

第三章 统计决策与贝叶斯估计

第3.1节 统计决策的基本概念

第3.2节 统计决策中的常用分布族

第3.3节 贝叶斯估计



前言

20世纪40年代，Wald提出了把统计推断问题看成是人与自然的一种博弈过程，由此建立了统计决策理论.

- 统计决策建立在统计分析和预测的基础上，是一种定量决策
- 统计决策是在不确定情况下，应用概率理论进行决策的计算和分析，是一种概率决策

贝叶斯估计是贝叶斯统计的主要部分，它是利用决策理论研究参数估计问题.本章将主要讨论贝叶斯方法在参数估计中的应用问题.



第3.1节 统计决策的基本概念

一、统计决策问题的三个要素

二、统计决策函数及其风险函数



一、统计决策问题的三个要素

在前几章讲的统计问题，都可以归结为一个统计决策问题，也就是建立所谓的统计决策函数.统计决策问题由三个因素组成，首先来看第一个因素：

1、样本空间和分布族

样本空间 设样本 $(X_1, X_2, \dots, X_n)^T$ 来自总体 $F(x, \theta)$, θ 未知, $\theta \in \Theta$, 则样本所有可能值组成的集合称为样本空间, 记为 X .



分布族 设样本 $(X_1, X_2, \dots, X_n)^T$ 来自总体 $F(x, \theta)$,
 θ 未知, $\theta \in \Theta$, 其联合分布为

$$F(x_1, x_2, \dots, x_n, \theta) = \prod_{i=1}^n F(x_i, \theta)$$

若记 $F^* = \{\prod_{i=1}^n F(x_i, \theta), \theta \in \Theta\}$, 则称 F^* 为样本 $(X_1, X_2, \dots, X_n)^T$ 的概率分布族, 简称分布族.



例1 (p81例3. 1) 设总体 X 服从两点分布 $B(1,p)$, p 为未知参数, $0 \leq p \leq 1$, $(X_1, X_2, \dots, X_n)^T$ 是取自总体 X 的样本, 试求其样本空间以及分布族.

解 由于是两点分布, 因而样本的取值只有0,1, 则样本空间为

$$X = \{(x_1, x_2, \dots, x_n) : x_i = 0, 1, \quad i = 1, 2, \dots, n\}$$

分布族为

$$F^* = \{p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}, x_i = 1, 0, \quad i = 1, 2, \dots, n, 0 \leq p \leq 1\}$$



2、决策空间(或称判决空间)

决策 对每个统计问题的具体回答, 就称为一个决策.

例如, 参数的点估计, 每一个估计值就是一个决策.

决策空间 一个统计问题中, 可能选取的全部决策组成的集合为决策空间, 记为 \mathcal{A} . 决策空间至少应含有两个决策, 否则就无需选择, 从而也形成不了一个统计决策问题.

例如, 设总体分布服从 $N(\mu, \sigma^2)$, 对未知参数 μ 进行估计, 由于 μ 在 $(-\infty, +\infty)$ 中取值, 因而其决策空间为 $\mathcal{A} = (-\infty, +\infty)$.



例2 (p82例3. 2) 某厂打算根据各年度市场的销售来决定下一年度应该扩大生产还是缩减生产，或者维持原状，这样其决策空间为

$$\mathcal{A} = \{\text{扩大生产}, \text{缩减生产}, \text{维持原状}\}$$

3、损失函数

通常情况下，做任何决策以后，总会有某种后果，由此可以带来某种收益和损失.为了以数量化的方式描述这种收益和损失，为此需要引入损失函数.

例3 (p82例3. 3) 设总体 X 服从正态分布 $N(\theta, 1)$, θ 为未知参数，参数空间为 $\Theta = (-\infty, +\infty)$ ，决策空间可以设



为 $\mathcal{A}=(-\infty,+\infty)$,此决策问题的损失函数可以为:

(1) 设 d 为 θ 的点估计值, 损失函数可以设为

$$L(\theta, d) = (\theta - d)^2$$

(2) 设 $[d_1, d_2]$ 为 θ 的区间估计值, 损失函数可以设为

$$L(\theta, d) = (d_2 - d_1)$$

也可以设为

$$L(\theta, d) = 1 - I_{[d_1, d_2]}(\theta) = \begin{cases} 0, & \theta \in [d_1, d_2] \\ 1, & \theta \notin [d_1, d_2] \end{cases}$$



常见的损失函数

(1) 线性损失函数

$$L(\theta, d) = \begin{cases} k_0(\theta - d), & d \leq \theta, \\ k_1(d - \theta), & d > \theta, \end{cases}$$

其中 k_0, k_1 为常数，它们可以反映大于或小于参数时带来不同的损失. 当 $k_0 = k_1 = 1$ 时

$$L(\theta, d) = |d - \theta|$$

此损失函数为绝对损失函数.



(2) 平方损失函数

$$L(\theta, d) = (\theta - d)^2$$

(3) 凸损失函数

$$L(\theta, d) = \lambda(\theta)W(|\theta - d|)$$

其中 $\lambda(\theta) > 0$ 是 θ 的已知函数，且有限， $W(t)$ 是 $t > 0$ 上的单调非降函数且 $W(0) = 0$.

(4) 多元二次损失函数

当参数 θ 以及决策的 d 为多维向量时，二次损失为



$$L(\theta, d) = (d - \theta)^T A (d - \theta)$$

其中 $\theta = (\theta_1, \theta_2, \dots, \theta_p)^T$, $d = (d_1, d_2, \dots, d_p)^T$, A 为 $p \times p$ 阶正定矩阵, p 为大于1 的自然数. 当 A 为对角矩阵时, 即 $A = \text{diag}(\omega_1, \omega_2, \dots, \omega_p)$, 则 p 元损失函数为

$$L(\theta, d) = \sum_{i=1}^p \omega_i (\theta_i - d_i)^2$$

注 由于在统计问题中, 进行的统计推断总是有误差, 因而损失一定存在, 因而一般都会假设损失函数为非负的. 二次损失为参数点估计常用的损失函数.



二、统计决策函数及其风险函数

1. 统计决策函数

给定统计决策问题的三要素后，在损失小的前提下，选择一个好决策函数就成为核心问题。

定义3.1 定义在样本空间 X 上，取值于决策空间 \mathcal{A} 内的函数 $d(x)$ ，称为统计决策函数，简称为决策函数。

注 决策函数其实就是决策问题的一个“行动方案”。
对于统计问题而言，决策函数为统计量。



例4 (p84) 设总体 X 服从正态分布 $N(\mu, \sigma^2)$, σ^2 为已知, $(X_1, X_2, \dots, X_n)^T$ 取自 X 的样本, 试求参数 μ 点估计和区间估计的决策函数.

解根据上一章的结论, 参数 μ 点估计的决策函数为

$$d(x) = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

参数 μ 区间估计的决策函数为

$$d(x) = \left[\bar{x} - u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$



2. 风险函数

由于损失函数 L 与决策函数 $d(\mathbf{x})$ 有关,而决策函数是随机变量,因而损失函数也为随机变量。这样损失函数与样本 \mathbf{X} 的取值有关,因而需要构造一个更好的指标来衡量决策函数的好坏. 这就是风险函数.

定义3.2 设样本空间和分布族分别为 \mathcal{X} 和 \mathcal{F}^* , 决策空间为 \mathcal{A} , 损失函数为 $L(\theta, d)$, 决策函数为 $d(X)$, 则参数 θ 的决策函数 $d(X)$ 引起的风险函数为 $R(\theta, d)$ 定义为

$$R(\theta, d) = E_{\theta}(L(\theta, d(X))) = E_{\theta}(L(\theta, d(X_1, X_2, \dots, X_n)))$$



注 由定义可以看到，风险函数是决策 d 的平均损失。从定义可以看到，风险越小，决策越好，由此可以给出判断决策函数优良性准则。

定义3.3 设 $d_1(X)$ 和 $d_2(X)$ 为统计决策问题的两个决策函数，若其风险函数满足不等式

$$R(\theta, d_1) \leq R(\theta, d_2), \quad \forall \theta \in \Theta$$

且存在一些 θ 使得不等式严格成立，即 $R(\theta, d_1) < R(\theta, d_2)$ ，则称决策函数 d_1 一致优于 d_2 ，如果等式成立即 $R(\theta, d_1) = R(\theta, d_2), \quad \forall \theta \in \Theta$ ，则二者等价。



定义3.4 设 $D = \{d(X)\}$ 是一切定义在样本空间上取值于决策空间 \mathfrak{R} 上的决策函数的全体, 若存在一个决策函数 $d^*(X)(d^*(X) \in D)$, 使得对任意一个 $d(X) \in D$, 都有

$$R(\theta, d^*) \leq R(\theta, d), \quad \forall \theta \in \Theta$$

则称决策函数 $d^*(X)$ 为一致最小风险决策函数, 或称为一致最优决策函数.

注 从上述定义可以看到, 决策函数的优良性与损失函数有关, 因而优良性会因损失函数而变化.



例5 (p85例3.4) 设总体 X 服从正态分布 $N(\mu, 1)$, μ 为未知参数, 参数空间为 $\Theta = (-\infty, +\infty)$, $(X_1, X_2, \dots, X_n)^T$ 取自 X 的样本, 若选取损失函数为平方损失

$$L(\mu, d) = (d - \mu)^2$$

试求参数 μ 任一估计 $d(X)$ 的风险函数?

解 根据风险函数的定义可知

$$R(\mu, d) = E_{\mu}(L(\mu, d(X))) = E_{\mu}(d - \mu)^2$$

若 $E(d(X)) = \mu$, 则其风险函数为

$$R(\mu, d) = E_{\theta}(d - E(d))^2 = D_{\mu}(d(X))$$



若 $d(X) = \bar{X}$, 则 $R(\mu, d) = D(\bar{X}) = \frac{1}{n}$,

若 $d(X) = X_1$, 则 $R(\mu, d) = D(X_1) = 1$,

显然, 当 $n > 1$ 时, 后者的风险大于前者的风险,
因而 \bar{X} 在平方损失的条件下优于 X_1 .

例6 (p85例3.5) 设 x_1 和 x_2 是从下列分布中获得
两个观察值,

$$P\{X = \theta - 1\} = P\{X = \theta + 1\} = 0.5, \quad \theta \in \Theta = R$$

决策空间为 $\mathcal{A} = (-\infty, +\infty)$, $(X_1, X_2, \dots, X_n)^T$ 取自 X
的样本, 若选取损失函数为



$$L(\theta, d) = 1 - I(d), \text{ 其中 } I(d) = \begin{cases} 1, & d = \theta, \\ 0, & d \neq \theta, \end{cases}$$

试求参数 θ 的估计 $d_1 = \frac{x_1 + x_2}{2}, d_2 = x_1 - 1,$

$$d_3 = \begin{cases} \frac{x_1 + x_2}{2}, & x_1 \neq x_2 \\ x_1 - 1, & x_1 = x_2 \end{cases} \text{ 的风险函数?}$$

解 根据风险函数的定义可知

$$R(\theta, d_1) = (1 - I(d_1 = \theta)) \square P\{d_1 = \theta\} + \\ (1 - I(d_1 \neq \theta)) \square P\{d_1 \neq \theta\}$$



$$= P\{d_1 \neq \theta\} = 1 - P\{d_1 = \theta\} = 1 - P\{x_1 \neq x_2\} = 0.5$$

$$R(\theta, d_2) = 1 - P\{d_2 = \theta\} = 1 - P\{x_1 = \theta + 1\} = 0.5$$

$$\begin{aligned} R(\theta, d_3) &= 1 - P\{d_3 = \theta\} \\ &= 1 - P\{x_1 \neq x_2 \text{ 或 } x_1 = \theta + 1\} = 0.25 \end{aligned}$$

显然， d_3 一致优于 d_1 和 d_2 ，这个优良性依赖于损失函数以及决策函数的范围。



Thank You!

