EXISTENCE OF MAXIMIN PRIORS AND STATISTICAL MINIMAX THEOREMS

Paul Delatte

University of Southern California, *delatte@usc.edu*

We provide general conditions for the existence of maximin priors in statistical minimax games over arbitrary parameter spaces. These results complete and extend statistical minimax theorems that only guarantee the existence of minimax and Γ-minimax decision rules. To verify these conditions in practice, a careful choice of topology on the priors is needed. When restricting the search to maximin probability measures, we show that the weak topology is the most suitable choice. We obtain readily verifiable conditions for existence that apply broadly in this case. By contrast, we prove that the vague topology is generally too coarse to obtain direct existence results over finite measures. In spite of these negative results, we show that the vague topology can still be profitably used to verify the general existence conditions for maximin probability measures under the weak topology. This leads to subtle existence arguments that apply even in games with non-tight families of priors over non-compact parameter spaces. All these results are illustrated in the normal mean problem under L^p loss with $p \in [1, +\infty)$ where we recover known existence results and derive new ones. For these games, new properties of the Bayes risk are obtained as well as an extension of the existence of maximin sparse priors from $p = 2$ to any $p \in [1, +\infty)$.

This version: March 26, 2026. We would like to thank Timothy B. Armstrong, Iain M. Johnstone, Éric Marchand, José Luis Montiel Olea, Hyungsik Roger Moon, Geert Ridder, Gilles Stoltz, and the participants of the statistics seminar at the Laboratoire de Mathématiques d'Orsay for helpful comments.

1 Introduction

Statistical games, pitting Nature choosing parameters or priors over these parameters against a Statistician choosing decision rules, are cornerstones of the decision-theoretic approach to statistics. The idea is to recover statistical procedures as the solutions of optimization problems that depend on a more or less adversarial choice by Nature. This game-theoretic foundation, with deep roots in the frequentist paradigm and a bias for conservative procedures, has evolved into a versatile framework. Many statistical problems can be reframed as statistical games by appropriately restricting the strategies of Nature and varying the payoffs of the Statistician.

In this paper, we are concerned with a general mathematical transcription of these games as minimax and maximin optimization problems. Given some measure π chosen by Nature over a measurable space $(\Theta, \mathcal{B}_\Theta)$, the Statistician chooses a Markov kernel from a measurable space $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ to another measurable space $(\mathcal{A}, \mathcal{B}_\mathcal{A})$ to minimize the functional

$$B(\delta, \pi) := \int_\Theta \left(\int_\mathcal{X} \int_\mathcal{A} \ell(a, \theta)\, d\delta(x, a)\, dP_\theta(x) \right) d\pi(\theta)\,,$$

where $\ell \colon \mathcal{A} \times \Theta \to [0, +\infty]$ is a measurable function with respect to $\mathcal{B}_\mathcal{A} \otimes \mathcal{B}_\Theta$ measuring the quality of the random decision $a \in \mathcal{A}$ drawn from $\delta(x, \cdot)$ at any fixed $x \in \mathcal{X}$. The measure π is called a prior, the set Θ the parameter space, the set $\mathcal{X}$ the sample space, the set $\mathcal{A}$ the action space, the kernel δ a decision rule, and the function ℓ the loss. The functional B is called the average risk and the term in parentheses, $r(\delta, \theta) := \int_\mathcal{X} \int_\mathcal{A} \ell(a, \theta)\, d\delta(x, a)\, dP_\theta(x)$, the (frequentist) risk. All these notions and notations are collected in Appendix A. By viewing Nature as playing second in the game, the Statistician faces the following minimax optimization problem

$$\inf_{\delta \in \Delta} \sup_{\pi \in \Gamma} B(\delta, \pi)\,,$$

where Δ is the set of possible decision rules for the Statistician and Γ is the set of possible priors for Nature. A solution to this problem is called a Γ-minimax rule over Δ. By viewing the Statistician as playing second, Nature faces the opposite maximin optimization problem

$$\sup_{\pi \in \Gamma} \inf_{\delta \in \Delta} B(\delta, \pi)\,.$$

A solution to this problem is called a Δ-maximin prior over Γ. For practical and conceptual reasons, it is natural to look for conditions under which the game has an equilibrium in the sense that neither Nature nor the Statistician is advantaged by playing first, that is,

$$V = \inf_{\delta \in \Delta} \sup_{\pi \in \Gamma} B(\delta, \pi) = \sup_{\pi \in \Gamma} \inf_{\delta \in \Delta} B(\delta, \pi).$$

This equality is guaranteed to hold under what are known as statistical minimax theorems, which usually derive from more general minimax theorems. Additional assumptions may be needed to guarantee the existence of a Γ-minimax rule, a Δ-maximin prior, or both.

Most statistical textbooks follow the early contributions of Wald (1950) and Blackwell and Girshick (1954) and provide minimax theorems under restrictive assumptions on the parameter space Θ and the loss ℓ that are often too stringent to be broadly used in applications. In most cases, the loss is assumed continuous in both variables and the parameter space is assumed to be either finite or a compact metric space. Under these conditions, the existence of Γ-minimax rules and Δ-maximin priors is generally guaranteed. Much more general statistical minimax theorems holding under arbitrary parameter spaces were obtained in Le Cam (1955) and reviewed in Le Cam (1986). These results were revisited in Brown (1974) and Strasser (1985) and more recently in Johnstone (2019), but the proofs underlying these results are hard to access and to parse. The first contribution of our paper is to provide an accessible reference to these constructions when $\mathcal{A}$ is a compact metric space, Θ is a Polish metric space (not necessarily compact), and the family $\{P_\theta : \theta \in \Theta\}$ is dominated by a common σ-finite measure. In Proposition 2.3, we recover a statistical inf-sup theorem over any convex subset Γ of positive Radon measures $\mathcal{M}_+(\Theta)$ that is general enough to be used in most games. From it, we directly obtain in Corollary 2.4 the "minimax theorem" of Le Cam for the frequentist risk that obtains when the set of priors Γ is rich enough. From the topological side, these results rely exclusively on the topology constructed for the decision rules. As a result, the existence of Γ-minimax rules is generally guaranteed, but not the existence of Δ-maximin priors. In order to guarantee their existence, topological conditions on the set of priors Γ need to be considered. As far as we know, the problem has never been tackled directly in the context of statistical minimax games. To address it, we start by reviewing in Section 2.2 weak existence conditions commonly considered in infinite-dimensional optimization problems such as in the calculus of variations. This leads us to state a weak version of the Weierstrass theorem that can be directly applied to the Bayes risk $B_\Delta(\cdot) := \inf_{\delta\in\Delta} B(\delta, \cdot)$ to obtain the existence of maximin priors. These conditions are collected in Proposition 2.6: a maximin prior exists over some set $\Gamma \subseteq \mathcal{M}_+(\Theta)$ if there is a metrizable topology on Γ such that a maximizing sequence for the Bayes risk $B_\Delta(\cdot)$ has a subsequence converging to a measure in Γ and $B_\Delta(\cdot)$ is upper semi-continuous at this limit. This generality is needed to cover games where Γ is not compact for the topology. These conditions can hardly be relaxed further given the natural semi-continuity properties of the frequentist risk in most statistical applications. Under these weak existence conditions, we obtain in Proposition 2.7 a general statistical minimax theorem where a saddle point always exists. Additional properties of the saddle points are derived under additional regularity of the games.

From the weak Weierstrass theorem, the existence problem for maximin priors translates into verifying that some subsets of the set of positive Radon measures $\mathcal{M}_+(\Theta)$ are compact and the Bayes risk $B_\Delta(\cdot)$ is upper semi-continuous. Because the set of positive Radon measures $\mathcal{M}_+(\Theta)$ is infinite-dimensional when Θ is not discrete, the choice of a topology on the priors then matters: a coarser topology would yield more compact sets but fewer semi-continuous functions than a finer one. When restricting the search to maximin priors that are finite measures (and, in particular, probability measures), two topologies have commonly been considered: the weak topology, defined by duality with the set of bounded and continuous functions $C_b(\Theta)$, and the vague topology, defined by duality with the set of compactly supported continuous functions $C_c(\Theta)$. To help automate the derivation of existence results in applications, we make clear in Section 3 when it is possible to verify the weak Weierstrass theorem for the Bayes risk $B_\Delta(\cdot)$ under these two topologies. In Section 3.1, we obtain expedient sufficient conditions under the weak topology defined over some subset of the set $\mathcal{M}_1(\Theta)$ of probability measures. These conditions are of wide applicability and cover games with non-compact parameter spaces provided the set of priors satisfies some tightness conditions. When these tightness conditions cannot be verified, the vague topology defined over some subsets of the set $\mathcal{M}_f(\Theta)$ of finite measures offers a natural remedy as a coarser topology. However, we show in Section 3.2 that the weak Weierstrass theorem cannot generally hold in this case as the Bayes risk $B_\Delta(\cdot)$ fails to be vaguely upper semi-continuous in most problems of interest. When restricting attention to subsets of the set $\mathcal{M}_{\leq 1}(\Theta)$ of subprobability measures, this result has a natural and intuitive counterpart: it is (almost) never optimal for Nature to use priors without full mass. These impossibility results are collected in Proposition 3.4, Proposition 3.6, and Proposition 3.7. In spite of these negative answers, we show that the vague topology is still of major interest for obtaining the existence of proper maximin priors, that is, maximin priors that are probability measures. This leads to subtle existence arguments that apply in statistical games with non-compact parameter spaces and priors not satisfying any tightness condition. The idea is to use the vague topology to verify the weak Weierstrass theorem under the weak topology. Indeed, since the two topologies coincide when restricted to measures with the same total mass, it is possible to use the vague topology to obtain vaguely converging subsequences of any maximizing sequence of $B_\Delta(\cdot)$ at no cost. If the limit can be shown to be a probability measure in the set Γ, then the result follows directly. We provide an expedient condition that obtains in practice and from which existence automatically derives without the need to introduce vaguely converging subsequences. To complete these topological considerations, we review and clarify in Section 3.3 some fundamental equivalences and interpretational challenges that obtain when considering different compactifications of the parameter space as an alternative proof strategy. These remarks are of independent interest and should apply to infinite-dimensional optimization problems over finite measures beyond the problem we consider. They help clarify, in particular, similar constructions used in the minimization of the Fisher information.

In Section 4, we illustrate these results by revisiting a fundamental set of statistical games central to estimation: the one-dimensional normal mean problem under L^p loss with $p \in [1, +\infty)$. The problem has been a cornerstone of statistical theory due to its simplicity and some fundamental equivalences with more complicated estimation problems – see Nussbaum (1996) and Johnstone (2019). The games are defined through the normal family $\{N(\theta, 1) : \theta \in \Theta\}$ where $\Theta = \mathbb{R}$. Without loss of generality, the action space $\mathcal{A} = \mathbb{R} \cup \{\infty\}$ is taken to be the one-point compactification of the real line. The decisions are evaluated according to the L^p loss ℓ_p defined for any $p \in [1, +\infty)$ by $\ell_p(a, \theta) = |a - \theta|^p$ and $\ell_p(\infty, \theta) = +\infty$ for all $a \in \mathbb{R}$ and all $\theta \in \Theta$. No restrictions are placed on the Statistician in the sense that her set of strategies Δ is taken to be the set $\mathcal{D}$ of all decision rules for the problem. For these games, some fundamental results on the existence of proper maximin priors have been derived. When no restrictions are placed on Nature in the sense that $\Gamma = \mathcal{M}_1(\mathbb{R})$, it is known that no maximin prior exists. The situation changes dramatically when Nature is appropriately constrained. For priors whose support lies in some compact set, the results of Wald (1950) and

Blackwell and Girshick (1954) apply directly – see, e.g., Ghosh (1964), Casella and Strawderman (1981), Bickel (1981). To go beyond this restrictive setting, the results of Section 3 can be used. For the sake of illustration, we will consider two classical sets of games where the strategies of Nature are no longer restricted to compactly supported priors: (1) the set of priors with uniformly bounded moments; (2) the set of sparse priors (that can be expressed as the mixture of a Dirac measure at zero and any probability measure). Showing that proper maximin priors exist for these two classes is not an exercise in futility, but is motivated by varied statistical applications of independent interest – see, for instance, Feldman (1991) and Donoho and Johnstone (1994) for the first class and Mallows (1980), Bickel (1983), or Donoho, Johnstone, Hoch, and Stern (1992) for the second one. The existence of proper maximin priors for any $p \in [1, +\infty)$ over the first class was proved in Donoho and Johnstone (1994). We recover these results in Section 4.2. The existence of proper maximin priors over sparse priors was proved in the case $p = 2$ in Bickel and Collins (1983). We recover and extend this result to any $p \in [1, +\infty)$ in Section 4.3. To obtain the existence results over the two classes, we derive some essential properties of the Bayes risk for any $p \in [1, +\infty)$. Some of these properties are new and should also be of independent interest. They are collected in Section 4.1. For both classes, additional properties for the maximin priors are derived. In particular, it is shown that the conditions of the general statistical minimax theorem of Section 2.3 are verified, hence guaranteeing that the maximin priors are saddle points for some Γ-minimax rules.

2 From Le Cam's topology to statistical minimax theorems

2.1 Statistical inf-sup theorems à la Le Cam

We review in this section the topology on the set of decision rules introduced in Le Cam (1955, 1986). We use it to recover a general statistical inf-sup theorem and derive from it a simple version of Le Cam's Minimax Theorem. We follow closely the construction of Brown (1974) in the dominated case. The assumptions we use are not the weakest possible, but they yield simpler results and proofs that are general enough to be used in most applications. Throughout the paper, we assume that Θ is a Polish metric space and that all measurable spaces are endowed with their Borel σ-algebra. All notions, notations, and conventions used in the main text of the paper are collected in Appendix A.

Definition 1 (Regular Statistical Decision Problem)**.** A statistical decision problem $\big(\mathcal{X}, \{P_\theta : \theta \in \Theta\}, \mathcal{A}, \ell(\cdot,\cdot)\big)$ is said to be regular if:

1. the action space $\mathcal{A}$ is a compact metric space;

2. the family of probability measures $\{P_\theta : \theta \in \Theta\}$ is dominated by a σ-finite measure P_0 such that the space $L^1(\mathcal{X}, P_0)$ is a separable Banach space.

Lemma 2.1. *Let $\big(\mathcal{X}, \{P_\theta : \theta \in \Theta\}, \mathcal{A}, \ell(\cdot,\cdot)\big)$ be a regular statistical decision problem and $\mathcal{D}$ the set of all decision rules for the problem. Then there exists a topological vector space such that $\mathcal{D}$ is a compact and convex subset of this space and the induced topology on $\mathcal{D}$ is metrizable.*

Proof. The result follows from Theorem 3.9 in Brown (1974) (p.219). We sketch the proof for reference. Consider the space $\mathscr{B}$ of bounded bilinear functionals on $L^1(\mathcal{X}, P_0) \times C(\mathcal{A})$ with the operator norm $\|b\|_{\mathscr{B}}$. The main idea is to identify $\mathcal{D}$ with the closed subset $\mathscr{B}_1^+$ of $\mathscr{B}$ composed of functionals that are positive and satisfy $b(f, \mathbf{1}) = \|f\|_1$ for all $f \geq 0$ where $\mathbf{1}$ is the constant function equal to 1. This set is then seen to be contained in the unit ball of $\mathscr{B}$ which is compact in the weak* topology by the Banach–Alaoglu theorem. The identification is done by way of the bijective mapping $\delta \mapsto b_\delta(f, c) := \int_{\mathcal{X}} \int_{\mathcal{A}} f(x)c(a)\, d\delta(x, a)\, dP_0(x)$. The injectivity of the map is obtained by first deriving the equality $\int_{\mathcal{A}} c(a)\, d\delta_1(x, a) = \int_{\mathcal{A}} c(a)\, d\delta_2(x, a)$ for P_0-a.e. x and then using the separability of $C(\mathcal{A})$ from the compactness of $\mathcal{A}$ to show that $\delta_1(x, \cdot)$ and $\delta_2(x, \cdot)$ agree P_0-a.e. on a dense set of continuous functions. The surjectivity of the map is obtained by first noting that any $b(f, c) \in \mathscr{B}_1^+$ induces a partial map $b_c(f) = b(f, c)$ which is a bounded linear functional on $L^1(\mathcal{X}, P_0)$. From the Riesz representation theorem, there exists a function $h(c, \cdot) \in L^\infty(\mathcal{X}, P_0)$ such that $b(f, c) = \int_{\mathcal{X}} f(x)h(c, x)\, dP_0(x)$. By separability of $C(\mathcal{A})$, the function $h(\cdot, x)$ is seen to be linear for P_0-a.e. x and extends uniquely to a positive continuous linear functional on $C(\mathcal{A})$. From the Riesz representation theorem on the compact space $\mathcal{A}$ and the normalization of b, this induces a unique Borel probability measure $\delta(x, \cdot)$ on $\mathcal{A}$ such that $h(c, x) = \int_{\mathcal{A}} c(a)\, d\delta(x, a)$. The separability of $L^1(\mathcal{X}, P_0)$ (and that of $C(\mathcal{A})$ from the compactness of $\mathcal{A}$) ensures the metrizability of the unit ball of $\mathscr{B}$ endowed with the weak* topology. □

Reference. The idea of the construction can be traced back to Le Cam (1955). For a version in the non-dominated case, see Section 42 in Strasser (1985). Our version for metric spaces in the dominated case is adapted from Section 3 in Brown (1974) following Appendix A in Johnstone (2019). It extends straightforwardly to more general topological spaces. In this case, $\mathcal{A}$ should be assumed second countable and the σ-algebra on $\mathcal{A}$ should be taken as the Baire σ-algebra. The topology induced on $\mathcal{D}$ is often called weak-* or weak or Le Cam's topology. To avoid any confusion with the topology on the priors, we will call it the regular topology.

Remark 2.1. The compactness constraint on the action space $\mathcal{A}$ can be accommodated in most problems at little cost, even in estimation problems with infinite parameter spaces. If Θ is compact, then $\mathcal{A}$ can be directly taken as Θ. If Θ is a locally compact non-compact metric space, then $\mathcal{A}$ can be taken as the one-point compactification of Θ, that is, $\mathcal{A} = \Theta \cup \{\infty\}$. For all natural losses, the action ∞ can be penalized to be always suboptimal. This extension is often inconsequential but may require additional technical considerations since the compactification may affect the convex properties of Θ (provided it had some). An illustration of these facts is given in the normal mean problem under L^p losses in Section 4 where $\mathcal{A} = \mathbb{R} \cup \{\infty\}$ – see notably Remark B.2. The compactification of $\mathcal{A}$ should be contrasted with the redefinition of the problem where the parameter space Θ is itself compactified. This opposite strategy (favoring Nature) is not equivalent and has consequential effects that are both interpretational and mathematical – see Section 3.3.

Lemma 2.2. *Let $\big(\mathcal{X}, \{P_\theta : \theta \in \Theta\}, \mathcal{A}, \ell(\cdot, \cdot)\big)$ be a regular statistical decision problem where the set of all decision rules $\mathcal{D}$ is endowed with the regular topology. If the loss function $\ell(\cdot, \theta)$ is lower semi-continuous on $\mathcal{A}$ for each $\theta \in \Theta$, then the integrated risk $B(\cdot, \pi)$ is lower semi-continuous on $\mathcal{D}$ for any Radon measure $\pi \in \mathcal{M}_+(\Theta)$.*

Proof. The result follows from Corollary 3.11 in Brown (1974) (p.221) and an application of Fatou's lemma. For reference, we complete the proof. Since $\mathcal{A}$ is a compact metric space and $\ell(\cdot, \theta)$ is lower semi-continuous, there is an increasing sequence of continuous functions such that $\lim_{n\to\infty} c_n(a) = \ell(a, \theta)$ for all $a \in \mathcal{A}$. By the monotone convergence theorem, it follows that $r(\delta, \theta) = \sup\{b_\delta(f_\theta, c) : c \in C(\mathcal{A}),\ c(a) \leq \ell(a, \theta) \text{ for all } a \in \mathcal{A}\}$, where f_θ is the density of P_θ with respect to P_0. The functions $\delta \mapsto b_\delta(f_\theta, c)$ are continuous by definition of the regular topology. The risk function $r(\cdot, \theta)$ is then lower semi-continuous as the pointwise supremum of continuous functions. Let $\delta_n \to \delta$ in the regular topology for $\mathcal{D}$. Then $\liminf_{n\to\infty} \int_\Theta r(\delta_n, \theta)\, d\pi(\theta) \geq \int_\Theta \liminf_{n\to\infty} r(\delta_n, \theta)\, d\pi(\theta) \geq \int_\Theta r(\delta, \theta)\, d\pi(\theta)$, where the first inequality follows from lower semi-continuity of the risk function and the second from Fatou's lemma. By definition, we then have $\liminf_{n\to\infty} B(\delta_n, \pi) \geq B(\delta, \pi)$, and so the average risk function $B(\cdot, \pi)$ is lower semi-continuous on $\mathcal{D}$. □

Proposition 2.3 (Statistical Inf-Sup Theorem)**.** *Let $\big(\mathcal{X}, \{P_\theta : \theta \in \Theta\}, \mathcal{A}, \ell(\cdot,\cdot)\big)$ be a regular statistical decision problem. Let $\Delta \subseteq \mathcal{D}$ be a non-empty subset of decision rules endowed with the regular topology. Let $\Gamma \subseteq \mathcal{M}_+(\Theta)$ be a non-empty subset of positive Radon measures. Suppose that:*

1. *Δ is closed and convex;*
2. *Γ is convex;*
3. *the loss function $\ell(\cdot, \theta)$ is lower semi-continuous on $\mathcal{A}$ for each $\theta \in \Theta$.*

Then it holds that

$$\inf_{\delta\in\Delta} \sup_{\pi\in\Gamma} B(\delta, \pi) = \sup_{\pi\in\Gamma} \inf_{\delta\in\Delta} B(\delta, \pi)\,.$$

Moreover, there exists a Γ-minimax decision rule δ^* over Δ, that is,

$$\sup_{\pi\in\Gamma} B(\delta^*, \pi) = \inf_{\delta\in\Delta} \sup_{\pi\in\Gamma} B(\delta, \pi)\,.$$

Proof. The result is a direct application of Kneser's minimax theorem in Kneser (1952). For reference, we complete the proof. By standard properties of the integral, the function $B(\delta, \cdot)$ is affine on the convex set Γ. From the same argument, the function $r(\cdot, \theta)$ is affine on the convex set Δ, and so the function $B(\cdot, \pi)$ is also affine on Δ by linearity of the integral. Since Δ is a closed subset of the compact space $\mathcal{D}$, it is also compact. By Lemma 2.2, the average risk $B(\cdot, \pi)$ is lower semi-continuous. The conditions of Kneser's minimax theorem are thus verified and yield the inf-sup equality. The lower semi-continuity of $B(\cdot, \pi)$ ensures the lower semi-continuity of $\sup_{\pi\in\Gamma} B(\cdot, \pi)$. Since Δ is compact, we obtain the existence of a Γ-minimax rule. □

If Γ is a subset of $\mathcal{M}_1(\Theta)$ that contains all Dirac measures on Θ, then Proposition 2.3 yields a version of the "minimax theorem" of Le Cam – see Theorem 1 in Chapter 2 of Le Cam (1986) or Theorem 46.3 in Strasser (1985). For completeness, we state and prove the version we obtain in this case.

Corollary 2.4 (Inf-Sup Theorem à la Le Cam)**.** *Suppose the assumptions of Proposition 2.3 are verified. If, moreover, $\Gamma \subseteq \mathcal{M}_1(\Theta)$ is a subset of probability measures that contains all Dirac* *measures on* Θ*, then it holds that*

$$\inf_{\delta\in\Delta}\sup_{\theta\in\Theta} r(\delta,\theta) = \sup_{\pi\in\Gamma}\inf_{\delta\in\Delta} B(\delta,\pi)\,.$$

Furthermore, there exists a Θ*-minimax rule* δ_m *over* Δ*, that is,*

$$\sup_{\theta\in\Theta} r(\delta_m,\theta) = \inf_{\delta\in\Delta}\sup_{\theta\in\Theta} r(\delta,\theta)\,.$$

Proof. We directly have that $\sup_{\pi\in\Gamma} B(\delta,\pi) \le \sup_{\theta\in\Theta} r(\delta,\theta)$ for any $\delta\in\Delta$. Without loss of generality, suppose that $\sup_{\theta\in\Theta} r(\delta,\theta) < +\infty$. By definition of the supremum, for any $\varepsilon > 0$, there exists $\theta_\varepsilon \in \Theta$ such that $r(\delta,\theta_\varepsilon) > \sup_{\theta\in\Theta} r(\delta,\theta) - \varepsilon$. We then have that $B(\delta,\mathfrak{d}_{\theta_\varepsilon}) = r(\delta,\theta_\varepsilon) > \sup_{\theta\in\Theta} r(\delta,\theta) - \varepsilon$. By taking a supremum over Γ, we get $\sup_{\pi\in\Gamma} B(\delta,\pi) \ge \sup_{\theta\in\Theta} r(\delta,\theta)$. It follows that $\sup_{\pi\in\Gamma} B(\delta,\pi) = \sup_{\theta\in\Theta} r(\delta,\theta)$ for any $\delta\in\Delta$ and so $\inf_{\delta\in\Delta}\sup_{\pi\in\Gamma} B(\delta,\pi) = \inf_{\delta\in\Delta}\sup_{\theta\in\Theta} r(\delta,\theta)$. We conclude by applying the inf-sup equality of Proposition 2.3. Since $r(\cdot,\theta)$ is lower semi-continuous, $\sup_{\theta\in\Theta} r(\cdot,\theta)$ is lower semi-continuous. The existence of a Θ-minimax rule over Δ is then guaranteed by the compactness of Δ. □

Remark 2.2. The terminology of "minimax theorem" for Corollary 2.4 follows from Le Cam. It corresponds, more precisely, to a min-sup theorem (or what is more often called an inf-sup theorem). Indeed, the existence of a maximin prior on the right-hand side is not guaranteed (nor is the existence of a maximin parameter for that matter). When a Δ-maximin prior π^* exists and the conditions of Corollary 2.4 are verified, then any Θ-minimax rule over Δ has to be a Bayes rule with respect to π^*. This follows from the semi-continuity properties of $r(\cdot,\theta)$, the inf-sup equality of Corollary 2.4, and the fact that $B(\delta,\pi^*) \le \sup_{\theta\in\Theta} r(\delta,\theta)$ is always true for any $\delta\in\Delta$. This is, indeed, enough to guarantee that $\arg\min(\sup_{\theta\in\Theta} r(\cdot,\theta)) \subseteq \arg\min B(\cdot,\pi^*)$. The converse inclusion is generally not true even under the conditions of Corollary 2.4: a Bayes rule for π^* need not be a Θ-minimax rule.

2.2 A general existence result for maximin priors

To bridge the gap in the inf-sup theorems, general conditions for the existence of maximin priors are needed. Given the standard semi-continuity properties of the risk $r(\cdot,\cdot)$, it is natural to derive sufficient conditions from the Weierstrass theorem. Given that the sets of priors under consideration can be infinite-dimensional, relaxing the Weierstrass theorem to its weakest version over metric spaces is required to obtain expedient sufficient conditions for existence in practice.

Lemma 2.5 (Weak Weierstrass Theorem)**.** *Let* E *be a non-empty metric space. If a maximizing sequence* $\{x_n\}$ *for a function* $f : E \to [0,+\infty]$ *has a subsequence converging to a limit* $x^* \in E$ *and* f *is upper semi-continuous at* x^**, then* f *attains its supremum on* E *and* $\sup_{x\in E} f(x) = f(x^*)$.

Proof. Let $\{x_n\}$ denote the subsequence converging to x^* (where the indices are relabeled). Since the original sequence is maximizing, we have $\lim_{n\to\infty} f(x_n) = \sup_{x\in E} f(x)$. Since f is upper semi-continuous at x^*, we have $\limsup_{n\to\infty} f(x_n) \le f(x^*)$, and so $\sup_{x\in E} f(x) \le f(x^*)$. By definition of the supremum, we have $f(x^*) \leq \sup_{x \in E} f(x)$. By combining the two inequalities, we directly get $f(x^*) = \sup_{x \in E} f(x)$. □

This simple lemma is at the basis of all versions of the Weierstrass theorem and is well-known in the infinite-dimensional optimization literature. No direct reference seems to be available, but it is implicitly used, for instance, in the direct method for the calculus of variations – see, e.g., Dacorogna (2008). It implies both the classical Weierstrass theorem when f is upper semi-continuous and E is compact as well as the standard weakened version when E is not compact but f is sup-compact or upper semi-compact (in the sense that its superlevel sets are compact or relatively compact) – see, for instance, Moreau (1967) or Aubin (1979).

Remark 2.3. If E is a finite-dimensional topological vector space, then the choice of the metrizable topology on E is immaterial for existence results. This is no longer the case for infinite-dimensional topological vector spaces such as $\mathcal{M}(\Theta)$. In this case, there exists a fundamental trade-off between the two requirements in the weak Weierstrass theorem: a coarser topology implies the existence of more convergent subsequences but fewer upper semi-continuous functions, and vice versa for a finer topology. This is a cornerstone of infinite-dimensional optimization problems: the choice of a topology matters – see, e.g., Hiriart-Urruty (2012) for a general discussion of the trade-off.

For reference, we can state a general existence result for maximin priors by applying Lemma 2.5 to the Bayes risk B_Δ. The proof is naturally omitted. By way of this result, the difficult part of the existence problem for maximin priors is shifted to verifying some continuity and compactness conditions. Because the set of priors Γ is typically infinite-dimensional, the choice of a topology on Γ matters for existence results as made clear in Remark 2.3. The inherent trade-off between continuity and compactness motivates the discussion and results of Section 3 where different topologies are considered for the existence of proper maximin priors.

Proposition 2.6 (Existence)**.** *Let $\left(\mathcal{X}, \{P_\theta : \theta \in \Theta\}, \mathcal{A}, \ell(\cdot,\cdot)\right)$ be a statistical decision problem. Let $\Delta \subseteq \mathcal{D}$ be a subset of decision rules and $\Gamma \subset \mathcal{M}_+(\Theta)$ a subset of Radon measures endowed with some metrizable topology. If a maximizing sequence $\{\pi_n\}$ for the Bayes risk B_Δ has a subsequence converging to a measure $\pi^* \in \Gamma$ and B_Δ is upper semi-continuous at π^*, then the measure π^* is a Δ-maximin prior over Γ.*

Remark 2.4. The existence condition for maximin priors in Proposition 2.6 does not require any topological conditions on the set of decision rules Δ (nor convex conditions). It is then natural to wonder if it is possible to obtain an inf-sup theorem mirroring Proposition 2.3 but with all topological conditions on the priors (and none on the decision rules). The answer is naturally positive if Γ is compact and $B(\delta, \cdot)$ is upper semi-continuous on Γ for all $\delta \in \Delta$. These conditions can be slightly relaxed (to the existence of a relatively compact superlevel set for $B(\delta_0, \cdot)$ where δ_0 is some decision rule), but not up to the conditions of Proposition 2.6 for the existence of a maximin prior.

2.3 A general statistical minimax theorem

By combining Proposition 2.3 and Proposition 2.6, we can state a general statistical minimax theorem. The existence of a maximin prior in this context guarantees that the problem admits a saddle point. If Γ is sufficiently rich, this also ensures that any Θ-minimax rule over Δ has to be a Bayes rule with respect to the maximin prior forming the saddle point – see Remark 2.2.

Proposition 2.7 (Statistical Minimax Theorem). *Let $\big(\mathcal{X}, \{P_\theta : \theta \in \Theta\}, \mathcal{A}, \ell(\cdot,\cdot)\big)$ be a regular statistical decision problem. Let $\Delta \subseteq \mathcal{D}$ be a non-empty subset of decision rules endowed with the regular topology. Let $\Gamma \subseteq \mathcal{M}_+(\Theta)$ be a non-empty subset of Radon measures endowed with some metrizable topology. Suppose that:*

1. *Δ is closed and convex;*
2. *Γ is convex;*
3. *the loss function $\ell(\cdot,\theta)$ is lower semi-continuous on $\mathcal{A}$ for all $\theta \in \Theta$;*
4. *a maximizing sequence $\{\pi_n\}$ for the Bayes risk B_Δ has a subsequence converging to a measure $\pi^* \in \Gamma$ and B_Δ is upper semi-continuous at π^*.*

Then $\pi^* \in \Gamma$ is a Δ-maximin prior and there exists a Γ-mininax rule $\delta^* \in \Delta$ such that

$$B(\delta^*, \pi^*) = \min_{\delta\in\Delta} \sup_{\pi\in\Gamma} B(\delta,\pi) = \max_{\pi\in\Gamma} \inf_{\delta\in\Delta} B(\delta,\pi)\,. \tag{2.3.1}$$

In particular, the pair $(\delta^*,\pi^*) \in \Delta\times\Gamma$ is a saddle point in the sense that for all $\delta \in \Delta$ and all $\pi \in \Gamma$,

$$B(\delta^*,\pi) \le B(\delta^*,\pi^*) \le B(\delta,\pi^*)\,. \tag{2.3.2}$$

Proof. From Proposition 2.3, we know that the inf-sup equality holds. We denote the common value by V. We know, moreover, that the infimum is attained by some rule $\delta^* \in \Delta$, that is, $\sup_{\pi\in\Gamma} B(\delta^*,\pi) = \inf_{\delta\in\Delta}\sup_{\pi\in\Gamma} B(\delta,\pi)$. From Proposition 2.6, the supremum is also attained by some rule $\pi^* \in \Gamma$, that is, $\inf_{\delta\in\Delta} B(\delta,\pi^*) = \sup_{\pi\in\Gamma}\inf_{\delta\in\Delta} B(\delta,\pi)$. From the inf-sup equality, we have $V = \inf_{\delta\in\Delta} B(\delta,\pi^*) = \sup_{\pi\in\Gamma} B(\delta^*,\pi)$. It follows that $B(\delta^*,\pi^*) \le V \le B(\delta^*,\pi^*)$, and so $V = B(\delta^*,\pi^*)$. We also get directly from the definition of the supremum and the infimum that for all $\pi \in \Gamma$ and $\delta \in \Delta$, $B(\delta^*,\pi) \le B(\delta^*,\pi^*)$ and $B(\delta^*,\pi^*) \le B(\delta,\pi^*)$. □

The saddle point characterization (2.3.2) in Proposition 2.7 naturally rewrites as $B(\delta^*,\pi^*) = \sup_{\pi\in\Gamma} B(\delta^*,\pi) = \inf_{\delta\in\Delta} B(\delta,\pi^*)$. Being a saddle point for any arbitrary pair $(\delta^*,\pi^*) \in \Delta\times\Gamma$ directly implies that δ^* is a Γ-minimax rule over Δ and π^* is a Δ-maximin prior over Γ. This follows directly by combining the two inequalities implied by the saddle point characterization. Moreover, any decision rule δ^* that forms a saddle point (δ^*,π^*) with π^* is necessarily a Bayes rule for π^*. This follows directly from the second equality in the saddle point characterization.

Remark 2.5. The converse is generally not true: a Bayes rule for π^* does not always form a saddle point with π^*, even under the conditions of Proposition 2.7. Indeed, $B(\delta,\pi^*) = \sup_{\pi\in\Gamma} B(\delta,\pi)$ is not necessarily verified for all Bayes rules δ of π^*. If the Bayes rule for π^* is P_0-a.e. unique (as a probability measure) where P_0 is the dominating measure in Definition 1, the converse is true. A sufficient set of conditions for the P_0-a.e. uniqueness of Bayes rules over finite measures is provided in Lemma B.1 and is applied to the normal mean problem of Section 4 in Lemma B.2.

Remark 2.6. The conditions of Lemma B.1 not only guarantee that the Bayes rule of any finite measure is unique P_0-a.e. but also that it is non-randomized P_0-a.e.. The conditions under which a Γ-minimax rule can be taken to be non-randomized (everywhere) are significantly weaker. Suppose the assumptions of Proposition 2.7 hold and $\Delta = \mathcal{D}$. If, moreover, $\mathcal{A}$ is a convex subset of a Fréchet space and the loss function $\ell(\cdot, \theta)$ is convex for every $\theta \in \Theta$, then we are guaranteed that a non-randomized rule δ^* forms a saddle point with π^*. Indeed, for any rule $\delta \in \mathcal{D}$, we can define the non-randomized rule as the Dirac δ_T at $T^*(x) := \int_{\mathcal{A}} a \, d\delta(x, a)$ for any fixed $x \in \mathcal{X}$. This rule is well-defined and belongs to $\mathcal{D}$ by convexity and compactness of $\mathcal{A}$. Then, by convexity of the loss and Jensen's inequality, we directly obtain that $B(\delta_T, \pi) \leq B(\delta, \pi)$ for all $\pi \in \Gamma$. This holds, in particular, for π^* and δ^* in Proposition 2.7, and so $B(\delta_{T^*}, \pi^*) = \inf_{\delta \in \Delta} B(\delta, \pi^*)$ where $T^* := \int_{\mathcal{A}} a \, d\delta^*(x, a)$. Moreover, it is seen that the min-max equality holds also when $\mathcal{D}$ is restricted to non-randomized rules. That is, we have that $B(\delta_{T^*}, \pi^*) = \min_{\delta \in \mathcal{D}_{\text{det}}} \sup_{\pi \in \Gamma} B(\delta, \pi) = \max_{\pi \in \Gamma} \inf_{\delta \in \mathcal{D}_{\text{det}}} B(\delta, \pi)$ where $\mathcal{D}_{\text{det}}$ denotes the restriction of $\mathcal{D}$ to non-randomized rules.

Remark 2.7. A maximin prior is not guaranteed to be unique under the conditions of Proposition 2.7. The Bayes risk B_Δ is only guaranteed to be concave as the pointwise infimum of affine functions, not strictly concave. Conditions for the strict concavity of B_Δ can only be derived under restrictive assumptions on the statistical experiments. An example is provided in Proposition 4.3 in the normal mean problem under L^2 loss. If the maximin prior is unique, then P_0-a.e. uniqueness and non-randomization of the Γ-minimax rule is guaranteed under the conditions of Lemma B.1. Few general uniqueness results can be derived beyond these restrictive conditions.

3 Topologies for the existence of proper maximin priors

We are concerned in this section with the existence of proper maximin priors. As made clear in Section 2, the choice of a topology on the priors matters for the existence of maximin priors. Since the norm topology is too fine to deliver sufficiently many relatively compact sets, the existing literature has resorted to coarser topologies: the weak topology and the vague topology. The goal of this section is to make clear how these two topologies can be validly used to verify the general existence conditions of Proposition 2.6. If the set of priors under consideration contains measures that are not probability ones, then Proposition 2.6 only guarantees the existence of a maximin prior. Another step is needed to ensure that this maximin prior is a probability measure and hence proper.

We consider throughout this section an arbitrary statistical decision problem $\big(\mathcal{X}, \{P_\theta : \theta \in \Theta\}, \mathcal{A}, \ell(\cdot, \cdot)\big)$ and a non-empty set $\Delta \subseteq \mathcal{D}$ of decision rules. We will not make use of topological conditions on the set $\mathcal{D}$ of decision rules to deliver the results of this section.

3.1 Weak topology: direct methods for existence

The weak topology is naturally suited to deliver the existence of proper maximin priors by virtue of: (a) its dual construction making integral functionals of bounded continuous functions directly continuous; and (b) the fundamental link between compactness and tightness provided by Prokhorov's theorem – see Theorem 4.1 in Billingsley (1971). This has made the use of the weak topology prevalent in the existing literature, in particular for problems exhibiting some form of "boundedness".

To make this clear, we shall exhibit expedient sufficient conditions to verify the conditions of Proposition 2.6 when B_Δ is defined over some convex set $\Gamma \subseteq \mathcal{M}_1(\Theta)$ endowed with the weak topology. The restriction to a set of probability measures forces the maximin prior to be proper when it exists. We provide a version of this result for completeness. The proof is immediate and thus omitted. The conditions we provide to verify it will be applied in Section 4 to the normal mean problem where we will recover known existence results and obtain new ones.

Corollary 3.1. *Let $\Gamma \subseteq \mathcal{M}_1(\Theta)$ be a non-empty subset of probability measures endowed with the weak topology. If a maximizing sequence $\{\pi_n\}$ for the Bayes risk B_Δ has a subsequence weakly converging to a measure $\pi^* \in \Gamma$ and B_Δ is weakly upper semi-continuous at π^*, then the measure π^* is a proper Δ-maximin prior over Γ.*

Remark 3.1. If, in addition, the assumptions of Proposition 2.7 are verified. Then we are guaranteed that the inf-sup equality holds and that there exists a Γ-minimax rule over Δ that forms a saddle point with π^*. This notably requires the set Γ of probability measures to be convex.

We start with three sets of conditions of increasing generality under which B_Δ is upper semi-continuous on the whole set Γ. The first idea is to obtain conditions for the upper semi-continuity of the average risk $B(\delta, \cdot)$ and the second idea is to relax these conditions by noticing that the upper semi-continuity of the average risk $B(\delta, \cdot)$ only needs to be verified for a subset of decision rules that approximate the optimal solutions. This last argument can be traced back to Donoho and Johnstone (1994). These conditions are labeled (**U1**), (**U2**), and (**U3**) and are proved below.

(**U1**) For every $\delta \in \Delta$, the function $r(\delta, \cdot)$ is upper semi-continuous and bounded on Θ.

Proof. If $r(\delta, \cdot)$ is upper semi-continuous and bounded on Θ, then by the Portmanteau lemma, we have that for any sequence $\pi_n \in \Gamma$ converging weakly to $\pi_0 \in \Gamma$, $\limsup_{n\to\infty} \int_\Theta r(\delta, \theta)\, d\pi_n(\theta) \leq \int_\Theta r(\delta, \theta)\, d\pi_0(\theta)$. This implies that $B(\delta, \cdot)$ is upper semi-continuous on Γ. Since B_Δ is the pointwise infimum of upper semi-continuous functions, it is also upper semi-continuous on Γ. □

(**U2**) For every $\delta \in \Delta$, the function $r(\delta, \cdot)$ is upper semi-continuous on Θ and uniformly integrable with respect to every compact subset of Γ.

Proof. Pick $\pi_0 \in \Gamma$ and consider a sequence $\{\pi_n\}$ in Γ weakly converging to π_0. By convergence, the set $\Gamma_0 = \{\pi_n\}_{n=1}^{\infty} \cup \{\pi_0\}$ is immediately compact in Γ. For any $M > 0$, define the truncated risk $r_M(\delta, \theta) = \min\{r(\delta, \theta), M\}$. The function $r_M(\delta, \cdot)$ is upper semi-continuous and bounded on Θ. It follows again by the Portmanteau lemma that $\limsup_{n\to\infty} \int_\Theta r_M(\delta, \theta)\, d\pi_n(\theta) \leq \int_\Theta r_M(\delta, \theta)\, d\pi_0(\theta)$.

Since $r(\delta,\cdot)$ is uniformly integrable with respect to Γ_0, for any $\varepsilon > 0$, we can choose M large enough such that $\sup_{\pi\in\Gamma_0} \int_{\{\theta:r(\delta,\theta)>M\}} r(\delta,\theta)\, d\pi(\theta) < \varepsilon$. It follows that $B(\delta,\pi_n) \leq \int_\Theta r_M(\delta,\theta)\, d\pi_n(\theta)+\varepsilon$. We conclude by taking superior limits, $\limsup_{n\to\infty} B(\delta,\pi_n) \leq \int_\Theta r_M(\delta,\theta)\, d\pi_0(\theta)+\varepsilon \leq B(\delta,\pi_0)+\varepsilon$. Since ε is arbitrary, this is enough to conclude that $B(\delta,\cdot)$ is upper semi-continuous on Γ. For the same reason as above, the Bayes risk B_Δ is also upper semi-continuous on Γ. □

(**U3**) There exists a subset Δ_T of Δ such that: for every $\delta \in \Delta_T$, the function $r(\delta,\cdot)$ is upper semi-continuous and bounded on Θ; and, for all $\varepsilon > 0$ and all $\pi \in \Gamma$, there is a rule $\delta_\varepsilon \in \Delta_T$ such that $B(\delta_\varepsilon,\pi) \leq B_\Delta(\pi)+\varepsilon$.

Proof. From the same argument, we again have for any $\delta \in \Delta_T$ and any sequence $\{\pi_n\}$ weakly converging to some $\pi_0 \in \Gamma$ that $\limsup_{n\to\infty} B(\delta,\pi_n) = \limsup_{n\to\infty} \int_\Theta r(\delta,\theta)\, d\pi_n(\theta) \leq \int_\Theta r(\delta,\theta)\, d\pi_0(\theta) = B(\delta,\pi_0)$. We now use the density argument that for any $\varepsilon > 0$, there is a rule $\delta_\varepsilon \in \Delta_T$ such that $B(\delta_\varepsilon,\pi_0) \leq B_\Delta(\pi_0)+\varepsilon$. We trivially have $B_\Delta(\pi_n) \leq B(\delta_\varepsilon,\pi_n)$. This implies by taking limit superiors that $\limsup_{n\to\infty} B_\Delta(\pi_n) \leq B(\delta_\varepsilon,\pi_0) \leq B_\Delta(\pi_0)+\varepsilon$. Since ε is arbitrary, we can conclude that B_Δ is upper semi-continuous on Γ. □

Remark 3.2. The difficult part of condition (**U3**) comes from unearthing a class of decision rules Δ_T that is wide enough for the Bayes risk to approximate the infimum over Δ for any possible prior in Γ. In general, the class is simply obtained by truncation provided the distributions for the model $\{P_\theta : \theta \in \Theta\}$ have sufficiently many bounded moments. This is directly illustrated in Proposition 4.1. It may be noted that the boundedness condition of the risk in (**U3**) can be relaxed to uniform integrability as in (**U2**). This relaxation is often immaterial as the rules Δ_T can often be chosen to have bounded risks by appropriate truncation while preserving the approximation.

We now provide four sets of conditions to ensure that maximizing sequences for the Bayes risk B_Δ have at least one subsequence that weakly converges and whose limit is in Γ. To ensure that the limit stays in Γ, we assume throughout the rest of this section that Γ is weakly closed (if it is not already so by construction). The conditions are in increasing order of generality and are all based on Prokhorov's theorem. They are labeled (**C1**), (**C2**), (**C3**), and (**C4**) and are proved below.

(**C1**) The parameter space Θ is compact.

Proof. If Θ is compact, then $\mathcal{M}_1(\Theta)$ is trivially tight and hence relatively compact by Prokhorov's theorem. Since $\mathcal{M}_1(\Theta)$ is closed under the weak topology, we directly have that $\mathcal{M}_1(\Theta)$ is compact. Since $\Gamma \subseteq \mathcal{M}_1(\Theta)$ is assumed closed, it is also compact. The existence of a subsequence of a maximizing sequence for B_Δ that converges in Γ is then guaranteed by sequential compactness. □

(**C2**) The set of priors Γ is tight.

Proof. If Γ is tight, it is relatively compact by Prokhorov's theorem. Since Γ is also assumed closed, it is hence compact. We conclude similarly as for (C1). □

(**C3**) There exists $c \in [0, +\infty]$ such that $\{\pi \in \Gamma : B_\Delta(\pi) \geq c\}$ is non-empty and B_Δ is c-coercive in the sense that $\limsup_{n\to\infty} B_\Delta(\pi_n) < c$ for any non-tight sequence π_n of priors in Γ.

Proof. Let $c \in [0, +\infty]$ satisfying the conditions and define $S_c := \{\pi \in \Gamma : B_\Delta(\pi) \geq c\}$. Suppose by contradiction that S_c is not tight. Then it contains a non-tight sequence of priors π_n. By assumption, $\limsup_{n\to\infty} B_\Delta(\pi_n) < c$. Since $\pi_n \in S_c$ for all $n \in \mathbb{N}$, we also have $\limsup_{n\to\infty} B_\Delta(\pi_n) \geq c$. A contradiction, hence S_c is tight and so relatively compact by Prokhorov's theorem. This is enough to guarantee the existence of a maximizing sequence for B_Δ with a convergent subsequence. We conclude that the limit of this subsequence is in Γ since Γ is assumed closed. □

(**C4**) There is a tight maximizing sequence for B_Δ.

Proof. By Prokhorov's theorem, this guarantees that the sequence is relatively compact and thus has a convergent subsequence. Since Γ is assumed closed, the limit of the subsequence is in Γ. □

Remark 3.3. The second condition in (**C3**) is typically verified by risk functions that are c-coercive for at least one decision rule in the sense that there is $\delta \in \Delta$ such that for all $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subset \Theta$ such that $\sup_{\theta\in\Theta\setminus K_\varepsilon} r(\delta, \theta) \leq c^* + \varepsilon$ with $c^* < c$. To see this, consider a non-tight sequence π_n in S_c. Then $B_\Delta(\pi_n) \leq \int_\Theta r(\delta, \theta)\, d\pi_n(\theta) \leq \int_{K_\varepsilon} r(\delta, \theta)\, d\pi_n(\theta) + \int_{\Theta\setminus K_\varepsilon} r(\delta, \theta)\, d\pi_n(\theta)$. By taking superior limits, we have $\limsup_{n\to\infty} \int_{K_\varepsilon} r(\delta, \theta)\, d\pi_n(\theta) = 0$ and so we find that $\limsup_{n\to\infty} B_\Delta(\pi_n) \leq c^* + \varepsilon$. Since ε is arbitrary, it follows that $\limsup_{n\to\infty} B_\Delta(\pi_n) < c$.

The conditions above are expedient but do not exhaust the range of valid strategies to verify the conditions of Corollary 3.1. In particular, on the compactness side, indirect methods that do not rely on Prokhorov's theorem can be profitably used. The idea is to consider subsequences converging in a coarser topology than the weak one (which are guaranteed to converge if the topology is coarse enough) and to show that the limit of these subsequences belong to Γ. If the weak topology and the coarser topology coincide on Γ, then the result follows. Indirect arguments of this form allow us to use Corollary 3.1 beyond "bounded" or "tight" problems. These arguments will be considered in Section 3.2 for the vague topology and illustrated in Section 4.3 for the normal mean problem. Before making these arguments clear, we should consider the merits of a more direct strategy consisting in directly verifying the conditions of Proposition 2.6 under the coarser topology.

3.2 Vague topology: impossibility results and indirect methods

If the previous compactness conditions cannot be readily verified under the weak topology, it is natural to consider coarser topologies on the priors in the hope of verifying the existence conditions of Proposition 2.6 under them – see Remark 2.3. The most natural candidate is the vague topology[1] – see, e.g., Lemma 3.2 below. In this case, the set of priors is extended to the vague closure of Γ in the space of subprobability measures $\mathcal{M}_{\leq 1}(\Theta)$.

We first make clear that this direct approach cannot bear fruit: the Bayes risk fails to be vaguely upper semi-continuous in most problems of interest. This is proved in Proposition 3.4 below.

[1] For the vague topology to be well-defined, we assume throughout the rest of the paper that Θ is locally compact.

Therefore, the existence conditions of Proposition 2.6 cannot be verified for the vague topology. As a direct and intuitive counterpart to this result, we show that a strict subprobability measure cannot be a maximin prior in most problems of interest. This is proved in Proposition 3.6 and Proposition 3.7 below.

While these results differ from related optimization problems over finite measures (see Remark 3.5), they do not prevent the use of the vague topology to prove the existence of proper maximin priors. They call, however, for subtle and indirect existence arguments. As suggested at the end of the previous section, the idea is to use the coarser topology to help obtain the compactness condition in Proposition 2.6 for the finer topology (that is, Corollary 3.1). This approach is feasible under the vague topology by virtue of Lemma 3.2 and Lemma 3.3 below. For the strategy to be practical, additional subtleties need to be considered. This will be made clear in the form of condition (**V**) at the end of this section.

Lemma 3.2. *Any subset of* $\mathcal{M}_{\leq 1}(\Theta)$ *is relatively compact with respect to the vague topology.*

Proof. This is proved in Corollary 31.3 in Bauer (2011). □

Lemma 3.3. *If* $\Gamma_0 \subseteq \mathcal{M}_{\leq 1}(\Theta)$ *is tight, then the vague and the weak topologies coincide on* Γ_0.

Proof. This is proved in Theorem 30.8 in Bauer (2011). □

Proposition 3.4. *Let* π *be some measure in* $\Gamma \subseteq \mathcal{M}_{\leq 1}(\Theta)$. *If there exists a sequence* $\{\nu_n\}$ *converging vaguely to the measure zero such that* $\liminf_{n\to\infty} B_\Delta(\nu_n) > 0$ *and* $\pi + \nu_n \in \Gamma$ *for all* $n \in \mathbb{N}$, *then the Bayes risk* B_Δ *is not vaguely upper semi-continuous at* π.

Proof. Define $\{\pi_n\}$ by $\pi_n = \pi + \nu_n$ for all $n \in \mathbb{N}$. By construction, $\{\pi_n\}$ is in Γ and $\pi_n \to \pi$ vaguely. By super-additivity of the infimum, we directly have $B_\Delta(\pi_n) \geq B_\Delta(\pi) + B_\Delta(\nu_n)$. By taking inferior limits, we obtain $\liminf_{n\to\infty} B_\Delta(\pi_n) \geq B_\Delta(\pi) + \liminf_{n\to\infty} B_\Delta(\nu_n)$. Since $\liminf_{n\to\infty} B_\Delta(\nu_n) > 0$ by assumption, we obtain $\liminf_{n\to\infty} B_\Delta(\pi_n) \geq B_\Delta(\pi)$. It follows that $\limsup_{n\to\infty} B_\Delta(\pi_n) \geq B_\Delta(\pi)$ and we conclude by the sequential characterization of upper semi-continuity. □

The previous result directly illustrate the two main conditions under which semi-continuity fails: (a) the unrestricted minimax problem is not trivial; (b) the set Γ allows sequences with escaping masses. The tightness conditions of Section 3.1 typically ensure that (b) fails in the case of probability measures. If we restrict attention to invariant problems, we can provide a practical condition under which (a) holds. This condition makes clear that under invariance semi-continuity fails without tightness constraint on Γ as soon as the minimax problem is non-trivial. To avoid unnecessary complications, we prove the result for $\Theta = \mathbb{R}$. The construction directly generalizes to non-compact locally compact Polish topological groups and their respective group actions.

Lemma 3.5. *Let* $\Theta = \mathbb{R}$ *and* $\mathcal{A} = \mathbb{R} \cup \{\infty\}$. *Suppose that for any* $\delta \in \Delta$ *and any* $c \in \mathbb{R}$, *there exists* $\delta(c) \in \Delta$ *such that* $r(\delta(c), \theta + c) = r(\delta, \theta)$ *for all* $\theta \in \mathbb{R}$. *If there exists a measure* $\nu \in \mathcal{M}_{\leq 1}(\mathbb{R})$ *with compact support such that* $B_\Delta(\nu) > 0$, *then there exists a sequence of measures* $\{\nu_n\}$ *in* $\mathcal{M}_{\leq 1}(\mathbb{R})$ *converging vaguely to the zero measure such that* $\liminf_{n\to\infty} B_\Delta(\nu_n) > 0$.

Proof. Define $\{\nu_n\}$ by $\nu_n(B) = \nu(B - n)$ for any Borel set $B \subseteq \mathbb{R}$. This defines a valid sequence of subprobability measures on $\mathbb{R}$. For any $f \in C_c(\mathbb{R})$, we have $\int_{\mathbb{R}} f(\theta)\, d\nu_n(\theta) = \int_{\mathbb{R}} f(\theta + n)\, d\nu(\theta)$ by a simple change of variable. Since f and ν have compact support, there is a sufficiently large $n \in \mathbb{N}$ such that $f(\theta + n) = 0$ on the support of ν. It follows that $\lim_{n\to\infty} \int_{\mathbb{R}} f(\theta)\, d\nu_n(\theta) = 0$. By the same change of variable, we get $B_\Delta(\nu_n) = \inf_{\delta\in\Delta} \int_{\mathbb{R}} r(\delta, \theta + n)\, d\nu(\theta)$. By the invariance assumption, we obtain that $\inf_{\delta\in\Delta} \int_{\mathbb{R}} r(\delta, \theta + n)\, d\nu(\theta) \geq \inf_{\delta\in\Delta} \int_{\mathbb{R}} r(\delta, \theta)\, d\nu(\theta) = B_\Delta(\nu)$. By taking inferior limits, we finally obtain $\liminf_{n\to\infty} B_\Delta(\nu_n) \geq B_\Delta(\nu) > 0$. □

Proposition 3.4 has an intuitive counterpart: it never makes sense for Nature to eschew mass that could be used to create a harder problem for the Statistician. The next two propositions formalize the idea: a strict subprobability measure cannot be maximin. The condition in the first proposition intuitively says that if Nature can feasibly add a non-trivial problem ν to an existing one, then the original problem π cannot be maximin. This condition applies broadly but it may not allow the subproblem ν to be a Dirac at a point (since the minimax risk is often zero in this case). For more generality, we consider another condition that can handle such subproblems. The result intuitively says that if a specific sequence of decision rules solves the larger problem $\pi + \nu$ but fails to solve the subproblem ν perfectly, then the larger problem is strictly harder than the original one.

Proposition 3.6. *Let $\Gamma \subseteq \mathcal{M}_{\leq 1}(\Theta)$ and $\pi \in \Gamma$. If there exists $\nu \in \Gamma$ such that $\pi + \nu \in \Gamma$ and $B_\Delta(\nu) > 0$, then π cannot be a maximizer of the Bayes risk B_Δ.*

Proof. The proof is immediate from the super-additivity of the infimum. We have $B_\Delta(\pi + \nu) \geq B_\Delta(\pi) + B_\Delta(\nu)$. Since $B_\Delta(\nu) > 0$, it follows directly that $B_\Delta(\pi + \nu) > B_\Delta(\pi)$. □

Proposition 3.7. *Let $\Gamma \subseteq \mathcal{M}_{\leq 1}(\Theta)$ and $\pi \in \Gamma$. If there exists $\nu \in \mathcal{M}_{\leq 1}(\Theta)$ such that $\pi + \nu \in \Gamma$ and $\liminf_{n\to\infty} \int_\Theta r(\delta_n, \theta)\, d\nu(\theta) > 0$ for some minimizing sequence $\{\delta_n\}$ in Δ for $B_\Delta(\pi + \nu)$, then π cannot be a maximizer of the Bayes risk B_Δ.*

Proof. We have $B(\delta_n, \pi + \nu) \geq B_\Delta(\pi) + \int_\Theta r(\delta_n, \theta)\, d\nu(\theta)$. Since $\{\delta_n\}$ is a minimizing sequence for $B_\Delta(\pi + \nu)$, we directly get by taking inferior limits that $B_\Delta(\pi + \nu) \geq B_\Delta(\pi) + \liminf_{n\to\infty} \int_\Theta r(\delta_n, \theta)\, d\nu(\theta)$. By assumption, we directly get that $B_\Delta(\pi + \nu) > B_\Delta(\pi)$. □

Remark 3.4. A minimizing sequence is always guaranteed to exist. When a Bayes rule $\delta(\pi + \nu)$ for $\pi + \nu$ exists, then the condition of Proposition 3.7 simplifies to $\int_\Theta r(\delta(\pi + \nu), \theta)\, d\nu(\theta) > 0$.

The previous set of results makes clear that there is little hope in verifying the conditions of Proposition 2.6 under the vague topology. The topology is too coarse to deliver the semi-continuity of the Bayes risk. Nonetheless, the vague topology can still be profitably used to verify the compactness conditions of Proposition 2.6 under the weak topology, that is, Corollary 3.1. This leads to subtle existence arguments of the following nature:

1. given a maximizing sequence $\{\pi_n\}$ of B_Δ in $\Gamma_0 \subseteq \mathcal{M}_1(\Theta)$, we can always isolate a subsequence converging vaguely with limit π^* in the vague closure of Γ_0 by virtue of Lemma 3.2;

2. if we can prove that $\pi^* \in \Gamma_0$, then the conditions of Proposition 2.6 are verified by virtue of Lemma 3.3.

This argument comes with a subtle caveat: since B_Δ is not vaguely upper semi-continuous on the vague closure of Γ_0, we cannot use the value of $B_\Delta(\pi^*)$ to prove that $\pi^* \in \Gamma_0$. In practice, the following condition obtains and the argument derives directly.

(V) There exists some function G such that $B_\Delta = G$ on the weakly closed set $\Gamma_0 \subseteq \mathcal{M}_1(\Theta)$, the function G is vaguely upper semi-continuous on the vague closure of Γ_0, and there is a maximizer of G (which is guaranteed to exist) that is a proper probability measure.

Proof. Since G is vaguely upper semi-continuous and the vague closure of Γ_0 is necessarily vaguely compact (from Lemma 3.2), the function G attains its supremum on the vague closure of Γ_0. If one of the maximizers is a probability measure, then it necessarily belongs to weakly closed set Γ_0. Given that $B_\Delta = G$ on Γ_0, the maximizer of G is also a maximizer of B_Δ on Γ_0. □

From condition **(V)**, we directly obtain a proper maximin prior without even the need of introducing maximizing sequences. In the background, however, it is exactly what is happening: we obtained from G a maximizing sequence for B_Δ with a convergent subsequence converging vaguely (and hence weakly) to a proper probability measure in Γ_0. If no function G can be unearthed, we can always obtain a subsequence converging vaguely to some limit π^* in the vague closure of Γ_0. However, we cannot use the value $G(\pi^*)$ to show that π^* necessarily belongs to Γ_0. In this case, only the properties of maximizing sequences can be used to prove that π^* is a proper probability measure.

Remark 3.5. The argument exhibited in condition (**V**) underlies the strategy used in Bickel and Collins (1983) to prove the existence of proper maximin priors in some Gaussian statistical games that we will revisit and extend in Section 4.3. The key component of the proof is to find the right extension G of the Bayes risk B_Δ. In the problem considered by Bickel and Collins (1983), this is given by a subtle closed-form equality linking B_Δ to the Fisher information for the marginal distribution – see Lemma 4.2. The results of Bickel and Collins (1983) then build on the existence of minimizers of the Fisher information as proved in Huber (1964). In contrast with the maximization of the Bayes risk, the minimization of the Fisher information behaves well under the vague topology because the Fisher information is vaguely semi-continuous on the whole space of subprobability measures – see Lemma 4.5 and the references below. As we will show, the existence result of Bickel and Collins (1983) can be extended beyond the case where the closed-form equality with the Fisher information exists by finding the right extension G for B_Δ – see Proposition 4.7 and above.

3.3 Equivalences by compactification

The previous results on the vague topology have immediate consequences for another proof strategy commonly encountered in optimization problems over finite measures. The strategy is based on the natural idea of compactifying the parameter space Θ by adding a point at infinity $\{\infty\}$ and considering the weak topology on the set $\mathcal{M}_1(\Theta \cup \{\infty\})$ of probability measures on the compact space $\Theta \cup \{\infty\}$. In this case, any subset of $\mathcal{M}_1(\Theta \cup \{\infty\})$ is relatively compact using the same argument as (**C1**) in Section 3.1 based on Prokhorov's theorem. The following well-known result makes clear that this strategy is strictly equivalent to endowing the space of subprobability measures $\mathcal{M}_{\leq 1}(\Theta)$ with the vague topology.

Lemma 3.8. *Let Θ be a locally compact Polish metric space and denote by $\Theta \cup \{\infty\}$ its one-point compactification. The space $\mathcal{M}_1(\Theta \cup \{\infty\})$ endowed with the weak topology is affinely homeomorphic to the space $\mathcal{M}_{\leq 1}(\Theta)$ endowed with the vague topology.*

Proof. This result is known but no reference seems to be directly available in the literature. For reference, we complete the proof. We show that the mapping $\Psi\colon \mathcal{M}_1(\Theta \cup \{\infty\}) \to \mathcal{M}_{\leq 1}(\Theta)$ where $\Psi(\pi) = \pi_{|\Theta}$ defines an affine homeomorphism. The inverse of Ψ is seen to be given by $\Psi^{-1}(\nu) = \nu + (1 - \nu(\Theta))\mathfrak{d}_\infty$ for any $\nu \in \mathcal{M}_{\leq 1}(\Theta)$. The map Ψ is directly seen to be affine by restriction. For any $\pi \in \mathcal{M}_1(\Theta \cup \{\infty\})$ and any $\nu \in \mathcal{M}_{\leq 1}(\Theta)$, it is directly seen that $\Psi^{-1}(\Psi(\pi)) = \pi$ and $\Psi(\Psi^{-1}(\nu)) = \nu$, so that Ψ is bijective. Consider now a sequence $\pi_n \to \pi$ weakly in $\mathcal{M}_1(\Theta \cup \{\infty\})$ and a sequence $\nu_n \to \nu$ vaguely in $\mathcal{M}_{\leq 1}(\Theta)$. We have that $\int_\Theta f\, d\Psi(\pi_n) = \int_{\Theta \cup \{\infty\}} f\, d\pi_n \to \int_{\Theta \cup \{\infty\}} f\, d\pi = \int_\Theta f\, d\Psi(\pi)$ for any function $f \in C_c(\Theta)$, so that $\Psi(\pi_n) \to \Psi(\pi)$ vaguely. By Proposition 4.36 in Folland (1999), we have that any function $f \in C(\Theta \cup \{\infty\})$ can be decomposed as $f = g + c$ where $g \in C_0(\Theta)$ and $c = f(\infty)$ is a constant. Since $\Psi^{-1}(\nu_n)(\Theta \cup \{\infty\}) = 1$, we have $\int_{\Theta \cup \{\infty\}} f\, d\Psi^{-1}(\nu_n) = \int_\Theta g\, d\nu_n + c$ for any $g + c = f \in C(\Theta \cup \{\infty\})$. Since $C_c(\Theta)$ is dense in $C_0(\Theta)$ and the sequence $\{\nu_n\}$ is uniformly bounded, it follows from the definition of vague convergence that $\int_\Theta g\, d\nu_n \to \int_\Theta g\, d\nu$ for any $g \in C_0(\Theta)$. We thus have $\int_{\Theta \cup \{\infty\}} f\, d\Psi^{-1}(\nu_n) \to \int_{\Theta \cup \{\infty\}} f\, d\Psi^{-1}(\nu)$ for any function $f \in C(\Theta \cup \{\infty\})$, and so $\Psi^{-1}(\nu_n) \to \Psi^{-1}(\nu)$ weakly. □

It follows from Lemma 3.8 that all the results of Section 3.2, both positive and negative, have direct and immediate translations for B_Δ defined on some subset Γ of $\mathcal{M}_1(\Theta \cup \{\infty\})$ endowed with the weak topology. Proposition 3.4 is equivalent to the failure of the weak semi-continuity of B_Δ at limit points supported on $\{\infty\}$. Proposition 3.6 and Proposition 3.7 have direct counterparts in terms of the impossibility for priors supported on $\{\infty\}$ to be maximin. Condition (**V**) has a direct equivalent consisting in the existence of some function G weakly continuous on all priors in Γ and such that $G = B_\Delta$ for all priors in Γ not supported on $\{\infty\}$. The choice of working with $\mathcal{M}_{\leq 1}(\Theta)$ endowed with the vague topology instead of $\mathcal{M}_1(\Theta \cup \{\infty\})$ endowed with the weak topology is thus immaterial for the existence of maximin priors.

Remark 3.6. In practice, working with measures supported on $\{\infty\}$ is not necessarily as foolproof as working with subprobability measures. See, for instance, Remark 4.8 for an illustration in the application of condition (V) on the compactified parameter space in Bickel and Collins (1983). The approach also suffers from interpretational difficulties: in the context of statistical games, it is not immediately clear what strategy the point $\{\infty\}$ is for Nature, especially when the Statistician is also allowed to pick $\{\infty\}$ as an action – see Remark 2.1. When working with subprobability measures supported on Θ, the problem never arises: Nature either uses all its credits to hurt the Statistician ($\|\pi\| = 1$) or spare some ($\|\pi\| < 1$). It is then intuitive that any prior such that $\|\pi\| < 1$ can hardly be an optimal strategy for Nature. The argument is less intuitive when Nature has to use all its credits

($\|\pi\| = 1$) but can put mass at $\{\infty\}$. This approach also leads to a number of error-prone subtleties in the choice of the compactification which we now clarify.

The result of Lemma 3.8 is only valid for the one-point compactification of Θ. However, it is not uncommon in the literature to encounter other compactifications of the parameter space – for instance, the two-point compactification $[-\infty, +\infty]$ when $\Theta = \mathbb{R}$. We can show that the difference between different compactifications is immaterial for the maximization of the Bayes risk B_Δ. Whatever the compactification Θ_C of Θ, it always holds that $\mathcal{M}_1(\Theta \cup \{\infty\})$ is a quotient space for $\mathcal{M}_1(\Theta_C)$ in the weak topology by collapsing all points in the boundary $\Theta_C \setminus \Theta$ to $\{\infty\}$. The addition of points in the compactification may theoretically help obtain the semi-continuity of B_Δ, but it should not in practice. Indeed, we have that B_Δ is weakly continuous on $\mathcal{M}_1(\Theta \cup \{\infty\})$ if and only if B_Δ is weakly continuous on $\mathcal{M}_1(\Theta_C)$ and B_Δ is invariant to redistribution of masses between the points in $\Theta_C \setminus \Theta$. Since any compactification should be chosen so that B_Δ is invariant to redistribution of masses within $\Theta_C \setminus \Theta$, there is no possible gain in considering other compactifications than $\Theta \cup \{\infty\}$.

Remark 3.7. Any compactification Θ_C of the parameter space Θ suffers from the same interpretational problem as $\Theta \cup \{\infty\}$: it is always arcane for Nature to play strategies on the points $\Theta_C \setminus \Theta$.

Remark 3.8. These topological and interpretational considerations extend directly to other optimization problems over finite measures by virtue of the generality of Lemma 3.8. This applies in particular to the problem of minimizing the Fisher information (which should be invariant to redistribution of masses on $\Theta_C \setminus \Theta$ when defined for measures on any compactification Θ_C). In his seminal paper, Huber (1964) directly considered subsets of $\mathcal{M}_{\leq 1}(\mathbb{R})$ endowed with the vague topology. In Bickel and Collins (1983), the authors considered some particular subsets implicitly embedded in $\mathcal{M}_1([-\infty, +\infty])$ endowed with the weak topology. These differences are immaterial for the problem and its solutions. As for the maximization of the Bayes risk, there is no loss but also no reward in adding more than one point in the compactification of the parameter space for the minimization of the Fisher information. In Huber and Ronchetti (2009), the authors informally claimed that "[strict subprobability measures are equivalent] to probability measures that put nonzero mass at $\pm\infty$" (p.76). The equivalence is not properly correct from a topological viewpoint, but the error is inconsequential in any case. What matters for the minimization of the Bayes risk in the problem considered by Bickel and Collins (1983) is the difference of semi-continuity properties between the Fisher information and the Bayes risk on measures supported on $\Theta_C \setminus \Theta$ (or, equivalently, between probability measures and subprobability measures) – see again Section 4.3 and Remark 4.8.

4 Applications: proper maximin priors in the normal mean problem

As an application, we now use the results of Section 3 to automate the derivation of existence results for proper maximin priors in the (univariate) normal mean problem. This location model is simple but a workhorse of statistics by virtue of fundamental equivalences – see, e.g., Nussbaum (1996) and, more generally, Section 3.11 in Johnstone (2019). The nature of the problem is to recover the mean of a normal distribution with known variance equal to 1 based on a single observation, that is,

$$X \sim N(\theta, 1), \quad \theta \in \Theta \subseteq \mathbb{R}.$$

This defines a statistical experiment with $P_\theta = N(\theta, 1)$ and $\mathcal{X} = \mathbb{R}$ where $\Theta \subseteq \mathbb{R}$ is assumed convex. On the decision side, we take $\mathcal{A} = \mathbb{R} \cup \{\infty\}$ as the action space and consider the set $\mathcal{D}$ of all possible decision rules for the problem. Since the experiment is regular in the sense of Definition 1 with the Lebesgue measure λ on $\mathbb{R}$ as the common dominating measure for $\{N(\theta, 1) : \theta \in \mathbb{R}\}$, we can endow $\mathcal{D}$ with the regular topology under which it is compact. In what follows, we should interchangeably use the notations Δ and $\mathcal{D}$. All decisions are then evaluated using the L^p loss for $p \in [1, +\infty)$ defined by $\ell_p(a, \theta) = |a - \theta|^p$ and $\ell_p(\infty, \theta) = +\infty$ for all $a \in \mathbb{R}$ and all $\theta \in \mathbb{R}$.

When Nature is unconstrained in the sense that $\Gamma = \mathcal{M}_1(\mathbb{R})$, it is known that no proper maximin prior exists. To obtain existence results, Nature needs to be constrained further. A simple approach is to consider priors whose supports are contained in some compact subset of $\mathbb{R}$. This leads to existence results directly covered by the results of Wald (1950) and Blackwell and Girshick (1954) – see, for instance, Ghosh (1964); Casella and Strawderman (1981); Bickel (1981). The compactness assumption can be relaxed in several ways. For the sake of exposition, we should be concerned with two classical classes of priors whose supports are not contained in some compact sets:

1. priors with uniformly bounded moments: see Feldman (1991) and Donoho and Johnstone (1994);
2. sparse priors: see Bickel and Collins (1983), Donoho, Johnstone, Hoch, and Stern (1992), and Johnstone (1994).

These classes of priors have independent statistical motivations beyond the mathematical problems they generate – see the references above for more on each class. The existence of proper maximin priors over the first class is known and was proved in full generality in Donoho and Johnstone (1994). We recover this result in Section 4.2. The existence of proper maximin priors over sparse priors is known when $p = 2$ and was proved in this case in Bickel and Collins (1983) (see the clarification of Remark 3.8). We recover this result and extend it to any $p \in [1, +\infty)$ in Section 4.3. The extension is new and should be of independent interest. Some of the intermediary properties of the Bayes risk we obtain to derive these results are also new or more general than previously available and should be of independent interest as well. They are collected in Section 4.1 below.

Remark 4.1. The normal mean problem under L^p loss, $p \in [1, +\infty)$, is a regular statistical experiment for the Lebesgue measure λ on $\mathbb{R}$ – see the proof of Lemma B.2 (where regularity holds unchanged for $p = 1$). It follows that for any finite measure $\pi \in \mathcal{M}_f(\mathbb{R})$, a Bayes rule for π is guaranteed to exist. Since the minimax rule given by the Dirac at x for every $x \in \mathbb{R}$ has constant risk, any Bayes rule for $\pi \in \mathcal{M}_f(\mathbb{R}) \setminus \{0\}$ is supported on some subset of $\mathbb{R}$ for λ-a.e. $x \in \mathbb{R}$ – see the proof of Lemma B.2. When $p \in (1, +\infty)$, the loss is strictly convex and so the Bayes rule δ_π for $\pi \in \mathcal{M}_f(\mathbb{R}) \setminus \{0\}$ is guaranteed to be unique and non-randomized in the sense that $\delta_\pi(x, \cdot)$ is a unique Dirac measure for λ-a.e. $x \in \mathbb{R}$. This is proved in Lemma B.2. Because of this result, we will abuse notation in this case and directly write Bayes rules as measurable functions $\delta_\pi : \mathbb{R} \to \mathbb{R}$. When $p = 2$, a closed-form expression known as Tweedie's formula is available for the Bayes rule

– see (1.2.2) in Brown (1971): for any $\pi \in \mathcal{M}_f(\mathbb{R})$, the Bayes rule δ_π is uniquely given for λ-a.e. $x \in \mathbb{R}$ by $\delta_\pi(x) = x + m'_\pi(x)/m_\pi(x)$ where $m_\pi(x) = \int_{\mathbb{R}} \phi(x-\theta)\, d\pi(\theta)$ is the density of the marginal distribution P_π of X (which is seen to be the density of $\pi * \gamma$). When $p = 1$, the loss is not strictly convex but convex. The λ-a.e. uniqueness of Bayes rules in this case cannot be guaranteed without direct arguments that are beyond the scope of this paper.

Remark 4.2. The choice of $\mathcal{A} = \mathbb{R} \cup \{\infty\}$ for the actions is without loss of generality. When Θ is a strict subset of $\mathbb{R}$, optimal actions are naturally restricted to the closure of Θ in the one-point compactification $\mathbb{R} \cup \{\infty\}$ – see Remark B.2 and Lemma B.2.

Remark 4.3. Given the weak upper semicontinuity of the Bayes risk under L^p losses obtained in Section 4.1, a plethora of new existence results for maximin priors in the normal mean problem can be (artificially) derived by choosing constraints on the priors leading to each condition of Section 3. For the sake of exposition, we will not consider such classes further but restrict attention to the two classical problems introduced above. As we shall see, the existence results for these two classes directly illustrate conditions (**U3**), (**C2**), and (**V**) of Section 3.

4.1 Properties of the Bayes risk function under L^p loss

We start by showing that the Bayes risk $B_\Delta \colon \Gamma \to [0, +\infty]$ is weakly upper semi-continuous over the whole space $\Gamma = \mathcal{M}_1(\mathbb{R})$ under the L^p loss ℓ_p for any $p \in [1, +\infty)$. For reference, we isolate the case $p = 2$ by virtue of a well-known equality that we refer to as the Brown–Tweedie identity. We show, moreover, that in the case $p = 2$ the Bayes risk B_Δ is strictly concave on $\mathcal{M}_1(\mathbb{R})$.

Remark 4.4. If Θ is a convex subset of $\mathbb{R}$, we can identify $\mathcal{M}_1(\Theta)$ with a convex subset of $\mathcal{M}_1(\mathbb{R})$ of measures supported on Θ. This defines an isomorphism that is both affine and continuous for the weak topology. It follows that the positive results on weak semi-continuity and concavity on $\mathcal{M}_1(\mathbb{R})$ directly extends to $\mathcal{M}_1(\Theta)$ viewed both as a subspace of $\mathcal{M}_1(\mathbb{R})$ or as its own space. The same naturally holds for the Brown–Tweedie formula when $p = 2$. It should be noted that the isomorphism is not a homeomorphism under the weak topology unless Θ is closed.

Proposition 4.1. *Let $p \in [1, +\infty)$. The Bayes risk B_Δ is weakly upper semi-continuous on $\mathcal{M}_1(\mathbb{R})$.*

Proof. We exhibit an approximating class Δ_T of decision rules satisfying (**U3**) for the problem. Define Δ_T as the class of non-randomized decision rules that are equal to x outside of a compact set and bounded otherwise, that is, $\Delta_T = \{x \mapsto \delta(x) = x + g(x) : g \text{ bounded, measurable, and with compact support}\}$. We start by showing that for any $\delta \in \Delta_T$, the function $r(\delta, \cdot)$ is continuous and bounded. By definition of Δ_T, there is $K, M > 0$ such that $g(x) = 0$ for $|x| > K$ and $|g(x)| \leq M$ for all $x \in \mathbb{R}$. We also have $r(\delta, \theta) = \int_{\mathbb{R}} |x + g(x+\theta)|^p \phi(x)\, dx$. We then directly get that $r(\delta, \theta) \leq \int_{\mathbb{R}} |x + M|^p \phi(x)\, dx < \infty$. For continuity, the risk rewrites as $r(\delta, \theta) = \int_{-K}^{K} |\delta(x) - \theta|^p \phi(x-\theta)\, dx + \int_{|x+\theta|>K} |x|^p \phi(x)\, dx$. Both the compact part and the tail part are then seen to be continuous by dominated convergence. It remains to show that the Bayes risk B_Δ can be approximated by rules in Δ_T for any prior $\pi \in \mathcal{M}_1(\mathbb{R})$. By definition of the infimum and positivity, for any prior $\pi \in \mathcal{M}_1(\mathbb{R})$, there is a rule $\delta_\pi \in \Delta$ such that

$B(\delta_\pi, \pi) \leq B_\Delta(\pi) + \varepsilon/2$. Moreover, $B(\delta_\pi, \pi) < \infty$ since the rule $\delta(x) = x$ always has finite risk. It follows then that δ_π is finite almost everywhere. For some $K, M > 0$, define δ^* as $\delta^*(x) = x$ if $|x| > K$ and $\delta^*(x) = \max(-M, \min(M, \delta_\pi(x)))$. It is clear that $\delta^* \in \Delta_T$. Since δ_π is finite almost everywhere, we have that $\delta^*(x) \to \delta_\pi(x)$ almost everywhere as $K, M \to \infty$. Using the bound $|\delta^*(x) - \theta|^p \leq 2^{p-1}(|\delta_\pi(x) - \theta|^p + |x - \theta|^p)$, we can directly conclude by dominated convergence that $B(\delta^*, \pi) \to B(\delta_\pi, \pi)$ as $K, M \to \infty$. This is enough to conclude that $B(\delta^*, \pi) \leq B_\Delta(\pi) + \varepsilon$. The upper semi-continuity of B_Δ then follows from (**U3**). □

Remark 4.5. The proof of upper semi-continuity is a direct illustration of (**U3**). By Remark 4.4, the weak upper semi-continuity holds for any convex subset $\Theta \subseteq \mathbb{R}$. If $\Theta = [-m, m]$ for some $m > 0$, then the proof can be made simpler as it falls directly under (**U1**).

Lemma 4.2. *Let $p = 2$. Then the Brown–Tweedie identity holds on $\mathcal{M}_1(\mathbb{R})$, that is, for any $\pi \in \mathcal{M}_1(\mathbb{R})$,*

$$B_\Delta(\pi) = 1 - I(\pi * \gamma)\,.$$

Moreover, the Bayes risk B_Δ is weakly continuous on $\mathcal{M}_1(\mathbb{R})$.

Proof. This is proved in Proposition 4.5 and Lemma 4.7 in Johnstone (2019). For completeness, we revisit the proof. For any $\pi \in \mathcal{M}_1(\mathbb{R})$, the Bayes rule δ_π for π exists, is unique, and non-randomized (λ-a.e.) – see Remark 4.1. From Tweedie's formula, we have that $\delta_\pi(x) = x + m'_\pi(x)/m_\pi(x)$ for λ-a.e. $x \in \mathbb{R}$ where $m_\pi(x) = \int_{\mathbb{R}} \phi(x-\theta)\, d\pi(\theta)$ is the density of $\pi * \gamma$ and the marginal density of X. It follows then that $B_\Delta(\pi) = \int_{\mathbb{R}} \int_{\mathbb{R}} (\delta_\pi(x) - \theta)^2 \phi(x - \theta)\, dx\, d\pi(\theta)$. By expanding the square and integration by parts, we directly obtain the identity $B_\Delta(\pi) = 1 - \int_{\mathbb{R}} ((m'_\pi(x))^2/m_\pi(x))\, dx = 1 - I(\pi * \gamma)$. We now show that $I(\pi * \gamma)$ is weakly continuous by taking a sequence $\{\pi_n\}$ converging to a limit π_0. The integrand $(m'_{\pi_n}(x))^2/m_{\pi_n}(x)$ of $I(\pi_n * \gamma)$ is seen to converge pointwise by definition of weak convergence applied to the bounded and continuous functions $\phi(x - \cdot)$ and $\phi'(x - \cdot)$. Moreover, by the Cauchy–Schwarz inequality, we have that $(m'_{\pi_n}(x))^2/m_{\pi_n}(x) \leq \int_{\mathbb{R}} (x - \theta)^2 \phi(x - \theta)\, d\pi_n(\theta) =: H(x, \pi_n)$. Again, by definition of weak convergence, we have that $H(x, \pi_n)$ converges pointwise. Since $\int_{\mathbb{R}} H(x, \pi)\, dx = 1$ for any $\pi \in \mathcal{M}_1(\mathbb{R})$, we can conclude that $I(\pi_n * \gamma) \to I(\pi * \gamma)$ by the generalized dominated convergence theorem (also known as Young's theorem) – see Theorem 2.8.8 in Bogachev (2007). This yields the weak continuity of B_Δ and concludes the proof. □

Proposition 4.3. *Let $p = 2$. The Bayes risk B_Δ is strictly concave on $\mathcal{M}_1(\mathbb{R})$.*

Proof. The average risk $B(\delta, \cdot)$ is affine over $\mathcal{M}_1(\mathbb{R})$. The Bayes risk B_Δ is then concave over $\mathcal{M}_1(\mathbb{R})$ as the pointwise infimum of affine functions. We then have that B_Δ fails to be strictly concave if and only if there exist two probability measures $\pi_1 \neq \pi_2$, some convex weight $t \in (0, 1)$, and a decision rule δ that is simultaneously Bayes for π_1, π_2, and $t\pi_1 + (1 - t)\pi_2$. We prove that this is impossible by showing that the map $\pi \to \delta_\pi$ is injective. The marginal density m_π for π is the density of $\pi * \gamma$. Since the Fourier transform of a Gaussian is non-zero everywhere, the Fourier transform of π rewrites uniquely in terms of the inverse of the Fourier transform of γ. By uniqueness of characteristic functions, the map $\pi \to m_\pi$ is necessarily injective. We now prove that the map $m_\pi \to \delta_\pi$ is also injective. For any $\pi \in \mathcal{M}_1(\mathbb{R})$, we have by Tweedie's formula that the Bayes rule for π is given by $\delta_\pi(x) = x + m'_\pi(x)/m_\pi(x)$ for λ-a.e. $x \in \mathbb{R}$. If $\delta_{\pi_1} = \delta_{\pi_2}$, then $m_{\pi_1}(x) = e^C m_{\pi_2}(x)$ after integrating. Since the marginal densities are densities of probability measures, we necessarily have $C = 0$. This yields the injectivity of $m_\pi \to \delta_\pi$ and we can conclude that B_Δ is strictly concave. □

Remark 4.6. The strict concavity of the Bayes risk B_Δ for general L^p losses, $p \in (1, +\infty)$, is likely valid but requires convoluted analytical arguments that are beyond the objective of this paper. To see the difficulty, we sketch a valid extension when $p = 4, 6, \ldots$ is an even integer. In this case, Tweedie's formula is no longer valid. Nonetheless, we directly retain the concavity of B_Δ, the (a.e.) uniqueness, non-randomization, and analyticity of the Bayes rule, and the injectivity of the map $\pi \to m_\pi$. Only the injectivity of the map $m_\pi \to \delta_\pi$ needs to be proved. For $p = 4, 6, \ldots$ an even integer, the optimality of the Bayes rule leads to the integral equation $\int_{\mathbb{R}} (\delta(x) - \theta)^{p-1} e^{x\theta} e^{-\theta^2/2}\, d\pi(\theta) = 0$ for all $x \in \mathbb{R}$. By binomial expansion and identification of the terms, this leads to a $(p-1)$-th order ODE for the marginal density m_π in terms of the Bayes rule δ_π which can be seen as a natural extension of Tweedie's formula. The uniqueness of the solution obtains by virtue of the constraints on m_π as a probability density obtained from convolution with a Gaussian. For general $p \in (1, +\infty)$, the implicit integral expression for the Bayes rule does not lead to a simple ODE for m_π and requires working directly with the integral equation for the prior π.

We now show that, for any $p \in [1, +\infty)$, the Bayes risk fails to be vaguely upper semi-continuous at any measure in $\mathcal{M}_{\le 1}(\mathbb{R}) \setminus \mathcal{M}_1(\mathbb{R})$. Since the Fisher information is vaguely lower semi-continuous on $\mathcal{M}_{\le 1}(\mathbb{R})$, this implies, when $p = 2$, that the Brown–Tweedie identity fails for any measure in $\mathcal{M}_{\le 1}(\mathbb{R}) \setminus \mathcal{M}_1(\mathbb{R})$. These negative results are a direct illustration of the results of Section 3.2.

Proposition 4.4. *Let $p \in [1, +\infty)$. The Bayes risk B_Δ is not vaguely upper semi-continuous at any measure in $\mathcal{M}_{\le 1}(\mathbb{R}) \setminus \mathcal{M}_1(\mathbb{R})$.*

Proof. Consider the sequence $\nu_n = N(0, n)$ for any $n \in \mathbb{N}$. If f is a continuous function whose support is contained in $[-K, K]$ for some $K > 0$, then $\left|\int_{\mathbb{R}} f(\theta)\, d\nu_n(\theta)\right| \le \sup_{\theta \in [-K,K]} |f(\theta)| \int_{-K}^{K} n^{-1/2} \phi(n^{-1/2}\theta)\, d\theta \le \sup_{\theta\in[-K,K]} |f(\theta)| 2K(2\pi n)^{-1/2}$. It follows that ν_n converges vaguely to the zero measure 0. Since the loss is symmetric and convex for any $p \ge 1$ and the posterior distribution under a Gaussian prior is itself Gaussian, the Bayes rule for ν_n is (λ-a.e.) unique and given pointwise by the posterior mean. By direct calculations, it follows that $B_\Delta(\nu_n) = (n/(n+1))^{p/2} m_p$ where $m_p > 0$ is the p-th absolute moment of the standard normal distribution. By taking limits, we have that $\lim_{n\to\infty} B_\Delta(\nu_n) = m_p$. Consider now any $\pi \in \mathcal{M}_{\le 1}(\mathbb{R}) \setminus \mathcal{M}_1(\mathbb{R})$. Define $\alpha := \|\pi\| < 1$ and $\pi_n := \pi + (1-\alpha)\nu_n$ for all $n \in \mathbb{N}$. The sequence $\{\pi_n\}$ satisfies the conditions of Proposition 3.4 and so B_Δ is not upper semi-continuous at π. □

Remark 4.7. We constructed the sequence $\nu_n = N(0, n)$ from scratch, but, by invariance, we could also have used Lemma 3.5 to directly guarantee the existence of such a sequence.

Lemma 4.5. *The function $I(\cdot * \gamma)$ is vaguely lower semi-continuous on $\mathcal{M}_{\le 1}(\mathbb{R})$.*

Proof. This follows from Theorem 4.2 in Huber and Ronchetti (2009) and the fact that convolution with the Gaussian measure γ is vaguely continuous on $\mathcal{M}_{\leq 1}(\mathbb{R})$. For completeness, we provide a self-contained proof. Since, for any fixed $x \in \mathbb{R}$, the functions $y \mapsto \phi(x-y)$ and $y \mapsto \phi'(x-y)$ belong to $C_0(\mathbb{R})$ and $\int_{\mathbb{R}} \phi(x-y)\, d\pi(y)$ is strictly positive whenever $\pi \neq 0$, we can conclude directly by definition of vague convergence, pointwise convergence of the integrand, and Fatou's lemma. □

Corollary 4.6. *Let $p = 2$. The Brown–Tweedie identity does not hold on any measure in $\mathcal{M}_{\leq 1}(\mathbb{R}) \setminus \mathcal{M}_1(\mathbb{R})$. That is, for any $\pi \in \mathcal{M}_{\leq 1}(\mathbb{R}) \setminus \mathcal{M}_1(\mathbb{R})$,*

$$B_\Delta(\pi) \neq 1 - I(\pi * \gamma)\,.$$

Proof. Since the functional $I(\cdot * \gamma)$ obtained as the Fisher information of finite measures convoluted with the Gaussian one is vaguely lower semi-continuous, the equality cannot hold for $\pi \in \mathcal{M}_{\leq 1}(\mathbb{R}) \setminus \mathcal{M}_1(\mathbb{R})$ by virtue of Proposition 4.4. The result follows also directly from noticing that both $B_\Delta(\cdot)$ and $I(\cdot * \gamma)$ scale linearly for measures $\alpha\pi$ with $\pi \in \mathcal{M}_1(\mathbb{R})$ and $\alpha \in [0,1]$. The linearity of the scaling then prevents the inequality from holding whenever $\alpha \neq 1$. □

Remark 4.8. In view of the equivalence provided by Lemma 3.8 and the discussion that follows, the previous result directly has an equivalent formulation for measures defined on $\mathbb{R} \cup \{\infty\}$ or $[-\infty, +\infty]$. In other words, Corollary 4.6 implies that the Brown–Tweedie formula is invalid for measures supported on $\{\infty\}$ or $\{-\infty, +\infty\}$. This difference has often been overlooked in the existence literature – see, for instance, (1.8) in Bickel and Collins (1983). Fortunately, this technical detail has been without consequences for existence results by virtue of the valid application of condition (**V**) in the background – see Lemma 4.11 and Corollary 4.12.

When $p = 2$, we have the existence of some function G_2 such that $B_\Delta = G_2$ on $\mathcal{M}_1(\mathbb{R})$ and G_2 is vaguely upper semi-continuous on $\mathcal{M}_{\leq 1}(\mathbb{R})$. A function satisfying these properties is given by $G_2(\cdot) = 1 - I(\cdot * \gamma)$ from the Brown–Tweedie identity – see Lemma 4.2 and Lemma 4.5. It is natural to wonder if similar extensions are valid for any $p \in [1, +\infty)$. The existence of closed-form functions for any $p \neq 2$ is unlikely – see Remark 4.6. However, it is still possible to find a vaguely upper semi-continuous extension of B_Δ on $\mathcal{M}_{\leq 1}(\mathbb{R})$. We define for any $p \in [1, +\infty)$, the function G_p defined for any $\pi \in \mathcal{M}_{\leq 1}(\mathbb{R})$ by

$$G_p(\pi) := B_\Delta(\pi) + (1 - \|\pi\|)m_p$$

where m_p is the p-th absolute moment of the standard normal distribution. It is also seen that $m_p = r^\Theta(\Delta)$ is the minimax risk for the problem. For $p = 2$, it is seen that $1 - I(\pi * \gamma) = G_2(\pi)$, hence justifying our earlier notations. The fact that for any $p \in [1, +\infty)$, $B_\Delta = G_p$ on $\mathcal{M}_1(\mathbb{R})$ is immediate by construction. The vague upper semi-continuity of G_p on $\mathcal{M}_{\leq 1}(\mathbb{R})$ requires more work as shown below. We first note that G_p has a convenient "regret" formulation, namely, for any $\pi \in \mathcal{M}_{\leq 1}(\mathbb{R})$, $G_p(\pi) = \inf_{\delta \in \Delta} \int_\Theta (r(\delta, \theta) - m_p)\, d\pi(\theta) + m_p$.

Proposition 4.7. *Let $p \in [1, +\infty)$. The function G_p is vaguely upper semi-continuous on $\mathcal{M}_{\leq 1}(\mathbb{R})$.*

Proof. The result for $\pi = 0$ is direct by definition of m_p. It now suffices to prove that $\pi \mapsto \inf_{\delta\in\Delta} \int_\Theta (r(\delta,\theta) - m_p)\, d\pi(\theta)$ is vaguely upper semi-continuous on $\mathcal{M}_{\leq 1}(\mathbb{R}) \setminus \{0\}$. Consider a sequence $\{\pi_n\}$ in $\mathcal{M}_{\leq 1}(\mathbb{R})$ vaguely converging to some limit $\pi \neq 0$. Fix $\varepsilon > 0$. By definition of the infimum, there exists a rule $\delta_\varepsilon \in \mathcal{D}$ such that $B(\delta_\varepsilon, \pi) \leq B_\Delta(\pi) + \varepsilon$. For any $M > 0$, define $\delta_M(x) = \delta_\varepsilon(x)$ if $|x| \leq M$ and $\delta_M(x) = x$ otherwise. By construction, we have that $\lim_{|\theta|\to\infty} r(\delta_M, \theta) = m_p$ and so $h\colon \theta \mapsto r(\delta_M, \theta) - m_p$ belongs to $C_0(\mathbb{R})$. By definition of the infimum, we have $G_p(\pi_n) \leq \int_\Theta (r(\delta_M,\theta) - m_p)\, d\pi_n(\theta) + m_p$. Since $\pi_n \to \pi$ vaguely and $h \in C_0(\mathbb{R})$, we have that $\limsup_{n\to\infty} G_p(\pi_n) \leq \int_\Theta (r(\delta_M,\theta) - m_p)\, d\pi(\theta) + m_p$. Since $r(\delta_M, \theta) \to r(\delta_\varepsilon, \theta)$ pointwise as $M \to \infty$, we have by dominated convergence that $\limsup_{n\to\infty} G_p(\pi_n) \leq \int_\Theta (r(\delta_\varepsilon,\theta) - m_p)\, d\pi(\theta) + m_p$. By definition of δ_ε, we then obtain $\limsup_{n\to\infty} G_p(\pi_n) \leq \inf_{\delta\in\Delta} \int_\Theta (r(\delta,\theta) - m_p)\, d\pi(\theta) + m_p + \varepsilon$. Since $\varepsilon > 0$ is arbitrary, this proves that G_p is vaguely upper semi-continuous on $\mathcal{M}_{\leq 1}(\mathbb{R})$. □

Remark 4.9. Based on the definition of G_p, it is natural to introduce the function g_p, which is the penalized version of B, defined for any $\delta \in \mathcal{D}$ and any $\pi \in \mathcal{M}_{\leq 1}(\mathbb{R})$ by $g_p(\delta, \pi) := B(\delta, \pi) + (1 - \|\pi\|)m_p$. The penalization by $(1 - \|\pi\|)m_p$ is independent of δ and is affine in π, so that Proposition 2.7 extends directly to the functions g_p and G_p without modification of the proof. In this case, condition (4) bears on G_p and the minimax value and saddle point property hold for g_p. By cancellation of the penalty on the right side of the saddle point property for g_p, it is seen that any decision rule δ^* that forms a saddle point with π^* for g_p is necessarily a Bayes rule for π^*. Similar considerations as stated in Remark 2.5 obtain in this case.

4.2 Maximin priors with uniformly bounded moments under L^p loss

In this section, we consider a classical class of priors all constrained by a uniform moment bound. For any $m > 0$ and any $r > 0$, we define

$$\Gamma_m^r := \left\{ \pi \in \mathcal{M}_1(\mathbb{R}) : \int_{\mathbb{R}} |\theta|^r\, d\pi(\theta) \leq m \right\}.$$

This class satisfies a tightness condition from the uniform moment constraint that makes it weakly compact. This is a direct illustration of condition (**C2**) of Section 3.1 using the generalized Markov inequality. This is proved below in Lemma 4.8 where the convexity of the class is also obtained.

Lemma 4.8. *For any $m > 0$ and any $r > 0$, the set Γ_m^r is convex and weakly compact.*

Proof. The convex combination of two probability measures on $\mathbb{R}$ is a probability measure on $\mathbb{R}$. By linearity of the integral, the convex combination of two measures in Γ_m^r has uniformly bounded absolute r moments and so belongs to Γ_m^r. Consider now a sequence $\{\pi_n\}$ in Γ_m^r converging weakly to some limit π. Since $|\theta|^r$ is continuous and bounded from below, we have by the Portmanteau theorem that $\liminf_{n\to\infty} \int |\theta|^r\, d\pi_n(\theta) \geq \int |\theta|^r\, d\pi(\theta)$. By definition of Γ_m^r, it directly follows that $\int |\theta|^r\, d\pi(\theta) \leq m$ and so $\pi \in \Gamma_m^r$. We finally show that Γ_m^r is tight. From the definition of Γ_m^r and the generalized Markov inequality, we have that for any $\pi \in \Gamma_m^r$ and any $K > 0$, $\pi(\mathbb{R} \setminus [-K, K]) \leq m/K^r$.

By taking K large enough, this proves that Γ_m^r is tight. This is enough to conclude that Γ_m^r is weakly compact as a direct application of condition (**C2**). □

In combination with the general upper semi-continuity of B_Δ under L^p losses, $p \in [1, +\infty)$, obtained in Proposition 4.1, we readily recover well-known existence results for maximin priors over the class Γ_m^r of priors with uniformly bounded moments. Similar existence results were first stated in Feldman (1991) without proof[2] in the case $p = 2$. A proof for any $p \in [1, +\infty)$ was provided in Donoho and Johnstone (1994) (where p plays the role of q in their paper). For reference, we collect these existence results below. The proof is an immediate consequence of Corollary 3.1.

Corollary 4.9. *Let $p \in [1, +\infty)$. For any $m > 0$ and any $r > 0$, a proper maximin prior over Γ_m^r exists.*

Remark 4.10. Since the statistical experiment is regular (see Remark 4.1) and $\Delta = \mathcal{D}$, the conditions of Proposition 2.7 are verified and we naturally get more than the mere existence of proper maximin priors provided by Corollary 4.9. We have that the min-max equality is satisfied at any of these maximin priors. In particular, we are guaranteed that there exist Γ_m^r-minimax rules over $\mathcal{D}$ with which the proper maximin priors form saddle points. In the case $p = 2$, we know that the pair is unique from Proposition 4.3 and Lemma B.2. Additional properties of the maximin prior and Γ_m^r-minimax rule can be obtained in this case – see Feldman (1991).

4.3 Maximin sparse priors under L^p loss

We now consider the class of sparse priors (which is also a version of Huber's ε-contamination model) with $\Theta = \mathbb{R}$. Some motivation for this class can be found in Huber (1964), Mallows (1980), Bickel (1983), Bickel and Collins (1983), Donoho, Johnstone, Hoch, and Stern (1992), or Johnstone (1994) – see also Sections 2.4 and 8.7 in Johnstone (2019). For any $\varepsilon \in [0, 1]$, we define the set of sparse priors by

$$\Gamma(\varepsilon) := \left\{ \varepsilon \mathfrak{d}_0 + (1 - \varepsilon)\nu : \nu \in \mathcal{M}_1(\mathbb{R}) \right\}.$$

The two boundary cases correspond to $\Gamma(0) = \mathcal{M}_1(\mathbb{R})$ and $\Gamma(1) = \{\mathfrak{d}_0\}$. We also introduce the following notation

$$\overline{\Gamma}(\varepsilon) := \left\{\varepsilon \mathfrak{d}_0 + (1 - \varepsilon)\nu : \nu \in \mathcal{M}_{\leq 1}(\mathbb{R})\right\},$$

with the two boundary cases $\overline{\Gamma}(0) = \mathcal{M}_{\leq 1}(\mathbb{R})$ and $\overline{\Gamma}(1) = \{\mathfrak{d}_0\}$. We start by considering the problem when $p = 2$ since the L^2 loss ℓ_2 leads to a known existence result leveraging the Fisher information representation of B_Δ given by the Brown–Tweedie identity. We then extend this existence result to any $p \in [1, +\infty)$ by leveraging the vague upper semi-continuous extension G_p of B_Δ introduced at the end of Section 4.1. These existence results are new and should be of independent interest. They all directly illustrate the use of condition (**V**) of Section 3.2.

[2]The author simply claimed that "[b]y standard methods [they] can prove that there exists a solution". It is true, indeed, that the existence result in the case $p = 2$ is much simpler by virtue of the Brown–Tweedie identity. The weak semi-continuity of the Fisher functional is, however, not as trivial as one may expect – see Lemma 4.2.

Lemma 4.10. *For any $\varepsilon \in [0,1]$, the set of sparse priors $\Gamma(\varepsilon)$ is convex and weakly closed. For any $\varepsilon \in [0,1)$, the set $\Gamma(\varepsilon)$ is not weakly compact and $\overline{\Gamma}(\varepsilon)$ is its vague closure.*

Proof. The cases $\varepsilon = 0, 1$ are immediate. Consider $\varepsilon \in (0,1)$. The set $\Gamma(\varepsilon)$ is convex as the sum of two convex sets. It is also weakly closed as the convex combination of a weakly compact set, $\{\mathfrak{d}_0\}$, and a weakly closed set, $\mathcal{M}_1(\mathbb{R})$. Consider now the sequence $\pi_n = \varepsilon\mathfrak{d}_0 + (1-\varepsilon)N(0,n)$ in $\Gamma(\varepsilon)$. For any compact subset $K \subset \mathbb{R}$, $\lim_{n\to\infty} \pi_n(K) = \varepsilon\mathfrak{d}_0(K) \le \varepsilon$, and so, for any $\eta \le 1-\varepsilon$, we have $\inf_{\pi\in\Gamma(\varepsilon)} \pi(K) \le 1-\eta$. The set $\Gamma(\varepsilon)$ is not tight and hence not weakly relatively compact by Prokhorov's theorem. We now show that $\overline{\Gamma}(\varepsilon) \subseteq \mathrm{cl}(\Gamma(\varepsilon))$ where $\mathrm{cl}(\Gamma(\varepsilon))$ denotes the vague closure of $\Gamma(\varepsilon)$. Consider a measure $\pi \in \overline{\Gamma}(\varepsilon)$ where $\|\pi\| = \alpha \le 1$. Then $\pi_n = \pi + (1-\alpha)N(0,n)$ satisfies $\pi_n \in \Gamma(\varepsilon)$ and $\pi_n \to \pi$ vaguely. The reverse inclusion $\mathrm{cl}(\Gamma(\varepsilon)) \subseteq \overline{\Gamma}(\varepsilon)$ is immediate by the linearity of limits, Lemma 3.3, and Lemma 30.3 in Bauer (2011). □

Lemma 4.11. *For any $\varepsilon \in (0,1)$, the function $I(\cdot * \gamma)$ attains its infimum on $\overline{\Gamma}(\varepsilon)$ at a unique probability measure $\pi^* \in \Gamma(\varepsilon)$.*

Proof. The proof follows directly from Theorem 3 in Bickel and Collins (1983) by virtue of the equivalence stated in Remark 3.8 (where uniqueness follows from the uniqueness of characteristic functions). For completeness, we provide a proof[3] directly using the vague topology on $\overline{\Gamma}(\varepsilon)$. From Lemma 4.10 and Lemma 4.5, $I(\cdot * \gamma)$ is vaguely lower semi-continuous on the vaguely compact set $\overline{\Gamma}(\varepsilon)$. It attains its infimum at some measure $\pi^* \in \overline{\Gamma}(\varepsilon)$. Equivalently, $\pi^* * \gamma$ is a minimizer for $I(\cdot)$ over the set $\mathcal{Q} = \{Q = \pi * \gamma : \pi \in \overline{\Gamma}(\varepsilon)\} \subseteq \mathcal{M}_{\le 1}(\mathbb{R})$. Since $\mathcal{Q}$ is convex and the density of $\pi^* * \gamma$ is strictly positive everywhere on $\mathbb{R}$, we have by Proposition 4.5 in Huber and Ronchetti (2009) that the minimizer $Q^* = \pi^* * \gamma$ is unique. By uniqueness of characteristic functions and strict positivity of the Laplace transform for γ, we get that π^* is the unique minimizer of $I(\cdot * \gamma)$. Suppose now by contradiction that $1 - \|\pi^*\| = \alpha > 0$. Consider the new measure $\bar{Q} = Q^* + \alpha(\nu * \gamma)$ where $\nu \in \mathcal{M}_1(\mathbb{R})$. We have that $\bar{Q} \in \mathcal{Q}$ and $I(\bar{Q}) < \infty$. By the variational representation of Section 4.5 in Huber and Ronchetti (2009) with $\nu = \mathfrak{d}_\theta$ for any $\theta \in \mathbb{R}$, we obtain $\int_{\mathbb{R}} (-2\psi'(x) + \psi^2(x))\phi(x-\theta)\,dx \le 0$ where $\psi(x) = -q'(x)/q(x)$ and q is the density of Q^*. Then by applying Stein's unbiased risk formula to $\delta'(x) = x - \psi(x)$, we obtain $r(\delta', \theta) \le 1$ for all $\theta \in \mathbb{R}$. Since $\delta^*(x) = x$ is the unique minimax estimator with constant risk equal to 1, we have a contradiction. It follows that $\|\pi^*\| = 1$ and so $\pi^* \in \Gamma(\varepsilon)$. By Lemma 4.2, it follows that π^* is also a maximizer of B_Δ on $\Gamma(\varepsilon)$. □

The result of Lemma 4.11 in combination with Lemma 4.2 ensures that condition (**V**) of Section 3.2 is satisfied. This directly guarantees the existence of a proper maximin prior over the class $\Gamma(\varepsilon)$, $\varepsilon \in (0,1)$, under L^2 loss. While existence is directly guaranteed by construction, we can still show that the conditions of Proposition 2.6 are verified as in Section 3.2.

Corollary 4.12. *For any $\varepsilon \in (0,1)$, there exists a maximizing sequence for the Bayes risk B_Δ over $\Gamma(\varepsilon)$ that admits a subsequence weakly converging to some limit $\pi^* \in \Gamma(\varepsilon)$.*

[3] This adapts a corrected proof of Proposition 8.19 in Johnstone (2019) that was privately communicated by I.A. Johnstone. The contradiction to prove that a minimizer is necessarily with mass 1 is identical to the argument in Bickel and Collins (1983). We use a similar but different argument for the generalization in Proposition 4.13.

Proof. Let $\{\pi_n\}$ be a maximizing sequence for B_Δ over $\Gamma(\varepsilon)$. The set $\overline{\Gamma}(\varepsilon) := \{\varepsilon\mathfrak{d}_0 + (1-\varepsilon)\nu : \nu \in \mathcal{M}_{\leq 1}(\mathbb{R})\}$ is the vague closure of $\Gamma(\varepsilon)$. By Lemma 3.2, $\{\pi_n\}$ has a vaguely converging subsequence in $\Gamma(\varepsilon)$ with limit π^* in $\overline{\Gamma}(\varepsilon)$. By Lemma 4.2, this convergent subsequence of $\{\pi_n\}$ is also a minimizing sequence for $I(\cdot * \gamma)$. By Lemma 4.11, it necessarily holds that $\pi^* \in \Gamma(\varepsilon)$. The subsequence of $\{\pi_n\}$ (vaguely and weakly) converging to $\pi^* \in \Gamma(\varepsilon)$ satisfies the required condition. □

To generalize these results beyond the L^2 loss, we resort to the vaguely upper semi-continuous extension G_p, $p \in [1, +\infty)$, we obtained at the end of Section 4.1. From the vague upper semi-continuity of G_p, we directly obtain the existence of a maximizer on the vague closure $\overline{\Gamma}(\varepsilon)$. We now show that any maximizer is necessarily a probability measure and so belongs to $\Gamma(\varepsilon)$. The proof is based on the uniqueness of the minimax estimator in the normal mean problem under L^p losses. The uniqueness of the maximin prior (which we obtained in the case $p = 2$) is not guaranteed without the strict concavity of B_Δ – see Proposition 4.3 and Remark 4.6 for more on this point.

Proposition 4.13. *Let $p \in [1, +\infty)$. For any $\varepsilon \in (0, 1)$, the function G_p attains its supremum on $\overline{\Gamma}(\varepsilon)$ at a probability measure $\pi^* \in \Gamma(\varepsilon)$.*

Proof. From Proposition 4.7, we know that the function G_p is vaguely upper semi-continuous on $\overline{\Gamma}(\varepsilon)$. Since $\overline{\Gamma}(\varepsilon)$ is vaguely compact, the function G_p attains its supremum at some measure $\pi^* \in \overline{\Gamma}(\varepsilon)$. From Remark 4.9, we know that there is a decision rule $\delta^* \in \mathcal{D}$ such that (δ^*, π^*) forms a saddle point for the penalized average risk $g_p(\delta, \pi) = B(\delta, \pi) + (1 - \|\pi\|)m_p$. From Remark 4.9, we also know that δ^* is a Bayes rule for π^*. By definition of $\overline{\Gamma}(\varepsilon)$, there is $\nu^* \in \mathcal{M}_{\leq 1}(\mathbb{R})$ such that $\pi^* = \varepsilon\mathfrak{d}_0 + (1-\varepsilon)\nu^*$. In particular, $g_p(\delta^*, \pi^*) = \varepsilon r(\delta^*, 0) + (1-\varepsilon)\left[\int_{\mathbb{R}} r(\delta^*, \theta)\, d\nu^*(\theta) + (1 - \|\nu^*\|)m_p\right] = \varepsilon r(\delta^*, 0) + (1 - \varepsilon)g_p(\delta^*, \nu^*)$. Since π^* maximizes $g_p(\delta^*, \cdot)$ over $\overline{\Gamma}(\varepsilon)$ by the saddle point property, it necessarily holds that ν^* maximizes $g_p(\delta^*, \cdot)$ over $\mathcal{M}_{\leq 1}(\mathbb{R})$. Suppose by contradiction that $\|\nu^*\| < 1$ and that there is $\theta_0 \in \mathbb{R}$ such that $r(\delta^*, \theta_0) > m_p$. If we define $\nu_0 = \nu^* + (1 - \|\nu^*\|)\mathfrak{d}_{\theta_0} \in \mathcal{M}_{\leq 1}(\mathbb{R})$, we then obtain $g_p(\delta^*, \nu_0) = \int_{\mathbb{R}} r(\delta^*, \theta)\, d\nu^*(\theta) + (1 - \|\nu^*\|)r(\delta^*, \theta_0) > g_p(\delta^*, \nu^*)$: a contradiction. Suppose now that $\|\nu^*\| < 1$ and that $r(\delta^*, \theta) \leq m_p$ for all $\theta \in \mathbb{R}$. Since the rule $\delta_m(x) = x$ is unique minimax for the problem and has constant risk equal to m_p, the rules δ_m and δ^* must be equal almost everywhere. Since δ^* cannot be the Bayes rule for any finite measure, we have again a contradiction. It follows that $\|\pi^*\| = 1$ and so $\pi^* \in \Gamma(\varepsilon)$. □

Proposition 2.7 ensures the existence of a proper maximin prior over $\Gamma(\varepsilon)$ for any $\varepsilon \in (0, 1)$ and any $p \in [1, +\infty)$. This result is new and extends the case $p = 2$ to any $p \in [1, +\infty)$ where the closed-form Brown–Tweedie identity is no longer valid. It is a direct illustration of condition (**V**) in Section 3.2 using the extension G_p. The result is collected in Corollary 4.14 for reference.

Corollary 4.14. *Let $p \in [1, +\infty)$. For any $\varepsilon \in (0, 1)$, a proper maximin prior over $\Gamma(\varepsilon)$ exists.*

Proof. The proof is immediate from Proposition 4.13 and the identity $B_\Delta = G_p$ on $\mathcal{M}_1(\mathbb{R})$. □

Since the statistical experiment is regular and $\Delta = \mathcal{D}$, the conditions of Proposition 2.7 are verified and we directly obtain that the maximin priors of Corollary 4.14 form saddle points with some $\Gamma(\varepsilon)$-minimax rules. In the case $p = 2$, the maximin prior is also known to be unique, symmetric, and to have countably infinite support – see Bickel and Collins (1983) and Johnstone (2019). Uniqueness and symmetry likely hold under arbitrary $p \in [1, +\infty)$ but call for convoluted analytical arguments that are beyond the objective of this paper – see Remark 4.6. As shown in Proposition 4.15 below, some properties of the support can still be proved under arbitrary $p \in [1, +\infty)$ by direct arguments. The result should be of independent interest for the variational characterization of the saddle point property that extends the one available in the case $p = 2$. Other properties of the support cannot be salvaged without stronger analytical arguments as made clear in Remark 4.11 below.

Proposition 4.15. *Let $p \in [1, +\infty)$ and $\varepsilon \in (0, 1)$. If π^* is a maximin prior over $\Gamma(\varepsilon)$, then there is a Bayes rule δ_{π^*} for π^* such that*

$$\int_{\mathbb{R}} r(\delta_{\pi^*}, \theta)\, d(\pi_0 - \pi^*)(\theta) \leq 0$$

for any $\pi_0 \in \Gamma(\varepsilon)$. In particular, any maximin prior over $\Gamma(\varepsilon)$ has infinite support. If p is an even integer, then any maximin prior over $\Gamma(\varepsilon)$ has countably infinite support.

Proof. Let $p \in [1, +\infty)$ and $\varepsilon \in (0, 1)$. From Corollary 4.14, a maximin prior π^* over $\Gamma(\varepsilon)$ is guaranteed to exist. Since $\|\pi^*\| = 1$, it necessarily has the form $\pi^* = \varepsilon\mathfrak{d}_0 + (1 - \varepsilon)\nu^*$ for some $\nu^* \in \mathcal{M}_1(\mathbb{R})$. Since the conditions of Proposition 2.7 are verified, there is a decision rule δ_{π^*}, which is also a Bayes rule for π^*, such that (δ_{π^*}, π^*) is a saddle point for B. This directly implies that $\int_{\mathbb{R}} r(\delta_{\pi^*}, t)\, d(\pi_0 - \pi^*)(t) \leq 0$ for any $\pi_0 \in \Gamma(\varepsilon)$. If we apply this inequality to $\pi_\theta = \varepsilon\mathfrak{d}_0 + (1 - \varepsilon)\mathfrak{d}_\theta$ for any $\theta \in \mathbb{R}$, we get that $r(\delta_{\pi^*}, \theta) \leq \int_{\mathbb{R}} r(\delta_{\pi^*}, t) d\nu^*(t)$ for any $\theta \in \mathbb{R}$. It follows that ν^* is necessarily supported on the set of global maxima of $\theta \mapsto r(\delta_{\pi^*}, \theta)$. Suppose now that the support S of ν^* is finite. Then the posterior distribution is supported on $S \cup \{0\}$ and the Bayes rule δ_{π^*} is uniformly bounded by $\max_{\theta \in S \cup \{0\}} |\theta|$. By direct computations, it follows that $r(\delta_{\pi^*}, \theta) \to \infty$ as $\theta \to \infty$. This contradicts the fact that $r(\delta_{\pi^*}, \theta) \leq m_p/(1 - \varepsilon)$ for any $\theta \in \mathbb{R}$ as obtained in the proof of Proposition 4.13. It follows that the support S of ν^* is infinite. Suppose now that p is an even integer. Then, by binomial expansion, we have $r(\delta_{\pi^*}, \theta) = \sum_{i=0}^{p} \binom{p}{i} (-\theta)^{p-i} \int_{\mathbb{R}} (\delta_{\pi^*}(x))^i \phi(x - \theta)\, dx$ for any $\theta \in \mathbb{R}$. By the uniform bound on $r(\delta_{\pi^*}, \theta)$ and Lemma 2.8 in Brown (1986), it follows that $\theta \mapsto r(\delta_{\pi^*}, \theta)$ is analytic. Suppose now that S has an accumulation point. Then, by the identity theorem for analytic functions, we have that $r(\delta_{\pi^*}, \theta)$ is constant for all $\theta \in \mathbb{R}$. This implies that δ_{π^*} is Θ-minimax: a contradiction by uniqueness of the minimax rule in the normal mean problem and $\|\pi^*\| = 1$. It follows that the supports of ν^* and π^* are countably infinite. □

Remark 4.11. The analyticity of the risk function $\theta \mapsto r(\delta_{\pi^*}, \theta)$ in the case where p is not an even integer cannot directly rely on Lemma 2.8 in Brown (1986) by algebraic expansion. It requires direct analytical arguments to extend Lemma 2.8 in Brown (1986) and is beyond the scope of this paper.

References

Aubin, J.-P. (1979): *Mathematical methods of game and economic theory*. North-Holland.

Bauer, H. (2011): *Measure and integration theory*, vol. 26. Walter de Gruyter.

Bickel, P. J. (1981): "Minimax estimation of the mean of a normal distribution when the parameter space is restricted," *The Annals of Statistics*, 9(6), 1301–1309.

——— (1983): "Minimax estimation of the mean of a normal distribution subject to doing well at a point," in *Recent Advances in Statistics*, pp. 511–528. Elsevier.

Bickel, P. J., and J. R. Collins (1983): "Minimizing Fisher information over mixtures of distributions," *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 1–19.

Billingsley, P. (1971): *Weak convergence of measures: Applications in probability*, vol. 5. SIAM.

Blackwell, D., and M. Girshick (1954): *Theory of games and statistical decisions.* Wiley.

Bogachev, V. I. (2007): *Measure theory*. Springer.

Brown, L. D. (1966): "On the admissibility of invariant estimators of one or more location parameters," *The Annals of Mathematical Statistics*, 37(5), 1087–1136.

——— (1971): "Admissible estimators, recurrent diffusions, and insoluble boundary value problems," *The Annals of Mathematical Statistics*, 42(3), 855–903.

——— (1974): "Notes on decision theory," *Unpublished lecture notes*.

——— (1986): *Fundamentals of statistical exponential families with applications in statistical decision theory*, vol. 9. Institute of Mathematical Statistics.

Casella, G., and W. E. Strawderman (1981): "Estimating a bounded normal mean," *The Annals of Statistics*, 9(4), 870–878.

Dacorogna, B. (2008): *Direct methods in the calculus of variations*. Springer.

DeGroot, M., and M. Rao (1963): "Bayes estimation with convex loss," *The Annals of Mathematical Statistics*, pp. 839–846.

Donoho, D. L., and I. M. Johnstone (1994): "Minimax risk over l_p-balls for l_q-error," *Probability Theory and Related Fields*, 99, 277–303.

Donoho, D. L., I. M. Johnstone, J. C. Hoch, and A. S. Stern (1992): "Maximum entropy and the nearly black object," *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(1), 41–67.

Farrell, R. (1964): "Estimators of a location parameter in the absolutely continuous case," *The Annals of Mathematical Statistics*, pp. 949–998.

Feldman, I. (1991): "Constrained minimax estimation of the mean of the normal distribution with known variance," *The Annals of Statistics*, 19(4), 2259–2265.

Folland, G. B. (1999): *Real analysis: modern techniques and their applications*, vol. 40. John Wiley & Sons.

Ghosh, M. (1964): "Uniform approximation of minimax point estimates," *The Annals of Mathematical Statistics*, pp. 1031–1047.

Hiriart-Urruty, J.-B. (2012): *Bases, outils et principes pour l'analyse variationnelle*, vol. 70. Springer.

Huber, P. J. (1964): "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, pp. 73–101.

Huber, P. J., and E. Ronchetti (2009): *Robust statistics*, Wiley series in probability and statistics. Wiley.

Johnstone, I. M. (1994): "On minimax estimation of a sparse normal mean vector," *The Annals of Statistics*, pp. 271–289.

——— (2019): "Gaussian estimation: Sequence and wavelet models," *Unpublished lecture notes*, GE_09_16_19.

Kneser, H. (1952): "Sur un théoreme fondamental de la théorie des jeux," *CR Acad. Sci. Paris*, 234, 2418–2420.

Le Cam, L. (1955): "An extension of Wald's theory of statistical decision functions," *The Annals of Mathematical Statistics*, 26(1), 69–81.

——— (1986): *Asymptotic methods in statistical decision theory*. Springer.

Mallows, C. L. (1980): "Some theory of nonlinear smoothers," *The Annals of statistics*, pp. 695–715.

Moreau, J.-J. (1967): "Fonctionnelles convexes," *Séminaire Jean Leray (Collège de France, Paris)*, 2, 1–108.

Nussbaum, M. (1996): "Asymptotic equivalence of density estimation and Gaussian white noise," *The Annals of Statistics*, pp. 2399–2430.

Strasser, H. (1985): *Mathematical theory of statistics: statistical experiments and asymptotic decision theory*, vol. 7. Walter de Gruyter.

Wald, A. (1950): *Statistical Decision Functions*. Wiley.

Appendices

Appendix A Definitions, notations, and conventions

Let $(E, \mathcal{E}, \mu)$ be a measure space. For $1 \le p \le \infty$, we denote by $L^p(E, \mu)$ the usual Lebesgue space of measurable functions $f\colon E \to \mathbb{R}$ such that $\|f\|_p < \infty$ where $\|f\|_p^p = \int_E |f|^p \, d\mu$ if $1 \le p < \infty$ and $\|f\|_\infty = \inf\{C \ge 0 : |f(x)| \le C \text{ for a.e. } x\}$.

Let E be a Polish metric space endowed with its Borel σ-algebra $\mathcal{B}_E$. We denote by $C(E)$ the set of continuous functions $f\colon E \to \mathbb{R}$, by $C_b(E)$ its restriction to bounded functions, by $C_0(E)$ its restriction to functions vanishing at infinity, and by $C_c(E)$ its restriction to compactly supported functions. A Dirac measure at a point $x \in E$ is denoted $\mathfrak{d}_x$. We call Borel measure any measure that is defined on $(E, \mathcal{E})$. We call Radon measure any σ-finite measure on $(E, \mathcal{B}_E)$ that is inner regular and locally finite. We denote by $\mathcal{M}(E)$ the set of all Radon measures on E and $\mathcal{M}_+(E)$ the restriction of this set to positive measures. We denote by $\mathcal{M}_f(E)$ the set of all Borel positive finite measures on E. We then denote by $\mathcal{M}_1(E) = \{\mu \in \mathcal{M}_f(E) : \mu(E) = 1\}$ the set of Borel probability measures on E and by $\mathcal{M}_{\le 1}(E) = \{\mu \in \mathcal{M}(E) : \mu(E) \le 1\}$ the set of Borel subprobability measures on E. Naturally, $\mathcal{M}_1(E) \subseteq \mathcal{M}_{\le 1}(E) \subseteq \mathcal{M}_f(E)$. Moreover, since E is assumed Polish, we have that $\mathcal{M}_f(E) \subseteq \mathcal{M}_+(E)$ – see Theorem 26.3 in Bauer (2011). For $\mu \in \mathcal{M}_f(E)$, we denote by $\|\mu\| = \mu(E)$ the (restriction of the) total variation norm (to the cone) of positive finite measures $\mathcal{M}_f(E)$. A sequence of measures $\{\mu_n\}$ in $\mathcal{M}_f(E)$ is said to converge weakly to μ if $\int_E f \, d\mu_n \to \int_E f \, d\mu$ for all bounded continuous functions $f \in C_b(E)$. This derives from a completely metrizable topology on $\mathcal{M}_f(E)$ called the weak topology (or narrow topology) – see Theorem 8.9.4 in Bogachev (2007). When E is also locally compact, a sequence of measures $\{\nu_n\}$ in $\mathcal{M}_+(E)$ is said to converge vaguely to ν if $\int_E f \, d\nu_n \to \int_E f \, d\nu$ for all continuous functions with compact support $f \in C_c(E)$. This derives from a completely metrizable topology on $\mathcal{M}_+(E)$ called the vague topology – see Theorem 31.5 in Bauer (2011).

If E is a non-compact locally compact metric space, then it can always be made into a compact space by adding a single point $\{\infty_E\}$ – see Proposition 4.36 in Folland (1999) – the resulting space is called the one-point compactification of E and is denoted $E \cup \{\infty_E\}$. If $E = \mathbb{R}$, this leads to the one-point compactification $\mathbb{R} \cup \{\infty\}$. This space is topologically different from the two-point compactification $[-\infty, +\infty]$ of $\mathbb{R}$ obtained by homeomorphism from any non-empty closed interval. Both spaces will be considered in the main text and both are endowed with their respective Borel σ-algebras. We follow standard conventions for arithmetic operations involving $\{-\infty, +\infty, \infty\}$.

The Lebesgue measure on $\mathbb{R}$ is denoted by λ. The normal distribution with mean $m \in \mathbb{R}$ and variance $\sigma^2 > 0$ is denoted by $N(m, \sigma^2)$ and its measure by γ_{m,σ^2}. For the standard normal distribution $N(0, 1)$, its measure is directly denoted by γ, its density with respect to the Lebesgue measure by ϕ, and its cumulative distribution function by Φ. To avoid conflicting notations, we denote the Gamma function through its Gauss notation: for any $z > 0$, $\Pi(z) = \Gamma(z + 1)$. For any $p > 0$, the p-th absolute moment of $N(0, 1)$ is denoted by m_p and it is known that $m_p = \pi^{-1/2} 2^{p/2} \Gamma((p + 1)/2) = \pi^{-1/2} 2^{p/2} \Pi((p - 1)/2)$. If $p = 2$, $m_p = 1$. We denote by $*$ the convolution operator for finite Borel measures on $\mathbb{R}$. The Fisher information of a finite Borel measure π on $\mathbb{R}$ that admits an absolutely continuous density p with respect to the Lebesgue measure is defined by $I(\pi) := \int_{\mathbb{R}} ((p'(x))^2/p(x))\, dx$ whenever the integral is finite. For extensions and equivalent characterizations of the function I, see Section 4.4 and 4.5 in Huber and Ronchetti (2009). For any measure $\pi \in \mathcal{M}_f(\mathbb{R})$, the measure $\pi * \gamma \in \mathcal{M}_f(\mathbb{R})$ admits an absolutely continuous density with respect to the Lebesgue measure given by $p(x) = \int_{\mathbb{R}} \phi(x - y)\, d\pi(y)$ for all $x \in \mathbb{R}$.

A statistical experiment is a pair $\big((\mathcal{X}, \mathcal{B}_{\mathcal{X}}), \{P_\theta : \theta \in \Theta\}\big)$ where $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ is a measurable space and $\{P_\theta : \theta \in \Theta\}$ is a set of probability measures on $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$. The index set Θ for the set of probability measures is called the parameter space and is assumed to be a Polish metric space. We endow it with its Borel σ-algebra that we denote $\mathcal{B}_\Theta$. A statistical decision problem is a quadruple $\big((\mathcal{X}, \mathcal{B}_{\mathcal{X}}), \{P_\theta : \theta \in \Theta\}, (\mathcal{A}, \mathcal{B}_{\mathcal{A}}), \ell(\cdot, \cdot)\big)$ where $\big((\mathcal{X}, \mathcal{B}_{\mathcal{X}}), \{P_\theta : \theta \in \Theta\}\big)$ is a statistical experiment, $(\mathcal{A}, \mathcal{B}_{\mathcal{A}})$ is a measurable space called action space, and $\ell \colon \mathcal{A} \times \Theta \to [0, +\infty]$ is a function called loss that is measurable with respect to the product $\mathcal{B}_C A \otimes \mathcal{B}_\Theta$. We assume that $\mathcal{A}$ is a metric space and $\mathcal{B}_{\mathcal{A}}$ is the Borel σ-algebra for $\mathcal{A}$. When the measurable spaces are clear, we simply write a statistical decision problem as $\big(\mathcal{X}, \{P_\theta : \theta \in \Theta\}, \mathcal{A}, \ell(\cdot, \cdot)\big)$.

A decision rule for the statistical experiment is a Markov kernel $\delta(\cdot, \cdot)$ from $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ to $(\mathcal{A}, \mathcal{B}_{\mathcal{A}})$. The set of all decision rules for a given decision problem is denoted $\mathcal{D}$. If for every $x \in \mathcal{X}$, the decision rule δ is the Dirac measure at $T(x)$ for some measurable function $T \colon \mathcal{X} \to \mathcal{A}$, then the decision rule is said to be non-randomized or deterministic. Given a statistical decision problem, we define the risk function as the function $r \colon \mathcal{D} \times \Theta \to [0, +\infty]$ given by

$$r(\delta, \theta) := \int_{\mathcal{X}} \int_{\mathcal{A}} \ell(a, \theta)\, d\delta(x, a)\, dP_\theta(x).$$

If the decision rule is a non-randomized rule δ_T inducing a measurable function T, then we denote $r(T, \theta) := r(\delta_T, \theta)$.

A positive Radon measure $\pi \in \mathcal{M}_+(\Theta)$ on Θ is called a prior. If $\pi \in \mathcal{M}_1(\Theta)$, then π is called a proper prior, and an improper prior (or generalized prior) otherwise. We define the average risk (or integrated risk) as the function $B\colon \mathcal{D} \times \mathcal{M}_+(\Theta) \to [0, +\infty]$ given by

$$B(\delta, \pi) := \int_\Theta r(\delta, \theta)\, d\pi(\theta).$$

Given a subset $\Delta \subseteq \mathcal{D}$ of decision rules, we define the Bayes risk (or Bayesian risk) for Δ as the function $B_\Delta\colon \mathcal{M}_+(\Theta) \to [0, +\infty]$ given by

$$B_\Delta(\pi) := \inf_{\delta \in \Delta} B(\delta, \pi) = \inf_{\delta \in \Delta} \int_\Theta r(\delta, \theta)\, d\pi(\theta).$$

A Bayes rule (or Bayesian rule) over Δ with respect to a prior $\pi \in \mathcal{M}_+(\Theta)$ is any rule $\delta_\pi \in \Delta$ that minimizes the average risk at π, that is, $B(\delta_\pi, \pi) = \inf_{\delta \in \Delta} B(\delta, \pi) = B_\Delta(\pi)$. Given a set $\Gamma \subseteq \mathcal{M}_+(\Theta)$ of priors, a Γ-minimax rule over $\Delta \subseteq \mathcal{D}$ is any rule $\delta^* \in \Delta$ that minimizes the worst-case average risk among all rules in Δ, that is, $\sup_{\pi \in \Gamma} B(\delta^*, \pi) = \inf_{\delta \in \Delta} \sup_{\pi \in \Gamma} B(\delta, \pi)$. In this case, $B^\Gamma(\Delta) := \sup_{\pi \in \Gamma} B(\delta^*, \pi)$ is called the Γ-minimax average risk over Δ. Given a set of decision rules $\Delta \subseteq \mathcal{D}$, a Δ-maximin prior (or Δ-least favorable prior) over Γ is a prior $\pi^* \in \Gamma$ that maximizes the Bayes risk for Δ among all priors in Γ, that is, $B_\Delta(\pi^*) = \sup_{\pi \in \Gamma} B_\Delta(\pi) = \sup_{\pi \in \Gamma} \inf_{\delta \in \Delta} B(\delta, \pi) = \inf_{\delta \in \Delta} B(\delta, \pi^*)$. In this case, $B_\Delta(\Gamma) = B_\Delta(\pi^*)$ is called the Δ-maximin average risk over Γ. If $\pi^* \in \mathcal{M}_1(\Theta)$, then π^* is called a proper Δ-maximin prior. If the sets Γ and Δ are clear, all the prefix notations can be dropped.

Remark A.1. The notion of Γ-minimax rule is different from the frequentist notion of Θ-minimax rule. A Θ-minimax rule (or minimax rule for the risk) over $\Delta \subseteq \mathcal{D}$ is any rule $\delta_m \in \Delta$ that minimizes the worst-case risk, that is, $\sup_{\theta \in \Theta} r(\delta_m, \theta) = \inf_{\delta \in \Delta} \sup_{\theta \in \Theta} r(\delta, \theta)$. In this case, $r^\Theta(\Delta) := \inf_{\delta \in \Delta} \sup_{\theta \in \Theta} r(\delta, \theta)$ is called the Θ-minimax risk over Δ. The notion of minimax rule (for the risk) is a frequentist notion that does not involve priors. It is, however, intimately related to the Bayesian notion of Γ-minimax rule. Indeed, it always holds that the Γ-minimax average risk over Δ is a lower bound for the Θ-minimax risk over the same set of decision rules, that is, $B^\Gamma(\Delta) \le r^\Theta(\Delta)$. If the set Γ is rich enough (e.g., it contains all Dirac measures on Θ), then a Θ-minimax rule is a Γ-minimax rule and the inequality above is an equality.

Appendix B Bayes rules: strict convexity and uniqueness a.e.

The existence of Bayes rules is guaranteed by the topological constructions on $\mathcal{D}$ and the semi-continuity of the loss as shown in Section 2.1. Other useful properties of Bayes rules are not necessarily guaranteed under these conditions. The following results provide sufficient conditions for the uniqueness and non-randomization of the Bayes rules for some of the problems considered in the main text. As far as we know, these results are not directly available in the literature. For comparable results, see, e.g., DeGroot and Rao (1963), Farrell (1964), or Brown (1966).

Lemma B.1. *Let $\pi \in \mathcal{M}_f(\Theta) \setminus \{0\}$. Let $\big(\mathcal{X}, \{P_\theta : \theta \in \Theta\}, \mathcal{A}, \ell(\cdot,\cdot)\big)$ be a regular statistical decision problem where P_0 is the dominating measure for $\{P_\theta : \theta \in \Theta\}$. Suppose that $\mathcal{A}$ is a compact and convex subset of a Fréchet space. Suppose that $\Delta = \mathcal{D}$ is endowed with the regular topology. Suppose that the loss function $\ell(\cdot, \theta)$ is strictly convex and lower semi-continuous on $\mathcal{A}$ for all $\theta \in \Theta$. Suppose that $B_\Delta(\pi) < \infty$ and P_0 is dominated by the marginal distribution P_π defined by $P_\pi(B) = \int_\Theta P_\theta(B)\, d\pi(\theta)$ for all Borel sets $B \subseteq \mathcal{X}$. Then the Bayes rule δ_π for π exists, is unique, and non-randomized in the sense that $\delta_\pi(x, \cdot)$ is a unique Dirac measure for P_0-almost all $x \in \mathcal{X}$.*

Proof. By Lemma 2.2, a Bayes rule δ for π is guaranteed to exist. Define for any $x \in \mathcal{X}$, $T(x) := \int_\mathcal{A} a\, d\delta(x, a)$. Since $\mathcal{A}$ is closed and convex, $T(x) \in \mathcal{A}$ for all $x \in \mathcal{X}$. Since $\mathcal{A}$ is separable, the function $T\colon \mathcal{X} \to \mathcal{A}$ is measurable. We denote by δ_T the kernel equal to the Dirac measure at $T(x)$ for any fixed $x \in \mathcal{X}$. By Jensen's inequality and the strict convexity of the loss, we have $\ell(T(x), \theta) \le \int_\mathcal{A} \ell(a, \theta)\, d\delta(x, a)$ for all $\theta \in \Theta$ and all $x \in \mathcal{X}$. Equality holds for any fixed $\theta \in \Theta$ and any fixed $x \in \mathcal{X}$ only if $\delta(x, \cdot) = \delta_T(x, \cdot)$. By integrating over P_θ and π, we obtain $B(\delta_T, \pi) \le B(\delta, \pi) = B_\Delta(\pi) < \infty$. Since $\delta_T \in \mathcal{D}$, we necessarily have $B(\delta_T, \pi) = B(\delta, \pi)$ by definition of the infimum. From the regularity of $\{P_\theta : \theta \in \Theta\}$, the joint measure is well-defined and admits a density with respect to the product measure $P_0 \otimes \pi$ which we denote p. The equality $B(\delta_T, \pi) = B(\delta, \pi)$ then rewrites as $\int_\mathcal{X} \int_\Theta \ell(T(x), \theta) p(x, \theta)\, d\pi(\theta)\, dP_0(x) = \int_\mathcal{X} \int_\Theta \int_\mathcal{A} \ell(a, \theta)\, d\delta(x, a) p(x, \theta)\, d\pi(\theta)\, dP_0(x)$. Since the integrands are positive, we have that $\int_\Theta \ell(T(x), \theta) p(x, \theta)\, d\pi(\theta) = \int_\Theta \int_\mathcal{A} \ell(a, \theta)\, d\delta(x, a) p(x, \theta)\, d\pi(\theta)$ for P_0-a.e. $x \in \mathcal{X}$. Denote by $\mathcal{N}_0$ the null set on which the equality is false. Denote by $\mathcal{N} \subseteq \mathcal{X}$ the set for which $\int_\Theta p(x, \theta)\, d\pi(\theta) = 0$. Suppose that $x \notin \mathcal{N}_0 \cup \mathcal{N}$. Then $\ell(T(x), \theta) = \int_\mathcal{A} \ell(a, \theta)\, d\delta(x, a)$ for $\pi(\cdot|x)$-a.e. $\theta \in \Theta$ where $\pi(\cdot|x)$ is the posterior measure. The posterior measure is a well-defined probability measure by normalization and the fact that Θ is Polish and endowed with its Borel σ-algebra. This implies that there is $\theta_0 \in \Theta$ such that $\ell(T(x), \theta_0) = \int_\mathcal{A} \ell(a, \theta_0)\, d\delta(x, a)$. Since $\mathcal{N}_0 \cup \mathcal{N}$ is negligible for the marginal distribution P_π, it follows that $\delta(x, \cdot) = \delta_T(x, \cdot)$ for P_π-a.e. $x \in \mathcal{X}$. Because P_π dominates P_0 by assumption, the result holds for P_0-a.e. $x \in \mathcal{X}$. Suppose now that there exist two non-randomized Bayes rules T_1 and T_2 for π that differ on some set with strictly positive P_0-measure. Then the Dirac measure at $(T_1(x) + T_2(x))/2$ for any fixed $x \in \mathcal{X}$ is a valid decision rule $\delta^* \in \mathcal{D}$ by convexity of $\mathcal{A}$. By strict convexity of the loss and the fact that P_0 is dominated by P_π, it follows directly that $B(\delta^*, \pi) < B_\Delta(\pi)$: a contradiction. □

Remark B.1. If the assumption that P_0 is dominated by the marginal distribution P_π does not hold, then the result remains valid for P_π-a.e. $x \in \mathcal{X}$. This is generally not enough to ensure that any two Bayes rules δ and δ' for a given finite measure π satisfy $r(\delta, \theta) = r(\delta', \theta)$ for all $\theta \in \Theta$.

Remark B.2. In practice, the compactness of the action space $\mathcal{A}$ may be obtained by compactifying a non-compact Fréchet space Θ. In this case, the action space $\mathcal{A}$ may fail to be a vector space and Lemma B.1 cannot be directly applied. However, the result remains valid if it holds that:

(a) for every $\theta \in \Theta$, the loss function $\ell(\cdot, \theta)$ is lower semi-continuous on the compactified space $\mathcal{A}$ and strictly convex on Θ;

(b) any Bayes rule δ_π for any finite measure $\pi \in \mathcal{M}_f(\Theta) \setminus \{0\}$ with finite Bayes risk must be supported P_0-a.e. on the non-compact space Θ;

(c) the barycenter $\int_\Theta a \, d\delta_\pi(x, a)$ is well-defined and belongs to Θ for P_0-a.e. $x \in \mathcal{X}$.

Under these conditions, it is seen that the proof of Lemma B.1 holds unchanged. This guarantees then that the Bayes rule δ_π for any finite measure $\pi \in \mathcal{M}_f(\Theta) \setminus \{0\}$ exists, is unique, and non-randomized in the sense that $\delta_\pi(x, \cdot)$ is a unique Dirac measure supported on Θ for P_0-a.e. $x \in \mathcal{X}$. We verify these conditions in the normal mean problem of Section 4 in Lemma B.2 below.

Lemma B.2. *Let $p \in (1, +\infty)$ and define $\ell_p(\infty, \theta) := +\infty$ and $\ell_p(a, \theta) := |a - \theta|^p$ for all $a \in \mathbb{R}$ and all $\theta \in \mathbb{R}$. Consider the statistical experiment $\big(\mathbb{R}, \{N(\theta, 1) : \theta \in \mathbb{R}\}, \mathbb{R} \cup \{\infty\}, \ell_p(\cdot, \cdot)\big)$. For any measure $\pi \in \mathcal{M}_f(\mathbb{R}) \setminus \{0\}$, the Bayes rule δ_π for π exists, is unique, and non-randomized in the sense that $\delta_\pi(x, \cdot)$ is a unique Dirac measure supported on $\mathbb{R}$ for λ-a.e. $x \in \mathbb{R}$.*

Proof. The statistical experiment is directly seen to be regular for $P_0 = \lambda$ the Lebesgue measure on $\mathbb{R}$. This guarantees that $\mathcal{D}$ can be endowed with the regular topology under which it is compact. For any $\theta \in \mathbb{R}$, the loss $\ell_p(\cdot, \theta)$ is continuous on $\mathbb{R}$. Since $\ell_p(\infty, \theta) = \infty$ and $\lim_{a \to \infty} \ell_p(a, \theta) = \infty$ for any $\theta \in \mathbb{R}$, the loss $\ell_p(\cdot, \theta)$ is lower semi-continuous on $\mathbb{R} \cup \{\infty\}$. By Lemma 2.2, this guarantees the existence of a Bayes rule δ_π for any finite measure $\pi \in \mathcal{M}_f(\mathbb{R}) \setminus \{0\}$. Moreover, the marginal distribution P_π satisfies $P_\pi(B) = \int_B \int_{\mathbb{R}} \phi(x - \theta) \, d\pi(\theta) \, dx$ for all Borel sets $B \subseteq \mathbb{R}$. Since $\pi \neq 0$ and $\phi(\theta) > 0$ for all $\theta \in \Theta$, we necessarily have $m_\pi(x) := \int_{\mathbb{R}} \phi(x - \theta) \, d\pi(\theta) > 0$ for all $x \in \mathbb{R}$. This is enough to conclude that the marginal distribution P_π dominates the Lebesgue measure λ on $\mathbb{R}$. Since the minimax rule $\delta^*(x) = x$ has constant risk, it has finite integrated risk for π, and we necessarily have $B_\Delta(\pi) < \infty$. Since $\ell_p(\infty, \theta) = \infty$ for all $\theta \in \mathbb{R}$ and $B_\Delta(\pi) < \infty$, we necessarily have $\delta_\pi(x, \{\infty\}) = 0$ for P_π-a.e. $x \in \mathbb{R}$. Since λ is dominated by P_π, it follows that $\delta_\pi(x, \{\infty\}) = 0$ for λ-a.e. $x \in \mathbb{R}$. From the standard inequality $|a|^p \leq 2^{p-1}(|a - \theta|^p + |\theta|^p)$ for all $a \in \mathbb{R}$ and all $\theta \in \mathbb{R}$ and integration over the measure $\delta_\pi(x, \cdot)$ and the posterior distribution $\pi(\cdot|x)$, we have for λ-a.e. $x \in \mathbb{R}$ that $2^{1-p} \int_{\mathbb{R}} |a|^p \, d\delta_\pi(x, a) \leq \int_{\mathbb{R}} \int_{\mathbb{R}} |a - \theta|^p \, d\delta_\pi(x, a) \, d\pi(\theta|x) + \int_{\mathbb{R}} |\theta|^p \, d\pi(\theta|x)$. The first integral on the right is finite since $B_\Delta(\pi) < \infty$ and the second integral is finite since the posterior distribution $\pi(\cdot|x)$ has finite absolute moments of any order for all $x \in \mathbb{R}$. This is enough to conclude that $\int_{\mathbb{R}} a \, d\delta_\pi(x, a) < \infty$ for λ-a.e. $x \in \mathbb{R}$. The result then follows from Remark B.2 by following the proof of Lemma B.1. □