

Coverage Guarantees as Ambiguity Sets: Coherent Decision Making under Statistical Uncertainty

Selman Erol*
Rakesh Vohra†

August 26, 2026

Abstract

Confidence and conformal sets are used to inform decisions. To employ them, a decision maker must give it a probabilistic interpretation, and that interpretation determines an ambiguity set: the probability distributions over states she regards as possible. This paper characterizes when a proffered confidence/conformal set admits a coherent probabilistic interpretation, a property we call *rationalizability*. We then study decision making over a rationalizable confidence/conformal set. A common practice assigns zero probability to states outside the proffered set. This replaces the original coverage requirement with one that can become incompatible with additional probabilistic information. The rationalizable set, however, preserves the original guarantee and remains compatible with information refinements. As information accumulates and the ambiguity set shrinks toward a singleton, the associated robust decision rule converges to Bayesian expected-loss minimization. This consistency under information refinement is our normative justification for the coherent interpretation.

Keywords: ambiguity sets, confidence sets, credible sets, conformal prediction, robust decision making, partial identification

JEL Codes: D81; C44; C11.

1 Introduction

Every operational use of a statistical uncertainty set implicitly defines an *ambiguity set*: the collection of probability distributions over an unknown state that the decision maker considers plausible. Consider a decision maker who must choose an action whose loss depends

*Tepper School of Business, Carnegie-Mellon University. Email: selman.erol@gmail.com.

†Department of Economics & Department of Electrical and Systems Engineering, University of Pennsylvania. Email: rvohra@seas.upenn.edu.

on an unknown state i drawn from a finite set S , having observed a feature vector f —a covariate profile, a test result, a signal—from a finite set F . An analyst hands her a set $C_f \subseteq S$, a confidence set, say, meant to contain the true state with probability at least $1 - \alpha$. If she minimizes worst-case expected loss over the states in C_f , she is treating C_f as her ambiguity set. The question this paper addresses is: what ambiguity set is *coherently induced* by a coverage guarantee, given what the decision maker observes about the data? The answer matters because different interpretations of the same set induce different ambiguity sets. The standard practice, which [Manski \[2021\]](#) calls *as-if decisions with set estimates*, treats the reported set as a *hard support restriction*: all probability mass is placed on states within C_f and none outside it. This *as-if interpretation* conflicts with the frequentist interpretation of a confidence set: under repeated sampling, the random set covers the truth at least $1 - \alpha$ of the time. A *credible set* is a statement about a distribution over states given what was observed: conditional on f , the states in the set carry probability at least $1 - \alpha$. The as-if decision maker treats the proffered confidence set as a credible set with a hard support restriction.

The first question we address is whether the as-if interpretation of confidence sets is coherent at all. Given the observables—the marginal state frequencies $\{\pi_i\}_{i \in S}$, the marginal feature frequencies $\{p_f\}_{f \in F}$, and the sets $\{C_f\}_{f \in F}$ —is there a joint distribution over state-feature pairs consistent with them and the coverage guarantee? If not, no data generating process could have produced these sets with the claimed coverage. We call this a *rationalizability* question, after the revealed-preference tradition [[Afriat, 1973](#), [Chambers and Echenique, 2016](#)]. Rationalizability in our context is the statement that *some* Bayes model consistent with the observed marginals makes the entire reported family $\{C_f\}_{f \in F}$ credible at the claimed level. If a rationalizing distribution exists, the coherent ambiguity set $\mathcal{X}$ is the set of all of them, and the decision maker may proceed.

In treating the marginals as population quantities we follow the partial identification literature [[Manski, 2003](#), [Beresteanu and Molinari, 2008](#)]: the marginals π_i and p_f are observed and treated as point-identified, while the joint distribution $x(i, f)$ is only partially identified by the coverage constraint.

Two interpretations of π should be highlighted that vary depending on whether one is discussing a confidence set or a conformal set. When i is a label realized in the data, as in conformal prediction, π_i is interpreted as an observed frequency and the rationalizability check is a test of the reported sets against data. When i is a parameter, as in the confidence-set case, there is no repeated realization of i , and π is interpreted as the decision maker’s prior.

The conditions under which a credible set can be interpreted as a confidence set are well

known (see the Bernstein–von Mises Theorem). However, there is no exact converse. The precise relationship of rationalizability to this issue is discussed in Section 2.

The second question we address is how rationalizable sets should inform decisions. The as-if interpretation induces an ambiguity set $\tilde{\mathcal{X}}$ assigning zero probability to states outside C_f and this set may be incompatible with other probabilistic information the decision maker holds or later receives: the observed marginal state frequencies $\{\pi_i\}_{i \in S}$, group membership, or sequential updates. Keeping the coverage guarantee as an inequality rather than collapsing it to the support restriction $\Pr(i \in C_f \mid f) = 1$ and fixing the observed marginals, yields an ambiguity set $\mathcal{X}$ that absorbs such information without contradiction. This ability to incorporate new information—*consistency under information refinement*—is the paper’s normative foundation. As $\mathcal{X}$ shrinks toward a singleton under accumulating information, the associated robust decision rule converges to Bayesian expected-loss minimization (Theorem 7.1). The as-if rule does not converge, however much information arrives, because its distortion of the coverage constraint leaves it unable to absorb that information. We propose optimizing minimax expected loss over $\mathcal{X}$ itself, and call the resulting decision maker the **rationalizable as-if** decision maker. This is the only rule consistent with everything the decision maker observes that also takes the coverage guarantee at face value. Two recent papers propose minimax rules for coverage-based sets: Andrews and Chen [2025] for confidence sets and Kiyani et al. [2025] for conformal sets. Each optimizes over a set larger than $\mathcal{X}$; the as-if rule discards the marginals entirely and treats C_f as the full support.

Section 2 discusses related work. Section 3 introduces the setting. Sections 4 and 5 characterize $\mathcal{X}$ for confidence and conformal sets. Section 6 studies the decision problem over $\mathcal{X}$, and Section 7 establishes the convergence result.

2 Related Literature

Confidence sets and credible sets

The question of when a credible set is a confidence set is answered in the affirmative within a regular parametric mode by the Bernstein–von Mises theorem. It shows that posterior credible sets attain frequentist coverage $1 - \alpha + o(1)$ asymptotically [van der Vaart, 1998, Ch. 10]. Exact or higher-order agreement requires the prior to be selected for the purpose, which is the theory of probability matching priors initiated by Welch and Peers [1963] and surveyed in Datta and Mukerjee [2004].

Outside this case, credible sets can have coverage bounded away from the nominal level [Freedman, 1999, Szabó et al., 2015], and in partially identified models Bayesian credible sets

are asymptotically strictly narrower than frequentist confidence sets and remain sensitive to the prior on the non-identified directions [Moon and Schorfheide, 2012, Kline and Tamer, 2016]. The robust Bayesian solution to that last problem is Giacomini and Kitagawa [2021], who replace the prior on the non-identified directions by a set of priors and report the resulting posterior bounds. It is a close relative of what we do: their ambiguity set is generated by the identified set, ours by the reported coverage guarantee and the observed marginals.

A second literature asks when may a confidence set be interpreted as a credible set? In general it may not. Buehler [1959] and Robinson [1977] show that standard confidence procedures admit *relevant subsets*, events recognizable from the data on which conditional coverage departs from the nominal level, so that a post-data probabilistic reading is untenable. Pratt [1961] and Fraser [2011] make the point from other directions, and Balch et al. [2019] show that any epistemic probability built from a coverage guarantee assigns high probability to some false hypotheses. Heath and Sudderth [1978] give a sharp coherence statement: a procedure escapes a sure-loss bet exactly when it is Bayes against some finitely additive prior. Closest to us is Müller and Norets [2016], who call a confidence set *bet-proof* if no informed inspector can profit by betting against it, show that containing a $1 - \alpha$ credible set for *some* prior suffices for bet-proofness, and construct short bet-proof sets in nonstandard problems. On the design side, Hoff [2023] constructs prediction regions that are Bayes-optimal subject to a frequentist coverage constraint, which is the converse exercise to ours: he builds sets that admit both interpretations, we test reported sets for the second.

Our contribution is to make the existential quantifier operational and testable. Müller and Norets [2016] ask whether some prior—any prior—renders a given set credible, in a continuous parametric model, one set at a time. We ask whether some joint distribution with the *observed* state marginal π and feature marginal p renders the *entire reported family* $\{C_f\}_{f \in F}$ credible simultaneously. Disciplining the prior by the observed frequencies and requiring the interpretation to hold across all features at once turns the question into a finite system of linear inequalities. Theorem 4.1 solves that system. The answer is a subset condition checkable by a single maximum-flow computation, and Corollary 4.2 reports the highest credibility the data permit. The decision-theoretic half of the paper then asks what a decision maker should do once the interpretation is supported.

Minimax decision rules

Minimax criteria for decision making under uncertainty have a long history. In the general form, the decision maker holds a set of probability distributions over states rather than a

single prior and chooses an action to minimize worst-case expected loss over that set. This is the framework of robust Bayes analysis [Berger, 1985, Wasserman and Kadane, 1992], imprecise probabilities [Walley, 1991], and maximin expected utility [Gilboa and Schmeidler, 1989]. A decision maker reaches for it whenever she distrusts a point prior—because the prior is not identified from data, because the data generating process may be adversarial, or because she guards against misspecification [Hansen and Sargent, 2008]. The set she entertains may be defined by moment conditions [Beresteanu, 2006], by Huber neighbourhoods of a reference distribution [Huber and Ronchetti, 2009], or by the sharp identification region of a partially identified model [Manski, 2003, Molchanov and Molinari, 2018]. This paper sits in that tradition: the decision maker minimizes worst-case expected loss over the rationalizable set $\mathcal{X}$, a polyhedron of joint distributions fixed by observed summary statistics and the coverage constraint. What sets the paper apart is that $\mathcal{X}$ is not given but is derived from a prior consistency check on the reported sets.

Partial identification and treatment assignment

The partial identification literature [Manski, 2003, Beresteanu and Molinari, 2008, Molchanov and Molinari, 2018] studies what can be learned about a functional of a distribution when the data are consistent with a set of distributions rather than one. Its decision-theoretic counterpart asks how a planner should act when the relevant parameter is only partially identified. Manski [2004] and Stoye [2011] are the canonical references for treatment assignment: a planner forms an identified interval for the average treatment effect in each covariate cell and chooses a rule to minimize worst-case regret or expected loss over all data generating processes consistent with that interval. This exercise and ours share a focus on minimax expected loss over a set of distributions derived from partial identification, with the decision variable a rule mapping a signal to an action. The differences are substantive. First, the source of the identified set differs. In treatment assignment, the set comes from the fundamental problem of causal inference: potential outcomes are never jointly observed, and the identified interval is the sharp consequence of that limitation under maintained assumptions such as independence or monotone selection. The set is non-empty by construction—the true process always lies in it, since the set is what the data cannot rule out. Here $\mathcal{X}$ comes from a consistency requirement on reported sets: it is the joint distributions over state-feature pairs compatible with the observed marginals and the claimed coverage under the as-if interpretation, and it may be empty. Emptiness is exactly incoherence. The rationalizability check—is $\mathcal{X}$ non-empty?—has no analogue in treatment assignment. Second, the decision variable differs. The treatment planner assigns subjects to treatments and aggregates welfare over a

population; she chooses for others. Here the decision maker chooses an action at a realized feature vector, with no population to assign. The problem is single-agent, not policy design with a population dimension. Third, the role of the signal differs. In treatment assignment, covariates are features of subjects on which rules may condition, but they are not tied to the uncertainty being quantified. Here the feature vector f is itself the object that generates C_f , and the rationalizability condition links the sets across features through the marginal structure of the joint distribution.

Distributionally robust optimization

The framework connects to distributionally robust optimization (DRO), a large literature in operations research and machine learning that minimizes worst-case expected loss over an ambiguity set [Delage and Ye, 2010, Wiesemann et al., 2014, Kuhn et al., 2019]. DRO ambiguity sets are typically moment sets, Wasserstein balls around an empirical distribution, or ϕ -divergence balls. Our coherent set $\mathcal{X}$ is a DRO ambiguity set fixed by marginal moment constraints and coverage constraints, and the rationalizable-as-if decision maker solves a DRO problem over it. The connection is more than formal. DRO has recently asked how ambiguity sets should be built from data to support reliable decisions [Rahimian and Mehrotra, 2019]. We answer this for coverage-based procedures: given a confidence or conformal guarantee, what ambiguity set does coherence require? Our answer, $\mathcal{X}$, has a clean linear-programming characterization that differs from the Wasserstein and moment-based sets, because it derives from a rationalizability condition rather than a distance from an empirical distribution. The as-if interpretation corresponds to a degenerate DRO set placing all mass within the reported set; our point that such a set may be incompatible with side information is the DRO concern that a badly designed ambiguity set may be too conservative or inconsistent with side information. The set $\mathcal{X}$ avoids both by construction.

Justifications of as-if decision making

Two papers justify as-if decisions in settings close to ours. Andrews and Chen [2025] work with confidence sets when the decision maker has a prior over joint distributions and a default action of fixed loss in all states. They optimize over all joint distributions consistent with conditional coverage and summing to one, without fixing marginals, and characterize the optimal mixing probability between the as-if action and the default, bounding the downside risk of the as-if rule. Kiyani et al. [2025] work with conformal sets and optimize over all joint distributions consistent with marginal coverage and summing to one, again without fixing marginals; they show the as-if rule is optimal when the conformal sets are themselves

chosen to solve a joint optimization tailored to the loss function. Neither paper fixes the marginal frequencies at the decision stage. A decision maker who used those frequencies to establish rationalizability should, for consistency, carry them into the decision problem. The rationalizable-as-if decision maker does: she optimizes over $\mathcal{X}$, which fixes both marginals and the coverage constraint. The as-if rule imposes $\Pr(i \in C_f \mid f) = 1$ and does not fix marginals. The extra devices in each paper serve to bound the as-if rule’s performance: for [Andrews and Chen \[2025\]](#), the default action and the tailored prior; for [Kiyani et al. \[2025\]](#), the requirement that the sets solve a loss-dependent optimization. Our starting point is different: we take the as-if interpretation as given and ask what consistency then demands, which leads to fixing the observed marginals and the coverage constraint together, giving $\mathcal{X}$.

Demand rationalizability

Our rationalizability question is in the spirit of demand rationalizability, begun by [Afriat \[1973\]](#) and surveyed in [Chambers and Echenique \[2016\]](#). That literature asks whether observed choices are consistent with some utility-maximizing model; if not, the model is rejected. We ask the analogous question for reported sets: are they consistent with some joint distribution over state-feature pairs satisfying the coverage guarantee—conditional for confidence sets, marginal for conformal sets? If not, the sets are rejected as incoherent. The novelty is the application to statistical outputs rather than economic choices, and the link from the rationalizability check to the decision problem. The analogy is worth pressing one step further, because it disciplines what [Corollary 4.1](#) claims. Afriat’s theorem certifies that *some* utility function rationalizes the observed choices, not that the consumer has that utility function. Likewise, rationalizability here certifies that *some* Bayes model consistent with the marginals makes the reported sets credible. It does not certify that the analyst’s procedure was Bayes, nor that the decision maker’s own prior is among those that work.

3 Model of the Data

An analyst provides a coverage-based uncertainty set to a decision maker, who will use it as the ambiguity set for a minimax decision problem. Let S be the finite set of possible states—values of a parameter, or labels such as cancer/no cancer. The analyst observes feature vectors and uses them to build the set. Let F be the finite set of feature vectors. To each $f \in F$ corresponds a set $C_f \subseteq S$, the analyst’s uncertainty set at f . We call C_f a *conditional* $1 - \alpha$ confidence set if, for each f separately, the true state lies in C_f with probability at least $1 - \alpha$ given f under the data generating process. This is a *conditional*

guarantee: it holds feature by feature, not merely on average. Before acting, the decision maker wants to check that the sets are reliable. Throughout, she observes:

1. π_i : the frequency of state $i \in S$ in the data, or, when i is a parameter, the prior she brings. See Remark 1.
2. p_f : the frequency of feature vector $f \in F$ in the data.
3. the sets $\{C_f\}_{f \in F}$ (or $\{K_f\}_{f \in F}$ for conformal sets) and the claimed level $1 - \alpha$.

She does not observe the joint distribution $\{x(i, f)\}$. The rationalizability check asks whether any joint distribution is consistent with the observables; the decision problem then optimizes over all such distributions.

REMARK 1 (TWO INTERPRETATIONS OF π) *When i is a label realized in the data, as in the conformal prediction case, π_i is an observed frequency and the rationalizability check is a test of the reported sets against data. When i is a parameter, as in the confidence-set case, and π is a prior the decision maker brings. Under this interpretation, Corollary 4.1 is a statement about which priors make the reported sets credible, and Theorem 4.1 tells the decision maker whether the prior she holds is one of them.*

REMARK 2 (ON TREATING MARGINALS AS KNOWN) *We treat π_i and p_f as population quantities, standard in the partial identification literature [Manski, 2003, Beresteanu and Molinari, 2008]: there the marginal of observed variables is point-identified, while the joint distribution of observed and unobserved variables is partially identified, and the identified set is derived holding the marginal fixed. Our setup is identical: π_i and p_f are the observed marginals and $x(i, f)$ is partially identified by the coverage constraint. When the marginals are estimated, we recommend reporting the estimated credibility ceiling $\hat{\kappa}$ of Corollary 4.2 with a one-sided confidence bound, rather than a binary verdict on the emptiness of $\hat{\mathcal{X}}$. Near the boundary, small sampling noise in $(\hat{\pi}, \hat{p})$ flips emptiness on and off, while κ moves continuously. Testing $H_0 : \kappa \geq 1 - \alpha$ is the better behaved formulation, and the minimum-of-ratios form of κ puts it within reach of standard inference on a minimum, e.g. Chernozhukov et al. [2013].*

For the features to be informative, the analyst posits an association, possibly stochastic, between states and features, modelled as an unobserved joint distribution over state-feature pairs. The analyst’s task is to reconstruct it from the data. Given π , p , $\alpha \in (0, 1)$, and $\{C_f\}_{f \in F}$, we ask: is there a joint distribution over state-feature pairs consistent with these? If none, we reject the hypothesis that the reported sets are reliable; the as-if interpretation is incoherent. If one exists, the sets are rationalized by it, and the as-if interpretation is at least consistent.

4 Rationalizability for Confidence Sets

Write $\{x(i, f)\}_{i \in S, f \in F}$ for the joint probability of state i and feature f . Given $\pi, p, \alpha \in (0, 1)$, and $\{C_f\}_{f \in F}$, a consistent joint distribution exists if and only if the following system is feasible:

$$\sum_{f \in F} x(i, f) = \pi_i \quad \forall i \in S \quad (1)$$

$$\sum_{i \in S} x(i, f) = p_f \quad \forall f \in F \quad (2)$$

$$\sum_{i \in C_f} x(i, f) \geq (1 - \alpha)p_f \quad \forall f \in F \quad (3)$$

$$x(i, f) \geq 0 \quad \forall i, f \quad (4)$$

Equation (1) fixes the state marginal, (2) the feature marginal. Equation (3) is the conditional coverage constraint: dividing by p_f gives $\Pr(i \in C_f \mid f) \geq 1 - \alpha$ for each f separately. This is what distinguishes confidence sets from conformal sets, which impose only one constraint summed across features. Read from right to left, that same division says something else. Under any x satisfying (1)–(4), the pair (π, x) is a Bayes model—prior π over states, joint x —in which C_f is a $1 - \alpha$ credible set at every f . Feasibility of the system is therefore the question of whether the reported confidence sets admit a credible-set reading consistent with the observed marginals. Corollary 4.1 records this.

THEOREM 4.1 *System (1–4) is feasible if and only if for all $T \subseteq F$:*

$$(1 - \alpha) \sum_{f \in T} p_f \leq \sum_{i \in \cup_{f \in T} C_f} \pi_i. \quad (5)$$

Note: At $T = F$ the condition reads $(1 - \alpha) \leq \sum_{i \in \cup_{f \in F} C_f} \pi_i$: the covered states carry π -mass at least $1 - \alpha$. At $T = \{f\}$ it reads $(1 - \alpha)p_f \leq \pi(C_f)$: the coverage required at each feature cannot exceed the marginal mass of the states in its set.

Proof. Necessity is immediate. Let x satisfy (1)–(4) and fix $T \subseteq F$. Since the sets $\{C_f\}_{f \in T}$ all sit inside $\cup_{f \in T} C_f$ and $x \geq 0$,

$$(1 - \alpha) \sum_{f \in T} p_f \leq \sum_{f \in T} \sum_{i \in C_f} x(i, f) \leq \sum_{i \in \cup_{f \in T} C_f} \sum_{f \in F} x(i, f) = \pi(\cup_{f \in T} C_f),$$

which is (5).

For sufficiency, we construct a feasible x from a maximum flow. Theorem A.1 of Appendix A builds a network $\mathcal{N}$ with source capacities $(1 - \alpha)p_f$ and sink capacities π_i , shows

that its minimum cut is $\min_T[(1-\alpha) \sum_{f \in T} p_f + \pi(\cup_{f \notin T} C_f)]$, and shows that a flow saturating every source edge can be completed to a solution of (1)–(4). Condition (5), applied at $F \setminus T$, is exactly the statement that every cut has capacity at least $1 - \alpha$. ■

Condition (5) is a Gale–Hoffman deficiency condition, and the proof just given is the corresponding flow argument. Feasibility can therefore be checked in time polynomial in $|S|$ and $|F|$; see Appendix A. Appendix ?? gives an alternative and self-contained proof by the Farkas lemma and total unimodularity, which some readers may prefer and which we use again in Appendix B.

COROLLARY 4.1 (CREDIBLE-SET READING) *The following are equivalent.*

- (i) *The reported family $\{C_f\}_{f \in F}$ is rationalizable at level $1 - \alpha$, i.e. $\mathcal{X} \neq \emptyset$.*
- (ii) *There is a joint distribution over $S \times F$ whose state marginal is π and whose feature marginal is p under which, for every $f \in F$, C_f is a $1 - \alpha$ credible set: $\Pr(i \in C_f \mid f) \geq 1 - \alpha$.*
- (iii) *$(1 - \alpha) \sum_{f \in T} p_f \leq \pi(\cup_{f \in T} C_f)$ for every $T \subseteq F$.*

Proof. (i) $\Leftrightarrow$ (iii) is Theorem 4.1. For (i) $\Leftrightarrow$ (ii), a joint distribution with marginals π and p is exactly a prior π over states together with a likelihood $\Pr(f \mid i)$. Dividing (3) by $p_f > 0$ gives $\Pr(i \in C_f \mid f) \geq 1 - \alpha$, and multiplying back recovers (3). ■

COROLLARY 4.2 (CREDIBILITY CEILING) *Let*

$$\kappa(\pi, p, \{C_f\}) := \min_{\emptyset \neq T \subseteq F} \frac{\pi(\cup_{f \in T} C_f)}{\sum_{f \in T} p_f}.$$

Then $\{C_f\}$ is rationalizable at level $1 - \alpha$ if and only if $1 - \alpha \leq \kappa$. Hence $\kappa \wedge 1$ is the highest credibility the observed marginals permit the analyst to claim for the reported sets, and any claimed level above it is refuted by the marginals alone.

Proof. Immediate from (iii) of Corollary 4.1, since $T = \emptyset$ is vacuous and the condition is monotone in $1 - \alpha$. ■

REMARK 3 (POSSIBILITY, NOT IDENTIFICATION) *Corollary 4.1 certifies that some Bayes model consistent with the marginals sustains the credible-set reading, not that the analyst’s procedure was Bayes or that the decision maker’s own prior is among those that work.*

REMARK 4 (RELATION TO BET-PROOFNESS) [Müller and Norets \[2016\]](#) show that a set containing a $1 - \alpha$ credible set for some prior is bet-proof: no inspector who observes the data can construct a bet against it with uniformly positive expected winnings. Corollary 4.1 is the finite-state analogue with the prior disciplined by the observed state frequencies and the requirement imposed on the whole reported family at once. Without it, any family with $\cup_f C_f \neq \emptyset$ can be made credible by a prior concentrated on the covered states. Condition (5) is exactly what remains of the question once the prior must reproduce π .

4.1 A Numerical Example in Which a Subset Condition Binds

Theorem 4.1 imposes one inequality per subset of F . The instances $T = \{f\}$ and $T = F$ are transparent, and it is natural to ask whether the rest are ever the binding ones. They are, and the mechanism is worth seeing: two distinct features that certify the *same* state can each be individually modest in its demands and jointly impossible.

Let $S = \{1, 2, 3\}$ and $F = \{f_1, f_2, f_3\}$, with

$$C_{f_1} = C_{f_2} = \{1\}, \quad C_{f_3} = \{2\},$$

$\pi = (0.30, 0.60, 0.10)$, $p = (0.25, 0.25, 0.50)$, and a claimed level $1 - \alpha = 0.9$. Two different signals, f_1 and f_2 , each certify state 1; the common signal f_3 certifies state 2; state 3 is certified by nothing. Condition (5) at each non-empty T requires $(1 - \alpha) \sum_{f \in T} p_f \leq \pi(\cup_{f \in T} C_f)$:

$$\begin{array}{llll} T = \{f_1\} : & 0.225 \leq 0.30 & \checkmark \\ T = \{f_2\} : & 0.225 \leq 0.30 & \checkmark \\ T = \{f_3\} : & 0.450 \leq 0.60 & \checkmark \\ T = \{f_1, f_2\} : & 0.450 \leq 0.30 & \text{fails} \\ T = \{f_1, f_3\} : & 0.675 \leq 0.90 & \checkmark \\ T = \{f_2, f_3\} : & 0.675 \leq 0.90 & \checkmark \\ T = F : & 0.900 \leq 0.90 & \checkmark \quad (\text{with equality}) \end{array}$$

Every single-feature check passes, the grand check passes exactly, and the reported sets are nonetheless not rationalizable. The pair $\{f_1, f_2\}$ is the sole binding constraint. The reason is visible in the arithmetic. Coverage at f_1 and at f_2 both draw on the same state, and together they promise $0.9 \times 0.5 = 0.45$ units of joint mass in state 1, while the state marginal supplies only $\pi_1 = 0.30$. No per-feature check can detect this, because neither feature alone over-promises.

Corollary 4.2 prices the failure. Here

$$\kappa = \min_{\emptyset \neq T \subseteq F} \frac{\pi(\cup_{f \in T} C_f)}{\sum_{f \in T} p_f} = \min \left\{ 1.2, 1.2, 1.2, \frac{0.30}{0.50}, 1.2, 1.2, 0.9 \right\} = 0.60,$$

attained at $T = \{f_1, f_2\}$. The observed marginals support a credible-set reading of these sets at no level above 60%, though every feature taken by itself would support any level whatever, and the union condition would support 90%. The maximum-flow construction of Appendix A confirms this directly. Source capacities are $(0.225, 0.225, 0.45)$, sink capacities are $(0.30, 0.60)$, and the maximum flow is $0.75 < 0.9$. The minimum cut it returns is exactly the subset $\{f_1, f_2\}$ that fails, so the check does not merely reject but names the features responsible.

The pattern generalizes. The subset conditions bind precisely when a group of features shares a small pool of certified states: duplicated tests for the same condition, several covariate cells whose confidence sets all name the same parameter value, a battery of screens all pointing to one diagnosis. In such settings a family of individually defensible guarantees can be jointly indefensible, and Theorem 4.1 is the check that catches it.

REMARK 5 *The one-feature instance of the condition is familiar from diagnostic testing. With $C_{f_+} = \{\text{diseased}\}$ a singleton, the instance $T = \{f_+\}$ of Theorem 4.1 reads $1 - \alpha \leq \pi_D/p_+$, and π_D/p_+ is precisely the largest positive predictive value the marginals allow. A claimed conditional coverage for a singleton set is therefore a claim about PPV, and the check asks only that the claim not exceed what prevalence and the positivity rate can deliver. Theorem 4.1 extends this familiar one-feature bound to necessary and sufficient conditions across all $T \subseteq F$, where, as the example above shows, the binding constraint need not be any single feature. Casual observers may interpret a reported sensitivity as though it were the conditional coverage of the singleton set $\{\text{diseased}\}$. It is not: sensitivity is $\Pr(f_+ \mid \text{diseased})$ and conditional coverage here is $\Pr(\text{diseased} \mid f_+)$. The check above says what the marginals will and will not support under this interpretation.*

5 Rationalizability for Conformal Sets

Conformal sets carry a weaker guarantee. Rather than coverage at each f separately—as (3) requires—a conformal collection $\{K_f\}_{f \in F}$ need only cover the true state with probability $1 - \alpha$ on average over features. Conditional coverage implies marginal coverage but not the

reverse. Formally, $\{K_f\}_{f \in F}$ has marginal coverage $1 - \alpha$ if

$$\sum_{f \in F} \sum_{i \in K_f} x(i, f) \geq 1 - \alpha, \quad (6)$$

a single constraint in place of the $|F|$ constraints (3). It allows wide variation across features: K_f may cover the truth with probability well below $1 - \alpha$ at some f , provided other features compensate. Given $\pi, p, \alpha \in (0, 1)$, and $\{K_f\}_{f \in F}$, we ask whether some joint distribution gives $\{K_f\}$ marginal coverage. This is feasibility of:

$$-\sum_{f \in F} x(i, f) \geq -\pi_i \quad \forall i \in S \quad (7)$$

$$\sum_{i \in S} x(i, f) \geq p_f \quad \forall f \in F \quad (8)$$

$$\sum_{f \in F} \sum_{i \in K_f} x(i, f) \geq 1 - \alpha \quad (9)$$

$$x(i, f) \geq 0 \quad \forall i \in S, f \in F \quad (10)$$

Summing (7) over i gives $\sum_{i, f} x(i, f) \leq 1$ and summing (8) over f gives $\sum_{i, f} x(i, f) \geq 1$, so both hold with equality and the marginal inequalities are in fact equalities; the system has the same feasible set as the one with $\sum_f x(i, f) = \pi_i$ and $\sum_i x(i, f) = p_f$.¹

THEOREM 5.1 *Assume $K_f \neq \emptyset$ for all $f \in F$. System (7–10) is feasible if and only if for all $T \subseteq F$:*

$$(1 - \alpha) \leq \sum_{f \in T} p_f + \pi \left(\bigcup_{f \notin T} K_f \right). \quad (11)$$

The conformal analogue of Corollary 4.2 is immediate. The highest level at which $\{K_f\}$ admits a credible-set reading consistent with the marginals is

$$\kappa^K = \min_{T \subseteq F} \left[\sum_{f \in T} p_f + \pi \left(\bigcup_{f \notin T} K_f \right) \right]. \quad (12)$$

See Appendix C for the proof.

The instance $T = \emptyset$ is the union condition $(1 - \alpha) \leq \pi(\bigcup_{f \in F} K_f)$, easy to meet when the sets cover most of S . The instance $T = F$ is trivial. The bite is in $T \neq \emptyset, F$. When the p_f are equal, the right-hand side varies little with T , and the union condition is nearly

¹This step is legitimate because (7) and (8) bound the full sums $\sum_f x(i, f)$ and $\sum_i x(i, f)$.

sufficient. The extra instances bind when (i) some features are rare (p_f small) and (ii) those rare features are the only cover for certain states. Taking T to be the rare features then makes both $\sum_{f \in T} p_f$ and $\pi(\cup_{f \notin T} K_f)$ small, and the capacity can fall below $1 - \alpha$ even when the union covers all of S . Conformal rationalizability is easier to meet than its confidence counterpart because it pools coverage across features. The exact statement of how far the pooling goes is (12): the maximum coverage attainable at the observed marginals is κ^K , and $\{K_f\}$ is rationalizable exactly when $1 - \alpha \leq \kappa^K$. It is tempting to summarize this by the simpler condition $\sum_{f \in F} \min(p_f, \pi(K_f)) \geq 1 - \alpha$, on the grounds that $\min(p_f, \pi(K_f))$ units can be routed through each feature. That is incorrect unless the sets $\{K_f\}$ are pairwise disjoint. Different features can draw on the same states, and the state marginal is a shared budget. With $S = \{1, 2\}$, $F = \{f_1, f_2\}$, $K_{f_1} = K_{f_2} = \{1\}$, $\pi = (0.5, 0.5)$ and $p = (0.5, 0.5)$, the sum of minima is 1, but total coverage is $\sum_f x(1, f) \leq \pi_1 = 0.5$, and indeed $\kappa^K = 0.5$, attained at $T = \emptyset$.

Medical testing illustration. Let f_1 be a positive test and f_2 a negative one, with $K_{f_1} = \{\text{diseased}\}$, $K_{f_2} = \{\text{healthy}\}$, $\pi_{\text{diseased}} = \pi_{\text{healthy}} = 0.5$, $p_{f_1} = 0.9$, $p_{f_2} = 0.1$. Here $\pi(\cup_f K_f) = 1$, so the instances $T = \emptyset$ and $T = F$ of (11) are both slack and the threshold is determined by the two singletons:

$$\kappa^K = \min(\pi(K_{f_1}) + p_{f_2}, p_{f_1} + \pi(K_{f_2})) = \min(0.6, 1.4) = 0.6.$$

Any $1 - \alpha > 0.6$ is not rationalizable. The bottleneck is that the negative result, the only certificate of health, occurs with probability only 0.1. When the reported sets do not exhaust S , the instance $T = \emptyset$ can itself be the binding one, as in the two-state example just above, so no reduction to singletons is available in general.

Comparison with confidence sets. Confidence rationalizability needs $(1 - \alpha) \sum_{f \in T} p_f \leq \pi(\cup_{f \in T} C_f)$ for every T , a proportionality between coverage and feature mass at every subset. The conformal condition asks only that each partition’s combined capacity reach $1 - \alpha$, a weaker aggregate. The practical difference is pooling: a common feature with a large set compensates for a rare feature with a small one. Conformal rationalizability fails only when pooling cannot, because rare features uniquely cover states of significant π -mass. The conformal literature strengthens marginal coverage to hold within subgroups—demographic strata or other partitions of the feature space—often on fairness grounds, so that the average is not met by under-covering some groups. But if group membership is observable at prediction time, as it must be to state or enforce such a guarantee, it can be folded into the feature vector. This gives the following.

COROLLARY 5.1 (GROUP-CONDITIONAL COVERAGE) *Let G index groups, let $\tilde{F} = F \times G$, and suppose the reported sets satisfy coverage at level $1 - \alpha$ within each cell (f, g) . If the decision maker observes the cell frequencies $\{p_{(f,g)}\}$ and the state frequencies $\{\pi_i\}$, then the sets are rationalizable if and only if (5) holds for every $T \subseteq \tilde{F}$.*

A group-conditional guarantee is a fully conditional one on the enriched space, which is the confidence constraint (3) on $\tilde{F}$, so Theorem 4.1 applies verbatim with $\tilde{F}$ in place of F . The hypothesis on the observables is not automatic: if the decision maker observes only p_f and the group distribution, the cell frequencies $p_{(f,g)}$ are themselves only partially identified and the enriched check is not directly available.

The literature keeps the two guarantees distinct for statistical, not conceptual, reasons. Standard conformal prediction attains marginal coverage in finite samples by exchangeability of the calibration data. Conditional coverage—even on group alone—is not attainable exactly without further assumptions, since each group’s conditional distribution must be estimated from that group’s data. Group-conditional coverage is an intermediate target: reachable with moderate per-group samples, while fully conditional coverage may need prohibitive per-cell counts. This is a finite-sample point, not a difference in kind. Our framework applies to whatever guarantee is reported, at any coarseness; see Zargarbashi et al. [2024], Gauthier et al. [2025], and Jung et al. [2023] for recent constructions.

6 From Sets to Decisions

Confidence and conformal sets estimate where a parameter may lie. If a decision’s payoff depends on that parameter, how should the set inform the choice? The preceding sections settled the first of our two questions—when the as-if interpretation is coherent, as encoded in Theorems 4.1 and 5.1. We now turn to the second: given rationalizable sets, how should they drive decisions? The main result is Theorem 7.1: the rationalizable-as-if rule converges to Bayesian expected-loss minimization as ambiguity shrinks, while the as-if rule does not, however much information arrives. Recall that the as-if rule treats the set estimate as accurate. Let A be the action set and $L(i, a) \in [0, 1]$ the loss of action a in state i . Given feature f and confidence set C_f , the as-if rule chooses

$$a_f \in \arg \min_{a \in A} \max_{i \in C_f} L(i, a).$$

Minimax here is justified by ambiguity aversion [Gilboa and Schmeidler, 1989]. With $R_f = \min_a \max_{i \in C_f} L(i, a)$,

$$\Pr(L(i, a_f) \leq R_f \mid f) \geq \Pr(i \in C_f \mid f) \geq 1 - \alpha.$$

Replacing C_f with K_f and R_f with $R_f^c = \min_a \max_{i \in K_f} L(i, a)$ gives the marginal (not conditional) guarantee $\Pr(L(i, a_f^c) \leq R_f^c) \geq 1 - \alpha$, consistent with conformal sets' weaker, average-case coverage.

The two rules as two credibility levels. It is worth naming the three decision makers in the vocabulary of Section 2 before comparing them, because the comparison is then almost self-evident. Each reads C_f as a credible set. They differ in the credibility they assign and in whether they retain the prior that licensed the reading.

- The *as-if* decision maker reads C_f as a credible set at credibility *one*: $\Pr(i \in C_f \mid f) = 1$. She uses the level $1 - \alpha$ to establish that the reading is licensed and then discards it when she acts. She also discards π , the prior that made the licensing question meaningful.
- The *data-dependent-as-if* decision maker reads C_f as a credible set at credibility $1 - \alpha$, but discards π .
- The *rationalizable-as-if* decision maker reads C_f as a credible set at credibility $1 - \alpha$ and keeps π . Her ambiguity set is the set of all Bayes models under which the reading is valid, which is $\mathcal{X}$ by Corollary 4.1.

Put this way, the difficulty with the as-if rule needs no theorem to see. The level at which the sets were certified is not the level at which they are used, and the prior against which they were certified plays no part in the decision. What requires a theorem, namely Theorem 7.1, is that this is not merely inelegant but forecloses convergence to the Bayes rule under any accumulation of information.

As-if decisions treat states outside C_f as impossible given f . In terms of $\mathcal{X}$, this replaces the conditional coverage constraint (3) with the equality $\Pr(i \in C_f \mid f) = 1$. For conformal sets, it replaces (6) with the requirement that all mass lie within $\cup_f K_f$. Either distortion forgoes hedging by randomization and, as we show, discards information. Two papers study minimax rules for these sets: Andrews and Chen [2025] for confidence sets and Kiyani et al. [2025] for conformal sets. Neither fixes the marginals, so each optimizes over a set larger than $\mathcal{X}$; the as-if rule imposes $\Pr(i \in C_f \mid f) = 1$. These are modeling choices, not errors. We examine each in turn.

6.1 Justification for Confidence Sets

Andrews and Chen [2025] introduce a default action of loss D in every state and restrict the decision maker at each f to a_f or the default. Let q_f be the probability of a_f . An auxiliary problem over all joint distributions consistent with the sets, under a prior μ , determines q_f and bounds the downside risk of the as-if rule. Their problem (P1) is:

$$\begin{aligned} \min_{(q_f)_{f \in F}} \quad & E_{x \sim \mu} \left\{ \sum_{f \in F} \sum_{i \in S} x(i, f) [q_f L(i, a_f) + (1 - q_f) D] \right\} \\ \text{s.t.} \quad & \max_{x \in X, i \in S} \sum_{f \in F} x(i, f) [q_f L(i, a_f) + (1 - q_f) D] \leq D + \tau \\ & q_f \in [0, 1] \quad \forall f \in F \end{aligned}$$

for $\tau \in (0, 1 - D]$. The set X is all joint distributions consistent with conditional coverage (3) and summing to one, without fixed marginals—strictly larger than $\mathcal{X}$. Andrews and Chen do not use the marginal information that establishes rationalizability. For a suitable μ , they show that if $\tau = u(1 - \alpha)D$ for some $u \in (0, 1)$, the optimum is $q_f = u$ when $R_f \leq D$ and zero otherwise, and for any $x \in X$:

$$\sum_f x(i, f) L(i, a_f) \leq D + u\alpha(1 - D) \quad \forall i \in S. \quad (13)$$

Three features limit the generality of this justification.

1. A default action with state-independent loss, which may not exist.
2. The restriction at each f to a_f or the default.²
3. A prior μ chosen partly to deliver the bound rather than from first principles.

It is easy to see that the first two features are the main drivers of the bound in equation (13). Given feature vector f , the decision maker can always compare $(1 - \alpha)R_f + \alpha \max_{i \notin C_f} L(i, a_f)$ with D . If the first is smaller, choose action a_f ; otherwise choose the default action. In this way, the decision maker is guaranteed to lose no more than D , and the guarantee holds for every x satisfying conditional coverage, since $(1 - \alpha)R_f + \alpha \max_{i \notin C_f} L(i, a_f)$ is the worst-case conditional expected loss of a_f at f over all such x .

If we eliminate these three features but maintain the mini-max criterion, what is the optimization problem the decision maker should solve? Adoption of the mini-max criterion

²Relaxed in the appendix of Andrews and Chen [2025].

presumes an ‘adversarial nature’; therefore, there is no reason to restrict attention to deterministic rules for selecting actions. To this end, let q_{fa} denote the probability with which action $a \in A$ is chosen at feature $f \in F$, i.e., we allow the decision maker to randomize. The decision maker solves:

$$\min_q \max_x \sum_{i \in S} \sum_{f \in F} [\sum_{a \in A} L(i, a) q_{fa}] x(i, f) \quad (14)$$

$$\text{s.t.} \quad \sum_{i \in S} \sum_{f \in F} x(i, f) = 1 \quad (15)$$

$$\sum_{i \in C_f} x(i, f) \geq (1 - \alpha) \sum_{i \in S} x(i, f) \quad \forall f \in F \quad (16)$$

$$x(i, f) \geq 0 \quad \forall i, f \quad (17)$$

$$\sum_{a \in A} q_{fa} = 1 \quad \forall f \in F \quad (18)$$

$$q_{fa} \in [0, 1] \quad \forall f \in F, a \in A \quad (19)$$

Theorem 6.1 below shows that the optimal objective function value of this program is determined, up to tie breaks, by one feature vector, call it f^* . This has implications for how one chooses confidence/conformal sets. Some papers argue that one should focus on statistical methods that shrink the expected size of these sets (see for example [Dhillon et al. \[2024\]](#)). Theorem 6.1 suggests this is misleading. What matters to first order is the identity of C_{f^*} , not the average set size. A large set of states on which all actions do about equally badly is harmless; a small set straddling a sharp disagreement between actions is not.

For each $f \in F$ let

$$g_f = \max_{z \in \Delta(S)} \min_{a \in A} \sum_{i \in S} L(i, a) z_i \quad \text{s.t.} \quad \sum_{i \in C_f} z_i \geq 1 - \alpha. \quad (20)$$

This is Nature’s problem at f . She chooses a conditional distribution over states that respects the coverage guarantee and maximizes the decision maker’s loss, the decision maker best-responding. Notice that the constraint is an inequality: Nature is required to allocate mass *at least* $1 - \alpha$ to states in C_f , and mass at most α to states outside C_f .

Each choice of x by Nature induces a probability distribution over F . Hence, Nature would choose x to assign probability 1 to the feature f that maximizes the decision maker’s loss. Hence, the optimal objective function value of (14- 19) will be $\max_f g_f$. This intuition is formalized in the next result.

THEOREM 6.1 *The optimal objective function value of problem (14- 19) is $\max_{f \in F} g_f$.*

Proof. By the mini-max Theorem, problem (14-19) is equivalent to

$$\max_x \min_q \sum_{i \in S} \sum_{f \in F} [\sum_{a \in A} L(i, a) q_{fa}] x(i, f)$$

subject to (15)–(19). Write $x(i, f) = \beta_f z_i^f$ with $\beta \in \Delta(F)$ and $z^f \in \Delta(S)$; this is without loss for any f with $\beta_f > 0$, and features with $\beta_f = 0$ contribute nothing. Constraint (16) becomes $\sum_{i \in C_f} z_i^f \geq 1 - \alpha$ for each f with $\beta_f > 0$, since the factor β_f appears on both sides. The decision maker best-responds at each f , so the objective becomes

$$\sum_{f \in F} \beta_f \min_{a \in A} \sum_{i \in S} L(i, a) z_i^f,$$

and maximizing over the z^f subject to their coverage constraints gives $\sum_{f \in F} \beta_f g_f$. This is linear in β with coefficients g_f , so it is maximized over $\Delta(F)$ by placing all mass on some $f^* \in \arg \max_f g_f$. Hence the optimal value is $\max_{f \in F} g_f$. ■

Observe why the decomposition works here. The coverage guarantee is attached to each feature in proportion to its probability, so Nature’s per-feature problem is positively homogeneous in β_f and she optimally concentrates on the single worst feature. Section 6.2 shows this fails for conformal sets, where the guarantee is an aggregate budget.

The **rationalizable-as-if** decision maker solves problem (21–27), over $\mathcal{X}$, specifically:

$$\min_q \max_x \sum_{i \in S} \sum_{f \in F} [\sum_{a \in A} L(i, a) q_{fa}] x(i, f) \tag{21}$$

$$\text{s.t. } \sum_{f \in F} x(i, f) = \pi_i \quad \forall i \in S \tag{22}$$

$$\sum_{i \in S} x(i, f) = p_f \quad \forall f \in F \tag{23}$$

$$\sum_{i \in C_f} x(i, f) \geq (1 - \alpha) p_f \quad \forall f \in F \tag{24}$$

$$x(i, f) \geq 0 \quad \forall i, f \tag{25}$$

$$\sum_{a \in A} q_{fa} = 1 \quad \forall f \in F \tag{26}$$

$$q_{fa} \in [0, 1] \quad \forall f \in F, a \in A \tag{27}$$

Constraints (22)–(23) fix the marginals; (24) is conditional coverage. Nature’s feasible set is

exactly $\mathcal{X}$. This program differs from Andrews-Chen in two ways: it fixes the marginals, and it randomizes over the full action set rather than a_f or a default. Unlike the marginals-free problem, this one does not decompose across f . The failure to decompose is not a technical nuisance. It is the substance. Fixing the state marginal couples the features, because mass Nature moves into a state at one feature must come out of that state at another. Example 7.1 below turns on exactly this coupling, and it is why the rationalizable-as-if rule can differ from the as-if rule at a feature where the two ambiguity sets look identical.

COROLLARY 6.1 *The optimal objective function value of problem (21–27) is at most $\max_{f \in F} g_f$.*

Proof. Any $x \in \mathcal{X}$ satisfies $\sum_{i \in C_f} x(i, f) \geq (1 - \alpha)p_f = (1 - \alpha) \sum_{i \in S} x(i, f)$, so it is feasible for (15)–(17). Hence $\mathcal{X} \subseteq X$, the inner maximum in (21) is taken over a subset of the feasible set of (14) for each q , and the value of the min-max is no larger. Apply Theorem 6.1. ■

The gap between the two values measures what the marginals are worth to the decision maker.

6.2 Justification for Conformal Sets

Kiyani et al. [2025] justify the as-if rule for conformal sets. Given $\{K_f\}$, let

$$a_f^c \in \arg \min_{a \in A} \max_{i \in K_f} L(i, a), \quad R_f^c = \min_{a \in A} \max_{i \in K_f} L(i, a).$$

They pose a minimax problem over all joint distributions consistent with marginal coverage (6)—not conditional coverage (3). Because marginal coverage is one aggregate constraint, the set is much larger than for confidence sets, giving Nature more room. Their result: as-if decisions are optimal when the conformal sets solve a joint optimization in which the sets are themselves decision variables. To describe the joint optimization problem, encode, as before, a decision maker’s rule for selecting actions given a feature by $\{q_{fa}\}_{f \in F, a \in A}$, which is the probability of choosing action $a \in A$ given feature $f \in F$. Kiyani et al. [2025] restrict q to being deterministic. Let t_f denote a ‘target’ loss at feature $f \in F$. Then, the optimization problem Kiyani et al. [2025] introduce is the following, called (P2):

$$\begin{aligned} & \max_{q, x, t} \sum_{f \in F} \sum_{i \in S} x(i, f) t_f \\ \text{s.t.} \quad & \sum_{(i, f): \sum_{a \in A} L(i, a) q_{fa} \geq t_f} x(i, f) \geq 1 - \alpha \end{aligned}$$

$$\begin{aligned}
\sum_{i \in S} \sum_{f \in F} x(i, f) &= 1 \\
x(i, f) &\geq 0 \quad \forall i \in S, f \in F \\
\sum_{a \in A} q_{fa} &= 1 \quad \forall f \in F \\
q_{fa} &\in \{0, 1\} \quad \forall f \in F, a \in A
\end{aligned}$$

If, for each $f \in F$, we set $q_{fa_f^c} = 1$ and $t_f = R_f^c$, we obtain a feasible but not necessarily optimal solution to the optimization problem above. However, and this is the key result in [Kiyani et al. \[2025\]](#), one can reformulate the problem with the choice of conformal sets rather than loss thresholds being the decision variables. To define this problem, let $c_{if} = 1$ if $i \in S$ is to be included in the conformal set associated with feature $f \in F$ and zero otherwise. The joint optimization problem is:

$$\begin{aligned}
&\min_{c, x, q, y} \sum_{f \in F} \sum_{i \in S} x(i, f) y_f \\
&\text{s.t. } y_f \geq c_{if} L(i, a) q_{fa} \quad \forall i \in S \quad \forall f \in F \quad \forall a \in A \\
&\sum_{f \in F} \sum_{i \in S} x(i, f) c_{if} \geq 1 - \alpha \\
&\sum_{f \in F} \sum_{i \in S} x(i, f) = 1 \\
&x(i, f) \geq 0 \quad \forall i \in S, f \in F \\
&\sum_{a \in A} q_{fa} = 1 \quad \forall f \in F \\
&q_{fa} \in \{0, 1\} \quad \forall f \in F, a \in A \\
&c_{if} \in \{0, 1\} \quad \forall i \in S, f \in F
\end{aligned}$$

Observe that for each $f \in F$, $y_f = \min_a \max_{i: c_{if}=1} L(i, a)$. Hence, the as-if decision rule is justified provided one has chosen the conformal sets in the ‘right’ way. We find this justification unattractive because it tailors the choice of conformal sets to the loss function of the decision maker. If the analyst has access to the decision maker’s loss function, why not conduct the analysis directly on the loss function? Typically, the output of a statistical analysis is not customized to a particular decision maker. So, let us suppose this. As before, we denote by q_{fa} the probability with which action $a \in A$ is chosen at feature $f \in F$. Then,

in keeping with the mini-max spirit, the decision maker should solve:

$$\min_q \max_x \sum_{i \in S} \sum_{f \in F} [\sum_{a \in A} L(i, a) q_{fa}] x(i, f) \quad (28)$$

$$\text{s.t.} \quad \sum_{i \in S} \sum_{f \in F} x(i, f) = 1 \quad (29)$$

$$\sum_{f \in F} \sum_{i \in K_f} x(i, f) \geq 1 - \alpha \quad (30)$$

$$x(i, f) \geq 0 \quad \forall i, f \quad (31)$$

$$\sum_{a \in A} q_{fa} = 1 \quad \forall f \in F \quad (32)$$

$$q_{fa} \in [0, 1] \quad \forall f \in F, a \in A \quad (33)$$

If $\cup_{f \in F} K_f = S$, then constraint (30) is redundant and optimization problem (28-33) is a standard mini-max problem for a zero-sum game. The decision maker has a pure strategy for each (f, a) pair and Nature has one pure strategy for each (i, f) pair. Now, focus on the case $\cup_{f \in F} K_f \subset S$. One might expect, by analogy with Theorem 6.1, that the optimal objective function value of problem (28-33) is again determined by a single feature. It is not, and the reason is instructive. Under conditional coverage the guarantee is attached to each feature in proportion to its probability. Under marginal coverage the guarantee is a single transferable budget, and Nature is free to spend it unevenly: she buys coverage at the features where coverage is cheap for her and economizes where it is dear.

To state the correct result, for $c \in [0, 1]$ and $f \in F$ define

$$h_f(c) = \max_{z \in \Delta(S)} \min_{a \in A} \sum_{i \in S} L(i, a) z_i \quad \text{s.t.} \quad \sum_{i \in K_f} z_i \geq c, \quad (34)$$

Nature's best conditional distribution at f when she is required to place mass c inside K_f . Each h_f is concave and non-increasing on $[0, 1]$, with $h_f(0) = V := \max_{z \in \Delta(S)} \min_a \sum_i L(i, a) z_i$ common to all f . Write $\hat{g}_f = h_f(1 - \alpha)$ for the value when the full coverage requirement is imposed at f alone.

THEOREM 6.2 *Suppose $\cup_{f \in F} K_f \subset S$. Then the optimal objective function value of problem (28-33) is*

$$\max \left\{ \sum_{f \in F} \beta_f h_f\left(\frac{m_f}{\beta_f}\right) : \beta \in \Delta(F), 0 \leq m_f \leq \beta_f \quad \forall f, \sum_{f \in F} m_f \geq 1 - \alpha \right\},$$

with the convention $\beta_f h_f(m_f/\beta_f) = 0$ when $\beta_f = 0$. The maximand is concave, each summand being the perspective of a concave function, so this is a concave program. In particular $\max_f \hat{g}_f$ is a lower bound on the value, attained by concentrating all mass on one feature, and the bound is strict whenever the optimal allocation (β, m) has m_f/β_f non-constant across the features it uses.

Proof. By the mini-max Theorem, problem (28-33) is equivalent to

$$\max_x \min_q \sum_{i \in S} \sum_{f \in F} \left[\sum_{a \in A} L(i, a) q_{fa} \right] x(i, f)$$

subject to (29)–(33). Write $x(i, f) = \beta_f z_i^f$ with $\beta \in \Delta(F)$ and $z^f \in \Delta(S)$, and set $m_f = \beta_f z^f(K_f)$, the coverage mass delivered at f . Constraint (30) is $\sum_f m_f \geq 1 - \alpha$ and nothing more, and $0 \leq m_f \leq \beta_f$ by construction. The decision maker best-responds at each f , so the objective is $\sum_f \beta_f \min_a \sum_i L(i, a) z_i^f$, whose maximum over z^f given (β_f, m_f) is $\beta_f h_f(m_f/\beta_f)$. Concavity of h_f is concavity of the value of a concave program in the right-hand side of a constraint, and the perspective of a concave function is concave. ■

The single-feature reduction of Theorem 6.1 is a property of conditional coverage, not of minimax decision making with coverage-based sets, and the design implication drawn after Theorem 6.1—that the identity of C_{f^*} is what matters—does extend to conformal sets. A small example shows the gap is real. Let $S = \{1, 2, 3\}$, $F = \{f_1, f_2\}$, $K_{f_1} = \{1\}$, $K_{f_2} = \{3\}$, $\alpha = \frac{1}{4}$, $A = \{a_1, a_2\}$, $L(\cdot, a_1) = (0, 1, \frac{1}{4})$ and $L(\cdot, a_2) = (1, 0, \frac{1}{4})$, so that $\cup_f K_f = \{1, 3\} \subsetneq S$. Take $x(1, f_1) = x(2, f_1) = \frac{1}{4}$ and $x(3, f_2) = \frac{1}{2}$, all other entries zero. Marginal coverage holds with equality, $x(1, f_1) + x(3, f_2) = \frac{3}{4}$, and the decision maker’s best responses give $\frac{1}{4}$ at f_1 and $\frac{1}{8}$ at f_2 , for a total of $\frac{3}{8}$. But $\hat{g}_{f_1} = \frac{1}{4}$ and $\hat{g}_{f_2} = \frac{5}{16}$, so $\max_f \hat{g}_f = \frac{5}{16} < \frac{3}{8}$. Nature does strictly better by spreading: she buys full coverage at f_2 , where coverage costs her nothing because state 3 lies in K_{f_2} and is a state she likes, and buys only a third of the required coverage at f_1 , where coverage is expensive.

7 Prior Information

To judge whether a confidence or conformal set is useful, first ask what the decision maker would do without it. That is the benchmark for any procedure built on set estimates. How should she use prior information given confidence sets? Suppose she has the marginals $\{\pi_i\}$ and $\{p_f\}$, and let q_{fa} be the probability of action a at f . She evaluates a rule by worst-case expected loss over all $x \in \mathcal{X}$, i.e., problem (21–27). Suppose two states with

$\pi_1 - (1 - \alpha) = \delta$ for some $\delta \geq 0$, and $C_f = \{1\}$ for all f . Knowing π_1 , the decision maker learns that for each f , $\frac{x(1,f)}{x(1,f)+x(2,f)} - (1 - \alpha) = \delta_f$ for some $\delta_f \geq 0$. Ignoring π_1 , the as-if maker takes an action in $\arg \min_a L(1, a)$ regardless of δ and f . The rationalizable-as-if maker, though ignorant of $x(1, f)$, knows the p_f -weighted δ_f sum to δ , that is $\sum_f p_f \delta_f = \delta$, since $\pi_1 = \sum_f x(1, f) = \sum_f p_f [(1 - \alpha) + \delta_f]$. To find the optimal strategy $\{b_f\}$, she solves:

$$\begin{aligned} \min_{\{b_f\}} \max_{(\delta_f)} \sum_f p_f [(1 - \alpha)L(1, b_f) + \alpha L(2, b_f) + \delta_f (L(1, b_f) - L(2, b_f))] \\ \text{s.t. } \sum_f p_f \delta_f = \delta, \quad \delta_f \geq 0 \quad \forall f. \end{aligned}$$

The conditional probabilities at f are $(1 - \alpha) + \delta_f$ on state 1 and $\alpha - \delta_f$ on state 2, which is where the δ_f term comes from.

When δ is large, $1 - \alpha$ is small, and the sets carry little. As δ shrinks, the sets and the prior reinforce each other. For fixed L , finite A , and any F , there is $\delta^* > 0$ such that for $\delta < \delta^*$ the rationalizable-as-if maker chooses what a Bayesian with full knowledge of x would choose: when δ is small each δ_f is small, so the conditional probability of state 1 at f , $\frac{x(1,f)}{x(1,f)+x(2,f)} \approx 1 - \alpha \approx \pi_1$, and with A finite the same action is optimal under either. This presumes that the Bayes-optimal action at the belief $(\pi_1, 1 - \pi_1)$ is unique, which holds for all π_1 outside a finite set given finite A , since ties occur only where two of the $|A|$ affine functions $\pi_1 L(1, a) + (1 - \pi_1)L(2, a)$ cross. The as-if maker stays at $\arg \min_a L(1, a)$.

This shows the cost of the distortion from $\mathcal{X}$ to $\tilde{\mathcal{X}}$. The rationalizable-as-if maker works within $\mathcal{X}$, which absorbs new information without contradiction. The as-if maker works within $\tilde{\mathcal{X}}$, and prior information I_0 consistent with $\mathcal{X}$ may be inconsistent with $\tilde{\mathcal{X}}$, forcing her to drop it. The trouble is not that she chooses to ignore I_0 ; her distortion of coverage— $\Pr(i \in C_f \mid f) = 1$ in place of $\geq 1 - \alpha$ —leaves her unable to use it.

Example 7.1 (Divergence of the two rules) Let $S = \{1, 2\}$, $F = \{f_1, f_2\}$, $\alpha = 1/2$, $\pi_1 = 3/4$, $\pi_2 = 1/4$, $p_{f_1} = p_{f_2} = 1/2$, with $C_{f_1} = \{1\}$ and $C_{f_2} = \{1, 2\}$: only state 1 is plausible at f_1 , both at f_2 . The action set is $A = \{a, b\}$ with

$$L(1, a) = 0, \quad L(2, a) = 1, \quad L(1, b) = 0.6, \quad L(2, b) = 0.$$

The ambiguity set. $\mathcal{X}$ is one-dimensional. Writing $s = x(2, f_1)$,

$$x(1, f_1) = \frac{1}{2} - s, \quad x(1, f_2) = \frac{1}{4} + s, \quad x(2, f_2) = \frac{1}{4} - s,$$

and the constraints $x \geq 0$ together with coverage at f_1 , $x(1, f_1) \geq (1 - \alpha)p_{f_1} = \frac{1}{4}$, give exactly

$s \in [0, \frac{1}{4}]$. Coverage at f_2 is automatic because $C_{f_2} = S$. Observe what s is: the location of state 2's marginal mass $\pi_2 = \frac{1}{4}$. Nature chooses how much of it to place at f_1 and how much at f_2 , and cannot place more in total than π_2 . This is the coupling the as-if maker cannot see.

As-if rule. At f_1 she treats $C_{f_1} = \{1\}$ as exhaustive and minimizes $L(1, \cdot)$, choosing a with $R_{f_1} = 0$. At f_2 she minimizes $\max_i L(i, \cdot) = \min(1, 0.6) = 0.6$, choosing b . Her rule is (a, b) , and her expected loss at x is $s + 0.6(\frac{1}{4} + s) = 0.15 + 1.6s$, so her worst case over $\mathcal{X}$ is 0.55, attained at $s = \frac{1}{4}$.

Rationalizable-as-if rule. The worst-case losses of the four deterministic rules are

$$(a, a) : s + (\frac{1}{4} - s) = \frac{1}{4}; \quad (a, b) : 0.15 + 1.6s \leq 0.55; \quad (b, a) : 0.55 - 1.6s \leq 0.55; \quad (b, b) : 0.45.$$

Randomization does not improve on $\frac{1}{4}$: with q_1, q_2 the probabilities of a at f_1 and f_2 , the expected loss is $0.45 - 0.3q_1 + 0.1q_2 + 1.6(q_1 - q_2)s$, whose maximum over $s \in [0, \frac{1}{4}]$ is $0.45 - 0.3q_1 + 0.1q_2 + 0.4(q_1 - q_2)^+$, minimized at $q_1 = q_2 = 1$ with value $\frac{1}{4}$. So the rationalizable-as-if maker plays a at both features, with worst-case loss $\frac{1}{4}$, less than half the as-if rule's 0.55.

Why. Playing a everywhere is a perfect hedge against Nature's only choice. Since $L(\cdot, a)$ is the indicator of state 2, the rule (a, a) incurs $x(2, f_1) + x(2, f_2) = \pi_2 = \frac{1}{4}$ whatever s is. The state marginal, which the rationalizable-as-if maker knows, pins her loss down exactly. The as-if maker discards π , so at f_2 , where the reported set is uninformative, she has nothing to hedge with and buys insurance at price 0.6 against a state that the marginal has already capped at $\frac{1}{4}$. The rules diverge at f_2 , the feature at which the two confidence sets are identical. The divergence comes entirely from the marginal, not from the sets.

REMARK 6 *The example demonstrates that one cannot diagnose the as-if rule feature by feature, because $\mathcal{X}$ ties the features together through π . An as-if action can be optimal at every feature considered in isolation and still be dominated overall.*

7.1 A Formalization

We restrict to information that bounds ensembles of joint probabilities. A **statistical information set** (SIS) I is a collection $I = \{(A_n, b_n)\}_{n \in N}$ with $A_n \subset S \times F$, $b_n \in [0, 1]$, and

$$X(I) := \{x \in \Delta(S \times F) : \sum_{(i,f) \in A_n} x(i, f) \geq b_n \ \forall n \in N\} \neq \emptyset.$$

Restricting to the simplex matters. Without $x \geq 0$ and $\sum_{i,f} x(i, f) = 1$ the set $X(I)$ is an unbounded polyhedron, and the compactness invoked in the proof of Theorem 7.1 is

unavailable. Confidence and conformal sets are SISs. Suppose the decision maker holds an SIS $I_0 = \{(A_{0,n}, b_{0,n})\}_{n \in N_0}$ and then receives confidence or conformal sets I_C . The marginal constraints $\sum_f x(i, f) = \pi_i$ and $\sum_i x(i, f) = p_f$ are themselves SIS constraints in I_0 when encoded as two opposed inequalities.³ For confidence sets, I_C is the conditional coverage constraints (3); for conformal sets, the single constraint (6).

We compare three decision makers. For expositional convenience we describe them here with deterministic a_f ; Theorem 7.1 below is proved for randomized rules q_{fa} , which contain the deterministic ones as a special case. The rationalizable-as-if decision maker uses both I_0 and the full I_C :

$$\min_{(a_f)} \max_{x \in X(I_0 \cup I_C)} \sum_{(i,f)} L(i, a_f) x(i, f).$$

The **data-dependent-as-if decision maker** ignores I_0 but keeps the full I_C :

$$\min_{(a_f)} \max_{x \in X(I_C)} \sum_{(i,f)} L(i, a_f) x(i, f).$$

The as-if decision maker ignores I_0 and distorts I_C to $\tilde{I}_C$: for confidence sets she replaces $\Pr(i \in C_f \mid f) \geq 1 - \alpha$ with $\Pr(i \in C_f \mid f) = 1$; for conformal sets she requires all mass within $\cup_f K_f$. She solves

$$\min_{(a_f)} \max_{x \in X(\tilde{I}_C)} \sum_{(i,f)} L(i, a_f) x(i, f).$$

7.1.1 Convergence to expected loss

We measure the ambiguity of an SIS by the size of $X(I)$ in $\mathbb{R}^{|S| \times |F|}$, denoted $|X(I)|$: the sum over (i, f) of the length of the projection of $X(I)$ onto $x(i, f)$, normalized by $|S||F|$. Zero size means I pins down the distribution. The **degree of complementarity** of I^1 and I^2 is $1 - |X(I^1 \cup I^2)|$; they are **perfect complements** if $|X(I^1 \cup I^2)| = 0$. This measure implies the hypothesis of the theorem below. If $x^* \in X(I)$ then $\max_{x \in X(I)} \|x - x^*\|_1 \leq |S||F| \cdot |X(I)|$, since each coordinate difference is at most the length of the corresponding projection. So $|X(I_t \cup I_C)| \rightarrow 0$ implies (35).

THEOREM 7.1 *Let A be finite and $L : S \times A \rightarrow [0, 1]$. Let x^* be the true joint distribution with $x^* \in X(I_C)$, and $\{I_t\}_{t \geq 1}$ a sequence of SISs each consistent with x^* , with $X(I_t \cup I_C)$ non-empty and*

$$\max_{x \in X(I_t \cup I_C)} \|x - x^*\|_1 \rightarrow 0 \quad \text{as } t \rightarrow \infty. \quad (35)$$

³An SIS as defined imposes only lower bounds $\sum_{A_n} x \geq b_n$, so encoding an equality requires the complementary bound $\sum_{A_n^c} x \geq 1 - b_n$. This is equivalent to an upper bound on $\sum_{A_n} x$ only on the simplex, which is a further reason to include $x \in \Delta(S \times F)$ in the definition.

Let q^t be optimal for the rationalizable-as-if maker given $I_t \cup I_C$, with value $V(I_t \cup I_C)$, and let $V(\{x^*\}) = \min_q \sum_{i,f} [\sum_{a \in A} L(i, a) q_{fa}] x^*(i, f)$. Then:

1. $V(I_t \cup I_C) \rightarrow V(\{x^*\})$.
2. For all $\epsilon > 0$ there is T with $\sum_{i,f} [\sum_{a \in A} L(i, a) q_{fa}^t] x^*(i, f) \leq V(\{x^*\}) + \epsilon$ for all $t \geq T$.
3. If the Bayesian optimal rule q^* is unique, then $q^t \rightarrow q^*$: for all $\epsilon > 0$ there is T such that $\|q_f^t - q_f^*\| \leq \epsilon$ for all $f \in F$ and all $t \geq T$.⁴

The as-if maker's rule is independent of I_t and so does not converge to q^* unless the as-if action already coincides with the Bayesian optimum.

Proof. Let $\phi(x, q) = \sum_{i,f} [\sum_{a \in A} L(i, a) q_{fa}] x(i, f)$, linear in each argument. Both the distributions and the mixed rules are compact and convex, so the mini-max theorem gives

$$V(I_t \cup I_C) = \min_q \max_{x \in X(I_t \cup I_C)} \phi(x, q) = \max_{x \in X(I_t \cup I_C)} \psi(x),$$

with $\psi(x) = \min_q \phi(x, q)$. As a minimum of finitely many linear functions, ψ is concave and 1-Lipschitz in ℓ_1 : $|\psi(x) - \psi(x')| \leq \|x - x'\|_1$ since $L \in [0, 1]$.

Part 1. Since $x^* \in X(I_t \cup I_C)$, $V(I_t \cup I_C) \geq \psi(x^*) = V(\{x^*\})$. For the upper bound, let x_t attain the max; then $V(I_t \cup I_C) = \psi(x_t) \leq \psi(x^*) + \|x_t - x^*\|_1 \leq V(\{x^*\}) + \max_x \|x - x^*\|_1 \rightarrow V(\{x^*\})$ by (35). *Part 2.* With q^t optimal and $x^* \in X(I_t \cup I_C)$, $\phi(x^*, q^t) \leq \max_x \phi(x, q^t) = V(I_t \cup I_C)$. By Part 1 this is $\leq V(\{x^*\}) + \epsilon$ for large t . *Part 3.* If q^* uniquely minimizes $\phi(x^*, \cdot)$, then by compactness, for any $\epsilon' > 0$ there is $\eta > 0$ with $\phi(x^*, q) \leq V(\{x^*\}) + \eta$ forcing q within ϵ' of q^* . Part 2 with $\epsilon = \eta$ gives $q^t \rightarrow q^*$. *As-if maker.* Her $a_f \in \arg \min_a \max_{i \in C_f} L(i, a)$ depends only on C_f and L , so her loss $\sum_{i,f} L(i, a_f) x^*(i, f)$ is fixed in t . It equals $V(\{x^*\})$ only if a_f is Bayes-optimal at x^* for every f with $p_f > 0$, a coincidence which the as-if rule cannot bring about by absorbing information, since it does not depend on I_t .

REMARK 7 *The Bernstein-von Mises theorem shows that data accumulate, the posterior concentrates and credible sets acquire frequentist coverage. Theorem 7.1 runs the other way. The frequentist object is given first, the decision maker interprets it as a credible set, and the question is whether the rule she builds acquires Bayesian optimality as her ambiguity set collapses. The rationalizable-as-if rule does. The as-if rule cannot, because a credibility of one is not the limit of anything: her ambiguity set does not shrink toward $\{x^*\}$, it sits at a*

⁴The space of mixed rules is finite-dimensional, so the choice of norm is immaterial.

fixed distortion of it. The requirement is thus the decision-theoretic counterpart of asking a posterior to be well calibrated. Not that the rule be correct at any finite t , but that nothing prevent it from becoming correct.

Hypothesis (35) formalizes perfect complementarity of I_t and I_C in the limit. It is weaker than $X(I_t \cup I_C) = \{x^*\}$ at any finite t , asking only that the ℓ_1 diameter shrink to zero. The two-state example meets it: as $\delta \rightarrow 0$ the coverage slack vanishes and the lone element of $X(I_t \cup I_C)$ is x^* . What it does not settle is when (35) holds for a given $\{I_t\}$ —a question about the design of I_t , not the decision problem. With two states, fixing π_1 via I_0 and letting the slack vanish via I_C suffices, since marginal and coverage then determine the joint distribution. With more states, knowing $\{\pi_i\}$ does not pin it down even as the slack vanishes: one needs further SISs on the cross-terms $x(i, f)$, and the rate depends on the geometry of those constraints relative to the decision-relevant directions. One sufficient condition is easy to state. Suppose I_t pins each sum $\sum_{f \in G} x(i, f)$ to within ϵ_t , for every $i \in S$ and every cell G of a partition $\mathcal{G}_t$ of F , and suppose $\mathcal{G}_t$ refines to singletons with $\epsilon_t \rightarrow 0$. Then for large t every coordinate $x(i, f)$ is pinned to within ϵ_t , so $\max_{x \in X(I_t \cup I_C)} \|x - x^*\|_1 \leq |S||F|\epsilon_t \rightarrow 0$ and (35) holds. General necessary and sufficient conditions on $\{I_t\}$ are left for future work. The uniqueness in Part 3 deserves a word. If the Bayesian optimum is not unique, individual rules need not converge, though the value and the set of near-optimal actions do—standard in minimax theory, and not a serious limitation, since a small perturbation of x^* restores uniqueness. Finally, the data-dependent-as-if maker, who ignores I_t but keeps I_C , has fixed value $V(I_C)$ and converges only if $X(I_C) = \{x^*\}$ already, so she too fails. The as-if maker compounds this: her distortion of I_C and her discard of I_t are not separable. Replacing $\Pr(i \in C_f \mid f) \geq 1 - \alpha$ with $\Pr(i \in C_f \mid f) = 1$ (or, for conformal sets, confining mass to $\cup_f K_f$) forces her to drop other information, since while $x^* \in X(I_t \cup I_C)$ it can happen that $X(I_t \cup \tilde{I}_C) = \emptyset$. The root error—zero probability on excluded states—forecloses convergence to expected-loss minimization.⁵

8 Conclusion

This paper studies the decision-theoretic consequences of the as-if interpretation of confidence and conformal sets: treating the realized set as a credible set at credibility one, so that all

⁵Here is an instance, which also shows that distorting I_t symmetrically does not help. Let $S = \{1, 2\}$, $F = \{f_1, f_2\}$, $\alpha = 1/2$, $C_{f_1} = C_{f_2} = \{1\}$, and let I_t carry the marginals $\pi = p = (1/2, 1/2)$. Then $x(i, f) = 1/4$ for all (i, f) lies in $X(I_t \cup I_C)$. Under $\tilde{I}_C$ we need $x(2, f) = 0$ at both features, whence $\sum_f x(2, f) = 0 < \pi_2$, so $X(I_t \cup \tilde{I}_C) = \emptyset$. Raising the bounds in I_t to one in the same way requires $\sum_f x(1, f) = \sum_f x(2, f) = 1$, which is impossible on the simplex, so $X(\tilde{I}_C \cup \tilde{I}_t) = \emptyset$ as well.

probability lies within it and none outside. This is not what confidence sets mean in the frequentist tradition, but it is the implicit model behind mini-max rules that treat the set as containing all plausible states. The paper is a study of that interpretation, not a critique of frequentist methods. The first contribution is a rationalizability criterion: given observed marginals and a reported set, does a joint distribution over state-feature pairs validate the as-if interpretation? Equivalently, and this is the form we would now emphasize: is there a prior consistent with the observed state frequencies under which every reported confidence set is a $1 - \alpha$ credible set? Theorem 4.1 answers it with a finite family of linear inequalities, and Corollary 4.2 converts the answer into the highest credibility the data will bear. The second is the rationalizable-as-if rule: minimize worst-case expected loss over the full set $\mathcal{X}$, using both the marginals and the coverage constraint. This places the decision maker in a smaller ambiguity set than Andrews and Chen [2025] and Kiyani et al. [2025], who do not fix marginals. The central justification is Theorem 7.1: as ambiguity shrinks—as further information complements the sets until $\mathcal{X}$ collapses toward a singleton—the robust rule converges to Bayesian expected-loss minimization. The as-if rule does not, however much information arrives, because its distortion of coverage leaves it unable to absorb that information. This consistency under information refinement is the right benchmark: a rule that does not become Bayesian as the ambiguity set collapses has a structural deficiency no information can remedy.

REMARK 8 (EXTENSION TO CONTINUOUS SPACES) *We restrict to finite state and feature spaces, which lets us invoke the max-flow min-cut theorem and total unimodularity in the feasibility arguments, and the mini-max theorem for finite-dimensional programs. Continuous spaces would need discretization or infinite-dimensional duality, and the mini-max theorem would have to be replaced by results for infinite-action games. We expect the feasibility half to survive. Condition (5) is a Gale–Hoffman deficiency condition, and its continuous analogue is Strassen [1965] on the existence of a coupling with prescribed marginals satisfying a set-valued constraint.*

Acknowledgements

We thank Aaron Roth who identified errors in earlier versions of the paper.

References

- S. N. Afriat. On a system of inequalities in demand analysis: An extension of the classical method. *International Economic Review*, 14(2):460–472, June 1973.

- Isaiah Andrews and Jiafeng Chen. Certified decisions, 2025. URL <https://arxiv.org/abs/2502.17830>.
- Michael Scott Balch, Ryan Martin, and Scott Ferson. Satellite conjunction analysis and the false confidence theorem. *Proceedings of the Royal Society A*, 475(2227):20180565, 2019.
- Arie Beresteanu. Nonparametric analysis of cost complementarities in the telecommunications industry. volume 38, pages 523–549. 2006.
- Arie Beresteanu and Francesca Molinari. Asymptotic properties for a class of partially identified models. *Econometrica*, 76(4):763–814, 2008.
- James O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, New York, 2nd edition, 1985.
- Robert J. Buehler. Some validity criteria for statistical inferences. *Annals of Mathematical Statistics*, 30(4):845–863, 1959.
- Christopher P. Chambers and Federico Echenique. *Revealed Preference Theory*. Number 56 in Econometric Society Monographs. Cambridge University Press, 2016. ISBN 9781107458116.
- Victor Chernozhukov, Sokbae Lee, and Adam M. Rosen. Intersection bounds: Estimation and inference. *Econometrica*, 81(2):667–737, 2013.
- Gauri Sankar Datta and Rahul Mukerjee. *Probability Matching Priors: Higher Order Asymptotics*, volume 178 of *Lecture Notes in Statistics*. Springer, New York, 2004.
- Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010. doi: 10.1287/opre.1090.0741.
- Guneet S. Dhillon, George Deligiannidis, and Tom Rainforth. On the expected size of conformal prediction sets, 2024. URL <https://arxiv.org/abs/2306.07254>.
- Jacqueline Dinnes, Jonathan J. Deeks, Sarah Berhane, Melissa Taylor, Adriano Adriano, Clare Davenport, Sabine Dittrich, Devy Emperador, Yemisi Takwoingi, Jane Cunningham, Sabine Beese, Julie Domen, Janine Dretzke, Lavinia Ferrante di Ruffano, Iain M. Harris, Malcolm J. Price, Sian Taylor-Phillips, Lotty Hooft, Mariska M. G. Leeflang, Matthew D. F. McInnes, René Spijker, Ann Van den Bruel, and Cochrane COVID-19 Diagnostic Test Accuracy Group. Rapid, point-of-care antigen tests for diagnosis of SARS-CoV-2

- infection. *Cochrane Database of Systematic Reviews*, (7):CD013705, 2022. doi: 10.1002/14651858.CD013705.pub3.
- D. A. S. Fraser. Is Bayes posterior just quick and dirty confidence? *Statistical Science*, 26(3):299–316, 2011.
- David Freedman. On the Bernstein–von Mises theorem with infinite-dimensional parameters. *Annals of Statistics*, 27(4):1119–1140, 1999.
- Etienne Gauthier, Francis Bach, and Michael I. Jordan. Backward conformal prediction, 2025. URL <https://arxiv.org/abs/2505.13732>.
- Raffaella Giacomini and Toru Kitagawa. Robust Bayesian inference for set-identified models. *Econometrica*, 89(4):1519–1556, 2021.
- Itzhak Gilboa and David Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153, 1989.
- Andrew V. Goldberg and Robert E. Tarjan. A new approach to the maximum-flow problem. *Journal of the ACM*, 35(4):921–940, 1988. doi: 10.1145/48014.61051.
- Lars Peter Hansen and Thomas J. Sargent. *Robustness*. Princeton University Press, Princeton, 2008.
- David Heath and William Sudderth. On finitely additive priors, coherence, and extended admissibility. *Annals of Statistics*, 6(2):333–345, 1978.
- Peter D. Hoff. Bayes-optimal prediction with frequentist coverage control. *Bernoulli*, 29(2):901–928, 2023.
- Peter J. Huber and Elvezio M. Ronchetti. *Robust Statistics*. Wiley, Hoboken, New Jersey, 2nd edition, 2009.
- Christopher Jung, Georgy Noarov, Ramya Ramalingam, and Aaron Roth. Batch multivalid conformal prediction. In *International Conference on Learning Representations*, 2023. URL <https://openreview.net/forum?id=Dk7QQp8jHEo>.
- Shayan Kiyani, George Pappas, Aaron Roth, and Hamed Hassani. Decision theoretic foundations for conformal prediction: Optimal uncertainty quantification for risk-averse agents, 2025. URL <https://arxiv.org/abs/2502.02561>.
- Brendan Kline and Elie Tamer. Bayesian inference in a class of partially identified models. *Quantitative Economics*, 7(2):329–366, 2016.

- Daniel Kuhn, Peyman Mohajerin Esfahani, Viet Anh Nguyen, and Soroosh Shafieezadeh-Abadeh. Wasserstein distributionally robust optimization: Theory and applications in machine learning. In Dimitris Bertsimas and Vishal Gupta, editors, *Operations Research & Management Science in the Age of Analytics*, pages 130–166. INFORMS, 2019. doi: 10.1287/educ.2018.0198.
- Charles F. Manski. *Partial Identification of Probability Distributions*. Springer, New York, 2003.
- Charles F. Manski. Statistical treatment rules for heterogeneous populations. *Econometrica*, 72(4):1221–1246, 2004.
- Charles F. Manski. Econometrics for decision making: Building foundations sketched by haavelmo and wald. *Econometrica*, 89(6):2827–2853, 2021.
- Ilya Molchanov and Francesca Molinari. *Random Sets in Econometrics*. Cambridge University Press, Cambridge, 2018.
- Hyungsik Roger Moon and Frank Schorfheide. Bayesian and frequentist inference in partially identified models. *Econometrica*, 80(2):755–782, 2012.
- Ulrich K. Müller and Andriy Norets. Credibility of confidence sets in nonstandard econometric problems. *Econometrica*, 84(6):2183–2213, 2016.
- John W. Pratt. Length of confidence intervals. *Journal of the American Statistical Association*, 56(295):549–567, 1961.
- Hamed Rahimian and Sanjay Mehrotra. Distributionally robust optimization: A review. *arXiv preprint arXiv:1908.05659*, 2019.
- G. K. Robinson. Conservative statistical inference. *Journal of the Royal Statistical Society, Series B*, 39(3):381–386, 1977.
- Jörg Stoye. Minimax regret treatment choice with finite samples. *Journal of Econometrics*, 166(1):138–156, 2011.
- Volker Strassen. The existence of probability measures with given marginals. *Annals of Mathematical Statistics*, 36(2):423–439, 1965.
- Botond Szabó, A. W. van der Vaart, and J. H. van Zanten. Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Annals of Statistics*, 43(4):1391–1428, 2015.

- A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998.
- Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.
- Larry Wasserman and Joseph B. Kadane. Symmetric upper probabilities. *Annals of Statistics*, 20(4):1720–1736, 1992.
- B. L. Welch and H. W. Peers. On formulae for confidence points based on integrals of weighted likelihoods. *Journal of the Royal Statistical Society, Series B*, 25(2):318–329, 1963.
- Wolfram Wiesemann, Daniel Kuhn, and Melvyn Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014. doi: 10.1287/opre.2014.1314.
- Soroush H. Zargarbashi, Simone Antonelli, and Aleksandar Bojchevski. Robust yet efficient conformal prediction sets. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, volume 235 of *Proceedings of Machine Learning Research*, pages 58452–58479, 2024.

A Maximum Flow Computation for the Rationalizability Check

Feasibility of system (1–4) can be settled by one maximum flow problem on a network polynomial in $|S|$ and $|F|$. Recall the condition of Theorem 4.1: feasible iff for all $T \subseteq F$,

$$(1 - \alpha) \sum_{f \in T} p_f \leq \sum_{i \in \cup_{f \in T} C_f} \pi_i. \quad (36)$$

We build a network whose maximum flow equals $1 - \alpha$ exactly when this holds for every T .

The network. Let $\mathcal{N} = (V_{\mathcal{N}}, E_{\mathcal{N}})$ have a source s , a sink t , a node v_f for each $f \in F$, and a node w_i for each $i \in \cup_{f \in F} C_f$. Edges: $s \rightarrow v_f$ with capacity $(1 - \alpha)p_f$; $v_f \rightarrow w_i$ with capacity $+\infty$ for $i \in C_f$; and $w_i \rightarrow t$ with capacity π_i . The network has $O(|S| + |F|)$ nodes and $O(|S||F|)$ edges, and total source capacity $\sum_f (1 - \alpha)p_f = 1 - \alpha$.

Proposition A.1 *The minimum s - t cut in $\mathcal{N}$ has capacity*

$$\min_{T \subseteq F} \left[(1 - \alpha) \sum_{f \in T} p_f + \sum_{i \in \cup_{f \notin T} C_f} \pi_i \right].$$

Proof. For a cut $(S_{\text{cut}}, T_{\text{cut}})$ set $T = \{f : v_f \in T_{\text{cut}}\}$. A finite cut must keep $w_i \in S_{\text{cut}}$ whenever some $f \notin T$ has $i \in C_f$, since $v_f \rightarrow w_i$ is infinite and $v_f \in S_{\text{cut}}$. The finite crossing edges are $s \rightarrow v_f$ for $f \in T$ (capacity $(1 - \alpha)p_f$) and $w_i \rightarrow t$ for $i \in \cup_{f \notin T} C_f$ (capacity π_i). Every T gives such a cut, and conversely, so the minimum cut is as stated.

States $i \notin \cup_{f \in F} C_f$ receive no node in $\mathcal{N}$; they are irrelevant to feasibility since they appear in neither side of condition (36) for any T .

By max-flow min-cut, $\text{MF}(\mathcal{N})$ equals this minimum.

THEOREM A.1 *System (1-4) is feasible if and only if the maximum flow saturates every source edge,*

$$\text{MF}(\mathcal{N}) = 1 - \alpha, \quad \text{equivalently} \quad \min_{T \subseteq F} \left[(1 - \alpha) \sum_{f \in T} p_f + \sum_{i \in \cup_{f \notin T} C_f} \pi_i \right] \geq 1 - \alpha.$$

Proof. Source capacity is $1 - \alpha$, so $\text{MF}(\mathcal{N}) \leq 1 - \alpha$ with equality iff the minimum cut is at least $1 - \alpha$. By Proposition A.1 this is $(1 - \alpha) \sum_{f \in T} p_f + \pi(\cup_{f \notin T} C_f) \geq 1 - \alpha$ for all T . Writing $1 - \alpha = (1 - \alpha) \sum_{f \in T} p_f + (1 - \alpha) \sum_{f \notin T} p_f$ and cancelling leaves $\pi(\cup_{f \notin T} C_f) \geq (1 - \alpha) \sum_{f \notin T} p_f$, which is (36) with T replaced by $F \setminus T$.

It remains to connect flows to joint distributions. Suppose $\text{MF}(\mathcal{N}) = 1 - \alpha$ and let $y(i, f) \geq 0$ be the flow on $v_f \rightarrow w_i$, so that $\sum_{i \in C_f} y(i, f) = (1 - \alpha)p_f$ for every f (all source edges are saturated) and $\sum_f y(i, f) \leq \pi_i$ for every i (capacity at the sink), with $y(i, f) = 0$ for $i \notin C_f$. Define residual supplies $\rho_i := \pi_i - \sum_f y(i, f) \geq 0$ and residual demands $\sigma_f := p_f - \sum_i y(i, f) = \alpha p_f \geq 0$. Then $\sum_i \rho_i = 1 - (1 - \alpha) = \alpha = \sum_f \sigma_f$, and the transportation problem with supplies ρ , demands σ and no forbidden arcs is feasible: take $r(i, f) = \rho_i \sigma_f / \alpha$. Setting $x(i, f) = y(i, f) + r(i, f)$ gives $\sum_f x(i, f) = \pi_i$, $\sum_i x(i, f) = p_f$, and $\sum_{i \in C_f} x(i, f) \geq \sum_{i \in C_f} y(i, f) = (1 - \alpha)p_f$, so x solves (1)-(4). Conversely, if x solves (1)-(4), scaling the restriction of x to the arcs $\{(i, f) : i \in C_f\}$ down to $y(i, f) = (1 - \alpha)p_f x(i, f) / \sum_{i' \in C_f} x(i', f)$ yields a feasible flow of value $1 - \alpha$.

Standard maximum-flow algorithms run in time polynomial in $|S|$ and $|F|$; the push-relabel method of Goldberg and Tarjan [1988], for example, runs in $O(|V|^2 \sqrt{|E|})$, here $O((|S| + |F|)^2 \sqrt{|S||F|})$. The same network computes the credibility ceiling of Corollary 4.2: κ is the

largest λ for which the network with source capacities λp_f has maximum flow λ , found by binary search or by a single parametric maximum-flow computation, and the minimum cut at $\lambda = \kappa$ returns the subset T responsible for the bound.

B Rationalizability with Feature-Specific Coverage Levels

In the main text, the level $1 - \alpha$ is common across features. In practice it may vary—a test may claim higher sensitivity in symptomatic than in asymptomatic individuals [Dinnes et al., 2022]. Let $\alpha_f \in (0, 1)$ be the miscoverage at f , so $\Pr(i \in C_f \mid f) \geq 1 - \alpha_f$. The system becomes:

$$\sum_{f \in F} x(i, f) = \pi_i \quad \forall i \in S \quad (37)$$

$$\sum_{i \in S} x(i, f) = p_f \quad \forall f \in F \quad (38)$$

$$\sum_{i \in C_f} x(i, f) \geq (1 - \alpha_f) p_f \quad \forall f \in F \quad (39)$$

$$x(i, f) \geq 0 \quad \forall i, f \quad (40)$$

THEOREM B.1 *System (37–40) is feasible if and only if for all $T \subseteq F$:*

$$\sum_{f \in T} (1 - \alpha_f) p_f \leq \sum_{i \in \cup_{f \in T} C_f} \pi_i. \quad (41)$$

When $\alpha_f = \alpha$ for all f , this is Theorem 4.1.

Proof. The argument of Theorem 4.1 applies unchanged. Necessity is the same two-line computation with $(1 - \alpha_f) p_f$ in place of $(1 - \alpha) p_f$. For sufficiency, run the construction of Appendix A on the network $\mathcal{N}$ with the source capacity on $s \rightarrow v_f$ set to $(1 - \alpha_f) p_f$ rather than $(1 - \alpha) p_f$; nothing else in Proposition A.1 or Theorem A.1 depends on the source capacities being proportional to a common level, and the residual demands become $\sigma_f = \alpha_f p_f$. ■

COROLLARY B.1 *A necessary condition is, for each f separately, $(1 - \alpha_f) p_f \leq \pi(C_f)$, the instance $T = \{f\}$ of (41).*

REMARK 9 *Feature-specific levels sharpen the example of Section 4.1. There f_1 and f_2 both certify state 1, and the binding instance $T = \{f_1, f_2\}$ becomes $(1 - \alpha_{f_1}) p_{f_1} + (1 - \alpha_{f_2}) p_{f_2} \leq \pi_1$.*

With $p_{f_1} = p_{f_2} = 0.25$ and $\pi_1 = 0.30$, the analyst may claim $1 - \alpha_{f_1}$ as high as he likes at f_1 only by lowering his claim at f_2 , and the two claims must average at most 0.6. The per-feature conditions of Corollary B.1 see none of this: each alone permits any level up to 1.2.

C Proof of Theorem 5.1

By the Farkas lemma, the system (7–10) is infeasible iff there exist $u_i \geq 0$, $v_f \geq 0$, $y \geq 0$, not all zero, with:

$$-\sum_{i \in S} \pi_i u_i + \sum_{f \in F} p_f v_f + (1 - \alpha)y > 0, \quad (42)$$

$$-u_i + v_f + y \cdot \mathbf{1}[i \in K_f] \leq 0 \quad \forall (i, f). \quad (43)$$

These define a cone. If a solution has $y > 0$, scale so $y = 1$; the $y = 0$ case is handled at the end. Fix $y = 1$. Constraint (43) becomes

$$u_i \geq v_f + \mathbf{1}[i \in K_f] \quad \forall (i, f). \quad (44)$$

We need not bound u_i, v_f : the feasible set is a cone, and we ask only whether the objective can be made positive. It suffices to test the extreme directions $v_f \in \{0, 1\}$, which we index by $A = \{f : v_f = 1\}$, and to take each u_i at its minimum feasible value $u_i = \max_f (\mathbf{1}[f \in A] + \mathbf{1}[i \in K_f])$.

Case $A \neq \emptyset$. Since $K_f \neq \emptyset$, each $f \in A$ has some $i \in K_f$ with $u_i = 2$; every other i has $i \notin K_f$ for some $f \in A$, giving $u_i = 1$. So $u_i = 2$ for $i \in \cup_{f \in A} K_f$ and $u_i = 1$ otherwise. The objective is

$$\begin{aligned} & -\sum_{i \in \cup_{f \in A} K_f} 2\pi_i - \sum_{i \notin \cup_{f \in A} K_f} \pi_i + \sum_{f \in A} p_f + (1 - \alpha) \\ & = -1 - \pi\left(\bigcup_{f \in A} K_f\right) + \sum_{f \in A} p_f + (1 - \alpha). \end{aligned}$$

This is positive iff $\sum_{f \in A} p_f > \alpha + \pi(\cup_{f \in A} K_f)$. Writing $T = F \setminus A$ and using $\sum_f p_f = 1$, this is

$$(1 - \alpha) > \sum_{f \in T} p_f + \pi\left(\bigcup_{f \notin T} K_f\right), \quad (45)$$

the violation of (11) at $T = F \setminus A$. Proper non-empty A gives non-trivial $T \subsetneq F$; $A = F$ gives $T = \emptyset$, that is, $(1 - \alpha) > \pi(\cup_f K_f)$. The case $A = \emptyset$ is redundant: it yields $u_i = 1$ for $i \in \cup_f K_f$ and $u_i = 0$ otherwise, hence the same inequality $(1 - \alpha) > \pi(\cup_f K_f)$ already obtained at $A = F$.

Case $y = 0$. Constraint (43) is $v_f \leq u_i$ for all (i, f) , so $v_f \leq \min_i u_i$ for every f . Hence $\sum_f p_f v_f \leq \min_i u_i \leq \sum_i \pi_i u_i$, since $\sum_i \pi_i = 1$, and the objective is never positive. So $y = 0$ yields no condition. Collecting the cases, the Farkas system is feasible—the primal infeasible—iff (11) fails for some $T \subseteq F$.