第三章　统计决策与贝叶斯估计

3.1 统计决策的基本概念

3.2 统计决策中常用分布族

3.3 贝叶斯估计

3.4 minimax估计

3.5 经验贝叶斯估计（略）

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.

贝叶斯估计是贝叶斯统计的主要部分，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.

贝叶斯

英国数学家，英国皇家学会会员。首先将归纳推理法用于概率论基础理论，创立了贝叶斯统计理论，对于统计决策函数、统计推断、统计的估算等做出了贡献。著有《机会的学说概论》。他的遗作《机会问题的解法》对于现代概率论和数理统计产生了重要的影响。贝叶斯所采用的许多术语被沿用至今。

Thomas Bayes
约1701-1761

从20世纪20~30年代开始，概率统计学出现了“频率学派”和“贝叶斯学派”的争论，至今，两派的恩恩怨怨仍在继续。

Wald

瓦尔德(Wald，Abrahom，1902～1950)美籍罗马尼亚数理统计学家

发展了统计决策理论，提出了一般的判决问题，引进了损失函数、风险函数、极大极小原则和最不利先验分布等重要概念

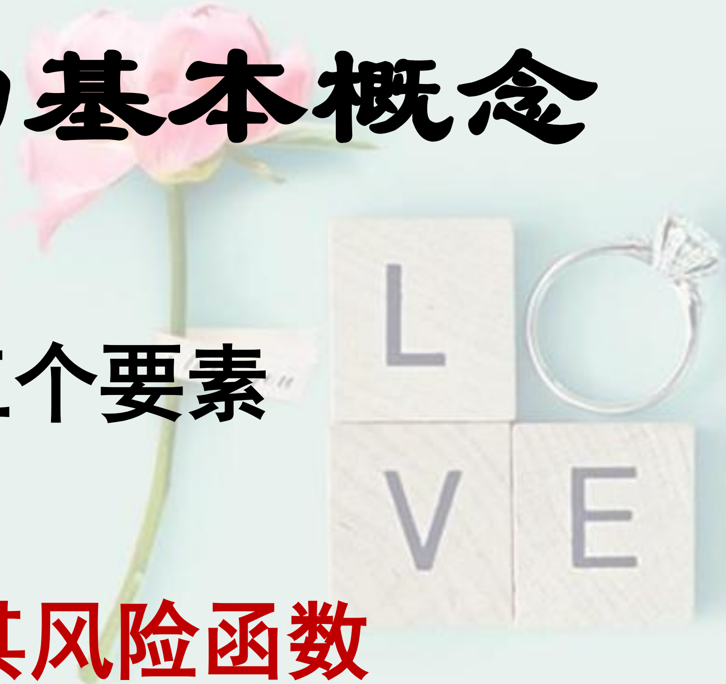
著有《统计决策函数论》(1950)

Wald，Abrahom，1902～1950

3.1 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

1、样本空间和分布族

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

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

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

若记$F^* = \{\prod_{i=1}^{n} F(x_i, \theta), \theta \in \Theta\}$,则称$F^*$为样本$(X_1, X_2, \cdots, X_n)^T$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,$0 < p < 1$,$(X_1, X_2, \cdots, X_n)^T$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（n维）

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

分布族为

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

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，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 $\mathcal{A}$.

例如，设总体服从$N(\mu,\sigma^2)$,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$\mathcal{A}=(-\infty,+\infty)$.

决策空间就是参数的取值范围。

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

$\mathcal{A}$={扩大生产，缩减生产，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（4种）

(1) 线性损失函数

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

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$k_0=k_1$时

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

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

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

注 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，一般损失最小是零，所以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 $d(x)$: 样本空间X $\rightarrow$ 决策空间$\mathcal{A}$,

称为统计决策函数，简称决策函数.

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$N(\mu,\sigma^2),\sigma^2$为已知,$(X_1,X_2,\cdots,X_n)^T$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

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

注： 由损失函数L的定义，

L是决策函数$d(x)$和参数θ的函数。

决策函数$d(x)$是样本的函数，是随机变量，

$\Rightarrow$ 损失函数L也是样本的函数，也是随机变量。

决策函数$d(x)$的好坏由损失函数L来评价。

损失函数L越小，说明决策函数$d(x)$越好。

但损失函数L是随机变量，对随机变量评价只能通过其均值的大小来衡量。

2. 风险函数

定义3.2 设样本空间和分布族分别为X和F^*,决策空间为$\mathcal{A}$,决策函数为$d(X)$,损失函数为$L(\theta,d)$,则定义

$$R(\theta,d)=E_X(L(\theta,d(X))=E_X(L(\theta,d(X_1,X_2,\cdots,X_n))$$

为参数θ的决策函数$d(X)$引起的风险函数。

注 $R(\theta,d)$ 是$L(\theta,d)$的均值，是决策d的平均损失.

$R(\theta,d)$越小，就是平均损失越小，决策就越好,

风险函数$R(\theta,d)$是θ的函数！

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

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

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$R(\theta,d_1) = R(\theta,d_2)$，$\forall\theta\in\Theta$，则二者等价.

一致最小风险决策函数

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$\mathcal{A}$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

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

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 （1）上述定义都是针对某个具体的损失函数而言，当损失函数改变了，相应的结论有可能就会变。

（2）定义3.4还针对某个决策函数类而言，一旦决策函数类改变了，一致最优性也可能就不具备了。

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$\Theta=(-\infty,+\infty)$,$(X_1,X_2,\cdots,X_n)^T$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

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

若$E(d(X))=\mu$,则其风险函数为

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

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样本,

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

决策空间$\mathcal{A}=R$,若取损失函数$L(\theta,d)$为

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

$$=\begin{cases}0, & d=\theta,\\ 1, & d\neq\theta,\end{cases}$$

试求参数θ的估计$d_1=\dfrac{X_1+X_2}{2}$， $d_2=X_1-1$，

$$d_3=\begin{cases}\dfrac{X_1+X_2}{2}, & X_1\neq X_2\\ X_1-1, & X_1=X_2\end{cases}$$ 的风险函数？

解 根据风险函数的定义可知 $L(\theta,d)=\begin{cases}0, & d=\theta,\\ 1, & d\neq\theta,\end{cases}$

$$R(\theta,d)=EL(\theta,d)=0\cdot P\{d=\theta\}+1\cdot P\{d\neq\theta\}$$

$$=P\{d\neq\theta\}$$

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

针对 $d_1=\dfrac{X_1+X_2}{2}$

$$R(\theta,d_1)=EL(\theta,d_1)=P\{d_1\neq\theta\}$$

$$=1-P\{d_1=\theta\}=1-P\{X_1+X_2=2\theta\}$$

$$=1-P\{X_1\neq X_2\}$$

$$=1-0.5=0.5$$

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

针对

$$d_2 = X_1 - 1,$$

$$R(\theta,d_2)=EL(\theta,d_2)$$
$$=P\{d_2\neq\theta\}=P\{X_1\neq\theta+1\}=0.5$$

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

针对 $d_3=\begin{cases}\dfrac{X_1+X_2}{2}, & X_1\neq X_2\\ X_1-1, & X_1=X_2\end{cases}$

$$R(\theta,d_3)=P\{d_3\neq\theta\}$$

$$=P\{\frac{X_1+X_2}{2}\neq\theta, X_1\neq X_2\}$$

$$+P\{X_1-1\neq\theta, X_1=X_2\}$$

$$=0+P\{X_1=\theta-1, X_2=\theta-1\}=0.25$$

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!