数理统计 · 2025 Spring

第五次习题课

2025 年 6 月

统计决策理论的三要素

- **统计模型**: 即样本分布族 $\{f(x \mid \theta), \theta \in \Theta\}$ 以及参数空间 Θ.
- **行动空间**: 即可能采取的行动所构成的非空集合 $\mathcal{A} = \{a = \delta(x)\}$.
- **损失函数**: 即参数为 θ 时采取行动 $a \in \mathcal{A}$ 所蒙受的损失 $L(a, \theta)$.
  - 平方损失 $L(a, \theta) = (a - \theta)^2$.
  - 绝对值损失 $L(a, \theta) = |a - \theta|$.
  - 检验问题中的损失 $L(a, \theta) = \begin{cases} 1, & \text{若 } a = 1,\ \theta \in \Theta_0, \\ c, & \text{若 } a = 0,\ \theta \in \Theta_1, \\ 0, & \text{其它}. \end{cases}$
- 对贝叶斯统计决策理论而言, 还需要指定参数 θ 的**先验分布族** $\{\pi(\theta)\}$.

$$\boxed{\text{观察到样本 } \boldsymbol{x} \Rightarrow \text{在行动空间 } \mathcal{A} \text{ 中采取行动 } a \Rightarrow \text{最小化损失 } L(a, \theta)}$$

Definition 1 (风险函数)

设 $\delta(\boldsymbol{x})$ 是 θ 的一个决策函数, $L(a,\theta)$ 是损失函数, 称平均损失

$$R(\delta,\theta)=\mathbb{E}_\theta\left[L(\delta(\boldsymbol{X}),\theta)\right]=\int_{\mathscr{X}} L(\delta(\boldsymbol{x}),\theta)f(\boldsymbol{x}\mid\theta)\mathrm{d}\boldsymbol{x}$$

为 $\delta(\boldsymbol{x})$ 的风险函数.

Example 1

设 $X_1,\ldots,X_n$ i.i.d. $\sim \mathrm{Exp}(\lambda)$, 讨论 λ 的 MLE $\widehat{\lambda}$ 及其纠偏版本 $\widetilde{\lambda}=c\widehat{\lambda}$ 在平方损失及损失 $L(a,\lambda)=\frac{\lambda}{a}-\log\frac{\lambda}{a}-1$ 下的可容许性.

Hint: 参数 λ 的 MLE 为 $\widehat{\lambda}=1/\overline{X}$, 其纠偏版本为 $\widetilde{\lambda}=\frac{n-1}{n\overline{X}}$. 在平方损失下 $\widehat{\lambda}$ 不可容许:

$$\mathbb{E}(\widehat{\lambda}-\lambda)^2=\frac{(n+2)\lambda^2}{(n-1)(n-2)}>\mathbb{E}(\widetilde{\lambda}-\lambda)^2=\frac{\lambda^2}{n-2},\quad \forall\lambda>0.$$

在损失 L 下 $\widetilde{\lambda}$ 不可容许: $R(\widetilde{\lambda},\lambda)-R(\widehat{\lambda},\lambda)=\frac{1}{n-1}-\log\frac{n}{n-1}>0.$

Definition 2 (一致最优决策函数)

1. 设 δ_1 和 δ_2 为 θ 的两个不同的决策函数，若
$$R(\delta_1,\theta) \le R(\delta_2,\theta), \quad \forall\theta \in \Theta,$$
且至少存在一个 $\theta_0 \in \Theta$ 使严格不等号成立，则称 δ_1 **优于** δ_2.
2. 若存在 $\delta^\star$，使得对 θ 的任意一个决策函数 δ，都有
$$R(\delta^\star,\theta) \le R(\delta,\theta), \quad \forall\theta \in \Theta,$$
则称 $\delta^\star$ 为**一致最优决策函数**.

- 对决策函数 δ，若不存在一致优于它的决策函数，则称它为**可容许**的决策函数，所有其它的决策函数都是**不可容许**的.
- 可容许决策不会比其他决策函数一致地好，也不会一致地差.

由于要求一致最优性, 这样的可容许的决策一般不存在, 人们更多采用

- **Minimax 决策**: 要求让最差的风险最小

$$\delta^\star = \arg\min_{\delta\in\mathcal{A}} \overline{R}(\delta) = \arg\min_{\delta\in\mathcal{A}} \sup_{\theta\in\Theta} R(\delta,\theta).$$

- **Bayes 决策**: 对 θ 赋予先验再让平均的风险函数 (称为**贝叶斯风险**) 最小

$$\delta^\star_\pi = \arg\min_{\delta\in\mathcal{A}} R_\pi(\delta) = \arg\min_{\delta\in\mathcal{A}} \int_\Theta R(\delta,\theta)\pi(\theta)\mathrm{d}\theta.$$

观察贝叶斯风险的形式:

$$
\begin{aligned}
R_\pi(\delta) &= \int_\Theta \boxed{R(\delta,\theta)}\pi(\theta)\mathrm{d}\theta \\
&= \int_\Theta \boxed{\int_{\mathscr{X}} L(\delta(\boldsymbol{x}),\theta)f(\boldsymbol{x}\mid\theta)\mathrm{d}\boldsymbol{x}}\,\pi(\theta)\mathrm{d}\theta \\
&= \int_{\mathscr{X}} \boxed{\int_\Theta L(\delta(\boldsymbol{x}),\theta)\pi(\theta\mid\boldsymbol{x})\mathrm{d}\theta}\,m(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \\
&\triangleq \int_{\mathscr{X}} \boxed{R_\pi(\delta(\boldsymbol{x})\mid\boldsymbol{x})}\,m(\boldsymbol{x})\mathrm{d}\boldsymbol{x},
\end{aligned}
$$

这里的 $R_\pi(\delta(\boldsymbol{x})\mid\boldsymbol{x})$ 是直接对损失函数按照 θ 的后验分布求期望的结果, 称之为**贝叶斯后验风险**. 所以要求贝叶斯决策可以直接最小化这个贝叶斯后验风险.

- 平方损失 $L(a,\theta)=(a-\theta)^2$ 下的 Bayes 估计为后验期望:

$$\widehat{\theta}_B(\boldsymbol{x})=\mathbb{E}(\theta\mid\boldsymbol{x}).$$

- 加权平方损失 $L(a,\theta)=w(\theta)(a-\theta)^2$ 下的 Bayes 估计为

$$\widehat{\theta}_B(\boldsymbol{x})=\frac{\mathbb{E}(\theta w(\theta)\mid\boldsymbol{x})}{\mathbb{E}(w(\theta)\mid\boldsymbol{x})}.$$

- 绝对值损失 $L(a,\theta)=|a-\theta|$ 下的 Bayes 估计为后验中位数.
- 假设检验中常用 0-k_i 损失 $L(a_i,\theta)=\begin{cases}0, & \theta\in\Theta_i,\\ k_i, & \theta\notin\Theta_i,\end{cases}\quad i=0,1,$

  当 $k_0\mathbb{P}(\theta\in\Theta_1\mid\boldsymbol{x})>k_1\mathbb{P}(\theta\in\Theta_0\mid\boldsymbol{x})$ 时接受假设 H_1.
- 多行动假设检验问题中, 计算诸行动的后验风险再比较即可.

Example 2 (7.3)

设 $X_1, \ldots, X_n$ 为抽自参数为 θ 的 Poisson 分布 $P(\theta)$ 的一组样本, 假如未知参数 θ 的先验分布为指数分布 $\mathrm{Exp}(\lambda)$. 在平方损失函数下, 求 θ 的 Bayes 估计.

Hint:

$$\theta \mid \boldsymbol{X} = \boldsymbol{x} \sim \Gamma(n\bar{x} + 1, n + \lambda) \;\Rightarrow\; \hat{\theta}_B(\boldsymbol{X}) = \mathbb{E}(\theta \mid \boldsymbol{X}) = \frac{n\bar{X} + 1}{n + \lambda}.$$

Example 2 (7.3)

设 $X_1, \ldots, X_n$ 为抽自参数为 θ 的 Poisson 分布 $P(\theta)$ 的一组样本, 假如未知参数 θ 的先验分布为指数分布 $\mathrm{Exp}(\lambda)$. 在平方损失函数下, 求 θ 的 Bayes 估计.

Hint:

$$\theta \mid \boldsymbol{X} = \boldsymbol{x} \sim \Gamma(n\bar{x} + 1, n + \lambda) \ \Rightarrow\ \hat{\theta}_B(\boldsymbol{X}) = \mathbb{E}(\theta \mid \boldsymbol{X}) = \frac{n\bar{X} + 1}{n + \lambda}.$$

Example 3 (7.8)

设随机变量 X 服从 Gamma 分布 $\Gamma(\alpha, 1/\theta)$, 其中 α 已知, θ 未知, 其先验分布为无信息先验, 即 $\pi(\theta) = (1/\theta) I_{(0,\infty)}(\theta)$. 设 $X_1, \ldots, X_n$ i.i.d. $\sim X$, 在加权平方损失 $L(d, \theta) = (d - \theta)^2/\theta^2$ 下, 求 θ 的 Bayes 估计.

Hint:

$$\theta \mid \boldsymbol{X} = \boldsymbol{x} \sim \Gamma^{-1}\left(n\alpha, \sum_{i=1}^{n} x_i\right) \ \Rightarrow\ \hat{\theta}_B(\boldsymbol{X}) = \frac{\mathbb{E}(\theta^{-1} \mid \boldsymbol{X})}{\mathbb{E}(\theta^{-2} \mid \boldsymbol{X})} = \frac{n\bar{X}}{n\alpha + 1}.$$

Example 4 (7.9)

设 θ 为某产品的废品率, 从该批产品中随机抽取 n 件, 其中废品数 $X \mid \theta \sim B(n,\theta)$, 令 θ 的先验分布为 $(0,1)$ 上的均匀分布. 若 $n=x=2$, 在绝对值损失函数下求 θ 的 Bayes 估计.

Hint:

$$\theta \mid X=x \sim \mathrm{Be}(x+1, n-x+1) \xrightarrow{n=x=2} \theta \mid X=2 \sim \mathrm{Be}(3,1).$$

即 θ 的后验分布函数为

$$\Pi(\theta \mid 2) = \begin{cases} 0, & \theta < 0, \\ \theta^3, & 0 \le \theta < 1, \\ 1, & \theta \ge 1. \end{cases}$$

所以

$$\hat{\theta}_B(2) = \Pi^{-1}_{\theta|X=2}\left(\frac{1}{2}\right) = 2^{-\frac{1}{3}}.$$

Example 5 (补充题, 2023 • 期末)

随机地从一批铁钉中抽取 16 枚, 测得它们的平均长度 (单位: cm) $\bar{x} = 3.035$. 已知铁钉长度 X 服从正态分布 $N(\mu, 0.1^2)$. 考虑假设检验问题

$$H_0 : \mu \leq 3 \longleftrightarrow H_1 : \mu > 3. \tag{$\star$}$$

1. 在 $\alpha = 0.05$ 水平下对检验问题 ($\star$) 进行检验, 并给出检验的 p 值. (8 分)
2. 如果行动 $a = 0$ 和 $a = 1$ 分别表示接受 H_0 和拒绝 H_0, μ 的先验分布为 $N(3, 0.1^2)$, 损失函数取为

$$L(a, \mu) = \begin{cases} 11, & a = 1, \mu \leq 3, \\ 1, & a = 0, \mu > 3, \\ 0, & \text{其他}. \end{cases}$$

   基于后验风险最小原则给出检验问题 ($\star$) 的最优决策行动, 并与 (1) 中检验结论进行对比. (7 分)

附表: $u_{0.025} = 1.960$, $u_{0.05} = 1.645$, $\Phi(1.36) = 0.9131$, $\Phi(1.4) = 0.9192$.

1. 给出检验统计量 $U(\boldsymbol{X})=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\overset{H_0}{\sim}N(0,1)$. (2 分)
由检验问题的形式知拒绝域为 $D=\{\boldsymbol{X}:U(\boldsymbol{X})>u_\alpha\}$. (1 分)
现有 $\overline{x}=3.035,\mu_0=3,\sigma=0.1,n=16$, 查表得 $u_{0.05}=1.645$, 因此
$$u(\boldsymbol{x})=\frac{3.035-3}{0.1/\sqrt{16}}=1.4<1.645.$$ (2 分)
因此不能拒绝原假设. (1 分)
检验的 p 值为 $p=\mathbb{P}(U(\boldsymbol{X})>u(\boldsymbol{x}))=1-\Phi(1.4)=0.0808$. (2 分)

2. 在先验分布 $\mu\sim N(\mu_0,\tau^2)$ 下, 正确计算出 μ 的后验分布
$$\mu\mid\boldsymbol{x}\sim N\left(\frac{n\tau^2\overline{x}+\sigma^2\mu_0}{n\tau^2+\sigma^2},\frac{\tau^2\sigma^2}{n\tau^2+\sigma^2}\right)\xlongequal{\text{代入数据}}N\left(\frac{1289}{425},\frac{1}{1700}\right).$$ (3 分)
后验风险 $R(a_0,\mu)=\mathbb{P}(\mu>3\mid\boldsymbol{x})=0.9131$, $R(a_1,\mu)=11\mathbb{P}(\mu\leq3\mid\boldsymbol{x})=0.9559$. (2 分)
结论: 按后验风险最小原则, 应采取行动 a_0, 即接受原假设 H_0. (1 分)
这与 (1) 中检验结果一致, 这是因为我们对拒绝 H_0 赋予了较大的惩罚. (1 分)

容许性、Minimax 决策与 Bayes 决策

- 若 $\delta^\star$ 是可容许决策, 其风险函数为常数, 则它是 Minimax 决策.
- 设 $\delta_\pi^\star$ 是先验 π 下的 Bayes 决策, 其风险函数为常数, 则它是 Minimax 决策.
- 设 $\{\pi_k\}$ 为一列先验分布, $\delta_k^\star$ 是 π_k 下的 Bayes 决策, 其贝叶斯风险为 r_k. 假设 $\lim\limits_{k\to\infty} r_k = r < \infty$, 决策函数 δ 满足最大风险 $\overline{R}(\delta) \le r$, 则 δ 是 Minimax 决策.
- 设 θ 的先验密度 $\pi(\theta)$ 是严格正的, Bayes 决策 $\delta_\pi^\star$ 有有限的贝叶斯风险, 且风险函数关于 θ 连续, 则它是可容许的.
- 最优 Bayes 决策 $R(\pi) := \inf\limits_{\delta\in\mathcal{A}} R_\pi(\delta)$ 不会大于最优 Minimax 决策 $\overline{R} := \inf\limits_{\delta\in\mathcal{A}} \overline{R}(\delta)$, 即

$$R(\pi) = \inf_{\delta\in\mathcal{A}} R_\pi(\delta) \le \sup_\pi \inf_{\delta\in\mathcal{A}} R_\pi(\delta) \le \inf_{\delta\in\mathcal{A}} \sup_\pi R_\pi(\delta) = \overline{R}.$$

  如果存在一个先验 $\pi^\star$ 使得 $R(\pi^\star) = \sup\limits_\pi \inf\limits_{\delta\in\mathcal{A}} R_\pi(\delta)$, 则称之为**最不利 (least favorable) 分布**.

Example 6 (7.15)

设随机变量 X 服从二项分布 $B(n,p)$, $0<p<1$, 证明: $d(x)=x/n$ 在损失函数 $L(d,p)=(d-p)^2/[p(1-p)]$ 下是 p 的 Minimax 估计.

Hint: 考虑 p 的先验分布为共轭先验 $\mathrm{Be}(a,b)$, 则

$$p \mid X=x \sim \mathrm{Be}(x+a, n-x+b).$$

从而在加权平方损失 $L(d,p)=(d-p)^2/[p(1-p)]$ 下, p 的 Bayes 估计为

$$\delta_{a,b}(x)=\frac{\mathbb{E}\left((1-p)^{-1} \mid X=x\right)}{\mathbb{E}\left([p(1-p)]^{-1} \mid X=x\right)}=\frac{x+a-1}{n+a+b-2}.$$

它的风险函数

$$R(\delta_{a,b},p)=\frac{1}{p(1-p)}\mathbb{E}\left(\frac{X+a-1}{n+a+b-2}-p\right)^2=\frac{n+\frac{[a-1-(a+b-2)p]^2}{p(1-p)}}{(n+a+b-2)^2}.$$

若取 $a=b=1$, 则风险函数为常数, 即在均匀先验 $U(0,1)$ 下, p 的 Bayes 估计 $\delta_{1,1}=x/n$ 为 p 的 Minimax 估计.

Example 7 (补充题)

对成功概率为 $1-p$ $(0<p<1)$ 的独立 Bernoulli 试验, 当试验进行到 r 次成功时, 令失败的总次数为随机变量 X, 其概率分布为

$$\mathbb{P}(X=x\mid p)=\binom{x+r-1}{r-1}p^x(1-p)^r,\quad x=0,1,2,\ldots.$$

令 $\theta=p/(1-p)>0$, 易知 $\mathbb{E}(X\mid\theta)=r\theta$, $\mathrm{Var}(X\mid\theta)=r\theta(1+\theta)$. 证明: $d(x)=x/r$ 在损失函数 $L(d,\theta)=(d-\theta)^2/[\theta(1+\theta)]$ 下是 θ 的 Minimax 估计.

Hint:

$$\mathbb{P}(X=x\mid\theta)=\binom{x+r-1}{r-1}\theta^x(1+\theta)^{-(x+r)},\quad x=0,1,2,\ldots.$$

考虑 θ 的先验分布为共轭先验 (二型贝塔分布, Wiki-link)

$$\pi(\theta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1+\theta)^{-(\alpha+\beta)},\quad \theta>0,$$

其中 $\alpha>0,\beta>0$ 为参数.

在共轭先验下, θ 的后验密度函数为

$$\pi(\theta \mid x) = \frac{\Gamma(x+\alpha+r+\beta)}{\Gamma(x+\alpha)\Gamma(r+\beta)}\theta^{x+\alpha-1}(1+\theta)^{-(x+\alpha+r+\beta)}, \quad \theta > 0.$$

在损失函数 $L(d,\theta) = (d-\theta)^2/[\theta(1+\theta)]$ 下, θ 的 Bayes 估计为

$$\delta_{\alpha,\beta}(x) = \frac{\mathbb{E}\big((1+\theta)^{-1} \mid X = x\big)}{\mathbb{E}\big([\theta(1+\theta)]^{-1} \mid X = x\big)} = \frac{X+\alpha-1}{r+\beta+1}.$$

它的风险函数

$$R(\delta_{\alpha,\beta}(x),\theta) = \frac{1}{\theta(1+\theta)}\mathbb{E}\left(\frac{X+\alpha-1}{r+\beta+1} - \theta\right)^2 = \frac{r + \frac{(\alpha-1-(\beta+1)\theta)^2}{\theta(1+\theta)}}{(r+\beta+1)^2}.$$

此时风险函数不可能为常数. 但受此启发, 考虑 $\alpha = 1, \beta = -1$, 由此考虑无信息先验 $\pi(\theta) = 1, \theta > 0$, 从而 Bayes 估计 $d(x) = x/r$ 是 θ 的 Minimax 估计.

Example 8 (7.22)

设 $X_1,\dots,X_n$ i.i.d. $N(\theta,1)$. 要估计 θ, 令损失函数为平方损失, 取估计量 $\hat{\theta}_n = c_1X_1+\cdots+c_nX_n$. 证明: 若 $c_1+\cdots+c_n=1$, 则除非 $c_1=\cdots=c_n=1/n$, 否则 $\hat{\theta}_n$ 是不可容许的.

Hint: 记 $\boldsymbol{X}=(X_1,\dots,X_n)^\top$, $\boldsymbol{c}=(c_1,\dots,c_n)^\top$, 有 $\boldsymbol{c}^\top\boldsymbol{1}=1$. 于是 $\hat{\theta}_n$ 的风险函数为

$$\begin{aligned} R(\hat{\theta}_n,\theta) &= \mathbb{E}\left[(\boldsymbol{c}^\top\boldsymbol{X}-\theta)^2\right] \\ &= \mathbb{E}\left[\boldsymbol{c}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{c}-2\theta\boldsymbol{c}^\top\boldsymbol{X}+\theta^2\right] \\ &= \boldsymbol{c}^\top(\boldsymbol{I}_n+\theta^2\boldsymbol{1}\boldsymbol{1}^\top)\boldsymbol{c}-2\theta^2+\theta^2 \\ &= \boldsymbol{c}^\top\boldsymbol{c}=c_1^2+\cdots+c_n^2. \end{aligned}$$

显然当 $c_1=\cdots=c_n=1/n$ 时风险函数取最小值.

Example 9 (2023・期末)

设总体 X 的密度函数为

$$f(x;\theta)=\begin{cases}\theta, & 0<x<1,\\ 1-\theta, & 1\le x<2,\\ 0, & \text{其他},\end{cases}$$

其中 $0<\theta<1$ 为未知参数. 现从此总体中抽出一样本量为 n 的样本 $X_1,\ldots,X_n$.

1. 试求参数 θ 的充分统计量, 并说明它是否为完全统计量?
2. 求假设检验问题 $H_0:\theta=\theta_0 \leftrightarrow H_1:\theta\neq\theta_0$ 的水平 α 似然比检验, 其中 $0<\theta_0,\alpha<1$ 已知.
3. 上述假设是否存在 UMPT? 为什么?

Example 10

袋中有 8 个球, 其中红球数未知, 在其中任取 3 个, 记录红球的个数为 X, 然后放回再任取 3 个, 记录红球个数, 然后放回, 如此反复进行了 112 次, 得到结果如下:

X	0	1	2	3
次数	1	31	55	25

试在 $\alpha = 0.05$ 水平下检验假设 H_0 : 红球的个数为 5.

预祝大家期末考取得好成绩!

Thanks for your time and attention!