$$H_0 : \theta \leq 1 \longleftrightarrow H_1 : \theta > 1.$$

## 复杂假设对复杂假设

- 由定理 3.2.2 (1), 算得  $\theta$  后验分布为  $N(1.4524, 0.3086^2)$ , 故  $\alpha_0 = P(\theta \leq 1|\mathbf{x}) = \Phi(-1.4660) = 0.0708$ ,  $\alpha_1 = 1 - \alpha_0 = 0.9292$ .
- 由先验分布  $N(0.5, 2)$  可算得  $H_0$  与  $H_1$  成立时的先验概率  $\pi_0 = \Phi\left(\frac{1-0.5}{\sqrt{2}}\right) = \Phi(0.3536) = 0.6368$ ,  $\pi_1 = 1 - \pi_0 = 0.3632$ .
- 因此叶斯因子为

$$B^\pi(\mathbf{x}) = \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1} = \frac{0.0762}{1.7533} = 0.0435.$$

可见, 支持  $H_0$  的贝叶斯因子也很小, 应否定  $H_0$ .

## 复杂假设对复杂假设

- 由定理 3.2.2 (1), 算得  $\theta$  后验分布为  $N(1.4524, 0.3086^2)$ , 故  $\alpha_0 = P(\theta \leq 1|\mathbf{x}) = \Phi(-1.4660) = 0.0708$ ,  $\alpha_1 = 1 - \alpha_0 = 0.9292$ .
- 由先验分布  $N(0.5, 2)$  可算得  $H_0$  与  $H_1$  成立时的先验概率  $\pi_0 = \Phi\left(\frac{1 - 0.5}{\sqrt{2}}\right) = \Phi(0.3536) = 0.6368$ ,  $\pi_1 = 1 - \pi_0 = 0.3632$ .
- 因此叶斯因子为

$$B^\pi(\mathbf{x}) = \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1} = \frac{0.0762}{1.7533} = 0.0435.$$

可见, 支持  $H_0$  的贝叶斯因子也很小, 应否定  $H_0$ .

## 复杂假设对复杂假设

- 由定理 3.2.2 (1), 算得  $\theta$  后验分布为  $N(1.4524, 0.3086^2)$ , 故  $\alpha_0 = P(\theta \leq 1 | \mathbf{x}) = \Phi(-1.4660) = 0.0708$ ,  $\alpha_1 = 1 - \alpha_0 = 0.9292$ .
- 由先验分布  $N(0.5, 2)$  可算得  $H_0$  与  $H_1$  成立时的先验概率  $\pi_0 = \Phi\left(\frac{1 - 0.5}{\sqrt{2}}\right) = \Phi(0.3536) = 0.6368$ ,  $\pi_1 = 1 - \pi_0 = 0.3632$ .
- 因此叶斯因子为

$$B^\pi(\mathbf{x}) = \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1} = \frac{0.0762}{1.7533} = 0.0435.$$

可见, 支持  $H_0$  的贝叶斯因子也很小, 应否定  $H_0$ .

## 复杂假设对复杂假设

**评论** 下面通过数值的例子将告诉我们，贝叶斯因子对样本数据变化的反映是灵敏的，而对先验信息变化的反映是迟钝的。由此可见贝叶斯因子  $B^\pi(\mathbf{x})$  的作用，突出反映数据  $\mathbf{x}$  支持零假设  $H_0$  的程度。

**表 4.3.2** 样本均值  $\bar{x}$  对贝叶斯因子的影响

$\bar{x}$	$\alpha_0$	$\alpha_1$	$\alpha_0/\alpha_1$	$\pi_0/\pi_1$	$B^\pi(\mathbf{x})$
1.5	0.0708	0.9292	0.0762	1.7533	0.0435
1.4	0.1230	0.8770	0.1403	1.7533	0.0800
1.3	0.1977	0.8023	0.2464	1.7533	0.1405
1.2	0.2946	0.7054	0.4176	1.7533	0.2382
1.1	0.4090	0.5910	0.6920	1.7533	0.3947
1.0	0.5319	0.4681	1.1363	1.7533	0.6481
0.9	0.6517	0.3483	1.8711	1.7533	1.0672
0.8	0.7549	0.2451	3.0800	1.7533	1.7567
0.7	0.8413	0.1587	5.3012	1.7533	3.0236
0.6	0.9049	0.0951	9.5152	1.7533	5.4271
0.5	0.9474	0.0526	18.0114	1.7533	10.2729



## 复杂假设对复杂假设

类似地，若样本容量  $n$  和样本均值  $\bar{x}$  保持不变，而让先验均值  $E(\theta)$  从 0.5 逐渐增加到 1.5，同样可算得后验机会比、先验机会比和贝叶斯因子 (见表4.3.3).

表 4.3.3 先验均值  $E(\theta)$  对贝叶斯因子的影响

$E(\theta)$	$\alpha_0$	$\alpha_0/(1 - \alpha_0)$	$\pi_0$	$\pi_0/(1 - \pi_0)$	$B^\pi(\mathbf{x})$
0.5	0.0708	0.0761	0.6368	1.7553	0.0435
0.6	0.0694	0.0746	0.6103	1.5782	0.0472
0.7	0.0668	0.0715	0.5832	1.3992	0.0511
0.8	0.0655	0.0701	0.5557	1.2507	0.0560
0.9	0.0630	0.0672	0.5279	1.1182	0.0601
1.0	0.0618	0.0658	0.5000	1.0000	0.0658
1.1	0.0594	0.0632	0.4721	0.8943	0.0707
1.2	0.0582	0.0618	0.4443	0.7996	0.0773
1.3	0.0559	0.0592	0.4168	0.7147	0.0828
1.4	0.0548	0.0580	0.3897	0.6336	0.0915
1.5	0.0526	0.0555	0.3632	0.5704	0.0973

## 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  $\theta$  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  $\varepsilon$  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  $\theta_0$  一个正概率  $\pi_0$ , 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  $\theta$  的先验密度为

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

## 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  $\theta$  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  $\varepsilon$  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  $\theta_0$  一个正概率  $\pi_0$ , 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  $\theta$  的先验密度为

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

## 简单假设对复杂假设

- 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:

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

## 简单假设对复杂假设

- 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:

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

## 简单假设对复杂假设

- 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：

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

## 简单假设对复杂假设

- 于是贝叶斯因子为

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

可见贝叶斯因子更简单.

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

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

- 例 4.3.6 设  $X \sim B(n, \theta)$ , 求下列检验

$$H_0: \theta = \frac{1}{2} \longleftrightarrow H_1: \theta \neq \frac{1}{2}.$$

设在  $\theta \neq 1/2$  上的密度  $g_1(\theta)$  为  $(0,1)$  上的均匀分布  $U(0,1)$ .

## 简单假设对复杂假设

- 于是贝叶斯因子为

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

可见贝叶斯因子更简单.

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

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

- 例 4.3.6 设  $X \sim B(n, \theta)$ , 求下列检验

$$H_0: \theta = \frac{1}{2} \longleftrightarrow H_1: \theta \neq \frac{1}{2}.$$

设在  $\theta \neq 1/2$  上的密度  $g_1(\theta)$  为  $(0,1)$  上的均匀分布  $U(0,1)$ .



## 简单假设对复杂假设

- 于是贝叶斯因子为

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

可见贝叶斯因子更简单.

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

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

- 例 4.3.6 设  $X \sim B(n, \theta)$ , 求下列检验

$$H_0: \theta = \frac{1}{2} \longleftrightarrow H_1: \theta \neq \frac{1}{2}.$$

设在  $\theta \neq 1/2$  上的密度  $g_1(\theta)$  为  $(0,1)$  上的均匀分布  $U(0,1)$ .

## 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

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

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

- 计算  $\Theta_i$  的后验概率

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

取其最大者, 则认为相应的假设成立.

- 例 4.3.7** (续例 4.3.3) 考虑对儿童做智力测验的情形, 求下列多重检验问题:

$$H_1 : \theta \leq 90, \quad H_2 : 90 < \theta < 110, \quad H_3 : \theta \geq 110.$$

## 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

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

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

- 计算  $\Theta_i$  的后验概率

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

取其最大者, 则认为相应的假设成立.

- 例 4.3.7 (续例 4.3.3) 考虑对儿童做智力测验的情形, 求下列多重检验问题:

$$H_1 : \theta \leq 90, \quad H_2 : 90 < \theta < 110, \quad H_3 : \theta \geq 110.$$

## 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

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

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

- 计算  $\Theta_i$  的后验概率

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

取其最大者, 则认为相应的假设成立.

- 例 4.3.7** (续例 4.3.3) 考虑对儿童做智力测验的情形, 求下列多重检验问题:

$$H_1 : \theta \leq 90, \quad H_2 : 90 < \theta < 110, \quad H_3 : \theta \geq 110.$$

1

## 4.1 贝叶斯点估计

- 4.1.1 条件方法
- 4.1.2 贝叶斯点估计
- 4.1.3 贝叶斯点估计的精度-估计的误差
- 4.1.4 贝叶斯点估计的几个例子
- 4.1.5 多参数情形

2

## 4.2 贝叶斯区间估计

- 4.2.1 可信区间定义及例子
- 4.2.2 最大后验密度 (HPD) 可信区间
- 4.2.3 大样本方法

3

## 4.3 假设检验

- 4.3.1 一般方法
- 4.3.2 贝叶斯因子
- 4.3.3 简单假设对简单假设情形
- 4.3.4 复杂假设对复杂假设情形
- 4.3.5 简单假设对复杂假设
- 4.3.6 多重假设检验

# 预测问题

- 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如下: 设  $\pi(\theta|x)$  为  $\theta$  的后验分布, 于是  $g(z|\theta)\pi(\theta|x)$  为给定  $X = x$  的条件下  $(Z, \theta)$  的联合分布, 把它对  $\theta$  积分得到给定  $X = x$  时  $Z$  的边缘分布密度作为预测密度.
- **定义 4.4.1** 设随机变量  $X$  的密度函数是  $f(x|\theta)$ ,  $\theta$  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

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

# 预测问题

- 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如下: 设  $\pi(\theta|x)$  为  $\theta$  的后验分布, 于是  $g(z|\theta)\pi(\theta|x)$  为给定  $X = x$  的条件下  $(Z, \theta)$  的联合分布, 把它对  $\theta$  积分得到给定  $X = x$  时  $Z$  的边缘分布密度作为预测密度.
- 定义 4.4.1 设随机变量  $X$  的密度函数是  $f(x|\theta)$ ,  $\theta$  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

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

# 预测问题

- 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如下: 设  $\pi(\theta|x)$  为  $\theta$  的后验分布, 于是  $g(z|\theta)\pi(\theta|x)$  为给定  $X = x$  的条件下  $(Z, \theta)$  的联合分布, 把它对  $\theta$  积分得到给定  $X = x$  时  $Z$  的边缘分布密度作为预测密度.
- **定义 4.4.1** 设随机变量  $X$  的密度函数是  $f(x|\theta)$ ,  $\theta$  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

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



## 例 4.4.1

- **例 4.4.1** 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.
- 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  $\theta$  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  $a = b = 1$ , 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

$$\begin{aligned} p(0|3) &= 0.1813, & p(1|3) &= 0.3022, & p(2|3) &= 0.2747, \\ p(3|3) &= 0.1649, & p(4|3) &= 0.0641, & p(5|3) &= 0.01282. \end{aligned}$$

## 例 4.4.1

- **例 4.4.1** 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.
- 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  $\theta$  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  $a = b = 1$ , 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

$$\begin{aligned} p(0|3) &= 0.1813, & p(1|3) &= 0.3022, & p(2|3) &= 0.2747, \\ p(3|3) &= 0.1649, & p(4|3) &= 0.0641, & p(5|3) &= 0.01282. \end{aligned}$$

## 例 4.4.1

- **例 4.4.1** 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.
- 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  $\theta$  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  $a = b = 1$ , 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

$$\begin{aligned} p(0|3) &= 0.1813, & p(1|3) &= 0.3022, & p(2|3) &= 0.2747, \\ p(3|3) &= 0.1649, & p(4|3) &= 0.0641, & p(5|3) &= 0.01282. \end{aligned}$$

## 例 4.4.1

- 由后验预测分布可见, 其概率集中在  $Z = 0, 1, 2, 3$  之间, 即

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

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.

- 例 4.4.2 一颗钻石在一架天平上重复称重  $n$  次, 其结果为  $X_1, \dots, X_n$ , 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?

## 例 4.4.1

- 由后验预测分布可见, 其概率集中在  $Z = 0, 1, 2, 3$  之间, 即

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

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.

- 例 4.4.2** 一颗钻石在一架天平上重复称重  $n$  次, 其结果为  $X_1, \dots, X_n$ , 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?

## 例 4.4.2

- 由定理 3.2.2 (1) 可知  $\theta$  的后验分布  $\pi(\theta|\bar{x})$  为  $N(\mu_1, \eta_1^2)$ , 即

$$\pi(\theta|\bar{x}) = \frac{1}{\sqrt{2\pi}\eta_1} \exp \left\{ -\frac{1}{2\eta_1^2}(\theta - \mu_1)^2 \right\},$$

- 其中样本分布  $X|\theta \sim N(\theta, \sigma_1^2)$ , 先验分布  $\theta \sim N(\mu, \tau^2)$ , 而

$$\begin{aligned}\mu_1 &= \frac{\tau^2}{\sigma_1^2/n + \tau^2} \bar{x} + \frac{\sigma_1^2/n}{\sigma_1^2/n + \tau^2} \mu, \\ \eta_1^2 &= \frac{\sigma_1^2/n \cdot \tau^2}{\sigma_1^2/n + \tau^2} = \frac{\sigma_1^2 \tau^2}{\sigma_1^2 + n\tau^2}.\end{aligned}$$

## 例 4.4.2

- 由定理 3.2.2 (1) 可知  $\theta$  的后验分布  $\pi(\theta|\bar{x})$  为  $N(\mu_1, \eta_1^2)$ , 即

$$\pi(\theta|\bar{x}) = \frac{1}{\sqrt{2\pi}\eta_1} \exp \left\{ -\frac{1}{2\eta_1^2}(\theta - \mu_1)^2 \right\},$$

- 其中样本分布  $X|\theta \sim N(\theta, \sigma_1^2)$ , 先验分布  $\theta \sim N(\mu, \tau^2)$ , 而

$$\begin{aligned}\mu_1 &= \frac{\tau^2}{\sigma_1^2/n + \tau^2} \bar{x} + \frac{\sigma_1^2/n}{\sigma_1^2/n + \tau^2} \mu, \\ \eta_1^2 &= \frac{\sigma_1^2/n \cdot \tau^2}{\sigma_1^2/n + \tau^2} = \frac{\sigma_1^2 \tau^2}{\sigma_1^2 + n\tau^2}.\end{aligned}$$

## 例 4.4.2

- 设第二架天平称得钻石重量为  $Z$ ,  $Z|\theta \sim N(\theta, \sigma_2^2)$ , 其密度为

$$g(z|\theta) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp \left\{ -\frac{1}{2\sigma_2^2}(z - \theta)^2 \right\}.$$

- $Z$  的后验预测密度为

$$\begin{aligned} p(z|\theta) &= \frac{1}{2\pi\eta_1\sigma_2} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \left[ A \left( \theta - \frac{B}{A} \right)^2 + \left( C - \frac{B^2}{A} \right) \right] \right\} d\theta, \\ &= \frac{1}{\sqrt{2\pi}\eta_1\sigma_2\sqrt{A}} \exp \left\{ -\frac{1}{2} \left( C - \frac{B^2}{A} \right) \right\} \\ &= \frac{1}{\sqrt{2\pi(\eta_1^2 + \sigma_2^2)}} \exp \left\{ -\frac{(z - \mu_1)^2}{2(\eta_1^2 + \sigma_2^2)} \right\}, \end{aligned}$$

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



## 例 4.4.2

- 设第二架天平称得钻石重量为  $Z$ ,  $Z|\theta \sim N(\theta, \sigma_2^2)$ , 其密度为

$$g(z|\theta) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp \left\{ -\frac{1}{2\sigma_2^2}(z - \theta)^2 \right\}.$$

- $Z$  的后验预测密度为

$$\begin{aligned} p(z|\theta) &= \frac{1}{2\pi\eta_1\sigma_2} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \left[ A \left( \theta - \frac{B}{A} \right)^2 + \left( C - \frac{B^2}{A} \right) \right] \right\} d\theta, \\ &= \frac{1}{\sqrt{2\pi}\eta_1\sigma_2\sqrt{A}} \exp \left\{ -\frac{1}{2} \left( C - \frac{B^2}{A} \right) \right\} \\ &= \frac{1}{\sqrt{2\pi(\eta_1^2 + \sigma_2^2)}} \exp \left\{ -\frac{(z - \mu_1)^2}{2(\eta_1^2 + \sigma_2^2)} \right\}, \end{aligned}$$

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

## 例 4.4.2 - 点预测

- 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

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

## 例 4.4.2 - 点预测

- 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

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

## 例 4.4.2 - 区间预测

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

$$P\left(-u_{\alpha/2} \leq \frac{Z - \mu_1}{\sqrt{\eta_1^2 + \sigma_2^2}} \leq u_{\alpha/2} \middle| \bar{x}\right) = 1 - \alpha,$$

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

- 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  $\mu_1$  和  $\eta_1^2$  分别如前给出.

## 例 4.4.2 - 区间预测

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

$$P\left(-u_{\alpha/2} \leq \frac{Z - \mu_1}{\sqrt{\eta_1^2 + \sigma_2^2}} \leq u_{\alpha/2} \middle| \bar{x}\right) = 1 - \alpha,$$

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

- 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  $\mu_1$  和  $\eta_1^2$  分别如前给出.

## 1 4.1 贝叶斯点估计

- 4.1.1 条件方法
- 4.1.2 贝叶斯点估计
- 4.1.3 贝叶斯点估计的精度-估计的误差
- 4.1.4 贝叶斯点估计的几个例子
- 4.1.5 多参数情形

## 2 4.2 贝叶斯区间估计

- 4.2.1 可信区间定义及例子
- 4.2.2 最大后验密度 (HPD) 可信区间
- 4.2.3 大样本方法

## 3 4.3 假设检验

- 4.3.1 一般方法
- 4.3.2 贝叶斯因子
- 4.3.3 简单假设对简单假设情形
- 4.3.4 复杂假设对复杂假设情形
- 4.3.5 简单假设对复杂假设
- 4.3.6 多重假设检验

# 引言

- 本章第三节讨论贝叶斯假设检验问题. 与假设检验问题密切相关的是贝叶斯模型选择问题. 贝叶斯模型选择可以认为是假设检验的特殊形式, 下面我们来加以说明.
- 设总体  $X \sim f(x|\theta)$ , 其中  $\theta$  为一未知参数且  $\theta \in \Theta \subset R^p$  感兴趣的假设

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

等价于比较两个模型

$M_0 : X$  有密度  $f(x|\theta)$ , 其中  $\theta \in \Theta_0$ ;

$M_1 : X$  有密度  $f(x|\theta)$ , 其中  $\theta \in \Theta_1$ ;

其中  $\Theta_0 = \Theta - \Theta_1$ .

# 引言

- 本章第三节讨论贝叶斯假设检验问题. 与假设检验问题密切相关的是贝叶斯模型选择问题. 贝叶斯模型选择可以认为是假设检验的特殊形式, 下面我们来加以说明.
- 设总体  $X \sim f(x|\theta)$ , 其中  $\theta$  为一未知参数且  $\theta \in \Theta \subset R^p$  感兴趣的假设

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

等价于比较两个模型

$M_0 : X$  有密度  $f(x|\theta)$ , 其中  $\theta \in \Theta_0$ ;

$M_1 : X$  有密度  $f(x|\theta)$ , 其中  $\theta \in \Theta_1$ ;

其中  $\Theta_0 = \Theta - \Theta_1$ .



# 引言

- 令  $g_i(\theta)$  表示给定真实模型为  $M_i$  下  $\theta$  的先验密度  $i = 0, 1$ . 则当有了样本  $\mathbf{X} = (X_1, \dots, X_n)$  后, 我们可以使用贝叶斯因子来比较  $M_0$  和  $M_1$ :

$$B_{01}^{\pi} = \frac{P(\Theta_0|\mathbf{x})}{P(\Theta_1|\mathbf{x})} \bigg/ \frac{\pi_0}{1 - \pi_0} = \frac{m_0(\mathbf{x})}{m_1(\mathbf{x})},$$

其中  $\pi_0 = P^{\pi}(M_0) = P^{\pi}(\Theta_0)$ ,  $P^{\pi}(M_1) = P^{\pi}(\Theta_1) = 1 - \pi_0$ ,

- 而

$$m_i(\mathbf{x}) = \int_{\Theta_i} f(\mathbf{x}|\theta)g_i(\theta)d\theta, \quad i = 0, 1.$$

# 引言

- 令  $g_i(\theta)$  表示给定真实模型为  $M_i$  下  $\theta$  的先验密度  $i = 0, 1$ . 则当有了样本  $\mathbf{X} = (X_1, \dots, X_n)$  后, 我们可以使用贝叶斯因子来比较  $M_0$  和  $M_1$ :

$$B_{01}^{\pi} = \frac{P(\Theta_0|\mathbf{x})}{P(\Theta_1|\mathbf{x})} \bigg/ \frac{\pi_0}{1 - \pi_0} = \frac{m_0(\mathbf{x})}{m_1(\mathbf{x})},$$

其中  $\pi_0 = P^{\pi}(M_0) = P^{\pi}(\Theta_0)$ ,  $P^{\pi}(M_1) = P^{\pi}(\Theta_1) = 1 - \pi_0$ ,

- 而

$$m_i(\mathbf{x}) = \int_{\Theta_i} f(\mathbf{x}|\theta)g_i(\theta)d\theta, \quad i = 0, 1.$$

# 引言

- 因此

$$P(M_0|\mathbf{x}) = \left[ 1 + \frac{1 - \pi_0}{\pi_0} \cdot \frac{1}{B_{01}^{\pi_1}(\mathbf{x})} \right]^{-1}.$$

从而，如果先验密度  $g_0, g_1$  被指定，则可以仅仅使用贝叶斯因子  $B_{01}^{\pi_1}$  进行模型选择.

- 进一步，如果  $\pi_0$  被指定，则可以计算得到模型  $M_0$  和  $M_1$  的后验机会比  $P(\Theta_0|\mathbf{x})/P(\Theta_1|\mathbf{x})$ . 因而也可以使用后验机会比进行模型选择.
- 但贝叶斯因子或后验机会比未必总是容易计算的，可能是得不出积分值. 此时，可以使用 Laplace 逼近方法计算贝叶斯因子. 当样本边缘密度难以计算时，还可以用蒙特卡洛抽样方法或 MCMC 方法获得贝叶斯因子的模拟结果.

# 引言

- 因此

$$P(M_0|\mathbf{x}) = \left[ 1 + \frac{1 - \pi_0}{\pi_0} \cdot \frac{1}{B_{01}^{\pi_1}(\mathbf{x})} \right]^{-1}.$$

从而，如果先验密度  $g_0, g_1$  被指定，则可以仅仅使用贝叶斯因子  $B_{01}^{\pi_1}$  进行模型选择.

- 进一步，如果  $\pi_0$  被指定，则可以计算得到模型  $M_0$  和  $M_1$  的后验机会比  $P(\Theta_0|\mathbf{x})/P(\Theta_1|\mathbf{x})$ . 因而也可以使用后验机会比进行模型选择.
- 但贝叶斯因子或后验机会比未必总是容易计算的，可能是得不出积分值. 此时，可以使用 Laplace 逼近方法计算贝叶斯因子. 当样本边缘密度难以计算时，还可以用蒙特卡洛抽样方法或 MCMC 方法获得贝叶斯因子的模拟结果.

# 引言

- 因此

$$P(M_0|\mathbf{x}) = \left[ 1 + \frac{1 - \pi_0}{\pi_0} \cdot \frac{1}{B_{01}^{\pi_1}(\mathbf{x})} \right]^{-1}.$$

从而，如果先验密度  $g_0, g_1$  被指定，则可以仅仅使用贝叶斯因子  $B_{01}^{\pi_1}$  进行模型选择.

- 进一步，如果  $\pi_0$  被指定，则可以计算得到模型  $M_0$  和  $M_1$  的后验机会比  $P(\Theta_0|\mathbf{x})/P(\Theta_1|\mathbf{x})$ . 因而也可以使用后验机会比进行模型选择.
- 但贝叶斯因子或后验机会比未必总是容易计算的，可能是得不出积分值. 此时，可以使用 Laplace 逼近方法计算贝叶斯因子. 当样本边缘密度难以计算时，还可以用蒙特卡洛抽样方法或 MCMC 方法获得贝叶斯因子的模拟结果.

# 1. 一般方法

- 假设我们感兴趣的是从候选模型  $M_1, \dots, M_r$  中选择一个最佳模型，则在样本  $\mathbf{x}_n = (X_1, \dots, X_n)$  给定的条件下，模型  $M_k$  的后验概率为

$$P(M_k|\mathbf{x}_n) = \frac{P(M_k) \int f_k(\mathbf{x}_n|\boldsymbol{\theta}_k) \pi_k(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}{\sum_{\alpha=1}^r P(M_\alpha) \int f_\alpha(\mathbf{x}_n|\boldsymbol{\theta}_\alpha) \pi_\alpha(\boldsymbol{\theta}_\alpha) d\boldsymbol{\theta}_\alpha}, \quad (4.5.5)$$

- 其中  $f_k(\mathbf{x}_n|\boldsymbol{\theta}_k)$  为样本  $\mathbf{x}_n$  在模型  $M_k$  下的密度， $P(M_k)$  为模型  $M_k$  的先验概率。  $P(M_k)$  和相应的先验密度  $\pi_k(\boldsymbol{\theta}_k)$  表示我们对模型  $M_k$  的初始认识，当有了样本  $\mathbf{x}_n$  后，对模型  $M_k$  的认识就更新为后验概率  $P(M_k|\mathbf{x}_n)$ 。

# 1. 一般方法

- 假设我们感兴趣的是从候选模型  $M_1, \dots, M_r$  中选择一个最佳模型，则在样本  $\mathbf{x}_n = (X_1, \dots, X_n)$  给定的条件下，模型  $M_k$  的后验概率为

$$P(M_k|\mathbf{x}_n) = \frac{P(M_k) \int f_k(\mathbf{x}_n|\boldsymbol{\theta}_k) \pi_k(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}{\sum_{\alpha=1}^r P(M_\alpha) \int f_\alpha(\mathbf{x}_n|\boldsymbol{\theta}_\alpha) \pi_\alpha(\boldsymbol{\theta}_\alpha) d\boldsymbol{\theta}_\alpha}, \quad (4.5.5)$$

- 其中  $f_k(\mathbf{x}_n|\boldsymbol{\theta}_k)$  为样本  $\mathbf{x}_n$  在模型  $M_k$  下的密度， $P(M_k)$  为模型  $M_k$  的先验概率。  $P(M_k)$  和相应的先验密度  $\pi_k(\boldsymbol{\theta}_k)$  表示我们对模型  $M_k$  的初始认识，当有了样本  $\mathbf{x}_n$  后，对模型  $M_k$  的认识就更新为后验概率  $P(M_k|\mathbf{x}_n)$ 。

## 一般方法

- 从本质上看, 贝叶斯模型选择方法是选择后验概率最大的模型. 因此, 后验模型概率  $P(M_1|\mathbf{x}_n), \dots, P(M_r|\mathbf{x}_n)$  即为模型选择中我们感兴趣的量. 等价于寻求最大化

$$P(M_k) \int_{\Theta_k} f_k(\mathbf{x}_n|\boldsymbol{\theta}_k) \pi_k(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k, \quad k = 1, \dots, r \quad (4.5.6)$$

的模型.

- 其中积分

$$\int_{\Theta_k} f_k(\mathbf{x}_n|\boldsymbol{\theta}_k) \pi_k(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k = P(\mathbf{x}_n|M_k) \quad (4.5.7)$$

为样本  $\mathbf{x}_n$  在模型  $M_k$  下的边际概率函数 (边际似然), 它度量了指定的先验分布对样本的拟合程度.



## 一般方法

- 从本质上看, 贝叶斯模型选择方法是选择后验概率最大的模型. 因此, 后验模型概率  $P(M_1|\mathbf{x}_n), \dots, P(M_r|\mathbf{x}_n)$  即为模型选择中我们感兴趣的量. 等价于寻求最大化

$$P(M_k) \int_{\Theta_k} f_k(\mathbf{x}_n|\boldsymbol{\theta}_k) \pi_k(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k, \quad k = 1, \dots, r \quad (4.5.6)$$

的模型.

- 其中积分

$$\int_{\Theta_k} f_k(\mathbf{x}_n|\boldsymbol{\theta}_k) \pi_k(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k = P(\mathbf{x}_n|M_k) \quad (4.5.7)$$

为样本  $\mathbf{x}_n$  在模型  $M_k$  下的边际概率函数 (边际似然), 它度量了指定的先验分布对样本的拟合程度.

## 一般方法

- 从本质上看, 贝叶斯模型选择方法是选择后验概率最大的模型. 因此, 后验模型概率  $P(M_1|\mathbf{x}_n), \dots, P(M_r|\mathbf{x}_n)$  即为模型选择中我们感兴趣的量. 等价于寻求最大化

$$P(M_k) \int_{\Theta_k} f_k(\mathbf{x}_n|\boldsymbol{\theta}_k) \pi_k(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k, \quad k = 1, \dots, r \quad (4.5.6)$$

的模型.

- 其中积分

$$\int_{\Theta_k} f_k(\mathbf{x}_n|\boldsymbol{\theta}_k) \pi_k(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k = P(\mathbf{x}_n|M_k) \quad (4.5.7)$$

为样本  $\mathbf{x}_n$  在模型  $M_k$  下的边际概率函数 (边际似然), 它度量了指定的先验分布对样本的拟合程度.

## 一般方法

- 对模型的先验概率来说，常用的一种指定为均匀分布

$$P(M_k) = 1/r, \quad k = 1, \dots, r.$$

显然这个无信息先验，表示我们对所有候选模型偏好相同。

- 在此先验下，而且后验概率简化为

$$P(M_k | \mathbf{x}_n) = \frac{\int f_k(\mathbf{x}_n | \boldsymbol{\theta}_k) \pi_k(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}{\sum_{k=1}^r \int f_k(\mathbf{x}_n | \boldsymbol{\theta}_k) \pi_k(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}. \quad (4.5.9)$$

尽管均匀先验应用很方便，但有时我们仍偏好非均匀先验。

- 贝叶斯因子是在贝叶斯框架下用于检验假设和比较模型的一个量。它在评价备选模型的拟合程度方面起重要的角色。它允许我去考虑对模型进行逐对的比较。比如说基于后验概率 (4.5.5) 去比较  $M_k$  和  $M_j$ 。

## 一般方法

- 对模型的先验概率来说，常用的一种指定为均匀分布

$$P(M_k) = 1/r, \quad k = 1, \dots, r.$$

显然这个无信息先验，表示我们对所有候选模型偏好相同。

- 在此先验下，而且后验概率简化为

$$P(M_k | \mathbf{x}_n) = \frac{\int f_k(\mathbf{x}_n | \boldsymbol{\theta}_k) \pi_k(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}{\sum_{k=1}^r \int f_k(\mathbf{x}_n | \boldsymbol{\theta}_k) \pi_k(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}. \quad (4.5.9)$$

尽管均匀先验应用很方便，但有时我们仍偏好非均匀先验。

- 贝叶斯因子是在贝叶斯框架下用于检验假设和比较模型的一个量。它在评价备选模型的拟合程度方面起重要的角色。它允许我去考虑对模型进行逐对的比较。比如说基于后验概率 (4.5.5) 去比较  $M_k$  和  $M_j$ 。

## 一般方法

- 对模型的先验概率来说，常用的一种指定为均匀分布

$$P(M_k) = 1/r, \quad k = 1, \dots, r.$$

显然这个无信息先验，表示我们对所有候选模型偏好相同。

- 在此先验下，而且后验概率简化为

$$P(M_k | \mathbf{x}_n) = \frac{\int f_k(\mathbf{x}_n | \boldsymbol{\theta}_k) \pi_k(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}{\sum_{k=1}^r \int f_k(\mathbf{x}_n | \boldsymbol{\theta}_k) \pi_k(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k}. \quad (4.5.9)$$

尽管均匀先验应用很方便，但有时我们仍偏好非均匀先验。

- 贝叶斯因子是在贝叶斯框架下用于检验假设和比较模型的一个量。它在评价备选模型的拟合程度方面起重要的角色。它允许我去考虑对模型进行逐对的比较。比如说基于后验概率 (4.5.5) 去比较  $M_k$  和  $M_j$ 。

## 2. 贝叶斯因子方法

- 由 (4.5.5) 式和与上一节类似于的讨论可知, 模型  $M_k$  和  $M_j$  的贝叶斯因子为

$$\begin{aligned} B_{kj}^{\pi} &= \frac{P(M_k|\mathbf{x}_n)}{P(M_j|\mathbf{x}_n)} \bigg/ \frac{P(M_k)}{P(M_j)} = \frac{P(\mathbf{x}_n|M_k)}{P(\mathbf{x}_n|M_j)} \\ &= \int_{\theta_k} f_k(\mathbf{x}_n|\theta_k)\pi_k(\theta_k)d\theta_k \bigg/ \int_{\theta_j} f_j(\mathbf{x}_n|\theta_j)\pi_j(\theta_j)d\theta_j. \quad (4.5.10) \end{aligned}$$

当先验概率  $P(M_k)$  和  $P(M_j)$ , 都相等时, 贝叶斯因子简化为后验概率比  $\frac{P(M_k|\mathbf{x}_n)}{P(M_j|\mathbf{x}_n)}$ .

- 由 (4.5.10) 可将模型  $M_k$  的后验概率通过  $B_{kj}^{\pi}$  表示为

$$P(M_k|\mathbf{x}_n) = \left[ \sum_{j=1}^r \frac{P(M_j)}{P(M_k)} \cdot \frac{1}{B_{jk}^{\pi}} \right]^{-1}.$$

因此可以根据模型的后验概率或者所有候选模型进行两两比较的贝叶斯因子, 从中选出最优的模型.

## 2. 贝叶斯因子方法

- 由 (4.5.5) 式和与上一节类似于的讨论可知, 模型  $M_k$  和  $M_j$  的贝叶斯因子为

$$\begin{aligned} B_{kj}^{\pi} &= \frac{P(M_k|\mathbf{x}_n)}{P(M_j|\mathbf{x}_n)} \bigg/ \frac{P(M_k)}{P(M_j)} = \frac{P(\mathbf{x}_n|M_k)}{P(\mathbf{x}_n|M_j)} \\ &= \int_{\theta_k} f_k(\mathbf{x}_n|\theta_k)\pi_k(\theta_k)d\theta_k \bigg/ \int_{\theta_j} f_j(\mathbf{x}_n|\theta_j)\pi_j(\theta_j)d\theta_j. \quad (4.5.10) \end{aligned}$$

当先验概率  $P(M_k)$  和  $P(M_j)$ , 都相等时, 贝叶斯因子简化为后验概率比  $\frac{P(M_k|\mathbf{x}_n)}{P(M_j|\mathbf{x}_n)}$ .

- 由 (4.5.10) 可将模型  $M_k$  的后验概率通过  $B_{kj}^{\pi}$  表示为

$$P(M_k|\mathbf{x}_n) = \left[ \sum_{j=1}^r \frac{P(M_j)}{P(M_k)} \cdot \frac{1}{B_{jk}^{\pi}} \right]^{-1}.$$

因此可以根据模型的后验概率或者所有候选模型进行两两比较的贝叶斯因子, 从中选出最优的模型.

## 4.5 假设检验与模型选择\*

## 4.5.2 正常先验下的贝叶斯模型选择方法

## 贝叶斯因子方法

表 4.5.1 Jeffreys 对贝叶斯因子  $B_{kj}^{\pi}$  取值的解释

贝叶斯因子	解释
$B_{kj}^{\pi} < 1$	否定模型 $M_k$
$1 \leq B_{kj}^{\pi} < 3$	对模型 $M_k$ 的支持证据微乎其微
$3 \leq B_{kj}^{\pi} < 10$	较强的证据支持 $M_k$
$10 \leq B_{kj}^{\pi} < 30$	强烈的证据支持 $M_k$
$30 \leq B_{kj}^{\pi} < 100$	非常强烈的证据支持 $M_k$
$100 \leq B_{kj}^{\pi}$	肯定支持 $M_k$



## 例子

**例 4.5.1** 假设我们研究8名司机的事故次数，且假设各司机是否发生事故相互独立，事故次数服从参数为  $\lambda$  的普阿松分布.  $\lambda$  的先验分布为  $\Gamma(\alpha, \beta)$ , 考虑两种先验分布：

$M_1$  :  $\lambda$  的先验分布为  $\Gamma(2, 2)$ , 这个先验反映了  $\lambda$  的均值为 1 的认识.

$M_2$  :  $\lambda$  的先验分布为  $\Gamma(1, 1)$ , 这个先验和  $M_1$  的均值相同，但是反映了较强的认知  $\lambda$  的方差比  $M_1$  的大.

在被调查的 8 个司机中有 3 个人没有出过事故，4 个人出过 1 次事故，1 个人出过 3 次事故. 试对这两种模型进行选择.

## 4.5 假设检验与模型选择\*

## 4.5.2 正常先验下的贝叶斯模型选择方法

## 例子

- **例 4.5.2** 假设我们玩一个打赌游戏  $n = 10$  次, 各次游戏是独立进行的, 且假设每次游戏打赌结果服从伯努利 (Bernoulli) 分布  $B(1, p)$ , 其中  $0 < p < 1$  为赌赢的概率. 如果我们对游戏很有信心, 则我们会期望  $p > 0.5$ , 否则  $p < 0.5$ . 考虑如下几种先验选择:

$$M_1 : p \sim Be(0.1, 4); \quad M_2 : p \sim Be(2, 4);$$

$$M_3 : p \sim Be(4, 4); \quad M_4 : p \sim Be(8, 4);$$

现知在完成整个游戏后, 我们胜利了 2 次. 基于此信息对先验进行选择.

- **注 4.5.1** 如果先验  $\pi_k(\theta_k)$  为不正常先验, 则边际似然也不是正常的. 各种“非正常先验下的贝叶斯因子”被提出, 例如后验贝叶斯因子 (Aitkin, 1991)、潜在贝叶斯因子 (Berger and Pericchi, 1996)、分数贝叶斯因子 (O’ Hagan, 1995)、基于交叉验证的拟贝叶斯因子 (Gelfand et al., 1992) 等. 详细的介绍请看韦和张 (2013) § 7.3 节.

## 例子

- **例 4.5.2** 假设我们玩一个打赌游戏  $n = 10$  次, 各次游戏是独立进行的, 且假设每次游戏打赌结果服从伯努利 (Bernoulli) 分布  $B(1, p)$ , 其中  $0 < p < 1$  为赌赢的概率. 如果我们对游戏很有信心, 则我们会期望  $p > 0.5$ , 否则  $p < 0.5$ . 考虑如下几种先验选择:

$$M_1 : p \sim Be(0.1, 4); \quad M_2 : p \sim Be(2, 4);$$

$$M_3 : p \sim Be(4, 4); \quad M_4 : p \sim Be(8, 4);$$

现知在完成整个游戏后, 我们胜利了 2 次. 基于此信息对先验进行选择.

- **注 4.5.1** 如果先验  $\pi_k(\theta_k)$  为不正常先验, 则边际似然也不是正常的. 各种“非正常先验下的贝叶斯因子”被提出, 例如后验贝叶斯因子 (Aitkin, 1991)、潜在贝叶斯因子 (Berger and Pericchi, 1996)、分数贝叶斯因子 (O’ Hagan, 1995)、基于交叉验证的拟贝叶斯因子 (Gelfand et al., 1992) 等. 详细的介绍请看韦和张 (2013) § 7.3 节.

# 1. 评价的重要性

- 贝叶斯模型的基本想法是指定抽样分布 (似然函数) 和所有未知参数的先验分布, 则贝叶斯模型的任何推断是基于后验分布进行. 但是, 后验推断结果的质量严重依赖于指定的模型. 因此对模型的评价是不可忽略的一个重要方面. 本节介绍两种常用的贝叶斯模型评价准则, 包括贝叶斯预测信息准则 (**BPIC**) 和偏差信息准则 (**DIC**).
- 由于 AIC 和 BIC 准则是经典统计方法中常用的模型选择的准则, 下面首先对其作简要的介绍, 然后引入 BPIC 和 DIC 准则, BPIC 准则和 DIC 准则是对 AIC 和 BIC 准则的推广.

# 1. 评价的重要性

- 贝叶斯模型的基本想法是指定抽样分布 (似然函数) 和所有未知参数的先验分布, 则贝叶斯模型的任何推断是基于后验分布进行. 但是, 后验推断结果的质量严重依赖于指定的模型. 因此对模型的评价是不可忽略的一个重要方面. 本节介绍两种常用的贝叶斯模型评价准则, 包括贝叶斯预测信息准则 (**BPIC**) 和偏差信息准则 (**DIC**).
- 由于 AIC 和 BIC 准则是经典统计方法中常用的模型选择的准则, 下面首先对其作简要的介绍, 然后引入 BPIC 和 DIC 准则, BPIC 准则和 DIC 准则是对 AIC 和 BIC 准则的推广.

## 2. AIC 和 BIC 准则

- 极大似然原理是统计学中参数推断的一种重要方法，日本统计学家 Akaike (1974) 基于这个原理提出了模型选择的一个准则，称为 AIC 准则，即

$$AIC = -2 \ln f(\mathbf{x}_n | \hat{\theta}_{MLE}) + 2p,$$

其中  $\hat{\theta}_{MLE}$  是  $\theta$  的极大似然估计 (MLE),  $p$  是参数向量的维数,  $2p$  是惩罚项. 最优模型可以通过最小化 AIC 得到.

- Schwartz (1978) 在一定的条件下引入 BIC 准则，即

$$BIC = -2 \ln f(\mathbf{x}_n | \hat{\theta}_{MLE}) + 2p \ln n,$$

其中  $p$  是参数向量的维数,  $n$  为样本大小,  $p \ln n$  是惩罚项. 最优模型可以通过最小化 BIC 得到.

## 2. AIC 和 BIC 准则

- 极大似然原理是统计学中参数推断的一种重要方法，日本统计学家 Akaike (1974) 基于这个原理提出了模型选择的一个准则，称为 AIC 准则，即

$$AIC = -2 \ln f(\mathbf{x}_n | \hat{\theta}_{MLE}) + 2p,$$

其中  $\hat{\theta}_{MLE}$  是  $\theta$  的极大似然估计 (MLE),  $p$  是参数向量的维数,  $2p$  是惩罚项. 最优模型可以通过最小化 AIC 得到.

- Schwartz (1978) 在一定的条件下引入 BIC 准则，即

$$BIC = -2 \ln f(\mathbf{x}_n | \hat{\theta}_{MLE}) + 2p \ln n,$$

其中  $p$  是参数向量的维数,  $n$  为样本大小,  $p \ln n$  是惩罚项. 最优模型可以通过最小化 BIC 得到.

### 3. 贝叶斯预测信息准则 (BPIC)

- 考虑下列两个假设： (a) 参数模型  $f(\mathbf{x}|\boldsymbol{\theta})$  包含了真实的模型  $g(\mathbf{x}) = f(\mathbf{x}; \boldsymbol{\theta}_0)$ ,  $\boldsymbol{\theta}_0 \in \Theta$ , 且指定的模型并不远离真实模型. (b) 对数先验的阶为  $\ln \pi(\boldsymbol{\theta}) = O_p(1)$ .
- 在上述两个假定和某些正则条件下, Ando (2007) 提出贝叶斯预测信息准则 (BPIC):

$$BPIC = -2 \int_{\Theta} \ln f(\mathbf{x}_n|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}|\mathbf{x}_n) d\boldsymbol{\theta} + 2p,$$

其中  $p$  为模型中的参数个数. 最优模型通过最小化 BPIC 得到.



### 3. 贝叶斯预测信息准则 (BPIC)

- 考虑下列两个假设： (a) 参数模型  $f(\mathbf{x}|\boldsymbol{\theta})$  包含了真实的模型  $g(\mathbf{x}) = f(\mathbf{x}; \boldsymbol{\theta}_0)$ ,  $\boldsymbol{\theta}_0 \in \Theta$ , 且指定的模型并不远离真实模型. (b) 对数先验的阶为  $\ln \pi(\boldsymbol{\theta}) = O_p(1)$ .
- 在上述两个假定和某些正则条件下, Ando (2007) 提出贝叶斯预测信息准则 (BPIC):

$$BPIC = -2 \int_{\Theta} \ln f(\mathbf{x}_n|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}|\mathbf{x}_n) d\boldsymbol{\theta} + 2p,$$

其中  $p$  为模型中的参数个数. 最优模型通过最小化 BPIC 得到.

### 3. 贝叶斯预测信息准则 (BPIC)

- 在实际应用中，对数似然的后验均值往往没有解析表达，此时可以使用蒙特卡洛逼近：

$$\int_{\Theta} \ln f(\mathbf{x}_n | \boldsymbol{\theta}) \pi(\boldsymbol{\theta} | \mathbf{x}_n) d\boldsymbol{\theta} \approx \frac{1}{L} \sum_{j=1}^L \ln f(\mathbf{x}_n | \boldsymbol{\theta}^{(j)}),$$

其中  $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(L)}$  为从后验分布  $\pi(\boldsymbol{\theta} | \mathbf{x}_n)$  中抽取的后验样本.

- 这个准则适合于具有较弱的先验信息的情形.

### 3. 贝叶斯预测信息准则 (BPIC)

- 在实际应用中，对数似然的后验均值往往没有解析表达，此时可以使用蒙特卡洛逼近：

$$\int_{\Theta} \ln f(\mathbf{x}_n | \boldsymbol{\theta}) \pi(\boldsymbol{\theta} | \mathbf{x}_n) d\boldsymbol{\theta} \approx \frac{1}{L} \sum_{j=1}^L \ln f(\mathbf{x}_n | \boldsymbol{\theta}^{(j)}),$$

其中  $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(L)}$  为从后验分布  $\pi(\boldsymbol{\theta} | \mathbf{x}_n)$  中抽取的后验样本.

- 这个准则适合于具有较弱的先验信息的情形.

## 4. 偏差信息准则 (DIC)

- 令  $D(\theta) = -2 \ln f(\mathbf{x}_n | \theta)$ , 它是常用的模型偏差的一种度量. Spiegelhalter 等 (2002) 指出对数似然期望的后验期望,  $\bar{D} = E[D(\theta) | \mathbf{x}_n]$ , 可以作为模型拟合程度的一个贝叶斯度量. 一个模型拟合数据的程度越高,  $\bar{D}$  越小. 下面定义有效参数个数来刻画模型的复杂程度:

$$p_D = \bar{D} - D(\bar{\theta}_n) = 2 \ln f(\mathbf{x}_n | \bar{\theta}_n) - 2 \int_{\Theta} \ln f(\mathbf{x}_n | \theta) \pi(\theta | \mathbf{x}_n) d\theta,$$

其中  $\bar{\theta}_n$  为  $\theta$  的后验分布的均值.

- Spiegelhalter 等 (2002) 定义偏差信息准则 (DIC) 为

$$DIC = \bar{D} + p_D = -2 \int_{\Theta} \ln f(\mathbf{x}_n | \theta) \pi(\theta | \mathbf{x}_n) d\theta + p_D.$$

其中第一项, 即  $\bar{D}$  可解释为模型拟合程度的一个度量, 越小越好; 第二项, 即  $p_D$  被认为是模型复杂性的一种度量.

## 4. 偏差信息准则 (DIC)

- 令  $D(\theta) = -2 \ln f(\mathbf{x}_n | \theta)$ , 它是常用的模型偏差的一种度量. Spiegelhalter 等 (2002) 指出对数似然期望的后验期望,  $\bar{D} = E[D(\theta) | \mathbf{x}_n]$ , 可以作为模型拟合程度的一个贝叶斯度量. 一个模型拟合数据的程度越高,  $\bar{D}$  越小. 下面定义有效参数个数来刻画模型的复杂程度:

$$p_D = \bar{D} - D(\bar{\theta}_n) = 2 \ln f(\mathbf{x}_n | \bar{\theta}_n) - 2 \int_{\Theta} \ln f(\mathbf{x}_n | \theta) \pi(\theta | \mathbf{x}_n) d\theta,$$

其中  $\bar{\theta}_n$  为  $\theta$  的后验分布的均值.

- Spiegelhalter 等 (2002) 定义偏差信息准则 (DIC) 为

$$DIC = \bar{D} + p_D = -2 \int_{\Theta} \ln f(\mathbf{x}_n | \theta) \pi(\theta | \mathbf{x}_n) d\theta + p_D.$$

其中第一项, 即  $\bar{D}$  可解释为模型拟合程度的一个度量, 越小越好; 第二项, 即  $p_D$  被认为是模型复杂性的一种度量.

## 偏差信息准则 (DIC)

- 上述定义的 DIC 可以改写为

$$DIC = D(\bar{\theta}_n) + 2p_D = -2 \ln f(\mathbf{x}_n | \bar{\theta}_n) + 2p_D,$$

其中  $\bar{\theta}_n$  为后验期望. 从形式上看它与 AIC 很相似, 因此可以认为 DIC 是  $AIC = D(\hat{\theta}_{MLE}) + 2p$  的一个推广, 此处  $\hat{\theta}_{MLE}$  为  $\theta$  的极大似然估计. 对非分层模型而言, 当  $n$  充分大时有  $p \approx p_D$ ,  $\hat{\theta}_{MLE} \approx \bar{\theta}_n$ , 从而  $DIC \approx AIC$ .

- 上述定义的另一个优点是 DIC 可以通过 MCMC 方法 (将在第 6 章介绍) 容易计算其结果. DIC 准则可被用于各种贝叶斯模型选择问题.

## 偏差信息准则 (DIC)

- 上述定义的 DIC 可以改写为

$$DIC = D(\bar{\theta}_n) + 2p_D = -2 \ln f(\mathbf{x}_n | \bar{\theta}_n) + 2p_D,$$

其中  $\bar{\theta}_n$  为后验期望. 从形式上看它与 AIC 很相似, 因此可以认为 DIC 是  $AIC = D(\hat{\theta}_{MLE}) + 2p$  的一个推广, 此处  $\hat{\theta}_{MLE}$  为  $\theta$  的极大似然估计. 对非分层模型而言, 当  $n$  充分大时有  $p \approx p_D$ ,  $\hat{\theta}_{MLE} \approx \bar{\theta}_n$ , 从而  $DIC \approx AIC$ .

- 上述定义的另一个优点是 DIC 可以通过 MCMC 方法 (将在第 6 章介绍) 容易计算其结果. DIC 准则可被用于各种贝叶斯模型选择问题.