Contents

Much of this content comes from Lehmann and Casella [1998, section 5.7] and Ferguson [1967, chapter 2].

1 Basic definitions

Let $(\mathcal{X}, \mathcal{A}, P_\theta)$ be a probability space for each $\theta \in \Theta$.

$\mathcal{X}$ is the sample space.

Θ is the parameter space.

$\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is the model.

Definition 1 (loss function). *A loss function for estimating $g(\theta) \in g(\Theta)$ is a function from $\Theta \times g(\Theta) \to \mathbb{R}^+$ such that*

$$L(\theta, d) \geq 0 \ \forall \theta \in \Theta, \ d \in g(\Theta),$$
$$L(\theta, g(\theta)) = 0 \ \forall \theta \in \Theta.$$

Definition 2 (non-randomized estimator). *A non-randomized estimator $\delta(x)$ of $g(\theta)$ is any function from $\mathcal{X}$ to $g(\Theta)$ such that $L(\theta, \delta)$ is a measurable function of x for all $\theta \in \Theta$.*

This means that $\{x : L(\theta, \delta(x)) \in B\} \in \mathcal{A}$ for all $B \in \mathcal{B}(\mathbb{R})$.
Why do we need this? So we can integrate the loss function to get the risk.

Definition 3 (risk). *The risk function $R(\theta, \delta)$ of an estimator δ is the expected loss as a function of $\theta \in \Theta$:*

$$R(\theta, \delta) = \mathrm{E}[L(\theta, \delta)|\theta] = \int L(\theta, \delta(x)) P_\theta(dx)$$

These definitions extend to randomized estimators, that is, decision procedures that proceed as follows:

1. Sample $X \sim P_\theta$, where $\theta \in \Theta$ is unknown;
2. Simulate $U \sim$ uniform[0,1] , independently of X;
3. Make decision $\delta(X, U)$.

Alternatively, you can view a randomized estimator as one that selects a non-randomized estimator at random:

1. Simulate $U \sim$ uniform [0,1] ;
2. Use the estimator $\delta(\cdot, U)$, meaning
   (a) sample $X \sim P_\theta$, where $\theta \in \Theta$ is unknown;
   (b) Make decision $\delta(X, U)$

Definition 4 (randomized estimator)**.** *A randomized estimator $\delta(x, u)$ is function from $\mathcal{X} \times [0, 1]$ to $g(\Theta)$ such that $L(\theta, \delta)$ is a measurable function of (x, u) for all $\theta \in \Theta$.*

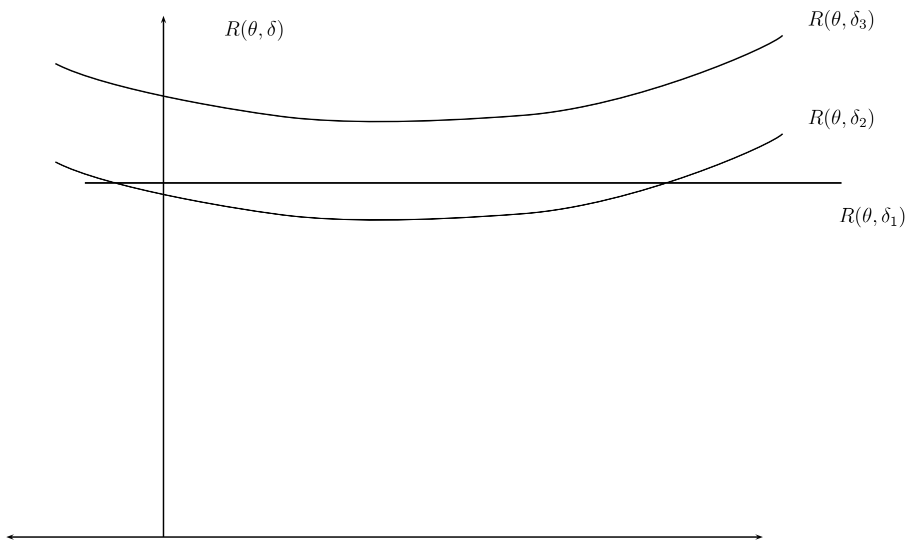
The risk function of a randomized estimator can still be written as $R(\theta, \delta) = \mathrm{E}[L(\theta, \delta)|\theta]$, but now the expectation is over X and U.

1.1 Admissibility

If $R(\theta, \delta_1) < R(\theta, \delta_2)$ for some θ-values, but
$R(\theta, \delta_1) > R(\theta, \delta_2)$ for others,

then it is not clear which one is to be preferred.

But what if δ_1 is as good as δ_2 for all θ, and better for some θ?



Definition 5 (dominate)**.** *An estimator δ_1 dominates another estimator δ_2 if*

$$\forall \theta \in \Theta,\ R(\theta, \delta_1) \leq R(\theta, \delta_2),$$
$$\exists \theta \in \Theta,\ R(\theta, \delta_1) < R(\theta, \delta_2).$$

It seems that if δ_1 dominates δ_2, then δ_2 would be eliminated from consideration. If δ_2 is not dominated by any estimator, then we may want to at least consider it.

Definition 6 (admissible). *An estimator δ is admissible if it is not dominated.*

As we've discussed, admissibility doesn't mean an estimator is any good:

Example: $\delta(x) = \theta_0$ is generally admissible if $\theta_0 \in \Theta$.

Exercise: Find a situation in which the above estimator is not admissible.

However, admissibility is a place to start when looking for good estimators.

Some interesting results

- $\boldsymbol{X} \sim N_p(\boldsymbol{\theta}, I)$, $p > 2$, then $\delta(\boldsymbol{x}) = \boldsymbol{x}$ is UMVUE for estimating $\boldsymbol{\theta} \in \mathbb{R}^2$ but inadmissible.
- If δ is admissible then it is non-randomized (in some cases).
- If δ is admissible then it is Bayes (in some cases).
- If δ is Bayes then it is admissible (in some cases).

2 Complete classes

When can we say that every admissible estimator is
Bayes?
nonrandomized?
a function of a sufficient statistic?

Let $(\mathcal{X}, \mathcal{A}, P)$ be a probability space for all $P \in \mathcal{P}$.
Let $\mathcal{D}$ be the set of estimators, including randomized estimators.

Definition 7 (Complete class). *A class of estimators $\mathcal{C} \subset \mathcal{D}$ is complete if $\forall \delta' \in \mathcal{C}^c$, $\exists \delta \in \mathcal{C}$ that dominates δ'.*

Exercise: Draw a Venn diagram of a complete class, the admissible estimators and the inadmissible estimators, indicating that

1. $\mathcal{C}$ contains all admissible estimators;
2. $\mathcal{C}^c$ contains only inadmissible estimators;
3. $\mathcal{C}$ may contain inadmissible estimators.

In the spirit of making $\mathcal{C}$ as small as possible, we can eliminate from $\mathcal{C}$ any estimators with "redundant" risk functions: If $R(\theta, \delta_1) = R(\theta, \delta_2)$, we don't really need to consider both δ_1 and δ_2.

Definition 8 (essentially complete class). *$\mathcal{C}$ is essentially complete if $\forall \delta' \in \mathcal{C}^c$, $\exists \delta \in \mathcal{C}$ such that $R(\theta, \delta) \leq R(\theta, \delta')$ $\forall \theta$.*

So an essentially complete class does not need to contain all admissible estimators, is just needs to contain all admissible risk functions.

Examples:

Recall a sufficient statistic $t(x)$ is a function of x such that $\Pr(X \in A|t(X), \theta_0) = \Pr(X \in A|t(X), \theta_1)$ for all $\theta_0, \theta_1 \in \Theta$, i.e. $\Pr(X \in A|t(X), \theta)$ does not depend on θ.

One neat thing about a sufficient statistic is that you can use them to simulate new datasets that have the same distribution as the original one, using the following procedure:

1. Sample $X \sim P_\theta$;
2. compute $t(X)$;
3. Simulate X^* from the distribution defined by $\Pr(X^* \in A|t(X))$.

Then $X^* \stackrel{d}{=} X$, i.e. $\Pr(X^* \in A) = P_\theta(A)$.

Example: $\boldsymbol{X} = (X_1, \ldots, X_n) \sim$ i.i.d. $N(\mu, \sigma^2)$. Sufficient statistics include

- $t_1(\boldsymbol{x}) = \boldsymbol{x}$;
- $t_2(\boldsymbol{x}) = \{\bar{x}, s^2\}$.

Lemma 1 (LC lem 5.7.5i). *Let $t(x)$ be any sufficient statistic. Then the class of estimators that are functions of $t(x)$ is essentially complete.*

Proof.
Given a sufficient statistic $t(x)$ and a random variable X,
create $X^* = X^*(t(X), U)$, $U \sim$ uniform[0,1] such that $X^* \stackrel{d}{=} X$.
Then for any estimator $\delta(X)$, $R(\theta, \delta(X)) = R(\theta, \delta(X^*))\ \forall\theta$.
Note that $\delta(X^*) = \delta^*(t(X), U)$, a randomized estimator based on $t(X)$. □

Lemma 2 (LC lem 5.7.5ii). *If $L(\theta, d)$ is strictly convex in $d \in g(\Theta)$ then the class of non-randomized estimators is complete.*

Proof.
The result follows if every randomized estimator is dominated by a non-randomized one.

Let $\delta'(x, u)$ be a randomized estimator, and let $\delta(x) = \mathrm{E}[\delta'(x, U)]$.

Since the expectation doesn't require knowledge of θ, this is a valid non-randomized estimator.

Since $L(\theta, d)$ is strictly convex in d by Jensen's inequality we have

$$\begin{aligned} \mathrm{E}[L(\theta, \delta'(x, U))] &> L(\theta, \mathrm{E}[\delta'(x, U)]) \\ &= L(\theta, \delta(x)), \end{aligned}$$

where the expectation is over U. The result follows, since

$$R(\theta, \delta') = \int \mathrm{E}[L(\theta, \delta'(x, U))]\ P_\theta(dx) > \int L(\theta, \delta(x))\ P_\theta(dx) = R(\theta, \delta),$$

and so δ' is dominated by a non-randomized estimator. □

The class of non-randomized estimators, or the class of estimators based on sufficient statistics, are both quite large, and of course contain lots of inadmissible estimators. Can we do better? Ideally, we'd like to identify a complete class without any inadmissible estimators in it.

Definition 9 (minimal complete class). *$\mathcal{C}$ is minimally complete if any proper subset $\mathcal{A}$ of $\mathcal{C}$ is not complete.*

Lemma 3 (LC lem 5.7.7). *If a minimal complete class exists, then it equals the class of admissible estimators.*

Definition 10 (minimal essentially complete class). *$\mathcal{C}$ is MEC if it EC and $\mathcal{A}$ is not EC for any proper subset $\mathcal{A}$ of $\mathcal{C}$.*

Exercise: Draw some Venn diagrams for that relate a complete class to a corresponding EC class, an MC class and an MEC class.

From the perspective of risk alone, we can restrict attention to an EC class.
For efficiency in our search, we should restrict attention to an MEC class, if possible.

3 Admissibility of Bayes estimators

Let π be a probability measure on Θ and δ be an estimator.

Definition 11 (Bayes risk). *The Bayes risk of δ with respect to π is*

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

Comparing estimators based on Bayes risk is attractive, as comparisons are based on numbers as opposed to functions. However, you have to decide what prior π to use.

Definition 12 (Bayes estimator)**.** *An estimator δ is Bayes with respect to a prior π on Θ if*

$$R(\pi, \delta) = \int R(\theta, \delta)\pi(d\theta) \leq \int R(\theta, \delta')\pi(d\theta) = R(\pi, \delta') \quad \forall \delta' \in \mathcal{D}.$$

Are Bayes estimators admissible? Can they be dominated? Suppose

1. δ is Bayes w.r.t. π.
2. δ' dominates δ.

Then we have

$$R(\pi, \delta') = \int R(\theta, \delta')\pi(d\theta) \leq \int R(\theta, \delta)\pi(d\theta) = R(\pi, \delta).$$

Thus the only thing that can dominate a Bayes estimator is another Bayes estimator. If there is only one Bayes estimator for a given prior, then it must be admissible.

Theorem 1. *(LC thm 5.2.4) Any unique Bayes estimator is admissible.*

Proof.
Let δ be Bayes for π and δ' be any estimator with $R(\theta, \delta') \leq R(\theta, \delta)$ $\forall \theta \in \Theta$. Then

$$R(\pi, \delta') = \int R(\theta, \delta')\pi(d\theta) \leq \int R(\theta, \delta)\pi(d\theta) = R(\pi, \delta),$$

and so δ' is also Bayes.
Since δ is unique Bayes by assumption, we must have $\delta(x) = \delta'(x)$ a.e. P_Θ.

□

In many problems the Bayes estimators are unique and therefore admissible. We will discuss conditions under which the estimator is unique when we discuss specific models and loss functions.

Admissibility without uniqueness

Theorem 2. *Let $\Theta = \{\theta_1, \ldots, \theta_K\}$, and suppose δ_π is Bayes for $\pi = \{\pi_1, \ldots, \pi_K\}$ where $\pi_k > 0\ \forall k \in \{1, \ldots, K\}$. Then δ_π is admissible.*

Proof. If δ were to dominate δ_π, then

$$R(\theta, \delta) \leq R(\theta, \delta_\pi)\ \forall \theta \in \{\theta_1, \ldots, \theta_K\}$$
$$R(\theta_0, \delta) < R(\theta_0, \delta_\pi) \text{ for some } \theta_0 \in \theta \in \{\theta_1, \ldots, \theta_K\}.$$

If so, then

$$R(\pi, \delta) = \sum_{k=1}^{K} \pi_k R(\theta_k, \delta) < \sum_{k=1}^{K} \pi_k R(\theta_k, \delta_\pi),$$

because $\pi_k > 0\ \forall k \in \{1, \ldots, K\}$. This contradicts the assumption that δ_π is the Bayes estimator.

□

We can extend this result to $\Theta \subset \mathbb{R}^p$ in some situations. However, we need to find an analogue to the "$\pi_k > 0$" condition.

Definition 13 (support)**.** *$\theta_0 \in \mathbb{R}^p$ is in the support of a prior distribution π if $\pi(\{\theta : |\theta - \theta_0| < \epsilon\}) > 0\ \forall \epsilon > 0$.*

Theorem 3. *Let Θ be an open subset of $\mathbb{R}^K$, and suppose*

1. *support$(\pi) = \Theta$;*
2. *$R(\theta, \delta)$ is continuous in $\theta\ \forall \delta \in \mathcal{D}$.*

If δ_π is Bayes for π with finite Bayes risk, then δ_π is admissible.

See LC Theorem 5.7.9 for essentially the same theorem (LC require the prior density to be everywhere nonzero).

Proof. As before, if δ were to dominate δ_π, then

$$R(\theta, \delta) \leq R(\theta, \delta_\pi)\ \forall \theta \in \Theta$$
$$R(\theta_0, \delta) < R(\theta_0, \delta_\pi) \text{ for some } \theta_0 \in \Theta.$$

Let $\eta = R(\theta_0, \delta_\pi) - R(\theta_0, \delta)$. By continuity of $R(\cdot, \delta)$,

$$\exists \epsilon > 0 : R(\theta, \delta_\pi) - R(\theta, \delta) > \eta/2 \ \forall \theta : |\theta - \theta_0| < \epsilon$$

Draw the picture. Now let $A = \{\theta : |\theta - \theta_0| < \epsilon\}$. Then

$$\begin{aligned}
R(\pi, \delta_\pi) - R(\pi, \delta) &= \int [R(\theta, \delta_\pi) - R(\theta, \delta)]\pi(d\theta) \\
&= \int_A [R(\theta, \delta_\pi) - R(\theta, \delta)]\pi(d\theta) + \int_{A^c} [R(\theta, \delta_\pi) - R(\theta, \delta)]\pi(d\theta) \\
&\geq \int_A [R(\theta, \delta_\pi) - R(\theta, \delta)]\pi(d\theta) \\
&> \tfrac{\eta}{2}\pi(A) > 0,
\end{aligned}$$

which contradicts that δ_π is Bayes.

□

Note: The assumption that $R(\theta, \delta)$ is continuous in θ for all $\delta \in \mathcal{D}$ seems very restrictive: "all $\delta \in \mathcal{D}$" is a lot of functions to be continuous for. However, this condition is actually met in several useful circumstances:

- exponential families (LC example 5.7.10);
- bounded, continuous loss functions and some other conditions.

We've shown that if δ_π is Bayes and

- unique, or
- support$(\pi) = \Theta$ and Θ is finite, or
- support$(\pi) = \Theta$ and $R(\theta, \delta)$ is continuous in θ,

then δ_π is admissible. Are there other types of admissible estimators? Or are all admissible estimators Bayes with respect to some prior? In other words, do the Bayes estimators form a complete class?

---

4 Geometry of admissible estimators

Suppose $\Theta = \{\theta_1, \ldots, \theta_K\}$ is a finite parameter space.
The risk function of an estimator δ is then a point in $\mathbb{R}^K$:

$$R(\Theta, \delta) = \begin{pmatrix} R(\theta_1, \delta) \\ \vdots \\ R(\theta_K, \delta) \end{pmatrix}$$

The *risk set* is the set of all possible risk functions:

Definition 14 (risk set)**.** *The risk set $\mathcal{S}$ for a finite-dimensional parameter space Θ is the set of all risk functions,*

$$\mathcal{S} = \{\boldsymbol{s} \in \mathbb{R}^K : \exists \delta \in \mathcal{D}, s_k = R(\theta_k, \delta), k = 1, \ldots, K\} = \text{range}_{\delta \in \mathcal{D}} R(\Theta, \delta).$$

Theorem 4. *The risk set is convex.*

Proof. Let $\boldsymbol{s}_a, \boldsymbol{s}_b \in \mathcal{S}$. Then

$$\boldsymbol{s}_a = (R(\theta_1, \delta_a), \ldots, R(\theta_K, \delta_a)) \text{ for some } \delta_a \in \mathcal{D}$$
$$\boldsymbol{s}_b = (R(\theta_1, \delta_b), \ldots, R(\theta_K, \delta_b)) \text{ for some } \delta_b \in \mathcal{D}$$

Define the randomized estimator δ_c as

$$\delta_c(X, \omega) = \delta_a(X) \times 1(\omega < p) + \delta_b(X) \times 1(\omega \geq p)$$

or equivalently,

$$\delta_c(X) = \begin{cases} \delta_a(X) & \text{w.p. } p \\ \delta_b(X) & \text{w.p. } 1-p \end{cases}$$

Then

$$R(\theta_k, \delta_c) = pR(\theta_k, \delta_a) + (1-p)R(\theta_k, \delta_b)$$

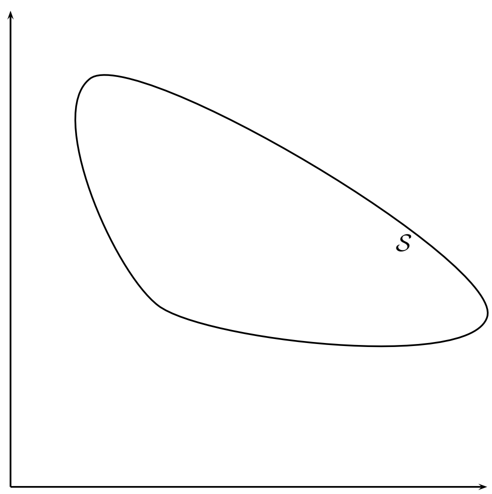


Figure 1: A risk set.

and so

$$\boldsymbol{s}_c = p\boldsymbol{s}_a + (1-p)\boldsymbol{s}_b \in \mathcal{S},$$

as it is the risk function of $\delta_c \in \mathcal{D}$.

□

Geometrically, $\mathcal{S}$ is the convex hull of the risk functions of the non-randomized estimators.

Bayes risk and Bayes rules:

A prior distribution provides an ordering of the estimators based on their Bayes risk;

$$R(\pi, \delta) = \sum \pi_k R(\theta_k, \delta) = \pi \cdot \boldsymbol{s}_\delta$$

All points with the same Bayes risk are "equally good" by the Bayesian criteria.

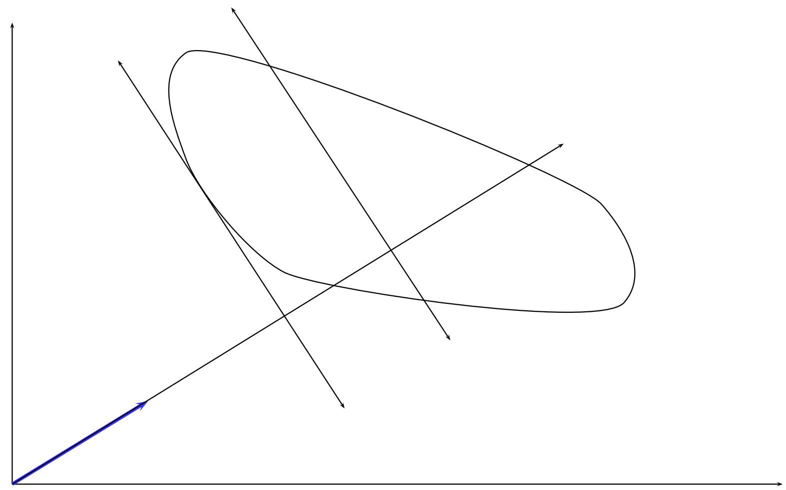
$$\delta_1 \sim_\pi \delta_2 \ \text{ if } \ \pi \cdot \boldsymbol{s}_{\delta_1} = \pi \cdot \boldsymbol{s}_{\delta_2}$$

All estimators with Bayes risk equal to $r \in \mathbb{R}^+$ have risk functions that lie on a hyperplane defined by

$$\{\boldsymbol{x} \in \mathbb{R}^K : \pi \cdot \boldsymbol{x} = r\}.$$

Risk functions that are "equally good" are obtained by intersecting this hyperplane with the risk set:

$$\mathcal{S} \cap \{\boldsymbol{x} \in \mathbb{R}^K : \pi \cdot \boldsymbol{x} = r\}$$



Which risk functions (elements of $\mathcal{S}$) correspond to Bayes estimators?
Which risk functions are admissible?

We will now show that the admissible risk functions those that are Bayes w.r.t. some prior.

Definition 15 (lower quantant)**.** *The lower quantant of a point* $\boldsymbol{s} \in \mathbb{R}^K$ *is the set*

$$Q(\boldsymbol{s}) = \{\boldsymbol{x} \in \mathbb{R}^K : x_k \leq s_k, k = 1, \ldots, K\}$$

Draw the picture of $Q(\boldsymbol{s})$ for admissible and inadmissible estimators.

Lemma 4. *If* δ *is admissible, then its risk function* $\boldsymbol{s} = R(\Theta, \delta)$ *satisfies* $Q(\boldsymbol{s}) \cap \mathcal{S} = \{\boldsymbol{s}\}$.

Proof.
$\boldsymbol{s} \in Q(\boldsymbol{s})$ by definition.
$\boldsymbol{s} \in \mathcal{S}$ as $\boldsymbol{s}$ is the risk function of $\delta \in D$.
Thus $\boldsymbol{s} \in Q(\boldsymbol{s}) \cap \mathcal{S}$.

Now let $\boldsymbol{x} \in Q(\boldsymbol{s})$, and $\boldsymbol{x} \neq \boldsymbol{s}$.
Then $x_j \leq s_j \ \forall j \in \{1, \ldots, K\}$, and $x_j < s_j$ for some $j \in \{1, \ldots, K\}$.
Then $\boldsymbol{x} \notin \mathcal{S}$: If it were, then it would correspond to the risk function of an estimator that dominated δ. □

Now we should be able to identify all admissible estimators on the diagram.
Next, we need to show that these are all Bayes estimators.

Lemma 5 (separating hyperplane theorem)**.** *Let S_1, S_2 be two disjoint convex subsets of $\mathbb{R}^K$. Then there exists a vector $\boldsymbol{w} \in \mathbb{R}^K$, $\boldsymbol{w} \neq 0$ s.t. $\boldsymbol{w} \cdot \boldsymbol{x}_1 \leq \boldsymbol{w} \cdot \boldsymbol{x}_2 \ \ \forall \boldsymbol{x}_1 \in S_1, \ \boldsymbol{x}_2 \in S_2$.*

Draw the picture. Such a separating hyperplane is given by

$$\{\boldsymbol{x} \in \mathbb{R}^K : \boldsymbol{w} \cdot \boldsymbol{x} = c\},$$

where $\sup_{x_1 \in S_1} \boldsymbol{w} \cdot \boldsymbol{x}_1 \leq c \leq \inf_{x_2 \in S_2} \boldsymbol{w} \cdot \boldsymbol{x}_2$. The vector $\boldsymbol{w}$ is normal to the hyperplane.

Theorem 5 (complete class theorem I)**.** *If δ is admissible and Θ is finite, then δ is Bayes (w.r.t. some prior distribution).*

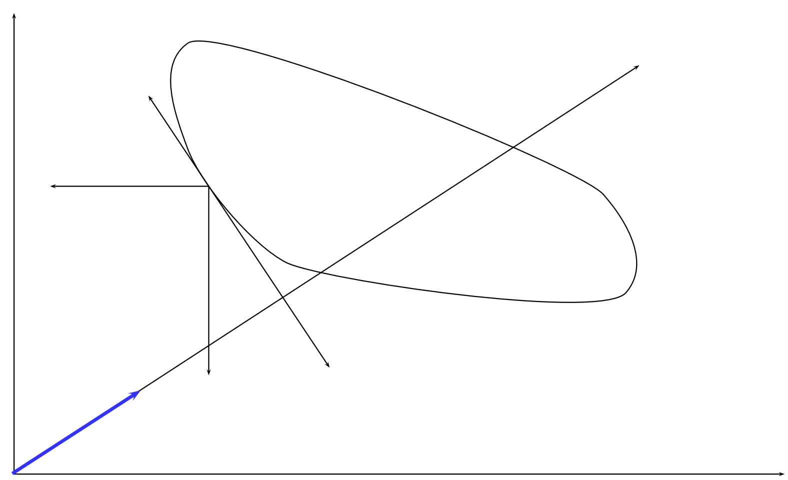
Proof.
If δ is admissible then $Q(\boldsymbol{s}) \cap \mathcal{S} = \{\boldsymbol{s}\}$, where $\boldsymbol{s} = R(\Theta, \delta)$.
Now let $\tilde{Q}(\boldsymbol{s}) = Q(\boldsymbol{s}) \setminus \{\boldsymbol{s}\}$.
$\tilde{Q}(\boldsymbol{s})$ is convex, as is $\mathcal{S}$, and $\tilde{Q}(\boldsymbol{s}) \cap \mathcal{S} = \emptyset$.
Thus there is a hyperplane separating them, i.e. there is a vector $\boldsymbol{w}$, $\boldsymbol{w} \neq 0$ s.t.

$$\boldsymbol{w} \cdot \boldsymbol{x} \leq \boldsymbol{w} \cdot \boldsymbol{s}' \text{ for all } \boldsymbol{x} \in \tilde{Q}(\boldsymbol{s}) \text{ and } \boldsymbol{s}' \in \mathcal{S}.$$

No coordinate of $\boldsymbol{w}$ can be negative: If $w_j < 0$, we could "move" x_j to be arbitrarily negative with $\boldsymbol{x}$ still in $\tilde{Q}(\boldsymbol{s})$, making $\boldsymbol{w} \cdot \boldsymbol{x}$ arbitrarily large. Hence $w_j \geq 0 \; \forall j \in \{1, \ldots, K\}$.
Now let $\pi = \boldsymbol{w}/(\boldsymbol{w} \cdot \mathbf{1})$, a probability distribution.
We have $\pi \cdot \boldsymbol{x} \leq \pi \cdot \boldsymbol{s}' \;\; \forall \boldsymbol{x} \in \tilde{Q}(\boldsymbol{s}), \boldsymbol{s}' \in \mathcal{S}$. Therefore, for any $\boldsymbol{s}' \in \mathcal{S}$

$$\pi \cdot \boldsymbol{s}' \geq \sup_{x \in \tilde{Q}(s)} \boldsymbol{x} \cdot \pi = \boldsymbol{s} \cdot \pi,$$

as $\boldsymbol{s}$ is a limit point of $\tilde{Q}(\boldsymbol{s})$.
Thus $\boldsymbol{s} \cdot \pi \leq \pi \cdot \boldsymbol{s}' \; \forall \boldsymbol{s}' \in \mathcal{S}$, so δ is Bayes w.r.t. π. □



The result says that all admissible procedures must be Bayes, and so the Bayes procedures form a complete class, *if there is such a class.* Situations where a complete class may not exist occur when $\mathcal{S}$ is not closed (or not "closed from below"). In such cases, there may not exist any admissible estimators (or Bayes estimators). However, $\mathcal{S}$ is closed in in most problems in the literature in which Θ is finite. For example, $\mathcal{S}$ is closed if the decision space D is finite.

However, not all Bayes procedures are admissible: Recall our Theorem 2 on admissibility required $\pi_k > 0 \; \forall k$. The set of admissible risk functions therefore equals the risk functions of *admissible* Bayes procedures.

Theorem (complete class theorem II). *If Θ is finite and $\mathcal{S}$ is closed then the class of Bayes rules is complete and the admissible rules form a minimal complete class.*

For more on this, see Ferguson [1967, section 2.10].

5 Complete class theorems for Euclidean parameter spaces

The above theorem is straightforward to prove and visualize, but not much use in practice, as it requires the parameter space to be finite. For more general parameter spaces, there is a large literature on complete class theorems that we will not go through, but we do provide some results for some particularly important cases, including exponential family models. These results should at least indicate that when searching for a good estimator, Bayes estimators are a good starting point.

Finite Θ: admissible Bayes rules form a (minimal) complete class.

Euclidean Θ: admissible Bayes rules generally do not form a complete class.

Draw the picture.
What estimators are admissible, but not Bayes? It turns out that in many useful cases, any admissible estimator must be "nearly" Bayes. Here are two types of estimators that are (in different senses) nearly Bayes:

Definition 16 (generalized Bayes estimator). *Let π be a measure on Θ (not necessarily a probability measure). An estimator δ_π is generalized Bayes w.r.t. π if*

$$\int R(\theta, \delta_\pi)\ \pi(d\theta) \leq \int R(\theta, \delta)\ \pi(d\theta) \quad \forall \delta \in \mathcal{D}.$$

Definition 17 (limiting Bayes). *Let $\{\pi_n\}$ be a sequence of probability measures and $\{\delta_{\pi_n}\}$ the corresponding sequence of Bayes estimators. An estimator δ is limiting Bayes w.r.t. $\{\pi_n\}$ if*

$$\delta(x) = \lim_{n\to\infty} \delta_{\pi_n}(x) \text{ a.e. } \mathcal{P}_\theta.$$

We now discuss three theorems relating such estimators to admissibility. The first is a useful tool for proving an estimator is admissible, but does not characterize the admissible estimators.

Theorem 6 (LC 5.7.13). *Suppose $\Theta \subset \mathbb{R}^p$ is open, and that $R(\theta, \delta)$ is continuous in θ for all $\delta \in \mathcal{D}$. Let δ be an estimator and $\{\pi_n\}$ be a sequence of measures such that $R(\pi_n, \delta_{\pi_n}) < \infty$ and for any open set $\Theta_0 \subset \Theta$,*

$$\frac{R(\pi_n, \delta) - R(\pi_n, \delta_{\pi_n})}{\pi_n(\Theta_0)} \to 0 \text{ as } n \to \infty.$$

Then δ is admissible.

Proof.
Suppose δ' dominates δ.
Then $R(\theta, \delta) - R(\theta, \delta') \geq 0 \ \forall \theta \in \Theta$ and there exists $\theta_0 \in \Theta$ such that $R(\theta_0, \delta) - R(\theta_0, \delta') > 0$.
By continuity of risk, $\exists \epsilon > 0$ and an open $\Theta_0 \subset \Theta$ s.t. $R(\theta, \delta) - R(\theta, \delta') > \epsilon \ \forall \theta \in \Theta_0$.
Therefore,

$$\begin{aligned}
R(\pi_n, \delta) - R(\pi_n, \delta') &= \int [R(\theta, \delta) - R(\theta, \delta')]\pi_n(d\theta) \\
&\geq \int_{\Theta_0} [R(\theta, \delta) - R(\theta, \delta')]\pi_n(d\theta) \\
&> \epsilon \int_{\Theta_0} \pi_n(d\theta) \text{ , implying that} \\
\frac{R(\pi_n, \delta) - R(\pi_n, \delta')}{\pi_n(\Theta_0)} &> \epsilon \ \forall n, \\
\frac{R(\pi_n, \delta) - R(\pi_n, \delta_{\pi_n})}{\pi_n(\Theta_0)} + \frac{R(\pi_n, \delta_{\pi_n}) - R(\pi_n, \delta')}{\pi_n(\Theta_0)} &> \epsilon \ \forall n
\end{aligned}$$

Since the left-hand ratio converges to zero (and is always nonnegative), we can find an n such that it is less than, say $\epsilon/2$, giving

$$R(\pi_n, \delta_{\pi_n}) - R(\pi_n, \delta') > \pi_n(\Theta_0)\epsilon/2$$

which contradicts the assumption that δ_{π_n} is the minimizer of $R(\pi_n, \delta_{\pi_n})$. □

Using this theorem to prove admissibility is often called "Blyth's method," named after C.R. Blyth. As we will see, this result implies that limits of Bayes estimators are admissible. For example, we will have cases where π_n converges to something other than a probability measure, and δ is not Bayes for any prior, yet δ will satisfy the above conditions and therefore be admissible.

Example: Univariate normal mean

Let $X \sim N(\mu, 1)$. Then X is the MLE and UMVUE of μ. Is it admissible under squared error loss? Let's try using Blyth's method. First, we need to come up with a sequence of priors for which the minimum Bayes risks are "close" to the Bayes risks of δ. As we will soon show, if

- $\mu \sim N(0, \tau^2) = \pi_\tau$, and
- $X \sim N(\mu, 1)$,

then the Bayes estimator and its Bayes risk are

$$\delta_{\tau^2}(x) = \frac{\tau^2}{1+\tau^2}x \ , \ R(\pi_\tau, \delta_{\tau^2}) = \frac{\tau^2}{1+\tau^2}.$$

The Bayes risk of $\delta(x) = x$ is given by

$$\begin{aligned} R(\pi_\tau, \delta) &= \int R(\mu, \delta)\pi_\tau(d\mu) \\ &= \int 1 \times \pi_\tau(d\mu) = 1, \end{aligned}$$

which is close to $\frac{\tau^2}{1+\tau^2}$ when τ is large. This is promising - let's try to apply the theorem: For any open $\Theta_0 \subset \Theta$,

$$
\begin{aligned}
0 \leq \lim_{\tau\to\infty} \frac{R(\pi_\tau, X) - R(\pi_\tau, \delta_{\tau^2})}{\pi_\tau(\Theta_0)} &= \lim_{\tau\to\infty} \frac{(1-\frac{\tau^2}{1+\tau^2})}{\pi_\tau(\Theta_0)} \\
&= \lim_{\tau\to\infty} \frac{1}{(1+\tau^2)\pi_\tau(\Theta_0)} \leq \lim_{\tau\to\infty} \frac{1}{\tau^2\pi_\tau(\Theta_0)}.
\end{aligned}
$$

Therefore, the limit is zero if

$$\lim_{\tau^2\to\infty} \tau^2\pi_\tau(\Theta_0) = \infty.$$

We have

$$
\begin{aligned}
\tau^2\pi_\tau(\Theta_0) &= \tau^2 \int_{\Theta_0} (2\pi\tau^2)^{-1/2} \exp(-\mu^2/[2\tau^2])\, d\mu \\
&= (2\pi)^{-1/2} \times (\tau^2)^{1-1/2} \times \int_{\Theta} \exp(-\mu^2/[2\tau^2])\, d\mu \\
&(2\pi)^{-1/2} \times \text{(a)} \times \text{(b)}.
\end{aligned}
$$

Now take the limit as $\tau \to \infty$:

$$
\begin{aligned}
(a) &\to \infty \text{ as } \tau \to \infty \\
(b) &\to \text{Vol}(\Theta_0) > 0 \quad \text{as} \quad \tau \to \infty,
\end{aligned}
$$

and so the desired limit is achieved. By the theorem, $\delta(x) = x$ is admissible for estimating μ.

Exercise: Extend this result to show that $\bar{X}$ is admissible under squared error loss in the case that $X_1, \ldots, X_n \sim$ i.i.d. $N(\mu, \sigma^2)$ with σ^2 unknown.

The next two theorems are complete class theorems, in that they provide characterizations of the set of admissible estimators.

(**A**) Assumptions:

1. Let $\mathcal{P} = \{p(x|\theta) : \theta \in \Theta\}$ be a model for $(\mathcal{X}, \mathcal{A})$ dominated by a σ-finite measure μ.

2. Let $p(x|\theta) > 0 \ \forall x \in \mathcal{X}, \theta \in \Theta$.

3. Let $L(\theta, d)$ be continuous and strictly convex in d for every θ, and satisfy

$$\lim_{|d|\to\infty} L(\theta, d) = \infty \ \forall \theta \in \Theta.$$

Theorem 7 (LC 5.7.15)**.** *Under assumptions (**A**), for every admissible estimator δ there is a sequence of priors $\{\pi_n\}$, each with support on a finite set (and hence giving finite Bayes risk), s.t.*

$$(*) \qquad \delta_{\pi_n} \to \delta \text{ a.s. } \mu,$$

where δ_{π_n} is the Bayes estimator under π_n.

This means that every admissible estimator is the limit of some Bayes estimator.

Corollary 1 (LC 5.7.16)**.** *Under assumptions (**A**), the class of all estimators δ for which $(*)$ holds is a complete class.*

Again, keep in mind that a complete class may contain inadmissible estimators.

The above theorem characterizes admissible estimators as limits of Bayes estimators. Thus admissible estimators don't need to be Bayes, but they need to be "very close" to being Bayes.

The final complete class theorem concerns generalized Bayes estimators. Note that every Bayes estimator is generalized Bayes, but not vice versa. In particular, π could be "uniform" Lebesgue measure on $\mathbb{R}^K$. This is not a probability distribution, and the resulting minimizer of the integrated risk is not (generally) a Bayes estimator for any real prior distribution. Nevertheless, the generalized Bayes estimators form a complete class (in some cases).

Theorem 8 (LC 5.7.17)**.** *If the model $\mathcal{P}$ is an exponential family and condition (**A**).3 is satisfied, then any admissible estimator is generalized Bayes, and so the generalized Bayes estimators form a complete class.*

Once again, a complete class may contain inadmissible estimators.

---

6 Complete class theorems for tests

Recall the standard hypothesis testing problem:

$X \sim P_\theta, \ \theta \in \Theta$

$H_0 : \theta \in \Theta_0 \subset \Theta$

$H_1 : \theta \in \Theta_1 = \Theta \cap \Theta_0^c$

Task: decide between H_0 and H_1 based on X.

A *test function* is a measurable map $\phi : \mathcal{X} \to [0,1]$.

- The procedure "reject H_0 with probability $\phi(X)$" is a test of H_0 (versus H_1).
- If $\phi : \mathcal{X} \to \{0,1\}$ then ϕ is a non-randomized test (otherwise is randomized).
- If $\mathrm{E}[\phi(X)|\theta] \leq \alpha \ \forall \theta \in \Theta_0$, then ϕ is a level-α test.
- If ϕ maximizes $\mathrm{E}[\phi(X)|\theta_1]$ among all level-α tests, then it is a "most powerful" (MP) level-α test of Θ_0 versus θ_1 ($\mathrm{E}[\phi(X)|\theta_1]$ is the *power* at θ_1).

The workhorse theorem of classical testing theory is the Neyman-Pearson lemma, which identifies the most powerful test of a simple null versus a simple alternative:

Theorem. *[Lehmann and Romano, 2005, Theorem 3.2.1] A test function* $\phi(x) : \mathcal{X} \to [0,1]$ *is most powerful for testing* $H_0 : X \sim P_{\theta_0}$ *versus* $H_1 : X \sim P_{\theta_1}$ *among all tests with level* $\mathrm{E}[\phi(X)|\theta_0]$ *if and only if it satisfies*

$$\phi(x) = \begin{cases} 1 & \text{when } p(x|\theta_1) > kp(x|\theta_0) \\ 0 & \text{when } p(x|\theta_1) < kp(x|\theta_0). \end{cases} ,$$

for some $k \in \{0\} \cup \mathbb{R}^+ \cup \{\infty\}$ *(and interpreting* $0 \times \infty$ *as* 0*).*

As we will see, many tests of more general hypotheses can be reduced down this simple versus simple case.

The proof of this theorem is not difficult (please go through it), but doing everything in terms of level and power is a bit clunky. Alternatively, by viewing the testing problem as a decision problem we can essentially prove the Neyman-Pearson lemma using the complete class theorems we've already covered, while also gaining some insight into the risks of level-α tests.

6.1 Complete class theorem for simple hypotheses

Deciding between simple hypotheses H_0 versus H_1 is essentially the same as estimating θ in the case where Θ and D each contain two points:

$X \sim P_\theta,\ \theta \in \{\theta_0, \theta_1\}$

$D = \{d_0, d_1\} = \{$ "say θ_0", "say θ_1" $\}$

$L(\theta, d) =$

	d_0	d_1
$\theta \in \Theta_0$	0	l_1
$\theta \in \Theta_1$	l_0	0

l_1 is the loss from false rejection, l_0 the loss from false acceptance. The risk of any estimator/decision rule is

$$\begin{aligned} R(\theta, \delta) &= \mathrm{E}_{X|\theta}[l_1 1(\theta = \theta_0, \delta(X) = \theta_1) + l_0 1(\theta = \theta_1, \delta(X) = \theta_0)] \\ &= l_1 1(\theta = \theta_0) \Pr(\delta(X) = \theta_1|\theta_0) + l_0 1(\theta = \theta_1) \Pr(\delta(X) = \theta_0|\theta_1) \end{aligned}$$

What are the admissible decision rules? Note that both the parameter space and decision space are finite. Our theorems regarding this case are

- If δ is Bayes w.r.t. a prior $\pi_0 = 1 - \pi_1$, $0 < \pi_0 < 1$, then δ is admissible.
- If δ is admissible, then δ is Bayes w.r.t. some prior.

In other words, the admissible Bayes procedures form a complete class. Let's identify the form of the Bayes procedures. The Bayes risk of a procedure δ under prior π is

$$\begin{aligned} R(\pi, \delta) &= \sum_{\Theta} R(\theta, \delta)\pi(\theta) \\ &= \sum_{\Theta} \int_{\mathcal{X}} L(\theta, \delta(x))\, p(x|\theta)\mu(dx)\, \pi(\theta) \\ &= \int_{\mathcal{X}} \sum_{\Theta} L(\theta, \delta(x))\, \pi(\theta|x)\, p(x)\mu(dx), \end{aligned}$$

where

- $p(x) = \pi_0 p(x|\theta_0) + \pi_1 p(x|\theta_1)$ is the marginal density of X,
- $\pi(\theta|x)$ is the conditional/posterior distribution/density of θ given $X = x$:

$$\pi(\theta_0|x) = \frac{\pi_0 p(x|\theta_0)}{\pi_0 p(x|\theta_0) + \pi_1 p(x|\theta_1)}\ ,\ \pi(\theta_1|x) = \frac{\pi_1 p(x|\theta_1)}{\pi_0 p(x|\theta_0) + \pi_1 p(x|\theta_1)}$$

Exercise: Show that if $\delta(x)$ minimizes

$$\sum_{\Theta} L(\theta, d)\, \pi(\theta|x)$$

in $d \in D$ for each x, then it is Bayes under π.

For our problem,

$$\sum_{\Theta} L(\theta, d)\, \pi(\theta|x) = \left[1(d = \theta_1) l_1 \pi_0 p(x|\theta_0) + 1(d = \theta_0) l_0 \pi_1 p(x|\theta_1)\right] / p(x).$$

For a given x, we would want to "say θ_1" if

$$l_1 \pi_0 p(x|\theta_0) < l_0 \pi_1 p(x|\theta_1),$$

or equivalently, if

$$\frac{\pi_1 p(x|\theta_1)}{\pi_0 p(x|\theta_0)} > \frac{l_1}{l_0},$$

and we would say "say θ_0" otherwise. This result should be intuitive:

- $[\pi_1 p(x|\theta_1)]/[\pi_0 p(x|\theta_0)]$ are the posterior odds of $\theta = \theta_1$ versus $\theta = \theta_0$;
- l_1/l_0 are the relative costs of type I versus type II errors.

Rewriting the Bayes rule yet again, we have that any Bayes rule is of the form

$$\delta_k(x) = \begin{cases} 1 & \text{when } p(x|\theta_1) > kp(x|\theta_0) \\ 0 & \text{when } p(x|\theta_1) < kp(x|\theta_0). \end{cases}$$

where $k = [\pi_0 l_1]/[\pi_1 l_0]$. For x values such that $p(x|\theta_1) = kp(x|\theta_0)$, $\delta(x)$ can be anything between 0 and 1 and still be Bayes for π.

Are such tests admissible? Recall all Bayes procedures for a finite parameter space are admissible if $\pi(\theta) > 0$ for all $\theta \in \Theta$. Therefore, all tests of the above form are admissible for $0 < k < \infty$.

If $\pi_0 \in \{0, 1\}$ then there may not be a unique Bayes procedure. Some Bayes procedures may be inadmissible, but there are admissible ones (there must be, as the Bayes procedures form a complete class). You can show that a minimal complete class consists of the procedures described above, in addition to the following two tests:

$$\delta_\infty(x) = \begin{cases} 1 & \text{if } p(x|\theta_0) = 0 \\ 0 & \text{if } p(x|\theta_0) > 0 \end{cases} \quad , \ \delta_0(x) = \begin{cases} 1 & \text{if } p(x|\theta_1) > 0 \\ 0 & \text{if } p(x|\theta_1) = 0 \end{cases}$$

Exercise: Show that these tests are both admissible, and are Bayes under the priors $\pi_0 = 1$ and $\pi_1 = 1$ respectively. Find a situation where other Bayes estimators will not be admissible.

We have now characterized the admissible tests (the minimal complete class) as

$$\{\delta_k : k \in \{0\} \cup (0, \infty) \cup \{\infty\}\}.$$

How does this relate to a test being most powerful for its size? We will show:

- an admissible test is most powerful for its size, but
- a test most powerful for its size is not necessarily admissible.

The result follows from the definitions. A test δ is most powerful for its size if

$$\mathrm{E}[\delta|\theta_1] \geq \mathrm{E}[\delta'|\theta_1] \ \forall \delta' : \mathrm{E}[\delta|\theta_0] \geq \mathrm{E}[\delta'|\theta_0].$$

On the other hand, the risk function of a test δ is given by

$$R(\theta_0, \delta) = l_1 \mathrm{E}[\delta|\theta_0] \ , \ R(\theta_1, \delta) = l_0 \mathrm{E}[1 - \delta|\theta_1].$$

Lemma 6. *An admissible test is most powerful for its size.*

Proof.
Let δ be admissible, and let δ' be any test such that $\mathrm{E}[\delta|\theta_0] \geq \mathrm{E}[\delta'|\theta_0]$
For such a test, $R(\theta_0, \delta') \leq R(\theta_0, \delta)$.
Since δ is admissible, δ' can't dominate it.
We can't have $R(\theta_1, \delta') < R(\theta_1, \delta)$, so we must have

$$\begin{aligned} R(\theta_1, \delta') &\geq R(\theta_1, \delta) \\ l_0 \mathrm{E}[1 - \delta'|\theta_1] &\geq l_0 \mathrm{E}[1 - \delta|\theta_1] \\ \mathrm{E}[\delta|\theta_1] &\geq \mathrm{E}[\delta'|\theta_1]. \end{aligned}$$

□

On the other hand, suppose δ is MP for its size. There could exist a δ' such that

- $\mathrm{E}[\delta|\theta_0] > \mathrm{E}[\delta'|\theta_0]$ (δ' has lower type I error, but is still "level $\mathrm{E}[\delta|\theta_0]$") ,
- $\mathrm{E}[\delta|\theta_1] = \mathrm{E}[\delta'|\theta_1]$ (the tests have the same power).

Both tests are technically level $\mathrm{E}[\delta|\theta_0]$ with the same power, but δ' has lower risk at θ_0 and so dominates δ.

Example: Let $0 < a < b < c < d$, and

- $P_0 = \mathrm{uniform}(0, c)$
- $P_1 = \mathrm{uniform}(b, d)$

Consider the two tests

$$\delta(x) = 1(X > a) \ , \ \delta'(x) = 1(X > b)$$

They both have power 1, and

$$\mathrm{E}[\delta(X)|\theta_0] = \Pr(X > a|\theta_0) = (c-a)/c > (c-b)/c = \mathrm{E}[\delta'(X)|\theta_0],$$

so both "have level" $(c-a)/c$, so δ is MP at its level. But δ' dominates δ in terms of risk. However, as shown in LR theorem 3.2.1, such dominated MP tests can only occur when they also have power 1: If δ and δ' have equal power less than 1, but δ' has lower size, then δ can't be MP at its level because we could create a test $\tilde{\delta}$ with size equal to that of δ but with greater power simply by mixing δ' with a procedure that randomly rejects with some small probability.

Anyway, the point is not to alert you about the existence of power-1 low-risk tests. Rather, the point is to illustrate that risk considerations can provide a more complete picture than than just thinking in terms level and power.

Discussion: Recall that our MP/admissible tests are essentially of the form

$$\delta_k(X) = 1(p_1(X)/p_0(X) > k)$$

where

- k is chosen to set a type-I error rate, or
- k is chosen to reflect prior opinion and loss: $k = [\pi_0 l_1]/[\pi_1 l_0]$.

Suppose $X = (X_1, \ldots, X_n)$, with these elements being i.i.d. under both p_0 and p_1. Questions:

1. For fixed α, what(typically) happens to k as $n \to \infty$?
2. Does it make sense for any of $\{\pi_0, \pi_1, l_0, l_1\}$ to depend on sample size?

References

Thomas S. Ferguson. *Mathematical statistics: A decision theoretic approach*. Probability and Mathematical Statistics, Vol. 1. Academic Press, New York, 1967.

E. L. Lehmann and George Casella. *Theory of point estimation*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 1998. ISBN 0-387-98502-6.

E. L. Lehmann and Joseph P. Romano. *Testing statistical hypotheses*. Springer Texts in Statistics. Springer, New York, third edition, 2005. ISBN 0-387-98864-5.