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). \( r_d(\theta) = r(\theta, d) = \mathbb{E}_{x \sim P_\theta} [L(\theta, d(X))] \quad \theta \in \Theta, \ d \in D_0 \)

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

\[
\begin{align*}
\text{Defn. } \delta_0 & \text{ minimax iff } \sup_{\pi} r(\pi, \delta_0) = \inf_{\pi} \sup_{\delta} r(\pi, \delta) \\
& \text{Bayes risk} \ \text{minimax risk}
\end{align*}
\]

\[
\begin{align*}
\text{Defn. } \delta_0 & \text{ Bayes iff } (\exists \pi) \ r(\pi, \delta_0) = \inf_{\delta} r(\pi, \delta) \\
& \text{Bayes risk}
\end{align*}
\]

\[
\begin{align*}
\text{Defn. } \delta_0 & \text{ admissible iff } \neg (\exists \delta) \ r_\delta \leq r_{\delta_0} \ \text{and } r_\delta \neq r_{\delta_0}
\end{align*}
\]
Thm. (Wald) If the parameter space $\Theta$ is finite, then admissible implies Bayes.

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

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

Define the risk set $S$ of $D$: 

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

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

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

\[ Q(\delta_0) \quad \text{lower quantant} \]

\[ S \quad \text{risk set} \]

\[ r_{\delta_0} \quad \text{(admissible estimator)} \]
Generalizing admissibility and Bayes

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

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

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 $\Theta$ 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 $\sigma$-finite measure $\mu$. Assume the action space $A$ is a closed convex subset of Euclidean space. Assume the loss $L(\theta, a)$ is lower semicontinuous and strictly convex in $a$ for every $\theta$, and satisfies

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

Then, for every admissible decision procedure $\delta$, there exists a sequence $\pi_n$ 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 $\delta^{\pi_n}$ is a Bayes procedure with respect to $\pi_n$.

**Defn.** Suppose $P$ admits densities $(f_\theta)_{\theta \in \Theta}$ with respect to a $\sigma$-finite measure $\mu$. An estimator $\delta_0$ is **generalized Bayes** with respect to a $\sigma$-finite measure $\pi$ on $\Theta$ if it minimizes the unnormalized posterior risk

$$\int L(\theta, \delta_0(x)) f_\theta(x) \pi(d\theta)$$

for $\mu$-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 $\theta$, 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)
\[
\phi(\delta, \pi) = (\forall \delta' \in D) (r(\pi, \delta) \leq r(\pi, \delta')) \\
\phi'(\delta) = (\exists \pi \in P) (\phi(\delta, \pi))
\]

Example. (normal-location model) Define \(\delta_B(v, x) = \frac{v}{v+1} x\). Then \(\delta_B(\nu, \cdot)\) is Bayes w.r.t. \(N(0, vI)\) prior, for all \(v > 0\). In logic,
\[
\phi'' = (\forall \nu \in \mathbb{R}_{>0}) \phi(\delta_B(\nu, \cdot), N(0, vI))
\]

Nonstandard analysis

Three mechanisms: \textit{extension, transfer, }\kappa\textit{-saturation}.

**Defn (internal).** An element \(b\) is \textit{internal} if \(b \in \star A\) for some standard set \(A\).

**Defn (\(\kappa\)-saturation).** Let \(\langle A_i \rangle_{i \in J}\) be a collection of \textit{internal} sets, where \(J\) has cardinality less than \(\kappa\). 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.,

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

Key point: Without saturation, we get nothing new.

Example (saturation). Assume $\mathbb{N}_1$-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 $v \in A \subset ^{*}\mathbb{R}$ is an infinite positive real and so $^{*}\mathbb{R} \supset \mathbb{R}$.

$\epsilon = \frac{1}{v}$ 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 $\delta \in D$ and $C \subset D$.

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

**Thm.** Bayes among $C$ $\implies$ extended admissible among $C$.

**Nonstandard Bayes optimality**

**Defn.** Let $\delta \in D$ and $C \subset D$.

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

**Theorem (Haosui–Roy).**

$\ast \delta_0$ nonstandard Bayes among $C^\sigma = \{ \ast \delta : \delta \in C \}$ $\implies$ $\delta_0$ extended admissible among $C$. 
Example (normal-location model). Choose $v \in \mathbb{R}_{>0}$ infinite.

By transfer, $*\delta_B(v, \cdot)$ is Bayes w.r.t. $*\mathcal{N}(0, vI)$. By transfer, $r(\mathcal{N}(0, vI), *\delta_B(v, \cdot)) = d \frac{v}{v+1}$.

By transfer, $r(\mathcal{N}(0, vI), *\delta_{MLE}) = d$. But $d \approx d \frac{v}{v+1}$ implies $*\delta_{MLE}$ is nonstandard Bayes among $D^* = \{ *\delta : \delta \in D \}$. Hence it is extended admissible among $D$.

$*\mathcal{N}(0, vI_d)$ is "flat": $\frac{(2\pi)^{-\frac{d}{2}} v^{-\frac{d}{2}}}{(2\pi)^{-\frac{d}{2}} v^{-\frac{d}{2}} \exp\left\{-\frac{1}{v} ||x||_2^2\right\}} \approx 1$, for $x \in \text{NS}(\mathbb{R}^d)$. 
*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 hyperfinite set $T \subset \mathbb{N}^*$ such that $\emptyset \subset T$.

**Proof.** For every finite set $A \subset \emptyset$, let $\phi(A)$ be the sentence

There is a hyperfinite set containing $A$.

By saturation, there is a hyperfinite set containing $\emptyset$ as a subset.

**Main result**

**Theorem (Haosui–Roy).**

$\delta_0$ extended admissible among $D$ $\iff$ $\delta_0$ nonstandard Bayes among $D^\sigma$.

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

Define the **hyperdiscretized risk set induced by $C \subset \emptyset$**: $S^C = \{x \in I(\mathbb{R}^{J_\emptyset} : (\exists \Delta \in C) (\forall k \leq J_\emptyset) x_k = *r(t_k, \Delta)\}$.

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

Define $(D_0^\sigma)_{FC} = \bigcup * \text{conv}(C)$

where $C$ ranges over finite subsets of $D_0^\sigma$.

Note: $(D_0^\sigma)_{FC} \supseteq D^\sigma$ and is convex over $\mathbb{R}$ and $S(D_0^\sigma)_{FC} = \bigcup S^C$. 

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

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

Claim. If $\delta_0$ is extended admissible among $D$, then $Q(\delta_0)_n$ and $S^{(D_0)_{fc}}$ 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 \( \text{st}(r) \).

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

**Example.** Consider a set \( E \subset \mathbb{R} \):

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

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

**Internal probability measures ⇒ standard probability measures**

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

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

Then \( \pi_p \) is a standard Borel probability measure.

\( \pi_p \) is called the **push down of** \( \pi \).

**Example.** Let \( N \in {}^*\mathbb{N} \setminus \mathbb{N} \). Let \( \pi \) be an internal probability measure concentrating on \( \{ \frac{1}{N} \} \). Then \( \pi_p \) is the degenerate measure on \( \{0\} \).

**Theorem** (Haosui–Roy). Suppose \( \Theta \) is compact Hausdorff and risk functions are continuous. Let \( \pi \) be an internal probability measure on \( T_{\Theta} \) and let \( \pi_p \) 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 \( \Theta \) is compact Hausdorff and risk functions are continuous. Then \( \delta_0 \) is extended admissible among \( \mathcal{D} \) if and only if \( \delta_0 \) is Bayes among \( \mathcal{D} \).
A nonstandard Blyth's method

Defn. For $x, y \in \ast \mathbb{R}$, 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 $\epsilon \in \ast \mathbb{R}_{>0}$. An internal probability measure $\pi$ on $\ast \Theta$ is $\epsilon$-regular if, for every $\theta_0 \in \Theta$ and non-infinitesimal $r > 0$,

$$\pi(\{t \in \ast \Theta : \ast d(t, \theta_0) < r\}) \gg \epsilon.$$

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

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

$$\frac{(2\pi)^{-d/2} \nu^{-d/2}}{(2\pi)^{-d/2} \nu^{-d/2} \exp\{-\frac{1}{\nu} ||x||^2\}} \approx 1, \quad \text{for } x \in \text{NS}(\mathbb{R}^d).$$

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

For $d = 1$, $\ast \mathcal{N}(0, \nu I_d)$ is $\epsilon$-regular.
Thus $\delta_{MLE}$ is admissible for $d = 1$, as is well known.

For $d \geq 2$, $\ast \mathcal{N}(0, \nu I_d)$ is not $\epsilon$-regular. Thus theorem is silent.

Thm (Stein). $\delta_{MLE}$ is admissible only for $d = 1, 2$.
Summary of main results

Theorem (Haosui–Roy).
\( \delta_0 \) extended admissible among \( D \iff *\delta_0 \) nonstandard Bayes among \( D^\sigma \).

Lemma (Haosui–Roy).
\( \delta_0 \) extended Bayes among \( D \iff *\delta_0 \) nonstandard Bayes among \( *D \).

Lemma (Haosui–Roy).
\( \delta_0 \) generalized Bayes among \( D \implies *\delta_0 \) nonstandard Bayes among \( D^\sigma \).

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 $\mathcal{L}(x, y) = (g(x) - y)^2$.

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 semicontinuous. 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 $c > 0$.

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