第7章
统计决策概要

§7.1　风险函数

7.1.1 风险函数与一致最优决策函数

我们知道状态集$\Theta=\{\theta\}$、行动集$A=\{a\}$和损失函数$L(\theta,a)$是描述决策问题的三个基本要素。在第 5 章我们用状态的先验分布$\pi(\theta)$和后验分布$\pi(\theta\,|\,\mathbf{x})$定义了两种期望损失，即行动$a$的先验期望损失

$$\bar{L}(a)=E^{\theta}[L(\theta,a)]=\int_{\Theta}L(\theta,a)\pi(\theta)d\theta$$

和决策函数$\delta(\mathbf{x})$的后验期望损失（后验风险）

$$R(\delta\,|\,\mathbf{x})=E^{\theta|\mathbf{x}}[L(\theta,\delta(\mathbf{x}))]=\int_{\Theta}L(\theta,\delta(\mathbf{x}))\pi(\theta\,|\,\mathbf{x})d\theta$$

在样本$\mathbf{x}$给定的情况下，先验期望损失$\bar{L}(a)$和后验风险$R(\delta\,|\,\mathbf{x})$都是一个数量，即一个行动对应一个先验期望损失$\bar{L}(a)$或一个决策函数$\delta(\mathbf{x})$对应一个后验风险$R(\delta\,|\,\mathbf{x})$。然后，根据这个数量的大小来评定和比较一个行动或一个决策函数的优良性。

现在我们假设样本$\mathbf{x}=(x_1,\cdots,x_n)$还没有获得，那么后验分布$\pi(\theta\,|\,\mathbf{x})$就还无法确定而损失函数$L(\theta,\delta(\mathbf{x}))$就是随机的。为了消除这种不确定性，假定样本分布$p(\mathbf{x}|\theta)$已知，例如当总体分布为$p(x|\theta)$而且样本为简单随机样本时样本分布

$$p(\mathbf{x}|\theta)=\prod_{i=1}^{n}p(x_i|\theta)$$

我们将损失函数$L(\theta,\delta(\mathbf{x}))$对样本分布$p(\mathbf{x}|\theta)$求期望，于是形成风险函数概念，现在正式定义如下。

定义 7.1 设$\delta(\mathbf{x})$是某个统计决策问题中的决策函数，样本$\mathbf{x}=(x_1,\cdots,x_n)$， $\mathbf{X}=\{\mathbf{x}\}$为抽样空间，那么损失函数$L(\theta,\delta(\mathbf{x}))$对样本分布$p(\mathbf{x}|\theta)$的期望

$$R(\theta,\delta)=E^{\mathbf{x}|\theta}[L(\theta,\delta(\mathbf{x}))]=\int_{\mathbf{X}}L(\theta,\delta(\mathbf{x}))\,p(\mathbf{x}|\theta)d\mathbf{x}$$

称为**决策函数$\delta(\mathbf{x})$的风险函数**。

注：决策函数$\delta(\mathbf{x})$的风险函数是将它的损失函数对样本分布积分的结果，因此风险函数已经与样本无关了，仅是状态（参数）θ与决策函数δ（看成一个自变量）的函数。说的高大上一点，当θ固定不变时，风险函数是决策函数δ的泛函。

下面给出比较风险函数大小的定义。

定义 7.2 设$\delta_1(\mathbf{x})$和$\delta_2(\mathbf{x})$是某统计决策问题中的两个决策函数，假如它们的风险函数在参数空间Θ上一致地有

$$R(\theta,\delta_1)\le R(\theta,\delta_2),\forall\theta\in\Theta$$

且存在至少一个$\theta\in\Theta$使上式中的严格不等式成立，则称**决策函数$\delta_1(\mathbf{x})$一致优于$\delta_2(\mathbf{x})$**；假如对任意$\theta\in\Theta$，风险函数间有

$$R(\theta,\delta_1)\equiv R(\theta,\delta_2),\forall\theta\in\Theta$$

则称**决策函数$\delta_1(\mathbf{x})$与$\delta_2(\mathbf{x})$等价**。

定义 7.3 设$D=\{\delta(\mathbf{x})\}$是某统计决策问题中决策函数全体。假如在决策函数类D中存在这样一个决策函数$\delta^*=\delta^*(\mathbf{x})$，使得对任一个决策函数$\delta(\mathbf{x})\in D$都有

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

则称$\delta^*(\mathbf{x})$为D中**一致最小风险决策函数**或**一致最优决策函数**（**一致最优解**），如所讨论的统计决策问题是点估计问题，则$\delta^*(\mathbf{x})$称为θ的**一致最小风险估计**（**一致最优估计**）。这个决策准则可称为**一致最优准则**。

7.1.2 统计决策框架中的经典推断

设 $\mathbf{x}=(x_1,\cdots,x_n)$ 是来自总体 $p(x|\theta)$ 的一个样本，在寻求参数 θ 的点估计问题中，常把行动集 A 就取为参数空间 Θ，估计量 $\hat{\theta}=\hat{\theta}(\mathbf{x})$ 就是从抽样空间 $\mathbf{X}=\{\mathbf{x}\}$ 到 A 上的一个决策函数，损失函数 $L(\theta,\hat{\theta})$ 就是用 $\hat{\theta}$ 去估计真值 θ 时所引起的损失，这样一来，点估计问题就是一个特殊的统计决策问题。

假设选用最常见的平方损失函数 $L(\theta,\hat{\theta})=(\hat{\theta}-\theta)^2$，那么风险函数就是估计量 $\hat{\theta}$ 的均方误差

$$R(\theta,\hat{\theta})=E^{\mathbf{x}|\theta}[\hat{\theta}(\mathbf{x})-\theta]^2=MSE[\hat{\theta}(\mathbf{x})]$$

这时最小均方误差估计就是 θ 的一致最优估计。遗憾的是，如果不对决策函数类 D 做任何限制，这样的估计在 D 中可能不存在！事实上，若这样的估计存在并记为 $\theta^*=\theta^*(\mathbf{x})$，我们可对 Θ 中任一点 θ_0 构造一个决策函数

$$\delta_0(\mathbf{x})\equiv\theta_0,\forall\mathbf{x}\in\mathbf{X}$$

从而风险函数 $R(\theta,\delta_0)$ 在 $\theta=\theta_0$ 处为零，而 θ^* 是一致最优（风险最小）决策函数，故 θ^* 在 $\theta=\theta_0$ 处的风险值 $R(\theta_0,\theta^*)=0$，由于 θ_0 的任意性，就有

$$R(\theta,\theta^*)=0,\forall\theta\in\Theta$$

这表明 $\theta^*(\mathbf{x})=\theta$ 处处成立，这样的 θ^* 不是统计量，更不是估计量。

现在转入考虑区间估计问题。在寻求参数θ的区间估计问题时，可把行动集取为某个给定的区间集合，例如直线上所有的有界区间组成的集合A。这时A中每个区间就是一个行动，决策函数就是定义在抽样空间$\mathbf{X}=\{\mathbf{x}\}$上而在$A$中取值的函数

$$\delta(\mathbf{x})=[\mathrm{d}_1,\mathrm{d}_2]=[\mathrm{d}_1(\mathbf{x}),\mathrm{d}_2(\mathbf{x})]$$

取如下损失函数

$$L(\theta,\delta(\mathbf{x}))=m_1(\mathrm{d}_2-\mathrm{d}_1)+m_2[1-I_{\delta(\mathbf{x})}(\theta)]$$

其中m_1和m_2为二个给定正常数，$I_B(\bullet)$表示集合B的示性函数，第一项表示区间$\delta(\mathbf{x})$长短引起的损失，长度越长则损失越大；第二项表示当θ不属于区间$\delta(\mathbf{x})$时引起的损失。这时风险函数为

$$R(\theta,\delta(\mathbf{x}))=m_1E^{\mathbf{x}|\theta}(\mathrm{d}_2-\mathrm{d}_1)+m_2P^{\mathbf{x}|\theta}(\theta\notin\delta(\mathbf{x}))$$

但是此时要寻求θ的一致最优区间估计不是容易的事。

现在讨论假设检验问题。设Θ_0与Θ_1为参数空间$\Theta=\{\theta\}$中两个不相交的非空子集，原假设H_0：$\theta\in\Theta_0$；备择假设H_1：$\theta\in\Theta_1$。在用统计决策语言描述假设检验问题时，常把行动集A取为仅有二个行动（接受与拒绝）组成的集合，即$A=\{0(接受H_0), 1(拒绝H_0)\}$，这时决策函数$\delta(\mathbf{x})$就是抽样空间$\mathbf{X}=\{\mathbf{x}\}$到$A$上的一个函数，所有这种决策函数的全体记为$D$。对给定的决策函数$\delta(\mathbf{x})\in D$，记

$$W=\{\mathbf{x};\delta(\mathbf{x})=1\}\subset\mathbf{X}$$

则决策函数$\delta(\mathbf{x})$可表示为拒绝域 W 上的示性函数，即

$$\delta(\mathbf{x})=I_W(\mathbf{x})=\begin{cases}1, & \mathbf{x}\in W\\ 0, & \mathbf{x}\notin W\end{cases}$$

反之，对于抽样空间$\mathbf{X}$的任一子集B，其示性函数$I_B(\mathbf{x})$就是该检验问题的一个决策函数。

最后来确定损失函数$L(\theta,\delta)$，它可看作θ为真时，采取行动$\delta=\delta(\mathbf{x})$所引起的损失，这种损失函数常采取 0-1 损失函数

$$L(\theta,0)=\begin{cases}0, & \theta\in\Theta_0\\ 1, & \theta\in\Theta_1\end{cases},\quad L(\theta,1)=\begin{cases}1, & \theta\in\Theta_0\\ 0, & \theta\in\Theta_1\end{cases}$$

到此，一个假设检验问题就可看作一个统计决策问题了，这时任一决策函数 $\delta(\mathbf{x})$ 的风险函数为

$$
\begin{aligned}
R(\theta,\delta) &= E^{\mathbf{x}|\theta}L(\theta,\delta(\mathbf{x})) \\
&= \int_W L(\theta,1)\,p(\mathbf{x}|\theta)d\mathbf{x} + \int_{\bar{W}} L(\theta,0)\,p(\mathbf{x}|\theta)d\mathbf{x} \\
&= \begin{cases} P^{\mathbf{x}|\theta}(\mathbf{x}\in W), & \theta\in\Theta_0 \\ P^{\mathbf{x}|\theta}(\mathbf{x}\notin W), & \theta\in\Theta_1 \end{cases}
\end{aligned}
$$

这就是说，当 $\theta\in\Theta_0$ 时 $\delta(\mathbf{x})$ 的风险函数值相当于犯第 I 类错误（拒真错误）的概率；当 $\theta\in\Theta_1$ 时其风险函数值相当于犯第 II 类错误（纳伪错误）的概率。

回顾经典概率论，我们知道奈曼-皮尔逊（Neyman-Pearson）假设检验理论的基本思想是在限制犯第 I 类错误的概率不超过某一个给定的正数 α 的条件下，寻找犯第 II 类错误概率尽可能小的拒绝域。在统计决策理论中这等价于寻找这样的决策函数 $\delta^*(\mathbf{x})$，其满足

$$\sup_{\theta\in\Theta_0} R(\theta,\delta^*(\mathbf{x})) \le a$$

而且对抽样空间 $\mathbf{X}$ 中任一子集的示性函数即决策函数 $\delta=\delta(\mathbf{x})$ 有

$$\forall\theta\in\Theta_1, R(\theta,\delta^*) \le R(\theta,\delta)$$

所以假设检验问题仍是特定损失函数下的统计决策问题。

注：至此我们已经看到，经典统计推断的三种基本形式都可纳入统计决策理论框架内，看作是特定行动集和特定损失函数下的统计决策问题。但是，从应用的角度看，这里的做法对于实际进行统计推断并没有什么帮助，其意义仅仅是理论上的。

§ 7.2　决策函数的容许性与最小最大准则

7.2.1 容许性

从上一节我们看到，对给定的统计决策问题，按照风险函数一致最小原则，在某个决策函数类 D 中寻求一致最优决策函数常常难以实现，这一般是由于对于 D 中的决策函数没有一定的要求引起的，所以，我们先对决策函数加上一个被称为容许性的要求，其一般定义如下。

定义 7.4 对给定的统计决策问题和决策函数类 D，对于决策函数 $\delta_1 = \delta_1(\mathbf{x})$，假如在 D 中存在另一个决策函数 $\delta_2 = \delta_2(\mathbf{x})$ 一致优于 δ_1，则称 δ_1 为**非容许决策函数**（**解**）。假如在 D 中不存在一致优于 δ_1 的决策函数 δ_2，则称 $\delta_1(\mathbf{x})$ 为**容许决策函数**（**解**）。在统计决策问题为参数估计问题时，相应的估计量分别称为**非容许估计**和**容许估计**。

注：从此定义可见，非容许决策函数肯定不是一致最优的。然而遗憾的是，即使决策函数类 D 中所有决策函数都是容许的，在 D 中也可能不存在一致最优决策函数，请看下面的例子。

例 7.1 某公司的产品每 100 件装成一箱运交客户。在向客户交货前，面临如下两个行动的选择

a_1：一箱中逐一检查产品，a_2：一箱中一件产品也不检查

如果公司选择行动a_1，则可保证交货时每件产品都是合格品，但因每件产品的检查费为 0.8 元，为此公司要支付检查费 80 元/箱。如果公司选择行动a_2，则无检查费要支付，但客户发现不合格品时，按合同不仅允许更换，而且每件要支付 12.5 元的赔偿费。在做决策之前，公司决定从仓库随机取出一箱并抽取两件产品进行检查，设θ表示一箱中的产品不合格率，X 为不合格产品数，则 $X \sim Bin(2,\theta)$（二项分布），将分布列写出就是

$$P_\theta(X = x) = p_\theta(x) = \binom{2}{x}\theta^x(1-\theta)^{2-x}, x = 0,1,2$$

这时公司的支付函数为

$$W(\theta,a) = \begin{cases} 80, & a = a_1 \\ 1.6 + 1250\theta, & a = a_2 \end{cases}$$

相应的损失函数为

$$L(\theta,a_1) = \begin{cases} 78.4 - 1250\theta, & \theta \le \theta_0 \\ 0, & \theta > \theta_0 \end{cases}, \quad L(\theta,a_2) = \begin{cases} 0, & \theta \le \theta_0 \\ -78.4 + 1250\theta, & \theta > \theta_0 \end{cases}$$

其中平衡值$\theta_0 = 0.06272$。

本例实际上就是例 5.9 中我们讨论过的二行动决策问题，在那里我们已得抽样空间 $\mathbf{X}=\{0,1,2\}$，行动集 $A=\{a_1,a_2\}$ 而且决策函数类 D 共有 8 个决策函数。我们还利用 θ 的先验分布 $U(0,0.12)$ 算出这 8 个决策函数的后验期望损失（即后验风险），得到 $\delta_5(x)$ 是后验风险准则下的最优决策函数。现在来计算这 8 个决策函数的风险函数，以 $\delta_5(x)$ 的风险函数为例

$$
\begin{aligned}
R(\theta,\delta_5) &= E^{x|\theta}L(\theta,\delta_5(x)) \\
&= L(\theta,\delta_5(0))\times p_\theta(0)+L(\theta,\delta_5(1))\times p_\theta(1)+L(\theta,\delta_5(2))\times p_\theta(2) \\
&= L(\theta,a_2)\times(1-\theta)^2+L(\theta,a_1)\times 2\theta(1-\theta)+L(\theta,a_1)\times\theta^2
\end{aligned}
$$

把损失函数代入，不难得到

$$
R(\theta,\delta_5)=\begin{cases}(78.4-1250\theta)[1-(1-\theta)^2], & \theta\le\theta_0 \\ (-78.4+1250\theta)(1-\theta)^2, & \theta>\theta_0\end{cases}
$$

类似地可算出其它七个决策函数的风险函数，例如

$$
R(\theta,\delta_2)=\begin{cases}(78.4-1250\theta)(1-\theta^2), & \theta\le\theta_0 \\ (-78.4+1250\theta)\theta^2, & \theta>\theta_0\end{cases}
\quad
R(\theta,\delta_8)=\begin{cases}0, & \theta\le\theta_0 \\ (-78.4+1250\theta), & \theta>\theta_0\end{cases}
$$

最后，将风险函数在不合格品率θ于 0 和 0.12 之间几个点位的值算出并编制成表 7.1。虽然只算出几个点位的值，从此表已经可以看出，这 8 个决策函数组成的类中没有一个是非容许决策函数（即全是容许决策函数），然而，这个决策函数类中并不存在一致最优决策函数，面对这 8 个容许决策函数，决策者仍然两眼茫茫无所适从，这从一个侧面表明一致最优准则过于严苛，不是一个适宜的准则。为了计算表 7.1 中的风险函数值，我们可以编写一个小程序来进行，以$\delta_5(x)$的风险函数为例，其风险函数值计算程序如下

```
f<-function(x){if(x<=0.06272){(78.4-1250*x)*(2*x-x^2)}
                    else {(-78.4+1250*x)*(1-x)^2}}
x<- c(0, 0.02, 0.04, 0.06, 0.08, 0.10, 0.12)
y<- c( )
for(t in 1:7){y[t]=f(x[t])}
y
[1]   0.00000   2.11464   2.22656   0.39576   18.28224   37.74600   55.44704
```

7.2.2 最小最大准则

从上小节我们已经看到即使一个决策函数类全由容许决策函数构成，一致最优决策函数也不见得存在，因为一致最优是要求最优决策函数 δ^* 的风险函数在任何状态下比任意别的决策函数的风险函数优良，这是非常苛刻的要求。如果只要求对决策函数的风险函数 $R(\theta,\delta)$ 的某一特征进行比较，则往往是可行的。例如，考虑风险函数在状态（参数）空间 Θ 上的最大值 $\max_{\theta\in\Theta} R(\theta,\delta)$，它表示决策者选用决策函数 $\delta=\delta(\mathbf{x})$ 后可能引发的最大风险。许多决策对象例如高铁、江河大坝、桥梁、海底隧道的安全要求很高，其决策者一般都是稳健或偏保守的，只能从最坏处做打算同时又希望结果最好，这就产生一种思想：让最大风险最小化。于是，最大风险最小化准则由此产生，由于这个准则的英文名是 Minimax，所以最大风险最小化准则又称为最小最大准则，其正式定义如下。

定义 7.5　在统计决策问题中，$\mathbf{x}=(x_1,\cdots,x_n)$ 是来自总体 $p(x|\theta)$ 的一个样本，$\theta\in\Theta$ (参数空间)，$D=\{\delta(\mathbf{x})\}$ 是决策函数类。那么风险值

$$\tilde{R}=\min_{\delta\in D}\max_{\theta\in\Theta} R(\theta,\delta)=\min_{\delta\in D}\max_{\theta\in\Theta} E^{\mathbf{x}|\theta}L(\theta,\delta(\mathbf{x}))$$

称为**损失函数** $L(\theta,\delta)$ **下的最小最大风险**。如果存在决策函数 $\delta^*\in D$，使得

$$\max_{\theta\in\Theta} R(\theta,\delta^*)=\tilde{R}$$

则称 δ^* 为该统计决策问题**在最小最大准则下的最优决策函数**，或称**最小最大决策函数**（**最小最大解**）。当该统计决策问题为参数估计或检验问题时，δ^* 还称为**最小最大估计**或**最小最大检验**。

例 7.2　在例 7.1 中有 8 个决策函数组成的类 D 而且参数空间 Θ 为[0, 0.12]即不合格品率 θ 在 0 到 0.12 之间。虽然表 7.1 不是全部风险函数值而只是几个点位上的风险值，但还是容易看出 D 中每个决策函数 $\delta=\delta(\mathbf{x})$ 所具有的最大风险（见表 7.2），从中立即可得 $\delta_5(x)$ 是最小最大风险准则下的最优决策函数（最小最大函数），而且采用 $\delta_5(x)$ 所引起的最大平均损失是 55.45 元/箱，这就是它的最小最大风险。

表 7.2　**8 个风险函数的最大值**

δ	$\max R(\theta,\delta)$	δ	$\max R(\theta,\delta)$
δ_1	78.4	δ_5	55.45
δ_2	78.4	δ_6	56.48
δ_3	78.4	δ_7	70.57
δ_4	78.4	δ_8	71.60

例 7.3 设 x 是从正态总体 $N(\theta,1)$ 抽取的容量为 1 的样本，参数 $\theta \in R$（实数集），决策函数类和损失函数分别为

$$D=\{\delta_c;\delta_c(x)=cx,c\in R\}，\quad L(\theta,\ \delta_c)=(cx-\theta)^2$$

（1）求参数 θ 的一致最小风险估计；（2）求参数 θ 的最小最大估计。

解：依条件这里样本分布就是总体分布 $N(\theta,1)$，若取平方损失作为损失函数，则 D 中任一个估计 δ_c 的风险函数为

$$R(\theta,\ \delta_c)=E^{x|\theta}(cx-\theta)^2=E^{x|\theta}[c(x-\theta)+(c-1)\theta)]^2=c^2+(c-1)^2\theta^2$$

（1）容易看出，当 $c=1$ 时，作为 θ 的函数，$R(\theta,\ \delta_1)\equiv 1$ 为一条直线，当 $|c|<1$ 时，风险函数 $R(\theta,\ \delta_c)$ 作为 θ 的函数是开口朝上的抛物线而且最小值为 $c^2<1$，因此都上穿直线 $R(\theta,\ \delta_1)\equiv 1$，换言之，这些风险函数没有一个是始终最小的，所以在决策函数类 D 中没有一致最小风险估计。

（2）决策函数 δ_c 的最大风险

$$M(\delta_c)=\max_{\theta\in R}R(\theta,\ \delta_c)=\max_{\theta\in R}[c^2+(c-1)^2\theta^2]=\begin{cases}1, & c=1\\ \infty, & c\neq 1\end{cases}$$

可见，按最小最大准则，$\delta_1(x)=x$ 是 θ 在 D 中的最小最大估计。

注：容易看出，当 $c=1$ 时，$R(\theta,\ \delta_1)\equiv 1$。对任意 $c>1$ 时，有

$$\forall\theta\in R,\ R(\theta,\ \delta_1)<R(\theta,\ \delta_c)$$

所以当 $c>1$ 时，δ_1 一致优于 δ_c。因此，当决策函数类取为 $D=\{\delta_c;\delta_c(x)=cx,c\in[1,\infty)\}$ 时，存在一致最小风险估计 $\delta_1(x)=x$。当然，这里把决策函数类 D 缩小了。

下面我们来讨论最小最大解与容许解的关系，有以下的两个定理。

定理 7.1　在给定的统计决策问题中，决策函数类为 D，如果 $\delta_0(\mathbf{x})$ 是唯一的最小最大解，则 $\delta_0(\mathbf{x})$ 是容许的。

证明：用反证法。若 $\delta_0=\delta_0(\mathbf{x})$ 是非容许的，则存在另一个决策函数 $\delta_1\neq\delta_0$，使得

$$\forall\theta\in\Theta, R(\theta,\ \delta_1)\le R(\theta,\ \delta_0)$$

且在 Θ 中至少存在一个 θ 使上述严格不等式成立，又 δ_0 是最小最大解，因此有

$$\max_{\theta\in\Theta}R(\theta,\delta_1)\le\max_{\theta\in\Theta}R(\theta,\delta_0)=\min_{\delta\in D}\max_{\theta\in\Theta}R(\theta,\delta)$$

从而 $\delta_1(x)$ 也是 θ 的最小最大解，这与 δ_0 是唯一的最小最大解矛盾。

定理 7.2　在一个统计决策问题中，决策函数类为 D，假如 $\delta_0(\mathbf{x})$ 是容许解，且在参数空间 Θ 上是常数风险，则 $\delta_0(\mathbf{x})$ 也是最小最大解。

证明：用反证法。由于 $\delta_0=\delta_0(\mathbf{x})$ 的风险函数是常数，故有

$$\max_{\theta\in\Theta}R(\theta,\delta_0)=R(\theta,\delta_0)\equiv c$$

若 δ_0 不是最小最大解，则存在另一个决策函数 $\delta_1\neq\delta_0$，其在 Θ 上的最大风险不应超过 δ_0 的常数风险，即

$$\max_{\theta\in\Theta}R(\theta,\delta_1)\le R(\theta,\delta_0)\equiv c$$

从而有

$$\forall\theta\in\Theta,\ R(\theta,\delta_1)\le R(\theta,\delta_0)\equiv c$$

这与 δ_0 是容许解矛盾。

§7.3 贝叶斯风险准则与贝叶斯解

7.3.1 贝叶斯风险准则

在本章前两节中，我们讨论的一致最优准则和最小最大准则都没有利用先验信息，因此属于经典统计决策的范畴，但是，我们在前面的各章中已经看到先验信息的重要作用，所以现在我们把先验信息加入到统计决策问题中来，从而把经典统计与贝叶斯统计联系起来，我们将看到一些重要而优美的结果。

定义 7.6 对给定的统计决策问题和给定的决策函数类 D，设决策函数 $\delta=\delta(\mathbf{x})\in D$ 的风险函数为 $R(\theta,\delta)$， θ 的先验分布为 $\pi(\theta)$，则风险函数对先验分布的期望

$$R(\delta)=E^{\pi}R(\theta,\delta)=\int_{\Theta}R(\theta,\delta)\pi(\theta)d\theta$$

称为 $\delta(\mathbf{x})$ 的**贝叶斯风险**，如果在决策函数类 D 中存在决策函数 $\delta^*(\mathbf{x})$ 满足

$$R(\delta^*)=\min_{\delta\in D}R(\delta)$$

则称 $\delta^*(\mathbf{x})$ 为**贝叶斯风险准则下的最优决策函数**，也称为**贝叶斯决策函数**或**贝叶斯解**，当决策问题是估计问题时则称为**贝叶斯估计**。

例 7.4 在例 7.1 中对 8 个决策函数分别计算出了相应的风险函数，若取均匀分布 $U[0,0.12]$ 作为 θ 的先验分布，计算它们的贝叶斯风险。

解：以计算 δ_1 与 δ_2 的贝叶斯风险作为例子，其余 6 个贝叶斯风险的计算结果列于表 7.3 中。

$$R(\delta_1)=\frac{1}{0.12}\int_0^{\theta_0}(78.4-1250\theta)d\theta=20.4885$$

$$R(\delta_2)=\frac{1}{0.12}\int_0^{\theta_0}(78.4-1250\theta)(1-\theta^2)d\theta+\frac{1}{0.12}\int_{\theta_0}^{0.12}(-78.4+1250\theta)\theta^2 d\theta$$
$$=20.6522$$

为了减轻计算负担，可以编个 R 程序来计算这些贝叶斯风险，以 δ_2 的贝叶斯风险为例，其计算程序为

```
f1<-function(x){(78.4-1250*x)*(1-x^2)}
f2<-function(x){(-78.4+1250*x)*x^2}
integrate(f1,0.0,0.06272)$value/0.12+integrate(f2,0.06272,0.12)$value/0.12
[1] 20.65221
```

表 7.3　例 7.1 中 8 个决策函数的贝叶斯风险

δ	$R(\delta)$	δ	$R(\delta)$
δ_1	20.4885	δ_5	14.6601
δ_2	20.6522	δ_6	14.8370
δ_3	22.7401	δ_7	16.9249
δ_4	22.9168	δ_8	17.0885

从表 7.3 容易得到，按照贝叶斯风险准则，该决策问题的最优决策函数（贝叶斯决策函数）是$\delta_5(x)$。比较这个结果与例 5.9 与例 5.9 续中用后验风险准则获得的结果，我们发现在两种不同的决策准则下得到的最优决策函数是一模一样的！我们不禁想问：这一现象是偶然的吗？下面就来讨论这个有趣而且重要的问题。

为了讨论这个问题，首先建立贝叶斯风险 $R(\delta)$ 与后验风险 $R(\delta \mid \mathbf{x})$ 之间的关系。由贝叶斯风险的定义以及风险函数的定义，我们有

$$R(\delta)=\int_{\Theta} R(\theta,\delta)\pi(\theta)d\theta=\int_{\Theta}\left[\int_{\mathbf{X}} L(\theta,\delta)\,p(\mathbf{x}|\theta)d\mathbf{x}\right]\pi(\theta)d\theta$$

交换积分次序并利用贝叶斯公式

$$p(\mathbf{x}|\theta)\pi(\theta)=\pi(\theta|\mathbf{x})m(\mathbf{x})$$

其中 $\pi(\theta|\mathbf{x})$ 为 θ 的后验密度函数， $m(\mathbf{x})$ 为样本 $\mathbf{x}$ 的边际密度函数，即得

$$R(\delta)=\int_{\mathbf{X}}[\int_{\Theta} L(\theta,\delta)\pi(\theta|\mathbf{x})d\theta]m(\mathbf{x})d\mathbf{x}=\int_{\mathbf{X}} R(\delta|\mathbf{x})m(\mathbf{x})d\mathbf{x}$$

这就是说，贝叶斯风险是后验风险对边际分布 $m(\mathbf{x})$ 的数学期望。为保证上述积分次序可交换，需要一定条件，可以证明这个条件就是贝叶斯风险在整个决策函数类 D 上的最小值是有限的，即

$$\min_{\delta\in D} R(\delta)<\infty$$

这个条件在实际应用中是容易满足的，因为当条件不满足时，那就意味着所有的决策函数 $\delta\in D$ 的贝叶斯风险为无穷大，这种决策函数类 D 无论在理论上或实际上都是意义不大的，是没有必要考虑的。

定理 7.3 对给定的统计决策问题和决策函数类 D，若先验分布 $\pi(\theta)$ 使贝叶斯风险满足条件 $\min\limits_{\delta\in D} R(\delta)<\infty$，则贝叶斯风险准则与后验风险准则等价，即使后验风险最小的决策函数同时也使贝叶斯风险最小，反之亦然。

证明：设 δ^* 为贝叶斯风险准则下的最优决策函数，δ^{**} 为后验风险准则下的最优决策函数。由 δ^* 定义和定理条件可知

$$R(\delta^*)=\min_{\delta\in D} R(\delta)=\min_{\delta\in D}\int_{\mathbf{X}} R(\delta|\mathbf{x})m(\mathbf{x})d\mathbf{x}\ge\int_{\mathbf{X}}\min_{\delta\in D} R(\delta|\mathbf{x})m(\mathbf{x})d\mathbf{x}$$

再由 δ^{**} 的定义知

$$R(\delta^{**}|\mathbf{x})=\min_{\delta\in D} R(\delta|\mathbf{x})$$

综上可得

$$R(\delta^*)\ge\int_{\mathbf{X}} R(\delta^{**}|\mathbf{x})m(\mathbf{x})d\mathbf{x}=R(\delta^{**})\ge\min_{\delta\in D} R(\delta)=R(\delta^*)$$

即

$$R(\delta^{**})=\min_{\delta\in D} R(\delta)=R(\delta^*)$$

这就是说，使后验风险最小的决策函数 δ^{**} 同时也使贝叶斯风险最小。

另外，由上述推理还得 $R(\delta^*)=R(\delta^{**})$，于是

$$\int_{\mathbf{X}}[R(\delta^*|\mathbf{x})-R(\delta^{**}|\mathbf{x})]m(\mathbf{x})d\mathbf{x}=\int_{\mathbf{X}} R(\delta^*|\mathbf{x})m(\mathbf{x})d\mathbf{x}-\int_{\mathbf{X}} R(\delta^{**}|\mathbf{x})m(\mathbf{x})d\mathbf{x}$$

$$=R(\delta^*)-R(\delta^{**})=0$$

由于δ^{**}是使后验风险最小的决策函数，所以上式被积函数

$$R(\delta^*|\mathbf{x})-R(\delta^{**}|\mathbf{x})\geq 0$$

于是

$$R(\delta^*|\mathbf{x})=R(\delta^{**}|\mathbf{x})=\min_{\delta\in D}R(\delta|\mathbf{x})$$

这就是说，使贝叶斯风险最小的决策函数δ^*同时也使后验风险最小。

注: 1. 贝叶斯风险（准则）把经典统计决策和贝叶斯统计决策有机联系在一起，而贝叶斯风险准则与后验风险准则的等价性使我们没有必要区分经典统计决策和贝叶斯统计决策。该等价性也是使这两种准则下的最优决策函数可以统称为贝叶斯决策函数（贝叶斯解、贝叶斯估计）的原因。

2. 在离散情形，定理 7.3 仍然是正确的，请读者自行证明。

7.3.2 贝叶斯解的性质

本节讨论贝叶斯解的性质，即贝叶斯解何时会是容许解或最小最大解。首先讨论贝叶斯解的容许性，我们将看到在一定的条件下许多贝叶斯解是容许的。

定理 7.4 在给定的统计决策问题中，决策函数类为 D，设 $\delta_0=\delta_0(\mathbf{x})$ 是一个贝叶斯解，如果先验分布 $\pi(\theta)$ 在 Θ 上任开子集的概率为正；δ_0 的风险函数 $R(\theta,\delta_0)$ 是 θ 的连续函数；δ_0 的贝叶斯风险 $\mathrm{R}(\delta_0)$ 是有限的，则 δ_0 是容许的。

证明：用反证法。若 δ_0 是非容许的，则存在另一个决策函数 δ_1=$\delta_1(\mathbf{x})$，满足

$$\forall\theta\in\Theta，R(\theta,\delta_1)\le R(\theta,\delta_0)$$

且至少对某个 $\theta_1\in\Theta$，使得上述严格不等式成立，即

$$R(\theta_1,\delta_1)<R(\theta_1,\delta_0)$$

由 $R(\theta,\delta_0)$ 是 θ 的连续函数知，存在一个正数 ε 以及 θ_1 的邻域 S_ε，使得

$$\forall\theta\in S_\varepsilon，R(\theta,\delta_1)<R(\theta,\delta_0)-\varepsilon$$

于是 δ_1 的贝叶斯风险

$$\begin{aligned}R(\delta_1)&=\int_{S_\varepsilon}R(\theta,\delta_1)\pi(\theta)d\theta+\int_{\bar{S}_\varepsilon}R(\theta,\delta_1)\pi(\theta)d\theta\\&\le\int_{S_\varepsilon}[R(\theta,\delta_0)\text{-}\varepsilon]\pi(\theta)d\theta+\int_{\bar{S}_\varepsilon}R(\theta,\delta_0)\pi(\theta)d\theta\\&=R(\delta_0)-\varepsilon P_\pi(\theta\in S_\varepsilon)\end{aligned}$$

由假设知 $P_\pi(\theta\in S_\varepsilon)>0$，故 $R(\delta_1)<R(\delta_0)$。这与 δ_0 是贝叶斯解矛盾。

定理 7.5　在给定的贝叶斯决策问题中，决策函数类为 D，如果在先验分布 $\pi(\theta)$ 下的贝叶斯解 $\delta_0=\delta_0(\mathbf{x})$ 是唯一的，则它是容许的。

证明： 用反证法。若 δ_0 是非容许的，则存在另一个决策函数 $\delta_1=\delta_1(\mathbf{x})$，满足

$$\forall \theta \in \Theta，\ R(\theta,\delta_1) \le R(\theta,\delta_0)$$

且严格不等式至少对某个 $\theta \in \Theta$ 成立。将上不等式两边对先验分布求期望，有

$$R(\delta_1) = E^{\theta} R(\theta,\delta_1) \le E^{\theta} R(\theta,\delta_0) = R(\delta_0)$$

这样，$\delta_1(\mathbf{x})$ 也是在先验分布 $\pi(\theta)$ 下的贝叶斯解，这与 δ_0 的唯一性矛盾。

接下来我们讨论给定先验分布下的贝叶斯解在什么条件下也是最小最大解以及别的有趣推论。

定理 7.6 在给定的贝叶斯决策问题中，决策函数类为 D，若 $\delta_0=\delta_0(\mathbf{x})$ 是先验分布 $\pi(\theta)$ 下的贝叶斯解且它的贝叶斯风险满足

$$R_\pi(\delta_0)=\max_{\theta\in\Theta}R(\theta,\delta_0)$$

那么，（1）δ_0 也是该决策问题的最小最大解。（2）如 δ_0 是唯一的贝叶斯解，则也是唯一的最小最大解。（3）对任意别的先验 $\pi_{any}(\theta)$ 有 $R_\pi(\delta_0)\ge R_{\pi_{any}}(\delta_0)$。

证明：（1）用反证法。若 δ_0 不是该决策问题的最小最大解，则存在决策函数 $\delta_1\in D$，对任 $\theta\in\Theta$ 满足

$$R(\theta,\delta_1)\le\max_{\theta\in\Theta}R(\theta,\delta_1)<\max_{\theta\in\Theta}R(\theta,\delta_0)=R_\pi(\delta_0)$$

于是

$$R_\pi(\delta_1)=E^{\pi(\theta)}R(\theta,\delta_1)\le\max_{\theta\in\Theta}R(\theta,\delta_1)<\max_{\theta\in\Theta}R(\theta,\delta_0)=R_\pi(\delta_0)$$

这就是说 δ_1 在先验分布 $\pi(\theta)$ 下的贝叶斯风险小于 δ_0 在先验分布 $\pi(\theta)$ 下的贝叶斯风险，这与 δ_0 是先验分布 $\pi(\theta)$ 下的贝叶斯解矛盾。

（2）如果存在另一个决策函数 $\delta_m \in D$ 也是最小最大解，则有

$$R_\pi(\delta_m) = E^{\pi(\theta)} R(\theta, \delta_m) \le \max_{\theta\in\Theta} R(\theta, \delta_m) \le \max_{\theta\in\Theta} R(\theta, \delta_0) = R_\pi(\delta_0)$$

由于 δ_0 是先验分布 $\pi(\theta)$ 下的贝叶斯解，故对任何 $\delta \in D$ 有 $R_\pi(\delta_0) \le R_\pi(\delta)$，从而对任何 $\delta \in D$ 有

$$R_\pi(\delta_m) \le R_\pi(\delta_0) \le R_\pi(\delta)$$

这就是说 δ_m 是先验分布 $\pi(\theta)$ 下的贝叶斯解，这与 δ_0 是唯一的贝叶斯解矛盾。

（3）如果结论不成立，则存在先验 $\pi_1(\theta)$ 使 $R_\pi(\delta_0) < R_{\pi_1}(\delta_0)$，于是由定理条件，有

$$\max_{\theta\in\Theta} R(\theta, \delta_0) = R_\pi(\delta_0) < R_{\pi_1}(\delta_0)$$

另一方面

$$R_{\pi_1}(\delta_0) = E^{\pi_1} R(\theta, \delta_0) \le E^{\pi_1} \max_{\theta\in\Theta} R(\theta, \delta_0) = \max_{\theta\in\Theta} R(\theta, \delta_0)$$

这就产生 $\max_{\theta\in\Theta} R(\theta, \delta_0)$ 要严格小于自身的荒唐结论。

注：从定理 7.6 的结论（3）我们看到满足定理条件的先验分布 $\pi(\theta)$ 并不那么招人喜爱，因为在其下的贝叶斯解的贝叶斯风险大于等于这个贝叶斯解在任意先验分布下的贝叶斯风险。因而，我们把这个先验 $\pi(\theta)$ 称为**最不讨喜先验**。

例 7.5　设二项分布总体 $Bin(n,\theta)$ 得到一个样本 x，参数 θ 的先验分布为贝塔分布 $Beta(\alpha,\beta)$，损失函数为 $L(\theta,\delta)=(\delta-\theta)^2$。（1）求参数 θ 的贝叶斯估计（解）;（2）该贝叶斯估计是容许的吗？;（3）该贝叶斯估计的风险函数何时为常数，从而估计也是最小最大估计;（4）将此最小最大估计与经典最大似然估计 $\delta_0(x)=x/n$ 进行比较。

解：（1）我们已经知道贝塔分布是参数 θ 的共轭先验，所以 θ 的后验分布为 $Beta(x+\alpha,n-x+\beta)$。当损失函数为 $L(\theta,\delta)=(\delta-\theta)^2$ 时，参数 θ 的贝叶斯估计为后验均值（参见定理 5. 3 和定理 7.3）

$$\delta_{(\alpha,\beta)}(x)=\frac{x+\alpha}{n+\alpha+\beta}$$

（2）不难验证定理 7.4 的条件全部满足，所以该贝叶斯估计是容许的。

（3）二项分布总体 X 的均值 $EX=n\theta$，方差 $\mathrm{var}X=n\theta(1-\theta)$。由于样本量为 1，所以样本分布等于总体分布，于是（1）中得到的参数 θ 的贝叶斯估计的风险函数

$$
\begin{aligned}
R(\theta,\delta_{(\alpha,\beta)}) &= E\left(\frac{X+\alpha}{n+\alpha+\beta}-\theta\right)^2 \\
&= E\left(\frac{X-EX}{n+\alpha+\beta}+(\frac{EX+\alpha}{n+\alpha+\beta}-\theta)\right)^2 \\
&= \mathrm{var}\left(\frac{X}{n+\alpha+\beta}\right)+\left(\frac{n\theta+\alpha}{n+\alpha+\beta}-\theta\right)^2 \\
&= \frac{n\theta(1-\theta)}{(n+\alpha+\beta)^2}+\left(\frac{\alpha-(\alpha+\beta)\theta}{n+\alpha+\beta}\right)^2 \\
&= \frac{n\theta(1-\theta)+[\alpha-(\alpha+\beta)\theta]^2}{(n+\alpha+\beta)^2}
\end{aligned}
$$

上式要为常数即分子要为常数，从而解得 $\alpha=\beta=0.5\sqrt{n}$，此时该常数风险为

$$
R(\theta,\delta_{(0.5\sqrt{n},0.5\sqrt{n})})=\frac{1}{4(1+\sqrt{n})^2}
$$

而 θ 的贝叶斯估计为

$$
\delta_{(0.5\sqrt{n},0.5\sqrt{n})}(x)=\frac{x+0.5\sqrt{n}}{n+\sqrt{n}}
$$

依推论 7.1，此时该贝叶斯估计也是参数 θ 的最小最大估计。

（4）我们首先求参数θ的最大似然估计$\delta_0(x)=x/n$的风险函数

$$R(\theta,\delta_0)=EL(\theta,\delta_0(X))=E(\theta-\frac{X}{n})^2=\frac{1}{n^2}E(n\theta-X)^2$$

$$=\frac{1}{n^2}E(X-EX)^2=\frac{1}{n^2}\mathrm{var}X=\frac{1}{n}\theta(1-\theta)$$

而由（3），参数θ的最小最大估计的风险函数

$$R(\theta,\delta_{(0.5\sqrt{n},0.5\sqrt{n})})=\frac{1}{4(1+\sqrt{n})^2}$$

现在将两个风险函数在平面上画出如图 7.1 所示，我们看到最大似然估计$\delta_0(x)=x/n$的风险函数有一部分是大于最小最大估计的风险函数（常数）的，所以该最大似然估计不是最小最大估计，但对于多数θ而言，最大似然估计的风险函数是小于最小最大估计的风险函数的，因此，人们可能选用最大似然估计而不是最小最大估计。

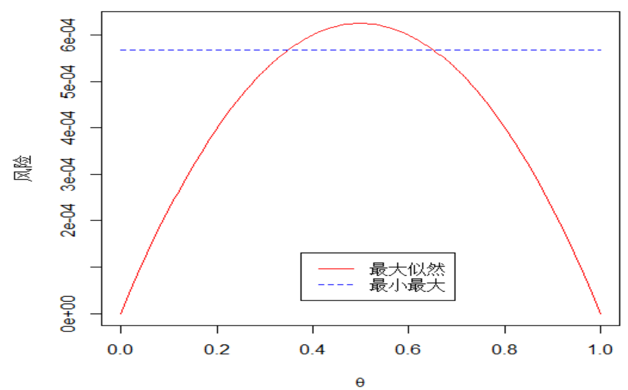

图 7.1 最大似然估计和最小最大估计的风险函数比较图

在 7.2.2 小节，我们提出了最小最大解（Minimax 解）的定义，但是，直接用定义判断一个决策函数是否为 Minimax 解一般而言是很困难的，下面我们给出判断一个决策函数是否为 Minimax 解的定理。

定理 7.7 在给定的贝叶斯决策问题中，决策函数类为 D，设 $\{\pi_k;k\geq 1\}$ 为 Θ 上的先验分布列，$\{\delta_k;k\geq 1\}$ 和 $\{R_k(\delta_k);k\geq 1\}$ 为相应的贝叶斯解列和贝叶斯风险列。如果 δ_0 是决策函数且它的风险函数满足

$$\max_{\theta\in\Theta} R(\theta,\delta_0)\leq \lim_{k\to\infty} R_k(\delta_k)<\infty$$

则 δ_0 是该决策问题的最小最大解。

证明： 用反证法。若 δ_0 不是最小最大解，则存在一个决策函数 $\hat{\delta}$，它的最大风险要小于 δ_0 的最大风险

$$\max_{\theta\in\Theta} R(\theta,\hat{\delta})<\max_{\theta\in\Theta} R(\theta,\delta_0)$$

另一方面，由定理条件知 δ_k 是在先验分布 π_k 下的贝叶斯解（$k\geq 1$），故其相应的贝叶斯风险最小，从而有

$$R_k(\delta_k)\leq R_k(\hat{\delta})=\int_\Theta \mathrm{R}(\theta,\hat{\delta})\pi_k(\theta)d\theta\leq \max_{\theta\in\Theta}\mathrm{R}(\theta,\hat{\delta})$$

于是

$$\lim_{k\to\infty} R_k(\delta_k)\leq \max_{\theta\in\Theta}\mathrm{R}(\theta,\hat{\delta})<\max_{\theta\in\Theta} R(\theta,\delta_0)$$

这与定理给定的条件矛盾。

例 7.6　设 $\mathbf{x}=(x_1,\cdots,x_n)$ 是来自总体 $N(\theta,1)$ 的简单随机样本，损失函数为平方损失函数 $L(\theta,\delta)=(\delta-\theta)^2$，试判断 $\bar{\mathbf{x}}=\sum_{i=1}^{n}x_i\big/n$ 是否为均值参数 θ 的最小最大估计。

解：我们利用定理 7.7 的思路来做。取 $\pi_k\sim N(0,k^2),k\ge 1$ 作为先验分布列，则依据例 2.1 和定理 5. 3，θ 的相应贝叶斯估计为后验期望

$$\delta_k(\mathbf{x})=E^{\pi_k(\theta|\mathbf{x})}(\theta)=\frac{nk^2}{1+nk^2}\bar{\mathbf{x}}$$

其风险函数为

$$R_k(\theta,\delta_k)=E^{\mathbf{x}|\theta}\left(\frac{nk^2}{1+nk^2}\bar{\mathbf{x}}-\theta\right)^2$$

由 $x_i\,|\,\theta\sim N(\theta,1),i=1,2,\cdots,n$ 易知 $\mathbf{x}\,|\,\theta\sim N(\boldsymbol{\theta},\mathbf{I}_n)$，其中 $\boldsymbol{\theta}=(\theta,\cdots,\theta)^T$ 为全部由 θ 组成的列向量，$\mathbf{I}_n$ 为单位矩阵。由概率论知

$$E^{\mathbf{x}|\theta}(\bar{\mathbf{x}})=\theta\text{，}\quad \mathrm{var}^{\mathbf{x}|\theta}(\bar{\mathbf{x}})=\frac{1}{n}$$

于是

$$R_k(\theta,\delta_k)=E^{\mathbf{x}|\theta}\left(\frac{nk^2}{1+nk^2}\bar{\mathbf{x}}-\theta\right)^2$$

$$=\frac{1}{(1+nk^2)^2}E^{\mathbf{x}|\theta}[nk^2(\bar{\mathbf{x}}-\theta)-\theta]^2$$

$$=\frac{1}{(1+nk^2)^2}E^{\mathbf{x}|\theta}[n^2k^4(\bar{\mathbf{x}}-\theta)^2-2nk^2\theta(\bar{\mathbf{x}}-\theta)+\theta^2]$$

$$=\frac{1}{(1+nk^2)^2}\{E^{\mathbf{x}|\theta}[n^2k^4(\bar{\mathbf{x}}-\theta)^2]+\theta^2\}$$

$$=\frac{1}{(1+nk^2)^2}[n^2k^4\operatorname{var}^{\mathbf{x}|\theta}(\bar{\mathbf{x}})+\theta^2]$$

$$=\frac{nk^4+\theta^2}{(1+nk^2)^2}$$

从而 δ_k 的贝叶斯风险

$$R_k(\delta_k)=E^{\theta}\left(\frac{nk^4+\theta^2}{(1+nk^2)^2}\right)=\frac{nk^4+k^2}{(1+nk^2)^2}=\frac{k^2}{1+nk^2}$$

显然 $\bar{\mathbf{x}}$ 的风险函数

$$R(\theta,\bar{\mathbf{x}})=\operatorname{var}^{\mathbf{x}|\theta}(\bar{\mathbf{x}})=\frac{1}{n}\le\lim_{k\to\infty}R_k(\delta_k)=\lim_{k\to\infty}\frac{k^2}{1+nk^2}=\frac{1}{n}$$

由定理 7.7， $\bar{\mathbf{x}}$ 为均值参数 θ 的最小最大估计。