Foundations of Statistical Inference

J. Berestycki & D. Sejdinovic

Department of Statistics
University of Oxford

MT 2019

Chapter 8: Decision Theory

Framework

Parameter space: $\Theta \subset \mathbb{R}^d$.
Model: $X|\theta \sim f(x;\theta)$ for some parametric family $\{f(x;\theta), \theta \in \Theta\}$, taking values in $\mathcal{X}$.
Action (decision) space: $\mathcal{A}$. Typical examples involve estimating $g(\theta)$ ($\mathcal{A} = g(\Theta)$), or selecting a hypothesis ($\mathcal{A} = \{0, 1\}$).
Loss function: $L : \Theta \times \mathcal{A} \to \mathbb{R}_+$. If we take action $a \in \mathcal{A}$ when the true parameter is $\theta \in \Theta$ then we incur the loss $L(\theta, a)$.
Set of decision rules: $\mathcal{D} \subseteq \{\delta : \mathcal{X} \to \mathcal{A}\}$. Rule δ specifies an action for each possible observed $x \in \mathcal{X}$.
(Frequentist) Risk: For a given rule $\delta \in \mathcal{D}$ and parameter $\theta \in \Theta$:

$$R(\theta, \delta) = \mathbb{E}_{X|\theta}[L(\theta, \delta(X))] = \int_{\mathcal{X}} L(\theta, \delta(x)) f(x;\theta) dx.$$

Risk functions

- Risk $R(\theta, \delta)$ is the expected loss of a decision rule δ assuming that the true parameter is θ
- Note that the definition of risk involves hypothetical repetition of the sampling mechanism that generated x
- The postulate of decision theory is that decision rules are compared through their risk functions (as functions of θ)
- Fundamental principles for selecting among the decision rules are Minimax and Bayes principles.

Risk functions: examples

- *Estimation example*: $\delta(x)$ is an estimator of $\theta \in \mathbb{R}$, and we use $L(\theta, a) = \|a - \theta\|^2$, so that $R(\theta, \delta) = \mathbb{E}_{X|\theta}\|\delta(X) - \theta\|^2$
- *Testing example*: We are testing $\theta \in H_0$ vs $\theta \in H_1$. Action space is $\mathcal{A} = \{0, 1\}$ and the 0/1 loss is

$$L(\theta, a) = \begin{cases} 1, & \theta \in H_0, a = 1, \\ 1, & \theta \in H_1, a = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Here, the risk is simply the probability of making the wrong action

$$R(\theta, \delta) = \begin{cases} \mathbb{P}(\delta(X) = 1|\theta), & \text{if } \theta \in H_0, \\ \mathbb{P}(\delta(X) = 0|\theta), & \text{if } \theta \in H_1. \end{cases}$$

corresponding to the standard notions of Type I and Type II errors.

Admissibility

Definition

We say that δ_2 strictly dominates δ_1 if

$$R(\theta, \delta_1) \geq R(\theta, \delta_2), \text{ for all } \theta \in \Theta$$

with $R(\theta, \delta_1) > R(\theta, \delta_2)$ for at least some θ.

A procedure δ_1 is inadmissible if there exists another procedure δ_2 such that δ_2 strictly dominates δ_1.

A procedure which is not inadmissible is admissible.

Admissibility

Definition

We say that δ_2 strictly dominates δ_1 if

$$R(\theta, \delta_1) \geq R(\theta, \delta_2), \text{ for all } \theta \in \Theta$$

with $R(\theta, \delta_1) > R(\theta, \delta_2)$ for at least some θ.

A procedure δ_1 is inadmissible if there exists another procedure δ_2 such that δ_2 strictly dominates δ_1.

A procedure which is not inadmissible is admissible.

Admissibility

Definition

We say that δ_2 strictly dominates δ_1 if

$$R(\theta, \delta_1) \geq R(\theta, \delta_2), \text{ for all } \theta \in \Theta$$

with $R(\theta, \delta_1) > R(\theta, \delta_2)$ for at least some θ.

A procedure δ_1 is inadmissible if there exists another procedure δ_2 such that δ_2 strictly dominates δ_1.

A procedure which is not inadmissible is admissible.

Example

Suppose $X \sim U(0, \theta)$. Consider estimators of the form $\widehat{\theta}(x) = ax$ (this is a family of decision rules indexed by a).

Show that $a = 3/2$ is a necessary condition for the rule $\widehat{\theta}$ to be admissible for quadratic loss.

$$\begin{aligned} R(\theta, \widehat{\theta}) &= \int_0^\theta (ax - \theta)^2 \frac{1}{\theta} dx \\ &= (a^2/3 - a + 1)\theta^2 \end{aligned}$$

and R is minimized at $a = 3/2$.

This does not show $\widehat{\theta}(x) = 3x/2$ is admissible here.

It only shows that all estimators with $a \neq 3/2$ are inadmissible. The estimator $\widehat{\theta}(x) = 3x/2$ may still be inadmissible relative to other estimators not in this class!

Example

Suppose $X \sim U(0, \theta)$. Consider estimators of the form $\widehat{\theta}(x) = ax$ (this is a family of decision rules indexed by a).

Show that $a = 3/2$ is a necessary condition for the rule $\widehat{\theta}$ to be admissible for quadratic loss.

$$\begin{aligned} R(\theta, \widehat{\theta}) &= \int_0^\theta (ax - \theta)^2 \frac{1}{\theta} dx \\ &= (a^2/3 - a + 1)\theta^2 \end{aligned}$$

and R is minimized at $a = 3/2$.

This does not show $\widehat{\theta}(x) = 3x/2$ is admissible here.

It only shows that all estimators with $a \neq 3/2$ are inadmissible. The estimator $\widehat{\theta}(x) = 3x/2$ may still be inadmissible relative to other estimators not in this class!

Example

Suppose $X \sim U(0, \theta)$. Consider estimators of the form $\widehat{\theta}(x) = ax$ (this is a family of decision rules indexed by a).

Show that $a = 3/2$ is a necessary condition for the rule $\widehat{\theta}$ to be admissible for quadratic loss.

$$
\begin{aligned}
R(\theta, \widehat{\theta}) &= \int_0^\theta (ax - \theta)^2 \frac{1}{\theta} dx \\
&= (a^2/3 - a + 1)\theta^2
\end{aligned}
$$

and R is minimized at $a = 3/2$.

This does not show $\widehat{\theta}(x) = 3x/2$ is admissible here.

It only shows that all estimators with $a \neq 3/2$ are inadmissible. The estimator $\widehat{\theta}(x) = 3x/2$ may still be inadmissible relative to other estimators not in this class!

Example

Suppose $X \sim U(0, \theta)$. Consider estimators of the form $\widehat{\theta}(x) = ax$ (this is a family of decision rules indexed by a).

Show that $a = 3/2$ is a necessary condition for the rule $\widehat{\theta}$ to be admissible for quadratic loss.

$$\begin{aligned} R(\theta, \widehat{\theta}) &= \int_0^\theta (ax - \theta)^2 \frac{1}{\theta} dx \\ &= (a^2/3 - a + 1)\theta^2 \end{aligned}$$

and R is minimized at $a = 3/2$.

This does not show $\widehat{\theta}(x) = 3x/2$ is admissible here.

It only shows that all estimators with $a \neq 3/2$ are inadmissible. The estimator $\widehat{\theta}(x) = 3x/2$ may still be inadmissible relative to other estimators not in this class!

Example

Suppose $X \sim U(0, \theta)$. Consider estimators of the form $\widehat{\theta}(x) = ax$ (this is a family of decision rules indexed by a).

Show that $a = 3/2$ is a necessary condition for the rule $\widehat{\theta}$ to be admissible for quadratic loss.

$$\begin{aligned} R(\theta, \widehat{\theta}) &= \int_0^\theta (ax - \theta)^2 \frac{1}{\theta} dx \\ &= (a^2/3 - a + 1)\theta^2 \end{aligned}$$

and R is minimized at $a = 3/2$.

This does not show $\widehat{\theta}(x) = 3x/2$ is admissible here.

It only shows that all estimators with $a \neq 3/2$ are inadmissible. The estimator $\widehat{\theta}(x) = 3x/2$ may still be inadmissible relative to other estimators not in this class!

Comments on admissibility

- It is a weak requirement (defined as an absence of negative attribute rather than a posession of positive one).
- We will see later in the course that some natural looking estimators are inadmissible (Stein phenomenon).

Minimax rules

Definition

A rule δ is a minimax rule if $\sup_\theta R(\theta, \delta) \leq \sup_\theta R(\theta, \delta')$ for any other rule δ'. It minimizes the maximum risk.

$$\delta^* = \arg\min_{\delta \in \mathcal{D}} \sup_{\theta \in \Theta} R(\theta, \delta).$$

Motivation: we do not know anything about the true θ, so we insure ourselves against the worst possible case.

It makes sense when the worst case scenario must be avoided, but can lead to poor performance on average.

Defines an order on decision rules, using a conservative point of view.

Minimax rules

Definition

A rule δ is a minimax rule if $\sup_\theta R(\theta, \delta) \leq \sup_\theta R(\theta, \delta')$ for any other rule δ'. It minimizes the maximum risk.

$$\delta^* = \arg\min_{\delta \in \mathcal{D}} \sup_{\theta \in \Theta} R(\theta, \delta).$$

Motivation: we do not know anything about the true θ, so we insure ourselves against the worst possible case.
It makes sense when the worst case scenario must be avoided, but can lead to poor performance on average.
Defines an order on decision rules, using a conservative point of view.

Minimax rules

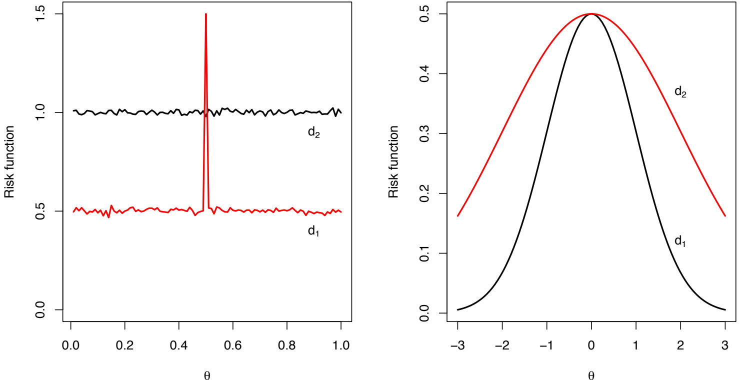
Definition

A rule δ is a minimax rule if $\sup_\theta R(\theta, \delta) \leq \sup_\theta R(\theta, \delta')$ for any other rule δ'. It minimizes the maximum risk.

$$\delta^* = \arg\min_{\delta \in \mathcal{D}} \sup_{\theta \in \Theta} R(\theta, \delta).$$

Motivation: we do not know anything about the true θ, so we insure ourselves against the worst possible case.
It makes sense when the worst case scenario must be avoided, but can lead to poor performance on average.
Defines an order on decision rules, using a conservative point of view.

Sometimes minimax does not produce a sensible choice of decision rule.



Bayes rules

Specify a prior $\pi(\theta)$ and introduce the Bayes (integrated) risk:

$$r(\pi, \delta) = \int_{\Theta} R(\theta, \delta)\pi(\theta)d\theta.$$

A decision rule δ is said to be a Bayes rule wrt π if it minimizes the Bayes risk:

$$r(\pi, \delta) = \inf_{\delta' \in \mathcal{D}} r(\pi, \delta') =: m_\pi$$

If the infimum is not attained, we can consider $\epsilon > 0$ and δ_ϵ such that $r(\pi, \delta) < m_\pi + \epsilon$. In this case δ_ϵ is said to be ϵ-Bayes wrt π.
A rule δ is said to be extended Bayes if $\forall \epsilon > 0$ there exists some π such that δ is ϵ-Bayes wrt π.

Are Bayes rules admissible ?

Definition (π-admissibility)

A procedure δ^* is said to be π-admissible iff for all other procedure δ, such that $R(\theta, \delta) \leq R(\theta, \delta^*)$ for all θ,

$$\pi\left(A_\delta\right) = 0,$$

where $A_\delta := \{\theta \ : \ R(\theta, \delta) < R(\theta, \delta^*)\}$.

Theorem

The rule which is Bayes wrt π is π-admissible.

Proof

If Bayes rule δ^π is not π-admissible then $\exists\delta$, s.t. $\pi\left(A_\delta\right) > 0$. Then

$$
\begin{aligned}
r(\pi,\delta) - r(\pi,\delta^\pi) &= \int_{A_\delta} [R(\theta,\delta) - R(\theta,\delta^\pi)]\pi(\theta)d\theta \\
&\quad + \int_{A_\delta^c} [R(\theta,\delta) - R(\theta,\delta^\pi)]\pi(\theta)d\theta \\
&\leq \int_{A_\delta} \underbrace{[R(\theta,\delta) - R(\theta,\delta^\pi)]}_{<0} \pi(\theta)d\theta < 0
\end{aligned}
$$

which contradicts δ^π being Bayes.

From the proof we see that Bayes rules are *easily* admissible.

For instance

1. If δ^π is unique almost surely and $r(\pi, \delta^\pi) < +\infty$ then it is admissible
2. If $\forall\delta$, $\theta \to R(\theta, \delta)$ is continuous, $r(\pi, \delta^\pi) < +\infty$ and π has a positive density wrt Lebesgue measure then δ^π is admissible.

From the proof we see that Bayes rules are *easily* admissible.
For instance

1. If δ^π is unique almost surely and $r(\pi, \delta^\pi) < +\infty$ then it is admissible
2. If $\forall \delta$, $\theta \to R(\theta, \delta)$ is continuous, $r(\pi, \delta^\pi) < +\infty$ and π has a positive density wrt Lebesgue measure then δ^π is admissible.

From the proof we see that Bayes rules are *easily* admissible.
For instance

1. If δ^π is unique almost surely and $r(\pi, \delta^\pi) < +\infty$ then it is admissible
2. If $\forall\delta$, $\theta \to R(\theta, \delta)$ is continuous, $r(\pi, \delta^\pi) < +\infty$ and π has a positive density wrt Lebesgue measure then δ^π is admissible.

Proofs

Proof of (1) If it is not admissible then $\exists \delta$ s.t. $R(\theta, \delta) \leq R(\theta, \delta^\pi)$ for all θ. This implies that

$$r(\pi, \delta) \leq r(\pi, \delta^\pi) \quad \Rightarrow \delta = \delta^\pi \quad \text{a.s.}$$

Proof of (2) If not admissible, then $\exists \delta$ s.t. $R(\theta, \delta) \leq R(\theta, \delta^\pi)$ for all θ and $A_\delta \neq \emptyset$. Since $\theta \to R(\theta, \delta) - R(\theta, \delta^\pi)$ is continuous then A_δ contains an open ($\neq \emptyset$) set and $\pi(A_\delta) > 0$, which is impossible.

Randomized decision rules

Suppose we have a collection of l decision rules $d_1, \ldots, d_l$. For probability weights $p_1, \ldots, p_l$ define d^* to be the rule *'select rule d_i with probability p_i and apply'*.

Definition

d^* is a randomized decision rule.

The risk function of a randomized decision rule is then

$$R(\theta, d^*) = \sum_{i=1}^{l} p_i R(\theta, d_i).$$

Minimax rules may be of this form.
Bayes rules are not randomized (if unique).

Randomized decision rules

Suppose we have a collection of l decision rules $d_1, \ldots, d_l$.For probability weights $p_1, \ldots, p_l$ define d^* to be the rule *'select rule d_i with probability p_i and apply'*.

Definition

d^* is a randomized decision rule.

The risk function of a randomized decision rule is then

$$R(\theta, d^*) = \sum_{i=1}^{l} p_i R(\theta, d_i).$$

Minimax rules may be of this form.
Bayes rules are not randomized (if unique).

Randomized decision rules

Suppose we have a collection of l decision rules $d_1, \ldots, d_l$.For probability weights $p_1, \ldots, p_l$ define d^* to be the rule *'select rule d_i with probability p_i and apply'*.

Definition

d^* is a randomized decision rule.

The risk function of a randomized decision rule is then

$$R(\theta, d^*) = \sum_{i=1}^{l} p_i R(\theta, d_i).$$

Minimax rules may be of this form.

Bayes rules are not randomized (if unique).

Randomized decision rules

Suppose we have a collection of l decision rules $d_1, \ldots, d_l$.For probability weights $p_1, \ldots, p_l$ define d^* to be the rule *'select rule d_i with probability p_i and apply'*.

Definition

d^* is a randomized decision rule.

The risk function of a randomized decision rule is then

$$R(\theta, d^*) = \sum_{i=1}^{l} p_i R(\theta, d_i).$$

Minimax rules may be of this form.
Bayes rules are not randomized (if unique).

Finite decision problem

Definition

A decision problem is said to be finite when the parameter space $\Theta = \{\theta_1, \dots, \theta_k\}$ is finite.

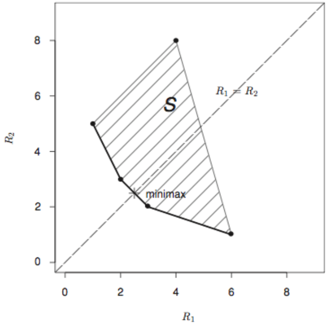
In this case the notions of admissible, minimax and Bayes can be given geometric interpretations.

Definition

The risk set $S \subset \mathbb{R}^k$ is the set of points $(R(\theta_1, d), \dots R(\theta_k, d))$ for some decision rule d.

Lemma

S is a convex set.



Finite decision problem

Definition

A decision problem is said to be finite when the parameter space $\Theta = \{\theta_1, \ldots, \theta_k\}$ is finite.

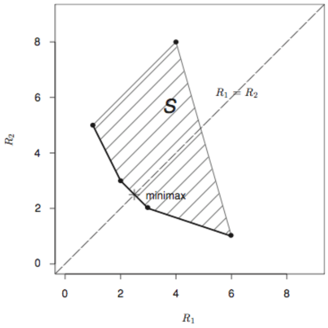
In this case the notions of admissible, minimax and Bayes can be given geometric interpretations.

Definition

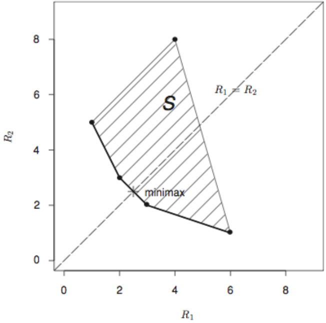
The risk set $S \subset \mathbb{R}^k$ is the set of points $(R(\theta_1, d), \ldots R(\theta_k, d))$ for some decision rule d.

Lemma

S is a convex set.

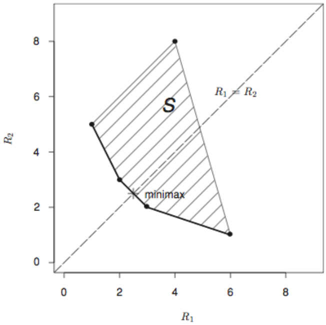


Finite decision problem



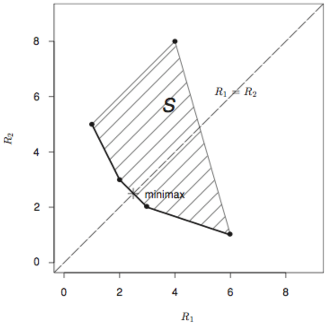
1. Extreme points = non-randomized rules.
2. Lower thick line= admissible rules.
3. Minimax is intersection with the square $\max(R_1, R_2) = c$. In this case with line $R_1 = R_2$.

Finite decision problem



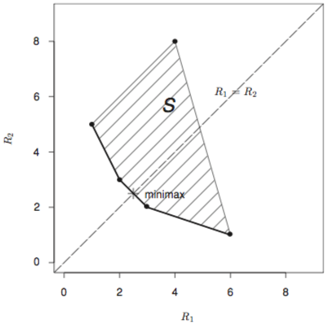
1. Extreme points = non-randomized rules.
2. Lower thick line= admissible rules.
3. Minimax is intersection with the square $\max(R_1, R_2) = c$. In this case with line $R_1 = R_2$.

Finite decision problem



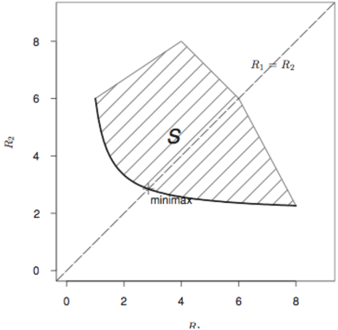
1. Extreme points = non-randomized rules.
2. Lower thick line= admissible rules.
3. Minimax is intersection with the square $\max(R_1, R_2) = c$. In this case with line $R_1 = R_2$.

Finite decision problem



1. Extreme points = non-randomized rules.
2. Lower thick line= admissible rules.
3. Minimax is intersection with the square $\max(R_1, R_2) = c$. In this case with line $R_1 = R_2$ (random) .

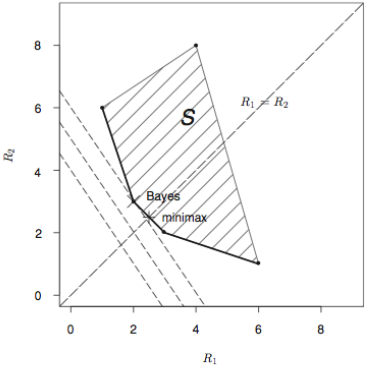
Finite decision problem



1. Extreme points = non-randomized rules.
2. Lower thick line= admissible rules.
3. Minimax is intersection with the square $\max(R_1, R_2) = c$. In this case with line $R_1 = R_2$ (non-random).

Finite decision problem

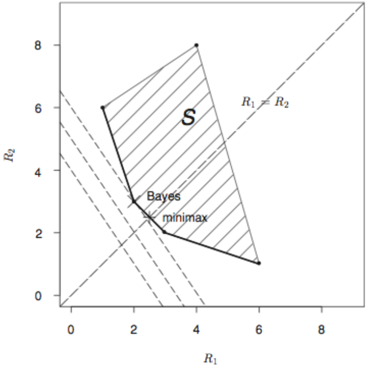
To find the Bayes rule; suppose prior is (π_1, π_2). For any c the line $\pi_1 R_1 + \pi_2 R_2 = c$ represents a class of decision rules with same Bayes risk c.



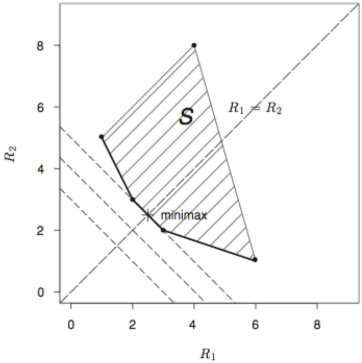
The Bayes rules is unique and therefore non-random

Finite decision problem

To find the Bayes rule; suppose prior is (π_1, π_2). For any c the line $\pi_1 R_1 + \pi_2 R_2 = c$ represents a class of decision rules with same Bayes risk c.



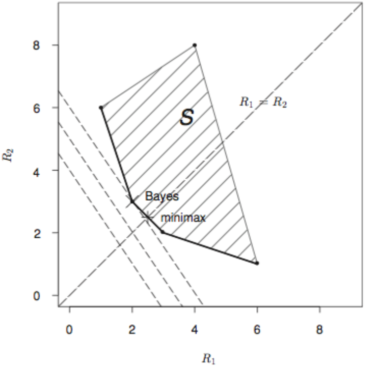
The Bayes rules is unique and therefore non-random



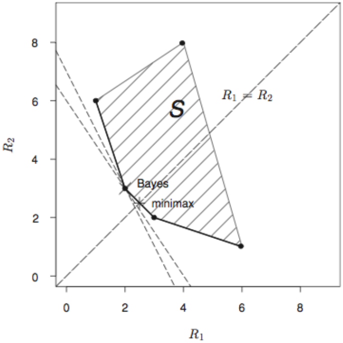
Bayes rule is not unique but can be chosen non-random.

Finite decision problem

To find the Bayes rule; suppose prior is (π_1, π_2). For any c the line $\pi_1 R_1 + \pi_2 R_2 = c$ represents a class of decision rules with same Bayes risk c.



The Bayes rules is unique and therefore non-random



How the prior influences the Bayes rule.

An important inequality

Theorem

Let $\mathcal{P}$ the set of all probability measures on Θ. Then

$$\sup_{\pi \in \mathcal{P}} r(\pi) \leq \inf_{\delta \in \mathcal{D}} \sup_{\theta \in \Theta} R(\theta, \delta) := \bar{R}$$

with $r(\pi) = r(\pi, \delta^\pi)$ and δ^π is the associated Bayes rule.

Proof Let $\pi \in \mathcal{P}$ then for all rules δ:

$$r(\pi, \delta^\pi) \leq r(\pi, \delta) = \int_\Theta R(\theta, \delta)\pi(\theta)d\theta \leq \sup_\Theta R(\theta, \delta).$$

Hence

$$r(\pi) \leq \bar{R} \quad \forall \pi \in \mathcal{P}.$$

Finding minimax rules that are Bayes

Theorem 1 If δ is a Bayes rule for prior π, with $r(\pi, \delta) = C$, and δ_0 is a rule for which $\max_\theta R(\theta, \delta_0) = C$, then δ_0 is minimax.

Proof If for some other rule δ', $\max_\theta R(\theta, \delta') = C - \epsilon$ for some $\epsilon > 0$ (so δ_0 is not minimax), then

$$\begin{aligned} r(\pi, \delta') &= \int R(\theta, \delta')\pi(\theta)d\theta \\ &\leq \int (C - \epsilon)\pi(\theta)d\theta \\ &= (C - \epsilon) \\ &< r(\pi, \delta) \end{aligned}$$

so δ is not the Bayes rule for π, a contradiction. [*This is an informal treatment which assumes the min and max exist - see Y&S Ch 2 Sec 2.6*]

Finding minimax rules that are Bayes

Theorem 1 If δ is a Bayes rule for prior π, with $r(\pi, \delta) = C$, and δ_0 is a rule for which $\max_\theta R(\theta, \delta_0) = C$, then δ_0 is minimax.

Proof If for some other rule δ', $\max_\theta R(\theta, \delta') = C - \epsilon$ for some $\epsilon > 0$ (so δ_0 is not minimax), then

$$
\begin{aligned}
r(\pi, \delta') &= \int R(\theta, \delta')\pi(\theta)d\theta \\
&\leq \int (C - \epsilon)\pi(\theta)d\theta \\
&= (C - \epsilon) \\
&< r(\pi, \delta)
\end{aligned}
$$

so δ is not the Bayes rule for π, a contradiction. [*This is an informal treatment which assumes the min and max exist - see Y&S Ch 2 Sec 2.6*]

Finding minimax rules that are Bayes

Theorem 1 If δ is a Bayes rule for prior π, with $r(\pi, \delta) = C$, and δ_0 is a rule for which $\max_\theta R(\theta, \delta_0) = C$, then δ_0 is minimax.

Proof If for some other rule δ', $\max_\theta R(\theta, \delta') = C - \epsilon$ for some $\epsilon > 0$ (so δ_0 is not minimax), then

$$\begin{aligned} r(\pi, \delta') &= \int R(\theta, \delta')\pi(\theta) d\theta \\ &\leq \int (C - \epsilon)\pi(\theta) d\theta \\ &= (C - \epsilon) \\ &< r(\pi, \delta) \end{aligned}$$

so δ is not the Bayes rule for π, a contradiction. [*This is an informal treatment which assumes the min and max exist - see Y&S Ch 2 Sec 2.6*]

Finding minimax rules that are Bayes

Theorem 1 If δ is a Bayes rule for prior π, with $r(\pi, \delta) = C$, and δ_0 is a rule for which $\max_\theta R(\theta, \delta_0) = C$, then δ_0 is minimax.

Proof If for some other rule δ', $\max_\theta R(\theta, \delta') = C - \epsilon$ for some $\epsilon > 0$ (so δ_0 is not minimax), then

$$\begin{aligned} r(\pi, \delta') &= \int R(\theta, \delta')\pi(\theta)d\theta \\ &\leq \int (C - \epsilon)\pi(\theta)d\theta \\ &= (C - \epsilon) \\ &< r(\pi, \delta) \end{aligned}$$

so δ is not the Bayes rule for π, a contradiction. [*This is an informal treatment which assumes the min and max exist - see Y&S Ch 2 Sec 2.6*]

Finding minimax rules that are Bayes

Theorem 1 If δ is a Bayes rule for prior π, with $r(\pi, \delta) = C$, and δ_0 is a rule for which $\max_\theta R(\theta, \delta_0) = C$, then δ_0 is minimax.

Proof If for some other rule δ', $\max_\theta R(\theta, \delta') = C - \epsilon$ for some $\epsilon > 0$ (so δ_0 is not minimax), then

$$\begin{aligned} r(\pi, \delta') &= \int R(\theta, \delta')\pi(\theta)d\theta \\ &\leq \int (C - \epsilon)\pi(\theta)d\theta \\ &= (C - \epsilon) \\ &< r(\pi, \delta) \end{aligned}$$

so δ is not the Bayes rule for π, a contradiction. [*This is an informal treatment which assumes the min and max exist - see Y&S Ch 2 Sec 2.6*]

Finding minimax rules that are Bayes

Theorem 1 If δ is a Bayes rule for prior π, with $r(\pi, \delta) = C$, and δ_0 is a rule for which $\max_\theta R(\theta, \delta_0) = C$, then δ_0 is minimax.

Proof If for some other rule δ', $\max_\theta R(\theta, \delta') = C - \epsilon$ for some $\epsilon > 0$ (so δ_0 is not minimax), then

$$\begin{aligned} r(\pi, \delta') &= \int R(\theta, \delta')\pi(\theta)d\theta \\ &\leq \int (C - \epsilon)\pi(\theta)d\theta \\ &= (C - \epsilon) \\ &< r(\pi, \delta) \end{aligned}$$

so δ is not the Bayes rule for π, a contradiction. [*This is an informal treatment which assumes the min and max exist - see Y&S Ch 2 Sec 2.6*]

Finding minimax rules that are Bayes

Theorem 1 If δ is a Bayes rule for prior π, with $r(\pi, \delta) = C$, and δ_0 is a rule for which $\max_\theta R(\theta, \delta_0) = C$, then δ_0 is minimax.

Proof If for some other rule δ', $\max_\theta R(\theta, \delta') = C - \epsilon$ for some $\epsilon > 0$ (so δ_0 is not minimax), then

$$\begin{aligned} r(\pi, \delta') &= \int R(\theta, \delta')\pi(\theta)d\theta \\ &\leq \int (C - \epsilon)\pi(\theta)d\theta \\ &= (C - \epsilon) \\ &< r(\pi, \delta) \end{aligned}$$

so δ is not the Bayes rule for π, a contradiction. [*This is an informal treatment which assumes the min and max exist - see Y&S Ch 2 Sec 2.6*]

Finding minimax rules

Theorem 2 If δ is a Bayes rule for prior π with the property that $R(\theta, \delta)$ does not depend on θ, then δ is minimax.

Proof (Y&S Ch 2) Let $R(\theta, \delta) = C$ (no θ dependence). This implies

$$r(\pi, \delta) = \int R(\theta, \delta)\pi(\theta)d\theta = C.$$

If for some other rule δ', $\max_\theta R(\theta, \delta') = C - \epsilon$ for some $\epsilon > 0$ (so δ_0 is not minimax), then we have $r(\pi, \delta') \leq C - \epsilon$. But $r(\pi, \delta) = C$ so δ is not the Bayes rule for π, a contradiction.

Finding minimax rules

Theorem 2 If δ is a Bayes rule for prior π with the property that $R(\theta, \delta)$ does not depend on θ, then δ is minimax.

Proof (Y&S Ch 2) Let $R(\theta, \delta) = C$ (no θ dependence). This implies

$$r(\pi, \delta) = \int R(\theta, \delta)\pi(\theta)d\theta = C.$$

If for some other rule δ', $\max_\theta R(\theta, \delta') = C - \epsilon$ for some $\epsilon > 0$ (so δ_0 is not minimax), then we have $r(\pi, \delta') \leq C - \epsilon$. But $r(\pi, \delta) = C$ so δ is not the Bayes rule for π, a contradiction.

Finding minimax rules

Theorem 2 If δ is a Bayes rule for prior π with the property that $R(\theta, \delta)$ does not depend on θ, then δ is minimax.

Proof (Y&S Ch 2) Let $R(\theta, \delta) = C$ (no θ dependence). This implies

$$r(\pi, \delta) = \int R(\theta, \delta)\pi(\theta)d\theta = C.$$

If for some other rule δ', $\max_\theta R(\theta, \delta') = C - \epsilon$ for some $\epsilon > 0$ (so δ_0 is not minimax), then we have $r(\pi, \delta') \leq C - \epsilon$. But $r(\pi, \delta) = C$ so δ is not the Bayes rule for π, a contradiction.

Finding minimax rules

Theorem 2 If δ is a Bayes rule for prior π with the property that $R(\theta, \delta)$ does not depend on θ, then δ is minimax.

Proof (Y&S Ch 2) Let $R(\theta, \delta) = C$ (no θ dependence). This implies

$$r(\pi, \delta) = \int R(\theta, \delta)\pi(\theta)d\theta = C.$$

If for some other rule δ', $\max_\theta R(\theta, \delta') = C - \epsilon$ for some $\epsilon > 0$ (so δ_0 is not minimax), then we have $r(\pi, \delta') \leq C - \epsilon$. But $r(\pi, \delta) = C$ so δ is not the Bayes rule for π, a contradiction.

Finding minimax rules

Theorem 2 tells us that

The Bayes estimator with constant risk is minimax.

This result is useful, as it gives an approach to finding minimax rules.

Bayes rules are sometimes easy to compute, so if we find a prior that yields a Bayes rule with constant risk for all θ we have the minimax rule.

Finding minimax rules

Theorem 2 tells us that

The Bayes estimator with constant risk is minimax.

This result is useful, as it gives an approach to finding minimax rules.

Bayes rules are sometimes easy to compute, so if we find a prior that yields a Bayes rule with constant risk for all θ we have the minimax rule.

Finding minimax rules

Theorem 2 tells us that

The Bayes estimator with constant risk is minimax.

This result is useful, as it gives an approach to finding minimax rules.

Bayes rules are sometimes easy to compute, so if we find a prior that yields a Bayes rule with constant risk for all θ we have the minimax rule.

Finding minimax rules

Theorem 2 tells us that

The Bayes estimator with constant risk is minimax.

This result is useful, as it gives an approach to finding minimax rules.

Bayes rules are sometimes easy to compute, so if we find a prior that yields a Bayes rule with constant risk for all θ we have the minimax rule.

Example : minimax estimator for quadratic loss

X is Binomial (n, θ), and the prior $\pi(\theta)$ is a Beta (α, β) distribution. For a quadratic loss function, the Bayes estimator is $(\alpha + X)/(\alpha + \beta + n)$

The risk function is

$$
\begin{aligned}
\mathbb{E}_{X|\theta}[(\widehat{\theta} - \theta)^2] &= \mathsf{MSE}(\widehat{\theta}) = [\mathsf{Bias}(\widehat{\theta})]^2 + \mathsf{Var}[\widehat{\theta}] \\
&= \left[\theta - \mathbb{E}\left(\frac{\alpha + X}{\alpha + \beta + n}\right)\right]^2 + \mathsf{Var}\left[\frac{\alpha + X}{\alpha + \beta + n}\right] \\
&= \left[\theta - \left(\frac{\alpha + n\theta}{\alpha + \beta + n}\right)\right]^2 + \frac{n\theta(1-\theta)}{(\alpha + \beta + n)^2} \\
&= \frac{[\theta(\alpha + \beta) - \alpha]^2 + n\theta(1-\theta)}{[\alpha + \beta + n]^2}
\end{aligned}
$$

The Bayes estimator with constant risk is minimax. This occurs when $\alpha = \beta = \sqrt{n}/2$, so the minimax estimator using quadratic loss is $(\alpha + x)/(\alpha + \beta + n) = (x + \sqrt{n}/2)/(n + \sqrt{n})$. This estimator is also admissible

Example : minimax estimator for quadratic loss

X is Binomial (n, θ), and the prior $\pi(\theta)$ is a Beta (α, β) distribution. For a quadratic loss function, the Bayes estimator is $(\alpha + X)/(\alpha + \beta + n)$

The risk function is

$$\begin{aligned}
\mathbb{E}_{X|\theta}[(\widehat{\theta} - \theta)^2] &= \text{MSE}(\widehat{\theta}) = [\text{Bias}(\widehat{\theta})]^2 + \text{Var}[\widehat{\theta}] \\
&= \left[\theta - \mathbb{E}\left(\frac{\alpha + X}{\alpha + \beta + n}\right)\right]^2 + \text{Var}\left[\frac{\alpha + X}{\alpha + \beta + n}\right] \\
&= \left[\theta - \left(\frac{\alpha + n\theta}{\alpha + \beta + n}\right)\right]^2 + \frac{n\theta(1-\theta)}{(\alpha + \beta + n)^2} \\
&= \frac{[\theta(\alpha + \beta) - \alpha]^2 + n\theta(1-\theta)}{[\alpha + \beta + n]^2}
\end{aligned}$$

The Bayes estimator with constant risk is minimax. This occurs when $\alpha = \beta = \sqrt{n}/2$, so the minimax estimator using quadratic loss is $(\alpha + x)/(\alpha + \beta + n) = (x + \sqrt{n}/2)/(n + \sqrt{n})$. This estimator is also admissible

Example : minimax estimator for quadratic loss

X is Binomial (n, θ), and the prior $\pi(\theta)$ is a Beta (α, β) distribution. For a quadratic loss function, the Bayes estimator is $(\alpha + X)/(\alpha + \beta + n)$

The risk function is

$$
\begin{aligned}
\mathbb{E}_{X|\theta}[(\widehat{\theta} - \theta)^2] &= \mathsf{MSE}(\widehat{\theta}) = [\mathsf{Bias}(\widehat{\theta})]^2 + \mathsf{Var}[\widehat{\theta}] \\
&= \left[\theta - \mathbb{E}\left(\frac{\alpha + X}{\alpha + \beta + n}\right)\right]^2 + \mathsf{Var}\left[\frac{\alpha + X}{\alpha + \beta + n}\right] \\
&= \left[\theta - \left(\frac{\alpha + n\theta}{\alpha + \beta + n}\right)\right]^2 + \frac{n\theta(1-\theta)}{(\alpha + \beta + n)^2} \\
&= \frac{[\theta(\alpha + \beta) - \alpha]^2 + n\theta(1-\theta)}{[\alpha + \beta + n]^2}
\end{aligned}
$$

The Bayes estimator with constant risk is minimax. This occurs when $\alpha = \beta = \sqrt{n}/2$, so the minimax estimator using quadratic loss is $(\alpha + x)/(\alpha + \beta + n) = (x + \sqrt{n}/2)/(n + \sqrt{n})$. This estimator is also admissible

Example : minimax estimator for quadratic loss

X is Binomial (n, θ), and the prior $\pi(\theta)$ is a Beta (α, β) distribution. For a quadratic loss function, the Bayes estimator is $(\alpha + X)/(\alpha + \beta + n)$

The risk function is

$$
\begin{aligned}
\mathbb{E}_{X|\theta}[(\widehat{\theta} - \theta)^2] &= \mathsf{MSE}(\widehat{\theta}) = [\mathsf{Bias}(\widehat{\theta})]^2 + \mathsf{Var}[\widehat{\theta}] \\
&= \left[\theta - \mathbb{E}\left(\frac{\alpha + X}{\alpha + \beta + n}\right)\right]^2 + \mathsf{Var}\left[\frac{\alpha + X}{\alpha + \beta + n}\right] \\
&= \left[\theta - \left(\frac{\alpha + n\theta}{\alpha + \beta + n}\right)\right]^2 + \frac{n\theta(1-\theta)}{(\alpha + \beta + n)^2} \\
&= \frac{[\theta(\alpha + \beta) - \alpha]^2 + n\theta(1-\theta)}{[\alpha + \beta + n]^2}
\end{aligned}
$$

The Bayes estimator with constant risk is minimax. This occurs when $\alpha = \beta = \sqrt{n}/2$, so the minimax estimator using quadratic loss is $(\alpha + x)/(\alpha + \beta + n) = (x + \sqrt{n}/2)/(n + \sqrt{n})$. This estimator is also admissible

Example : minimax estimator for quadratic loss

X is Binomial (n, θ), and the prior $\pi(\theta)$ is a Beta (α, β) distribution. For a quadratic loss function, the Bayes estimator is $(\alpha + X)/(\alpha + \beta + n)$

The risk function is

$$
\begin{aligned}
\mathbb{E}_{X|\theta}[(\widehat{\theta} - \theta)^2] &= \mathsf{MSE}(\widehat{\theta}) = [\mathsf{Bias}(\widehat{\theta})]^2 + \mathsf{Var}[\widehat{\theta}] \\
&= \left[\theta - \mathbb{E}\left(\frac{\alpha + X}{\alpha + \beta + n}\right)\right]^2 + \mathsf{Var}\left[\frac{\alpha + X}{\alpha + \beta + n}\right] \\
&= \left[\theta - \left(\frac{\alpha + n\theta}{\alpha + \beta + n}\right)\right]^2 + \frac{n\theta(1-\theta)}{(\alpha + \beta + n)^2} \\
&= \frac{[\theta(\alpha + \beta) - \alpha]^2 + n\theta(1-\theta)}{[\alpha + \beta + n]^2}
\end{aligned}
$$

The Bayes estimator with constant risk is minimax. This occurs when $\alpha = \beta = \sqrt{n}/2$, so the minimax estimator using quadratic loss is $(\alpha + x)/(\alpha + \beta + n) = (x + \sqrt{n}/2)/(n + \sqrt{n})$. This estimator is also admissible

Example : minimax estimator for quadratic loss

X is Binomial (n, θ), and the prior $\pi(\theta)$ is a Beta (α, β) distribution. For a quadratic loss function, the Bayes estimator is $(\alpha + X)/(\alpha + \beta + n)$

The risk function is

$$\begin{aligned}
\mathbb{E}_{X|\theta}[(\widehat{\theta} - \theta)^2] &= \mathsf{MSE}(\widehat{\theta}) = [\mathsf{Bias}(\widehat{\theta})]^2 + \mathsf{Var}[\widehat{\theta}] \\
&= \left[\theta - \mathbb{E}\left(\frac{\alpha + X}{\alpha + \beta + n}\right)\right]^2 + \mathsf{Var}\left[\frac{\alpha + X}{\alpha + \beta + n}\right] \\
&= \left[\theta - \left(\frac{\alpha + n\theta}{\alpha + \beta + n}\right)\right]^2 + \frac{n\theta(1-\theta)}{(\alpha + \beta + n)^2} \\
&= \frac{[\theta(\alpha + \beta) - \alpha]^2 + n\theta(1-\theta)}{[\alpha + \beta + n]^2}
\end{aligned}$$

The Bayes estimator with constant risk is minimax. This occurs when $\alpha = \beta = \sqrt{n}/2$, so the minimax estimator using quadratic loss is $(\alpha + x)/(\alpha + \beta + n) = (x + \sqrt{n}/2)/(n + \sqrt{n})$. This estimator is also admissible

Example : minimax estimator for quadratic loss

X is Binomial (n, θ), and the prior $\pi(\theta)$ is a Beta (α, β) distribution. For a quadratic loss function, the Bayes estimator is $(\alpha + X)/(\alpha + \beta + n)$

The risk function is

$$\begin{aligned}
\mathbb{E}_{X|\theta}[(\widehat{\theta} - \theta)^2] &= \mathsf{MSE}(\widehat{\theta}) = [\mathsf{Bias}(\widehat{\theta})]^2 + \mathsf{Var}[\widehat{\theta}] \\
&= \left[\theta - \mathbb{E}\left(\frac{\alpha + X}{\alpha + \beta + n}\right)\right]^2 + \mathsf{Var}\left[\frac{\alpha + X}{\alpha + \beta + n}\right] \\
&= \left[\theta - \left(\frac{\alpha + n\theta}{\alpha + \beta + n}\right)\right]^2 + \frac{n\theta(1-\theta)}{(\alpha + \beta + n)^2} \\
&= \frac{[\theta(\alpha + \beta) - \alpha]^2 + n\theta(1-\theta)}{[\alpha + \beta + n]^2}
\end{aligned}$$

The Bayes estimator with constant risk is minimax. This occurs when $\alpha = \beta = \sqrt{n}/2$, so the minimax estimator using quadratic loss is $(\alpha + x)/(\alpha + \beta + n) = (x + \sqrt{n}/2)/(n + \sqrt{n})$. This estimator is also admissible