On Extended Admissible Procedures and their Nonstandard Bayes Risk

Haosui "Kevin" Duanmu and Daniel M. Roy
University of Toronto

Fourth Bayesian, Fiducial, and Frequentist Conference
Harvard University

+

Statistical Decision Theory (Wald)

Defn (risk). $r_d(\theta) = r(\theta, d) = \mathop{\mathbb{E}}_{X \sim P_\theta}[L(\theta, d(X))] \quad \theta \in \Theta,\ d \in \mathcal{D}_0$

Defn (convex extension). $r(\pi, \delta) = \mathop{\mathbb{E}}_{\substack{\theta \sim \pi \\ d \sim \delta}}[r(\theta, d)] \quad \pi \in \mathcal{P},\ \delta \in \mathcal{D}$

Defn. δ_0 *minimax* iff $\sup_\pi r(\pi, \delta_0) = \underbrace{\inf_\delta \sup_\pi r(\pi, \delta)}_{\text{minimax risk}}$

Defn. δ_0 *Bayes* iff $(\exists \pi)\ r(\pi, \delta_0) = \underbrace{\inf_\delta r(\pi, \delta)}_{\text{Bayes risk}}$

Defn. δ_0 *admissible* iff $\neg(\exists \delta) \quad r_\delta \le r_{\delta_0} \quad \text{and } r_\delta \ne r_{\delta_0}$

Thm. (Wald) If the parameter space Θ is finite, then admissible implies Bayes.

Proof. Let $\Theta = \{\theta_1, \dots, \theta_J\}$. Then

$$r_\delta = (r_\delta(\theta_1), \dots, r_\delta(\theta_J)) \in \mathbb{R}^J.$$

Define the *risk set* $\mathcal{S}$ of $\mathcal{D}$:

$$\mathcal{S} = \{r_\delta \in \mathbb{R}^J : \delta \in \mathcal{D}\}.$$

Let δ_0 be admissible and define *lower quantant* $Q(\delta_0) = \{x \in \mathbb{R}^J : x \leq r_{\delta_0}\}$.

Claim. $\mathcal{S}$ and $Q(\delta_0)$ intersect at one point, r_{δ_0}.

Generalizing admissibility and Bayes

Defn. δ_0 *extended admissible* iff $(\forall \epsilon > 0) \quad \neg(\exists \delta) \quad r_\delta \leq r_{\delta_0} - \epsilon$

Defn. δ_0 *extended Bayes* iff $(\forall \epsilon > 0)\ (\exists \pi)\ \underbrace{r(\pi, \delta_0) - \inf_\delta r(\pi, \delta)}_{\text{excess risk}} \leq \epsilon$

Thm (BG54). Extended Bayes implies extended admissible.
Proof by picture. Via contrapositive: not extended admissible implies not extended Bayes.

Set-theoretic relationships

M : minimax
A : admissible
A_0 : extended admissible
B : Bayes
B_0 : extended Bayes

Thm. 5.5.3 (BG54). Assuming only that Θ is finite, the diagram (right) cannot be simplified.

Thm. 5.5.1 (BG54). Assuming only bounded risk, the diagram (left) cannot be simplified.

Thm. 5.5.3 (BG54). Assuming only bounded risk and minimax=maximin for every derived game, the diagram (right) cannot be simplified.

Results beyond bounded risk / finite Θ

Thm. (Wald, LeCam, Brown) Assume P admits strictly positive densities $(f_\theta)_{\theta\in\Theta}$ with respect to a σ-finite measure μ. Assume the action space $\mathcal{A}$ is a closed convex subset of Euclidean space. Assume the loss $L(\theta, a)$ is lower semicontinuous and strictly convex in a for every θ, and satisfies

$$\lim_{|a|\to\infty} L(\theta, a) = \infty \text{ for all } \theta \in \Theta.$$

Then, for every admissible decision procedure δ, there exists a sequence π_n of priors with support on a finite set, such that

$$\delta^{\pi_n}(x) \to \delta(x) \text{ as } n \to \infty \text{ for } \mu\text{-almost all } x,$$

where δ^{π_n} is a Bayes procedure with respect to π_n.

Defn. Suppose P admits densities $(f_\theta)_{\theta\in\Theta}$ with respect to a σ-finite measure μ. An estimator δ_0 is *generalized Bayes* with respect to a σ-finite measure π on Θ if it minimizes the unnormalized posterior risk $\int L(\theta, \delta_0(x))\, f_\theta(x)\, \pi(\mathrm{d}\theta)$ for μ-a.e. x.

Thm. (Berger–Srinivasan) Assume P is a multi-dimensional exponential family, and that the loss $L(\theta, a)$ is jointly continuous, strictly convex in a for every θ, and satisfies

$$\lim_{|a|\to\infty} L(\theta, a) = \infty \text{ for all } \theta \in \Theta.$$

Then every admissible estimator is generalized Bayes.

First-order logic

Example. (bounded quantifier formulas)

$$\phi(\delta, \pi) = (\forall \delta' \in \mathcal{D})\,(r(\pi, \delta) \leq r(\pi, \delta'))$$
$$\phi'(\delta) = (\exists \pi \in \mathcal{P})\,(\phi(\delta, \pi))$$

Example. (normal-location model) Define $\delta_B(v, x) = \frac{v}{v+1}x$. Then $\delta_B(v, \cdot)$ is Bayes w.r.t. $\mathcal{N}(0, vI)$ prior, for all $v > 0$. In logic,

$$\phi'' = (\forall v \in \mathbb{R}_{>0})\phi(\delta_B(v, \cdot), \mathcal{N}(0, vI))$$

Nonstandard analysis

Three mechanisms: **extension**, **transfer**, κ-**saturation**.

Defn (internal). An element b is *internal* if $b \in {}^*A$ for some standard set A.
Defn (κ-saturation). Let $(A_i)_{i \in J}$ be a collection of *internal* sets, where J has cardinality less than κ. Then $\bigcap_{i \in J} A_i$ is nonempty if $\bigcap_{i \in F} A_i$ is nonempty for every finite $F \subseteq J$.

Some elementary applications

Example (transfer). What do elements of ${}^*\mathbb{R}$ look like? They satisfy exactly the first order properties satisfied by the standard reals, $\mathbb{R}$.

How about the relation ${}^*\leq$? By transfer, we know it's a total order.

Example. (normal-location model) By transfer, ${}^*\phi''$ holds, i.e.,

$$(\forall v \in {}^*\mathbb{R}_{>0})((\forall \delta' \in {}^*\mathcal{D})({}^*r({}^*\mathcal{N}(0, vI), {}^*\delta_B(v, \cdot)){}^*\leq {}^*r({}^*\mathcal{N}(0, vI), \delta')))$$

Key point: Without saturation, we get nothing new.

Example (saturation). Assume $\aleph_1$ -saturation.
Let $A_k = \{v \in \mathbb{R}_{>0} : v \geq k\}$.
Then $A_k \supset A_{k+1}$ and $\bigcap_{n\in\mathbb{N}} A_n = \emptyset$. In contrast, $A = \cap_{k\in\mathbb{N}} {}^*A_k \neq \emptyset$.

An element $v \in A \subseteq {}^*\mathbb{R}$ is an *infinite positive real* and so ${}^*\mathbb{R} \supsetneq \mathbb{R}$.
$\epsilon = \frac{1}{v}$ is an *infinitesimal*, i.e., $\epsilon < \frac{1}{n}$ for all $n \in \mathbb{N}$.
Note that $\epsilon > 0$, hence $\epsilon^2 > 0$, by transfer.

Defn. For $x, y \in {}^*\mathbb{R}$, write $x \approx y$ if $|x - y|$ is infinitesimal.

Example (normal-location model). Recall the MLE estimator, $\delta_{MLE}(x) = x$. For v infinite, ${}^*\delta_B(v, x) = \frac{v}{v+1}x \approx x = {}^*\delta_{MLE}(x)$.

Recap: Bayes optimality

Defn. Let $\delta \in \mathcal{D}$ and $\mathcal{C} \subset \mathcal{D}$.

1. A *prior* is a probability measure $\pi \in \mathcal{P}$.
2. The *average risk* of δ with respect to a prior π is
$$r(\pi, \delta) = \int_\Theta r(\theta, \delta)\, \pi(\mathrm{d}\theta).$$
3. δ is *Bayes (optimal) among $\mathcal{C}$* if, for some prior π, $r(\pi, \delta) < \infty$ and $r(\pi, \delta) \le r(\pi, \delta')$ for all $\delta' \in \mathcal{C}$.

Thm. Bayes among $\mathcal{C}$ $\implies$ extended admissible among $\mathcal{C}$.

Nonstandard Bayes optimality

Defn. Let $\delta \in \mathcal{D}$ and $\mathcal{C} \subset \mathcal{D}$.

1. A *nonstandard prior* is a * probability measure $\pi \in {}^*\mathcal{P}$.
2. The internal *average risk* of $\Delta \in {}^*\mathcal{D}$ with respect to a nonstandard prior π is
$${}^*r(\pi, \Delta) = {}^*\!\!\int_{^*\Theta} {}^*r(\theta, \Delta)\, \pi(\mathrm{d}\theta).$$
3. Δ is *nonstandard Bayes among $\mathcal{C} \subseteq {}^*\mathcal{D}$* if, for some nonstandard prior π, ${}^*r(\pi, \Delta) < \infty$ and ${}^*r(\pi, \Delta) \lessapprox {}^*r(\pi, \Delta')$ for all $\Delta' \in \mathcal{C}$.

Theorem (Haosui–Roy).
${}^*\delta_0$ nonstandard Bayes among $\mathcal{C}^\sigma = \{{}^*\delta : \delta \in \mathcal{C}\}$
$\implies$ δ_0 extended admissible among $\mathcal{C}$.

Example (normal-location model). Choose $\nu \in {}^*\mathbb{R}_{>0}$ infinite.
By transfer, ${}^*\delta_B(\nu, \cdot)$ is * Bayes w.r.t. ${}^*\mathcal{N}(0, \nu I)$.
By transfer, ${}^*r({}^*\mathcal{N}(0, \nu I), {}^*\delta_B(\nu, \cdot)) = d\frac{\nu}{\nu+1}$.
By transfer, ${}^*r({}^*\mathcal{N}(0, \nu I), {}^*\delta_{MLE}) = d$.
But $d \approx d\frac{\nu}{\nu+1}$ implies ${}^*\delta_{MLE}$ is nonstandard Bayes among
$\mathcal{D}^\sigma = \{{}^*\delta : \delta \in \mathcal{D}\}$. Hence it is extended admissible among $\mathcal{D}$.

${}^*\mathcal{N}(0, \nu I_d)$ is "flat": $\dfrac{(2\pi)^{-\frac{d}{2}}\nu^{-\frac{d}{2}}}{(2\pi)^{-\frac{d}{2}}\nu^{-\frac{d}{2}}\exp\{-\frac{1}{\nu}\|x\|_2^2\}} \approx 1,$ for $x \in \mathrm{NS}(\mathbb{R}^d)$.

*finite (aka hyperfinite) sets

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

Lemma. There is a hypefinite set $T \subset {}^*\Theta$ such that $\Theta \subset T$.
Proof. For every finite set $A \subset \Theta$, let $\phi(A)$ be the sentence

There is a hyperfinite set containing A.

By saturation, there is a hyperfinite set containing Θ as a subset.

Main result

Theorem (Haosui–Roy).
δ_0 extended admissible among $\mathcal{D}$ $\iff$
${}^*\delta_0$ nonstandard Bayes among $\mathcal{D}^\sigma$.

Proof. By saturation, exists hyperfinite $T_\Theta \subset {}^*\Theta$ containing Θ.
Let $T_\Theta = \{t_1, \ldots, t_{J_\Theta}\}$.

Define the *hyperdiscretized risk set induced by $C \subseteq {}^*\mathcal{D}$*:

$$\mathcal{S}^C = \{x \in \mathrm{I}({}^*\mathbb{R}^{J_\Theta}) : (\exists \Delta \in C)\,(\forall k \leq J_\Theta)\, x_k = {}^*r(t_k, \Delta)\}.$$

Note $\mathcal{D}^\sigma$ is not convex over ${}^*\mathbb{R}$.

$$\text{Define} \quad (\mathcal{D}_0^\sigma)_{FC} = \bigcup_C {}^*\mathrm{conv}(C)$$

where C ranges over finite subsets of $\mathcal{D}_0^\sigma$.

Note: $(\mathcal{D}_0^\sigma)_{FC} \supseteq \mathcal{D}^\sigma$ and is convex over ${}^*\mathbb{R}$ and

$$\mathcal{S}^{(\mathcal{D}_0^\sigma)_{FC}} = \bigcup_C \mathcal{S}^C.$$

For $\Delta \in {}^*\mathcal{D}$ and $n \in {}^*\mathbb{N}$, define

$$Q(\Delta)_n = \{x \in \mathrm{I}({}^*\mathbb{R}^{J_\Theta}) : (\forall k \leq J_\Theta)(x_k \leq {}^*r(t_k, \Delta) - \frac{1}{n})\}.$$

Claim. If δ_0 is extended admissible among $\mathcal{D}$, then $Q({}^*\delta_0)_n$ and $\mathcal{S}^{(\mathcal{D}_0^\sigma)_{FC}}$ do not intersect.

A standard result via nonstandard theory

Defn Let $r \in {}^*\mathbb{R}$. If there exists $x \in \mathbb{R}$ such that $x \approx r$ then x is called the *standard part of r*, denoted $\mathsf{st}(r)$.

st is a partial function from ${}^*\mathbb{R}$ to $\mathbb{R}$, called the *standard part map*.

Example. Consider a set $E \subset \mathbb{R}$:

$$\mathsf{st}^{-1}(E) = \{r \in {}^*\mathbb{R} : (\exists x \in E)(r \approx x)\}.$$

$\mathsf{st}^{-1}(E)$, in general, is not an internal set.

Internal probability measures $\Rightarrow$ standard probability measures

Theorem (Cutland–Neves–Oliveira–Sousa-Pinto). Let $(X, \mathcal{B}[X])$ be a compact Hausdorff Borel measurable space. Let π be an internal probability measure on $({}^*X, {}^*\mathcal{B}[X])$. Define $\pi_p : \mathcal{B}[X] \to [0,1]$ by

$$\pi_p(B) = \sup\{\mathsf{st}(\pi(A)) : A \subset \mathsf{st}^{-1}(B) \wedge A \in {}^*\mathcal{B}[X]\}, \quad B \in \mathcal{B}[X].$$

Then π_p is a standard Borel probability measure.

π_p is called the *push down of* π.

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 $\{0\}$.

Theorem (Haosui–Roy). Suppose Θ is compact Hausdorff and risk functions are continuous. Let π be an internal probability measure on T_Θ and let π_p be its push-down. Let $\delta_0 \in \mathcal{D}$ be a standard decision procedure. Then ${}^*r(\pi, {}^*\delta_0) \approx r(\pi_p, \delta_0)$.

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 $\mathcal{D}$.

A nonstandard Blyth's method

Defn. For $x, y \in {}^*\mathbb{R}$, write $x \gg y$ when $\gamma x > y$ for all $\gamma \in \mathbb{R}_{>0}$.

Defn. Let (X, d) be a metric space, and let $\epsilon \in {}^*\mathbb{R}_{>0}$. An internal probability measure π on ${}^*\Theta$ is *ϵ-regular* if, for every $\theta_0 \in \Theta$ and non-infinitesimal $r > 0$, $\pi(\{t \in {}^*\Theta : {}^*d(t, \theta_0) < r\}) \gg \epsilon$.

Theorem. Suppose Θ is metric, risk functions are continuous, and let $\delta_0 \in \mathcal{D}$ and $\mathcal{C} \subseteq \mathcal{D}$. If there exists $\epsilon \in {}^*\mathbb{R}_{>0}$ such that ${}^*\delta_0$ is within ϵ of the optimal * Bayes risk among $\mathcal{C}^\sigma = \{{}^*\delta : \delta \in \mathcal{C}\}$ with respect to some ϵ-regular nonstandard prior, then δ_0 is admissible among $\mathcal{C}$.

Example (normal-location problem). Choose $v \in {}^*\mathbb{R}_{>0}$ infinite.
Recall that ${}^*\mathcal{N}(0, vI_d)$ is "flat" on $\mathrm{NS}(\mathbb{R}^d)$:

$$\frac{(2\pi)^{-\frac{d}{2}} v^{-\frac{d}{2}}}{(2\pi)^{-\frac{d}{2}} v^{-\frac{d}{2}} \exp\{-\frac{1}{v}\|x\|_2^2\}} \approx 1, \qquad \text{for } x \in \mathrm{NS}(\mathbb{R}^d).$$

Bayes risk of ${}^*\delta_B(v, \cdot)$ is $d\frac{v}{v+1}$.
Bayes risk of ${}^*\delta_{MLE}$ is d.
Thus, excess risk is $\epsilon = (v+1)^{-1}$.

For $d = 1$, ${}^*\mathcal{N}(0, vI_d)$ is ϵ-regular.
Thus δ_{MLE} is admissible for $d = 1$, as is well known.

For $d \geq 2$, ${}^*\mathcal{N}(0, vI_d)$ is not ϵ-regular. Thus theorem is silent.

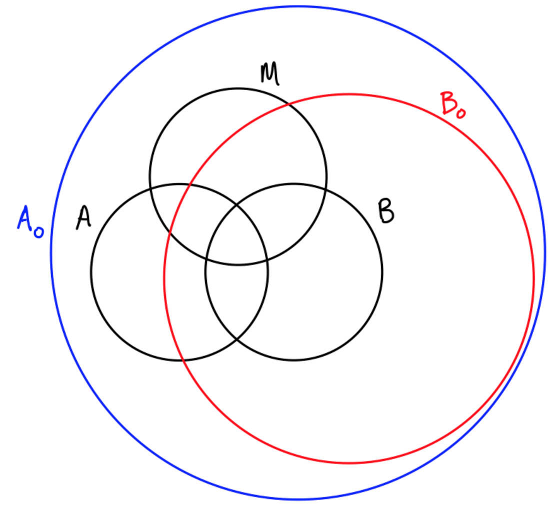
Thm (Stein). δ_{MLE} is admissible only for $d = 1, 2$.

Summary of main results

Theorem (Haosui–Roy).
δ_0 extended admissible among $\mathcal{D}$ $\iff$
${}^*\delta_0$ nonstandard Bayes among $\mathcal{D}^\sigma$.

Lemma (Haosui–Roy).
δ_0 extended Bayes among $\mathcal{D}$ $\iff$
${}^*\delta_0$ nonstandard Bayes among ${}^*\mathcal{D}$.

Lemma (Haosui–Roy).
δ_0 generalized Bayes among $\mathcal{D}$ $\implies$
${}^*\delta_0$ nonstandard Bayes among $\mathcal{D}^\sigma$.



Conclusion

- By working in a saturated models of the reals, a notion of Bayes optimality aligns perfectly with extended admissibility.
- Our results come without conditions other than saturation, and so they can be used to study infinite dimensional nondominated models with unbounded risk beyond the remit of existing results
- The nonstandard Blyth method points the way towards necessary conditions for admissibility.
- There's hope that more of frequentist and Bayesian theory can be aligned using similar techniques.

Another example

Example. Let $X = \{0, 1\}$ and $\Theta = [0, 1]$.
Define $g : [0, 1] \to [0, 1]$ by $g(x) = x$ for $x > 0$ and $g(0) = 1$.
Let $P_t = \text{Bernoulli}(g(t))$, for $t \in [0, 1]$.
Consider the loss function $\ell(x, y) = (g(x) - y)^2$.

Every nonrandomized decision procedure $\delta : \{0, 1\} \to [0, 1]$ corresponds with a pair $(\delta(0), \delta(1)) \in [0, 1]^2$.

Note: Loss is merely lower semicontinous. Model also not continuous.

Thm. $(0, 0)$ is an admissible non-Bayes estimator.

Thm. ${}^*(0, 0)$ is nonstandard Bayes with respect to any prior concentrating on some infinitesimal $\epsilon > 0$.

Lem. $(0, 0)$ is a generalized Bayesian estimator with respect to the improper prior $\pi(\mathrm{d}\theta) = \theta^{-2}\mathrm{d}\theta$.