STAT201B Lecture 19-20 Minimax Rule

1 Minimax rule

1.1 一个例子: 使用 Bayes rule

情境:

考虑是否购买高风险债券:

- 债券的购入价为 1000 美元
- 如果购买债券,
  - 到期时可以获得 500 美元的净收益
  - 但债券有 0.1 的概率出现违约 (default), 即投入的 1000 美元全部损失
- 如果不购买这些债券,
  - 将确保能利用资金获得 300 美元的净收益

问题:

1. 求 parameter space Θ 和 space of possible actions $\mathcal{A}$
2. 求 prior distribution
3. 对于所有的 $\theta \in \Theta$ 和 $a \in \mathcal{A}$, 求对应的 loss
4. 判断是否存在 inadmissible action

解答:

Question 1:

- $\Theta = \{\theta_1, \theta_2\}$, 其中 θ_1 表示 default, θ_2 表示 no default
- $\mathcal{A} = \{a_1, a_2\}$, 其中 a_1 表示 buy, a_2 表示 do not buy

Question 2:

Prior distribution 为

$$\mathbb{P}(\theta = \theta_1) = 0.1, \qquad \mathbb{P}(\theta = \theta_2) = 0.9$$

Question 3:

所有的 $\theta \in \Theta$ 和 $a \in \mathcal{A}$ 对应的 loss 可以被汇总至以下表格中:

	a_1	a_2
θ_1	1000	-300
θ_2	-500	-300

Remark

为了确保所有 loss 均非负, 我们可以对每个 loss 加上 500, 即:

	a_1	a_2
θ_1	1500	200
θ_2	0	200

Question 4:

由于没有 data, (frequentist) risk 即为 loss, 因此 a_1 和 a_2 对应的 Bayes risk 分别为:

$$\begin{aligned} r(f, a_1) &= R(\theta_1, a_1) \cdot \mathbb{P}(\theta = \theta_1) + R(\theta_2, a_1) \cdot \mathbb{P}(\theta = \theta_2) \\ &= 1000 \times 0.1 + (-500) \times 0.9 \\ &= -350 \\ r(f, a_2) &= -300 \end{aligned}$$

因此 a_1 为 Bayes rule
因此 a_2 为 inadmissible

> **Remark**
>
> 若我们使用非负的 loss, 也能得到一样的结论:
>
> $$\begin{aligned} r(f, a_1) &= R(\theta_1, a_1) \cdot \mathbb{P}(\theta = \theta_1) + R(\theta_2, a_1) \cdot \mathbb{P}(\theta = \theta_2) \\ &= 1500 \times 0.1 + 0 \times 0.9 \\ &= 150 \\ r(f, a_2) &= 200 \end{aligned}$$

1.2 一个例子: 使用 Minimax rule

> **Logic**
>
> 继续之前的例子, a_1 作为 Bayes rule 能 minimize 平均情况下的 risk
>
> 但假设 investor 非常 conservative, 即希望策略能在 "worst case scenario" 下使 risk 最小, 此时需要使用 minimax rule

记关于 action a 的 frequentist risk 为 $R(\theta, a)$, maximum risk 为 $\bar{R}(a) = \sup_\theta R(\theta, a)$, 则在上例中, 我们有

$$\bar{R}(a_1) = 1000, \qquad \bar{R}(a_2) = -300$$

因此 a_2 为 minimax rule

1.3 Definition: Minimax rule

令:

- possible actions 为 estimator $\hat{\theta}$
- maximum risk 为 $\bar{R}(\hat{\theta}) = \sup_\theta R(\theta, \hat{\theta})$

则 **minimax rule** $\hat{\theta}$ 会 minimize frequentist risk 的最大值:

$$\sup_\theta R(\theta, \hat{\theta}) = \inf_{\tilde{\theta}} \sup_\theta R(\theta, \tilde{\theta})$$

即

$$\bar{R}(\hat{\theta}) = \inf_{\tilde{\theta}} \bar{R}(\tilde{\theta})$$

> **Example**
>
> ***问题:***
> 令:
>
> - $X \sim \mathcal{N}(0, 1)$
> - 使用 squared error loss
> - 使用 estimator $\hat{\theta}_c(x) = cx$
>
> 求:

1. $\sup_\theta R(\theta, \hat{\theta}_c)$
2. minimax estimator of θ
3. 令 θ 的 prior distribution 为 $\mathcal{N}(a, b)$, 求 Bayes estimator of θ

解答:
Question 1:

首先求出 risk (计算过程见 STAT201B Lecture 17-18 Decision Theory):

$$R(\theta, \hat{\theta}) = \mathbb{E}_{X|\theta}[(\theta - cX)^2] = (c-1)^2\theta^2 + c^2$$

因此 maximum of frequentist risk 为:

$$\bar{R}(\hat{\theta}) = \sup_\theta R(\theta, \hat{\theta}) = \begin{cases} 1 & \text{if } c = 1 \\ \infty & \text{if } c \neq 1 \end{cases}$$

Question 2:

Minimax estimator of θ 为 $c = 1$

Question 3:

⚠ Remark: 一个错误思路

一个错误的思路是: 求出 θ 的 posterior distribution, 随后求出 posterior mean 作为 Bayes rule

$$\begin{aligned} f(\theta|X) &\propto \exp\left\{-\frac{(X-\theta)^2}{2}\right\} \cdot \exp\left\{-\frac{(\theta-a)^2}{2b}\right\} \\ &\propto \exp\left\{-\frac{b\theta^2 - 2b\theta X + \theta^2 - 2a\theta}{2b}\right\} \\ &\propto \exp\left\{-\frac{(b+1)\theta^2 - 2(bX+a)\theta}{2b}\right\} \\ &\propto \exp\left\{-\frac{(b+1)\left(\theta - \frac{bX+a}{b+1}\right)^2}{2b}\right\} \end{aligned}$$

注意到此时的 posterior mean 为 $\mathbb{E}_{\theta|X}[\theta] = \frac{b}{b+1}X + \frac{1}{b+1}a$, 是 prior mean 和 MLE 的 weighted sum, 但这并不是此处的 Bayes rule, 因为此处对于 estimator 的形式作出了限制

我们直接利用 frequentist risk 计算并 minimize 关于 c 的 Bayes risk:

$$\begin{aligned} r(f, \hat{\theta}_c) &= \mathbb{E}_\theta[R(\theta, \hat{\theta}_c)] \\ &= \mathbb{E}_\theta[(c-1)^2\theta^2 + c^2] \\ &= (c-1)^2(a^2+b) + c^2 \\ &= (a^2+b+1)c^2 - 2(a^2+b)c + a^2 + b \\ &= (a^2+b+1)\left(c - \frac{a^2+b}{a^2+b+1}\right)^2 + a^2 + b - \frac{(a^2+b)^2}{a^2+b+1} \end{aligned}$$

因此当 $c = \frac{a^2+b}{a^2+b+1}$ 时, $\hat{\theta}_c$ minimizes $r(f, \hat{\theta}_c)$, 即为 Bayes estimator

2 Bayes 和 Minimax Points 的 Geometry (Finite Parameter Space)

2.1 Definition: Risk set

考虑 **finite parameter space** $\Omega = \{\theta_1, \dots, \theta_k\}$, 则 risk set $S \subset \mathbb{R}^k$ 被定义为:

$$S = \{(y_1, \dots, y_k) : y_i = R(\theta_i, a) \text{ for } a \in \mathcal{A}\}$$

⚠ Remark

注意到 S 是一个 set of vectors:

- 每个 vector 对应了 $\mathcal{A}$ 中的某一个 action a
- 每个 vector 的每个 element 对应了某个 action a 和某个 parameter θ_i 的 (frequentist) risk

---

2.2 Lemma: Risk set 的 convexity

若 $\mathcal{A}$ 中为 randomized estimators, 则 risk set S 为 convex

Remark: randomized estimators

区别于之前的例子与定理, 此处考虑的 estimator 不再为 non-randomized, 我们可以通过 non-randomized estimators 构造 randomized estimators, 一个例子是:

对于 $a_1, a_2, v \in (0,1)$, 我们可以根据 a_1 和 a_2 构造以下 randomized estimator:

$$\tilde{a} = \begin{cases} a_1 & \text{with probability } v \\ a_2 & \text{with probability } 1-v \end{cases}$$

$\tilde{a}$ 的 risk 为:

$$\begin{aligned} R(\theta, \tilde{a}) &= \mathbb{E}_{X|\theta}[\mathcal{L}(\theta, \tilde{a})] \\ &= \mathbb{E}_{X|\theta}[v \cdot \mathcal{L}(\theta, a_1) + (1-v) \cdot \mathcal{L}(\theta, a_2)] \\ &= vR(\theta, a_1) + (1-v)R(\theta, a_2) \end{aligned}$$

令 $\mathcal{A} = \{a_1, a_2, \tilde{a} | v \in (0,1)\}$, 则 risk set S 为 $R(\theta, a_1)$ 和 $R(\theta, a_2)$ 的连线, 即为 convex

---

2.3 Bayes risk 的向量内积形式

在上述 setting 下, 由于 parameter space 为 finite, θ 的 prior 可以写作一个 finite vector:

$$\boldsymbol{\lambda}(\theta) = (\lambda_1, \dots, \lambda_k) = (\lambda(\theta_1), \dots, \lambda(\theta_k))$$

其中的 elements 作为 prior probability 需要满足: $\sum_{i=1}^{k} \lambda_i = 1$ 和 $\lambda_i \geq 0$

则 Bayes risk 可以写作:

$$r(\Lambda, a) = \sum_{i=1}^{k} \lambda_i R(\theta_i, a) = (\lambda_1, \dots, \lambda_k) \begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix}$$

2.4 Geometry of Bayes point and Minimax point

考虑 Bayes risk 相同的点组成的 hyperplane:

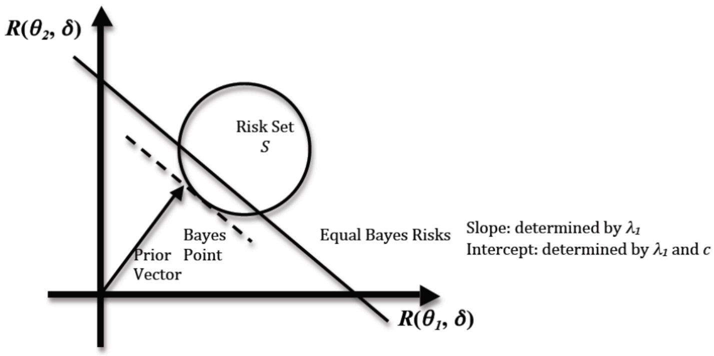
$$\{(y_1, \dots, y_k) : (\lambda_1, \dots, \lambda_k) \begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix} = c,\ y_i = R(\theta_i, a)\ \text{ for } a \in \mathcal{A}\}$$

Remark

若我们 fix prior vector $(\lambda_1, \dots, \lambda_k)$, 则 hyperplane 垂直于 prior vector (因为 hyperplane 上的点在 prior vector 方向的投影相同)

2.4.1 Geometry of Bayes point

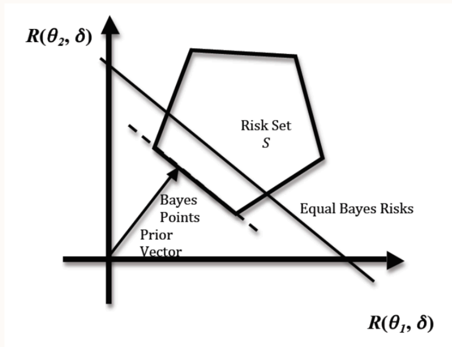
若给定 $\boldsymbol{\lambda}(\theta) = (\lambda_1, \dots, \lambda_k)$, 则 Bayes points 为 hyperplane 和 risk set 的相切点



Geometry of a Bayes Point for $k = 2$.

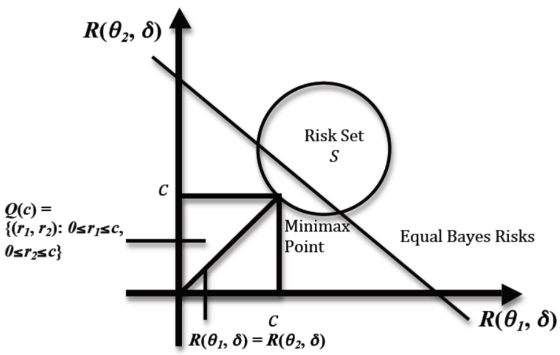
Remark

- 由于 hyperplane 上的点的 Bayes risk 均相同, 且随着 hyperplane 的上移, Bayes risk 逐渐增加, 因此最先和 risk set 触碰的点即为 Bayes rule
- 对于 $k = 2$ 的情况, hyperplane $\lambda_1 R(\theta_1, a) + \lambda_2 R(\theta_2, a) = c$ 的 slope 为 $-\frac{\lambda_1}{\lambda_2} = -\frac{\lambda_1}{1-\lambda_1}$, 和 x 轴的 intercept 为 $\frac{c}{\lambda_1}$
- Bayes points 可能不唯一:



2.4.2 Geometry of Minimax point

定义 Q_c 为 maximum risk 相等的 points 的集合: $Q_c = \{(R(\theta, \delta))_{\theta \in \Omega} : \sup\limits_{\theta \in \Omega} R(\theta, \delta) = c\}$ (形状为正方体), 逐步增加 c (扩大正方体), 则 minimax point(s) 为 Q_c 与 risk set S 最先接触的 point(s)



Remark

Minimax point 同样可能不唯一

Example

问题:
令:

- $X \sim \text{Ber}(p)$
- 考虑 hypothesis: $H_0 : p = p_0 = 0.3 \quad v.s. \quad H_1 : p = p_1 = 0.7$

- 考虑使用 zero-one loss: 若 a_i 为 decision rule, 则 $\mathcal{L}(p_i, a_j) = \begin{cases} 0, & p_i = a_j \\ 1, & p_i \neq a_j \end{cases}$

考虑以下四个 actions:

- a_1: always accept H_0
- a_2: accept H_0 if $X = 0$; accept H_1 if $X = 1$
- a_3: accept H_0 if $X = 1$; accept H_1 if $X = 0$
- a_4: always accept H_1

求:

1. 每个 decision rule 的 risk
2. 画出 risk set
3. minimax rule

解答:
Question 1:

我们逐个考虑所有 pairs (p_i, a_j) 的 risk:

$$
\begin{aligned}
&R(p_0,a_1) = \mathbb{E}_{X|p}[\mathcal{L}(p_0,a_1)] = \mathbb{E}_{X|p}[\mathcal{L}(0.3,0.3)] = 0 \\
&R(p_0,a_2) = \mathbb{E}_{X|p}[\mathcal{L}(p_0,a_2)] = \mathbb{E}_{X|p}[\mathcal{L}(0.3,a_2)|X=0]\cdot(1-p_0) + \mathbb{E}_{X|p}[\mathcal{L}(0.3,a_2)|X=1]\cdot p_0 = 0\times 0.7 + 1\times 0.3 = 0.3 \\
&R(p_0,a_3) = \mathbb{E}_{X|p}[\mathcal{L}(p_0,a_3)] = \mathbb{E}_{X|p}[\mathcal{L}(0.3,a_3)|X=0]\cdot(1-p_0) + \mathbb{E}_{X|p}[\mathcal{L}(0.3,a_3)|X=1]\cdot p_0 = 1\times 0.7 + 0\times 0.3 = 0.7 \\
&R(p_0,a_4) = \mathbb{E}_{X|p}[\mathcal{L}(p_0,a_4)] = \mathbb{E}_{X|p}[\mathcal{L}(0.3,0.7)] = 1 \\
&R(p_1,a_1) = \mathbb{E}_{X|p}[\mathcal{L}(p_1,a_1)] = \mathbb{E}_{X|p}[\mathcal{L}(0.7,0.3)] = 1 \\
&R(p_1,a_2) = \mathbb{E}_{X|p}[\mathcal{L}(p_1,a_2)] = \mathbb{E}_{X|p}[\mathcal{L}(0.7,a_2)|X=0]\cdot(1-p_1) + \mathbb{E}_{X|p}[\mathcal{L}(0.7,a_2)|X=1]\cdot p_1 = 1\times 0.3 + 0\times 0.7 = 0.3 \\
&R(p_1,a_3) = \mathbb{E}_{X|p}[\mathcal{L}(p_1,a_3)] = \mathbb{E}_{X|p}[\mathcal{L}(0.7,a_3)|X=0]\cdot(1-p_1) + \mathbb{E}_{X|p}[\mathcal{L}(0.7,a_3)|X=1]\cdot p_1 = 0\times 0.3 + 1\times 0.7 = 0.7 \\
&R(p_1,a_4) = \mathbb{E}_{X|p}[\mathcal{L}(p_1,a_4)] = \mathbb{E}_{X|p}[\mathcal{L}(0.7,0.7)] = 0
\end{aligned}
$$

即:

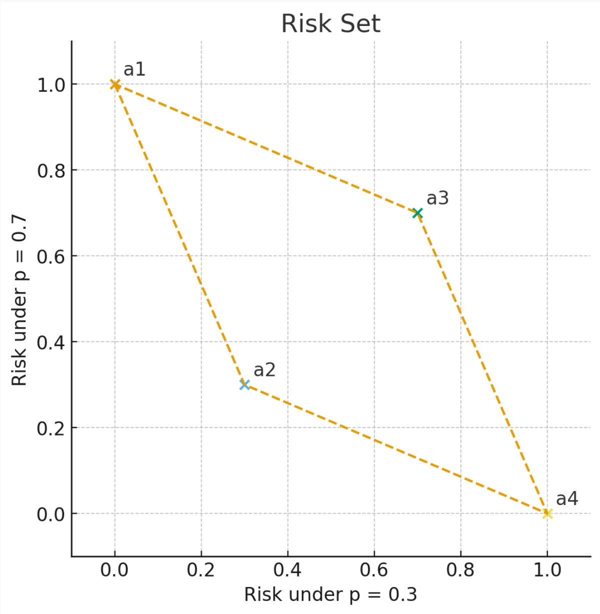
Action	$R(p_0)$	$R(p_1)$
a_1	0	1
a_2	0.3	0.3
a_3	0.7	0.7
a_4	1	0

Question 2:

Risk set 为:

$$S = \{(0,1), (0.3,0.3), (0.7,0.7), (1,0)\}$$

若允许 randomized estimator, 则 risk set 为这四个点的凸包



Question 3:

由于

$$\begin{aligned}
\bar{R}(a_1) &= 1 \\
\bar{R}(a_2) &= 0.3 \\
\bar{R}(a_3) &= 0.7 \\
\bar{R}(a_4) &= 1
\end{aligned}$$

则 minimax rule 为 a_2

> ⚠ **Remark**
>
> 我们也可以通过图像来判断 minimax rule: 从原点逐步扩大正方形, 最先触碰到的点为 a_2, 因此 minimax rule 为 a_2

Example

问题:

令:

- Possible states of nature 为 θ_1 和 θ_2
- X 为 random variable, 其 probability function $p(x|\theta)$ 为:

$$\mathbb{P}(X=0|\theta_1)=0.2, \quad \mathbb{P}(X=1|\theta_1)=0.8, \quad \mathbb{P}(X=0|\theta_2)=0.4, \quad \mathbb{P}(X=1|\theta_2)=0.6$$

- 两个 non-randomized actions a_1 和 a_2 的 loss 分别为:

$$\begin{aligned}
&\mathcal{L}(\theta_1,a_1(0))=1, \quad \mathcal{L}(\theta_1,a_1(1))=2, \quad \mathcal{L}(\theta_1,a_2(0))=4, \quad \mathcal{L}(\theta_1,a_2(1))=0 \\
&\mathcal{L}(\theta_2,a_1(0))=3, \quad \mathcal{L}(\theta_2,a_1(1))=1, \quad \mathcal{L}(\theta_2,a_2(0))=1, \quad \mathcal{L}(\theta_2,a_2(1))=4
\end{aligned}$$

求:

1. 给出并画出 risk set $S=\{(r_1,r_2): r_1=\lambda R(\theta_1,a_1)+(1-\lambda)R(\theta_1,a_2), r_2=\lambda R(\theta_2,a_1)+(1-\lambda)R(\theta_2,a_2), \lambda\in[0,1]\}$
2. 若 θ 的 prior distribution 为 $\Lambda(\theta)$: $\mathbb{P}(\theta=\theta_1)=0.9, \mathbb{P}(\theta=\theta_2)=0.1$, 求关于 Λ 的 Bayes rule
3. 求 minimax rule

解答:

(a) Based on the loss and likelihood function:

$R(\theta_1, a_1) = 1\times 0.2 + 2\times 0.8 = 1.8$

$R(\theta_2, a_1) = 3\times 0.4 + 1\times 0.6 = 1.8$

$R(\theta_1, a_2) = 4\times 0.2 + 0\times 0.8 = 0.8$

$R(\theta_2, a_2) = 1\times 0.4 + 4\times 0.6 = 2.8$

Therefore,

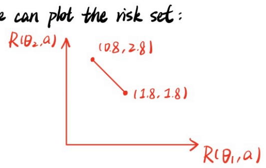
$r_1 = \lambda\cdot R(\theta_1, a_1) + (1-\lambda)\cdot R(\theta_1, a_2) = \lambda + 0.8$

$r_2 = \lambda\cdot R(\theta_2, a_1) + (1-\lambda)\cdot R(\theta_2, a_2) = -\lambda + 2.8$

The risk set is:

$S = \{(\lambda+0.8, -\lambda+2.8) : \lambda\in[0,1]\}$

We can plot the risk set:



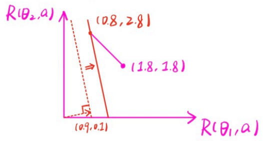
(b) The Bayes risks for a_1 and a_2 are:

$r(\Lambda, a_1) = E_\Lambda(R(\theta, a_1)) = 1.8\times 0.9 + 1.8\times 0.1 = 1.8$

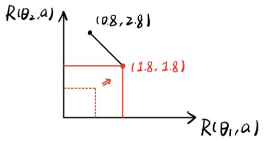
$r(\Lambda, a_2) = E_\Lambda(R(\theta, a_2)) = 0.8\times 0.9 + 2.8\times 0.1 = 1$

Since $r(\Lambda, a_2) < r(\Lambda, a_1)$, a_2 is the Bayes rule

Remark: Even if we consider the randomized actions, a_2 still is the Bayes rule based on the following plot:



(c) Based on the following plot, the minimax rule is a_1



Remark: we can also numerically derive the minimax rule:

Since for any $\lambda\in[0,1]$, $2.8-\lambda \geq \lambda + 0.8$

so the minimax rule would minimize $2.8-\lambda$, $\lambda\in[0,1]$ $\Rightarrow$ $\lambda = 1$ $\Rightarrow$ the minimax rule is a_1

Logic

若 parameter space 为 infinite, 则 minimax rule 一般较难求出, 此时一种求 minimax rule 的方法是利用 Bayes rule

2.5 Theorem: Minimax rule 的充分条件 (infinite parameter space)

若

- $\hat{\theta}$ 为某个 prior f 下的 Bayes estimator
- $\hat{\theta}$ 有 constant frequentist risk: $R(\theta, \hat{\theta}) = c$ (关于所有参数值 θ 为 constant)

则

- $\hat{\theta}$ 的 Bayes risk 等于 minimax risk, 即 $r(f, \hat{\theta}) = \inf_{\tilde{\theta}} \sup_{\theta} R(\theta, \tilde{\theta})$
- $\hat{\theta}$ 也为 minimax estimator

Remark: 借助 geometry of Bayes and Minimax point 的理解

假设我们顺着某个 prior vector 的方向移动 hyperplane, 并在和 risk set 相切的地方找到了一个 Bayes estimator, 且该 Bayes point 在各个轴的分量都相同 (frequentist risk 对所有参数值都是常数), 则显然这个点也是 Minimax point (如果我们逐步扩大正方形, 最先触碰到 risk set 的点也是这个点)

Proof

由于 $\hat{\theta}$ 有 constant frequentist risk $R(\theta, \hat{\theta}) = c$, 因此有:

$$\sup_{\theta}(R(\theta, \hat{\theta})) = \mathbb{E}_{\theta \sim f}[R(\theta, \hat{\theta})] = c$$

由于 $\hat{\theta}$ 为 Bayes rule, 因此其 Bayes risk 小于等于任意其他 estimator δ:

$$\mathbb{E}_{\theta \sim f}[R(\theta, \hat{\theta})] = r(f, \hat{\theta}) \leq r(f, \delta) = \mathbb{E}_{\theta \sim f}[R(\theta, \delta)]$$

而对于任意 estimator δ, 其 Bayes risk (平均而言的 risk) 不大于 maximum risk, 即:

$$\mathbb{E}_{\theta \sim f}[R(\theta, \delta)] \leq \mathbb{E}_{\theta \sim f}[\sup_{\theta} R(\theta, \delta)] = \sup_{\theta} R(\theta, \delta)$$

结合以上所有不等式, 即为:

$$\sup_{\theta}(R(\theta, \hat{\theta})) \leq \sup_{\theta} R(\theta, \delta), \quad \forall \delta$$

因此 $\hat{\theta}$ 为 minimax rule

Example

问题:
令:

- $X|p \sim \text{Bin}(n, p)$
- 使用 squared error loss

求:

1. 证明 $\hat{p} = X/n$ 不是 minimax
   Hint: 考虑 randomized estimator:

$$\tilde{p} = \begin{cases} X/n & \text{with probablity } 1 - \frac{1}{n+1} \\ 1/2 & \text{with probabilty } \frac{1}{n+1} \end{cases}$$

2. 考虑 prior distribution $p \sim \text{Beta}(a, b)$, 求出 a 和 b 使得 Bayes estimator 有 constant frequentist risk (因此该 estimator 为 minimax)

解答:

Question 1:

首先求出 $\hat{p}$ 的 (frequentist) risk:

$$\begin{aligned} R(p, \hat{p}) &= \mathbb{E}_X[(p - \hat{p})^2] \\ &= \mathbb{E}_X\left[\left(p - \frac{X}{n}\right)^2\right] \\ &= p^2 - 2p\mathbb{E}_X\left[\frac{X}{n}\right] + \mathbb{E}_X\left[\left(\frac{X}{n}\right)^2\right] \\ &= p^2 - 2p^2 + \frac{n^2p^2 + np(1-p)}{n^2} \\ &= \frac{p(1-p)}{n} \end{aligned}$$

因此 $\hat{p}$ 的 maximum risk 为:

$$\sup_{p} R(p, \hat{p}) = \sup_{p} \frac{p(1-p)}{n} = \frac{1}{4n}$$

接下来考虑 $\tilde{p}$ 的 (frequentist) risk:

$$
\begin{aligned}
R(p,\tilde{p}) &= \mathbb{E}_X[(p-\tilde{p})^2] \\
&= \mathbb{E}_X\left[\left(1-\frac{1}{n+1}\right)\left(p-\frac{X}{n}\right)^2+\frac{1}{n+1}\left(p-\frac{1}{2}\right)^2\right] \\
&= \frac{n}{n+1}\cdot\frac{p(1-p)}{n}+\frac{1}{n+1}\cdot\left(p^2-p+\frac{1}{4}\right) \\
&= \frac{1}{4(n+1)} < \frac{1}{4n}
\end{aligned}
$$

因此 $\hat{p}$ 不是 minimax

Question 2:

首先求出 p 的 posterior distribution:

$$
\begin{aligned}
f(p|X) &\propto f(X|p)\cdot f(p) \\
&\propto p^X(1-p)^{n-X}\cdot p^{a-1}(1-p)^{b-1} \\
&= p^{X+a-1}(1-p)^{n-X+b-1}
\end{aligned}
$$

注意到上式为 $\mathrm{Beta}(X+a, n-X+b)$ 的 kernel,
因此 Bayes estimator 即为 posterior mean:

$$
\bar{p} = \mathbb{E}_p[p|X] = \frac{X+a}{n+a+b}
$$

接下来求出该 Bayes estimator 的 frequentist risk:

$$
\begin{aligned}
R(p,\bar{p}) &= \mathbb{E}_X[(p-\bar{p})^2] \\
&= \mathbb{E}_X\left[\left(p-\frac{X+a}{n+a+b}\right)^2\right] \\
&= p^2-2p\cdot\frac{np+a}{n+a+b}+\frac{1}{(n+a+b)^2}np(1-p)+\left(\frac{np+a}{n+a+b}\right)^2 \\
&= \frac{((a+b)^2-n)p^2-(2a^2+2ab+n)p+a^2}{(a+b+n)^2}
\end{aligned}
$$

令其为 constant, 则有:

$$
\begin{cases}
(a+b)^2-n=0 \\
2a^2+2ab+n=0
\end{cases}
$$

解得:

$$
a=b=\frac{\sqrt{n}}{2}
$$