应用数理统计——第四章

4.1 统计决策理论

一、统计决策思想

统计学是一门“在不确定情况下做出决策的一门科学”，参数估计、假设检验等问题都可归结为“统计决策”理论。

1. 统计决策三要素

统计决策中存在三个关键性要素：

（1） 参数空间Θ：未知参数的全部可能取值的集合；

（2） 决策空间D：实际生活中所采取的行为构成的集合；

（3） 损失函数$L(\theta,d)$：参数为θ时采取行为d而引起的损失。

2. 损失函数

（1） 平方损失：$L(\theta,d)=(\theta-d)^2$

（2） 绝对损失：$L(\theta,d)=|\theta-d|$

3. 风险函数

采取某一行为d所对应的决策$d=\delta(X_1,...,X_n)$而产生的平均损失，称为该决策的风险函数。即，决策δ对应的损失函数的期望$R(\theta,\delta)=E[L(\theta,\delta(X_1,...,X_n))]$，称为该决策的风险函数。

二、统计决策观点下的估计与检验

1.参数点估计：

在参数点估计问题中，常选取平方损失对参数进行估计，此时的风险函数为估计量的均方误差。其中，一致最小方差无偏估计在无偏估计类中，在平方损失下的风险是最小的。

2.假设检验：

在假设检验问题中，常选取 0-1 损失作为损失函数。该损失函数下的风险函数，常用功效函数进行刻画。

$$L(\theta,k)=\begin{cases}0, & \text{如果}\theta\in\Theta_k\\ 1, & \text{如果}\theta\in\Theta_{k\text{-}1}\end{cases}\qquad R(\theta,\phi)=\begin{cases}\beta_\phi(\theta), & \text{当}\theta\in\Theta_0\\ 1-\beta_\phi(\theta), & \text{当}\theta\in\Theta_1\end{cases}$$

3.区间估计：

在区间估计问题中，需选取两类损失函数：一类为 0-1 损失，用于刻画区间的置信度；一类为区间长度，用于刻画区间的精度。

$$L_1(\theta,(a,b))=\begin{cases}0, & \text{当} a<\theta<b\\ 1, & \text{否则}\end{cases} \qquad L_2(\theta,(a,b))=b-a$$

对于任意一个区间估计 $\delta(x)=(\varphi_1(x),\varphi_2(x))$，都对应着刻画置信度的风险函数：

$$R_1(\theta,\delta)=1-P\{\varphi_1(x)<\theta<\varphi_2(x)\},$$

和刻画精度的风险函数：

$$R_2(\theta,\delta)=E[\varphi_2(x)-\varphi_1(x)],$$

该风险函数利用区间平均长度进行描述。

三、常见的两个决策原则

1.极小化极大方法

对每一个决策都考虑最坏的可能状态，从最坏的状态中选取最好的一个作为决策。

2.Bayes 决策

对于参数 θ 定义一个概率分布，即先验分布。把风险函数 $R(\theta,d)$ 关于这个先验分布取数学期望，即 Bayes 风险。在决策空间 D 中取使得 Bayes 风险最小的那一个作为决策。

4.2Bayes 理论

一、Bayes 公式与 Bayes 统计理论

1. Bayes 公式

$$P(B_i \mid A)=\frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n}P(B_j)P(A|B_j)}$$

2.Bayes 统计理论

（1）频率学派观点：

对于总体的未知参数θ来说，无论做点估计、区间估计还是假设检验，θ始终仅仅是一个常数。在抽取样本之前，对于参数θ没有任何了解，所有关于θ的信息都全部包含在样本之中。

（2）Bayes 学派观点：

在抽取样本之前，对参数θ已经有一些认识，即：先验知识。这种知识可通过一个概率分布体现出来，即：先验分布$h(\theta)$。先验分布综合了在抽样之前对θ的全部信息。

3.后验分布

（1）后验分布的基本思想

在随机向量(X,θ)中，我们只能通过对X的观察去间接地推导出θ。概率函数$f(x,\theta)$被看成样本x关于θ的条件分布$f(x|\theta)$，因此抽取一组样本x，也就意味着对θ得到了新的信息，它将修正原来的观点（θ的先验分布），这种对θ的新认识通过θ关于样本x的条件分布（后验分布）体现出来。

对θ的估计和检验都依据后验分布来进行。

（2）后验分布的构造

a.利用样本观察值，构建似然函数$f(x_1,\dots,x_n \mid \theta)$

b.根据先验分布和似然函数构建联合分布$f(x_1,\dots,x_n,\theta)=\pi(\theta)f(x_1,\dots,x_n \mid \theta)$

c.对联合分布求关于样本的边缘分布$\int_{\Theta}\pi(\theta)f(x_1,\dots,x_n|\theta)d\theta$

d.利用联合分布和边缘分布，求后验分布

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

二、Bayes 统计推断理论

得到θ关于样本x的后验分布$h(\theta|X)$以后，其余的如样本观察值、样本分布、先验分布等都不再有用，对θ的处理全部依赖于后验分布。

（1） Bayes 点估计：根据损失函数取成后验分布的中位数或者数学期望；
（2） Bayes 区间估计：直接由后验分布构造；
（3） Bayes 检验：由后验分步计算θ落在零假设或对立假设中的概率，哪个大就接受哪个。

三、先验分布的确定

（1）根据以往经验或过去资料确定；

（2）广义先验分布：同等无知原则，即：θ以相同的概率可以取参数空间Θ中任何一个。

四、Beta 分布与 Gamma 分布

1. Beta 分布与 Gamma 分布表

分布名称	密度函数	期望	方差	平方期望
Beta 分布	【$X\sim Beta(\alpha,\beta)$】 $f(x)=\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}$， $B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$\frac{\alpha(\beta-\alpha)}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Gamma 分布	【$X\sim\Gamma(\alpha,\lambda)$】 $f(x)=\begin{cases}\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, & x\ge 0\\ 0, & x<0\end{cases}$， $\Gamma(\alpha)=\int_0^{+\infty}x^{\alpha-1}e^{-x}dx$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$	$\frac{\alpha(\alpha+1)}{\lambda^2}$
逆 Gamma 分布	【$X\sim I\Gamma(\alpha,\lambda)$】 $f(x)=\begin{cases}\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{-\alpha-1}e^{-\frac{\lambda}{x}}, & x\ge 0\\ 0, & x<0\end{cases}$， $\Gamma(\alpha)=\int_0^{+\infty}x^{\alpha-1}e^{-x}dx$	$\frac{\lambda}{\alpha-1}$	$\frac{\lambda^2}{(\alpha-1)^2(\alpha-2)}$	$\frac{\lambda^2+\lambda(\alpha-1)(\alpha-2)}{(\alpha-1)^2(\alpha-2)}$

巴斯卡分布	【 $X \sim NB(r,p)$ 】 $P(x=k)=C_{k-1}^{r-1}p^{r-1}(1-p)^{k-r}$， $p,k=r,r+1,\ldots$	$\dfrac{k(1-p)}{p}$	$\dfrac{k(1-p)}{p^2}$	$\dfrac{k(1-p)^2}{p^2}$

2.常用结论

（1） $B(\alpha,\beta)=\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ 被称为 β 函数，是贝叶斯估计的常用化简公式;

（2）$\Gamma(\alpha+1)=\alpha\Gamma(\alpha),\quad \Gamma(\alpha+1)=\alpha!$

五、核技巧

如果函数 $\varphi(x)$ 与函数 $f(x)$ 只相差一个常数因子，则称 $\varphi(x)$ 为 $f(x)$ 的核，记为 $f(x)\propto\varphi(x)$。

如果函数 $\varphi(x)$ 是密度函数为 $f(x)$ 的分布的核，则 $\varphi(x)$ 也服从密度函数为 $f(x)$ 的分布。

4.3 题型总结

一、统计决策问题

1.题型描述： 利用已知条件，采用极小化极大方法或 Bayes 方法作出决策。

2.解题策略：

根据方法的定义对实际问题作出决策。

3.题解：

例：一位统计学家准备买一辆汽车，计划使用两年行驶 40,000 英里。他把选择的范围缩小到两部：汽车 A 价格＄5,000，平均每加仑汽油行驶 20 英里，汽车 B 价格＄6,700，平均每加仑汽油行驶 40 英里；假定汽油平均价格有三种，每加仑＄1，＄2，＄3，问这位统计学家应该选择哪一辆车？

解：

如果选择汽车 A 需要 2000 加仑汽油；如果选择汽车 B 需要 1000 加仑汽油；可列出下表：

油价	＄1	＄2	＄3
A 总费用	＄7000	＄9000	＄11000
B 总费用	＄7700	＄8700	＄9700

若采用极小化极大方法：

A 车的最大费用为＄11000，B 车的最大费用为＄9700；因此选择 B 车。

若采用 Bayes 方法：

给定油价的先验概率：

＄1	＄2	＄3
3/4	1/8	1/8

则 A 车的平均总费用为＄7750，B 车的平均总费用为＄8075；因此选择 A 车。

二、给定损失函数的 Bayes 估计问题

1.题型描述： 题设中给定除“平方损失”和“绝对损失”外的其他损失函数。

2.解题策略：

对风险函数求期望，期望的极小值即为参数的估计值。

3.题解：

例：设 $F(x;\theta)$ 为总体 ξ 的分布函数，θ 为未知参数，$\xi_1,\ldots,\xi_n$ 为 ξ 的样本。

证明：当损失函数为 $L(\theta,d)=\lambda(\theta)(\theta-d)^2, 0<\lambda(\theta)<\infty$ 时，如果

$$E\{\lambda(\theta)[\theta-d(\xi_1,\ldots,\xi_n)]^2 \mid \xi_1,\xi_2,\ldots,\xi_n\}<\infty,$$

则 θ 的贝叶斯估计量为

$$\tilde{d}(\xi_1,...,\xi_n)=E[\theta\lambda(\theta)\,|\,\xi_1,...,\xi_n]\,/\,E[\lambda(\theta)\,|\,\xi_1,...,\xi_n]$$

证明：

对损失函数求均值得：

$B(d)=E(L(\theta,d))=E[E(\lambda(\theta)(\theta-d(\xi_1,...,\xi_n))^2\,|\,\xi_1,...,\xi_n)]$，记 $\eta=\{\xi_1,...,\xi_n\}$

则 $B(d)=E[\lambda(\theta)(\theta-d(\eta))^2\,|\,\eta]$

$$=E[\lambda(\theta)\theta^2\,|\,\eta]-2d(\eta)E[\theta\lambda(\theta)\,|\,\eta]+d^2(\eta)E[\lambda(\theta)\,|\,\eta]$$

对 $d(\eta)$ 求导得 $B'(d)=-2E[\theta\lambda(\theta)\,|\,\eta]+2d(\eta)E[\lambda(\theta)\,|\,\eta]$

令 $B'(d)=0$ 得 $d(\eta)=E[\theta\lambda(\theta)\,|\,\eta]/E[\lambda(\theta)\,|\,\eta]$

因为 $B''(d)=2E[\lambda(\theta)\,|\,\eta]>0$，所以 $d(\eta)=E[\theta\lambda(\theta)\,|\,\eta]/E[\lambda(\theta)\,|\,\eta]$ 为极小值点。

即 $\tilde{d}(\xi_1,...,\xi_n)=E[\theta\lambda(\theta)\,|\,\xi_1,...,\xi_n]\,/\,E[\lambda(\theta)\,|\,\xi_1,...,\xi_n]$。

证毕。

三、连续型 Bayes 参数估计问题

1.题型描述：求解某一参数的 Bayes 估计，先验分布为连续型随机变量。

2.解题策略：

（1）利用 Bayes 公式，构建后验分布；

（2）利用核技巧将后验分布进行化简，常化为 Beta 函数；

（3）求解化简后函数（常为 Beta 函数和 Gamma 函数）的期望，即为参数估计。

3.题解：

例：设 $\xi_1,...,\xi_n$ 为总体 ξ 的样本，在平方差损失下，求下列情形参数总体 θ 的贝叶斯估计量

（1） $P_\theta\{\xi=x\}=C_{x-1}^{r-1}\theta^r(1-\theta)^{x-r},x=r,r+1,...,\theta\sim\beta(\alpha,\beta),$

（2） $P_\theta\{\xi=x\}=\theta(1-\theta)^x,x=0,1,2,...,\theta\sim\beta(\alpha,\beta),$

其中 α,β 为已知。

（1）解：

1）利用 Bayes 公式，构建后验分布：

$$h(\theta | x_1,\dots,x_n) = \frac{\pi(\theta) f(x_1,\dots,x_n | \theta)}{\int_{-\infty}^{+\infty} \pi(\theta) f(x_1,\dots,x_n | \theta) d\theta} \propto \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)} \prod_{i=1}^{n} C_{x_i-1}^{r-1} \theta^r (1-\theta)^{x_i-r}, 0<\theta<1$$

2）利用核技巧将后验分布化简为 Beta 函数

$$h(\theta | x_1,\dots,x_n) \propto \theta^{\alpha+nr-1}(1-\theta)^{\beta+n\overline{x}-nr-1}, 0<\theta<1$$

所以，$\theta \sim B(\alpha+rn, n\overline{x}-nr+\beta)$

3）求 Beta 函数的期望

解得：$\tilde{\theta}(\xi_1,\dots,\xi_n) = E[B(\alpha+rn, n\overline{x}-nr+\beta)] = \dfrac{\alpha+rn}{n\overline{\xi}-\alpha+\beta}$

（2）解：

1）利用 Bayes 公式，构建后验分布：

$$h(\theta | x_1,\dots,x_n) = \frac{\pi(\theta) f(x_1,\dots,x_n | \theta)}{\int_{-\infty}^{+\infty} \pi(\theta) f(x_1,\dots,x_n | \theta) d\theta} \propto \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)} \theta^n \prod_{i=1}^{n} (1-\theta)^{x_i}, 0<\theta<1$$

2）利用核技巧将后验分布化简为 Beta 函数

$$h(\theta | x_1,\dots,x_n) \propto \theta^{\alpha+n-1}(1-\theta)^{\beta+n\overline{x}-1}, 0<\theta<1$$

所以，$\theta \sim B(\alpha+n, \beta+n\overline{x})$

3）求 Beta 函数的期望

解得：$\tilde{\theta}(\xi_1,\dots,\xi_n) = E[B(\alpha+n, \beta+n\overline{x})] = \dfrac{\alpha+n}{\alpha+\beta+n+n\overline{x}}$

四、离散型 Bayes 参数估计问题

1.题型描述： 求解某一参数的 Bayes 估计，先验分布为离散型随机变量。

2.解题策略：

（1）利用 Bayes 公式，求出后验分布律；

（2）对后验分布率求期望

3.题解：

例：设总体 $\xi \sim P(\lambda), \xi_1,\dots,\xi_n$ 为 ξ 的样本。如果 λ 有概率函数

$\pi(y) = (p_0)^{\delta(y-1)}(1-p_0)^{\delta(y-2)}, p_0$ 已知。求 λ 的贝叶斯估计量 $\tilde{\lambda}$。

解：由题可知，λ 的先验分布为离散型分布，其分布律为：

λ	1	2
P	p_0	$1-p_0$

1）利用 Bayes 公式，求出后验分布律：

$$h(\lambda=1|\xi_1,...,\xi_n)=\frac{\pi(\lambda=1)f(\xi_1,...,\xi_n|\lambda=1)}{\pi(\lambda=1)f(\xi_1,...,\xi_n|\lambda=1)+\pi(\lambda=2)f(\xi_1,...,\xi_n|\lambda=2)}=\frac{p_0}{p_0+(1-p_0)2^{n\bar{k}}e^{-n}}$$

$$h(\lambda=2|\xi_1,...,\xi_n)=\frac{\pi(\lambda=2)f(\xi_1,...,\xi_n|\lambda=2)}{\pi(\lambda=1)f(\xi_1,...,\xi_n|\lambda=1)+\pi(\lambda=2)f(\xi_1,...,\xi_n|\lambda=2)}=\frac{(1-p_0)2^{n\bar{k}}e^{-n}}{p_0+(1-p_0)2^{n\bar{k}}e^{-n}}$$

所以，后验分布律为：

λ	1	2
P	$\frac{p_0}{p_0+(1-p_0)2^{n\bar{k}}e^{-n}}$	$\frac{(1-p_0)2^{n\bar{k}}e^{-n}}{p_0+(1-p_0)2^{n\bar{k}}e^{-n}}$

2）对后验分布求期望得

$$E[h(\lambda|\xi_1,...,\xi_n)]=\frac{p_0}{p_0+(1-p_0)2^{n\bar{k}}e^{-n}}+\frac{2(1-p_0)2^{n\bar{k}}e^{-n}}{p_0+(1-p_0)2^{n\bar{k}}e^{-n}}=\frac{p_0+2^{n\bar{k}+1}(1-p_0)e^{-n}}{p_0+(1-p_0)2^{n\bar{k}}e^{-n}}$$

即：$\tilde{\lambda}=\frac{p_0+2^{n\bar{k}+1}(1-p_0)e^{-n}}{p_0+(1-p_0)2^{n\bar{k}}e^{-n}}$