On Extended Admissible Decision Procedures and Their Nonstandard Bayes Risk

Haosui Duanmu
Department of Economics
University of California, Berkeley

July 10, 2021

Joint work with Daniel Roy

Overview of the Talk

1. Give sufficient and necessary conditions to characterize frequentist optimality using Bayesian notions.
2. Raised by Wald 80 years ago and there exist a huge literature on this problem but they are subject to technical conditions.
3. We have resolved the problem under complete generality using nonstandard analysis.
4. This is the starting point of a programme to rework the foundations of statistical decision theory.

Statistical decision theory framework

Definition

A **statistical decision problem** consists of:

1. Sample space X, parameter space Θ, and action space $\mathcal{A}$
2. Loss function $\ell : \Theta \times \mathcal{A} \to \mathbb{R}_{\geq 0}$
3. Model $\{P_\theta\}_{\theta \in \Theta}$ where $P_\theta \in \mathcal{M}_1(X)$.

Definition

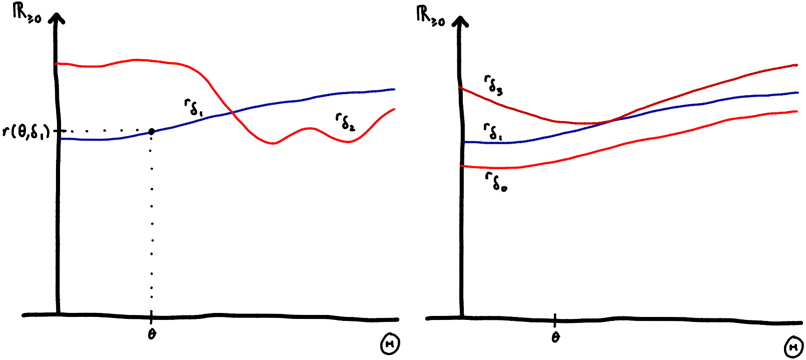
A **randomized decision procedure** is a map $\delta : X \to \mathcal{M}_1(\mathcal{A})$. Write $\delta(x, A)$ for $(\delta(x))(A)$.

$\mathcal{D}$: the set of randomized decision procedures.

Definition

The risk of δ at $\theta \in \Theta$ is $r(\theta, \delta) = \int_X [\int_A \ell(\theta, a)\delta(x, \mathrm{d}a)] P_\theta(\mathrm{d}x)$.

Frequentist admissibility



- δ is ϵ-**dominated** by δ' when

$$(\forall \theta \in \Theta)(r(\theta, \delta') \leq r(\theta, \delta) - \epsilon) \wedge (\exists \theta_0 \in \Theta)(r(\theta_0, \delta') < r(\theta_0, \delta) - \epsilon).$$

- δ is ϵ-**admissible** if it is not ϵ-dominated by any $\delta' \in \mathcal{D}$.
- δ is **admissible** if δ is 0-admissible.
- δ is **extended admissible** if δ is ϵ-admissible for all $\epsilon \in \mathbb{R}_{>0}$.

Bayes optimality

Definition

Let $\delta \in \mathcal{D}$.

- A **prior** is a probability measure on Θ.
- The **Bayes risk** of δ with respect to a prior π is

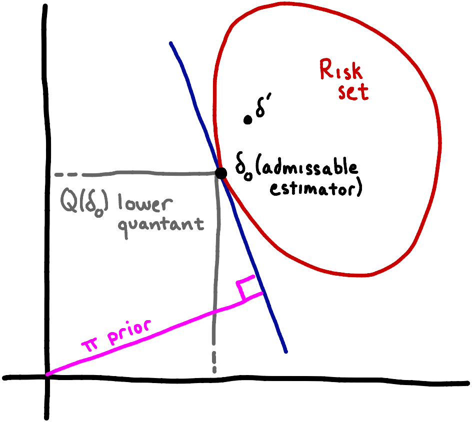
$$r(\pi, \delta) = \int_\Theta r(\theta, \delta)\, \pi(\mathrm{d}\theta).$$

- δ is **Bayes (optimal)** if there exists a prior π such that $r(\pi, \delta) < \infty$ and $r(\pi, \delta) \le r(\pi, \delta')$ for all $\delta' \in \mathcal{D}$.

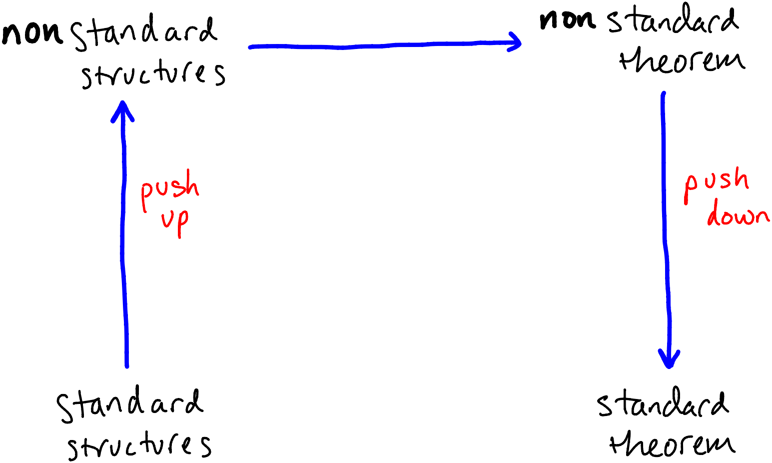
Theorem (Wald)

Suppose Θ is finite. If δ_0 is extended admissible then δ_0 is Bayes.

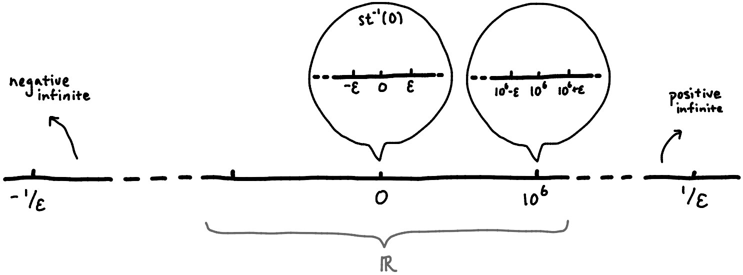
Graphic proof of complete class theorem



risk set $\mathcal{S} = \{(r(\theta_1, \delta), \ldots, r(\theta_n, \delta)) : \delta \in \mathcal{D}\}$
lower quantant $Q(\delta_0) = \{x \in \mathbb{R}^n : (\forall k \leq n)(x_k \leq r(\theta_k, \delta_0))\}$


nonstandard structures
nonstandard theorem
push up
push down
standard structures
standard theorem

Nonstandard real line



1. **Transfer Principle**:
   $\mathbb{R} \models \phi(x_1, ..., x_n) \Leftrightarrow {}^*\mathbb{R} \models {}^*\phi({}^*x_1, ..., {}^*x_n)$.
2. κ-**Saturation Principle**: If a collection of statements are finitely satisfiable, then the whole collection can be satisfied simultaneously.
3. **Infinite and Infinitesimal Number**: There exists $x \in {}^*\mathbb{R}$ such that $|x| > n$ for all $n \in \mathbb{N}$. Then $\frac{1}{x}$ is less than $\frac{1}{n}$ for all $n \in \mathbb{N}$

Two elements $x, y \in {}^*\mathbb{R}$ are **infinitely close**, written $x \approx y$, if $|x - y|$ is infinitesimal.

Hyperfinite Probability Space

Definition

A set A is **hyperfinite** if and only if there exists an internal bijection between A and $\{0, 1, ..., N-1\}$ for some $N \in {}^*\mathbb{N}$. This N is unique and is called the internal cardinality of A.

A hyperfinite probability space is a triple $(\Omega, I(\Omega), P)$ such that

1. Ω is a hyperfinite set.
2. $I(\Omega)$ is the collection of all hyperfinite subsets of Ω.
3. $P : I(\Omega) \rightarrow {}^*[0, 1]$ such that $P(\emptyset) = 0$, $P(\Omega) = 1$ and P is hyperfinitely additive.

Example

Let $T = \{\frac{1}{N}, \frac{2}{N}, \ldots, 1\}$ for some $N \in {}^*\mathbb{N} \setminus \mathbb{N}$. Let $P(\{\omega\}) = \frac{1}{N}$ for every $\omega \in T$. The Lebesgue measure of $A \subset [0, 1]$ is the same as "counting" the number of points in T that are infinitely close to some points in A.

Nonstandard Bayes optimality

Definition

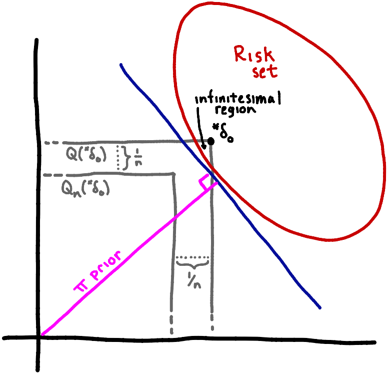
Let $\delta \in \mathcal{D}$.

- An **internal prior** is a *probability measure on $^*\Theta$.
- The **internal Bayes risk** of $\delta \in \mathcal{D}$ with respect to an internal prior π is $^*r(\pi, {}^*\delta) = \int_{^*\Theta} {}^*r(t, {}^*\delta)\, \pi(\mathrm{d}t)$.
- δ is **nonstandard Bayes** if there is an internal prior π such that $^*r(\pi, {}^*\delta) \lessapprox {}^*r(\pi, {}^*\delta')$ for all $\delta' \in \mathcal{D}$.

Theorem (Haosui–Roy)

δ_0 is extended admissible if and only if δ_0 is nonstandard Bayes.

Nonstandard complete class theorem



By saturation, $T_\Theta = \{t_1, t_2, \dots, t_K\}$ and $\Theta \subset T_\Theta$.
Define $Q(\Delta)_n = \{x \in I({}^*\mathbb{R}^K) : (\forall k \le K)(x_k \le {}^*r(t_k, \Delta) - \frac{1}{n}\}$.
By ext. admissibility, exists hyperplane separating $Q({}^*\delta_0)_n$ and risk set for all $n \in \mathbb{N}$.
By saturation, exists hyperplane Π separating $\bigcup_{n\in\mathbb{N}} Q({}^*\delta_0)_n$ and risk set.
Hence, ${}^*\delta_0$ is nonstandard Bayes w.r.t Π normalized.

Constructing Standard Measures from Hyperfinite Measures

Hyperfinite probability spaces can be extended to standard countably additive probability spaces. In particular, let Θ be a Hausdorff space endowed with Borel σ-algebra $\mathcal{B}[\Theta]$ and Π be a nonstandard probability measure on ${}^*\Theta$. We can construct:

1. A countably additive measure Π_p on $(\Theta, \mathcal{B}[\Theta])$. If Θ is compact, then Π_p is a probability measure.
2. A finitely additive probability measure Π^p on $(\Theta, \mathcal{B}[\Theta])$.

Example

Let $N \in {}^*\mathbb{N} \setminus \mathbb{N}$. Let Π be an internal probability measure concentrating on $\{\frac{1}{N}\}$. Then, Π_p is the degenerate measure on $\{0\}$ and Π^p is a finitely additive probability measure with $\Pi^p((0, \frac{1}{n})) = 1$ for every $n \in \mathbb{N}$.

Nonstandard Bayes risk and Standard Bayes risk

The following results establish connections between a nonstandard prior Π, Π_p and Π^p.

Theorem (Haosui–Roy)

Suppose Θ is compact Hausdorff and risk functions are continuous. Let Π be an internal probability measure on T_Θ. Let $\delta_0 \in \mathcal{D}$. Then ${}^*r(\Pi, {}^*\delta_0) \approx r(\Pi_p, \delta_0)$, i.e., the Bayes risk of δ_0 with respect to Π_p is equal to the nonstandard Bayes risk of δ_0 with respect to Π up to an infinitesimal.

Theorem (Haosui–Roy)

Suppose risk functions are bounded. Let Π be an internal probability measure on T_Θ. Let $\delta_0 \in \mathcal{D}$. Then ${}^*r(\Pi, {}^*\delta_0) \approx r(\Pi^p, \delta_0)$, i.e., the Bayes risk of δ_0 with respect to Π^p is equal to the nonstandard Bayes risk of δ_0 with respect to Π up to an infinitesimal.

Standard results

Theorem (Haosui–Roy)

δ_0 is extended admissible among $\mathcal{D}$ if and only if δ_0 is nonstandard Bayes among $\mathcal{D}$.

Strong connections between Π, Π_p and Π^p generate the following two purely standard results.

Theorem (Haosui–Roy)

Suppose Θ is compact Hausdorff and risk functions are continuous. Then δ_0 is extended admissible among $\mathcal{D}$ if and only if δ_0 is Bayes among $\mathcal{D}$.

Theorem (Sudderth–Heath)

Suppose Θ is measurable and risk functions are bounded. Then δ_0 is extended admissible among $\mathcal{D}$ if and only if δ_0 is f.a Bayes among $\mathcal{D}$.

Conclusion

In this talk, I have

1. defined the term of nonstandard Bayes optimality.
2. showed that extended admissibility is equivalent to nonstandard Bayes optimality.
3. established two standard results using our nonstandard theory.

This is the first step to rebuild foundation of statistical decision theory using nonstandard analysis.