

Lecture 4: Uniformly best rules; admissibility; linear estimators of a normal mean

Statistical Inference–II ■ Unit I: Decision Theory

Kapil Kumar

Department of Statistics
Chaudhary Charan Singh University, Meerut

M.Sc. Statistics, Semester III
Academic Session 2026–2027

Previous lecture

Lecture 3: Randomised decision rules; convexity of the risk set.

Today

- Comparison of Decision Rules
 - Uniformly Best Rules
 - Admissibility

Definition 4.1 (Uniformly better; uniformly best)

A rule δ_1 is said to be *uniformly better than* (or to *dominate*) a rule δ_2 if

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

with strict inequality for at least one $\theta \in \Theta$.

A rule $\delta_0 \in \mathcal{D}$ is *uniformly best* in $\mathcal{D}$ if

$$R(\theta, \delta_0) \leq R(\theta, \delta) \quad \text{for all } \theta \in \Theta \text{ and all } \delta \in \mathcal{D}.$$

A uniformly best rule, if it exists, is clearly the rule to use. Unfortunately such rules almost never exist in problems of practical interest.

Proposition 4.2 (Non-existence of a uniformly best estimator)

Let Θ contain at least two points, let $\mathcal{A} = \Theta$, let $L(\theta, a) = (a - \theta)^2$, and suppose that $\mathcal{D}$ contains every constant rule $\delta_c(\mathbf{x}) \equiv c$, $c \in \Theta$.

If the data never reveal θ without error (for example, if the densities $f(\mathbf{x} \mid \theta)$ are positive on a common support), then no uniformly best rule exists in $\mathcal{D}$.

Idea of the proof.

Suppose δ_0 were uniformly best. The constant rule $\delta_c \equiv c$ has risk $(c - \theta)^2$, which vanishes at $\theta = c$. Comparing δ_0 with δ_c at $\theta = c$,

$$0 \leq R(c, \delta_0) \leq R(c, \delta_c) = 0,$$

so $R(c, \delta_0) = 0$, and this holds for every $c \in \Theta$. Thus δ_0 would estimate θ *without error, whatever the true value of θ* — which is impossible unless the data reveal θ exactly.

(A complete proof, and an example showing why the common-support condition is needed: Lecture Notes, Appendix A.) □

This negative result is the engine of the whole subject. Since we cannot minimise the risk function uniformly, we must adopt one of the strategies on the next frame.

Three responses to the non-existence of a best rule

- (A) **Restrict the class $\mathcal{D}$** , for instance to unbiased rules or to equivariant rules, and seek a uniformly best rule within the restricted class (leading to UMVUE and to minimum-risk equivariant estimators).
- (B) **Reduce the risk function to a single number** by averaging it over a prior distribution on Θ (the *Bayes* principle).
- (C) **Reduce the risk function to a single number** by taking its supremum over Θ (the *minimax* principle).

Before pursuing (B) and (C), we develop the weakest and most universally accepted requirement: **admissibility**.

Definition 4.3 (Admissible and inadmissible rules)

A decision rule $\delta_0 \in \mathcal{D}$ is called *inadmissible* if there exists another rule $\delta \in \mathcal{D}$ which is uniformly better than δ_0 , that is, if there exists $\delta \in \mathcal{D}$ with

$$R(\theta, \delta) \leq R(\theta, \delta_0) \quad \text{for all } \theta \in \Theta,$$

and

$$R(\theta, \delta) < R(\theta, \delta_0) \quad \text{for at least one } \theta \in \Theta.$$

A rule which is not inadmissible is called *admissible*.

Thus δ_0 is admissible if no other rule in $\mathcal{D}$ dominates it.

Remark 4.4 (What admissibility does and does not say)

- (i) Admissibility is a *minimal* requirement, not a recommendation. An inadmissible rule should certainly not be used, since something uniformly better is available; but admissibility alone is very weak. In the estimation of a normal mean, the absurd rule $\delta_{17}(\mathbf{x}) \equiv 17$ is admissible under squared error loss.

It has zero risk at $\theta = 17$, and no rule that differs from 17 with positive probability can match that, so nothing dominates it (details: Lecture Notes, Appendix A).

Remark (cont.)

- (ii) Admissibility depends on the class $\mathcal{D}$, on the loss function, and on the model. A rule admissible in one formulation may be inadmissible in another.
- (iii) Domination is only a *partial* ordering of rules. Two admissible rules either have identical risk functions or are incomparable (each is strictly better somewhere).
- (iv) Admissibility can fail in surprising ways. Stein (1956) and James and Stein (1961) showed that in estimating the mean vector of a p -variate normal distribution under squared error loss, the sample mean $\bar{\mathbf{X}}$ is inadmissible for $p \geq 3$ — a result that overturned a long standing consensus.

Example 4.5 (Comparing linear estimators of a normal mean)

Let $X \sim N(\theta, 1)$ with $\theta \in \mathbb{R}$, and consider the class of rules $\delta_c(X) = cX$, $c \geq 0$, under squared error loss $L(\theta, a) = (a - \theta)^2$.

We compute the risk. Since $\mathbb{E}_\theta(X) = \theta$ and $\mathbb{E}_\theta(X^2) = \text{Var}_\theta(X) + \{\mathbb{E}_\theta(X)\}^2 = 1 + \theta^2$,

$$\begin{aligned} R(\theta, \delta_c) &= \mathbb{E}_\theta[(cX - \theta)^2] = \mathbb{E}_\theta[c^2X^2 - 2c\theta X + \theta^2] \\ &= c^2\mathbb{E}_\theta(X^2) - 2c\theta\mathbb{E}_\theta(X) + \theta^2 \\ &= c^2(1 + \theta^2) - 2c\theta^2 + \theta^2 \\ &= c^2 + \theta^2(c^2 - 2c + 1) \\ &= c^2 + (1 - c)^2\theta^2. \end{aligned}$$

Example 4.5 (cont.): reading off the consequences

- For $c = 1$: $R(\theta, \delta_1) = 1$ for every θ — a *constant* risk function.
- For $c > 1$: $R(\theta, \delta_c) = c^2 + (1 - c)^2\theta^2 \geq c^2 > 1 = R(\theta, \delta_1)$ for every θ , with strict inequality everywhere. Hence δ_c is *inadmissible* for every $c > 1$, being dominated by δ_1 .
- For $0 \leq c < 1$: the risk $c^2 + (1 - c)^2\theta^2$ equals $c^2 < 1$ at $\theta = 0$ but tends to ∞ as $|\theta| \rightarrow \infty$. Thus δ_c is better than δ_1 near the origin and worse far from it: the two are *incomparable*. More generally, for $0 \leq c_1 < c_2 \leq 1$ the risk functions of δ_{c_1} and δ_{c_2} cross exactly once on $\theta > 0$, at $\theta^2 = (c_1 + c_2)/(2 - c_1 - c_2)$ (and symmetrically on $\theta < 0$), so no member of the family $\{\delta_c : 0 \leq c \leq 1\}$ dominates another. Every δ_c with $0 \leq c \leq 1$ is therefore admissible *within this family*.

Remark 4.6 (A caution)

The example of linear estimators establishes admissibility of δ_c , $0 \leq c \leq 1$, only *relative to the family* $\{\delta_c\}$.

- Admissibility in the class of *all* rules is a stronger assertion and does not follow from the computation above.
- It is in fact true that $\delta_1(X) = X$ is admissible in the class of all estimators of a normal mean under squared error loss, but its proof requires an argument beyond comparison within the linear subclass.

Students should therefore not conclude admissibility in $\mathcal{D}$ from a calculation restricted to a smaller family.

Practice (Lecture Notes, Chapter 1)

- Short-answer questions (Section 1.12): Q5, Q6.

Next lecture

Lecture 5: The minimax principle; minimax action in a no-data problem.