

ON EXTENDED ADMISSIBLE PROCEDURES AND THEIR NONSTANDARD BAYES RISK

BY HAOSUI DUANMU AND DANIEL M. ROY

Department of Statistical Sciences, University of Toronto

For finite parameter spaces, among decision procedures with finite risk functions, a decision procedure is extended admissible if and only if it is Bayes. Various relaxations of this classical equivalence have been established for infinite parameter spaces, but these extensions are each subject to technical conditions that limit their applicability, especially to modern (semi and nonparametric) statistical problems. Using results in mathematical logic and nonstandard analysis, we extend this equivalence to *arbitrary* statistical decision problems: informally, we show that, among decision procedures with finite risk functions, a decision procedure is extended admissible if and only if it has infinitesimal excess Bayes risk. In contrast to existing results, our equivalence holds in complete generality, that is, *without* regularity conditions or restrictions on the model or loss function. We also derive a nonstandard analogue of Blyth's method that yields sufficient conditions for admissibility, and apply the nonstandard theory to derive a purely standard theorem: when risk functions are continuous on a compact Hausdorff parameter space, a procedure is extended admissible if and only if it is Bayes.

1. Introduction. Among the key contributions of statistical decision theory [3, 6, 7, 19, 23, 26, 28–34] are notions of frequentist and Bayesian optimality and connections between them. In one direction, under suitable technical conditions, every admissible procedure is Bayes under a carefully chosen prior, improper prior, or sequence thereof. The resulting (quasi-)Bayesian interpretation provides insight into the strengths and weaknesses of the procedure from an average-case perspective. In the other direction, necessary and/or sufficient conditions for admissibility expressed in terms of (generalized) priors point us toward Bayesian procedures with good frequentist properties.

As the literature stands, connections between frequentist and Bayesian optimality are subject to technical conditions, and these technical conditions (see Section 3) often rule out semiparametric problems and regularly rule out nonparametric problems. As a result, the relationship between frequentist and Bayesian optimality in the setting of modern statistical decision problems is often uncharacterized. Indeed, given the plethora of technical conditions, it would not be unreasonable to presume the connection between frequentist and Bayesian optimality was, to some extent, fragile and might, in general, fail to hold in nonparametric settings.

Using results in an area of mathematical logic known as nonstandard analysis, we identify an equivalence between the frequentist notion of extended admissibility (a necessary condition of both admissibility and minimaxity) and a novel notion of Bayesian optimality, and show that this equivalence holds in *arbitrary* decision problems *without* technical conditions: informally, our main result (Theorem 5.18) establishes that, among decision procedures with finite risk functions, a decision procedure δ is extended admissible if and only if it has infinitesimal excess Bayes risk.

Received June 2017; revised June 2020.

MSC2020 subject classifications. Primary 62C07, 62C10, 62A01; secondary 28E05.

Key words and phrases. Decision theory, complete class theorems, nonstandard analysis.

Our main result also yields new (standard) results in statistical decision theory, including characterizations of the class of Bayes procedures under additional regularity (Section 6), new insight into the class of finitely additive Bayes procedures (Section 7), and novel sufficient conditions for admissibility. Our findings suggest that the application of nonstandard analysis may yield other connections between frequentist and Bayesian optimality.

1.1. Nonstandard analysis. Rather than considering generalized priors or sequences of priors, we work within the standard Bayesian theory but carry out that work in an unusual setting. In particular, we rely on results in mathematical logic that establish the existence of *nonstandard* models satisfying three principles: *extension*, which associates every ordinary mathematical object with a nonstandard counterpart called its extension; *transfer*, which permits us to use first-order logic to relate standard and nonstandard objects; and *saturation*, which gives us a powerful mechanism for proving the existence of nonstandard objects in terms of finitely satisfiable collections of first-order formulas.

The power of nonstandard models for statistical decision theory comes from two sources: first, every nonstandard model possesses nonstandard real numbers, including infinitesimal / infinite positive real numbers that are, respectively, smaller than / larger than every standard positive real number. Nonstandard real numbers have found numerous applications within probability theory. Halpern [16], for example, uses nonstandard probabilities to define conditional probabilities for null events.

In our setting, we can use nonstandard real numbers to construct a much broader set of (nonstandard) prior distributions. For example, we can construct uniform probability measures over infinite intervals that contain the entire standard real line, or construct probability measures on the positive real line concentrating all their mass on a positive infinitesimal. As priors, nonstandard probability distributions can express extreme states of uncertainty that cannot be expressed by any standard prior.

As for the second source of power, in a suitably saturated model of the reals, the set of standard real numbers is contained within a *hyperfinite* set, that is, an infinite set that nonetheless possesses all the first-order properties of a standard finite set. Elementary arguments for finite problems can often be recast to the hyperfinite setting, yielding new results.

1.2. Our results. Using nonstandard analysis and nonstandard probability theory, we are able to establish a link between extended admissibility and Bayesian optimality without technical conditions. In particular, using a separating hyperplane argument in concert with the three principles outlined above, we show that a standard decision procedure δ is extended admissible if and only if it is *nonstandard Bayes*, that is, for some nonstandard prior, the Bayes risk of its nonstandard extension ${}^*\delta$ is within an infinitesimal of the minimum Bayes risk among all extensions.

This result achieves the longstanding goal of identifying a notion of Bayes optimality equivalent to a notion of frequentist admissibility. Although our notion of Bayes optimality is defined using nonstandard analysis, we can apply our main result to obtain new (standard) results in statistical decision theory. For example, Theorem 6.6 improves upon Wald's results for compact spaces: we show that, for statistical decision problems with compact Hausdorff parameter spaces and continuous risk functions, a decision procedure is extended admissible if and only if it is Bayes. Theorem 6.7 establishes that this result is tight in the sense that neither condition can be dropped. Another application of our result (Theorem 7.4) recovers a seminal result by Heath and Sudderth [17]: for statistical decision problems with bounded risk functions, a decision procedure is extended admissible if and only if it is Bayes under a finitely additive prior.

Using our nonstandard framework, we also derive novel sufficient conditions for admissibility, and demonstrate their use in two examples. By Theorem 8.1, a decision procedure is

admissible if it is the only procedure (up to equivalence in risk) that is nonstandard Bayes under some nonstandard prior. In Example 9.3, we use this technique to establish the admissibility of the empirical distribution function under L^2 loss in a nonparametric estimation problem. We do so using a nonstandard Dirichlet process prior, circumventing technical obstacles raised by Cohen and Kuo [8].

We also develop a nonstandard variant of Blyth's method, in which admissibility is witnessed by a single nonstandard prior, rather than by a sequence of improper priors. In particular, by Theorem 8.4, for any Hausdorff topology on the parameter space with respect to which risk functions are continuous, a procedure is admissible if it is nonstandard Bayes under a nonstandard prior that assigns "sufficient" mass to every standard open ball. In Example 9.1, we use this technique to establish the admissibility of the usual estimate of the mean in the classic normal location model, providing a familiar statistical example for our nonstandard techniques. We provide several other examples to further illustrate our new framework.

In the final section, we discuss promising directions and open problems, relating to admissibility, minimaxity, and our understanding of the frequentist–Bayesian interface. We note that a substantial portion of the content in this paper appears also in the first author's doctoral dissertation [11]. Appendices A–L can be found in the Supplementary Material [12].

2. Standard preliminaries. A (nonsequential) statistical decision problem is defined by a *parameter* space Θ , each element of which represents a possible state of nature; a set \mathbb{A} of *actions* available to the statistician; a function $\ell : \Theta \times \mathbb{A} \rightarrow \mathbb{R}_{\geq 0}$ characterizing the *loss* associated with taking action $a \in \mathbb{A}$ in state $\theta \in \Theta$; and finally, a family $P = \{P_\theta : \theta \in \Theta\}$ of probability measures on a measurable *sample* space X . Intuitively, the statistician observes a random element x of the sample space X such that the probability measure P_θ is the distribution of this observation when θ is the true (but unknown) state of nature. On the basis of the observation, x , the statistician takes a (potentially randomized) action a , and then suffers a loss $\ell(\theta, a)$. Note that we consider only the case where the loss function is nonnegative. (See [14] for an approach based on integrable loss functions.)

Formally, the action space \mathbb{A} is equipped with a σ -algebra with measurable singletons such that $\ell(\theta, \cdot)$ is a measurable function from \mathbb{A} to $\mathbb{R}_{\geq 0}$ for every $\theta \in \Theta$. Every possible response by the statistician is captured by a (*randomized*) *decision procedure*, that is, a map δ from X to the space $\mathcal{M}_1(\mathbb{A})$ of probability measures \mathbb{A} . As is customary, we will write $\delta(x, A)$ for $(\delta(x))(A)$. The expected loss, or *risk*, to the statistician in state θ associated with following a decision procedure δ is

$$(2.1) \quad r_\delta(\theta) = r(\theta, \delta) = \int_X \left[\int_{\mathbb{A}} \ell(\theta, a) \delta(x, da) \right] P_\theta(dx).$$

For the risk function to be well-defined, the maps $x \mapsto \int_{\mathbb{A}} \ell(\theta, a) \delta(x, da)$, for $\theta \in \Theta$, must be measurable, and so we do not consider those decision procedures not meeting this weak measurability criterion. We further restrict our attention to decision procedures δ with finite risk functions, that is, we require $r_\delta(\theta) \in \mathbb{R}_{\geq 0}$ for all $\theta \in \Theta$. Fixing a σ -algebra on Θ with measurable singletons, we let \mathcal{D} denote the set of all randomized decision procedures with well-defined finite risk functions $r_\delta : \Theta \rightarrow \mathbb{R}_{\geq 0}$ that are moreover measurable with respect to this σ -algebra.¹ Provided \mathcal{D} is nonempty, we refer to the tuple $(\Theta, \mathbb{A}, \ell, X, P, \mathcal{D})$ as a

¹The σ -algebra on Θ both (i) constrains the class \mathcal{D} of decision procedures and (ii) determines the set of (countably additive) probability measures on Θ . There is often a natural σ -algebra on Θ coming from a problem's particular topological structure. Even if there is no natural choice, choosing the power set of Θ , one obtains (i) the largest possible class \mathcal{D} and (ii) the set of available probability measures containing all purely atomic distributions. These are the only distributions if and only if certain set-theoretic conjectures hold (see [15], Theorem 1D).

statistical decision problem. Note that procedures are sometimes called *estimators* when, informally speaking, the goal is to recover one or more of the model parameters.

The set \mathcal{D} may be viewed as a convex subset of a vector space over \mathbb{R} . In particular, for all $\delta_1, \dots, \delta_n \in \mathcal{D}$ and $p_1, \dots, p_n \in \mathbb{R}_{\geq 0}$ with $\sum_i p_i = 1$, define $\sum_i p_i \delta_i : X \rightarrow \mathcal{M}_1(\mathbb{A})$ by $(\sum_i p_i \delta_i)(x) = \sum_i p_i \delta_i(x)$ for $x \in X$. Then $r(\theta, \sum_i p_i \delta_i) = \sum_i p_i r(\theta, \delta_i) < \infty$, and so we see that $\sum_i p_i \delta_i \in \mathcal{D}$ and $r(\theta, \cdot)$ is a linear function on \mathcal{D} for every $\theta \in \Theta$. For a subset $D \subseteq \mathcal{D}$, let $\text{conv}(D)$ denote the convex hull of D , the set of all finite convex combinations of decision procedures $\delta \in D$. Note that $\text{conv}(D)$ is the smallest convex set containing D , and, writing $S^{[<\infty]}$ for the set of all finite subsets of a set S , $\text{conv}(D) = \bigcup_{C \in D^{[<\infty]}} \text{conv}(C)$. This last characterization plays a key role.

For each $a \in \mathbb{A}$, let $\text{Dirac}(a)$ denote the probability measure assigning measure 1 to the singleton $\{a\}$. A decision procedure $\delta \in \mathcal{D}$ is *nonrandomized* if, there is an *underlying* map $x \mapsto d(x)$ such that $\delta(x) = \text{Dirac}(d(x))$ for every $x \in X$. It follows that $r(\theta, \delta) = \int_X \ell(\theta, d(x)) P_\theta(dx)$. Let $\mathcal{D}_0 \subseteq \mathcal{D}$ denote the subset of all nonrandomized decision procedures.

2.1. Admissibility. In general, decision procedures are incomparable, as one procedure may present greater risk in one state and less risk in another. Some cases, however, are clear cut: the notion of domination induces a partial order on the space of decision procedures: Let $\epsilon \in \mathbb{R}_{\geq 0}$ and $\delta, \delta' \in \mathcal{D}$. Then δ is ϵ -dominated by δ' when:

1. $(\forall \theta \in \Theta)(r(\theta, \delta') \leq r(\theta, \delta) - \epsilon)$, and
2. $(\exists \theta \in \Theta)(r(\theta, \delta') \neq r(\theta, \delta))$.

Note that the second condition is redundant when $\epsilon > 0$. By convention, δ is said to be *dominated* by δ' when δ is 0-dominated by δ' .

If a decision procedure is ϵ -dominated by another decision procedure, then, computational issues notwithstanding, the ϵ -dominated procedure is eliminated from consideration. Let $\mathcal{C} \subseteq \mathcal{D}$. More carefully, $\delta \in \mathcal{D}$ is ϵ -admissible among \mathcal{C} when there does not exist $\delta_0 \in \mathcal{C}$ such that δ is ϵ -dominated by δ_0 . By convention, δ is said to be *admissible among \mathcal{C}* when δ is 0-admissible among \mathcal{C} . Further, δ is said to be *extended admissible among \mathcal{C}* if δ is ϵ -admissible among \mathcal{C} for all $\epsilon > 0$.²

Clearly, admissibility implies extended admissibility. Admissibility gives rise to the notion of a complete class: Let $\mathcal{A} \subseteq \mathcal{C}$. Then \mathcal{A} is a *complete* subclass of \mathcal{C} when, for all $\delta \in \mathcal{C} \setminus \mathcal{A}$, there exists $\delta_0 \in \mathcal{A}$ such that δ_0 dominates δ . Similarly, \mathcal{A} is an *essentially complete* subclass of \mathcal{C} when, for all $\delta \in \mathcal{C} \setminus \mathcal{A}$, there exists $\delta_0 \in \mathcal{A}$ such that $r(\theta, \delta_0) \leq r(\theta, \delta)$ for all $\theta \in \Theta$. An *essentially complete class* is an essentially complete subclass of \mathcal{D} .

If a decision procedure $\delta \in \mathcal{C}$ is admissible among \mathcal{C} , then every complete subclass of \mathcal{C} must contain δ . Note that the term *complete class* often refers to a complete subclass of some essentially complete class (such as \mathcal{D} itself or \mathcal{D}_0 under the conditions described in Section 2.4).

The next lemma captures a key consequence of essential completeness.

LEMMA 2.1. *Suppose \mathcal{A} is an essentially complete subclass of \mathcal{C} , then extended admissible among \mathcal{A} implies extended admissible among \mathcal{C} .*

²The function $s(\delta, \delta') = \sup_{\theta \in \Theta} \{r_{\delta'}(\theta) - r_\delta(\theta)\}$, for $\delta, \delta' \in \mathcal{D}$, provides alternative characterizations: For $\epsilon > 0$, we have that δ is ϵ -admissible among \mathcal{C} if and only if, for all $\delta' \in \mathcal{C}$, $s(\delta, \delta') > -\epsilon$. In contrast, δ is admissible among \mathcal{C} if and only if, for all $\delta' \in \mathcal{C}$, either $s(\delta, \delta') > 0$ or $s(\cdot, \delta) = s(\cdot, \delta')$. Finally, δ is extended admissible among \mathcal{C} if and only if, for all $\delta' \in \mathcal{C}$, $s(\delta, \delta') \geq 0$.

The class of extended admissible procedures plays a central role in this paper. It is not hard, however, to construct statistical decision problems for which the class is empty, and thus not a complete class. (See Example F.1.) Thus, in order to obtain a complete class theorem, one must at least assert that the class of extended admissible procedures is nonempty. The following result gives conditions under which the class of extended admissible procedures are a complete class. (See [5], Sections 5.4–5.6 and Theorem 5.6.3, and [14], Section 2.6 Corollary 1, for related results for finite spaces.) The proof can be found in Appendix K.2.2.

THEOREM 2.2. *Let $\mathcal{C} \subseteq \mathcal{D}$. Suppose that, for all $\delta \in \mathcal{C}$, $\{\epsilon \geq 0 : (\exists \delta' \in \mathcal{C})(\delta \text{ is } \epsilon\text{-dominated by } \delta')\}$ is closed. Then the class of decision procedures that are extended admissible among \mathcal{C} form a complete subclass of \mathcal{C} .*

2.2. Bayes optimality. Consider now the Bayesian framework, in which one adopts a prior π , that is, a probability measure defined on the σ -algebra on Θ . Irrespective of the interpretation of π , we may define the *Bayes risk of δ under π* ,

$$(2.2) \quad r(\pi, \delta) = \int_{\Theta} r(\theta, \delta) \pi(d\theta),$$

that is, the expected risk under a parameter chosen at random from π . Note the Bayes risk is well-defined because $r(\cdot, \delta)$ is a nonnegative measurable function.

We may now define several notions of *Bayes optimality* in terms of the Bayes risk. We say δ is *ϵ -Bayes among \mathcal{C} under π* when $r(\pi, \delta) < \infty$ and, for all $\delta_0 \in \mathcal{C}$, we have $r(\pi, \delta) \leq r(\pi, \delta_0) + \epsilon$. By convention, δ is (*normal-form*) *Bayes among \mathcal{C} under π* when δ is 0-Bayes among \mathcal{C} under π . We say δ is *ϵ -Bayes among \mathcal{C}* (resp., *Bayes among \mathcal{C}*) when there exists a prior π_0 such that δ is ϵ -Bayes among \mathcal{C} under π_0 (resp., Bayes among \mathcal{C} under π_0). Finally, δ is said to be *extended Bayes among \mathcal{C}* when, for all $\epsilon > 0$, there exists a prior π_0 such that δ is ϵ -Bayes among \mathcal{C} under π_0 .

Note that the term *Bayes optimality* refers informally to any one of the notions of being Bayes introduced here, as well as those introduced later. Given a prior π , a *Bayes solution (among some class $\mathcal{C} \subseteq \mathcal{D}$) for π* refers to some procedure that is Bayes among \mathcal{C} under π . Other notions of Bayes solutions are defined analogously. The following well-known result establishes a basic connection between Bayes optimality and admissibility (see, e.g., [5], Theorem 5.5.1).

THEOREM 2.3. *Bayes among \mathcal{C} implies extended Bayes among \mathcal{C} implies extended admissible among \mathcal{C} .*

We give a proof for completeness in Appendix K.2.3. Note that neither extended admissibility nor admissibility imply Bayes optimality, in general.

2.3. Extensive-form Bayes optimality. The definition of a (normal-form) Bayes procedure above is given in terms of expected loss, that is, risk. Although this definition is common in the statistical decision theory setting, the traditional (*Bayesian*) definition of a Bayes procedure is given in terms of *posterior* expected loss, that is, *conditional* expected loss *given an observed sample*. In particular, an *extensive-form* Bayes procedure is defined pointwise: given an observed sample, choose the action (or distribution on actions) that minimizes the posterior expected loss. (The terminology of normal- and extensive-form Bayes optimality is due to Raiffa and Schlaifer [25].)

More carefully, let $\pi \otimes P$ be the distribution on $\Theta \times X$ given by $(\pi \otimes P)(A \times B) = \int_A P_{\theta}(B) \pi(d\theta)$ for measurable $A \subseteq \Theta$ and $B \subseteq X$, and let m_{π} be its projection onto X , that

is, $m_\pi(B) = (\pi \otimes P)(\Theta \times B)$ for measurable $B \subseteq X$. Then a *disintegration of $\pi \otimes P$* is a probability kernel κ from X to Θ satisfying $\int_B \kappa(x, A) m_\pi(dx) = (\pi \otimes P)(A \times B)$ for all measurable $A \subseteq \Theta$ and $B \subseteq X$.

Let $\mathcal{M}_1^P(\Theta)$ denote the set of priors π' such that there exists a disintegration of $\pi' \otimes P$ that is $m_{\pi'}$ -almost-everywhere unique, that is, all disintegrations agree on a $m_{\pi'}$ -measure one set. (So-called disintegration theorems describe mild regularity conditions that guarantee $\mathcal{M}_1^P(\Theta) = \mathcal{M}_1(\Theta)$). See Theorem J.1.)

Assume $\pi \in \mathcal{M}_1^P(\Theta)$. Then the *posterior expected loss of a (randomized) action $p \in \mathcal{M}_1(\mathbb{A})$ having observed $x \in X$* is $E_\pi^x[p] = \int [\int \ell(t, a) p(da)] \kappa(x, dt)$.

LEMMA 2.4. For all $\pi' \in \mathcal{M}_1^P(\Theta)$ and $\delta_0 \in \mathcal{D}$, $r(\pi', \delta_0) = \int E_{\pi'}^x[\delta_0(x)] m_{\pi'}(dx)$.

Let $B \subseteq X$. We say δ is *randomized extensive-form Bayes under π on B* when $E_\pi^x[\delta(x)] \leq E_\pi^x[p]$ for all $p \in \mathcal{M}_1(\mathbb{A})$ and all $x \in B$. We say δ is *(nonrandomized) extensive-form Bayes under π on B* when $E_\pi^x[\delta(x)] \leq E_\pi^x[\text{Dirac}(a)]$ for all $a \in \mathbb{A}$ and all $x \in B$. These relations specify a set on which the procedure is (conditionally) optimal. In the literature, *extensive-form Bayes under π* means (nonrandomized) extensive-form Bayes under π on B' , for some m_π -measure-one set B' .

Under mild regularity conditions, a procedure is extensive-form Bayes under π if and only if it is Bayes among \mathcal{D}_0 under π , and similarly for randomized extensive-form Bayes under π and Bayes among \mathcal{D} under π . This equivalence fails in some settings, including that of finitely additive priors, which we introduce in Section 3 and study further in Section 7. We work mainly with normal-form notions, and describe the connection with extensive-form notions in Appendix J.

2.4. *Essential completeness of \mathcal{D}_0 .* In many natural statistical decision problems, the nonrandomized decision procedures, \mathcal{D}_0 , form an essentially complete subclass of their convex hull, which we capture by the following formal condition.

CONDITION RU (Randomization unnecessary). \mathcal{D}_0 is an essentially complete subclass of $\text{conv}(\mathcal{D}_0)$.

One broad class of problems for which (RU) holds are those where the action space \mathbb{A} is itself a convex subset of a vector space over the field \mathbb{R} and the loss function is convex with respect to its second parameter (in \mathbb{A}). Given the convex structure of \mathbb{A} , and the fact that elements of $\text{conv}(\mathcal{D}_0)$ are finite mixtures of nonrandomized procedures, the *mean action* $\int_{\mathbb{A}} a \delta(x, da)$ is well-defined for every $\delta \in \text{conv}(\mathcal{D}_0)$ and $x \in X$. Given the convex loss, it is well known that the mean action will be no worse on average than the original randomized one. Besides the possibility of technical measurability obstacles, (RU) holds. See Appendix D for more details, and [2], Section 8.2.2, for additional conditions leading to (RU).

3. Prior work. In this section, we survey prior work in order to facilitate comparison with our nonstandard results in the following sections. In this section, we write *Bayes* as an abbreviation for *Bayes among \mathcal{D}* .

The first key results linking frequentist and Bayes optimality are due to Abraham Wald, who laid the foundation of sequential decision theory [30–34]. For finite parameters spaces, one can use intuitive geometric arguments to establish that every admissible decision procedure is Bayes (see, e.g., [14], Section 2.10, Theorem 1). In the other direction, elementary arguments show that every procedure that is Bayes under a prior with full support is admissible (see [14], Section 2.3, Theorems 2 and 3). This close relationship between admissibility

and Bayes optimality already breaks down when the parameter space is a finite-dimensional vector space over the reals: Here, admissible decision procedures are, in general, not Bayes procedures (see, e.g., Stein [28], Section 4). Under some regularity conditions, however, every admissible procedure is a *limit* of Bayes procedures [7, 23, 33], although, limits of Bayes procedures are in general neither admissible nor Bayes, as famously demonstrated by Stein [28] in the multivariate normal-location model (see also [19]). Under more stringent conditions, admissible procedures are *generalized Bayes* [3, 6, 26, 29], that is, procedures derived from the “formal” (i.e., mechanical) application of Bayes rule to *improper* priors. In the remainder of this section, we collect together these and other key results. In each case, the proof establishes the (essential) *completeness* of some class \mathcal{C} and the *containment* of \mathcal{C} within the class of procedures satisfying (some relaxed notion of) Bayes optimality. Every example requires technical conditions, even to establish the set-theoretic containment.

3.1. Results following from compactness, boundedness and continuity. In order to prove his early results, Wald introduced tailor-made notions of continuity for risk and loss functions: A sequence of parameters $\{\theta_i\}_{i \in \mathbb{N}}$ converges *in risk* to a parameter θ when $\sup_{\delta \in \mathcal{D}} |r(\theta_i, \delta) - r(\theta, \delta)| \rightarrow 0$ as $i \rightarrow \infty$, and converges *in loss* when $\sup_{a \in \mathbb{A}} |\ell(\theta_i, a) - \ell(\theta, a)| \rightarrow 0$ as $i \rightarrow \infty$. Similarly, a sequence of decision procedures $\{\delta_i\}_{i \in \mathbb{N}}$ in \mathcal{D} converges *in risk* to a decision procedure δ when $\sup_{\theta \in \Theta} |r(\theta, \delta_i) - r(\theta, \delta)| \rightarrow 0$ as $i \rightarrow \infty$. A sequence of actions $\{a_i\}_{i \in \mathbb{N}}$ converges *in loss* to an action $a \in \mathbb{A}$ when $\sup_{\theta \in \Theta} |\ell(\theta, a_i) - \ell(\theta, a)| \rightarrow 0$ as $i \rightarrow \infty$. Topologies on Θ , \mathbb{A} , and \mathcal{D} are generated by these notions of convergence.

In the following result and elsewhere, a model P is said to *admit (a measurable family of) densities* $(f_\theta)_{\theta \in \Theta}$ (with respect to a σ -finite measure ν) when $P_\theta(A) = \int_A f_\theta(x) \nu(dx)$ for every $\theta \in \Theta$ and measurable $A \subseteq X$. In terms of these densities, there is a *unique (extensive-form) Bayes solution* under a prior π on Θ when, for ν -almost every $x \in X$, there exists exactly one action $a^* \in \mathbb{A}$ for which the expression $\int_\Theta \ell(\theta, a) f_\theta(x) \pi(d\theta)$ takes its minimum value with respect to $a \in \mathbb{A}$. (Another notion of uniqueness used in the literature simply demands that the risk functions of two Bayes solutions agree.) The following result is due to Wald.

THEOREM 3.1 ([32], Theorems 4.11 and 4.14). *Suppose that Θ and \mathcal{D} are compact in risk; that Θ and \mathbb{A} are compact in loss; that P admits densities $(f_\theta)_{\theta \in \Theta}$ with respect to Lebesgue measure; and that these densities are strictly positive outside a Lebesgue measure-zero set. Then every extended admissible decision procedure is Bayes. If there is a unique Bayes solution for every prior π , the class of nonrandomized Bayes procedures form a complete class.*

Wald’s regularity conditions are quite strong; he essentially requires equicontinuity in each variable for both the loss and risk functions. For example, the standard normal-location problem under squared error does not satisfy these criteria.

Wald gives a similar result with more transparent hypotheses in [31].

THEOREM 3.2 ([31], Theorem 3.1). *Suppose that Θ is a compact subset of a Euclidean space; that \mathbb{A} is compact in loss; that P admits densities $(f_\theta)_{\theta \in \Theta}$; that the map $(x, \theta) \mapsto f_\theta(x)$ is jointly continuous; that $\ell(\theta, a)$ is a continuous function of θ for every a ; and that there is a unique Bayes solution for every prior π on Θ . Then every Bayes procedure is admissible and the collection of Bayes procedures form an essentially complete class.*

Berger [2], Section 8.8, Theorem 12, establishes the following complete class theorem by assuming the continuity of every risk function with respect to the parameter and restricting attention to Euclidean parameter spaces.

THEOREM 3.3. *Suppose that \mathbb{A} and Θ are compact subsets of Euclidean spaces; that all decision procedures have continuous risk functions; that $\ell(\theta, a)$ is a continuous function of a for every θ ; and that the model P admits densities $(f_\theta)_{\theta \in \Theta}$ w.r.t. either Lebesgue or counting measure such that the map $(x, \theta) \mapsto f_\theta(x)$ is jointly continuous; Then the collection of Bayes procedures form a complete class.*

3.2. Results under weaker hypotheses. In the noncompact setting, Bayes procedures generally do not form a complete class. With a view to generalizing the notion of a Bayes procedure and recovering a complete class, Wald [33] introduced the notion of extended Bayes, which he called “Bayes in the wide sense.” The statement of the following theorem is adapted from [14].

THEOREM 3.4. *Suppose that there exists a topology on \mathcal{D} such that \mathcal{D} is compact and $r(\theta, \delta)$ is lower semicontinuous in $\delta \in \mathcal{D}$ for all $\theta \in \Theta$. Then the set of extended Bayes procedures form an essentially complete class.*

Wald also studied the closure (in a suitable sense) of the collection of all Bayes procedures, and showed that every admissible procedure was contained in this new class. The first result of this form appears in [33] and is extended later in [23]. Brown [7], Appendix to Chapter 4, extended these results and gave a modern treatment. The following statement of Brown’s version is adapted from [24], Section 5, Theorem 7.15.

THEOREM 3.5. *Suppose that \mathbb{A} is a closed convex subset of Euclidean space; that P admits strictly positive densities $(f_\theta)_{\theta \in \Theta}$ w.r.t. a σ -finite measure ν ; and that $\ell(\theta, a)$ is lower semicontinuous and strictly convex in a for every θ , and satisfies $\lim_{|a| \rightarrow \infty} \ell(\theta, a) = \infty$ for all $\theta \in \Theta$. Then every admissible decision procedure δ is an a.e. limit of Bayes procedures, that is, there exists a sequence π_n of priors with finite supports, such that $\delta^{\pi_n}(x) \rightarrow \delta(x)$ as $n \rightarrow \infty$ for ν -almost all x , where δ^{π_n} is Bayes under π_n .*

In the normal-location model under squared error loss, the sample mean, while not a Bayes estimator in the strict sense, can be seen as a limit of Bayes estimators, for example, under normal priors of variance K as $K \rightarrow \infty$ or uniform priors on $[-K, K]$ as $K \rightarrow \infty$. (We revisit this problem in Example 9.1.) In his seminal paper, Sacks [26] observes that the sample mean is also the Bayes solution if the notion of prior distribution is relaxed to include Lebesgue measure on the real line. Sacks [26] raised the natural question: if δ is a limit of Bayes estimators, is there a measure m on the real line such that δ is “Bayes” under this measure? A solution in this latter form was termed a *generalized Bayes solution* by Sacks [26].

The following definitions are adapted from [29]: Let $\mathcal{C} \subseteq \mathcal{D}$ and let π be an *improper prior*, that is, a σ -finite nonnull measure on Θ . A decision procedure δ_0 is *normal-form generalized Bayes* among \mathcal{C} under π when $r(\pi, \delta_0) < \infty$ and δ_0 minimizes $r(\pi, \delta) = \int r(\theta, \delta)\pi(d\theta)$ over $\delta \in \mathcal{C}$. If P admits densities $(f_\theta)_{\theta \in \Theta}$ with respect to a σ -finite measure ν and the map $d : X \rightarrow \mathbb{A}$ underlying a nonrandomized decision procedure δ_0 minimizes the unnormalized posterior risk $\int \ell(\theta, d(x)) f_\theta(x) \pi(d\theta)$ for ν -a.e. x , then δ_0 is (*extensive-form*) *generalized Bayes* under π .

When a model admits densities, Stone [29] showed that every normal-form generalized Bayes procedure is also extensive-form. (Sacks defined generalized Bayes in extensive form, but demanded also that $\int f_\theta(\cdot)\pi(d\theta)$ be finite ν -a.e.) More is known for finite-dimensional exponential families, that is, models parameterized by vectors $\theta \in \mathbb{R}^k$, k finite, admitting densities (with respect to ν) of the form $f_\theta(x) = \beta(\theta) \exp(\theta'x)$ for $\theta \in \Theta = \{t \in \mathbb{R}^k :$

$\int \exp(t'x)v(dx) < \infty$. Under suitable conditions, one can show that every admissible estimator is generalized Bayes. The first such result was developed by Sacks [26] in his original paper: he proved that, for statistical decision problems where the model admits a density of the form $e^{x\theta}/Z_\theta$ with $Z_\theta = \int e^{x\theta}v(d\theta)$, every admissible estimator is generalized Bayes. Stone [29] extended this result to estimation of the mean in one-dimensional exponential families under squared error loss. These results were further generalized in similar ways by Brown [6], Section 3.1, and Berger and Srinivasan [3]. The following theorem is given in [3]. (We adapt the statement from [24].)

THEOREM 3.6 ([24], Section 5, Theorem 7.17). *Suppose P is a finite-dimensional exponential family and $\ell(\theta, a)$ is jointly continuous in (θ, a) , strictly convex in a for every θ , and satisfies $\lim_{|a| \rightarrow \infty} \ell(\theta, a) = \infty$ for all θ . Then every admissible estimator is extensive-form generalized Bayes.*

Finally, we turn to finitely additive priors. (We give a brief introduction to finitely additive measure theory in Appendix C. For a rigorous treatment, see [4]. Note that we assume the model, P , remain σ -additive here. See [20] for a discussion of the purely finitely additive case.) Let $\delta \in \mathcal{D}$ and $\mathcal{C} \subseteq \mathcal{D}$, and let π_0 be a finitely additive prior. Throughout this paper, we use the same symbol for finitely additive Bayes risk as the symbol for standard Bayes risk. The *finitely additive Bayes risk under π_0 of δ* , $r(\pi_0, \delta) = \int_{\Theta} r(\theta, \delta)\pi_0(d\theta)$, is defined exactly as in the standard countably additive setting, except that the Lebesgue integral is replaced by the Daniell integral. Similarly, δ is *finitely additive Bayes among \mathcal{C} under π_0* if $r(\pi_0, \delta) < \infty$ and, for all $\delta' \in \mathcal{C}$, we have $r(\pi_0, \delta) \leq r(\pi_0, \delta')$. Risk functions are well-defined by Theorem C.5. Heath and Sudderth [17] establish the following equivalence between extended admissibility and finitely additive Bayes optimality.

THEOREM 3.7 ([17], Theorem 2). *Let $\delta \in \mathcal{D}$ and $\mathcal{C} \subseteq \mathcal{D}$. If δ is finitely additive Bayes among \mathcal{C} then δ is extended admissible among \mathcal{C} . If the loss function is bounded and the class \mathcal{C} is convex, then δ is finitely additive Bayes among \mathcal{C} only if δ is extended admissible among \mathcal{C} .*

The simplicity of this statement is remarkable. However, the assumption of boundedness is very strong and rules out many standard estimation problems on unbounded spaces. In Section 7, we establish Theorem 3.7 as a special case of our main result, Theorem 5.18.

4. Nonstandard admissibility. As we have seen in the previous section, strong regularity appears to be necessary to align Bayes optimality and admissibility. In noncompact parameter spaces, the statistician must apparently abandon the strict use of probability measures in order to represent certain extreme states of uncertainty that correspond with admissible procedures. Even then, strong regularity conditions are required (such as domination of the model and strict positiveness of densities, ruling out estimation in some infinite-dimensional contexts). In the remainder of the paper, we describe a new approach using nonstandard analysis, in which the statistician has access to a much richer collection of real numbers with which to express their prior beliefs.

Let $(\Theta, \mathbb{A}, \ell, X, P, \mathcal{D})$ be a standard statistical decision problem.

We will assume the reader is familiar with basic concepts in nonstandard analysis. (See Appendices A and B for a review tailored to this paper.) For a set S , let $\mathcal{P}(S)$ be its power set and $\mathbb{V}(S) = S \cup \mathcal{P}(S) \cup \mathcal{P}(S \cup \mathcal{P}(S)) \cup \dots$. We assume that we are working within a nonstandard model $*$: $\mathbb{V}(S) \rightarrow \mathbb{V}(*S)$ where $S \supseteq \mathbb{R} \cup \Theta \cup \mathbb{A} \cup X$, and we assume the model is as saturated as necessary. We use $*$ to denote the nonstandard extension map taking elements,

sets, functions, relations, etc., to their nonstandard counterparts. In particular, ${}^*\mathbb{R}$ and ${}^*\mathbb{N}$ denote the nonstandard extensions of the reals and natural numbers, respectively. Note that ${}^*s = s$ for all $s \in S$. Hence, $\mathbb{R} \subseteq {}^*\mathbb{R}$, $\Theta \subseteq {}^*\Theta$, etc. For $x, y \in {}^*\mathbb{R}$, write $x \approx y$ if $|x - y|$ is infinitesimal and write $x \lesssim y$ to mean $x < y$ or $x \approx y$. For two internal sets A, B , let $I(B^A)$ denote the collection of all internal functions from A to B .

4.1. *Nonstandard extension of a statistical decision problem.* Recall that Θ is a measurable space such that all singletons are measurable, and so, all finite subsets of Θ are measurable. By transfer, all hyperfinite subsets of ${}^*\Theta$ are * measurable.

One should view the statistical model P as a function from Θ to the space $\mathcal{M}_1(X)$ of probability measures on X . Write *P_y for $({}^*P)_y$. For every $y \in {}^*\Theta$, the transfer principle implies that *P_y is a * probability measure on *X (defined on the extension of its σ -algebra). By Lemma B.1, ${}^*(P_\theta) = {}^*P_\theta$ for $\theta \in \Theta$, as one would expect from the notation.

Recall that standard decision procedures $\delta \in \mathcal{D}$ have finite nonnegative risk functions. Therefore, the risk map $(\theta, \delta) \mapsto r(\theta, \delta)$ is a function from $\Theta \times \mathcal{D}$ to $\mathbb{R}_{\geq 0}$. By the extension and transfer principles, the nonstandard extension *r is an internal function from ${}^*\Theta \times {}^*\mathcal{D}$ to ${}^*\mathbb{R}_{\geq 0}$ and ${}^*\delta \in {}^*\mathcal{D}$ if $\delta \in \mathcal{D}$. The transfer principle also implies that every $\Delta \in {}^*\mathcal{D}$ is an internal function from *X to ${}^*\mathcal{M}_1(\mathbb{A})$. The * risk function of $\Delta \in {}^*\mathcal{D}$ is the function ${}^*r(\cdot, \Delta)$ from ${}^*\Theta$ to ${}^*\mathbb{R}_{\geq 0}$. By the transfer of the equation defining risk, the following statement holds:

$$(4.1) \quad (\forall \theta \in {}^*\Theta) (\forall \Delta \in {}^*\mathcal{D}) \quad {}^*r(\theta, \Delta) = \int_{{}^*X} \left[\int_{{}^*\mathbb{A}} {}^*\ell(\theta, a)\Delta(x, da) \right] {}^*P_\theta(dx).$$

As is customary, we will simply write f for *f , provided the context is clear. (We will also drop * from the extensions of common functions and relations like addition, multiplication, less-than-or-equal-to, etc.)

4.2. *Nonstandard admissibility.* Let $\delta_0, \delta \in \mathcal{D}$, let $\epsilon \in \mathbb{R}_{\geq 0}$, and assume δ_0 is ϵ -dominated by δ . Then there exists $\theta_0 \in \Theta$ such that

$$(4.2) \quad (\forall \theta \in \Theta) \quad (r(\theta, \delta) \leq r(\theta, \delta_0) - \epsilon) \wedge (r(\theta_0, \delta) \neq r(\theta_0, \delta_0)).$$

By the transfer principle,

$$(4.3) \quad (\forall \theta \in {}^*\Theta) \quad ({}^*r(\theta, {}^*\delta) \leq {}^*r(\theta, {}^*\delta_0) - \epsilon) \wedge ({}^*r(\theta_0, {}^*\delta) \neq {}^*r(\theta_0, {}^*\delta_0)).$$

Because ${}^*r(\theta_0, {}^*\delta) = r(\theta_0, \delta)$ and similarly for ${}^*r(\theta_0, {}^*\delta_0)$, Lemma A.7.1 implies that ${}^*r(\theta_0, {}^*\delta) \not\approx {}^*r(\theta_0, {}^*\delta_0)$. These results motivate the following nonstandard version of domination.

DEFINITION 4.1. Let $\Delta, \Delta' \in {}^*\mathcal{D}$, let $\epsilon \in \mathbb{R}_{\geq 0}$, and let $R, S \subseteq {}^*\Theta$. Then Δ is ϵ - ${}^\circ$ dominated in R/S by Δ' when:

1. $(\forall \theta \in S) ({}^*r(\theta, \Delta') \leq {}^*r(\theta, \Delta) - \epsilon)$, and
2. $(\exists \theta \in R) ({}^*r(\theta, \Delta') \not\approx {}^*r(\theta, \Delta))$.

Write ${}^\circ$ dominated in R/S for 0 - ${}^\circ$ dominated in R/S , and write ϵ - ${}^\circ$ dominated on S for ϵ - ${}^\circ$ dominated in S/S .³

In this paper, we are only interested in the following three cases: $R = \Theta$ and $S = {}^*\Theta$; $R = S = \Theta$; and $R = S = {}^*\Theta$. These nonstandard notions of domination give rise to nonstandard notions of admissibility.

³We use the ${}^\circ$ prefix to highlight that, even when restricted to internal subsets R, S , the relation is not internal due to its reliance on the \approx relation.

DEFINITION 4.2. Let $\epsilon \in \mathbb{R}_{\geq 0}$, $R, S \subseteq {}^*\Theta$, $\mathcal{C} \subseteq {}^*\mathcal{D}$ and $\Delta \in {}^*\mathcal{D}$.

1. Δ is ϵ - $^\circ$ admissible in R/S among \mathcal{C} unless Δ is ϵ - $^\circ$ dominated in R/S by some $\Delta' \in \mathcal{C}$.
2. Δ is $^\circ$ admissible in R/S among \mathcal{C} if Δ is 0 - $^\circ$ admissible in R/S among \mathcal{C} .
3. Δ is ϵ - $^\circ$ admissible on S among \mathcal{C} if Δ is ϵ - $^\circ$ admissible in S/S among \mathcal{C} .
4. Δ is $^\circ$ extended admissible on S among \mathcal{C} if Δ is ϵ - $^\circ$ admissible on S among \mathcal{C} for every $\epsilon \in \mathbb{R}_{>0}$.

Note that Δ is $^\circ$ extended admissible on S among \mathcal{C} if and only if, for all $\Delta' \in \mathcal{C}$, $\sup_{\theta \in S} \text{st}({}^*r(\theta, \Delta') - {}^*r(\theta, \Delta)) \geq 0$. The following result is immediate upon inspection of the definitions above.

LEMMA 4.3. Let $\epsilon \leq \epsilon'$, $R \subseteq R'$, $S \subseteq S'$ and $\mathcal{A} \subseteq \mathcal{C}$. Then ϵ - $^\circ$ admissible in R'/S among \mathcal{C} implies ϵ' - $^\circ$ admissible in R/S' among \mathcal{A} . For $\epsilon > 0$, ϵ - $^\circ$ admissible on S among \mathcal{C} implies ϵ' - $^\circ$ admissible on S' among \mathcal{A} .

The analogous results for $^\circ$ admissible in R/S among \mathcal{C} and $^\circ$ extended admissible on S among \mathcal{C} follow directly. The connection between standard and nonstandard admissibility relies on the notion of the standard-part copy.

DEFINITION 4.4. The standard-part copy of $\mathcal{C} \subseteq \mathcal{D}$ is ${}^\sigma\mathcal{C} = \{{}^*\delta : \delta \in \mathcal{C}\}$.

Note that ${}^\sigma\mathcal{C} \subseteq {}^*\mathcal{C}$ and ${}^\sigma\mathcal{C}$ is an external set unless \mathcal{C} is finite. We can now connect standard and nonstandard admissibility. (See Appendix K.4.5 for the proof.)

THEOREM 4.5. Let $\epsilon \in \mathbb{R}_{\geq 0}$, $\delta_0 \in \mathcal{D}$, and $\mathcal{C} \subseteq \mathcal{D}$. The following are equivalent:

1. δ_0 is ϵ -admissible among \mathcal{C} .
2. ${}^*\delta_0$ is ϵ - $^\circ$ admissible in $\Theta/{}^*\Theta$ among ${}^\sigma\mathcal{C}$.
3. ${}^*\delta_0$ is ϵ - $^\circ$ admissible on Θ among ${}^\sigma\mathcal{C}$.

If $\epsilon > 0$, then the following are also equivalent:

4. ${}^*\delta_0$ is ϵ - $^\circ$ admissible on ${}^*\Theta$ among ${}^\sigma\mathcal{C}$.
5. ${}^*\delta_0$ is ϵ - $^\circ$ admissible on ${}^*\Theta$ among ${}^*\mathcal{C}$.

The following consequences for * extended admissibility are then immediate.

THEOREM 4.6. Let $\delta_0 \in \mathcal{D}$ and $\mathcal{C} \subseteq \mathcal{D}$. The following are equivalent:

1. δ_0 is extended admissible among \mathcal{C} .
2. ${}^*\delta_0$ is $^\circ$ extended admissible on Θ among ${}^\sigma\mathcal{C}$.
3. ${}^*\delta_0$ is $^\circ$ extended admissible on ${}^*\Theta$ among ${}^\sigma\mathcal{C}$.
4. ${}^*\delta_0$ is $^\circ$ extended admissible on ${}^*\Theta$ among ${}^*\mathcal{C}$.

Finally, we require a nonstandard analogue of essential completeness.

DEFINITION 4.7. Let $\mathcal{C} \subseteq {}^*\mathcal{D}$. A subset $\mathcal{A} \subseteq \mathcal{C}$ is an essentially $^\circ$ complete subclass of \mathcal{C} when, for all $\Delta \in \mathcal{C} \setminus \mathcal{A}$, there exists $\Delta' \in \mathcal{A}$ such that ${}^*r(\theta, \Delta') \lesssim {}^*r(\theta, \Delta)$ for all $\theta \in \Theta$.

The following result is the nonstandard analogue of Lemma 2.1. Its proof (Appendix K.4.8) is nearly identical.

LEMMA 4.8. Suppose \mathcal{A} is an essentially $^\circ$ complete subclass of $\mathcal{C} \subseteq {}^*\mathcal{D}$. Then $^\circ$ extended admissible on Θ among \mathcal{A} implies $^\circ$ extended admissible on Θ among \mathcal{C} .

5. Nonstandard Bayes. In this section, we start by defining several nonstandard analogues of Bayes optimality. We establish the equivalence between these notions in Theorem 5.5. By using a nonstandard version of hyperplane separation theorem, we establish the main result of this paper (Theorem 5.18), showing that a decision procedure δ is extended admissible if and only if its nonstandard extension ${}^*\delta$ has infinitesimal excess Bayes risk.

We now define the nonstandard counterparts to Bayes risk and optimality for the class ${}^*\mathcal{D}$ of internal decision procedures.

DEFINITION 5.1. Let $\Delta \in {}^*\mathcal{D}$, $\epsilon \in {}^*\mathbb{R}_{\geq 0}$ and $\mathcal{C} \subseteq {}^*\mathcal{D}$, and let $\Pi_0 \in {}^*\mathcal{M}_1(\Theta)$ be a *nonstandard prior*, that is, a * probability measure on $({}^*\Theta, {}^*\mathcal{B}_\Theta)$. The *internal Bayes risk* under Π_0 of Δ is ${}^*r(\Pi_0, \Delta) = \int {}^*r(\theta, \Delta)\Pi_0(d\theta)$.

1. Δ is ϵ -*Bayes among \mathcal{C} under Π_0* if ${}^*r(\Pi_0, \Delta)$ is hyperfinite and, for all $\Delta' \in \mathcal{C}$, we have ${}^*r(\Pi_0, \Delta) \leq {}^*r(\Pi_0, \Delta') + \epsilon$.
2. Δ is *S-Bayes among \mathcal{C} under Π_0* if ${}^*r(\Pi_0, \Delta)$ is hyperfinite and, for all $\Delta' \in \mathcal{C}$, we have ${}^*r(\Pi_0, \Delta) \lesssim {}^*r(\Pi_0, \Delta')$.

We will write *S-Bayes among \mathcal{C}* to mean *S-Bayes among \mathcal{C} under some nonstandard prior Π* . The same nomenclature will be used for ϵ -*Bayes among \mathcal{C}* . Note that the internal Bayes risk is precisely the extension of the standard Bayes risk. Similarly, if we consider the relation $\{(\delta, \epsilon, \mathcal{C}) \in \mathcal{D} \times \mathbb{R}_{\geq 0} \times \mathcal{P}(\mathcal{D}) : \delta \text{ is } \epsilon\text{-Bayes among } \mathcal{C}\}$, then its extension is $\{(\Delta, \epsilon, \mathcal{C}) \in {}^*\mathcal{D} \times {}^*\mathbb{R}_{\geq 0} \times {}^*\mathcal{P}(\mathcal{D}) : \Delta \text{ is } \epsilon\text{-}{}^*\text{Bayes among } \mathcal{C}\}$. Note, however, that our definition of “ ϵ -*Bayes among \mathcal{C}* ” allows the set $\mathcal{C} \subseteq {}^*\mathcal{D}$ to be external, and so it is not simply the transfer of the standard relation. The following lemma, whose proof can be found in Appendix K.5.2, relates the two nonstandard notions of Bayes optimality: Recall that our nonstandard model is κ saturated.

LEMMA 5.2. *Let $\mathcal{C} \subseteq {}^*\mathcal{D}$. If $\epsilon \approx 0$, then ϵ -*Bayes among \mathcal{C} under Π_0 implies S-Bayes among \mathcal{C} under Π_0 . In the other direction, if \mathcal{C} is either internal or has a fixed external cardinality less than κ , then S-Bayes among \mathcal{C} under Π_0 implies ϵ -*Bayes among \mathcal{C} under Π_0 for some $\epsilon \approx 0$.***

For standard decision procedures, we define the following novel notion of Bayes optimality.

DEFINITION 5.3. Let $\delta \in \mathcal{D}$, $\mathcal{C} \subseteq \mathcal{D}$, and let Π_0 be a nonstandard prior.

1. δ is *nonstandard Bayes among \mathcal{C} under Π_0* if ${}^*r(\Pi_0, {}^*\delta)$ is hyperfinite and, for all $\delta' \in \mathcal{C}$, we have ${}^*r(\Pi_0, {}^*\delta) \lesssim {}^*r(\Pi_0, {}^*\delta')$.
2. δ is *nonstandard Bayes among \mathcal{C}* if δ is nonstandard Bayes among \mathcal{C} under some nonstandard prior.

Intuitively, this relation allows for the use of hyperreals to expand the set of allowable priors, rather than the set of decision procedures. The following result is immediate from definitions.

LEMMA 5.4. *Let $\delta \in \mathcal{D}$ and $\mathcal{C} \subseteq \mathcal{D}$.*

1. *If π_0 is a standard prior, then δ is Bayes among \mathcal{C} under π_0 if and only if δ is nonstandard Bayes among \mathcal{C} under ${}^*\pi_0$.*
2. *If Π_0 is a nonstandard prior, then δ is nonstandard Bayes among \mathcal{C} under Π_0 if and only if ${}^*\delta$ is S-Bayes among ${}^*\mathcal{C}$ under Π_0 .*

Transfer remains a powerful tool for relating the optimality of standard procedures with that of their extensions. For example, by transfer, δ is ϵ -Bayes among \mathcal{C} under π if and only if $^*\delta$ is ϵ - * Bayes among $^*\mathcal{C}$ under $^*\pi$. (Recall that $^*\epsilon = \epsilon$ for a real ϵ , by extension.) Transfer also yields the following result, the proof of which can be found in Appendix K.5.5.

THEOREM 5.5. *Let $\delta_0 \in \mathcal{D}$ and $\mathcal{C} \subseteq \mathcal{D}$. The following are equivalent:*

1. δ_0 is extended Bayes among \mathcal{C} .
2. $^*\delta_0$ is ϵ - * Bayes among $^*\mathcal{C}$ for all $\epsilon \in ^*\mathbb{R}_{>0}$.
3. $^*\delta_0$ is ϵ_0 - * Bayes among $^*\mathcal{C}$ for some $\epsilon_0 \approx 0$.
4. $^*\delta_0$ is S-Bayes among $^*\mathcal{C}$.

In Appendix G, we characterize nonstandard priors witnessing (2)–(4). By Lemma 5.4 and Theorem 5.5, if δ_0 is extended Bayes among \mathcal{C} , then δ_0 is nonstandard Bayes among \mathcal{C} . In general, we would not expect the extension of a standard procedure to be 0- * Bayes among \mathcal{C} under Π for a generic nonstandard prior Π and class $\mathcal{C} \subseteq ^*\mathcal{D}$. The definition of S-Bayes provides infinitesimal slack. As in the standard universe, S-Bayes optimality implies nonstandard extended admissibility. The proof, which is similar to that of its standard counterpart, can be found in Appendix K.5.6.

THEOREM 5.6. *Let $\Delta_0 \in ^*\mathcal{D}$, let $\mathcal{C} \subseteq ^*\mathcal{D}$ and suppose that Δ_0 is S-Bayes among \mathcal{C} . Then Δ_0 is $^\circ$ extended admissible on $^*\Theta$ among \mathcal{C} .*

Theorems 4.6 and 5.6 immediately yield the following corollary.

COROLLARY 5.7. *Let $\delta \in \mathcal{D}$ and $\mathcal{C} \subseteq \mathcal{D}$. If δ is nonstandard Bayes among \mathcal{C} , then δ is extended admissible among \mathcal{C} .*

The above result raises several questions: Are extended admissible decision procedures also nonstandard Bayes? What is the relationship with admissibility and its nonstandard counterparts? We establish a converse to Theorem 5.6 in the remainder of this section, and return to admissibility in Section 8.

5.1. Hyperdiscretized risk set. In a statistical decision problem with a finite parameter space, one can use a separating-hyperplane argument to show that every admissible decision procedure is Bayes (see, e.g., [14], Section 2.10, Theorem 1). With the aid of extension, transfer, and saturation, we pursue a similar argument and arrive at our main result, Theorem 5.18, which gives a precise characterization of *extended* admissibility.

When relating extended admissibility and Bayes optimality among a subclass $\mathcal{C} \subseteq \mathcal{D}$ of decision procedures, the key structure is the *risk set (for \mathcal{C})*, that is, the set of all risk functions r_δ , for $\delta \in \mathcal{C}$. On a finite parameter space, the risk set for \mathcal{D} is a convex subset of a finite-dimensional vector space over \mathbb{R} . When the parameter space is not finite, one must grapple with infinite-dimensional function spaces. However, in a sufficiently saturated nonstandard model, there exists an internal set $T_\Theta \subseteq ^*\Theta$ that is hyperfinite and contains Θ . While the risk at all points in T_Θ does not suffice to characterize an arbitrary element of $^*\mathcal{D}$, it suffices to study the optimality of extensions of standard decision procedure *relative to other extensions*.

Let $J_\Theta \in ^*\mathbb{N}$ be the internal cardinality of T_Θ and let $T_\Theta = \{t_1, \dots, t_{J_\Theta}\}$. Recall that $I(^*\mathbb{R}^{T_\Theta})$ denotes the set of (internal) functions from T_Θ to $^*\mathbb{R}$. For $N \in ^*\mathbb{N}$, we write $IV(^*\mathbb{R}^N)$ for $I(^*\mathbb{R}^{\{0,1,\dots,N-1\}})$. Note that $IV(^*\mathbb{R}^N)$ is the set of internal N -dimensional vectors of elements in X . It is easy to see that there exists a natural bijection between $I(^*\mathbb{R}^{T_\Theta})$ and $IV(^*\mathbb{R}^{J_\Theta})$. Hence, every * probability measure on T_Θ can be identified with an element in $IV(^*\mathbb{R}^{J_\Theta})$. For an element $x \in IV(^*\mathbb{R}^{J_\Theta})$, we will write x_k for $x(k)$.

DEFINITION 5.8. The hyperdiscretized risk set induced by $D \subseteq {}^*\mathcal{D}$ is the set

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

The following result is straightforward. (For completeness, Appendix K.5.9 contains the proof.)

LEMMA 5.9. Let $D \subseteq {}^*\mathcal{D}$ be an internal * convex set. Then \mathcal{S}^D is an internal * convex set.

Having defined the (hyperdiscretized) risk set, we now describe a set whose intersection with the risk set captures the notion of $\frac{1}{n}$ - $^\circ$ domination, for some standard $n \in \mathbb{N}$. In that vein, for $\Delta \in {}^*\mathcal{D}$ and $n \in \mathbb{N}$, define the $\frac{1}{n}$ -quantant of Δ ,

$$(5.1) \quad Q(\Delta)_n = \left\{x \in \text{IV}({}^*\mathbb{R}^{J_\Theta}) : (\forall k \leq J_\Theta) \left(x_k \leq {}^*r(t_k, \Delta) - \frac{1}{n}\right)\right\}.$$

LEMMA 5.10. Let $\Delta \in {}^*\mathcal{D}$ and $n \in \mathbb{N}$. The set $Q(\Delta)_n$ is internal and * convex and $Q(\Delta)_m \subseteq Q(\Delta)_n$ for every $m < n$.

See Appendix K.5.10 for proof. The following is immediate from definitions.

LEMMA 5.11. Let $\mathcal{C} \subseteq {}^*\mathcal{D}$ and $n \in \mathbb{N}$. Then Δ is $\frac{1}{n}$ - $^\circ$ admissible on T_Θ among \mathcal{C} if and only if $Q(\Delta)_n \cap \mathcal{S}^\mathcal{C} = \emptyset$.

We now introduce an external notion of * convex hull.

DEFINITION 5.12. For every $\mathcal{C} \subseteq {}^*\mathcal{D}$, let $(\mathcal{C})_{FC} = \bigcup_{D \in \mathcal{C}^{[<\infty]}} {}^*\text{conv}(D)$ be the set of all finite * convex combinations of $\Delta \in \mathcal{C}$.

Let $\mathcal{C} \subseteq \mathcal{D}$, $\delta_1, \delta_2 \in \mathcal{C}$ and $p \in {}^*[0, 1]$. Put $\Delta' = p {}^*\delta_1 + (1 - p) {}^*\delta_2$. Then $\Delta' \in ({}^\circ\mathcal{C})_{FC}$, while $\Delta' \in {}^\circ\text{conv}(\mathcal{C})$ if $p \in [0, 1]$. It is easy to see that $({}^\circ\text{conv}(\mathcal{C}))_{FC} = ({}^\circ\mathcal{C})_{FC}$. Thus, ${}^\circ\mathcal{C} \subseteq ({}^\circ\text{conv}(\mathcal{C})) \subseteq ({}^\circ\text{conv}(\mathcal{C}))_{FC} = ({}^\circ\mathcal{C})_{FC} \subseteq ({}^*\text{conv}(\mathcal{C}))$. For $\mathcal{C} \subseteq {}^*\mathcal{D}$, it is straightforward to show that $(\mathcal{C})_{FC}$ is a * convex set containing \mathcal{C} .

The following result is critical to our main theorem, as it allows us to restrict our attention to a set of fixed external cardinality. Its proof can be found in Appendix K.5.13.

LEMMA 5.13. For every convex set $\mathcal{C} \subseteq \mathcal{D}$, ${}^\circ\mathcal{C}$ is an essentially $^\circ$ complete subclass of $({}^\circ\mathcal{C})_{FC}$.

5.2. *Nonstandard Bayes if extended admissible.* Recall that the parameter space Θ is a measurable space such that all singletons are measurable. Thus every hyperfinite subset of ${}^*\Theta$ is a * measurable set. We say that a nonstandard prior Π concentrates on a * measurable set T when $\Pi(T) = 1$. For vectors a, b in a (hyper)finite-dimensional Euclidean space, let $\langle a, b \rangle$ denote the inner product.

LEMMA 5.14. Fix $\Delta \in {}^*\mathcal{D}$ and nonempty $D \subseteq {}^*\mathcal{D}$, and suppose there exists a nonzero vector $v \in \text{IV}({}^*\mathbb{R}^{J_\Theta})$ such that $\langle v, x \rangle \leq \langle v, s \rangle$ for all $x \in \bigcup_{n \in \mathbb{N}} Q(\Delta)_n$ and $s \in \mathcal{S}^D$. There exists a nonstandard prior Π , concentrating on T_Θ and assigning mass $v_k / \|v\|_1$ to t_k for all $k \leq J_\Theta$, and Δ is S -Bayes among D under Π .

PROOF. We first establish that $v_k \geq 0$ for all k . Suppose otherwise, that is, $v_{k_0} < 0$ for some k_0 . Then we can pick a point x in $\bigcup_{n \in \mathbb{N}} Q(\Delta)_n$ whose k_0 -th coordinate is arbitrarily large and negative, causing $\langle v, x \rangle$ to be arbitrary large, a contradiction because $\langle v, s \rangle$ is hyperfinite for all $s \in \mathcal{S}^D$. Hence, all coordinates of v must be nonnegative.

Define $\bar{v} \in \text{IV}(*\mathbb{R}^{J_\Theta})$ by $\bar{v} = v/\|v\|_1$. Because $v \neq 0$ and $v \geq 0$, we have $\bar{v} \geq 0$ and $\|\bar{v}\|_1 = 1$. Therefore, \bar{v} induces a *probability measure Π on $(*\Theta, *\mathcal{B}_\Theta)$, concentrating on T_Θ , and assigning probability \bar{v}_k to t_k for all $k \leq J_\Theta$.

Let $\Delta' \in D$. By hypothesis, $\sum_{k \leq J_\Theta} v_k (*r(t_k, \Delta) - \frac{1}{n}) \leq \sum_{k \leq J_\Theta} v_k *r(t_k, \Delta')$ for all $n \in \mathbb{N}$. Dividing both sides by $\|v\|_1 > 0$, we obtain $*r(\Pi, \Delta) - \frac{1}{n} \leq *r(\Pi, \Delta')$ for all $n \in \mathbb{N}$. Thus, $*r(\Pi, \Delta) \lesssim *r(\Pi, \Delta')$. And so, Δ is S-Bayes among D under Π . \square

The previous result shows that, if a nontrivial hyperplane separates the risk set from every $\frac{1}{n}$ -quantant, for $n \in \mathbb{N}$, then the corresponding procedure is S-Bayes. In order to prove our main theorem, we require a nonstandard version of the hyperplane separation theorem, which we give here. We begin by stating the standard hyperplane separation theorem.

THEOREM 5.15 (hyperplane separation theorem). *For any $k \in \mathbb{N}$, let S_1 and S_2 be two disjoint convex subsets of \mathbb{R}^k , then there exists $w \in \mathbb{R}^k \setminus \{0\}$ such that, for all $p_1 \in S_1$ and $p_2 \in S_2$, we have $\langle w, p_1 \rangle \geq \langle w, p_2 \rangle$.*

Using a suitable encoding of this theorem in first-order logic, the transfer principle yields a hyperfinite version.

THEOREM 5.16. *Fix any $K \in *\mathbb{N}$. If S_1, S_2 are two disjoint internal *convex subsets of $\text{IV}(*\mathbb{R}^K)$, then there exists $W \in \text{IV}(*\mathbb{R}^K) \setminus \{0\}$ such that, for all $P_1 \in S_1$ and $P_2 \in S_2$, we have $\langle W, P_1 \rangle \geq \langle W, P_2 \rangle$.*

See Appendix A.2.1 for a proof.

Recall that our nonstandard model is κ -saturated for some infinite κ .

PROPOSITION 5.17. *Let $\mathcal{C} \subseteq *\mathcal{D}$ be a (necessarily finite or external) set with cardinality less than κ , let $\Delta_0 \in *\mathcal{D}$, and suppose Δ_0 is $^\circ$ -extended admissible on Θ among $(\mathcal{C})_{FC}$. Then, for every hyperfinite set $T \subseteq *\Theta$ containing Θ , Δ_0 is S-Bayes among $(\mathcal{C})_{FC}$ under some nonstandard prior concentrating on T .*

PROOF. Without loss of generality, take $T = T_\Theta$. Let $n \in \mathbb{N}$ and $D \in \mathcal{C}^{[<\infty]}$.

By Lemma 4.3, Δ_0 is $\frac{1}{n}$ - $^\circ$ -admissible on T_Θ among $(\mathcal{C})_{FC}$. Hence, by Lemma 5.11, $Q(\Delta_0)_n \cap \mathcal{S}^{(\mathcal{C})_{FC}} = \emptyset$. By the definition of $(\mathcal{C})_{FC}$, we have $Q(\Delta_0)_n \cap \mathcal{S}^{*\text{conv}(D)} = \emptyset$.

By Lemmas 5.9 and 5.10, $\mathcal{S}^{*\text{conv}(D)}$ and $Q(\Delta_0)_n$ are both internal *convex sets, hence, by Theorem 5.16, there is a nontrivial hyperplane that separates them.

For every $n \in \mathbb{N}$ and $D \in \mathcal{C}^{[<\infty]}$, let $\phi_n^D(v)$ be the formula

$$(v \in \text{IV}(*\mathbb{R}^{J_\Theta})) \wedge (v \neq 0 \wedge (\forall x \in Q(\Delta_0)_n) (\forall s \in \mathcal{S}^{*\text{conv}(D)}) \langle v, x \rangle \leq \langle v, s \rangle),$$

and let $\mathcal{F} = \{\phi_n^D(v) : n \in \mathbb{N}, D \in \mathcal{C}^{[<\infty]}\}$. By the above argument and the fact that (i) $\mathcal{C}^{[<\infty]}$ is closed under finite unions and (ii) the sets $Q(\Delta_0)_n$, for $n \in \mathbb{N}$, are nested, \mathcal{F} is finitely satisfiable. Note that \mathcal{F} has cardinality no more than κ , yet our nonstandard extension is κ -saturated by hypothesis. Therefore, by the saturation principle, there exists a nontrivial hyperplane v satisfying every sentence in \mathcal{F} simultaneously. That is, there exists $v \in \text{IV}(*\mathbb{R}^{J_\Theta})$ such that $v \neq 0$ and, for all $x \in \bigcup_{n \in \mathbb{N}} Q(\Delta_0)_n$ and for all $s \in \bigcup_{D \in \mathcal{C}^{[<\infty]}} \mathcal{S}^{*\text{conv}(D)} = \mathcal{S}^{(\mathcal{C})_{FC}}$, we have $\langle v, x \rangle \leq \langle v, s \rangle$.

Hence, by Lemma 5.14, the normalized vector $v/\|v\|_1$ is well-defined and induces a non-standard prior Π , concentrating on T_Θ , such that Δ_0 is S-Bayes among $(\mathcal{C})_{FC}$ under Π . \square

We are now able to state and prove our main result, identifying extended admissibility with nonstandard Bayes optimality.

THEOREM 5.18. *Let $\mathcal{C} \subseteq \mathcal{D}$ be an essentially complete subclass of $\text{conv}(\mathcal{C})$. For $\delta_0 \in \mathcal{D}$, the following are equivalent:*

1. δ_0 is extended admissible among \mathcal{C} .
2. δ_0 is nonstandard Bayes among \mathcal{C} .
3. $^*\delta_0$ is S-Bayes among $(^{\circ}\mathcal{C})_{FC}$.

Moreover, statement (2) (resp., (3)) is equivalent to the statement: for all hyperfinite sets $T \subseteq ^*\Theta$ containing Θ , there exists a nonstandard prior Π concentrating on T such that δ_0 is nonstandard Bayes among \mathcal{C} under Π , (resp., $^*\delta_0$ is S-Bayes among $(^{\circ}\mathcal{C})_{FC}$ under Π).

Note that a class is an essentially complete subclass of itself. Thus, the equivalence above holds for all convex subsets $\mathcal{C} \subseteq \mathcal{D}$ and, in particular, for $\mathcal{C} = \mathcal{D}$ and $\mathcal{C} = \text{conv}(\mathcal{D}_0)$. Under (RU), the equivalence also holds for $\mathcal{C} = \mathcal{D}_0$.

PROOF. By (1) and Lemma 2.1, δ_0 is extended admissible among $\text{conv}(\mathcal{C})$. Then by Theorem 4.6, $^*\delta_0$ is $^{\circ}$ extended admissible on Θ among $^{\circ}\text{conv}(\mathcal{C})$. It follows from Lemma 5.13 and Proposition 5.17 that, for all hyperfinite sets $T \subseteq ^*\Theta$ containing Θ , $^*\delta_0$ is S-Bayes among $(^{\circ}\mathcal{C})_{FC}$ under Π for some nonstandard prior Π concentrating on T . Hence (3) holds. Because $^{\circ}\mathcal{C} \subseteq (^{\circ}\mathcal{C})_{FC}$, (3) implies (2) by Lemma 5.4.

By Lemma 5.4 and Theorem 5.6, (2) implies $^*\delta_0$ is $^{\circ}$ extended admissible on $^*\Theta$ among $^{\circ}\mathcal{C}$. Then (1) follows from Theorem 4.6. \square

It follows that the class of procedures that are nonstandard Bayes among \mathcal{C} is complete (resp., essentially complete) if and only if the class of procedures that are extended admissible among \mathcal{C} is complete (resp., essentially complete). Theorem 2.2 provides sufficient conditions for the latter.

The remainder of the paper considers implications of this general result and the overall nonstandard Bayes framework. In Section 6, we apply Theorem 5.18 to decision problems with compact parameter spaces and continuous risk functions and show that a decision procedure is extended admissible if and only if it is Bayes. In Section 7, we apply Theorem 5.18 to decision problems with bounded risk functions and show that a decision procedure is extended admissible if and only if it is finitely additive Bayes. In both cases, we prove purely standard results by establishing connections between nonstandard Bayes optimality and (finitely additive) Bayes optimality. We establish connections between nonstandard Bayes optimality and other notions of Bayes optimality (extended Bayes, normal-form generalized Bayes and extensive-form generalized Bayes) in Appendices G to J. We then turn to admissibility in Section 8. By Theorem 5.18, admissible among \mathcal{C} implies nonstandard Bayes among \mathcal{C} for any (essentially complete subclass of a) convex class \mathcal{C} . We consider sufficient conditions in the sequel. Finally, we present examples and potential applications in Sections 9 and 10.

6. Application to compact statistical decision problems. In this section, we apply our main result to the case where (i) Θ is a compact Hausdorff space endowed with its Borel σ -algebra \mathcal{B}_Θ , and (ii) all risk functions are continuous (hence, bounded). Under these hypotheses, we show that the class of extended admissible procedures is precisely the class of

Bayes procedures. This result improves on longstanding results by removing technical conditions on the loss and model. The proof associates every nonstandard prior Π with a standard prior Π_p such that the corresponding Bayes risks are within an infinitesimal. The key tool is Loeb measure theory, which we briefly introduce. See Appendix B for more details.

Given a Hausdorff space (Y, T) and a subset $Z \subseteq {}^*Y$, let $\text{NS}(Z) \subseteq Z$ denote the subset of its near-standard elements and let $\text{st} : \text{NS}({}^*Y) \rightarrow Y$ denote the standard part map taking near-standard elements of *Y to their standard parts. The Hausdorff property ensures that the standard part map is well-defined. In both cases, the notation elides the underlying space Y and the topology T , because the space and topology will always be clear from context. As an abbreviation, we will write ${}^\circ x$ for $\text{st}(x)$. Note that ${}^\circ({}^*y) = y$ for $y \in Y$. Given an internal (hyperfinitely additive) probability space $(\Omega, \mathcal{F}, \mu)$, we write $(\Omega, \overline{\mathcal{F}}_\mu, \overline{\mu})$ for the associated *Loeb space*, that is, the completion of the (standard) probability space corresponding to the unique σ -additive extension of μ (viewed as a *standard* probability measure) to the σ -algebra generated by \mathcal{F} . The Loeb space is well-defined by Theorem B.4.

Every (standard) probability measure π corresponds to a $*$ probability measure, namely its extension ${}^*\pi$. We may then construct the corresponding Loeb space $({}^*\Theta, {}^*\overline{\mathcal{B}}_{\Theta, {}^*\pi}, {}^*\overline{\pi})$. The reverse process is characterized by the following result in Loeb measure theory, which is immediate from Theorems A.10 and B.8 and [10], page 162, Theorem 4.1.

LEMMA 6.1. *Let Θ be compact Hausdorff, let Π be an internal probability measure defined on $({}^*\Theta, {}^*\mathcal{B}_\Theta)$, and let $\mathcal{F} = \{F \subseteq \Theta : \text{st}^{-1}(F) \in {}^*\overline{\mathcal{B}}_{\Theta, \Pi}\}$. Define $\Pi_p : \mathcal{F} \rightarrow [0, 1]$ by $\Pi_p(F) = \overline{\Pi}(\text{st}^{-1}(F))$ for $F \in \mathcal{F}$. Then $\mathcal{B}_\Theta \subseteq \mathcal{F}$ and $(\Theta, \mathcal{F}, \Pi_p)$ is the completion of a regular Borel probability space.*

We call $(\Theta, \mathcal{F}, \Pi_p)$ the *pushdown* of $({}^*\Theta, {}^*\mathcal{B}_\Theta, \Pi)$. (When the underlying spaces are clear, we may refer directly to Π_p as the pushdown of Π .) This result links internal probability measures to (standard) countably additive ones. (See Examples F.2 and F.3 for intuition.)

The following two lemmas connect (standard) integration under a Loeb measure to $*$ integration under the corresponding internal probability measure and (standard) integration under the pushdown, respectively.

LEMMA 6.2 ([9], page 110, Corollary 6.1). *Suppose $(\Omega, \mathcal{F}, \mu)$ is an internal probability space, and $F : \Omega \rightarrow {}^*\mathbb{R}$ is an internal μ - $*$ integrable function such that $\text{st}(F)$ exists everywhere. Then $\text{st}(F)$ is integrable with respect to $\overline{\mu}$ and $\int F \, d\mu \approx \int \text{st}(F) \, d\overline{\mu}$.*

LEMMA 6.3 ([1], Proposition 8.4). *Let Θ be compact Hausdorff space, let Π be an internal probability measure on $({}^*\Theta, {}^*\mathcal{B}_\Theta)$, let Π_p be the pushdown of Π , and let $f : \Theta \rightarrow \mathbb{R}$ be a continuous function. Then $\int f \, d\Pi_p = \int f({}^\circ s) \overline{\Pi}(ds)$.*

By Lemma 6.3, taking f to be a standard risk function, the corresponding standard Bayes risk and nonstandard Bayes risks are infinitely close under a nonstandard prior Π and its pushdown Π_p , if f is continuous. To that end, we introduce the following condition.

CONDITION RC (Risk continuity). r_δ is continuous on Θ , for all $\delta \in \mathcal{D}$.

Theorems E.1 and E.2 give natural conditions that imply **(RC)**.

The nonstandard implications of continuity can be formalized in terms of S -continuity: Let Y and Z be topological spaces. A function $f : {}^*Y \rightarrow {}^*Z$ is S -continuous at $x \in {}^*Y$ if $f(y) \approx f(x)$ for all $y \approx x$. A fundamental result in nonstandard analysis links continuity and S -continuity.

LEMMA 6.4. *Let Y and Z be Hausdorff spaces, where Z is also locally compact, and let $D \subseteq Y$. If a function $f : Y \rightarrow Z$ is continuous on D , then its extension $*f$ is $NS(*Z)$ -valued and S -continuous on $NS(*D)$.*

(See Appendix A.2 for a proof of this classical result.)

We are now in place to compare the nonstandard Bayes risk under a nonstandard prior with the standard Bayes risk under the pushdown of the nonstandard prior.

LEMMA 6.5. *Suppose Θ is compact Hausdorff and (RC) holds. Let Π be a nonstandard prior, let Π_p be its pushdown, and let $\delta_0 \in \mathcal{D}$. Then $r(\cdot, \delta_0)$ is Π_p -integrable and $r(\Pi_p, \delta_0) \approx *r(\Pi, *\delta_0)$, that is, the Bayes risk of δ_0 under Π_p is within an infinitesimal of the internal Bayes risk $*\delta_0$ under Π .*

PROOF. Because Θ is compact Hausdorff, Lemma 6.1 implies that the pushdown Π_p is a well-defined probability measure on some extension of $(\Theta, \mathcal{B}_\Theta)$. By (RC) and Lemma 6.4, for all $t \in *\Theta$, we have $*r(t, *\delta_0) \approx *r(\circ t, *\delta_0) = r(\circ t, \delta_0)$. As Θ is compact and (RC) holds, $r(\cdot, \delta_0)$ is bounded and continuous. Thus, $*r(\cdot, *\delta_0)$ is Π -*integrable. By Lemma 6.2, $\circ(*r(\cdot, *\delta_0))$ is $\bar{\Pi}$ -integrable and

$$(6.1) \quad \int *r(t, *\delta_0)\Pi(dt) \approx \int \circ(*r(t, *\delta_0))\bar{\Pi}(dt) = \int *r(\circ t, *\delta_0)\bar{\Pi}(dt).$$

Finally, by Lemma 6.3, $\int *r(\circ t, *\delta_0)\bar{\Pi}(dt) = \int r(\theta, \delta_0)\Pi_p(d\theta)$. \square

We now present the main result of this section, which identifies extended admissibility and standard Bayes optimality under compactness and continuity conditions. The result is an almost immediate consequence of our general result (Theorem 5.18) and Lemma 6.5. A detailed proof can be found in Appendix K.6.6.

THEOREM 6.6. *Suppose Θ is compact Hausdorff and (RC) holds. Let $\mathcal{C} \subseteq \mathcal{D}$ be an essentially complete subclass of $\text{conv}(\mathcal{C})$. For $\delta_0 \in \mathcal{D}$, the following are equivalent:*

1. δ_0 is extended admissible among \mathcal{C} .
2. δ_0 is extended Bayes among \mathcal{C} .
3. δ_0 is Bayes among \mathcal{C} .

Again, a class of decision procedures is always an essentially complete class of itself and so the equivalence above holds for all convex subsets $\mathcal{C} \subseteq \mathcal{D}$. If (RU) also holds, then the equivalence extends to these statements with \mathcal{D}_0 . Clearly, if the parameter space is compact Hausdorff and the collection \mathcal{C} of decision procedures with continuous risk functions form a complete subclass of \mathcal{D} , then every admissible decision procedure is Bayes among \mathcal{D} .

The next results demonstrates that Theorem 6.6 is tight.

THEOREM 6.7. *The following two statements hold:*

1. *There is a statistical decision problem on a Hausdorff parameter space satisfying (RC) where extended admissibility does not imply Bayes.*
2. *There is a statistical decision problem on a compact Hausdorff parameter space where extended admissibility does not imply Bayes.*

PROOF. Example 9.1 and Example 9.4 provide examples witnessing the truthfulness of these two statements, respectively. \square

7. Application to decision problems with bounded risk. In this section, we consider the setting of bounded risk functions, and show that the class of extended admissible decision procedures are precisely the class of finitely additive Bayes procedures. This yields a new proof of a well-known result by Heath and Sudderth [17]. Thus every finitely additive Bayes procedure is also nonstandard Bayes with respect to some nonstandard prior. (See Appendix C for background on finitely additive probability theory.)

Let $\mathcal{C} \subseteq \mathcal{D}$ and Π be a nonstandard prior on ${}^*\Theta$. Assume δ is nonstandard Bayes among Π under \mathcal{C} . In this section, we construct, from Π , a finitely additive probability measure Π^P in such a way that the internal risk of ${}^*\delta$ under Π is infinitesimally close to the (finitely additive Bayes) risk of δ under Π^P . This then implies that δ is finitely additive Bayes among Π^P under \mathcal{C} and yields another standard characterization of extended admissible procedures.

DEFINITION 7.1. Let (Ω, \mathcal{F}) be a measurable space and let ν be an internal probability measure on $({}^*\Omega, {}^*\mathcal{F})$. The *internal pushdown* of ν is the set function $\nu^P : \mathcal{F} \rightarrow [0, 1]$ defined by $\nu^P(A) = \text{st}(\nu({}^*A))$.

It is straightforward to verify that the internal pushdown ν^P is a finitely additive probability measure on (Ω, \mathcal{F}) . The pushdown and *internal* pushdown of an internal probability measure may differ even when the underlying space is compact Hausdorff (see Example F.4). The following lemma establishes the key connection between an internal probability measure and its internal pushdown.

LEMMA 7.2 ([13], Theorem 5.4). *Let (Ω, \mathcal{F}) be a measurable space. Let ν be an internal probability measure on $({}^*\Omega, {}^*\mathcal{F})$ and let $f : \Omega \rightarrow \mathbb{R}$ be a bounded measurable function. Then we have $\int_{{}^*\Omega} {}^*f \, d\nu \approx \int_{\Omega} f \, d\nu^P$.*

This result motivates the following condition on risk functions.

CONDITION RB (Risk boundedness). r_{δ} is bounded for every $\delta \in \mathcal{D}$.

(RB) is implied by, for example, loss being bounded. By Theorem C.5, Bayes risk under finitely additive priors is then well-defined. The following lemma relates nonstandard and finitely additive Bayes optimality under (RB).

LEMMA 7.3. *Suppose (RB) holds. Let $\mathcal{C} \subseteq \mathcal{D}$, let Π be a nonstandard prior with internal pushdown Π^P , and let $\delta_0 \in \mathcal{D}$ be nonstandard Bayes among \mathcal{C} under Π . Then δ_0 is finitely additive Bayes among \mathcal{C} under Π^P .*

PROOF. Note that the internal pushdown Π^P is a finitely additive prior on Θ . Let $\delta \in \mathcal{C}$. By Lemma 7.2 and the definition of nonstandard Bayes,

$$(7.1) \quad r(\Pi^P, \delta_0) \approx \int_{{}^*\Theta} {}^*r(t, {}^*\delta_0) \Pi(dt) \lesssim \int_{{}^*\Theta} {}^*r(t, {}^*\delta) \Pi(dt) \approx r(\Pi^P, \delta).$$

Thus, δ_0 is finitely additive Bayes among \mathcal{C} under Π^P . \square

We are now in position to establish the main result of this section: The following theorem generalizes Theorem 3.7, due to Heath and Sudderth [17], Theorem 2, although their proof can be seen to establish the same result. The proof is essentially immediate from Lemma 7.3 (see Appendix K.7.4).

THEOREM 7.4. *Let $\mathcal{C} \subseteq \mathcal{D}$ and $\delta_0 \in \mathcal{D}$. If δ_0 is finitely additive Bayes among \mathcal{C} , then δ_0 is extended admissible among \mathcal{C} . If (RB) holds and \mathcal{C} is an essentially complete subclass of $\text{conv}(\mathcal{C})$, then the following are equivalent:*

1. δ_0 is extended admissible among \mathcal{C} .
2. δ_0 is finitely additive Bayes among \mathcal{C} .

For compact Hausdorff Θ , the next result gives an explicit construction of a standard prior from a finitely additive one in a way that preserves Bayes risk, when risk functions are continuous. See Appendix K.7.5 for the proof.

THEOREM 7.5. *Suppose Θ is compact Hausdorff and (RC) holds. Let ν be a finitely additive prior. Then $r((^*\nu)_p, \delta) = r(\nu, \delta)$ for all $\delta \in \mathcal{D}$. Hence, letting $\mathcal{C} \subseteq \mathcal{D}$, if δ_0 is finitely additive Bayes among \mathcal{C} under ν , then δ_0 is Bayes among \mathcal{C} under $(^*\nu)_p$.*

8. Admissibility of nonstandard Bayes procedures. Heretofore, we have focused on the connection between extended admissibility and nonstandard Bayes optimality. We now provide evidence that nonstandard Bayes optimality may be useful for establishing sufficient “Bayesian” conditions for admissibility. Recall that, if δ is the unique Bayes decision procedure under some prior, then δ is admissible. This result does not, in general, hold for improper priors. However, the following nonstandard analogue holds.

THEOREM 8.1. *Let Π be a nonstandard prior and suppose δ_0 is ϵ -*Bayes among $\mathcal{A} \subseteq \mathcal{D}$ under Π . If $r_\delta = r_{\delta_0}$ for every $\delta \in \mathcal{A}$ such that δ is ϵ -*Bayes among $\mathcal{A} \subseteq \mathcal{D}$ under Π , then δ_0 is admissible among \mathcal{A} .*

See Appendix K.8.1 for the proof, and Example 9.3 for an application.

On finite parameter spaces with bounded loss, it is known that procedures that are Bayes under priors assigning positive mass to every state are admissible. More generally, when risk functions are continuous, procedures that are Bayes under priors with full support are admissible. In the remainder of this section, we establish a nonstandard analogue of this result. In order to define the correct notion of support, we must compare infinitesimal numbers. To do so, for $x, y \in {}^*[0, \infty)$, write $x \gg y$ when $x > 0$ and $y/x \approx 0$.

DEFINITION 8.2. Suppose Θ is a Hausdorff space and let $\epsilon \in {}^*\mathbb{R}_{\geq 0}$. A nonstandard prior Π on ${}^*\Theta$ is ϵ -regular if, for every $\theta_0 \in \Theta$ and every open set U containing θ_0 , we have $\Pi({}^*U) \gg \epsilon$.

The following result establishes $^\circ$ admissibility from *Bayes optimality under conditions analogues to full support and continuity of the risk function. Its proof can be found in Appendix K.8.3.

LEMMA 8.3. *Suppose Θ is a Hausdorff space, let $\epsilon \in {}^*\mathbb{R}_{\geq 0}$, $\Delta_0 \in {}^*\mathcal{D}$ and $\mathcal{C} \subseteq {}^*\mathcal{D}$, and assume ${}^*r(\cdot, \Delta)$ is S -continuous on $\text{NS}({}^*\Theta)$ for all $\Delta \in \mathcal{C} \cup \{\Delta_0\}$. If Δ_0 is ϵ -*Bayes among \mathcal{C} under an ϵ -regular nonstandard prior Π , then Δ_0 is $^\circ$ admissible in $\Theta/{}^*\Theta$ among \mathcal{C} .*

The following theorem is an immediate consequence of Lemma 8.3 and is a nonstandard analogue of Blyth’s Method [24], Section 5, Theorem 7.13 (see also [24], Section 5, Theorem 8.7). In Blyth’s method, a sequence of (potentially improper) priors with sufficient support is used to establish the admissibility of a decision procedure. In contrast, a single nonstandard prior can witness the admissibility of a procedure that is merely nonstandard Bayes. The proof of the following theorem can be found in Appendix K.8.3.

THEOREM 8.4. *Suppose Θ is a Hausdorff space. Let $\mathcal{C} \subseteq \mathcal{D}$ be the collection of decision procedures with continuous risk functions. Let $\delta_0 \in \mathcal{C}$, and assume \mathcal{C} is a complete subclass of \mathcal{D} . If there exists $\epsilon \in {}^*\mathbb{R}_{\geq 0}$ such that ${}^*\delta_0$ is ϵ -*Bayes among ${}^*\mathcal{C}$ under an ϵ -regular nonstandard prior Π , then δ_0 is admissible among \mathcal{D} . In particular, δ_0 is admissible among \mathcal{D} if it is Bayes among \mathcal{C} under a prior with full support.*

It can be shown that Theorem 8.4 is no weaker than the standard Blyth method. In particular, the ϵ -regular nonstandard prior Π in Theorem 8.4 can be constructed—via truncation and normalization—from any infinite element in the nonstandard extension of any sequence of (possibly improper) priors that meets the hypotheses of the standard Blyth method. We close with an admissibility result requiring no topological structure, the proof of which can be found in Appendix K.8.5.

THEOREM 8.5. *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 ϵ -*Bayes among ${}^*\mathcal{C}$ under a nonstandard prior Π satisfying $\Pi\{\theta\} \gg \epsilon$ for all $\theta \in \Theta$, then δ_0 is admissible among \mathcal{C} .*

REMARK 8.6. The astute reader may notice that Theorem 8.5 is actually a corollary of Theorem 8.4 provided we adopt the discrete topology on Θ . Changing the topology changes the set of available prior distributions and also changes the set of ϵ -regular nonstandard priors.

9. Some examples. In this section, we highlight some properties of our nonstandard theory through examples. We begin with two classical statistical decision problems.

EXAMPLE 9.1 (Infinite-variance normal). Given one sample from a d -dimensional multivariate normal distribution, consider estimating the mean under squared error loss, assuming the covariance matrix is known to be the identity matrix. By the strict convexity of the loss, the nonrandomized procedures \mathcal{D}_0 are a complete class. We thus focus on this class.

Consider the best shift-invariant (and maximum likelihood) estimator $\delta^M(\mathbf{x}) = \mathbf{x}$, which is minimax (hence, extended admissible) and also (extensive-form) generalized Bayes with respect to right Haar measure (i.e., Lebesgue measure on \mathbb{R}^d , which is also left Haar measure / Jeffrey's prior, in this case). This improper prior is often interpreted as a “flat” or “uniform” prior on \mathbb{R}^d . By Theorem 5.18, extended admissibility and nonstandard Bayes optimality are equivalent, and so this procedure will be nonstandard Bayes.

Before we consider a nonstandard perspective, we recall some standard facts: First, for every $k > 0$, the decision procedure $\delta_k^B(\mathbf{x}) = k^2(k^2 + 1)^{-1}\mathbf{x}$ is Bayes under the full-support prior $\pi_k = \mathcal{N}(0, k^2 I_d)$ and its Bayes risk under π_k is $dk^2(k^2 + 1)^{-1}$. Second, the risk function of δ^M takes a constant value, d , and so this is also the Bayes risk under every prior.

Let $K \in {}^*\mathbb{N} \setminus \mathbb{N}$ and let ${}^*\delta^B$ and ${}^*\pi$ be the extensions of the sequences $\delta_1^B, \delta_2^B, \dots$ and π_1, π_2, \dots , respectively. Note that ${}^*\delta^M(\mathbf{x}) \approx ({}^*\delta^B)_K(\mathbf{x})$ for all $\mathbf{x} \in \text{NS}({}^*\mathbb{R}^d)$. The *normal prior $({}^*\pi)_K$ is “flat” on \mathbb{R}^d in the sense that, at every near-standard real number, the ratio of its probability density to $(2\pi)^{-\frac{d}{2}} K^{-d}$ is within an infinitesimal of 1.

By transfer, the Bayes risk of $({}^*\delta^B)_K$ under $({}^*\pi)_K$ is $dK^2(K^2 + 1)^{-1}$, while the Bayes risk of ${}^*\delta^M$ under $({}^*\pi)_K$ is d . Thus, ${}^*\delta^M$ is ϵ_K -*Bayes among ${}^*\mathcal{D}$ under $({}^*\pi)_K$ for $\epsilon_K = d(K^2 + 1)^{-1}$. Since $\epsilon_K \approx 0$ when K is infinite, this implies that ${}^*\delta^M$ is nonstandard Bayes among \mathcal{D} . This yields an alternative argument showing δ^M is extended admissible (and then minimax, because its risk function is constant).

For $d = 1$, it is easy to check that the prior $({}^*\pi)_K$ is ϵ_K -regular. Because risk functions are continuous [24], Example 7.10, Theorem 8.4 implies that δ^M is admissible for $d = 1$. On the other hand, the prior $({}^*\pi)_K$ is not ϵ_K -regular for $d \geq 2$, and so Theorem 8.4 is silent for

$d \geq 2$. In fact, Stein [28] famously showed that δ^M is admissible for $d = 2$ and inadmissible for $d \geq 3$.

EXAMPLE 9.2 (Invariant estimates of scale). Given n i.i.d. samples from a univariate normal distribution with unknown mean μ and standard deviation σ , consider estimating σ under relative-entropy loss, that is, $\ell((\mu, \sigma), \hat{\sigma}) = \psi(\hat{\sigma}^2/\sigma^2)$, where $\psi(r) = r - \log r - 1$. Again, strict convexity allows us to restrict our attention to the class \mathcal{D}_0 of nonrandomized procedures. Every shift-and-scale invariant decision procedure δ has constant risk and, for some $m > 0$, satisfies $\delta^2(x) = \delta_m^2(x) = m^{-1} \sum_{i=1}^n (x_i - \bar{x})^2$, where $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $\bar{x} = n^{-1} \sum_{i=1}^n x_i$. Consider δ_{n-1} and δ_n , which are (extensive-form) generalized Bayes under right Haar measure ($\sigma^{-1} d\mu d\sigma$) and left Haar measure ($\sigma^{-2} d\mu d\sigma$, that is, Jeffrey’s prior), respectively. The best invariant procedure is δ_{n-1} . By general results of Kudō [22], page 69, and Kiefer [21], δ_{n-1} is minimax and extended Bayes. Indeed, for every $k > 0$, δ_{n-1} is 2^{-k} -Bayes under some truncation (and normalization) π_k of right Haar measure. By saturation, it follows that δ_{n-1} is nonstandard Bayes with respect to $(*\pi)_K$, for every infinite $K \in *\mathbb{N} \setminus \mathbb{N}$, where $*\pi$ is the extension of the sequence (π_1, π_2, \dots) . Because δ_{n-1} has constant risk, a direct proof that δ_{n-1} is nonstandard Bayes among \mathcal{D}_0 would immediately imply its minimaxity. On the other hand, while δ_n is generalized Bayes, it is uniformly dominated by δ_{n-1} , hence not extended admissible (and thus not nonstandard Bayes). This highlights that generalized Bayes optimality does not, in general, imply extended admissibility. (We explore set-theoretic relationships between generalized Bayes and other classes in Appendix I.) In contrast, Theorem 5.18 implies that extended admissibility and nonstandard Bayes optimality are equivalent. For more details, see Appendix L.1.

The next example is a special case of a nonparametric distribution estimation problem studied by Cohen and Kuo [8]. We show how to establish the admissibility of the empirical cumulative distribution function (c.d.f.) using a nonstandard Dirichlet process prior. In fact, Cohen and Kuo [8] raise the possibility of establishing admissibility using a (standard) Dirichlet process prior only to suggest this route is blocked for technical reasons that we are able to circumvent using infinitesimals and hyperfinite sets. Instead, they establish admissibility by a direct inductive argument, that implicitly uses improper prior distributions on distributions with finitely many jumps, defined in terms of improper Dirichlet distributions.

EXAMPLE 9.3 (Nonstandard Dirichlet process prior). Given n i.i.d. samples from an unknown distribution, consider estimating its c.d.f. F under the loss function $\ell(F, \hat{F}) = \int (F - \hat{F})^2 dW$, where W is a nonnull, finite measure on \mathbb{R} and \hat{F} ranges over monotonically nondecreasing cadlag functions taking values in $[0, 1]$. Note that every risk function is bounded by $W(\mathbb{R})$.

Consider the decision procedure δ_n that estimates F by the empirical c.d.f. That is, $\delta_n(x)(t) = \frac{1}{n} \sum_{i=1}^n I(x_i \leq t)$ for every $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $t \in \mathbb{R}$. Then let Π be a *Dirichlet process prior with “base measure” parameter θF_0 , where $\theta \in *(0, \infty)$ and F_0 is a *probability measure concentrating on some hyperfinite set $S \subseteq *\mathbb{R}$. Using the transfer principle, it is straightforward to show that, if Δ is *Bayes under Π , then, for all $x \in S^n$, $\Delta(x) = \Delta_0(x)$ W -a.e., where, for all $t \in *\mathbb{R}$,

$$(9.1) \quad \Delta_0(x)(t) = \frac{\theta F_0(t) + n *\delta_n(x)(t)}{\theta + n}.$$

Note that $\Delta_0(x)(t) \approx *\delta_n(x)(t)$ if $\theta \approx 0$. It is straightforward to use this fact to establish that the excess *Bayes risk of $*\delta_n$ is infinitesimal, hence δ_n is nonstandard Bayes among \mathcal{D}_0 under Π , and thus, by Theorem 5.18, extended admissible. Choosing $\theta \approx 0$ does not, in general,

suffice to establish admissibility. However, for a careful choice of $\theta \approx 0$ and F_0 , we can establish admissibility. In particular, let $S \subseteq {}^*\mathbb{R}$ be a hyperfinite set containing \mathbb{R} , let F_0 be the $*$ uniform probability measure on S and let $\theta \in {}^*(0, \infty)$ satisfies $\theta^2 \ll |S|^{-n}$. By the above argument, δ_n is nonstandard Bayes among \mathcal{D}_0 under Π . By an inductive argument, however, one can show that, up to equivalence in risk, no other procedure is. Thus, by Theorem 8.1, δ_n is admissible. See Appendix L.2 for details.

The next example demonstrates that extended admissibility and Bayes optimality do not necessarily align if we drop the risk continuity assumption, even when the parameter space is compact. We describe a non-Bayes admissible estimator and then identify a nonstandard prior under which it is nonstandard Bayes. The examples hinges on the fact that, on ${}^*(0, 1]$, there are nonstandard priors that concentrate on a point infinitesimally close to zero. Relative to the model, these nonstandard priors behave as if they concentrate on zero, while no standard prior can do so. For a detailed analysis of this example, see Appendix L.3.

EXAMPLE 9.4 (Discontinuous risk functions). Let $X = \{0, 1\}$ and $\Theta = [0, 1]$, the latter viewed as a subset of Euclidean space, hence compact Hausdorff. Define $g : [0, 1] \rightarrow [0, 1]$ by $g(x) = x$ for $x > 0$ and $g(0) = 1$, and let $P_t = \text{Bernoulli}(g(t))$, for $t \in [0, 1]$, where $\text{Bernoulli}(p)$ denotes the distribution on $\{0, 1\}$ with mean $p \in [0, 1]$. Every nonrandomized decision procedure $\delta : \{0, 1\} \rightarrow [0, 1]$ thus corresponds with a pair $(\delta(0), \delta(1)) \in [0, 1]^2$, and so we will express nonrandomized decision procedures as pairs. Let the loss function be $\ell(x, y) = (g(x) - y)^2$. (For every x , the map $y \mapsto \ell(x, y)$ is convex but merely lower semi-continuous on $[0, 1]$. It follows from Lemma D.4 that nonrandomized procedures form an essentially complete class.) It can be shown that the decision procedure $(0, 0)$ is an admissible non-Bayes estimator. Moreover, $(0, 0)$ is nonstandard Bayes among all estimators under any nonstandard prior concentrating on some positive infinitesimal.

10. Promising directions and open problems.

(i) Our main result identifies extended admissibility with nonstandard Bayes optimality, without regularity conditions. Extended admissibility, however, is much weaker than admissibility. Can we also identify admissibility with some refinement of nonstandard Bayes optimality, without relying on regularity conditions? If so, this would provide a new perspective on how Bayes optimality differs from frequentist admissibility. Better understanding may simplify the process of designing admissible procedures and establishing the admissibility of procedures, especially in semiparametric and nonparametric problems, where many current results are inapplicable.

There is a sizable literature on necessary and sufficient conditions for admissibility, but there is no unifying theory unhindered by regularity conditions. One possible approach would be to establish a converse of Theorem 8.1 or Theorem 8.5. Several results in the literature may also point the way to a general result. Among these are necessary and sufficient conditions for admissibility established by Stein [27] and by Hsuan [18], both of which connect admissibility with Bayes optimality with respect to *multiple* priors. Stein's result relies on a notion of compactness on the space of decision procedures, while Hsuan's result relies on the finiteness of the parameter space and convexity. In both cases, we may be able to exploit generic tools from nonstandard analysis that allow one to extend arguments on finite or compact spaces to general spaces.

(ii) It is well known that there exist statistical decision problems in which the minimax equality does not hold or there is no least favorable prior. In some cases where minimax procedures are not Bayes, they are extended Bayes or generalized Bayes, but there is no general characterization in these terms. Yet, every minimax procedure is extended admissible,

and so, our main result already implies that every minimax procedure is nonstandard Bayes. Can we uncover a simpler characterization of the minimax equality using our nonstandard framework? If so, this may yield tighter lower bounds on minimax risk, new techniques to establish minimax optimality and new approaches to design such procedures.

(iii) The class of nonstandard Bayes procedures contains the class of Bayes, extended Bayes, finitely additive Bayes and normal-form generalized Bayes procedures. Obtaining necessary and sufficient conditions that separate or collapses these classes is an important open problem that will shed further light on extended admissibility. Settling the relationship between the class of extended admissible procedures and finitely additive Bayes procedures is of particular interest, given the role of finite additivity in axiomatic developments of Bayesian statistics.

In contrast, the class of nonstandard Bayes procedures is, in general, incomparable to the class of *extensive-form* generalized Bayes procedures. Indeed, Heath and Sudderth [17] show that improper priors could yield decision rules that failed to be extended admissible, even under bounded loss, while we demonstrate an extended admissible procedure that is not extensive-form generalized Bayes (Appendix I.1). Given the possibility of improper priors yielding uniformly inadmissible procedures, a characterization of the subclass of extensive-form generalized Bayes procedures that are extended admissible would provide confidence when using this particular form of Bayes optimality. We discuss some existing literature relating improper and proper priors in Appendix I.2.

Acknowledgments. The authors owe a debt of gratitude to William Weiss for detailed suggestions. We thank Gintarė Džiugaitė, Cameron Freer, and H. Jerome Keisler for early discussions and insights, and thank Nate Ackerman, Robert Anderson, Michael Evans, Jeffrey Negrea, Arno Pauly, and Aaron Smith for feedback on drafts and helpful discussions. We thank the anonymous referees for suggesting reformulations of Theorems 2.2, 5.18 and 6.6, and for the idea of pursuing the results in Section 7 and Appendix J. Finally, the authors would like to thank Peter Hoff for his course notes, which served as our first introduction to the topic. This work was done in part while the authors were visiting the Simons Institute for the Theory of Computing at UC Berkeley.

Funding. This research was made possible through an NSERC Discovery Grant, Connaught Award, and U.S. Air Force Office of Scientific Research Grant #FA9550-15-1-0074.

SUPPLEMENTARY MATERIAL

Supplement to “On extended admissible procedures and their nonstandard Bayes risk” (DOI: [10.1214/20-AOS2026SUPP](https://doi.org/10.1214/20-AOS2026SUPP); .pdf). Background in nonstandard analysis/probability theory, and finitely additive probability; convexity, continuity, some technical examples; other notions of Bayes optimality; proofs; and further details on examples.

REFERENCES

- [1] ANDERSON, R. M. (1982). Star-finite representations of measure spaces. *Trans. Amer. Math. Soc.* **271** 667–687. [MR0654856 https://doi.org/10.2307/1998904](https://doi.org/10.2307/1998904)
- [2] BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. [MR0804611 https://doi.org/10.1007/978-1-4757-4286-2](https://doi.org/10.1007/978-1-4757-4286-2)
- [3] BERGER, J. O. and SRINIVASAN, C. (1978). Generalized Bayes estimators in multivariate problems. *Ann. Statist.* **6** 783–801. [MR0478426](https://doi.org/10.1214/aos/1176344826)
- [4] BHASKARA RAO, K. P. S. and BHASKARA RAO, M. (1983). *Theory of Charges: A Study of Finitely Additive Measures. Pure and Applied Mathematics* **109**. Academic Press [Harcourt Brace Jovanovich, Publishers], New York. [MR0751777](https://doi.org/10.1080/00036818308839077)

- [5] BLACKWELL, D. and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions*. Wiley, New York. MR0070134
- [6] BROWN, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Stat.* **42** 855–903. MR0286209 <https://doi.org/10.1214/aoms/1177693318>
- [7] BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series 9*. IMS, Hayward, CA. MR0882001
- [8] COHEN, M. P. and KUO, L. (1985). The admissibility of the empirical distribution function. *Ann. Statist.* **13** 262–271. MR0773166 <https://doi.org/10.1214/aos/1176346591>
- [9] CUTLAND, N. J. (1997). Nonstandard real analysis. In *Nonstandard Analysis* (L. O. Arkeryd, N. J. Cutland and C. W. Henson, eds.). *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.* **493** 51–76. Kluwer Academic, Dordrecht. MR1603229 https://doi.org/10.1007/978-94-011-5544-1_2
- [10] CUTLAND, N. J., NEVES, V., OLIVEIRA, F. and SOUSA-PINTO, J. (1995). Loeb measure theory. In *Developments in Nonstandard Mathematics* (N. J. Cutland, V. Neves, F. Oliveira and J. Sousa-Pinto, eds.). *Pitman Research Notes in Mathematics Series 336* 151–176. Longman, Harlow. MR1394215
- [11] DUANMU, H. (2017). Applications of nonstandard analysis to Markov processes and statistical decision theory. Ph.D. thesis, Univ. Toronto. Available at <http://hdl.handle.net/1807/89711>. MR3858218
- [12] DUANMU, H. and ROY, D. M. (2021). Supplement to “On extended admissible procedures and their non-standard Bayes risk.” <https://doi.org/10.1214/20-AOS2026SUPP>
- [13] DUANMU, H. and WEISS, W. (2018). Finitely-additive, countably-additive and internal probability measures. *Comment. Math. Univ. Carolin.* **59** 467–485. MR3914713 <https://doi.org/10.14712/1213-7243.2015.270>
- [14] FERGUSON, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. *Probability and Mathematical Statistics 1*. Academic Press, New York. MR0215390
- [15] FREMLIN, D. H. (2009). Real-valued-measurable cardinals. Version 19.9.09. Available at <https://www1.essex.ac.uk/maths/people/fremlin/rvmc.pdf>. Accessed 2021-07-01.
- [16] HALPERN, J. Y. (2010). Lexicographic probability, conditional probability, and nonstandard probability. *Games Econom. Behav.* **68** 155–179. MR2577384 <https://doi.org/10.1016/j.geb.2009.03.013>
- [17] HEATH, D. and SUDDERTH, W. (1978). On finitely additive priors, coherence, and extended admissibility. *Ann. Statist.* **6** 333–345. MR0464450
- [18] HSUAN, F. C. (1979). A stepwise Bayesian procedure. *Ann. Statist.* **7** 860–868. MR0532249
- [19] JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 361–379. Univ. California Press, Berkeley, CA. MR0133191
- [20] KADANE, J. B., SCHERVISH, M. J. and SEIDENFELD, T. (1999). Statistical implications of finitely additive probability. In *Rethinking the Foundations of Statistics* 211.
- [21] KIEFER, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. *Ann. Math. Stat.* **28** 573–601. MR0092325 <https://doi.org/10.1214/aoms/1177706874>
- [22] KUDŌ, H. (1955). On minimax invariant estimates of the transformation parameter. *Natur. Sci. Rep. Ochanomizu Univ.* **6** 31–73. MR0077842
- [23] LECAM, L. (1955). An extension of Wald’s theory of statistical decision functions. *Ann. Math. Stat.* **26** 69–81. MR0067443 <https://doi.org/10.1214/aoms/1177728594>
- [24] LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. MR1639875
- [25] RAIFFA, H. and SCHLAIFER, R. (1961). *Applied Statistical Decision Theory*. *Studies in Managerial Economics*. Division of Research, Graduate School of Business Administration, Harvard Univ., Boston, MA. MR0117844
- [26] SACKS, J. (1963). Generalized Bayes solutions in estimation problems. *Ann. Math. Stat.* **34** 751–768. MR0150908 <https://doi.org/10.1214/aoms/1177704001>
- [27] STEIN, C. (1955). A necessary and sufficient condition for admissibility. *Ann. Math. Stat.* **26** 518–522. MR0070929 <https://doi.org/10.1214/aoms/1177728497>
- [28] STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Vol. I* 197–206. Univ. California Press, Berkeley and Los Angeles. MR0084922
- [29] STONE, M. (1967). Generalized Bayes decision functions, admissibility and the exponential family. *Ann. Math. Stat.* **38** 818–822. MR0226780 <https://doi.org/10.1214/aoms/1177698876>
- [30] WALD, A. (1939). Contributions to the theory of statistical estimation and testing hypotheses. *Ann. Math. Stat.* **10** 299–326. MR0000932 <https://doi.org/10.1214/aoms/1177732144>
- [31] WALD, A. (1947). An essentially complete class of admissible decision functions. *Ann. Math. Stat.* **18** 549–555. MR0023499 <https://doi.org/10.1214/aoms/1177730345>

- [32] WALD, A. (1947). Foundations of a general theory of sequential decision functions. *Econometrica* **15** 279–313. MR0024113 <https://doi.org/10.2307/1905331>
- [33] WALD, A. (1949). Statistical decision functions. *Ann. Math. Stat.* **20** 165–205. MR0044802 <https://doi.org/10.1214/aoms/1177730030>
- [34] WALD, A. (1950). *Statistical Decision Functions*. Wiley, New York. MR0036976